

Statistical Analysis of Survival Models Using Feature Quantification on Prostate Cancer Histopathological Images

Jian Ren¹, Eric A. Singer^{2,3}, Evita Sadimin², David J. Foran³, Xin Qi³

¹Department of Electrical and Computer Engineering, Rutgers University, Piscataway, NJ, USA, ²Department of Pathology and Laboratory Medicine, Section of Urologic Oncology, Rutgers Cancer Institute of New Jersey, New Brunswick, NJ, USA, ³Center for Biomedical Imaging and Informatics, Rutgers Cancer Institute of New Jersey, New Brunswick, NJ, USA

Received: 06 November 2018

Accepted: 14 June 2019

Published: 27 September 2019

Abstract

Background: Grading of prostatic adenocarcinoma is based on the Gleason scoring system and the more recently established prognostic grade groups. Typically, prostate cancer grading is performed by pathologists based on the morphology of the tumor on hematoxylin and eosin (H and E) slides. In this study, we investigated the histopathological image features with various survival models and attempted to study their correlations. **Methods:** Three texture methods (speeded-up robust features, histogram of oriented gradient, and local binary pattern) and two convolutional neural network (CNN)-based methods were applied to quantify histopathological image features. Five survival models were assessed on those image features in the context with other prostate clinical prognostic factors, including primary and secondary Gleason patterns, prostate-specific antigen levels, age, and clinical tumor stages. **Results:** Based on statistical comparisons among different image features with survival models, image features from CNN-based method with a recurrent neural network called CNN-long-short-term memory provided the highest hazard ratio of prostate cancer recurrence under Cox regression with an elastic net penalty. **Conclusions:** This approach outperformed the other image quantification methods listed above. Using this approach, patient outcomes were highly correlated with the histopathological image features of the tissue samples. In future studies, we plan to investigate the potential use of this approach for predicting recurrence in a wider range of cancer types.

Keywords: Histopathological image, image features, neural networks, prostate cancer, survival models

INTRODUCTION

Survival analysis is a means for predicting patient outcomes by providing invaluable information for selecting treatment. Predicting prostate cancer survival outcomes is a significant challenge. Following radical prostatectomy, men must be closely monitored for the evidence of recurrence. This is typically done via prostate-specific antigen (PSA) blood tests. A detectable or rising PSA after surgery is the evidence of biochemical recurrence. The measure of time from surgery to biochemical recurrence is biochemical recurrence-free survival (bRFS). Multiple studies examined predictors of bRFS using quantitative histopathological features with some survival models.^[1-4] However, numerous prediction tools^[5-11] utilized whole-slide images (WSIs) to assess prostate cancer recurrence and predicted the likely outcomes resulting from treatments. Few of these studies simultaneously considered clinical factors (primary and

secondary Gleason patterns, PSA value, age, tumor stage) and tissue WSIs to correlate with recurrence under different survival models.

The Gleason scoring system for prostate cancer remains one of the best predictors for prostate cancer progression and recurrence,^[12-15] despite significant interobserver reproducibility among pathologists.^[16-18] A more recently adapted grading system stratifies patients into five prognostic grade groups^[19] based on their Gleason patterns: grade Group 1 (Gleason $\leq 3 + 3 = 6$),

Address for correspondence: Dr. Xin Qi,

Center for Biomedical Imaging and Informatics, Rutgers Cancer Institute of New Jersey, New Brunswick, NJ, USA.
E-mail: qixi@rutgers.edu

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

For reprints contact: reprints@medknow.com

How to cite this article: Ren J, Singer EA, Sadimin E, Foran DJ, Qi X. Statistical analysis of survival models using feature quantification on prostate cancer histopathological images. *J Pathol Inform* 2019;10:30.

Available FREE in open access from: <http://www.jpathinformatics.org/text.asp?2019/10/1/30/268079>

Access this article online

Quick Response Code:



Website:
www.jpathinformatics.org

DOI:
10.4103/jpi.jpi_85_18

Grade Group 2 (Gleason 3 + 4 = 7), Grade Group 3 (Gleason 4 + 3 = 7), Grade Group 4 (Gleason 4 + 4 = 8, 3 + 5 = 8, and 5 + 3 = 8), and Grade Group 5 (Gleason 4 + 5 = 9, 5 + 4 = 9, and 5 + 5 = 10). Figure 1 shows an example of Giga-pixel WSI with different Gleason patterns. The green-framed patch contains Gleason pattern 3; the blue-framed patch contains Gleason pattern 4; and the red-framed patch contains Gleason pattern 5. In this study, we conducted experiments on public prostate cancer dataset using different feature quantification methods and recurrence analysis using different survival models. Histopathological image features were quantified through texture methods and neural network-based approaches. We focused on the prostate cancer grade groups of 1–4. The bRFS was applied as the time to recurrence for prostate cancer progression analysis.

MATERIALS AND METHODS

Materials

In this study, we used the prostate dataset from the Genomic Data Commons (GDC).^[20] The dataset included whole-slide histopathological images from patients and their corresponding clinical reports, including the primary and secondary Gleason pattern, patients’ PSA value, age, and tumor stage. All the image data, annotations of Gleason score, and clinical information were publicly available.

We selected the patients with low-risk (Gleason score 3 + 3), intermediate-risk (Gleason score 3 + 4 or 4 + 3), and high-risk prostate cancer (Gleason score 4 + 4) because those patient populations show a reasonable range of prognoses for our analysis. We excluded patients with Gleason Grade Group 5 patients in this study due to poor prognosis of their disease.^[21] Considering the high computational cost on the Giga-pixel tissue WSIs, existing WSIs classification and recurrence analysis approaches were focused on effectively utilizing the cropped patches from region of interests.^[22-27] For image preparation, we adopted the two-step cropping–selecting

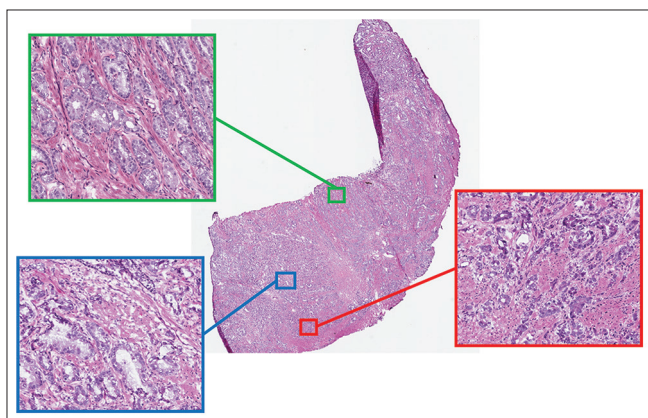


Figure 1: Example Giga-pixel whole-slide image with different Gleason patterns. The green framed patch contains Gleason pattern 3; the blue-framed patch contains Gleason pattern 4; and the red-framed patch contains Gleason pattern 5

process. First, original patches were automatically generated within each WSI under $\times 40$ with a patch size of 4096×4096 . Second, the patches with the tissue accounting for at least 20% of the whole area were selected for our experiments. The number of WSIs and cropped patches under different Gleason scores is shown in Table 1.

Methods

Initially, we utilized various quantification methods to extract image features from WSIs. Next, the recurrence analysis was performed on the combination of image features and clinical factors utilizing different survival models, as shown in Figure 2. Hazard ratios using different survival models were calculated to indicate the correlation between image features (or in context of clinical factors) and recurrence; the higher the hazard ratio, the higher the correlations.

Image feature quantification

We adopted five approaches for the purpose of feature quantification including unsupervised and supervised methods. The unsupervised texture methods consisted of speeded-up robust features (SURFs),^[28] histogram of oriented gradients (HOGs),^[29] and local binary pattern (LBP).^[30] The two supervised methods are based on convolutional neural networks (CNNs). For supervised methods, we randomly selected 20% of the cases as testing set, 10% as validation set, and the remaining as training set.

Texture features

We chose three texture methods for prostate cancer histopathological image analysis. They were rotation, translation, and scale- and intensity-invariant which were suitable for descriptions of the texture features within WSIs.

The SURF^[28] is partly inspired by the scale-invariant feature transform (SIFT) descriptors. The standard version of SURF is several times faster than SIFT and more robust against different image transformations than SIFT. The image is transformed into coordinates, using the multiresolution pyramid technique, to copy the original image with a pyramidal Gaussian or Laplacian pyramid shape to obtain an image with the same size but with reduced bandwidth. The HOG^[29] counts occurrences of gradient orientation in a local region of an image. It is similar to that of edge-orientation histograms, SIFT descriptors, and shape contexts but differs in that it is computed on a dense grid of uniformly spaced cells and uses overlapping local contrast normalization for improved accuracy. The LBP^[30] is used to

Table 1: The number of whole-slide images and their corresponding automatically selected patches under different Gleason scores composing from a sum of Gleason patterns 3+3, 3+4, 4+3, and 4+4 prostate prognostic grading groups

Gleason score	3+3	3+4	4+3	4+4
# WSIs	43	144	99	49
# patches	1229	4753	2997	1597

WSIs: Whole-slide images

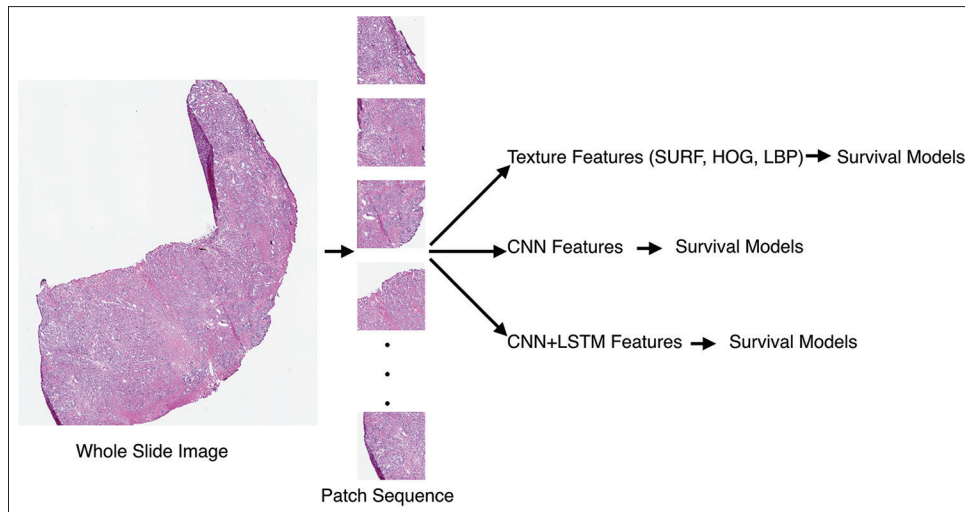


Figure 2: Outline of image feature quantification from whole-slide images and assessed by various survival models

model the image local features in texture spectrum units in a multiresolution gray-scale mode. It is based on recognizing local binary unit patterns for any quantization of the angular space and spatial resolution.

The image features for each patch were generated using a bag-of-words approach^[31] from the texture features of different texture methods. By treating image features as words, a bag of words is a sparse vector of occurrence counts (histogram) of a vocabulary of local image features. In the bag-of-word approach, it converts vector-represented texture features to *codewords*, which also produce a *codebook*. The image features are mapped to certain *codewords* through the clustering process, and the image is then represented by the histogram of the *codewords*. Empirically, we use 100 as the number of cluster centers to report the best performance for texture features. To select the texture features for WSIs, we apply principal component analysis (PCA)^[32] of the image features for all patches within a WSI due to correlations among the patches.

Convolutional neural network-based features

In recent years, with the advances of deep learning, studies using CNNs have demonstrated significant improvement on histopathological image classification^[27,33-36] and segmentation.^[33,34,37,38] For the WSIs, applications based on CNNs have been widely developed.^[39-41] In our study, we adopted two approaches to obtain CNN-based features. The first one was using the neural network to obtain image features for each patch, and then the features for WSIs were obtained by utilizing PCA on all patches. The CNN employed in the study is shown in Table 2. The input to the network was the cropped patches from prostate pathological WSIs. The activations from the second to the last layer were considered as the image features of the input samples. To train the network with patches, we assigned Gleason pattern as the ground truth annotation for the patch. The GDC WSIs have been previously graded with the primary and secondary patterns, as well as the final Gleason score given.

Table 2: The convolutional neural network applied in our approach

Layer	Filter size, stride	Output W × H × N
Input	-	256×256×3
Conv	11×11, 4	55×55×96
Max-pooling	3×3, 2	27×27×96
Conv	5×5, 1	27×27×256
Max-pooling	3×3, 2	13×13×256
Conv	3×3, 1	13×13×384
Conv	3×3, 1	13×13×384
Conv	3×3, 1	13×13×256
Max-pooling	3×3, 2	6×6 × 256
FC6	-	4096
FC7	-	4096
FC8, FC9	-	2, 4

All the Conv are followed by ReLU. For the FC, the FC6 and FC7 are followed by the ReLU and dropout layer with the dropout ratio as 0.5; FC8 and FC9 are both at the top of FC7. Conv: Convolution layers, ReLU: Rectified linear units, FC: Fully connected layers, W×H×N: Height×Width×Channel)

To model variations among Gleason patterns within a WSI, we used the multitask architecture to enable the network to learn as much information about the Gleason patterns from the patches of a WSI as possible. During the training process, we assigned the primary pattern and the sum of primary pattern and secondary pattern (Gleason score) as labels for each patch and use the following multitask loss function:

$$L_{\text{multi-task}} = -\sum_{i=0}^N t_i^p \log(\hat{t}_i^p) - \sum_{i=0}^N t_i^s \log(\hat{t}_i^s) \quad (1)$$

where for the i^{th} image within the batch of N images, t_i^p and t_i^s encoded the Gleason grading for the primary pattern and the sum score and \hat{t}_i^p and \hat{t}_i^s encoded the predicted grading of the model. The one-hot encoding is a process by which categorical variables are converted into a form that could be provided to CNN to do a better job in classification. The results suggested

that using the primary Gleason pattern and the Gleason score together achieved the best estimate of risk of recurrence by capturing local and global image feature distribution more efficiently than using either one alone.

For the second approach, we treated the cropped patches from the WSI as an image sequence and used one type of recurrent neural network (RNN) called long-short-term memory (LSTM) to explore the long-term dynamic information of the patch spatial sequence within the WSI. We denoted the method as CNN features with LSTM (CNN + LSTM). The LSTM could fully leverage the patch spatial sequence within a WSI to get the representative features that model the global Gleason score of the WSI and the distribution of the Gleason patterns among the WSI. Recently, the LSTM model has been successfully used in speech recognition,^[42,43] language translation models,^[44] image captioning,^[45] and video classification.^[46] Compared with the traditional RNNs, LSTM is more effectively in long-range and short-term spatial sequence modeling. In general, given an input feature sequence (x_1, x_2, \dots, x_T) , the LSTM outputs the output sequence (y_1, y_2, \dots, y_T) . The hidden layer of LSTM is computed recursively from $t = 1$ to $t = T$ with the following equations:

$$i_t = \sigma(W_{x_i} x_t + W_{h_i} h_{t-1} + W_{c_i} c_{t-1} + b_i) \quad (2)$$

$$f_t = \sigma(W_{x_f} x_t + W_{h_f} h_{t-1} + W_{c_f} c_{t-1} + b_f) \quad (3)$$

$$c_t = f_t c_{t-1} + i_t \tanh(W_{x_c} x_t + W_{h_c} h_{t-1} + b_c) \quad (4)$$

$$o_t = \sigma(W_{x_o} x_t + W_{h_o} h_{t-1} + W_{c_o} c_{t-1} + b_o) \quad (5)$$

$$h_t = o_t \tanh(c_t) \quad (6)$$

where x_i is the network activations of the i^{th} patch, h_t is the hidden vector, i_t , c_t , f_t , and o_t are, respectively, the activation vectors of the input gate, memory cell, forget gate, and output gate. W terms denoted the weight matrices connecting different units, b terms denoted the bias vectors, and σ is the logistic sigmoid function. From the above equations, we can see the memory cell c_t in LSTM having two inputs: the weighted sum of the current inputs and the previous memory cell units c_{t-1} , which enables the model to learn when to forget the old information and when to consider new information. The output gate o_t controls the propagation of information to the following step.

Since we utilized the spatial characteristic encoded features from CNN, the training process of LSTM of patches within WSIs was formed in a spatial format instead of time sequential manner. As shown in Figure 3, we used the image coordinates to indicate the location of each patch in the patch spatial sequence. In this way, we considered both the unique characteristics of each patch and the fine-grained variations between patches. For one prostate WSI, the patches were fed into the network to get the activations from the second to the last layer. Then, we utilized a one-layer LSTM to recursively

map the activations of each patch to a feature vector. In addition, the average pooling layer was applied on top of the network to get a feature vector as the computational image features for the WSI. The number of hidden units for each LSTM is 1024. During the training process, we applied the multitask loss and assign the primary pattern and the Gleason score for the WSIs.

Survival models

To evaluate the performance of various survival models using different image features quantified by textural and CNN-based methods on patients with prostate cancer, we used the bRFS since their initial treatment as a time-to-recurrence variable for survival models. Using survival models, we assessed the image features related to recurrence hazard risk scores in the context of other clinical prognostic factors, including the primary and the secondary Gleason patterns, PSA, age, and clinical tumor stage.

The hazard risk scores of image features in the context of clinical mean a measure of prostate cancer recurrence risk ratio, commonly in time-to-event analysis or survival analysis. The survival models evaluated in our study include multivariate Cox proportional-hazards model,^[47] Cox regression by an elastic net penalty (COX-EN),^[48] parametric proportional-hazard model (PH-EX),^[49] parametric proportional-hazard model with log-normal distance (PH-LogN),^[49] and parametric proportional-hazard model with log-logistic distance (PH-LogL).^[49]

For the high-dimensional data, univariate Cox regression was applied to the computational image features. Only those with Wald test, $P < 0.05$ is selected in conjunction with clinical factors as inputs of the survival models. The Cox proportional-hazards model is a popular regression model for the analysis of survival data. It is a semi-parametric method for adjusting survival rate estimates to quantify the effect of predictor variables. In contrast with parametric models, it makes no assumptions about the shape of the so-called baseline hazard function. It represents the effects of explanatory variables as a multiplier of a common baseline hazard function H_0 . Given the patients (t_i, l_i, x_i) , where $i = 1, 2, \dots, N$, we have the t_i as the patient's recurrence time for individual i ; l_i is the label of the censored data that equals 1 if the recurrence occurred at that time and 0 if the patient has been censored; and X_i as the vector of covariates of the selected image features and clinical factors.

The hazard function is the nonparametric part of the Cox proportional-hazards regression function corresponding to

$$H(X_i, l_i, t_i) = H_0(t) \exp \sum_{j=1}^p x_{ij} \beta_j \quad (7)$$

Here, x_{ij} is the image features j for patient i , where $j = 1, 2, \dots, p$ and β_j is the Cox regression parameter for each patient.

The hazard ratio is derived from $(X_i) = \frac{H(X_i, l_i, t)}{H_0}$,

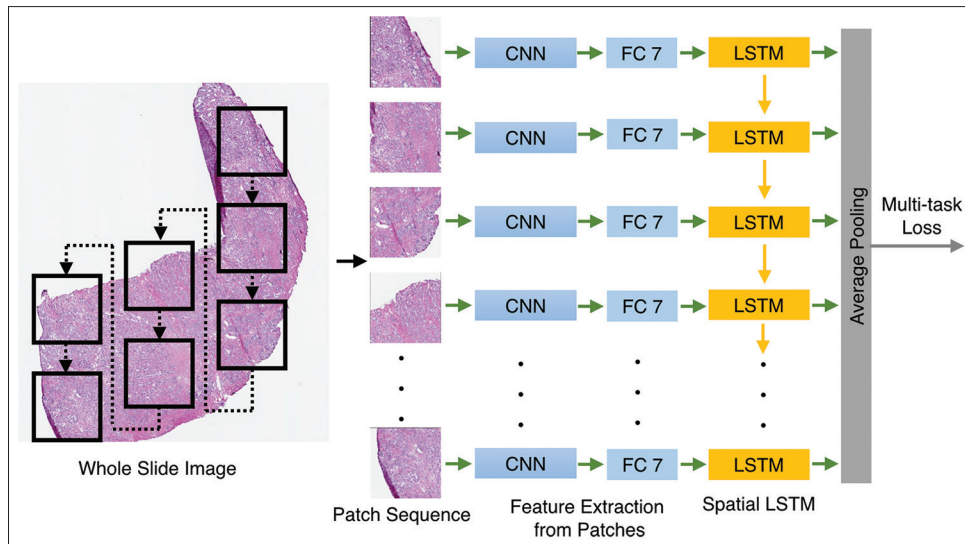


Figure 3: The multi-task neural network architecture for computational image features extraction from whole-slide images. The cropped patches are formed as a sequence by the image coordinates. The long-short-term memory is built on top of the convolutional neural network for the long-term spatial modeling of the activation sequence. An average pooling layer maps the activations into one feature vector

representing the relative risk of instant failure for patients having the predictive value X_i compared to the ones having the baseline values. Here, d_i is weighting parameters for each patient.

$$HR(X_i) = \sum_i^N d_i (X_i \beta_i - \log(\sum_i^p I[t_j - t_i] \exp[X_i \beta_i])) \quad (8)$$

For the COX-EN, the elastic net penalty $\hat{\beta}$ is given in the equation below. It is a mixture of the L1 (Lasso) and L2 (ridge regression) penalty. Here, is the ratio between L1 and L2 for elastic net.

$$\hat{\beta} = \operatorname{argmax}_{\beta} \left[\frac{2}{n} \left(\sum_{i=1}^N x_i^T \beta_i - \log \left(\sum_{j \in R_k} e^{x_j^T \beta_i} \right) \right) - \lambda P_{\alpha}^{\beta_i} \right] \quad (9)$$

where

$$\lambda P_{\alpha}^{\beta_i} = \lambda \left[\alpha \left(\sum_i^p |\beta_i| + \frac{1}{2} (1 - \alpha) \sum_i^p \beta_i^2 \right) \right] \quad (10)$$

Based on the assumption that the effect of the covariates is to increase or decrease the hazard by a proportionate amount at all durations, the parametric proportional-hazard model is a location-scale model for arbitrary transform of the time variable t_j , leading to accelerated failure time model with different penalty distance functions. The distance functions we use for parametric proportional-hazard models are exponential transformation (PH-EX), log-normal (PH-LogN), and log-logistic (PH-LogL) distances.

The survival model fitting to different image features were quantified by Akaike information criteria (AIC).^[50]

$$AIC = -2 \log(\text{likelihood}) + 2K \quad (11)$$

where *likelihood* is a measure-of-model fitness and K represents the number of model parameters. The smaller value of the AIC, the better the goodness of fit of the survival models.

EXPERIMENTAL RESULTS

In this section, we conducted the experiments on the public prostate cancer dataset to make statistical analysis on various survival models using different histopathological image feature quantification methods.

Implementation details

For the CNN-based approaches to extract image features, we first used the patches to train the CNN with multitask loss. Each patch was resized as 256×256 and assigned two labels according to the Gleason grading of the WSI: one being the primary pattern and another being the Gleason score. The CNN was trained with mini-batch stochastic gradient descent. The momentum is 0.9, and weight decay was 5×10^{-5} . The initial learning rate is 10^{-3} and annealed by 0.1 after 10^4 iterations. To train the LSTM, we set the same momentum, the weight decay, and the initial learning rate. The learning rate is annealed by 0.1 after 2×10^3 iterations. The implementation is based on the Caffe toolbox.^[51]

Comparison of image features

First, only using image features from tissue specimens, including clinical Gleason primary and secondary patterns and the quantified image features from various image methods, their Cox hazard ratios are shown in Table 3. CNN achieved better results than texture methods, including SURF,^[28] HOG,^[29] and LBP.^[30] Using CNN with LSTM to model the spatial relation of patches achieved the highest Cox hazard ratio, which indicated the best recurrence correlation for prostate cancer patients' recurrence data. On the other hand, the image features obtained from texture-based methods and CNN

approaches achieved higher Cox hazard ratios as compared to utilizing primary and secondary patterns alone.

Second, in addition to the image features, PSA levels, ages, and clinical tumor stages were included in the Cox survival model, besides the primary and the secondary Gleason patterns. The results of combining clinical factors and image features are shown in Table 4, demonstrating that the image features generated from CNN-based approaches were more representative than the texture features by having higher values of hazard ratio. In addition, those features were more representative than clinical prognostic factors. We also calculated the AIC values, as shown in Table 4. The smaller AIC value encodes the better goodness of fit of the survival model. CNN + LSTM achieved the best fitness on the Cox regression model compared to other image features quantification methods.

Finally, without any image features, we showed the Cox hazard ratios of the clinical factors, as shown in Table 5. From the results of Tables 3-5, we can see that primary Gleason patterns have higher Cox hazard ratios than the ones of other clinical factors, which was consistent with its high prediction power for prostate cancers.^[1,4]

Table 3: The Cox hazard ratios of only using clinical Gleason primary and secondary patterns and image features from different image analysis methods

Methods	Primary pattern	Secondary pattern	Image features
SURF	0.76	0.58	1.15
HOG	0.84	0.55	1.09
LBP	0.77	0.60	1.10
CNN	0.80	0.73	1.83
CNN + LSTM	0.90	0.71	3.54

The texture feature quantification methods include SURF,^[28] HOG,^[29] and LBP.^[30] Using CNN with LSTM to model the spatial relationship of patches achieves the highest Cox hazard ratio, which indicates the best performance on progression prediction for the recurrence data. Meanwhile the image features from texture and CNN approaches achieve the Cox hazard ratios compared to the ones from clinical Gleason primary and secondary patterns. SURF: Speeded-up robust features, HOG: Histogram of oriented gradient, CNN: Convolutional neural network, LSTM: Long-short-term memory, LBP: Local binary pattern

Table 4: The Cox hazard ratios and Akaike information criteria of using clinical factors including Gleason primary and secondary patterns, patients' prostate-specific antigen, age, and clinical tumor stages, and image features from different image analysis methods

Methods	Primary pattern	Secondary pattern	PSA	Age	Tumor stage	Image features	AIC
SURF	0.99	0.67	0.84	0.98	1.04	1.13	38.93
HOG	1.21	0.65	0.82	1.01	1.13	1.10	51.97
LBP	0.97	0.76	0.84	1.00	1.08	1.08	35.97
CNN	1.10	1.13	0.80	1.00	1.17	2.58	38.02
CNN + LSTM	1.38	0.75	0.76	0.97	1.14	7.10	35.60

The texture feature quantification methods include SURF,^[28] HOG,^[29] and LBP.^[30] Using CNN + LSTM achieves the highest Cox hazard ratio and lowest value of AIC, which indicates the best performance on progression prediction for the recurrence data. PSA: Prostate-specific antigen, AIC: Akaike information criteria, SURF: Speeded-up robust features, HOG: Histogram of oriented gradient, LBP: Local binary pattern, LSTM: Long-short-term memory

Ablation study on training strategies

Furthermore, considering the multiple Gleason patterns within WSIs, we designed two training strategies to train the CNN-based approaches. The first one was to use multitask loss to learn both the primary Gleason pattern and the sum of the primary and secondary patterns (namely, the Gleason score). The second one was to use the primary Gleason pattern or the Gleason score alone to learn the patterns within the patches or WSIs.

The performance of two CNN-based approaches on patient recurrence analysis was compared using different training strategies. The results are shown in Table 6. We can see that the multitask architecture achieved better correlation with patients' recurrence than training label using the primary Gleason pattern or Gleason score alone as it has much higher recurrence hazard ratios and lower AIC values. This is because the primary Gleason pattern and the Gleason score together could better reflect the local and global image features in the WSIs than use each alone.

Comparison of survival models

In this section, we performed statistical analysis on various survival models, including COX-EN,^[48] PH-EX,^[50] PH-LogN,^[50] and PH-LogL,^[50] using prostate images with Gleason score 6–8 and clinical factors. The Cox proportional-hazards model does not need an assumption of a particular survival distribution of the patients' survival data. The only assumption in the model is about the proportional hazards. Unlike the Cox proportional-hazards model, parametric models with different penalty distance functions (such as exponential, log-normal, and log-logistic) need to specify the hazard functions.^[52,53] Studies have indicated that under certain circumstances, such as strong effect or strong time trend in covariates or follow-up depending on covariates, the parametric models are good alternatives to the Cox regression model.^[53]

We assessed different survival models and show the hazard ratios of image features and patients' clinical prognostic factors, as shown in Table 7. Based on these results, first, we can see that the image features quantified from WSIs outperformed other clinical factors in all texture and CNN-based approaches. Second, CNN-based approaches achieved a better correlation

with patients' recurrence due to their higher hazard ratios than other texture methods for all survival models. Third, by comparing with Table 4, COX-EN achieved the lowest AIC value with image features obtained from CNN + LSTM, proving that the model was more suitable for recurrence analysis for prostate patients with low, intermediate, and high risk than other survival models.

DISCUSSION AND CONCLUSIONS

Table 5: The Cox hazard ratios of the clinical factors

Primary pattern	Secondary pattern	PSA	Age	Tumor stage
2.15	1.09	0.73	0.90	1.30

PSA: Prostate-specific antigen

In this paper, we presented three unsupervised texture methods (SURF, HOG, and LBP) and two supervised CNN-based methods to quantify the features from histopathological images. Five survival models were assessed on those image features along with prostate cancer clinical prognostic factors, including the primary and the secondary Gleason patterns, PSA, age, and clinical tumor stage to perform bPFS analyses.

Based on the statistical comparisons among different image feature quantification methods with survival models, the CNN-LSTM provided the highest hazard ratio of prostate cancer recurrence under COX-EN. COX-EN outperforms other image quantification methods with other survival models, respectively. In our approach, patient outcomes were better correlated with their histopathological image features. Due

Table 6: The Cox hazard ratios and Akaike information criteria of convolutional neural network-based approaches on patients' progression analysis using three different training strategies

Methods	Training Strategy	Primary Pattern	Secondary Pattern	PSA	Age	Tumor Stage	Image Features	AIC
CNN	Primary Pattern	1.11	1.12	0.80	1.00	1.16	1.34	46.13
CNN	Gleason Score	1.26	1.03	0.75	0.98	1.12	1.53	44.29
CNN	Multi-task	1.10	1.13	0.80	1.00	1.17	2.58	38.02
CNN + LSTM	Primary Pattern	1.35	0.84	0.78	0.98	1.14	1.63	44.27
CNN + LSTM	Gleason Score	1.09	0.66	0.81	0.99	1.11	2.76	41.47
CNN + LSTM	Multi-task	1.38	0.75	0.76	0.97	1.14	7.10	35.60

Using multitask architecture achieves the highest Cox hazard ratio and lowest AIC values than training using the primary Gleason pattern or Gleason score alone, which indicates the best performance on progression prediction for the recurrence data. CNN: Convolutional neural network, LSTM: Long-short-term memory, PSA: Prostate-specific antigen, AIC: Akaike information criteria

Table 7: The Cox hazard ratios and Akaike information criteria of different survival models using texture methods and convolutional neural network-based approaches

Survival models	Methods	Primary patterns	Secondary patterns	PSA	Age	Tumor stage	Image features	AIC
COX-EN	SURF	0.10	0.27	0.33	0.06	0.03	3.38	42.93
COX-EN	HOG	0.10	0.25	0.32	0.06	0.03	3.85	59.72
COX-EN	LBP	0.10	0.19	0.30	0.06	0.03	2.40	39.83
COX-EN	CNN	0.23	0.21	0.33	0.06	0.04	7.57	29.86
COX-EN	CNN + LSTM	0.13	0.27	0.36	0.06	0.03	15.85	29.83
PH-EX	SURF	0.07	0.09	0.29	0.03	0.03	1.94	41.26
PH-EX	HOG	0.05	0.12	0.29	0.04	0.03	2.41	61.56
PH-EX	LBP	0.07	0.06	0.28	0.03	0.03	1.49	41.22
PH-EX	CNN	0.08	0.07	0.29	0.04	0.04	4.50	35.60
PH-EX	CNN + LSTM	0.08	0.10	0.29	0.04	0.03	10.22	31.22
PH-LogN	SURF	0.18	0.22	0.30	0.02	0.08	2.03	47.27
PH-LogN	HOG	0.18	0.23	0.30	0.02	0.08	2.70	47.58
PH-LogN	LBP	0.21	0.18	0.29	0.02	0.08	1.38	45.99
PH-LogN	CNN	0.16	0.15	0.30	0.02	0.08	4.33	42.51
PH-LogN	CNN + LSTM	0.20	0.18	0.31	0.02	0.08	11.92	33.31
PH-LogL	SURF	0.11	0.15	0.29	0.02	1.89	1.89	43.74
PH-LogL	HOG	0.07	0.20	0.28	0.02	2.91	2.91	44.45
PH-LogL	LBP	0.79	0.29	1.09	0.77	1.46	1.46	44.39
PH-LogL	CNN	0.09	0.08	0.29	0.03	4.39	4.39	35.96
PH-LogL	CNN + LSTM	0.12	0.13	0.29	0.02	9.92	9.92	33.02

The survival models include COX-EN,^[48] PH-EN,^[50] PH-LogN,^[50] and PH-LogL.^[50] CNN: Convolutional neural network, LSTM: Long-short-term memory, PSA: Prostate-specific antigen, AIC: Akaike information criteria, SURF: Speeded-up robust features, HOG: Histogram of oriented gradient, LBP: Local binary pattern

to the limited size of the public prostate dataset, the results achieved from our experiments were preliminary. To further validate its generalizability of our approach, more prostate images from local institutions are needed to perform extensive experiments.

In the future, besides using tissue WSIs for patients' bRFS analysis, integrating patients' genomic information and tissue histopathology images will be investigated as a means for providing additional predictive power. Doing so would provide a more quantitative and accurate clinical decision-making support system for patients with prostate cancer.

Financial support and sponsorship

This research was funded, in part, by grants from NIH contracts 4R01LM009239-08, 4R01CA161375-05, 1UG3CA225021-01, and P30CA072720.

Conflicts of interest

Dr. Singer is the principal investigator on an investigator-initiated clinical trial that is funded by Astellas/Medivation (NCT02885649) (<http://clinj.org/clinical-trials/index?show=trial&p=081604>). The other authors declare that they have no competing interests.

REFERENCES

- Madabhushi A, Agner S, Basavanahally A, Doyle S, Lee G. Computer-aided prognosis: Predicting patient and disease outcome via quantitative fusion of multi-scale, multi-modal data. *Comput Med Imaging Graph* 2011;35:506-14.
- Lee G, Singanamalli A, Wang H, Feldman MD, Master SR, Shih NN, *et al*. Supervised multi-view canonical correlation analysis (SMVCCA): Integrating histologic and proteomic features for predicting recurrent prostate cancer. *IEEE Trans Med Imaging* 2015;34:284-97.
- Lee G, Veltri RW, Zhu G, Ali S, Epstein JI, Madabhushi A. Nuclear shape and architecture in benign fields predict biochemical recurrence in prostate cancer patients following radical prostatectomy: Preliminary findings. *Eur Urol Focus* 2017;3:457-66.
- Leo P, Lee G, Shih NN, Elliott R, Feldman MD, Madabhushi A. Evaluating stability of histomorphometric features across scanner and staining variations: Prostate cancer diagnosis from whole slide images. *J Med Imaging (Bellingham)* 2016;3:047502.
- Kattan MW, Eastham JA, Stapleton AM, Wheeler TM, Scardino PT. A preoperative nomogram for disease recurrence following radical prostatectomy for prostate cancer. *J Natl Cancer Inst* 1998;90:766-71.
- Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, *et al*. Assessing the performance of prediction models: A framework for traditional and novel measures. *Epidemiology* 2010;21:128-38.
- Hull GW, Rabbani F, Abbas F, Wheeler TM, Kattan MW, Scardino PT. Cancer control with radical prostatectomy alone in 1,000 consecutive patients. *J Urol* 2002;167:528-34.
- Kattan MW, Wheeler TM, Scardino PT. Postoperative nomogram for disease recurrence after radical prostatectomy for prostate cancer. *J Clin Oncol* 1999;17:1499-507.
- Cooperberg MR, Broering JM, Carroll PR. Time trends and local variation in primary treatment of localized prostate cancer. *J Clin Oncol* 2010;28:1117-23.
- Ren J, Sadimin ET, Wang D, Epstein JI, Foran DJ, Qi X. Computer aided analysis of prostate histopathology images Gleason grading especially for Gleason score 7. *Conf Proc IEEE Eng Med Biol Soc* 2015;2015:3013-6.
- Ren J, Sadimin E, Foran DJ, Qi X. Computer aided analysis of prostate histopathology images to support a refined Gleason grading system. *Proc SPIE Int Soc Opt Eng* 2017. pii: 101331V.
- Egevad L, Granfors T, Karlberg L, Bergh A, Stattin P. Prognostic value of the Gleason score in prostate cancer. *BJU Int* 2002;89:538-42.
- Gleason DF, Mellinger GT. Prediction of prognosis for prostatic adenocarcinoma by combined histological grading and clinical staging. *J Urol* 1974;111:58-64.
- Epstein JI, Partin AW, Sauvageot J, Walsh PC. Prediction of progression following radical prostatectomy. A multivariate analysis of 721 men with long-term follow-up. *Am J Surg Pathol* 1996;20:286-92.
- Billis A, Guimaraes MS, Freitas LL, Meirelles L, Magna LA, Ferreira U. The impact of the 2005 international society of urological pathology consensus conference on standard Gleason grading of prostatic carcinoma in needle biopsies. *J Urol* 2008;180:548-52.
- Allsbrook WC Jr., Mangold KA, Johnson MH, Lane RB, Lane CG, Amin MB, *et al*. Interobserver reproducibility of Gleason grading of prostatic carcinoma: Urologic pathologists. *Hum Pathol* 2001;32:74-80.
- Allsbrook WC Jr., Mangold KA, Johnson MH, Lane RB, Lane CG, Epstein JI. Interobserver reproducibility of Gleason grading of prostatic carcinoma: General pathologist. *Hum Pathol* 2001;32:81-8.
- Glaessgen A, Hamberg H, Pihl CG, Sundelin B, Nilsson B, Egevad L. Interobserver reproducibility of modified Gleason score in radical prostatectomy specimens. *Virchows Arch* 2004;445:17-21.
- Pierorazio PM, Walsh PC, Partin AW, Epstein JI. Prognostic Gleason grade grouping: Data based on the modified Gleason scoring system. *BJU Int* 2013;111:753-60.
- Kandath C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, *et al*. Mutational landscape and significance across 12 major cancer types. *Nature* 2013;502:333-9.
- Makino T, Miwa S, Koshida K. Impact of Gleason pattern 5 on outcomes of patients with prostate cancer and iodine-125 prostate brachytherapy. *Prostate Int* 2016;4:152-5.
- Wang H, Xing F, Su H, Stromberg A, Yang L. Novel image markers for non-small cell lung cancer classification and survival prediction. *BMC Bioinformatics* 2014;15:310.
- Yao J, Wang S, Zhu X, Huang J. Imaging Biomarker Discovery for Lung Cancer Survival Prediction. In: Ourselin S, Joskowicz L., Sabuncu M., Unal G., Wells W. (eds) *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*. MICCAI 2016. Springer, Cham: Lecture Notes in Computer Science 2016;9901:649-57.
- Yu KH, Zhang C, Berry GJ, Altman RB, Ré C, Rubin DL, *et al*. Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. *Nat Commun* 2016;7:12474.
- Zhu X, Yao J, Huang J. Deep convolutional neural network for survival analysis with pathological images. In: *IEEE International Conference on Bioinformatics and Biomedicine. BIBM*; 2016. p. 544-7.
- Zhu X, Yao J, Luo X, Xiao G, Xie Y, Gazdar A, *et al*. Lung Cancer Survival Prediction from Pathological Images and Genetic Data – An Integration Study. *IEEE 13th International Symposium on Biomedical Imaging*; 2016. p. 1173-6.
- Hou L, Samaras D, Kurc TM, Gao Y, Davis JE, Saltz JH, *et al*. Patch-based convolutional neural network for whole slide tissue image classification. *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit* 2016;2016:2424-33.
- Bay H. SURF: Speeded up robust features. *Comput Vis Image Underst* 2008;110:346-59.
- Dalal N, Triggs B. Histograms of Oriented Gradients for Human Detection. *CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition*; 2005. p. 886-93.
- Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans Pattern Anal Mach Intell* 2002;24:971-87.
- Fei L, Perona PA. Bayesian hierarchical model for learning natural scene categories. In: *CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2005. p. 524-31.
- Wold S, Esbensen K, Geladi P. Principal component analysis. *Chemometr Intell Lab Syst* 1987;2:37-52.
- Ing N, Ma Z, Li J, Salemi H, Arnold C, Kundsens BS, *et al*. Semantic segmentation for prostate cancer grading by convolutional neural networks. In: *Proceedings of Medical Imaging; Digital Pathology 2018*. p. 105811B.
- Xu J, Luo X, Wang G, Gilmore H, Madabhushi A. A deep convolutional neural network for segmenting and classifying epithelial and stromal

- regions in histopathological images. *Neurocomputing* 2016;191:214-23.
35. Hou L, Singh K, Samaras D, Kurc TM, Gao Y, Seidman RJ, *et al*. Automatic histopathology image analysis with CNNs, Proceedings of 2016 New York Scientific Data Summit (NYSDS), New York 2016. p. 1-6
 36. Cruz-Roa A, Basavanthally A, Gonzalez F, Gilmore H, Fledman M, Ganesan S, *et al*. Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks. In: Proceedings of SPIE Medical Imaging: International Society for Optics and Photonics; 2014. p. 904103-1.
 37. Pan X, Li L, Yang H, Liu Z, Zhao L, Fan Y. Accurate segmentation of nuclei in pathological images via sparse reconstruction and deep convolutional networks. *Neurocomputing* 2017;229:88-99.
 38. Naik S, Doyle S, Agner S, Madabhushi A, Feldman M, Tomaszewski J. Automated gland and nuclei segmentation for grading of prostate and breast cancer histopathology. In: ISBI: 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro. Paris, France; 2008. p. 284-7.
 39. Kothari S, Phan JH, Stokes TH, Wang MD. Pathology imaging informatics for quantitative analysis of whole-slide images. *J Am Med Inform Assoc* 2013;20:1099-108.
 40. Roullier V, Lézoray O, Ta VT, Elmoataz A. Multi-resolution graph-based analysis of histopathological whole slide images: Application to mitotic cell extraction and visualization. *Comput Med Imaging Graph* 2011;35:603-15.
 41. Toth R, Shih N, Tomaszewski JE, Feldman MD, Kutter O, Yu DN, *et al*. Histostitcher: An informatics software platform for reconstructing whole-mount prostate histology using the extensible imaging platform framework. *J Pathol Inform* 2014; 5-8.
 42. Graves A, Mohamed AR, Hinton GE. Speech recognition with deep recurrent neural networks. In: ICASSP 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. 2013. p.6645- 9.
 43. Graves A, Jaitly N. Towards end- to- end speech recognition with recurrent neural networks. In: Proceedings of the 31st International Conference on Machine Learning (ICML-14). 2014. p. 1764-72s.
 44. Sutskever I, Vinyals O, Le QV. Sequence to sequence learning with neural networks. In: Advances in Neural Information Processing Systems 27 (NIPS) 2014. p. 3104-12
 45. Donahue J, Hendricks LA, Rohrbach M, Venugopalan S, Guadarrama S, Saenko K, *et al*. Long-term recurrent convolutional networks for visual recognition and description. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2017;39:677-91.
 46. Wu Z, Wang X, Jiang YG, Ye H, Xue X. Modeling spatial- temporal clues in a hybrid deep learning framework for video classification. In: Proceedings of the 23rd ACM International Conference on Multimedia. Brisbane, Australia 2015. p. 461- 70.
 47. Therneau TM, Grambsch P. Modeling Survival Data: Extending the Cox Model, Chapter of estimating the survival and hazard functions, Springer-Verlag New York, 2000. p. 7-39.
 48. Yang Y, Zou H. A cocktail algorithm for solving the elastic net penalized Cox's regression in high dimensions. *Stat Interface* 2012;6:167-73.
 49. Kalbfleisch JD, Prentice RL. The Statistical Analysis of Failure Time Data. Chapter of relative risk cox regression model, John Wiley & Sons; 2011. p. 95.
 50. Moghimi-Dehkordi B, Safaee A, Pourhoseingholi MA, Fatemi R, Tabeie Z, Zali MR, *et al*. Statistical comparison of survival models for analysis of cancer data. *Asian Pac J Cancer Prev* 2008;9:417-20.
 51. Jia Y, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, *et al*. Caffe: Convolutional architecture for fast feature embedding. In: Proceedings of the 22nd ACM International Conference on Multimedia. Orland, Florida 2014. p. 675-8.
 52. Cleves M, Gould W, Gutierrez RG, Marchenko Y. An Introduction to Survival Analysis Using Stata, 3rd, StataCorp LP, 2010.
 53. Cox DR, Oakes D. Analysis of Survival Data. London: Chapman&Hall; 1984.