A deep multiple instance model to predict prostate cancer metastasis from nuclear morphology

Nathan Ing, Beatrice S. Knudsen, Arkadiusz Gertych
Cedars-Sinai Medical Center

Jakub M. Tomczak, Max Welling
University of Amsterdam

Isla P. Garraway, Eric Miller
UCLA

Abstract

We consider the problem of identifying the patients who are diagnosed with high-grade prostate cancer using the histopathology of tumor in a prostate needle biopsy and are at a very high risk of lethal cancer progression. We hypothesize that the morphology of tumor cell nuclei in digital images from the biopsy can be used to predict tumor aggressiveness and posit the presence of metastasis as a surrogate for disease specific mortality. For this purpose, we apply a compositional multi-instance learning approach which encodes images of nuclei through a convolutional neural network, then predicts the presence of metastasis from sets of encoded nuclei. Through experiments on prostate needle biopsies (PNBX) from a patient cohort with known presence (M1 stage, $n = 85$) or absence (M0 stage, $n = 86$) of metastatic disease, we obtained an average area under the receiver operating characteristic curve of 0.71 ± 0.08 for predicting metastatic cases. These results support our hypothesis that information related to metastatic capacity of prostate cancer cells can be obtained through analysis of nuclei and establish a baseline for future research aimed at predicting the risk of future metastatic disease at a time when it might be preventable.

Introduction

While the presence of high grade prostate cancer (HGPC) in prostate needle biopsies (PNBX) is a risk factor in lethal progression [1], many patients die of other causes. Thus, there is an immediate, unmet need to stratify high-risk patients and identify the subgroup that is at very high risk of cancer specific mortality and requires aggressive treatment. Unfortunately, to date neither pathological nor genomic markers have succeeded in predicting the presence of clinically significant metastasis at diagnosis or of occult metastasis that may progress to a lethal form of cancer following primary treatment. We hypothesize that cancer nuclei visible in the PNBX of patients with metastatic or non-metastatic clinical stage hold unrealized information associated with aggressive behavior of cancer cells and likelihood of lethal metastatic prostate cancer progression.

Nuclear morphologic and texture patterns visualized through special Feulgen staining have been previously associated with prostate cancer aggressiveness [2, 3], however, this approach is currently not used to subtype aggressive HGPC. Therefore, we decided to develop a nuclear analysis approach for slides stained with Hematoxylin and Eosin (H&E), which is broadly used as part of the regular clinical workup for microscopic diagnosis. To test the hypothesis that nuclei associated with HGPC contain information pertaining to the metastatic potential of the cancer we adopt the assumptions and framework of multiple instance learning (MIL) [4, 5]. We reason that the nuclei within regions of HGPC in each PNBX are an exchangeable set, each nucleus with an unknown share in predicting the ultimate outcome [6]. Following this rationale, we fit a function to compute the probability that a

1st Conference on Medical Imaging with Deep Learning (MIDL 2018), Amsterdam, The Netherlands.

set of nuclei comes from a case of metastatic disease. For this propose we develop a convolutional neural network (CNN), trained end-to-end through backpropagation and stochastic gradient descent, in a manner consistent with supervised multiple instance learning from weakly labelled data.

Predicting the presence of metastasis from tumor nuclei

Data and model architecture

Data from a cohort of 171 patients with HGPC within the Los Angeles Veterans Affairs Healthcare System was collected between 2000 and 2013 [1]. For each patient, an H&E stained PNBX was digitized at $40\times$ on an Aperio AT Turbo slide scanner (Leica Biosystems) and labelled with the metastatic state of the patient, Y. Y is a binary state with $Y = 1$ indicating distant metastasis at the time of biopsy ($n = 85$ cases), and $Y = 0$ indicating the patient did not develop metastatic cancer for at least 5 years after diagnosis ($n = 86$ cases). To focus analysis on tumor nuclei, we first collected a set of 30 non-overlapping tiles from predominantly cancerous areas of each PNBX.

For nuclear segmentation in H&E images, we implemented a pipeline based on [7]. Since factors such as the degree of chromatin clumping, staining intensity, and the average nuclear diameter have a dramatic impact on the performance of this method, the parameters for nuclear mask refinement and segmentation were fine-tuned by hand on a case-by-case basis. Nuclear images were extracted from the RGB color space H&E stained images as a 128 pixel square window, centered at the centroid of each nuclear contour. During training and testing, each nuclear image was randomly cropped to 112 pixel sub-images ($x \in \mathbb{R}^{\times 112 \times 112 \times 3}$). Note that since we use the whole contents of the square, the subjective details of the nuclear contour are rendered unimportant. In all, 153,916 nuclei (average 900 ± 194 nuclei per case) were collected for training and validation.

We predicted the case-level metastatic state from a set of observed nuclei. From each PNBX, a small fraction ($\approx 10\%$) of nuclei were sampled and individually passed through a shared encoding CNN, denoted $E_\phi(\cdot)$. The architecture of $E_\phi(\cdot)$ was based on AlexNet with filter sizes, number of filters and padding amended to suit our nuclear images. The encoder terminated with two additional fully connected hidden layers with dropout, yielding a latent representation of the encoded nuclear image, $z = E_\phi(x)$. To describe the set of nuclei X with K observations we take the average of each latent vector, and the final encoding is expressed as: $\hat{z} = \frac{1}{K}\sum_{k=1}^{K} E_\phi(x_k)$. Note that the averaging operation is permutation invariant, so that the observation order has no bearing on $\hat{z}$ [8]. We can treat the hyperparameter K as a constant, a random variable, or predict an optimal value through an external decision making system. After combination, $\hat{z}$ was passed to a two layer feed forward neural network classifier, $C_\theta(\cdot)$, which results in a vector $\hat{y}$ that is a binary encoding for the target, Y. The parameters of neural networks, θ and ϕ, were tuned by stochastic gradient descent and a standard cross entropy loss function. The architecture is schematically represented in Fig.1a.

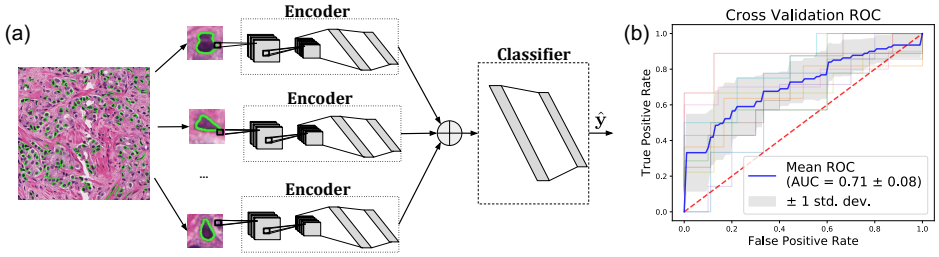

Figure 1: **(a)** Schematic representation of the proposed model. **(b)** Cross validation Receiver Operating Characteristic curves.

Results

To evaluate our framework, we determined the model performance by 10-fold cross validation. The following hyperparameters were fixed for all subsequent experiments. Training proceeded with a batch size of 1 case, from which K nuclei were uniformly sampled. To learn a model invariant to the number of observed nuclei, K was uniformly sampled from the range $K \in [50, 100]$ for each

training example. Latent vector dimensionality was set to $\mathbb{R}^{64}$. The learning rate was set to a fixed 0.0001 for the Adam optimization algorithm.

The 171 cases were divided into 10 non-overlapping training and testing sets so that each case was represented in exactly one testing set. For each fold, 20% of the training set was set aside as a validation set. The accuracy and loss were determined using the validation set after each epoch of training. A snapshot of the model state was saved if the validation accuracy exceeded the previous best. If no improvement was observed in the validation set for 75 consecutive epochs, the training was halted, and the test accuracy evaluated using the most recent snapshot. During testing, a single bag of $K = 200$ nuclei was drawn from each test case. By setting $Y = 1$ as the positive class, and varying the predictive threshold, we obtained an average Receiver Operating Characteristic (ROC) curve with an area under the curve (AUC) of 0.71 ± 0.08 (Fig.1b).

Conclusions

We demonstrated a system to predict the risk of metastatic prostate cancer at diagnosis using a unique dataset of HGPC biopsy images. Through a modernization of research dating to the 1990's, we were able to cast the morphology of tumor nuclei as weak predictors for tumor aggressiveness. A multiple instance learning framework based on a CNN feature extractor was developed and applied to classify sets of nuclei associated with an aggressive subtype of HGPC. Preliminary experiments have resulted in an average cross-validation AUC of 0.71, partially confirming our hypothesis that nuclei do indeed carry morphological information associated with tumor aggressiveness. While CNN's are known to be capable of making association between histologic image patches with abstract clinical outcomes [9], we have shown that similar results are obtainable when focusing on specific nuclear constituents, which is a step towards interpreting the complex features learned by prognosticating CNN's.

Acknowledgements

JMT is funded by the European Commission within the MSC-IF (Grant No. 702666).

References

[1] Eric T Miller, Karim Chamie, Lorna Kwan, Michael S Lewis, Beatrice S Knudsen, and Isla P Garraway. Impact of treatment on progression to castration-resistance, metastases, and death in men with localized high-grade prostate cancer. *Cancer medicine*, 6(1):163–172, 2017.

[2] Alan W Partin, Robert H Getzenberg, Marne J CarMichael, Don Vindivich, John Yoo, Jonathan I Epstein, and Donald S Coffey. Nuclear matrix protein patterns in human benign prostatic hyperplasia and prostate cancer. *Cancer research*, 53(4):744–746, 1993.

[3] Robert W Veltri, M Craig Miller, Alan W Partin, Donald S Coffey, and Jonathan I Epstein. Ability to predict biochemical progression using gleason score and a computer-generated quantitative nuclear grade derived from cancer cell nuclei. *Urology*, 48(5):685–691, 1996.

[4] Murat Dundar, Balaji Krishnapuram, RB Rao, and Glenn M Fung. Multiple instance learning for computer aided diagnosis. In *Advances in neural information processing systems*, pages 425–432, 2007.

[5] Gwenolé Quellec, Guy Cazuguel, Béatrice Cochener, and Mathieu Lamard. Multiple-instance learning for medical image and video analysis. *IEEE reviews in biomedical engineering*, 10:213–234, 2017.

[6] Maximilian Ilse, Jakub M Tomczak, and Max Welling. Attention-based deep multiple instance learning. *arXiv preprint arXiv:1802.04712*, 2018.

[7] Arkadiusz Gertych, Anika O Joseph, Ann E Walts, and Shikha Bose. Automated detection of dual p16/ki67 nuclear immunoreactivity in liquid-based pap tests for improved cervical cancer risk stratification. *Annals of biomedical engineering*, 40(5):1192–1204, 2012.

[8] Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Ruslan R Salakhutdinov, and Alexander J Smola. Deep sets. In *Advances in Neural Information Processing Systems*, pages 3394–3404, 2017.

[9] Dmitrii Bychkov, Nina Linder, Riku Turkki, Stig Nordling, Panu E Kovanen, Clare Verrill, Margarita Walliander, Mikael Lundin, Caj Haglund, and Johan Lundin. Deep learning based tissue analysis predicts outcome in colorectal cancer. *Scientific reports*, 8(1):3395, 2018.