Prostate Cancer Diagnosis and Grading in Whole Slide Images of Core Needle Biopsies

Nitin Singhal NITIN.SINGHAL@AIRAMATRIX.COM
Nilanjan Chattopadhyay
Pranab Samanta
Saikiran Bonthu
Advanced Technology Group, AIRAMATRIX PVT. LTD., Mumbai, India

Editors: Under Review for MIDL 2022

Abstract

Gleason grading is a risk stratification procedure for prostate cancer that is subjective and based on the reporting pathologist's experience and skill. Deep Learning (DL) algorithms have showed potential in improving Gleason grading objectivity and efficiency. On Whole Slide Images (WSI) from a source other than training data, however, DL networks show domain shift and poor performance. Using a novel training process that learns domain agnostic features, we propose a DL approach for segmenting and grading epithelial tissue. When utilised as an aid for core needle biopsy (CNB) evaluation, our DL approach has the potential to increase grading consistency and accuracy, leading to better patient outcomes.
Keywords:

1. Introduction

Prostate cancer is the most frequent malignancy in men and one of the main causes of cancer-related death. Prostate cancer is generally non-aggressive, making it difficult to determine if the disease poses a high enough risk to patients to warrant therapies such as surgery (prostatectomy) or radiation therapy. Gleason grades assigned histopathologically are one of the most powerful prognostic indicators in prostate cancer. Gleason grading, on the other hand, is difficult and subjective, with high inter- and intra-observer variability (Ozkan et al., 2016).Prostate cancer is classified into risk groups based on the Gleason score, which ranges from 3+3 (low risk) to 5+5 (high risk) (high risk) (Epstein et al., 2016).

In whole slide image (WSI) of prostate CNB, we provide a Deep Learning method for detecting cancer areas and predicting Gleason grades. Our method is a semi-supervised method that selects samples for annotation using Active Learning and uncertainty measurements. A Convolutional Neural Network (CNN) architecture that learns domain-agnostic features for enhanced generalisation is also described. With a consensus reference standard, we test our method on a broad group of patients. We compared the outcomes of our technique to the consensus reference using quadratic kappa statistic. We also compared our method to two external test sets published by RUMC and Karolinska (Walter et al., 2022). More details about the dataset, approach, and results can be found in (Singhal et al., 2022).

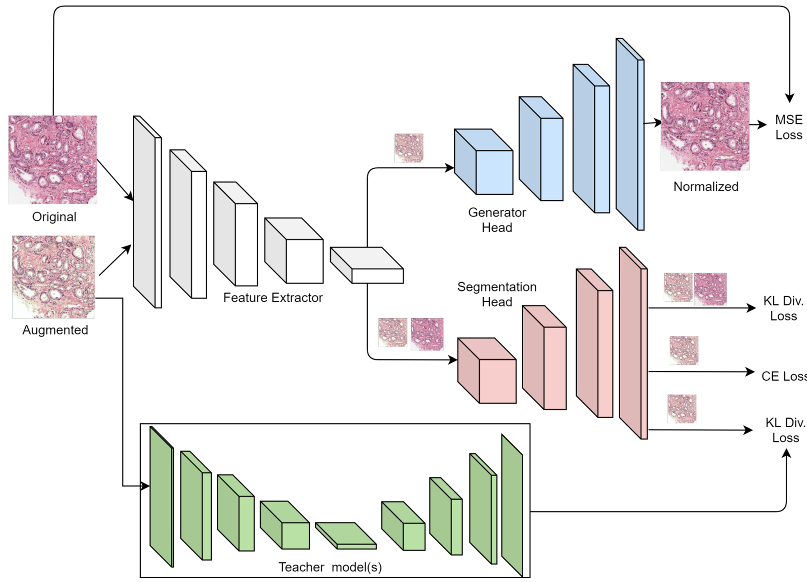

Figure 1: The model learns from three supervisory signals during the training phase: reconstruction loss between raw and colour augmented image logits, KL divergence loss between raw and colour augmented image logits, and segmentation DICE loss.

2. Training Setup

Dataset: A total of 6670 WSI were used in this investigation, sourced from Muljibhai Patel Urological Hospital (MPUH), Radboud University Medical Center, and Karolinska Institute. Cases were randomly assigned to either the development (training/tuning) or independent validation datasets.

Method: Three primary steps are included in the proposed technique for automated Gleason grading. For tumour glands, a pathologist selects and annotates a small number of images with Gleason grades 3+3, 4+4, and 5+5. A fully convolutional network (FCN) is trained to segment the epithelial tissue. The predicted tumour areas are given the pathologist's Gleason grade. These predictions are then used as labels in the next stage. The system next simulates iterative active learning-based data labelling in step two. In the first iteration, a FCN semantic segmentation model is trained for Gleason grade group identification using the initial labelled dataset. Following that, unlabeled images are input into the trained FCN, and an uncertainty measure is produced for each sample. The most uncertain samples are next annotated and added to the initial training set by a pathologist. The FCN gets re-trained with a new batch of annotated images in the next iteration. This process is continued until there are no samples with an uncertainty measure greater than the threshold.

Using a multi-task paradigm, we provide a domain-neutral training methodology for greater generalizability on unseen dataset. As shown in Figure 1, we train a segmentation model with a stain-normalization network as its feature extractor (Generator Head). We demonstrate the multi-task model with a pair of raw and color-augmented images that approximate stain colour variations over time. We penalise the gap between the logits for

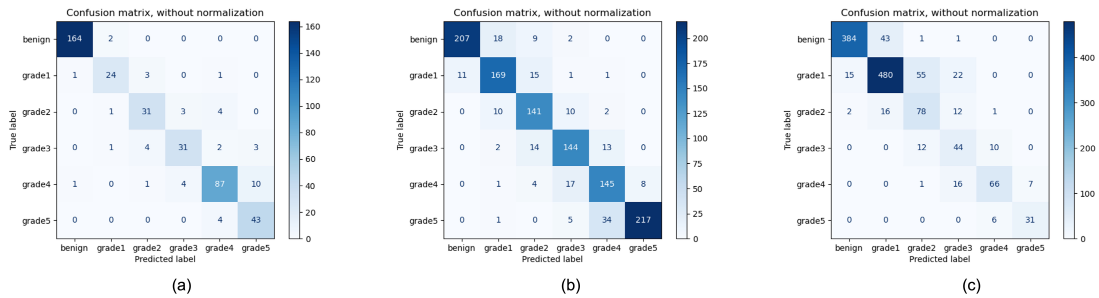

Figure 2: Gleason grade group confusion matrices for test sets (a) MPUH; (b) Radboud; (c) Karolinska.

colour augmented and raw images to enforce learning stain robust features. As a result, the model is steered through stain normalisation and detects the augmented image with accuracy.

3. Results and Discussion

The accuracy of our suggested approach was 89.4 percent, with 0.92 agreement (kquad) with the consensus on grading the 425 biopsies in the test set. The accuracy of the external Radboud test set was 85.3 percent with a kquad of 0.96. The model had an accuracy of 83.1 percent and a kquad of 0.93 on an external test set from Karolinska, where the system had not been exposed to this data during training. Figure 2 depicts the confusion matrix between core-level annotations derived from several test sets and predictions provided by the proposed system.

4. Conclusion and Future work

We developed a novel Deep Learning system to detect tumour areas and predict Gleason grades based on the WSI of prostatic CNBs. Our method can be utilised as a support tool for prostate CNB evaluation, with the potential for more consistent and accurate grading, leading to better patient outcomes.

References

J. I. Epstein et al. The 2014 international society of urological pathology (ISUP) consensus conference on gleason grading of prostatic carcinoma. *Am. J. Surg. Pathol.*, 40:244–252, 2016.

T. A. Ozkan et al. Interobserver variability in gleason histological grading of prostate cancer. *Scand. J. Urol.*, 50:420–424, 2016.

N. Singhal et al. A deep learning system for prostate cancer diagnosis and grading in whole slide images of core needle biopsies. *Sci Rep*, 12(3383), 2022.

B. Walter et al. Artificial intelligence for diagnosis and gleason grading of prostate cancer: the panda challenge. *Nature Medicine*, 28:154–163, 2022.