Physical Color Calibration of Digital Pathology Scanners for Deep Learning Based Diagnosis of Prostate Cancer

Xiaoyi Ji[1], Richard Salmon[2], Nita Mulliqi[1], Henrik Olsson[1], Lars Egevad[3], Pekka Ruusuvuori[4,5], Martin Eklund[1], Kimmo Kartasalo[1,5]

1. Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden.
2. Department of Life Sciences, FFEI Ltd, Hemel Hempstead, UK.
3. Department of Oncology and Pathology, Karolinska Institutet, Stockholm, Sweden.
4. Institute of Biomedicine, Cancer Research Unit and FICAN West Cancer Centre, University of Turku and Turku University Hospital, Turku, Finland.
5. Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland.

Introduction

It has been observed that there exists a drop in performance on deep learning based diagnosis of prostate cancer caused by the technical instability introduced by whole slide images (WSIs) acquired using different scanners in various institutions (Fig. 1). To solve this problem, Sierra slide from FFEI (Fig. 2) is introduced to estimate the color profiles of scanners and transfer the color of the corresponding images to a standard profile.

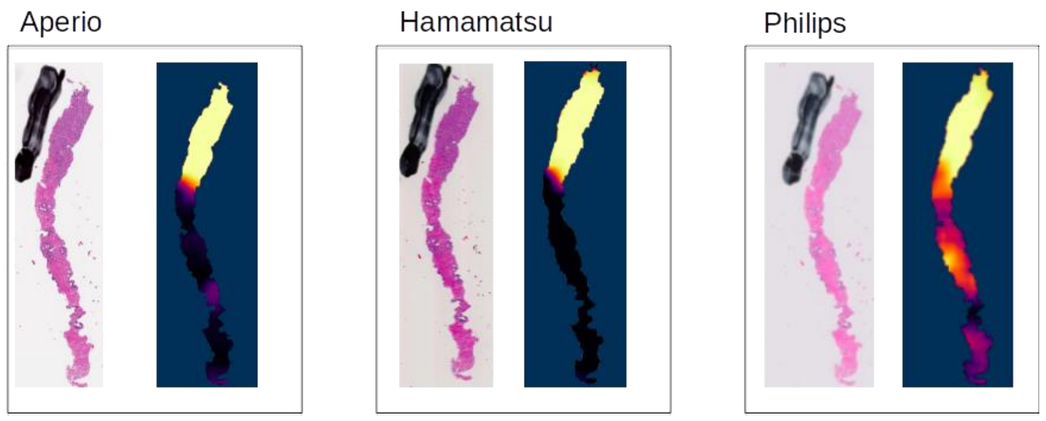

Figure 1: Heatmaps indicating the cancer grades estimation by AI. The results show an observable variance on the same slide scanned with three different machines.

Figure 2: Sierra Color Calibration slide with uniquely stained bio-polymer covers the full range of pathology stains used.

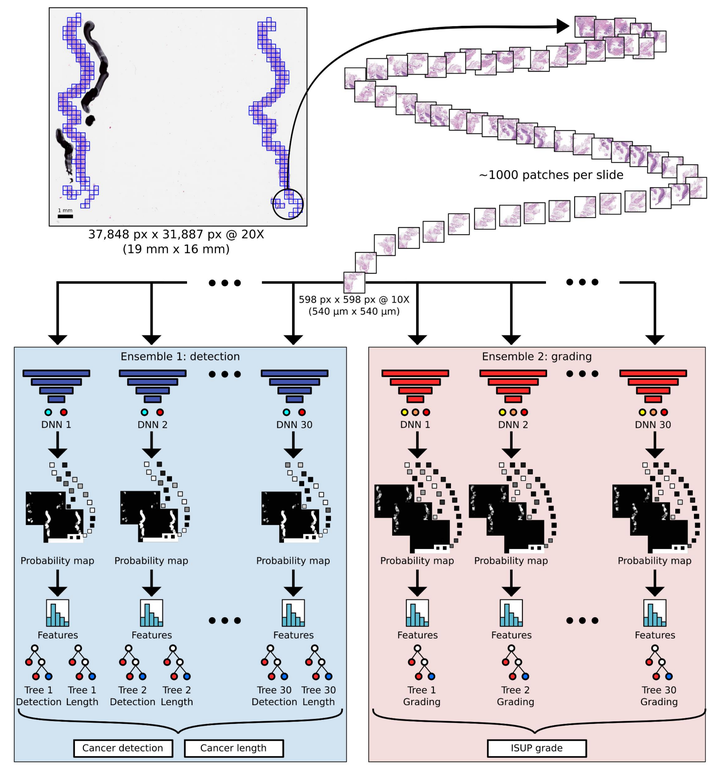

Figure 3: Overview of the artificial intelligence system

Methods

A total of 3,653 biopsies scanned by Aperio ScanScope AT2 were used for training, while 100 tuning and 230 testing slides were scanned with a Hamamatsu NanoZoomer S360 C13220 01. One Sierra slide is scanned on both scanners to estimate their scanner-specific ICC color profiles for the color transformation, where the profile of standard RGB has been chosen as the target. Ensembles of Inception V3 DNNs are used for patch level training and prediction, and gradient boosted trees implemented in Xgboost for WSI level prediction of cancer presence. The DNNs are trained and evaluated both with and without calibration and their corresponding prediction performances will be compared. Approximate annotations of malignant regions and Gleason grading were performed by a single pathologist (L.E.).

Result

As a preliminary result, we have successfully applied color calibration on several images, resulting in visually more natural colors (Fig. 4).

By evaluating the results of the prediction by the two models in terms of AUC compared to the pathologist's diagnosis, we expect to observe if color calibration using the Sierra slide improves the performance of deep learning for prostate cancer detection and Gleason grading on WSIs.

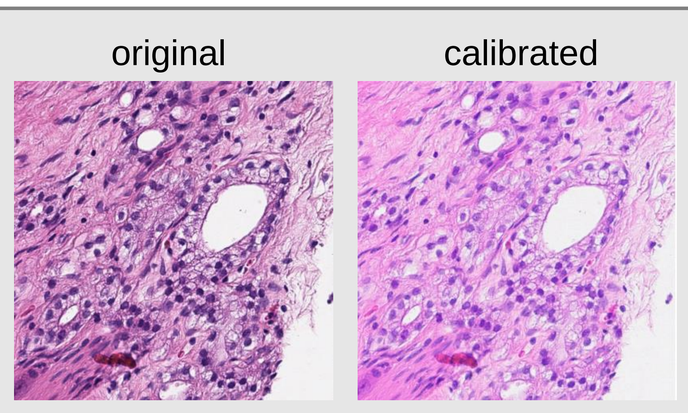

Figure 4: One patch from a WSI scanned by Aperio ScanScope AT2 gets color calibrated to standard RGB with the assistance of the Sierra slide.

Karolinska Institutet
Department of Medical Epidemiology and Biostatistics
Nobels väg 12, Solna, Sweden

Xiaoyi Ji
Research Assistant
E-mail: xiaoyi.ji@ki.se

Karolinska Institutet