AI-Assisted Thyroid Malignancy Prediction From Whole-Slide Images

David Dov, Shahar Z. Kovalsky, Jonathan Cohen, Danielle Elliott Range, Ricardo Henao, Lawrence Carin

With a rapidly increasing incidence rate, thyroid cancer is among the most common cancers [1]. Among the various tests used to diagnose thyroid malignancy prior to surgery, a key test is fine needle aspiration biopsy (FNAB). In this test, follicular (thyroid) cells are smeared onto a glass slide, stained and manually inspected under a microscope by an expert in cytopathology [2]. In many cases, however, FNABs result in indeterminate diagnoses leading to unnecessary surgeries. In this work, we establish a dataset of whole-slide digital thyroid cytopathology images and propose a deep-learning-based algorithm for computational preoperative prediction of thyroid malignancy. Our algorithm was trained on 799 whole-slide scans and tested on a set of 109 scans never seen during the training procedure. Experimental results show that the proposed algorithm achieves human-level thyroid malignancy predictions (compared to three expert cytopathologists). In particular, the algorithm provides a correct and conclusive diagnosis, with zero false (either positive or negative) predictions, for 35% of the cases, demonstrating its potential use as a screening tool. We further show that the proposed algorithm may be used as an assistive diagnostic tool, helping to improve pathologists’ decisions in indeterminate cases. The full description of the medical problem we address in this research and the mathematical details of the proposed algorithm are presented in [3] and [4], respectively.

Background

The Bethesda System for Reporting Thyroid Cytopathology (TBS) is the universally accepted system for thyroid FNAB analysis [5]. It provides guidelines for the analysis of an FNAB slide according to which a cytopathologist assigns a score, known as a TBS category, indicating the risk of thyroid malignancy. This assessment is based on the structure of groups of follicular (thyroid) cells as well as the characterization of individual cells (e.g., size, nuclear features, etc.). There are six TBS categories: TBS 1 indicates an inadequately prepared slide, and is excluded from this work; TBS 2 indicates benign findings and requires no treatment but surveillance; TBS 6 indicates malignancy and is referred to surgery; lastly, the intermediate categories TBS 3, 4, and 5 indicate inconclusive findings with an increased risk of malignancy.

We propose a supervised deep-learning approach for the prediction of thyroid malignancy in cytopathology images. Related deep-learning approaches have been studied for the prediction of thyroid malignancy in ultrasound imaging [6–11] as well as for histopathologic analysis performed after surgery [12]. [13–18] are concerned with the computational analysis of preoperative thyroid cytopathology; their scope, however, is restricted to regions (or individual cells) carefully preselected by an expert. As such, many of the challenges associated with fully automated cytopathological analysis of whole-slide thyroid FNAB images remain open.

Figure 1: (Top left) Whole-slide scan. (Bottom left) Heat map of the predictions of regions containing follicular cells. (Right) Zoom-in on the region marked by the red rectangle.

* D. David, R. Henao and L. Carin are with the Department of Electrical and Computer Engineering, Duke University, Durham, NC 27708, USA (e-mail: david.dov@duke.edu; ricardo.henao@duke.edu; lcarin@duke.edu).
† S. Kovalsky is with the Department of Mathematics, Duke University, Durham, NC 27708, USA (e-mail: shaharko@math.duke.edu).
‡ J. Cohen is with the Department of Surgery, Duke University Medical Center, Durham, NC 27710, USA (e-mail: jonathan.m.cohen@duke.edu).
§ D. Elliott Range is with the Department of Pathology, Duke University Medical Center, Durham, NC 27710, USA (e-mail: danielle.range@duke.edu).
Methods

Dataset and sample selection We have established a dataset including all FNABs with a subsequent thyroidectomy surgery from June 2008 through June 2017, as documented in the institutional databases. The dataset comprises 908 samples each consisting of a whole slide scan with a typical resolution of $\sim 150,000 \times 100,000$ pixels; postoperative histopathology diagnosis used as the gold standard (ground truth) in this study; and preoperative TBS category assigned by a cytopathologist, as recorded in the medical files. The dataset was split into a training and a test set of 799 and 109 samples, respectively. To compare the algorithm to human performance, the slides in the test set were annotated by three expert cytopathologists.

Proposed Algorithm Most indicative for the diagnosis of thyroid malignancy are groups of follicular cells, whose architecture, texture size and color are among the main characteristics determining the TBS category. Follicular groups, however, are only sparsely distributed on the slide whereas most of the slide is covered by blood cells, which are diagnostically irrelevant and are considered background. Inspired by the work of pathologists, the proposed algorithm addresses this challenge via a cascade of two convolutional neural networks. The first network is trained using a supervised procedure to identify informative regions of the scan containing follicular cells, distinguishing them from background regions. The second network aggregates local decisions, based on the informative regions, into a global, slide-level, prediction of thyroid malignancy.

We further propose to simultaneously predict TBS category along with the malignancy prediction from a single output of the network. Specifically, the TBS category is predicted by comparing the output of the network to a set of threshold values via an ordinal regression framework that associates higher TBS categories with increased probabilities of malignancy. These TBS predictions regularize the training process, improve the prediction of thyroid malignancy, and could be further employed as diagnostic screening tool.

Results

Fig. 1 demonstrates how the first neural network successfully identifies follicular cells (bright colors in the heat map), distinguishing them from background regions. We further compare the proposed algorithm to human decisions based on medical records (MR TBS), as well as those of three expert cytopathologists (Expert 1 to 3). To evaluate human performance, we refer to TBS categories 2 to 6 assigned by the cytopathologist as corresponding to malignancy predictions with increasing probabilities. The results are presented in Fig. 2 in the form of receiver operating characteristic (ROC) and precision-recall curves. Compared to the cytopathologists, the proposed algorithm achieves comparable AUC scores and improved AP scores.

We further advocate the use of TBS predictions obtained with the proposed algorithm as a screening tool. Out of 109 tested cases, the algorithm provides 29 predictions of TBS 2 and 10 predictions of TBS 6, all of which (39 total) correctly correspond to benign and malignant cases, respectively. Moreover, these 39 cases include 11 cases that were previously classified as indeterminate by a pathologist; thus suggesting that the proposed algorithm can be used as an assistive diagnostic tool, helping pathologists resolve indeterminate cases.

Implications for improving the value of care

While we plan to validate the performance of the proposed algorithm on a substantially increased cohort, current results already demonstrate potential for improving the value of care as follows: i) The Bethesda categories predicted
by the algorithm may be used as a screening tool: all 39 cases predicted by the algorithm as TBS 2 and 6 are indeed benign and malignant, respectively. ii) The algorithm may be used in remote areas where the availability of expert cytopathologists is limited. iii) The algorithm may improve pathologists’ decisions in indeterminate cases.

References


