





# Antibody Supervised Training of a Deep Learning Based Algorithm for Leukocyte Segmentation in Papillary Thyroid Carcinoma

Sebastian Stenman , Dmitrii Bychkov , Hakan Küçük, Nina Linder, Caj Haglund, Johanna Arola , and Johan Lundin 

**Abstract**—The quantity of leukocytes in papillary thyroid carcinoma (PTC) potentially have prognostic and treatment predictive value. Here, we propose a novel method for training a convolutional neural network (CNN) algorithm for segmenting leukocytes in PTCs. Tissue samples from two retrospective PTC cohort were obtained and representative tissue slides from twelve patients were stained with hematoxylin and eosin (HE) and digitized. Then, the HE slides were destained and restained immunohistochemically (IHC) with antibodies to the pan-leukocyte anti CD45 antigen and scanned again. The two stain-pairs of all representative tissue slides were registered, and image tiles of regions of interests were exported. The image tiles were processed and the 3,3'-diaminobenzidine (DAB) stained areas representing anti CD45 expression were turned into binary masks. These binary masks were applied as annotations on the HE image tiles and used in the training of a CNN algorithm. Ten whole slide images (WSIs) were used for training using a five-fold cross-validation and the remaining two slides were used as an independent test set for the trained model. For visual

evaluation, the algorithm was run on all twelve WSIs, and in total 238,144 tiles sized  $500 \times 500$  pixels were analyzed. The trained CNN algorithm had an intersection over union of 0.82 for detection of leukocytes in the HE image tiles when comparing the prediction masks to the ground truth anti CD45 mask. We conclude that this method for generating antibody supervised annotations using the destain-restain IHC guided annotations resulted in high accuracy segmentations of leukocytes in HE tissue images.

**Index Terms**—Digital pathology, artificial neural network, antibody-supervised learning, papillary thyroid carcinoma, tumor-infiltrating lymphocytes.

## I. INTRODUCTION

PAPILLARY thyroid carcinoma (PTC), the most common variant of thyroid cancer, shows an increase in incidence and is about three times more common in women [1]–[3]. In the US, about 52,000 new cases of PTC are diagnosed annually. However, treated with surgery and radioiodine ablation therapy, the vast majority of patients are cured, and the 5-year survival rate for PTC is over 98% [1]–[4].

The immune response plays a crucial role in the defense against the development of cancer. However, there is also evidence that inflammatory cells can be actively tumor promoting [5]. The inflammatory milieu of PTC plays a crucial role in tumor progression, metastasis and recurrence of thyroid cancer [6], [7]. The presence of immune cells has been shown to correlate with a favorable outcome of PTC [8], [9]. The prognostic significance of specific immune cells in PTC has also been studied by analyzing immunological parameters specific to certain cells [10]. Several specific immunological markers have been shown to be prognostically significant, including CD8 and PD-L1 [11].

Tumor-infiltrating lymphocytes (TILs) predict a more favorable survival in numerous types of cancers; e.g. breast cancer [12], colon cancer [13], and melanoma [14]. Immune cells are currently to the largest extent quantified by pathologists through microscopy of tissue sections [15]. However, this method is time consuming, has high inter- and intraobserver variability, and consequently a poor reproducibility. Thus, new and more objective methods for immune cell quantification are needed.

A class of artificial intelligence methods showing great performance in various image recognition tasks in digital pathology is deep learning-based algorithms [16]. Convolutional neural

Manuscript received December 6, 2019; revised March 26, 2020 and May 8, 2020; accepted May 10, 2020. Date of publication May 29, 2020; date of current version February 4, 2021. This work was supported in part by Sigrid Jusélius Foundation (C.H., J.L.), in part by Medicinska Understödsföreningen Liv och Hälsa (J.A., C.H., N.L., J.L.), and in part by Flnska Läkaresällskapet. (Corresponding author: Sebastian Stenman.)

Sebastian Stenman is with the Institute for Molecular Medicine Finland – FIMM, University of Helsinki, 00290 Helsinki, Finland, and with the HUSLAB Pathology Department, Helsinki University Hospital, 00290 Helsinki, Finland, and also with the Department of Surgery, Helsinki University Hospital, 00290 Helsinki, Finland (e-mail: sebastian.stenman@helsinki.fi).

Dmitrii Bychkov and Hakan Küçük are with the Institute for Molecular Medicine Finland – FIMM, University of Helsinki, 00290 Helsinki, Finland (e-mail: dmitrii.bychkov@helsinki.fi; hakan.kucukel@helsinki.fi).

Nina Linder is with the Institute for Molecular Medicine Finland – FIMM, University of Helsinki, 00290 Helsinki, Finland, and also with the Department of Women's and Children's Health, International Maternal and Child Health, Uppsala University, 75310 Uppsala, Sweden (e-mail: nina.linder@helsinki.fi).

Caj Haglund is with the Department of Surgery, Helsinki University Hospital, 00290 Helsinki, Finland, and also with the Research Programs Unit in Translational Cancer Medicine, University of Helsinki, 00290 Helsinki, Finland (e-mail: caj.haglund@helsinki.fi).

Johanna Arola is with the HUSLAB Pathology Department, University of Helsinki, 00290 Helsinki, Finland (e-mail: johanna.arola@hus.fi).

Johan Lundin is with the Institute for Molecular Medicine Finland – FIMM, University of Helsinki, 00290 Helsinki, Finland, and also with the Department of Global Public Health, Karolinska Institutet, 171 77 Solna, Sweden (e-mail: johan.lundin@helsinki.fi).

This article has supplementary downloadable material available at <https://ieeexplore.ieee.org>, provided by the authors.

Digital Object Identifier 10.1109/JBHI.2020.2994970

networks (CNNs) have been applied to many tasks in pathology, including cell detection [17], outcome prediction [18], as well as analyzing complex spatial patterns within tumors [19]. Also, deep learning algorithms have been applied to a wide range of tasks in image cytometry [20]. Indeed, CNNs have already shown promising results in quantifying TILs in hematoxylin and eosin (HE) stained tissue samples [21]. Furthermore, reproducibility can even further be improved when using leukocyte-specific immunohistochemical (IHC) stains as a reference when annotating leukocytes in HE stained samples [22], [23]. However, the tissue morphology might significantly change in consecutive tissue sections, particularly on cell level. This can prove to be problematic when using one section as an annotation reference for another. Therefore, sequential staining and digitization of the same tissue section, as has been proposed in previous works, would be preferred for referencing purposes [24], [25].

In the present proof-of-concept study, we propose a method for generating antibody-supervised annotations for training of CNN algorithms. Our aim was to assess the feasibility of a machine-learning based method for segmenting leukocytes. In the present study, we trained the model on HE stained tissue sections since it is the most widely used stain in routine diagnostics. The annotations were generated using a novel destain-retain protocol where the pan-leukocyte anti CD45 antibody staining formed the ground truth for the HE stained samples.

## II. MATERIAL AND METHODS

### A. Patient Cohort

The twelve patient cases used in the present study derived from two different patient cohorts. Five patients were originally included in a cohort consisting of 65 PTC patients treated between 1973 and 1996 [26], [27]. The remaining seven cases derived from a newer series of PTC patients treated between 2003 and 2013. All patients were treated at the Helsinki University Hospital. These representative cases visually contained varying amounts of leukocyte infiltration were selected to be used in training and testing of the CNN algorithm. As no clinical records were retrieved for this study, and the study contained no personal identifiers, no written consent was required according to the Ministry of Social Affairs and Health, Finland Act on the Medical Use of Human Organs, Tissues and Cells (Amendments including and up to 227/2013)

### B. Staining Protocol and Digitization of Tissue Samples

Two researchers (S.S., J.A.) reviewed all available original tissue glass slides of the twelve patients included in the training and the independent test set. One formalin-fixed and paraffin-embedded (FFPE) tissue block containing the most representative tumor material was selected for each of the twelve patients. The selected FFPE blocks were retrieved from the archives of Helsinki University Hospital Laboratory (HUSLAB, Helsinki, Finland). Sections ( $0.3 \mu\text{m}$ ) were freshly cut and fixed on glass slides. The tissue slides were then stained with HE according to standard procedures. The HE stained samples were digitized

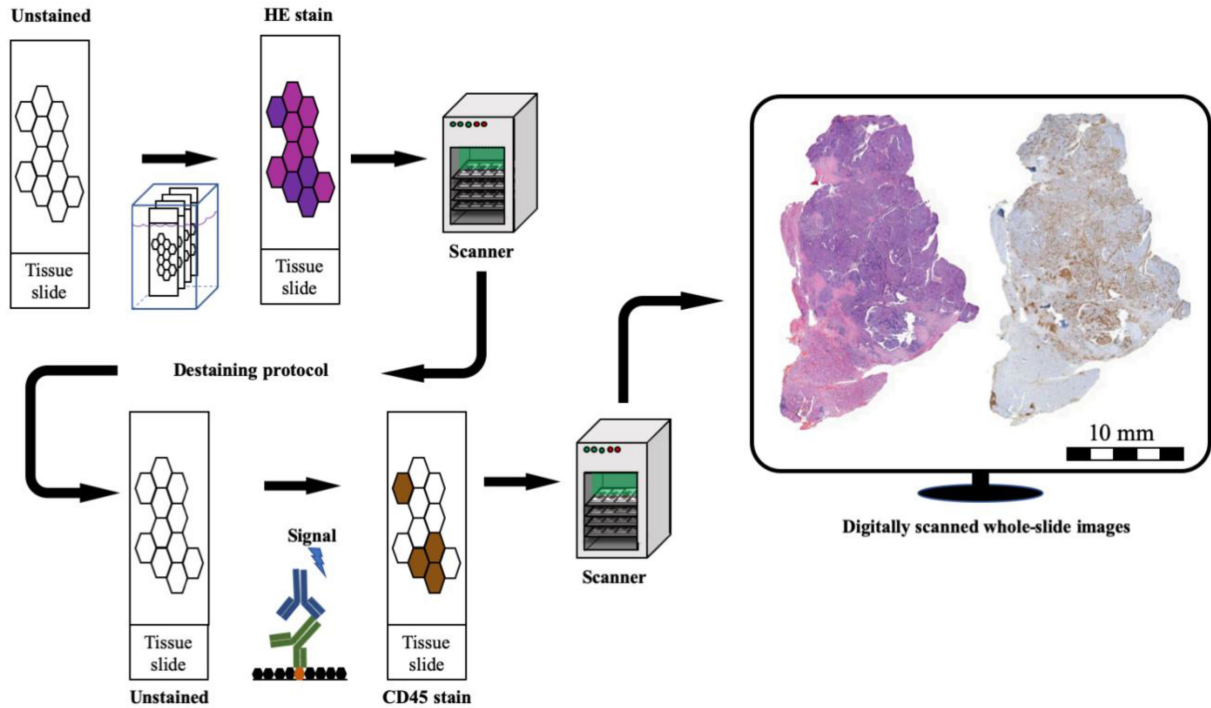
with a whole-slide image scanner (Pannoramic 250, 3DHitech, Hungary). The HE procedure as well as the scanner specifics are described in detail in the appendix. The scanned images ( $0.24 \mu\text{m}/\text{pixel}$ ) were then imported to an image management platform (WebMicroscope, Aiforia Technologies Oy, Helsinki, Finland). After digitization of the HE slides, the coverslips were soaked off in xylene and the sections were rehydrated. Then the HE was boiled off during antigen retrieval for 20 min in 99 °C 10 mM Tris/ 1mM EDTA pH9 solution. For antigen retrieval, slides were pre-treated with Cell Conditioning 1 buffer (Ventana Medical Systems, Inc., Arizona, USA) for 20 min. The tissue sections were incubated with the primary anti CD45 antibody (RTU CD45, clone 2B11PD7/26, Ventana) for 44 min at room temperature and the tissue sections were then processed using an automated staining system (BenchMark ULTRA system, Ventana). A 3,3'-diaminobenzidine (DAB) kit (UltraVIEW, Ventana) was used as the detection. Finally, the anti CD45 stained tissue sections were digitized into whole-slide images (WSIs) using the same scanner as with the HE stained samples (Fig. 1).

### C. Creation of Binary Masks

For creating binary masks, we first created a custom software using C# in Dotnet (Microsoft, Redmond, WA) and Windows Presentation Foundation (WPF, Microsoft, Redmond, WA) frameworks. The five WSI pairs of HE and anti CD45 DAB antibody stained tissue sections were imported to the software and registered. Using the abovementioned custom software, we manually registered the WSIs using the HE as base layer and the ground truth anti CD45 DAB immunostaining as the top layer. First, the slides were roughly registered using morphological landmarks that could be seen in the tissue slides. Following this, regions of interest (ROIs) were exported as  $5000 \times 5000$ -pixel sized tiles. These tiles were then re-layered and re-registered, now on a cell level, and further tiled into smaller  $500 \times 500$ -pixel sized image tiles. This two-step matching and tiling procedure was performed to limit the impact of potential stitching shifts that occur in the digitization of the slides (Fig. 2). During tiling and exporting of the  $500 \times 500$  sized images, the ground truth mask, the anti CD45 DAB stained sample, was turned into a binary mask in multiple steps. First, the image tiles were split into red, green and blue color channels. Since the blue color channel had the best contrast between DAB stained anti CD45 positive regions and the background, the red and green color channels were discarded. The blue color channel image was then converted into a binary mask by manually selecting the threshold matrix prior to export. After this, the binary mask was further processed by blurring. Finally, noise was filtered out by discarding areas smaller than a total area of 350 pixels (Fig. 3).

### D. Image Datasets

A total of 1,738  $500 \times 500$ -pixel image tiles of the twelve destained-restrained WSIs (range: 48–321 tiles per tissue slide) were selected and used in the training and testing of the CNN algorithm. The tiles of ten of these slides ( $n = 1,387$ ) were used for training and validating the CNN algorithm. The trained model was then tested on an independent test set comprising 351



**Fig. 1.** Flowchart of protocol used in the study. First, tissue sections were stained with hematoxylin and eosin. Then, the sections were digitized into whole-slide images (WSIs) before being destained. The sections were then restained with anti CD45 antibody immunostaining and again scanned into WSIs. This staining protocol yielded HE and anti CD45 antibody stains of the same tissue section.

500 × 500-pixel tiles exported from the remaining two destain-restained tissue slides (Fig. 4). The entire WSIs in the training set were analyzed for internal validation outside the tiled regions used for training.

The WSIs were analyzed in 500 × 500-pixel image tiles. Before performing the analysis, we discarded the all-white tiles which did not include any tissue. Excluding the training and testing tiles, 236,377 tiles of 500 × 500 pixels (mean 19,698 tiles per slide, range: 11,464–30,476 tiles) were analyzed with the CNN algorithm. Of these, 197,071 were from the ten WSIs included in the cross-validation training. The 39,306 remaining tiles were from the WSIs in the independent test set.

### E. Deep Convolutional Neural Network Image Analysis

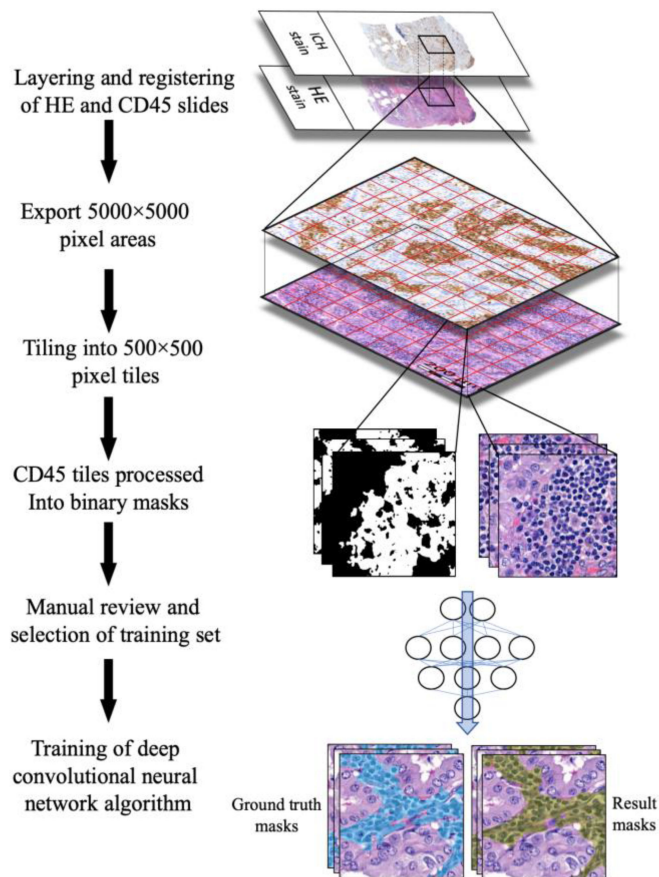
The U-Net architecture [28] is a CNN tailored to solve various image segmentation tasks in the medical domain. Specifically, the U-Net performs dense semantic segmentation, where each pixel of the input image is assigned a corresponding class label. In our study, we adapted the U-Net architecture with ImageNet [29] pre-trained weights and ResNet-18 [30] backbone to perform binary segmentation, i.e. separation of leukocytes from the rest of the tissue. The upward path was left identical to the original U-Net architecture. We used the default learning rate of 0.001. Batch normalization was not used, nor was dropout or L1/L2 regularization. Also, both encoder and decoder weights were optimized at training phase. We utilized on the fly data augmentation by applying random horizontal and vertical flips to the image-mask pairs as well as random shear up to 15 percent. Training was done by presenting augmented HE tissue samples

as input and corresponding anti CD45 DAB binary masks as output. A five-fold cross-validation method was implemented. The folds were made by preserving the percentage of tiles from each of the WSIs in both training and validation splits. In each fold, an average of 1,110 tiles were used for training and an average 277 tiles were used for validation (Fig. 4). After the network models reached their best performance on the validation set, it was evaluated on held-out image-mask pairs to validate that the trained model generalizes and performed well on unseen data. Each of the five models trained in cross-validation were applied to the held-out set and the results were averaged. Evaluation of the results was done both quantitatively, by calculating segmentation accuracy metrics, as well as qualitatively, by visual examination and comparison of ground truth labels with predicted segmentation masks. For quantitative assessment of algorithm performance, the prediction map created by the algorithm was turned into binary masks which were compared to the anti CD45 DAB ground truth binary masks. For visual assessment, the probability score for each pixel generated by the algorithm was turned into a color intensity score which resulted in a heatmap directly based on the probability map. The generated heatmap was then registered with the HE stain for each corresponding WSI for visual assessment. At training an adaptive learning rate optimization algorithm [31] was minimizing the Jaccard index in mini batches of size 16 over 45 epochs.

### F. Ethical Statement

The Ethics Committee at the University of Helsinki approved the study protocol (226/E6/2006, extension 17.4.2013). The





**Fig. 2.** The hematoxylin and eosin (HE) and pan-leukocyte anti CD45 antibody stained tissue slides were layered and manually registered using a custom software. Larger,  $5000 \times 5000$ -pixel image tiles of regions of interest were then exported. Following this, the tiles were again layered and matched and further tiled into  $500 \times 500$ -pixel image tiles. During the last export, the registered positively DAB stained anti CD45 regions were processed in multiple steps and turned into binary masks. The binary masks were used as annotations for the HE stained images and a convolutional neural network (CNN) algorithm was trained based on the patterns of these masks. Results of the algorithm is illustrated as heatmaps.

National Authority of Welfare and Health approved the retrospective study (Valvira Dnro 10041/06.01.03.01/2012).

### III. RESULTS

For performance evaluation, the trained CNN algorithm was tested on 351 image tiles sized  $500 \times 500$  pixels. For each of these tiles, the probability masks generated by the CNN algorithm were turned into binary masks and compared to the ground truth mask based on the anti CD45 DAB staining. Thus, in pixel-wise comparisons of the algorithm result mask and the ground truth mask, a total of 87.8 million pixels were compared. Based on this independent test set, we observed an intersection over union (IoU) of 0.82. By averaging results from the five models trained in cross-validation, we observed a receiver operating characteristics area under the curve (ROC AUC) of 0.96 on the held-out data when comparing the anti CD45 DAB ground truth mask to the algorithm result mask on pixel level (Fig. 5).

For visual performance assessment, the HE WSIs of all twelve destain-restained tissue sections were analyzed by the trained algorithm. For convenient visual evaluation, we registered the algorithm result heatmap and the HE stain of the ten WSIs included in the training set. For the test WSIs, we registered the HE stain, the algorithm result heatmap, as well as the anti CD45 DAB ground truth stain. The algorithm results could be visually compared to the anti CD45 DAB antibody stain by moving through the layers (Fig. 6). The WSIs subject to the visual performance assessment can be explored via the following URL: <https://tinyurl.com/qorlnlg>

### IV. DISCUSSION

In this proof-of-concept study, we used a novel method for antibody-supervised training of a CNN algorithm. The trained algorithm was highly accurate both measured on a pixel-level in the test set image tiles and through visual examination of the analyzed WSIs, which indicates that the proposed method of generating training sets is feasible.

Machine learning-based tools have previously been used to quantify leukocytes within tumors [21], [22]. However, these methods use supervised learning and rely on manual annotations. This method is both laborious and subjective and thus has poor reproducibility. We propose a method that require no manual annotations and is more objective. To the best of our knowledge, the proposed method for applying antibody supervised annotations derived from binary masks using a destain-remain protocol of the same tissue section has not been described.

Previously, a method for transferring annotations from an IHC stain to HE has been proposed [25]. However, this method differs from the present method in a few ways. First, in our method, we suggest using binary masks when transferring the annotations in between stains which allows for pixel-wise training of the algorithm. In contrast, in the paper by Tellez et al, they train the first CNN patch-wise using  $100 \times 100$ -pixel tiles. Secondly, in the method proposed in this paper, no manual annotations are required, compared to an average of 2 hours per observer. However, we manually reviewed and selected the exported tiles which introduced a manual element in our method as well.

In order to train an algorithm for segmentation tasks, the annotated or labeled training material has to be as precise as possible. Thus, creating high quality training material is a time-consuming process. Our proposed method replaces the manual annotation task with the IHC staining mask that is directly applied as annotations to the HE-stained training material. This trains the CNN model to output a virtual IHC stain e.g. a “digital biomarker” that mimics the performance of the particular antibody used as the mask.

The proposed method for generating training data is fast; matched and marked areas are tiled, processed, and exported in a matter of minutes and can easily be extended to whole WSIs that are hard to fully label by a human annotator. Also, as compared to manual annotation, the antibody-supervised annotation method is likely to be more reproducible.

We explored several different thresholding methods for the image processing, including automatic methods such as Otsu’s

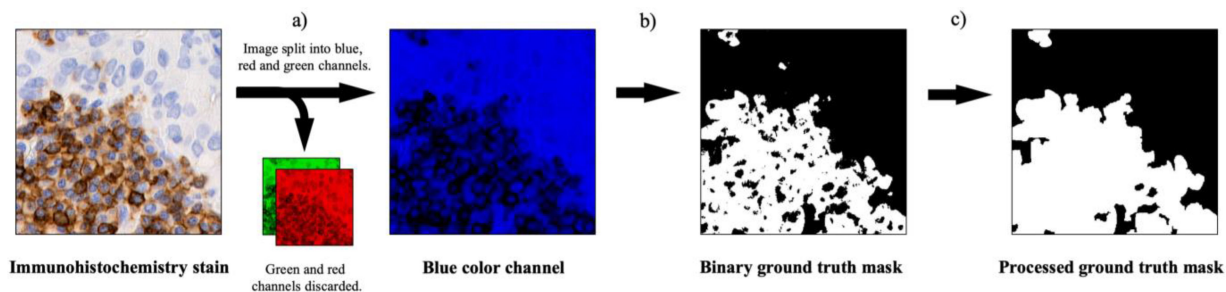


Fig. 3. Image processing protocol. First, the images were split into red, green and blue color channels, and the red and green channels were discarded (a). Since the blue color channel had the best contrast between anti CD45 DAB positive and background regions, we used this for further processing. We then turned the blue channel image into a binary mask (b). This image was then blurred and noise smaller than a total area of 350 pixels were filtered out (c).

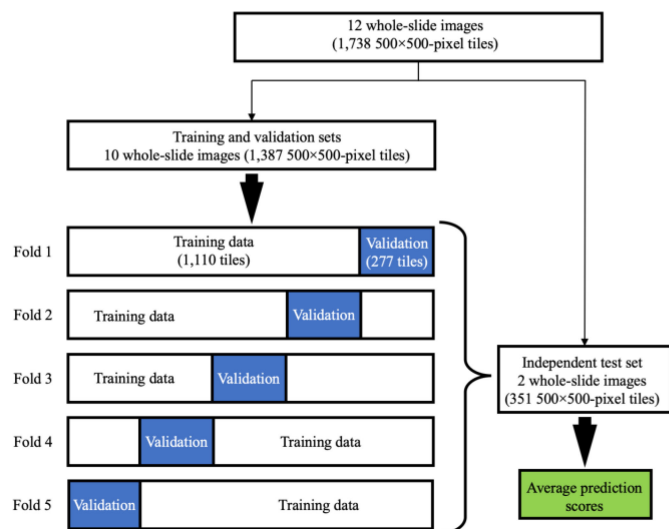


Fig. 4. A consort diagram showing the datasets used in the study. A total of twelve destained-restained tissue slides were used. Ten WSIs were used in a five-fold cross validation training protocol. The tiles from all ten WSIs were pooled and divided up in five batches so that each fold had an equal number of tiles from all ten WSIs averaging 1,110 tiles for training and 277 for validation. The performance of the five models generated were averaged and run on unseen image tiles from the two WSIs in the test set and the performance was evaluated both quantitatively and visually.

method. The masks created by various thresholding methods were visually reviewed and we decided on a manual thresholding method since it gave superior results compared to the other methods. This, in turn, means that the threshold has to be manually selected and thus introduces some subjectivity. However, compared to a fully manual way of creating training annotations, this method of antibody-supervised training is still a more objective way of creating annotations. In our proposed method we selected the blue color channel for further image processing. This color channel was selected because it offered good contrast with the positively DAB stained areas and the rest of the tissue and thus could easily be converted to a binary mask. A method for color deconvolution has previously been proposed [32]. We also explored this method which resulted in similar binary masks as with our proposed method (supplementary Fig. 1). However, using an image with only one-color channel, as we propose, decreases the size of the files being processed to

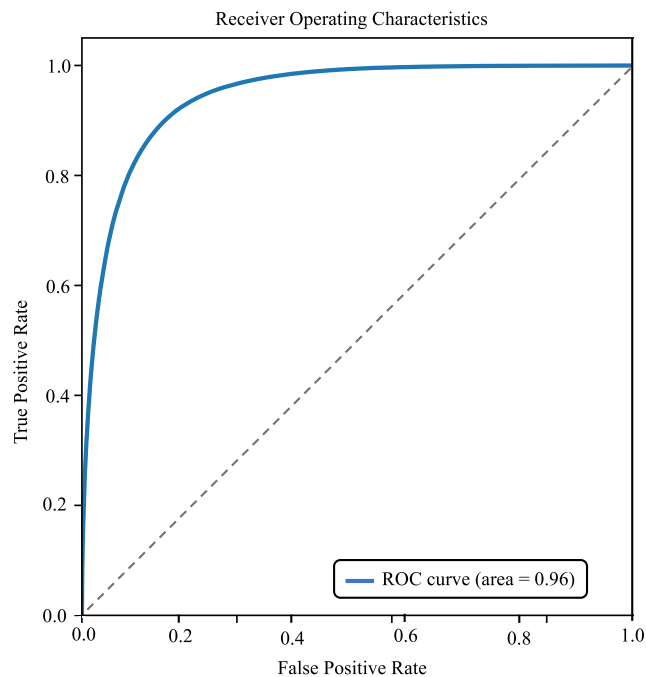
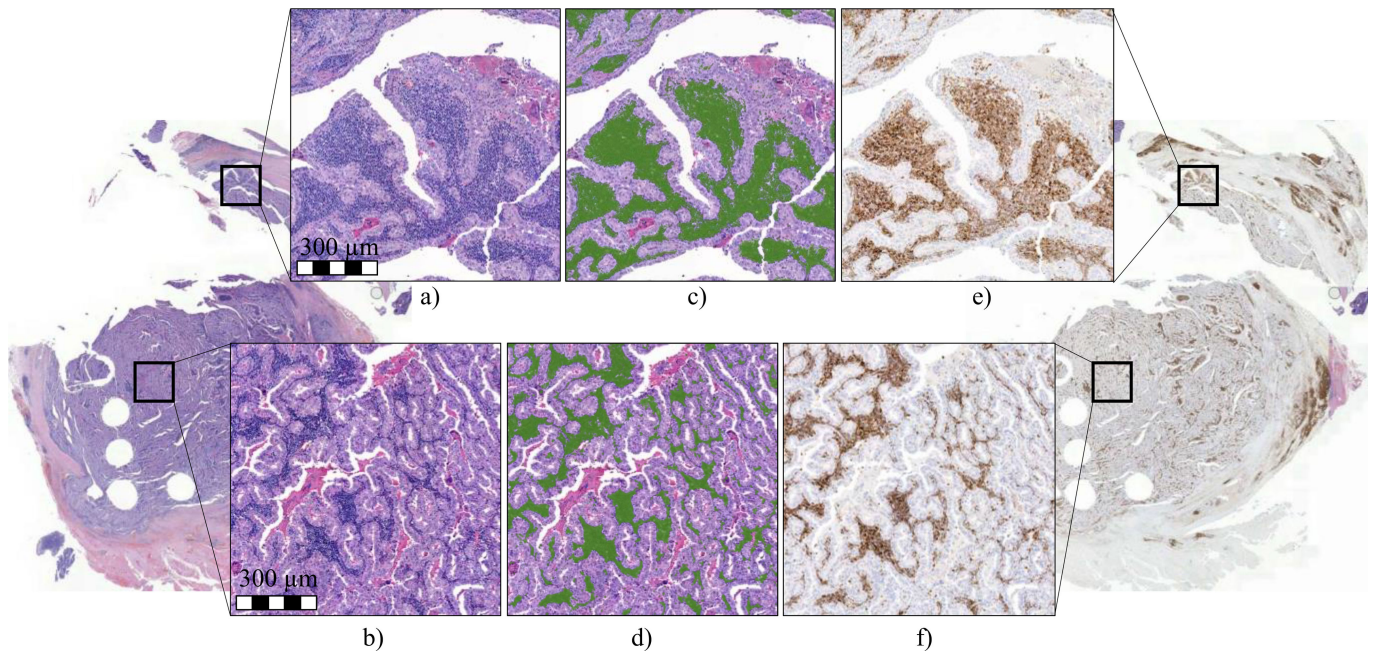


Fig. 5. Average receiver operating characteristics (ROC) curve of the five-fold cross validation algorithm. The pixel-level predictions scores of the five models were averaged and turned into binary masks which were then compared with the anti CD45 DAB ground truth masks of the image tiles in the independent test set. The independent test set comprised of 351 tiles of  $500 \times 500$  pixels. Thus, 87.8 million pixels were evaluated. The area under the curve was 0.96.

one third compared to using RGB images. This is a significant decrease in data being processed and using color deconvolution could drastically slow down the image processing step in bigger datasets.

Since the annotations of the training material is directly based on the IHC stains, the staining quality, in turn, directly affects algorithm performance. In the present study, we limited this by manually reviewing all exported image tiles and selecting the training data. Another issue that needs to be taken into consideration is the scanner stitching artifacts. When digitizing a tissue slide, the slide is first being scanned in tiles and then stitched together. This might cause imperfect alignment and thus some shift between the scanned tiles. This effect is further amplified when layering multiple WSIs and had to be considered





**Fig. 6.** Analysis of one of the whole slide images in the independent test set. Two separate regions of interest zoomed-in from the hematoxylin and eosin slide (a and b) and the corresponding areas in the immunohistochemically stained slide (e and f). The images in between these (c and d) represents heatmaps of the analysis result of the trained deep convolutional neural network algorithm. The rest of the WSIs can be explored in detail via the following URL: <https://tinyurl.com/qorlnlg>.

in the layering and tiling process. Therefore, we used a two-step matching and tiling protocol previously described in the methods section to limit the impact of the stitching shifts (Fig. 2). Also, training images with misalignments due to stitching shift were discarded in the manual review process.

We trained the algorithm using a five-fold cross validation on 1,387 image tiles cropped from ten WSIs and annotated using the proposed antibody-supervised method. In the quantitative performance evaluation, the algorithm accurately segmented leukocytes in the HE stained WSIs when comparing algorithm prediction masks to the ground truth masks. For visual performance evaluation, we then ran the algorithm on all twelve WSIs. The material in the training and testing comprised of only 0.73% of the entire WSI tissue area analyzed by the algorithm. Even though the algorithm was trained on a relatively small training set and few individual tissue slides, the algorithm accurately segmented leukocytes in the test WSIs (URL: <https://tinyurl.com/qorlnlg>). Here, the algorithm can also be seen working well on slides containing few leukocytes (supplementary Fig. 2). However, due to several factors, the accuracy tends to drop when testing algorithms on samples coming from different centers [33]. Thus, including more samples representing a larger variety of fixation, staining and scanning protocols from multiple centers in the training set needs to be evaluated in further studies.

The prognostic and predictive value of the quantity of peritumoral and tumor infiltrating immune cells in PTC is still largely unclear. The HE stain is widely used for identification of leukocytes in tumor tissue in addition to identification of specific antibodies [34]. Therefore, we used HE as the selected staining for training of the CNN algorithm. Our aim was to study whether an algorithm can be trained to detect leukocytes in HE stained

samples guided by a leukocyte-specific antibody. The anti CD45 pan-leukocyte marker was used as the immune cell marker in the present study, but the method needs to be evaluated using other immunohistochemistry stainings in future studies.

In conclusion, we show that a destain-re stain protocol and transfer of the anti CD45 DAB stain to be used as a mask on HE stained samples is a feasible method for quickly generating annotations for training accurate machine learning models.

## APPENDIX

The freshly cut tissue glass slides were stained with hematoxylin & eosin according to standard procedure. First, the tissue sections were deparaffinized and rehydrated after which they were incubated for 10 min at room temperature in Mayer's hemalum solution (Merck 1.09249, Merck Life Science, Darmstadt, Germany). The slides were then rinsed under tap water for 5 min and the staining was differentiated by dipping the slides twice in 70% EtOH/1% HCl solution. The slides were then rinsed with tap water for 5 min more after which they were incubated in aqueous 1% eosin solution (Sigma E4382, Sigma Corporation, USA) for 2 min. After dehydration in alcohol and xylene, the slides were mounted in DPX mounting medium (Sigma, USA).

Both the HE and the immunohistochemistry stained samples were digitized with a scanner (Pannoramic 250 FLASH 3DHISTECH Ltd., Budapest, Hungary) equipped with a plan-apochromat 20x objective (NA 0.8), a VCC-F52U24CL camera (CIS, Tokyo, Japan) with three image sensors ( $1,224 \times 1,624$ ;  $4.4 \times 4.4 \mu\text{m}/\text{pixels}$ ), and a 1.0 adapter. After digitization, the scanned images were compressed into a wavelet format

(Enhanced Compressed Wavelet, ECW, ER Mapper, Intergraph, Atlanta, GA) with a compression ratio of 1:9 and imported to an image management platform (WebMicroscope, Aiforia Technologies Oy, Helsinki, Finland).

### ACKNOWLEDGMENT

The author would like to thank the FIMM Digital Microscopy and Molecular Pathology Unit supported by Helsinki University and Biocenter Finland for outstanding assistance.

### DISCLOSURE STATEMENT

Johan Lundin is a Founder and Chief Scientific Officer at Aiforia Oy, Helsinki, Finland.

### REFERENCES

- [1] A. M. Noone *et al.*, "SEER Cancer Statistics Review 1975-2016. National Cancer Institute," Bethesda, MD, USA, based on Nov. 2017 SEER data submission, posted to the SEER web site, Apr. 2018. [Online]. Available: [https://seer.cancer.gov/csr/1975\\_2015/](https://seer.cancer.gov/csr/1975_2015/)
- [2] L. Davies and H. G. Welch, "Current thyroid cancer trends in the United States," *Otolaryngology—Head Neck Surgery*, vol. 140, no. 4, pp. 317–322, Apr. 2014, doi: [10.1001/jamaoto.2014.1](https://doi.org/10.1001/jamaoto.2014.1).
- [3] R. V. Lloyd, R. Y. Osamura, and G. Kloppel, *WHO Classification of Tumours: Pathology and Genetics of Tumours of Endocrine Organs*, 4th ed. Lyon, France: IARC, 2017.
- [4] D. S. McLeod, A. M. Sawka, and D. S. Cooper, "Controversies in primary treatment of low-risk papillary thyroid cancer," *Lancet*, vol. 381, no. 9871, pp. 1046–1057, Mar. 23, 2013, doi: [10.1016/S0140-6736\(12\)62205-3](https://doi.org/10.1016/S0140-6736(12)62205-3).
- [5] D. Hanahan and R. A. Weinberg, "The hallmarks of cancer," *Cell*, vol. 100, no. 1, pp. 57–70, Jan. 2000, doi: [10.1016/S0092-8674\(00\)81683-9](https://doi.org/10.1016/S0092-8674(00)81683-9).
- [6] J. Modi, A. Patel, R. Terrell, R. M. Tuttle, and G. L. Francis, "Papillary thyroid carcinomas from young adults and children contain a mixture of lymphocytes," *J. Clin. Endocrinol. Metabolism*, vol. 88, no. 9, pp. 4418–4425, Sep. 2003, doi: [10.1210/jc.2003-030342](https://doi.org/10.1210/jc.2003-030342).
- [7] J. S. Jeong *et al.*, "Coexistence of chronic lymphocytic thyroiditis with papillary thyroid carcinoma: Clinical manifestation and prognostic outcome," *J. Korean Med. Sci.*, vol. 27, no. 8, pp. 883–889, Aug. 2012, doi: [10.3346/jkms.2012.27.8.883](https://doi.org/10.3346/jkms.2012.27.8.883).
- [8] S. K. Kim *et al.*, "Chronic lymphocytic thyroiditis and BRAF V600E in papillary thyroid carcinoma," *Endocrine-Related Cancer*, vol. 23, no. 1, pp. 27–34, Jan. 2016, doi: [10.1530/ERC-15-0408](https://doi.org/10.1530/ERC-15-0408).
- [9] L. L. Cunha and L. S. Ward, "Concurrent lymphocytic thyroiditis is associated to less aggressive papillary thyroid carcinomas," *Eur. Arch. Otorhinolaryngol.*, vol. 269, no. 2, pp. 699–700, Feb. 2012, doi: [10.1007/s00405-011-1764-y](https://doi.org/10.1007/s00405-011-1764-y).
- [10] L. L. Cunha *et al.*, "CD8+ tumour-infiltrating lymphocytes and COX2 expression may predict relapse in differentiated thyroid cancer," *Clin. Endocrinology (Oxf)*, vol. 83, no. 2, pp. 246–253, Aug. 2015, doi: [10.1111/cen.12586](https://doi.org/10.1111/cen.12586).
- [11] M. J. Aghajani, T. Yang, C. E. McCafferty, S. Graham, X. Wu, and N. Niles, "Predictive relevance of programmed cell death protein 1 and tumor-infiltrating lymphocyte expression in papillary thyroid cancer," *Surgery*, vol. 163, no. 1, pp. 130–136, Jan. 2018, doi: [10.1016/j.surg.2017.04.033](https://doi.org/10.1016/j.surg.2017.04.033).
- [12] P. Savas *et al.*, "Clinical relevance of host immunity in breast cancer: From TILs to the clinic," *Nature Rev. Clin. Oncol.*, vol. 13, no. 4, pp. 228–241, Apr. 2016, doi: [10.1038/nrclinonc.2015.215](https://doi.org/10.1038/nrclinonc.2015.215).
- [13] J. Galon *et al.*, "Towards the introduction of the 'Immunoscore' in the classification of malignant tumours," *J. Pathol.*, vol. 232, no. 2, pp. 199–209, Jan. 2014, doi: [10.1002/path.428](https://doi.org/10.1002/path.428).
- [14] S. M. Ferrari *et al.*, "Thyroid autoimmune disorders and cancer," *Semin. Cancer Biol.*, vol. 64, pp. 135–146, Aug. 2020, doi: [10.1016/j.semcancer.2019.05.019](https://doi.org/10.1016/j.semcancer.2019.05.019).
- [15] R. Salgado *et al.*, "The evaluation of tumor-infiltrating lymphocytes (TILs) in breast cancer: Recommendations by an International TILs Working Group 2014," *Ann. Oncology*, vol. 26, no. 2, pp. 259–271, Feb. 2015, doi: [10.1093/annonc/mdu450](https://doi.org/10.1093/annonc/mdu450).
- [16] D. Shen, G. Wu, and H. I. Suk, "Deep learning in medical image analysis," *Annu. Rev. Biomed. Eng.*, vol. 19, pp. 221–248, Jun. 2017, doi: [10.1146/annurev-bioeng-071516-044442](https://doi.org/10.1146/annurev-bioeng-071516-044442).
- [17] F. Xing and L. Yang, "Robust nucleus/cell detection and segmentation in digital pathology and microscopy images: A comprehensive review," *IEEE Rev. Biomed. Eng.*, vol. 9, pp. 234–263, 2016, doi: [10.1109/RBME.2016.2515127](https://doi.org/10.1109/RBME.2016.2515127).
- [18] D. Bychkov *et al.*, "Deep learning based tissue analysis predicts outcome in colorectal cancer," *Sci. Rep.*, vol. 8, no. 1, pp. 3395–018–21758-3, Feb. 2018, doi: [10.1038/s41598-018-21758-3](https://doi.org/10.1038/s41598-018-21758-3).
- [19] G. Corredor *et al.*, "Spatial architecture and arrangement of tumor-infiltrating lymphocytes for predicting likelihood of recurrence in early-stage non-small cell lung cancer," *Clin. Cancer Res.*, vol. 25, no. 5, pp. 1526–1534, Mar. 2019, doi: [10.1158/1078-0432.CCR-18-201](https://doi.org/10.1158/1078-0432.CCR-18-201).
- [20] A. Gupta *et al.*, "Deep learning in image cytometry: A review," *Cytometry A*, vol. 95, no. 4, pp. 366–380, Apr. 2019, doi: [10.1002/cyto.a.23701](https://doi.org/10.1002/cyto.a.23701).
- [21] N. Linder *et al.*, "Deep learning for detecting tumour-infiltrating lymphocytes in testicular germ cell tumours," *J. Clin. Pathol.*, vol. 72, no. 2, pp. 157–164, Feb. 2019, doi: [10.1136/jclinpath-2018-205328](https://doi.org/10.1136/jclinpath-2018-205328).
- [22] R. Turkki, N. Linder, P. E. Kovanen, T. Pellinen and J. Lundin, "Antibody-supervised deep learning for quantification of tumor-infiltrating immune cells in hematoxylin and eosin stained breast cancer samples," *J. Pathol. Informat.*, vol. 7, Sep. 2016, Art. no. 38, doi: [10.4103/2153-3539.189703](https://doi.org/10.4103/2153-3539.189703).
- [23] M. Valkonen *et al.*, "Cytokeratin-supervised deep learning for automatic recognition of epithelial cells in breast cancers stained for ER, PR, and Ki-67," *IEEE Trans. Med. Imaging*, vol. 39, no. 2, pp. 534–542, Feb. 2020, doi: [10.1109/TMI.2019.2933656](https://doi.org/10.1109/TMI.2019.2933656).
- [24] H. O. Helin, M. E. Lundin, M. Laakso, J. Lundin, H. J. Helin, and J. Isola, "Virtual microscopy in prostate histopathology: simultaneous viewing of biopsies stained sequentially with hematoxylin and eosin, and alpha-methylacyl-coenzyme A racemase/p63 immunohistochemistry," *J. Urology*, vol. 175, no. 2, pp. 495–499, Feb. 2006, doi: [S0022-5347\(05\)00164-3](https://doi.org/S0022-5347(05)00164-3).
- [25] D. Tellez *et al.*, "Whole-slide mitosis detection in H&E breast histology using PHH3 as a reference to train distilled stain-invariant convolutional networks," *IEEE Trans. Med. Imaging*, vol. 37, no. 9, pp. 2126–2136, Sep. 2018, doi: [10.1109/TMI.2018.2820199](https://doi.org/10.1109/TMI.2018.2820199).
- [26] P. Siironen *et al.*, "Prognostic factors in papillary thyroid cancer: an evaluation of 601 consecutive patients," *Tumour Biol.*, vol. 26, no. 2, pp. 57–64, Mar./Apr. 2005, [Online]. Available: <https://doi.org/10.1159/000085586>.
- [27] S. Stenman, P. Siironen, H. Mustonen, J. Lundin, C. Haglund, and J. Arola, "The prognostic significance of tall cells in papillary thyroid carcinoma: A case-control study," *Tumour Biol.*, vol. 40, no. 7, Jul. 2018, doi: [10.1177/1010428318787720](https://doi.org/10.1177/1010428318787720).
- [28] O. Ronneberger, P. Fischer and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing Computer-Assisted Intervention. Lecture Notes in Computer Science*, Springer, Cham, vol. 9351, 2015.
- [29] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vision Pattern Recogn.*, 2009, pp. 248–255.
- [30] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vision Pattern Recognition*, 2016, pp. 770–778.
- [31] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Representations*, San Diego, CA, USA, 2015, arXiv:1412.6980.
- [32] A. C. Ruifrok and D. A. Johnston, "Quantification of histochemical staining by color deconvolution," *Anal. Quantitative Cytology Histology*, vol. 23, no. 4, pp. 291–299, Aug. 2001.
- [33] F. Wang, L. P. Casalino and D. Khullar, "Deep learning in medicine—Promise, progress, and challenges," *Intemed*, vol. 179, no. 3, pp. 293–294, 2019, doi: [10.1001/jamainternmed.2018.7117](https://doi.org/10.1001/jamainternmed.2018.7117).
- [34] R. Salgado *et al.*, "The evaluation of tumor-infiltrating lymphocytes (TILs) in breast cancer: Recommendations by an International TILs working group 2014," *Ann. Oncology*, vol. 26, no. 2, pp. 259–271, Feb. 2015, doi: [10.1093/annonc/mdu450](https://doi.org/10.1093/annonc/mdu450).