# Artificial Intelligence-Based Digital Image Analysis for Pathological Categorization of Colonic Polyps

Mark A. Valasek, MD, PhD[1], Eli Brazowski, MD[2], Maha Al Khawaja, MD[1], Ran Shadmi[3], Roman Gluskin[3], Yuval Gabay[3], Ifat Rotbein[3], Amir Luchtenstein[3], Israel Ingber[3], Eliron Amir[3] and Lotan Chorev[3]

University of California San Diego, La Jolla, CA, USA[1]; Tel Aviv Medical Center, Tel Aviv, Israel[2]; Nucleai, Tel Aviv, Israel[3]

## Background

Digital image analysis has traditionally been used primarily for research purposes and for the classification of cytological specimens in clinical settings. However, the potential of digital image analysis to aid in the pathological categorization of surgical pathology biopsy specimens is a relatively unknown, but rapidly growing field. For example, previous studies have utilized deep learning to evaluate breast carcinoma and metastases (1, 2), but other features and cancers from other organ systems are being evaluated. In gastrointestinal pathology, classification of polyps is critical for the identification of cancer, determination of cancer risk, and decisions regarding colonoscopic surveillance intervals. Computerized nuclear morphometry with hierarchical clustering has been used to categorize polyps (3) and deep learning has been used to evaluate colonic crypts/glands and colon cancer (4, 5). Here, we utilized an artificial intelligence-based image analysis (AIIA) algorithm with convolutional neural networks (CNN) for the classification of colonic polyps (CP).

## Design

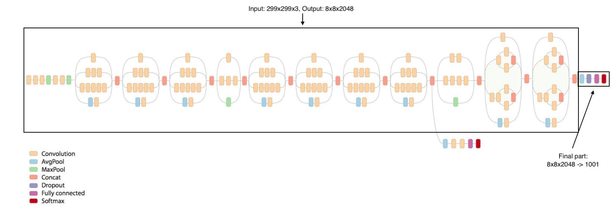


**Figure 1** – The architecture of the convolutional neural network used by Nucleai-CP-Algorithm

A partially annotated set of 417 CP whole slide images (WSI) was utilized to train an AIIA algorithm (Nucleai-CP-Algorithm) based on deep learning convolutional neural networks. Subsequently, we performed a retrospective analysis of a test sample set of 877 WSI including the following diagnoses: 1) normal colonic mucosa (n=87), 2) serrated polyp (i.e. microvesicular hyperplastic polyp and sessile serrated adenoma; n=175), 3) low grade dysplasia (i.e. tubular adenoma; n=577), and 4) high-risk polyp (includes high grade dysplasia and carcinoma; n=38). Alternatively, for binary classification, CP were placed in either non-dysplastic (groups 1 and 2) or dysplastic/malignant (groups 3 and 4) categories. A subset of 219 cases were independently evaluated by multiple (n=3) gastrointestinal pathologists

**CP Classification Algorithm - High Level Design**

The algorithm is based on 3 main sub-components:

1. Tissue Segmentation Algorithm
2. Tile Classification Algorithm
3. Slide Level Decision Algorithm

This was performed for several hundred slides. These slides were annotated by pathologists, and the results compared.

## Methods

***Tissue Segmentation Algorithm***

This algorithm is based on a segmentation CNN. To train this algorithm, hundreds of slides were annotated with tissue and background, and a segmentation CNN was train to segment the tissue vs. the background and any noise (including markers, scratches on the surface etc.). Therefore, the algorithm effectively identifies tissue for further evaluation.

***Tile Classification Algorithm***

1. Several hundred slides were scanned, and annotated in detail. The annotations included marking polygons according to several classes, including: stroma, normal tissue, serrations, low grade dysplasia, high grade dysplasia, etc.
2. A tile classification CNN was trained using tiles extracted from the annotated.
   a) The network receives as an input a tile, and outputs a probability vector which has the probability that the tile corresponds to each of the optional classes.
   b) There are several network models tested with different parameters, different augmentations etc. where the model with the highest performance was chosen.

After a first CNN was trained there were several iterations of enhancing the annotations by fixing mistakes and annotating more areas according to the areas that were less accurate in the CNN output. This was performed until satisfying results were achieved in the tile level classification accuracy.

***Slide Level Decision Algorithm***

To train the slide level algorithm the following steps were performed:

1. Several hundred slides were annotated only using slide level diagnosis (e.g. " Adenoma with Low Grade Dysplasia", "Adenoma with High Grade Dysplasia", "Hyperplastic Polyp" etc.
2. On each slide the tissue segmentation algorithm was executed, then tiles were extracted from the tissue.
3. The above tiles were inferred through the tile classification algorithm and the probability vector for each tile is saved. The output for this step is a 3 dimensional matrix, where each point (x,y) in the matrix represents a tile and contains the probability vector for this tile.
4. Several geometric features are extracted from the 3 dimensional matrix (e.g. largest area in pixels where a specific class crosses a predefined threshold).
5. A machine learning model is built where the input for the model are the extracted features and the output is the slide diagnosis and the confidence for the diagnosis.

Several models were tested (including Support Vector Machines and Logistic Regression), where the model with the best performance was chosen.

***Inference & Testing Phase***

During inference, the first step is the tissue segmentation algorithm, then tiles are extracted from the tissue, each tile is classified using the tile classification algorithm, and that serves as an input to the final step of the slide level decision. The slide level decision algorithm outputs the diagnosis and confidence for the slide.

## Results

| Pathologist / AIIA | Normal colonic mucosa | Serrated polyp | Adenoma with low grade dysplasia | High risk polyp |
|---|---|---|---|---|
| **Normal colonic mucosa** | 86% (67/78) | 5% (4/78) | 9% (7/78) | 0% (0/78) |
| **Serrated polyp** | 12% (21/172) | 85% (146/172) | 2% (4/172) | 1% (1/172) |
| **Adenoma with low grade dysplasia** | 2% (9/551) | 1% (5/551) | 92% (508/551) | 5% (29/551) |
| **High risk polyp** | 0% (0/36) | 3% (1/36) | 3% (1/36) | 94% (34/36) |

**Table 1 –** AIIA multiclass confusion matrix as compared to a reference pathologist (n=837 slides). Overall observed unweighted kappa = 0.8181 (almost perfect agreement).

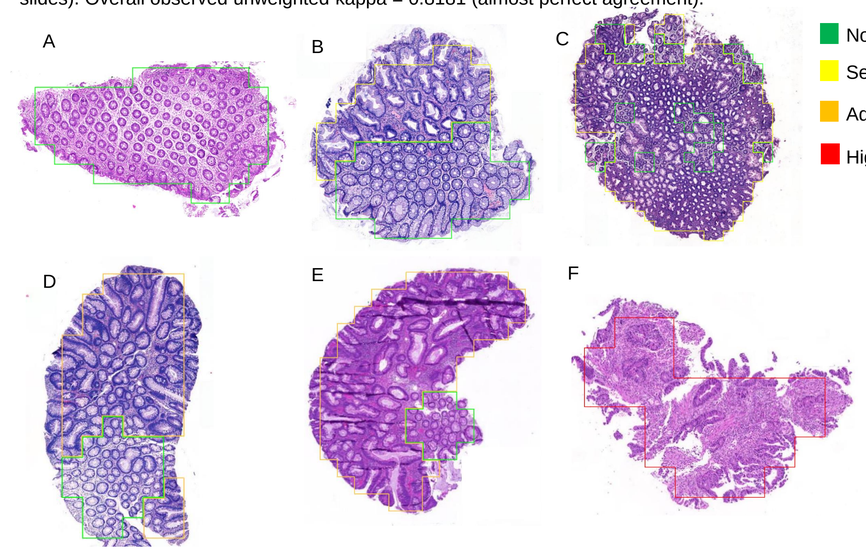


**Figure 2** - Digital display of a Colorectal Polyp H&E slide with automatically detection of findings by AIIA. The findings were marked by a color map as follows – green: normal colonic mucosa, yellow: serrated polyp, orange: adenoma with low grade dysplasia, red: high risk polyp.

Out of the 877 slides, 4.6% (n=40) slides were assessed as "unknown" by the algorithm. After exclusion of these slides, AIIA overall multiclass accuracy with the reference pathologist was 90.2% (ranging from 85% for serrated polyps to 94% for HG/malignant polyps) while binary classification concordance was 96.3% with a sensitivity and specificity of 96.3%. A subset of 219 cases were evaluated by 3 pathologists in which 198 had complete agreement. AIIA concordance with majority agreement diagnoses was 97% (multiclass) and 98% (binary), while concordance with complete agreement cases was 99% (multiclass) and 99.5% (binary). The majority of discrepant cases were between normal colonic mucosa and subtle hyperplastic polyps lacking significant serration.

## Conclusions

The findings demonstrate the capability of AIIA to successfully categorize colonic polyps. A potentially useful feature of the algorithm is that it can provide a digital display of scanned slides with superimposed heat maps thereby alerting the observer to areas of interest, such as high grade dysplasia. Therefore, AIIA may have a potential role in clinical diagnostic workflows in surgical pathology. AIIA could be used to make the diagnostic workflow substantially more efficient and cost effective, and to help reduce diagnostic error via enhanced quality assurance/control.

**References:**

1. Couture HD, Williams LA, Geradts J, Nyante SJ, Butler EN, Marron JS, Perou CM, Troester MA, Niethammer M. Image analysis with deep learning to predict breast cancer grade, ER status, histologic subtype, and intrinsic subtype. NPJ Breast Cancer. 2018 Sep 3;4:30.
2. Ehteshami Bejnordi B1, et al. Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. JAMA. 2017 Dec 12;318(22):2199-2210.
3. Martin DR, Braxton DR, Farris AB. Dysplasia discrimination in intestinal-type neoplasia of the esophagus and colon via digital image analysis. Virchows Arch. 2016 Oct;469(4):405-15.
4. Sirinukunwattana K, Snead DR, Rajpoot NM. A stochastic polygons model for glandular structures in colon histology images. IEEE Trans Med Imaging 2015;34:2366-78.
5. Sirinukunwattana K, Ahmed Raza SE, Yee-Wah Tsang, Snead DR, Cree IA, Rajpoot NM. Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. IEEE Trans Med Imaging 2016;35:1196-206.