**In vivo computer-aided diagnosis of colorectal polyps using white light endoscopy**

**Authors**
Ana García-Rodríguez1, Yael Tudela2, Henry Córdova1,3,4, Sabela Carballal1,3,4, Ingrid Ordás1,3,4, Leticia Moreira1,3,4, Eva Vaquero1,3,4, Oswaldo Ortiz1, Liseth Rivero1,3,4, F. Javier Sánchez2, Miriam Cuatrecasas3,4,5, Maria Pellisé1,3,4, Jorge Bernal2, Glòria Fernández-Esparrach1,3,4

**Institutions**
1 Endoscopy Unit. Gastroenterology Department. ICMDiM. Hospital Clínic of Barcelona. University of Barcelona, Barcelona, Catalonia, Spain
2 Computer Science Department. Autonomous University of Barcelona and Computer Vision Center, Barcelona, Catalonia, Spain
3 IDIBAPS, Barcelona, Catalonia, Spain
4 CIBEREHD, Spain
5 Pathology Department. Hospital Clínic of Barcelona. University of Barcelona, Barcelona, Catalonia, Spain

**ABSTRACT**

**Background and study aims** Artificial intelligence is currently able to accurately predict the histology of colorectal polyps. However, systems developed to date use complex optical technologies and have not been tested in vivo. The objective of this study was to evaluate the efficacy of a new deep learning-based optical diagnosis system, ATENEA, in a real clinical setting using only high-definition white light endoscopy (WLE) and to compare its performance with endoscopists.

**Methods** ATENEA was prospectively tested in real life on consecutive polyps detected in colorectal cancer screening colonoscopies at Hospital Clínic. No images were discarded, and only WLE was used. The in vivo ATENEA’s prediction (adenoma vs non-adenoma) was compared with the prediction of four staff endoscopists without specific training in optical diagnosis for the study purposes. Endoscopists were blind to the ATENEA output. Histology was the gold standard.

**Results** Ninety polyps (median size: 5 mm, range: 2–25) from 31 patients were included of which 69 (76.7 %) were adenomas. ATENEA correctly predicted the histology in 63 of 69 (91.3 %, 95 % CI: 82 %–97 %) adenomas and 12 of 21 (57.1%, 95% CI: 34%–78%) non-adenomas while endoscopists made correct predictions in 52 of 69 (75.4 %, 95 % CI: 60 %–85 %) and 20 of 21 (95.2 %, 95 % CI: 76 %–100 %), respectively. The global accuracy was 83.3 % (95 % CI: 74 %–90 %) and 80 % (95 % CI: 70 %–88 %) for ATENEA and endoscopists, respectively.

**Conclusion** ATENEA can accurately be used for in vivo characterization of colorectal polyps, enabling the endoscopist to make direct decisions. ATENEA showed a global accuracy similar to that of endoscopists despite an unsatisfactory performance for non-adenomatous lesions.

**Introduction**
Colorectal cancer (CRC) is the third most common cancer in both sexes and the second leading cause of death in the world [1]. Screening for CRC and the removal of neoplastic polyps by colonoscopy has led to a substantial improvement in survival [2]. Optical diagnosis aims to predict the histology of a polyp based on its endoscopic features. This practice could avoid pathological analysis in several cases and reduce the derived
processing times, which allows diagnosis in real time [13–16]. More recent studies, most of them DL approaches, have been prospectively conducted reporting higher accuracy with faster processing times, which allows diagnosis in real time [13–16]. However, to date, only CADx systems using complex optical technologies as endocytoscopy have been tested in vivo [13], that is, while the colonoscopy is being performed, which still hinders the adoption of this technology in daily practice. Furthermore, the newest CADx systems are currently using advanced imaging modalities, which clearly limits their implementation worldwide.

The quality and design of published CADx systems has varied over time. Initial studies were carried out retrospectively and were tested ex vivo using selected stored images [8–12]. More recent studies, most of them DL approaches, have been prospectively conducted reporting higher accuracy with faster processing times, which allows diagnosis in real time [13–16]. However, to date, only CADx systems using complex optical technologies as endocytoscopy have been tested in vivo [13], that is, while the colonoscopy is being performed, which still hinders the adoption of this technology in daily practice. Furthermore, the newest CADx systems are currently using advanced imaging modalities, which clearly limits their implementation worldwide.

The aim of this study was to evaluate the efficacy of a new CADx system based on DL called ATENEA for in vivo optical diagnosis in consecutive patients using only white light endoscopy (WLE) and compare its performance with endoscopists.

Methods

Development of ATENEA

As a first step, images of any polyp from routine colonoscopies performed at Hospital Clinic of Barcelona from January 2016 to December 2020 were prospectively collected. The images were captured with high-definition colonoscopes (CF-HQ180, CF-HQ-185, CF-HQ-190 and EVIS EXERA III videoprocessor, Olympus Europe, Hamburg, Germany) using an external computer with a frame grabber to ensure image acquisition with the highest quality. Only white light images without magnification or chromoendoscopy were used. Data from morphology, location, and size of the polyp were collected and periodically transferred by an assistant into the database along with the histological category of the lesion obtained after pathological analysis.

The development of ATENEA consisted of two stages: feature extraction and image classification. Feature extraction was made by using a faster region-based CNN with ResNet50 as backbone. For system training purposes, a region of interest (ROI) delineating the polyp was manually defined by clinicians using a program called GTCreator (Fig. 1) [17]. The extracted features, ROI, and actual histologic class of the polyp were used to train the system. ATENEA learned to classify images into adenoma or non-adenoma categories using an 80% (high confidence) threshold value (predictions with a confidence value < 80% were not considered valid to represent the actual performance of the system as they could not guarantee a robust performance).

ATENEA was trained and validated with a total of 1049 high-definition white light images of 483 polyps from 354 patients, with a maximum of three images of the same polyp (but with a different view or perspective). Images had variable quality but all had a visible mucosal pattern and only blurred images and polyps covered by mucus were excluded. Images containing patient data or without histological analysis were also excluded. About two-thirds were adenomas and one-third non-adenomas, following a similar proportion to what is found in real life, with a median size of 4.5 mm (range 2 mm–20 mm). These images were randomly split into training (n = 837) and validation sets (n = 212), with the condition that all the images of the same polyp were in the same set.

In vivo experiment

This observational prospective cohort study was performed in Hospital Clinic of Barcelona between January and March 2021 and included individuals from the fecal immunochemical test-based (cut-off of ≥20 μg of hemoglobin/g of feces) organized population CRC screening program in which all individuals aged 50 to 69 years are invited to participate. Colonoscopies were performed by four staff endoscopists with more than 1000 colonoscopies and adenoma detection rates of 47%,
During the study period, 90 polyps from 31 consecutive patients were included. Sixty-nine (76.7%) were adenomas with a median size of 5.0 mm (range: 2–25 mm). Characteristics of the polyps are described in Table 1. ATENEA provided an output for all 90 polyps and was able to correctly predict the histology in 63 of 69 adenomas (sensitivity: 91.3%, 95% CI: 82.9–97%) and 12 of 21 non-adenomas (specificity: 57.1%, 95% CI: 34.8–78%). Endoscopists were able to provide a correct optical diagnosis in 52 of 69 adenomas (sensitivity: 75.4%, 95% CI: 64.5–85%) and 20 of 21 non-adenomas (specificity: 95.2%, 95% CI: 76.1–100%), with an accuracy of 83.3% (95% CI: 74–90%) and 80% (95% CI: 70–88%) for CADx and endoscopists, respectively (Table 2).

With respect to diminutive polyps (≤5 mm), ATENEA was able to correctly predict the histology in 31 of 35 adenomas (sensitivity: 88.2%, 95% CI: 73–97%) and 11 of 18 non-adenomas (specificity: 61.1%, 95% CI: 36–83%) and endoscopists in 20 of 35 adenomas (sensitivity: 58.8%, 95% CI: 41–75%) and 17 of 18 non-adenomas (specificity: 94.4%, 95% CI: 73–100%), respectively, with an accuracy of 78.8% (95% CI: 65–89%) and 71.1% (95% CI: 57–83%) for ATENEA and endoscopists, respectively.

**Results**

All polyps were globally analyzed, independently of the endoscopist who performed the exploration. The numbers of polyps that were true positive (adenomatous polyps predicted to be adenomatous), true negative (non-adenomatous polyps predicted to be non-adenomatous), false positive (non-adenomatous polyps predicted to be adenomatous) or false negative (adenomatous polyps predicted to be non-adenomatous) were calculated. Sensitivity, specificity, positive predictive value, negative predictive value and accuracy with 95% confidence intervals were calculated. Comparisons of these metrics between ATENEA and endoscopists were performed with Chi squared test.

Receiver operator characteristics (ROC) curves for use of different prediction confidence values to determine ATENEA in vivo performance for diagnosis of adenoma were constructed with Matlab. The area under the curve (AUC) and optimal operating point with its sensitivity and specificity and 95% confidence intervals were calculated.

Two-sided P<0.05 was considered statistically significant. All calculations were performed with R version 4.1.2 and Matlab for Windows version 2020.

**Histopathology**

All polyps were removed using the usual techniques and sent separately for further evaluation to the Pathology Department, which was the gold standard. The diagnosis of dysplasia in neoplastic polyps was made based on the Modified Vienna Classification [19]. The pathologist was blinded to the predictions of both endoscopist and ATENEA.

**Statistical analysis**

All the polyps were globally analyzed, independently of the endoscopist who performed the exploration. The numbers of polyps that were true positive (adenomatous polyps predicted to be adenomatous), true negative (non-adenomatous polyps predicted to be non-adenomatous), false positive (non-adenomatous polyps predicted to be adenomatous) or false negative (adenomatous polyps predicted to be non-adenomatous) were calculated. Sensitivity, specificity, positive predictive value, negative predictive value and accuracy with 95% confidence intervals were calculated. Comparisons of these metrics between ATENEA and endoscopists were performed with Chi squared test.

Receiver operator characteristics (ROC) curves for use of different prediction confidence values to determine ATENEA in vivo performance for diagnosis of adenoma were constructed with Matlab. The area under the curve (AUC) and optimal operating point with its sensitivity and specificity and 95% confidence intervals were calculated.

Two-sided P<0.05 was considered statistically significant. All calculations were performed with R version 4.1.2 and Matlab for Windows version 2020.

**Results**

During the study period, 90 polyps from 31 consecutive patients were included. Sixty-nine (76.7%) were adenomas with a median size of 5.0 mm (range: 2–25 mm). Characteristics of the polyps are described in Table 1. ATENEA provided an output for all 90 polyps and was able to correctly predict the histology in 63 of 69 adenomas (sensitivity: 91.3%, 95% CI: 82.9–97%) and 12 of 21 non-adenomas (specificity: 57.1%, 95% CI: 34.8–78%). Endoscopists were able to provide a correct optical diagnosis in 52 of 69 adenomas (sensitivity: 75.4%, 95% CI: 64.5–85%) and 20 of 21 non-adenomas (specificity: 95.2%, 95% CI: 76.1–100%), with an accuracy of 83.3% (95% CI: 74–90%) and 80% (95% CI: 70–88%) for CADx and endoscopists, respectively (Table 2).

With respect to diminutive polyps (≤5 mm), ATENEA was able to correctly predict the histology in 31 of 35 adenomas (sensitivity: 88.2%, 95% CI: 73–97%) and 11 of 18 non-adenomas (specificity: 61.1%, 95% CI: 36–83%) and endoscopists in 20 of 35 adenomas (sensitivity: 58.8%, 95% CI: 41–75%) and 17 of 18 non-adenomas (specificity: 94.4%, 95% CI: 73–100%), respectively, with an accuracy of 78.8% (95% CI: 65–89%) and 71.1% (95% CI: 57–83%) for ATENEA and endoscopists, respectively.
endoscopists, respectively (Table 2). Results of prediction for small polyps (<10 mm) are also shown in Table 2.

ATENEA and the endoscopists disagreed in their prediction in 31 of 90 cases (34.5%). Endoscopists made their prediction with high confidence in 79 cases (87%) and with low confidence in 11. In all these 11 cases, ATENEA made a good prediction (Fig. 3).

The ROC curve for ATENEA in vivo performance (Fig. 4) showed an AUC of 0.782. The optimal operating point was achieved by using 74.2% as the threshold value with a sensitivity of 86.9% (95% CI: 79.9%–93.9%) and a specificity of 66.7% (95% CI: 56.9%–76.4%).

Discussion

This study demonstrates that ATENEA, a fully automatic optical diagnostic system, can accurately predict polyp histology in an in vivo setting using only white light, which answers one of the 10 key research questions identified by international experts related to AI implementation in colonoscopy [20]. ATENEA is particularly useful for identifying adenomatous lesions, showing its readiness to be used in a clinical environment.

Until now, the majority of CAD systems developed for characterizing colorectal polyps that use AI have required advanced optical diagnostic equipment, such as narrow band imaging (NBI) [9–12, 21] and endocytoscopy [8, 22, 23]. However, WLE, which is the most common endoscopic modality, has not been

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Characteristics of the polyps included in the study.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>In vivo test (n = 90)</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Histological type</strong></td>
<td>Adenoma 69 (76.7%)</td>
</tr>
<tr>
<td><strong>Size (in mm)</strong></td>
<td>≤ 5 mm 52 (57.8%)</td>
</tr>
<tr>
<td><strong>Location</strong></td>
<td>Rectum-sigma 22 (24.4%)</td>
</tr>
<tr>
<td><strong>Morphology</strong></td>
<td>0-Ip 5 (5.5%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Performance characteristics of ATENEA and the endoscopists for diagnosis of adenoma.</th>
</tr>
</thead>
<tbody>
<tr>
<td>**TP</td>
<td>FP</td>
</tr>
<tr>
<td>ATENEA, n = 90</td>
<td>63</td>
</tr>
<tr>
<td>Endoscopists, n = 90</td>
<td>52</td>
</tr>
<tr>
<td><strong>P value</strong></td>
<td>0.07</td>
</tr>
<tr>
<td>ATENEA diminutive polyps ≤ 5 mm, n = 52</td>
<td>30</td>
</tr>
<tr>
<td>Endoscopists diminutive polyps ≤ 5 mm, n = 52</td>
<td>20</td>
</tr>
<tr>
<td><strong>P value</strong></td>
<td>0.27</td>
</tr>
<tr>
<td>ATENEA small polyps &lt; 10 mm, n = 73</td>
<td>47</td>
</tr>
<tr>
<td>Endoscopists small polyps &lt; 10 mm, n = 73</td>
<td>36</td>
</tr>
</tbody>
</table>
| **P value** | 0.09 | 0.02 | 0.46 | 0.01 | 0.69 | TP, true positive; FP, false positive; TN, true negative; FN, false negative; PPV, positive predictive value; SENS, sensitivity; NPV, negative predictive value; Spec, specificity.
extensively studied in this context yet. For this reason, we used only white light images to allow for widespread use of ATENEA regardless of the manufacturer and model of endoscope used.

Our group previously developed a hand-crafted predictive model based on extraction of surface patterns (textons) over white light images. This system was validated in a small dataset containing only 225 high-quality images with a diagnostic accuracy of over 90% [24]. Unfortunately, the system was not fully automatic and could not perform in real time, and consequently, it could not be used in an in vivo experiment. In contrast, ATENEA can perform in vivo automatic polyp classification due to its capability of calculating the result in less than 40 ms (real time). It is important to make a difference between what real-time and in vivo means: the first refers to the time the system takes to process an image, whereas the second is related to where and when the system is applied (in vivo: in the exploration room versus ex vivo: off-line experiments). These concepts are commonly confused in the literature and, in most of the cases, real-time is used instead of in vivo without mentioning the actual processing time [13, 14].

The results obtained in the present experiment showed that ATENEA had better global accuracy than endoscopists, particularly for adenomatous lesions. Conversely, clinicians performed better for the non-adenoma category. It has to be pointed out that they had more information than the system because they knew the location and size of the lesion (variables for which ATENEA was blind), so this could lead to a major pretest probability of better diagnosis of non-adenomatous lesions when faced with a small or diminutive polyp in the rectosigmoid colon. In contrast, ATENEA performance depended solely on the number and variability of examples from each of the classes in the training set.

The lower performance of ATENEA for non-adenomatous lesions could limit implementation of a "leave in situ" policy. However, it must be stated that the training of the AI system was greatly affected by the reduced number of examples of this class in the dataset. There are two reasons for this: first, the prevalence of non-adenomatous lesions is generally lower and second, they are not systematically removed when located in the rectosigmoid colon, making their collection difficult. Hence, it is necessary to enlarge the dataset both in quantity and percentage of examples of the minority class.

CADx systems are not intended to replace endoscopists, but rather, to help them in their tasks. As stated by the ESGE review on advanced endoscopic imaging, the most likely scenario is that the intelligent systems will be used as a "second reader" to support the endoscopist’s final diagnosis [25]. In this sense, an important result from our study was that if endoscopists...
with low confidence had followed the output of the system, all their predictions would have been correct. This shows the potential of ATENEA to assist clinicians with less expertise.

Similar to endoscopists, AI systems do also provide a confidence value in their predictions using a percentage instead of a binary assessment (high versus low). By varying the confidence value, we represented different performances of ATENEA using a ROC curve. Our results show that if the confidence value is low, the sensitivity of the system is lower as it provides more outputs, some of them erroneous. Conversely, if we increase the allowed confidence value, less but more confident outputs will be provided, with the risk of missing some of the lesions without a prediction. The so-called Simple Optical Diagnosis Accuracy or SODA criteria, which were recently published by the ESGE Curricula Working Group [26], are more flexible than the PIVI criteria and emphasize the importance of not leaving any diminutive lesion with advanced neoplasia in situ. In accordance with SODA criteria, we considered sensitivity to be the most important outcome of the intelligent system. The AUC value obtained in our study is similar to the 0.84 reported in a recent meta-analysis using only WLE [27] (which is usually lower than the AUC in studies using chromoendoscopy or magnification techniques).

Unlike other studies in which low-quality images or the known “difficult for AI cases” were not included, the main strength of this study is that it was performed under real clinical practice conditions and polyps were included regardless of image quality (not centered polyps, blurred or covered by mucus). If we had excluded these polyps from the analysis, the ATENEA performance would have been better. Nevertheless, the lack of publicly available annotated datasets does not allow for a fair comparison between ATENEA and other CADx systems. In this sense, comparison of metrics in meta-analysis are difficult to understand as the number, quality, and class of the images are different in each study and affect training and testing stages of the validation of computational systems.

The study has the following limitations. First, the training dataset was small in terms of number of different polyps. To mitigate this, we used pre-trained weights from ImageNet [28], which is a more general dataset and we applied data augmentation operations (color transformation, rotation, horizontal and vertical flip) to enlarge its size. The collaboration with other centers and the public availability of other datasets could also be of great use to both enlarge the dataset and perform multicenter validations, increasing the robustness of the results. Second, our study did not consider a separate class for SSL due to the low number of examples in the dataset. The problem of SSLs has not adequately been addressed in other previous studies. It is well known that SSLs are neoplastic lesions and there is not an ideal optical method for their characterization [29]. Due to the clinical relevance of SSLs, some studies propose a classification between neoplastic and non-neoplastic lesions instead of using adenoma vs. non-adenoma. Following this logic, if we had included all SSL in the same category as adenomas, four of five SSLs would have been correctly classified.

Conclusions

In conclusion, ATENEA, a CADx system based only on WLE data, is ready to be used for in vivo and real-time characterization of colorectal polyps, enabling the endoscopist to make direct decisions. ATENEA achieved a global accuracy similar to endoscopists, despite having lower performance for non-adenomatous lesions.

Acknowledgments

This work was supported by the Instituto de Salud Carlos III (PI17/00894 and PI19/01050), co-funded by the European Union; the Spanish Ministry for Science and Innovation (MCIN) (ALETHEIA project, PID2020-120311RB-I00 funded by MCIN/AEI/10.13039/501100011033); Fundació Marató de TV3 (201932–30); Secretaria d’Universitats i Recerca de la Generalitat de Catalunya (2014-SGR-1470, 2014-SGR-135, SGR-2017–1669 and SGR-2017–653), and by CERCA Programme/Generalitat de Catalunya. Ana García-Rodríguez received a personal grant from the Societat Catalana de Digestologia. The authors thank Dr. Andrés Cárdenas for his English medical writing assistance and Roger Borrás for performing the statistical analysis.

Competing interests

Gloria Fernández-Esparrach was employed on an advisory board of Medtronic and is shareholder of MiWEndo Solutions S.L. Miriam Cuatrecas is shareholder of MiWEndo Solutions S.L. Maria Pellisé has served on clinical advisory boards for Fujifilm Europe; has served on the clinical advisory board and owns share options in MiWendo S.L.; reports speaker fees from Casen Recordati, Fujifilm, Medtronic and Olympus Europe; and has received research funding from Fujifilm Europe, Casen Recordati and Ziuz.

Funding

Fundació la Marató de TV3 201932-30
Iniciació a la recerca de la Societat Catalana de Digestologia
Generalitat de Catalunya CERCA Programme
Instituto de Salud Carlos III PI17/00894,PI19/01050,PID2020-120611RB-I00
Spanish Ministry for Science and Innovation (MCIN) PID2020-120311RB-I00

Clinical trial

ClinicalTrials.gov
NCT03775811
TRIAL REGISTRATION: Prospective trial at http://www.clinicaltrials.gov/
References


[27] Lui TKL, Guo CG, Leung WK. Accuracy of artificial intelligence on histology prediction and detection of colorectal polyps: a systematic review and meta-analysis. Gastrointest Endosc 2020; 92: 11–22.e6
