Abstract. Colorectal cancer (CRC) is the third most common cancer in both men and women. It is estimated that in 2004, nearly 147,000 cases of colon and rectal cancer will be diagnosed in the US, and approximately 57,000 people would die from the disease, accounting for about 11% of all cancer deaths. While there is wide consensus that screening patients is effective in decreasing advanced disease, only 44% of the eligible population undergoes any type of colorectal cancer screening. Multiple reasons have been identified for non-compliance, key being: patient comfort, bowel preparation and cost. Virtual Colonoscopy (VC) derived from computer tomographic (CT) images is gaining broader acceptance as a screening method for colorectal neoplasia. Our research suggests that Computer Aided Detection (CAD), as a second reader, has great potential in improving polyp detection. The developed CAD system, presented in this paper, has focused on the detection of polyps of sizes up to and including 20mm. Results of this study, with 150 patients, demonstrate that: the developed algorithm generalizes, the sensitivity for middle- to large-sized polyps is on the average 90% while the overall sensitivity is roughly 82%. The false positive rate is a manageable 4.5 per volume on average.

Keywords: CAD, automated polyp detection, virtual colonography, colon screening.

1. Introduction

Colorectal cancer is the third most common cancer in both men and women. In [1] it is estimated that in 2004, nearly 147,000 cases of colon and rectal cancer will be diagnosed in the US, and more than 56,730 people would die from colon cancer, accounting for about 11% of all cancer deaths. Table 1 illustrates the relation between early detection of colon cancer and five-year survival rate. In particular, since it is known that in over 90% of cases the progression stage for colon-cancer is from local (polyp adenomas) to advanced stages (colorectal cancer), it is really critical that major efforts be devoted to screening of colon cancer and remove lesions (polyp) when still in a local stage of the disease. In [2] a guideline for CRC screening is presented while a guideline for patient management based on CT Colonography (CTC) is given in [3].

<table>
<thead>
<tr>
<th>Staging of colon cancer at detection</th>
<th>Localized</th>
<th>Regional</th>
<th>Distant</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 year survival rate ('92–'99)</td>
<td>90%</td>
<td>66%</td>
<td>9%</td>
<td>62%</td>
</tr>
<tr>
<td>Diagnosed cancers at time detection</td>
<td>38%</td>
<td>38%</td>
<td>19%</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Colon cancer staging and survival rate & distribution of staging at time of diagnosis. This table, integrating information on Colon & Rectal from the Cancer Statistics 2004 [1], clearly illustrates the importance of early diagnosis and treatment of colon cancer.

As evidence of the impact of removing lesions, a study [4] on 1693 patients, followed over a period of ten years, demonstrated that colonoscopic polypectomy substantially reduced the incidence of colorectal cancer in the cohort when compared with that expected in the general population. While there is wide consensus that screening patients [5] is effective in decreasing advanced disease, only 44% of the eligible population undergoes any type of CRC screening [6].
Multiple reasons have been identified for non-compliance, key being: patient comfort, bowel preparation and cost [7]. As CTC is gaining broader acceptance as a screening method for colorectal neoplasia [8][9], results presented in this paper suggest that, in a rather near future, CAD could be employed as a second reader to aid in polyp detection.

This article first reviews a few studies that have demonstrated radiologists’ sensitivities when performing CTC. The Siemens’ Colon CAD system is presented as an integrated tool in the context of a clinical workflow. The following sections then present a study involving two centers and 150 patients demonstrating both the high level sensitivity and generalizability reached by the developed CAD system. The article concludes with a discussion on the performance of this CAD system when compared to physician’s sensitivity and to other CAD systems being developed.

2. CTC in studies and clinical practice

Virtual Colonography was introduced as a new practice by Vining [10] in 1994. Since then this technique has received increasing attention. Studies over the past five years, Table 2, have demonstrated radiologists’ ability to detect polyps using CT-Colonography (CTC). Specifically, sensitivity for small polyp is at most 60%, ranges between 70%-90% for medium-sized polyps and 90%-100% for large polyps. Thus, we can observe that physician’s sensitivity, while depending on many factors – patient preparation, distension and insufflation, image quality, level of expertise of the reader – is strongly correlated to the size of the polyps. This sensitivity correlates well with the clinical relevance of the lesions. In fact, it has been shown [11] that the malignancy likelihood for small polyps is less than 0.1%. These facts suggest that a CAD system, while attempting to achieve high overall sensitivity, would be very well served if it could provide highest level of sensitivity for the middle to large polyps thus reflecting the clinical relevance based on size as well as experts’ performance in the middle to large sizes.

<table>
<thead>
<tr>
<th>CTC Studies</th>
<th># Patients (polyps)</th>
<th>Small &lt; 6mm</th>
<th>Medium ≥ 6&amp; &lt;10</th>
<th>Large ≥ 10mm</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Felon 1999[12]</td>
<td>100(115)</td>
<td>55%</td>
<td>82%</td>
<td>91%</td>
<td>71%</td>
</tr>
<tr>
<td>Yee 2001 [13]</td>
<td>300(223)</td>
<td>59%</td>
<td>80%</td>
<td>90%</td>
<td>70%</td>
</tr>
<tr>
<td>Macari 2002 [14]</td>
<td>52(132)</td>
<td>12%</td>
<td>70%</td>
<td>93%</td>
<td>33%</td>
</tr>
<tr>
<td>Iannacone 2003 [15]</td>
<td>158(74)</td>
<td>51%</td>
<td>83%</td>
<td>100%</td>
<td>70%</td>
</tr>
<tr>
<td>Pickard 2003 [8]</td>
<td>1233(550)</td>
<td>89%</td>
<td>92%</td>
<td>89%</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Studies showing the sensitivity across different polyp categories. The sensitivity for middle and large polyps are the most important since these have strongest clinical relevance.

Recently, a meta-analysis study by Sosna [9] considered 14 studies, with a combined pool of 1324 patients and 1,411 polyps. The study reported that the per-polyp sensitivity was 81%, 62% and 43% for large, medium and small polyps respectively. The studies included in the Sosna’s meta-analysis study span a broader variety of studies and conditions but the results are consistent with those presented in Table 2.

These studies offer strong evidence of the value provided by CTC. However, given the large band of the population who could benefit from early detection (Table 1), CTC must become a ubiquitous practice. Therefore, it is necessary that the average community radiologist be able to reproduce the high levels of sensitivities in polyp detection shown in Table 2. As a study, presented in this paper and others suggest, Table 5, suggest a CAD tool (used as a second reader) could be employed to fulfill this role toward the detection of polyps.

3. CAD as second reader

CAD systems for mammography, R2-Tech’s ImageCheck™, lung nodule detection – R2-Tech’s LungCheck™, Deus Technologies, and Siemens’ LungCAD™ - are playing larger roles in the
automated detection arena allowing physicians to review cases more efficiently and with increased level of sensitivity. These systems are slowly evolving and demonstrating value not only in detection but also in proposing diagnostic evidence, using different type of markers, for the physician to determine the diagnostic value of the identified finding.

Figure 1: This figure shows, a snap-shot the prototype Colon CAD incorporated in the Siemens’ syngo Colonography at the moment that physician, after completing the review of the case is ready to review the additional CAD findings. The figures show (in addition to menus options on the top and right-hand-side), in the two top quadrants, coronal and axial MPRs of a supine view of CT dataset. The bottom left quadrant shows a global view of the colon in a translucent rendering onto which some findings located by the physicians appear (marks: 5a in yellow, 7a & 8a in red). The bottom right quadrant shows an endo-view for a given position of the colon.

In the area of colon cancer detection from CT images, the key aspect of a CAD system is the improvement in sensitivity with respect to detection of polyps that such system can offer. In order to put the role of colon CAD as a second reader in proper context, we outline it in the context of a complete clinical workflow.

A workflow with integrated Colon CAD system would consist of the following 4 stages:

- **Case Loading**: physician loads the case for review – CAD system begins processing in the background.
- **First read**: physician reviews the case, prone and supine, finalizes its findings.
- **CAD results are invoked** (CAD button is pressed): physician acknowledges she/he has completed the review of the case, illustrated in Figure 1.
• **Second read**: physician reviews additional CAD findings, e.g. shown in Figure 2, and rejects any considered false positives.

Figure 1 and Figure 2 illustrate the two key moments in the workflow for the Colon CAD prototype, as it would appear when incorporated as part of the Siemens’ syngo Colonography prototype.

**Figure 2**: The bottom right quadrant shows clearly the endo-view with one of the potential lesions found by the CAD system that had been missed by the physician. This finding is labelled “c9a” where the prefix “c” differentiates it from the physician findings. The bottom left quadrant (global view) shows other potential findings, also prefixed by the letter “c”.

4. Colon CAD: Study and Performance

The performance of the developed CAD system is presented next. It was determined as part of a study involving 150 datasets obtained from two sites NYU Medical Center (NYU) and the Cleveland Clinic Foundation (CCF).

4.1 Methods and Materials

The database consisted of 150 datasets, 292 volumes from high-resolution CT scanners. These included both patients with polyps (positive cases) (n=64) and patients without polyps (negative cases) (n=86). These cases were partitioned into working set (training set) and unseen set (test set). The sensitivity and specificity (number of false positives) of CAD as a tool to aid in polyp detection was established with respect to CTC by comparison to results from concurrent fiber-
optic colonoscopy. Cases were de-identified (all patient identification information was removed) and then exported to CD. The locations and dimensions of the lesions were then used in subsequent stages to automatically compute sensitivity and specificity with respect to polyp size. The patient protocols and acquisitions parameters are shown in Table 3.

**Patient preparation protocol**

<table>
<thead>
<tr>
<th></th>
<th>bowel preparation</th>
<th>insufflation</th>
</tr>
</thead>
<tbody>
<tr>
<td>NYU</td>
<td>phosphor-soda + evacuation prior to CT scans</td>
<td>room air (manual)</td>
</tr>
<tr>
<td>CCF</td>
<td>phosphor-soda (fleet) or Go-litely</td>
<td>room-air (manual)</td>
</tr>
</tbody>
</table>

**Acquisition parameters**

<table>
<thead>
<tr>
<th></th>
<th>System</th>
<th>KV</th>
<th>rotation</th>
<th>effective mAs</th>
<th>slice collimation</th>
<th>slice width</th>
<th>reconstr. increment</th>
<th>kernel</th>
</tr>
</thead>
<tbody>
<tr>
<td>NYU</td>
<td>Volume-Zoom</td>
<td>120kV</td>
<td>0.5sec</td>
<td>50mAs</td>
<td>0.75mm</td>
<td>1.5mm</td>
<td>1.0mm</td>
<td>B30f</td>
</tr>
<tr>
<td>CCF</td>
<td>Sensation 16</td>
<td>120kV</td>
<td>0.5sec</td>
<td>50mAs</td>
<td>1.0mm</td>
<td>1.25</td>
<td>1.0mm</td>
<td>B30f</td>
</tr>
</tbody>
</table>

Table 3: Patient preparation protocol and acquisition parameters for the datasets from NYU and CCF.

Our proprietary ColonCAD algorithm (patents pending) included the following phases: data pre-processing, candidate generation, feature extraction, and classification. The data pre-processing phase includes colon segmentation and a transformation to isotropic volume. During candidate generation, based on a simple shape filter, loci of detection are identified. These are sequentially processed during the following phase in which multiple features are extracted. The features are based on moments of tissue intensity, volumetric and surface shape as well as texture characteristics. Each candidate, uniquely identified, and the associated features are then fed to a classifier. Candidates are then evaluated and labelled as potential polyps. In the actual workflow, these would then be presented to the physician for review. The primary focus of the developed algorithms has been that of yielding high sensitivity and specificity. At present, the running time on a single volume, 600 slices with 512x512 axial resolution, was on average 4 minutes on a Pentium IV 3.06GHz dual-processor machine with 2GB of memory.

In order to automatically process all the cases, a flexible framework was developed which would allow loading the cases (read as DICOM images), and push them through the various stages of the algorithm, outputting intermediary results. The modularity of the system has allowed assessing the effectiveness of each component and hence both improve and independently refine them.

### 4.2 Feature Selection and Classification

**Training and Test Data:** The 150 datasets were randomly partitioned into two groups: training (n=88) and test (n=62) sets. The test-set was sequestered and only used to evaluate the performance of the final system. In order to automate the training and verification process, a database was developed to allow the software to automatically query the database and provide feedback as to whether the finding was a polyp or non-polyp and, in the case of polyps, also obtain its sizes. The training-set was used to design the classifier and to automatically select the relevant features as described below.

**Feature Selection:** The feature selection stage is a key component of our approach. We use the “wrapper” method for feature selection [23] in which the classifier decides which features are useful. Procedurally, the classifier is run iteratively on the training-set using different feature sets – during each iteration, one or more features are added, until the cross-validation error no longer improves.

**Test Results:** The system’s performance was evaluated on the 88 cases in the training-set using Leave-One Patient Out (LOPO) cross-validation. In this scheme, both the supine and prone views of a case from the training-set were left out. The classifier is trained using the volumes from the remaining 87 (i.e., 88-1) case, and tested on both volumes of the “left-out” case. This process was repeated 88 times, leaving out each of the 88 cases in the training set, and the resulting testing errors were averaged out to determine the LOPO error.
Results on Test Group: A classifier was trained on all 88 cases in the training-set using the 15 features selected in the "Feature Selection" phase. Only after this classifier was frozen, was the test group of 62 cases released for evaluation. This provided a very accurate estimate of system performance on completely unseen data – the only true test for a classification-based system. In the computation of the sensitivity, a polyp was considered as "found" if it was detected in at least one of the volumes (supine or prone) from the same patient.

4.3 Results

In this section we review the data characteristics and specific results both for training and test set. These are also summarized in Table 4.

4.3.1 Results on Training-Set

**Patient and Polyp Info:** There were 88 cases with 171 volumes (some cases did not have both prone and supine studies). A total of 53 polyps were identified in this set: 19 small-size (1-5 mm), 25 mid-size (6-9 mm), 9 large-size (10-20 mm).

**Candidate Generation:** The candidate generation stage generated an average of 48.2 candidates per volume while missing 3 small-size polyps.

**Classifier Results:** We obtained a false positive rate of 4.4 per volume. The sensitivities obtained for different ranges of polyps are as follows: small-sized = 63.1%, mid-sized = 92.0%, large sized = 88.9%, overall = 81.1% and overall >= 5mm 91.2%.

4.3.2 Results on Test-Set (previously unseen)

**Patient and Polyp Info:** There were 62 cases with 121 volumes. A total of 39 polyps were identified in this set: 18 small-size, 11 mid-size, 10 large-size.

**Candidate Generation:** The candidate generation stage generated an average of 51.4 candidates per volume while missing 5 small and 1 medium polyps.

**Classifier Results:** We obtained a false positive rate of 4 per volume. The sensitivities obtained for different ranges of polyps are as follows: small-sized = 66.7%, mid-sized = 81.8%, large sized = 100%, overall = 82.1% and overall >= 5mm 90.5%.

<table>
<thead>
<tr>
<th>#Patients (polyps) (fp)</th>
<th>Candidate Generation</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>#polyps_found_CG / #polyps</td>
<td>(#polyps_found_CG - missed_polyps)#/polyps</td>
</tr>
<tr>
<td></td>
<td>Small</td>
<td>Medium</td>
</tr>
<tr>
<td>Training 88(53)[9]</td>
<td>16/19</td>
<td>25/25</td>
</tr>
<tr>
<td>100%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Testing 62(39)[8]</td>
<td>13/18</td>
<td>10/11</td>
</tr>
<tr>
<td>95.2%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 4:** CAD Performance: The table summarizes the performance of the CAD system as indicated in terms of the two phases: Candidate Generation and Classification. The first row characterizes the training phase, while the second row captures the testing phase. For each of the polyp categories (small, medium, large), the sensitivity is expressed first as a fraction of polyps found over the total number of polyps and next as the percentage. In first row of the first column from the major column labelled candidate generation (CG), the 16/19 indicates that of the 88 cases used in the training phase, there were 19 small polyps of which 3 were missed during the CG phase. In the corresponding classification phase, the entry (16-4)/19 shows that 4 small polyps were additionally missed during this phase. The combined sensitivity for middle and large polyps are highlighted. Further analysis is presented in the text of the document.
5. Discussion

In the area of colon cancer detection from CT images, the key aspect of a CAD system is the improvement in sensitivity that such system can offer without adding too many false positives. The sensitivity of a CAD system’s can be determined by two types of studies:

- **Physician-comparative** - a CAD system is run on a set of cases, its sensitivity is computed and compared to the performance of physicians on: (a) the same set cases or (b) other study with similar case characteristics. Goal: to demonstrate a level of performance similar to that of physicians, and in particular to experts. If the sensitivity of CAD is similar to that of expert physicians, it can be inferred that: since there is a difference in sensitivity between experts and non-experts (community radiologist, gastro-enterologists, etc.), CAD should prove to be of help to non-expert by raising their sensitivity if they could use CAD (as a second reader).

- **Physician-additive** – a CAD system functions as a second reader. Goal: establish CAD’s additive value for a reader, be it expert or non-expert. This study can only be performed in conjunction with a reader reviewing the case first and then accepting CAD’s findings. Okamura [24] reports a first study in which CAD was been demonstrated to add value to all physicians regardless of the level of expertise when used as a second reader.

To date, with the exception to [24], the few studies involving CAD can be categorized as physician-comparative (b) and the results we presented fit in this category.

The CAD results presented compare, on the overall, favourably with other CAD systems for polyp detection, see Table 5. A direct comparison across the different CAD system is, however, difficult given validation methods, patient selection and preparation, data acquisition protocols and quality, and how the information reported on sensitivity (per-patient, per polyp, per volume) and computation of false positive vary.

<table>
<thead>
<tr>
<th>CAD Studies</th>
<th>Test Method</th>
<th># patients (polyps)[fp]</th>
<th>Polyp Sensitivity (found/#polyps)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Small &lt; 6mm</td>
</tr>
<tr>
<td>NIH 2003 [16]</td>
<td>Independent set trained</td>
<td>40(20)[3]</td>
<td>unk.</td>
</tr>
<tr>
<td>Chicago 2002 [18]</td>
<td>LOPO</td>
<td>71(35)[2]</td>
<td>0</td>
</tr>
<tr>
<td>Stanford 2001[19]</td>
<td>10-fold X-validation</td>
<td>40(40)</td>
<td>unk.</td>
</tr>
<tr>
<td>WFU 2004 [20]</td>
<td>unk.</td>
<td>(34)</td>
<td>unk.</td>
</tr>
<tr>
<td>Leuven 2003 [21]</td>
<td>unk.</td>
<td>50(33)[2.5]</td>
<td>2/6 (33%)</td>
</tr>
<tr>
<td>Siemens (training)</td>
<td>LOPO</td>
<td>88(53)[9]</td>
<td>12/19 63%</td>
</tr>
<tr>
<td>Siemens (testing)</td>
<td>621(39)[8]</td>
<td>14/18 67%</td>
<td>9/11 82%</td>
</tr>
</tbody>
</table>

**Table 5:** CAD Studies: The development of ColonCAD system have only begun recently and to date only a limited number of studies have begun to assess CAD sensitivity beyond just proving its feasibility on a handful of patients. Some of the tabulated values were estimated from the information available in the literature and information for entries marked as unk. (unknown) was not available.

The ability of the radiologist, and of the present CAD system, to detect polyps greatly depends on colon distension, insufflation and cleansing. Therefore, while it is reasonable to perform quality control on the data being processed for major breathing artefacts and poorly prepped cases, a
system intended to work in clinical setting should work on cases deemed suitable for diagnostic review or provide a graceful degradation. The data sets, we received from NYU and CCF, had, on the overall, good bowel preparation and distension – except for some cases in which in which the colon presented some tortuosity and stenosis or had collapsed. There were, also, some artefacts due to prosthetic implants, breathing as well as minimal peristaltic motion. Thus, while we noticed that these cases affected the overall performance, both reducing the sensitivity as well as increasing the false positives, all were kept in both training and test sets so as to comply with our desire to develop a CAD system which would (a) closely resemble the type of cases reviewed in a clinical settings and (b) be robust to future testing when presented with such cases.

The number of false positive remained at a manageable 4/volume often occurring in the small intestine and ileocecal valve – these can be reduced by more accurate pre-processing; however, these findings are readily eliminated by radiologists. Another source of false positive is due to the erroneous labelling of fecal matter, which may not be easily differentiated based on density or morphology. In such cases, it is assumed that motility may be used to weed out these findings; however, exception made for large displacement which could be approached thru prone-supine registration [25]. It was shown [26] that motion in filling defects should not be assumed to be indicative of fecal material. Thus, integration of other means, such as tagging or electronic cleansing, may be needed in order to completely handle these types of false positives.

6. Conclusions

The developed CAD system has focused on the detection of polyps of sizes up to and including 20mm. The results have demonstrated that:
(a) the developed ColonCAD algorithm generalizes well, when applied on a completely unseen test set;
(b) the sensitivity for middle- to large-sized polyps is on the average 90% while the overall sensitivity is roughly 82%;
(c) while the false positive rates can be improved, they have remained at a manageable 4 per volume
(d) the reported sensitivity for the ColonCAD algorithm are well in the range of the reported sensitivities of expert radiologists.

Additionally, 3 polyps larger than 6 mm and 1 smaller than 5 mm, which were detected prospectively by the CAD system, were only found retrospectively by the radiologist following OC. CAD system also detected 2 polyps larger than 6 mm that were missed by OC but found by radiologist, and later found in OC. These observations further demonstrate the value of this CAD system. While the above population also included 12 masses, these have not been included as part of the statistics reported, since they are rarely missed by radiologists.

References


