Evaluation of Electronic Biopsy for Clinical Diagnosis in Virtual Colonoscopy

Joseph Marino\textsuperscript{a}, Wei Du\textsuperscript{b}, Matthew Barish\textsuperscript{c}, Ellen Li\textsuperscript{c}, Wei Zhu\textsuperscript{b}, Arie Kaufman\textsuperscript{a}
\textsuperscript{a}Center for Visual Computing and Department of Computer Science
\textsuperscript{b}Department of Applied Mathematics and Statistics
\textsuperscript{c}University Medical Center
Stony Brook University, Stony Brook, NY 11794

ABSTRACT

Virtual colonoscopy provides techniques not available in optical colonoscopy, an exciting one being the ability to perform an electronic biopsy. An electronic biopsy image is created using ray-casting volume rendering of the CT data with a translucent transfer function mapping higher densities to red and lower densities to blue. The resulting image allows the physician to gain insight into the internal structure of polyps. Benign tissue and adenomas can be differentiated; the former will appear as homogeneously blue and the latter as irregular red structures. Although this technique is now common, is included with clinical systems, and has been used successfully for computer aided detection, there has so far been no study to evaluate the effectiveness of a physician using electronic biopsy in determining the pathological state of a polyp. We present here such a study, wherein an experienced radiologist ranked polyps based on electronic biopsy alone per scan (supine and prone), as well as both combined. Our results show a correct identification 77\% of the time using prone or supine images alone, and 80\% accuracy using both. Using ROC analysis based on this study with one reader and a modest sample size, the combined score is not significantly higher than using a single electronic biopsy image alone. However, our analysis indicates a trend of superiority for the combined ranking that deserves a follow-up confirmatory study with a larger sample and more readers. This study yields hope that an improved electronic biopsy technique could become a primary clinical diagnosis method.

Keywords: Virtual colonoscopy, electronic biopsy, colonic polyps

1. INTRODUCTION

Virtual endoscopic examinations of the colon have been a popular research topic for some time [1]. Virtual colonoscopy (VC) examinations using volume rendered images of CT data have become increasingly common in practice, a popular method for colon cancer screening, and have been shown to be an effective screening method [4,7]. VC provides several advantages over traditional optical colonoscopy (OC), including a gentler patient prep, a much easier, non-invasive, and quicker procedure for the patient, and the ability to view the entire colon mucosa. Electronic biopsy [9], also referred to as a translucent view, is another tool present in VC which has no analog in OC, and is commonly present in conventional VC systems. This view, while not equivalent to a true pathological biopsy, provides the physician with initial insight into the internal structure at the millimeter level, and can assist the physician in deciding whether a patient needs to be referred to an invasive OC procedure or not.

The electronic biopsy rendering uses a modified transfer function when ray casting through the volume data. Whereas the general endoluminal view is created by using an opaque isosurfacing transfer function to view the surface of the colon wall, the electronic biopsy transfer function is translucent, allowing the ray to enter the colon wall and accumulate values within. These accumulated values are integrated along the ray traversal path to yield the final electronic biopsy image. The density of the traversed tissue is mapped to color, with higher densities being mapped to red and lower densities to blue. The resulting color image can be used to differentiate between benign tissue, adenomas, and retained stool. Benign tissue appears as a region of homogeneous low density (blue) and retained stool appears as a region of homogeneous high intensity (red) due to barium tagging, while adenomas appear as regions of irregular high density (red) structure. Figure 1 shows the difference between surface rendering and electronic biopsy rendering and also illustrates how the electronic biopsy view can be used to help determine if a polyp is benign or an adenoma.
The purpose of the work presented in this paper is to evaluate the effectiveness of the electronic biopsy technique provided in current VC systems as a diagnostic tool in assisting the radiologist in determining the pathological state of polyps. Although the tool has been present for some time, there has been no user study that we are aware of which has sought to determine the true utility of this technique in the VC environment. The electronic biopsy technique has been suggested for use in computer aided detection (CAD) of polyps [2], and we have shown that on a sample base to date of 178 datasets, we have achieved 100% sensitivity in the identification of the polyps [3]. Based on these encouraging results, this study seeks to determine to what rate a human reader can use the current electronic biopsy technology to correctly predict the pathology of polyps.

![Figure 1. Examples of surface and electronic biopsy rendering. (a) Surface rendering of a benign hyperplastic polyp. (b) Electronic biopsy rendering of the benign polyp in (a), showing low density in blue. (c) Surface rendering of an adenoma. (d) Electronic biopsy rendering of the adenoma in (c), showing an irregular high density structure in red.](image_url)

### 2. SCORING

Being that our purpose in this study is to evaluate the correctness of the determinations made by a radiologist in using the electronic biopsy view, we limit our samples to those which contain a polyp identified in the VC examination and the corresponding polyp identified in the OC, and therefore a pathology biopsy report is extant. This pathology report is used as the ground truth in labeling a polyp as being benign or an adenoma. We have so far used 450 datasets in the NIH VC repository for this study. Out of these 450 datasets, we identified 89 unique polyps that met our criteria spanning a total of 76 patients. The translucent view provided by the most recent version of the Viatronix V3D-Colon software [11], with which the radiologist is extremely familiar with from clinical practice, was used for this study.

We recruited a qualified and VC-experienced radiologist to interrogate these 89 polyps and provide his diagnosis regarding the believed pathology of the polyp using the following scale:

1 – Definitely not an adenoma  
2 – Probably not an adenoma  
3 – Indeterminate  
4 – Probably an adenoma  
5 – Definitely an adenoma

A mobility score, based on the look of the polyp’s shape in both supine and prone endoluminal views, was also assigned to each polyp:

0 – No motion (i.e., adenoma)  
1 – Mobile with stalk (i.e., adenoma)  
2 – Mobile without stalk (i.e., not an adenoma)

Using these two scoring scales, the radiologist was asked to provide four scores for each polyp. The first two scores were to rank the polyp using the 1-5 pathology scale using only the electronic biopsy as a tool for this determination. The supine views and the prone views were each scored individually, with no correlation taken into account. The third
score provided was a rank of the mobility of the polyp based on its shape characteristics in comparing the surface rendered views of both the supine and prone scans. A final pathology score was then applied to each polyp wherein the radiologist was allowed to take all factors into account.

3. RESULTS

To compare the scoring results of the radiologist against the true pathology of the polyps, it is necessary to convert the five point scoring scale we used into a binary decision. Since calling a polyp an adenoma would result in a biopsy and we would want to err on the side of caution in a clinical setting, a score of 3 (indeterminate) and above was taken to be positive (i.e., an adenoma), whereas a score of 2 or lower was taken to be a negative (i.e., a non-adenoma).

We used receiver operating characteristic (ROC) analysis to analyze our results and calculated the area under the ROC curve (AUC) value for three cases: electronic biopsy on prone only, electronic biopsy on supine only, and total decision (using both electronic biopsies and mobility ranking). Our results show that a single electronic biopsy alone has an approximately 77% chance of correctly distinguishing an adenoma from benign tissue (AUC = 0.7756 for prone and AUC = 0.7696 for supine). The total score had an AUC of 0.7983. The $\chi^2$ test [5] revealed no difference between the supine and the prone scores ($p = 0.8085$). Furthermore, there was no significant difference between the total and the individual scores (Total vs. Prone: $p = 0.2955$; Total vs. Supine: $p = 0.1829$) for the given study, although given a larger sample and more readers, such differences may emerge as significant. The ROC curves for the results of all samples are shown in Figure 2.

![ROC curves](image.png)

**Figure 2.** ROC curves of our results for all samples. (a) ROC curve for electronic biopsy of prone only score. (b) ROC curve for electronic biopsy of supine only score. (c) ROC curve for total score.

In addition to analyzing the results for all samples which were scored, we removed any non-polyps from this group (e.g., unremarkable colonic mucosa) and also divided the polyps into three size groups: less than 6 mm, from 6 to 10 mm, and greater than 10 mm. The motivation behind this is that polyps in the 6 to 10 mm range are of greatest interest for this work. Generally, polyps less than 6 mm are not considered significant. On the other hand, polyps larger than 10 mm are extremely significant and a physical biopsy would always be recommended, no matter what the results of the electronic biopsy are. The results of our ROC analysis (including prone only, supine only, and total scores) on these three polyp groups are shown in Figures 3-5. Note that for polyps larger than 10 mm, we only have two score values (4 and 5) which were assigned by the radiologist, and thus the ROC curve in Figure 5 contains only one threshold and three points. For all other polyps, scores of 2 through 5 were recorded, and thus the ROC curves in Figures 3 and 4 contain three thresholds and five points. The corresponding AUC values for these curves are presented in Table 1. Again, the results for the total score are slightly better than the results for the individual prone and supine scores, but not significantly so.
Figure 3. ROC curves for polyps less than 6mm in size for (a) prone only, (b) supine only, and (c) total scores.

Figure 4. ROC curves for polyps in the 6mm to 10mm range for (a) prone only, (b) supine only, and (c) total scores.

Figure 5. ROC curves for polyps larger than 10mm for (a) prone only, (b) supine only, and (c) total scores.
Table 1. AUC values for the ROC curves shown in Figures 3-5 (determining adenoma from non-adenoma). Probability of a polyp being correctly identified as an adenoma or non-adenoma.

<table>
<thead>
<tr>
<th></th>
<th>Prone</th>
<th>Supine</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 6 mm</td>
<td>0.7406</td>
<td>0.7386</td>
<td>0.7874</td>
</tr>
<tr>
<td>6-10 mm</td>
<td>0.6850</td>
<td>0.7043</td>
<td>0.7139</td>
</tr>
<tr>
<td>&gt; 10 mm</td>
<td>0.8750</td>
<td>0.8750</td>
<td>0.8750</td>
</tr>
</tbody>
</table>

In addition to analyzing the ROC curves to determine the probability of a correct decision being made by the radiologist, we also determined the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) for each of the three scores and for each size group. The sensitivity gives the percentage of adenomas correctly identified as adenomas. The specificity gives the percentage of non-adenomas correctly identified as non-adenomas. The PPV gives the percentage of polyps identified as adenomas which were in fact adenomas. The NPV gives the percentage of polyps identified as non-adenomas which were in fact non-adenomas. The results of these values are shown in Tables 2-4.

Table 2. Sensitivity, specificity, PPV, and NPV values for the polyp results using the prone only score.

<table>
<thead>
<tr>
<th></th>
<th>Prone Only Score</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sensitivity</td>
<td>Specificity</td>
<td>PPV</td>
</tr>
<tr>
<td>&lt; 6 mm</td>
<td>0.875</td>
<td>0.429</td>
<td>0.724</td>
</tr>
<tr>
<td>6-10 mm</td>
<td>0.958</td>
<td>0.273</td>
<td>0.742</td>
</tr>
<tr>
<td>&gt; 10 mm</td>
<td>1.000</td>
<td>0.000</td>
<td>0.600</td>
</tr>
<tr>
<td>All Polyps</td>
<td>0.926</td>
<td>0.310</td>
<td>0.714</td>
</tr>
</tbody>
</table>

Table 3. Sensitivity, specificity, PPV, and NPV values for the polyp results using the supine only score.

<table>
<thead>
<tr>
<th></th>
<th>Supine Only Score</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sensitivity</td>
<td>Specificity</td>
<td>PPV</td>
</tr>
<tr>
<td>&lt; 6 mm</td>
<td>0.880</td>
<td>0.385</td>
<td>0.733</td>
</tr>
<tr>
<td>6-10 mm</td>
<td>0.923</td>
<td>0.300</td>
<td>0.774</td>
</tr>
<tr>
<td>&gt; 10 mm</td>
<td>1.000</td>
<td>0.000</td>
<td>0.600</td>
</tr>
<tr>
<td>All Polyps</td>
<td>0.912</td>
<td>0.296</td>
<td>0.732</td>
</tr>
</tbody>
</table>

Table 4. Sensitivity, specificity, PPV, and NPV values for the polyp results using the total score.

<table>
<thead>
<tr>
<th></th>
<th>Total Score</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sensitivity</td>
<td>Specificity</td>
<td>PPV</td>
</tr>
<tr>
<td>&lt; 6 mm</td>
<td>0.875</td>
<td>0.643</td>
<td>0.808</td>
</tr>
<tr>
<td>6-10 mm</td>
<td>0.962</td>
<td>0.300</td>
<td>0.781</td>
</tr>
<tr>
<td>&gt; 10 mm</td>
<td>1.000</td>
<td>0.000</td>
<td>0.600</td>
</tr>
<tr>
<td>All Polyps</td>
<td>0.929</td>
<td>0.429</td>
<td>0.765</td>
</tr>
</tbody>
</table>
4. DISCUSSION

Since the overall accuracy when using the combined score is only 77%, electronic biopsy is not yet ready for use as a primary diagnostic tool in determining polyp pathology. When looking at the sensitivity and specificity values, it becomes clear that much of the error comes from false positive results. Indeed, the overall sensitivity is quite good for the total score at 93%. However, the specificity of 43% is extremely low. A probable reason for this comes from the fact that polyps scored as a 3 were taken to be a positive (adenoma) diagnosis in our binary scoring scheme. The good part from this is that the chances of an actual adenoma being identified as a non-adenoma is relatively low, albeit at the expense of a low sensitivity. As these results would be used to determine whether or not to send a patient for an OC, and not for actual disease treatment, this holds hope for some improvement over simply sending every patient who is found to have polyps for the OC procedure.

Our results also give credence to the importance of using both supine and prone scans in making diagnoses. Although not statistically significant based on our sample size, there is a trend that the combined total score is superior to either the individual supine or prone scores. There was also one case where a polyp could be located in the supine view but not in the prone view.

The use of digital cleansing of tagged materials should also be further scrutinized in relation to the electronic biopsy mode, as the cleansing can modify the density values on the surface of the polyps [10]. On multiple occasions, the radiologist made notes as to his belief that the result of the electronic biopsy was affected by either the tagged material or the digital cleansing process. For one polyp, the notation was made that the supine view was covered in tagging material, and a score of 3 (indeterminate) was assigned for that view. For the prone view, a score of 5 (definitely an adenoma) was assigned. The true pathology was a tubular adenoma.

A more extreme example of the effect of digital cleansing was present on another dataset. For this polyp, a score of 5 was assigned for the prone electronic biopsy and a score of 2 was assigned for the supine electronic biopsy. The polyp was indeed a tubulovillous adenoma. The radiologist made a note on this case that he did not believe the supine image and that the erroneous electronic biopsy view was probably the result of a subtraction artifact due to the electronic cleansing. Screenshots of this polyp are shown in Figure 6. While electronic cleansing methods for tagged materials are popular, they can obviously have effects on the electronic biopsy technique, which relies solely on the intensity values from the CT data.

![Figure 6](image.png)

**Figure 6.** Example of a discrepancy between prone and supine views of a polyp. (a) The prone electronic biopsy view shows significant high density throughout the polyp (red coloration). (b) The supine electronic biopsy view shows much lower density within the center of the polyp (green coloration) and less prevalent high density around the periphery. (c) The polyp is located within a pool of tagged material in the supine, leading the radiologist to believe that the observed discrepancy is the result of subtraction artifacts from the digital cleansing.
5. CONCLUSION

In this work we present for the first time a rigorous study of the effectiveness of the electronic biopsy technique in a standard VC environment. The goal of this work is to observe what rate of accuracy can be achieved in using the electronic biopsy technique to make a definitive call on whether a polyp is benign or an adenoma. The implications from this work are twofold. Based on the results, a physician can have a better understanding of what degree of accuracy to expect from using the electronic biopsy. Secondly, this data can be used to analyze the polyps that were misidentified to assist in developing an improved electronic biopsy technique that will increase the rate of correct diagnosis.

Our results, showing that the radiologist was able to correctly predict the pathological state 77% of the time using each imaging analysis alone and 80% when combined, illustrate that while this electronic biopsy tool can be a useful aide, it is not ready in its current state to act as the sole or the primary indicator of pathological state for a polyp. In general, the sensitivity in identifying an adenoma as an adenoma was quite good, but the specificity was extremely poor (due to classifying indeterminate scores as adenomas). Our results also showed that combining both the supine and prone acquisitions could be superior than using each imaging acquisition alone. These results yield hope that an improved electronic biopsy technique could become a primary diagnosis method and help reduce the number of patients being referred for OC.

We plan on expanding this study to include further datasets from the publicly available NIH databases, as well as more readers. In addition to further samples, we hope to observe the continuance of the trend of increasing accuracy when using both scores over a single score. We are also expanding our studies to observe the difference in diagnostic accuracy between using the system with and without the electronic biopsy mode. A further analysis of electronic cleansing methods might also be required to observe how closely they maintain the correct density values. As the high density regions of the polyps are typically on the surface, and since this is the region most affected by the electronic cleansing, this is a real concern.

For our expanded study, we expect to use the publicly available datasets provided by the National Biomedical Imaging Archive (NBIA) through the NIH. With an increased number of readers and samples, we seek to reach a level where a significant difference between total and individual scores can be obtained. Using all identified polyps in the NIH datasets, we will have a sample size of approximately 489 polyps. If we focus only on the polyps in the major interest range of 6 to 10 mm, then we will have a sample size of approximately 301. Based on these, the number of readers necessary to reach significance can be calculated [6]. To obtain a statistically significant result (p < 0.05) with a power of 0.800 and using only a single reader, we would require approximately 3364 samples in the case of total vs. prone and 2134 samples in the case of total vs. supine (based on our preliminary results). Using six readers, we require approximately 246 samples under the conditions of moderate accuracy, small difference, and a ratio of 1/1. This number is easily achievable using the publicly available datasets.

We plan to utilize our knowledge from this and future studies to improve the abilities of a virtual colonoscopy system to provide accurate diagnostic information on the pathology of colonic polyps. In addition to modifying the translucent transfer function used in generating the electronic biopsy images, another possibility is to use a segmentation of the colon outer boundary [8]. With this, the ray integration through the colon would be limited so as to not reach any structures beyond the colon. The hope is that at a future date the electronic biopsy can be applied as a primary means of diagnosis, allowing for a reduction in the number of patients that are referred for the invasive OC procedure based on VC findings.

ACKNOWLEDGEMENTS

This work has been supported by NIH grants R01EB007530, R01HL091939, and R01MH090134. The datasets have been obtained through the NIH, courtesy of Dr. Richard Choi, Walter Reed Army Medical Center. We thank Viatronix for the use of their V3D-Colon system in this study.
REFERENCES


