Boosted Accuracy of Colorectal Cancer Screening Through Supervised Ensemble Machine Learning Algorithms

Anirudh Kamath

Introduction

• 150,000 diagnosed per year, nearly a third die
• 55% of patients are diagnosed with higher than stage II CRC where survival rates are between 10-50%
• Current process: screen, then colonoscopy to confirm, then diagnosis
• Colonoscopy is painful and long - shouldn’t be taken if unnecessary
• Only existing noninvasive prescreen tests for colorectal cancer find hidden blood in the stool and make a diagnosis solely based on that biomarker.
• These tests, and other such blood count tests approved by the FDA, are too generalized and therefore diagnose GI illnesses, but not always colorectal cancer itself.

Approach

• Biomarkers communicate a very important piece of information, but with more information, a more practical and accurate screening result can be achieved
• Combine biomarker test with a machine learning-based predisposition model that calculates risk factors of colon cancer
• Predisposition model based on ensemble learning for boosted accuracy
• Risk factors are then combined with actual biomarker for cheap, accurate, and accessible colon cancer screening.

Data/Results

<table>
<thead>
<tr>
<th>Method</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fecal Immunochemical Test (FIT)</td>
<td>Sensitivity: 74% Specificity: 96% Cost: $5 OTC Results in five minutes Too many false negatives</td>
</tr>
<tr>
<td>Stool DNA Test</td>
<td>Sensitivity: 92% Specificity: 89% Cost: $640 Prescription-only Two weeks to get results Not for people with a history of lesions/polyps</td>
</tr>
<tr>
<td>Virtual CT Scan</td>
<td>Sensitivity: 84% Specificity: 88% Cost: $400 - not covered by most insurance</td>
</tr>
<tr>
<td>Predisposition Model</td>
<td>Sensitivity: 87% Specificity: 90% AUC: .97 Precision: 91% F1: .91 Cost: $10 Time: 5 minutes</td>
</tr>
</tbody>
</table>

Impact

• Fastest, cheapest, most accessible, and most accurate noninvasive screening method
• Allows for rapid and accurate screening in developing countries with less technology

Acknowledgements: My family, which includes a recent survivor of stage III colon cancer and the inspiration for this project, the American Cancer Society, and the Colon Cancer Alliance for their data and the work they do in the field