Cervical cancer is one of the most prevalent forms of cancer in women. It is reported by the American Cancer Society that an estimate of 13,800 women will be diagnosed with cervical cancer and for 4,290 women will die as a result of it in 2020. When cervical cancer is in its preliminary stage it rarely shows symptoms and its rate of formation is slow compared to other cancers. If it is caught when it is still confined to its cervix it has a 92 percent survival rate. In its early stages it is one of the most treatable form of cancers but if it spreads to other parts of the body treatment becomes more difficult. The early detection of the precancerous cells can greatly increase an individual's chance of survival.

The current method used to analyze cervical cells is a pap smear test. A pap smear test allows a cytopathologist to perform a microscopic analysis of the collected cervical cells. When conducting this screening the cytopathologist is looking for any abnormalities in the collected sample that would suggest that there are precancerous cells present. This method is susceptible human error due to fact that the cytopathologist must conclusively state the result of the test based solely on a qualitative analysis. Image analysis is an alternative method that can be used to evaluate the morphology of the collected cervical cells quantitatively which in turn can increase screening accuracy. In this study, we propose an algorithm that segments the nucleus of the cervical cells and utilizes the texture analysis tool gray level co-matrix to determine whether there are precancerous cells present.

**Methodology**

**Dataset:** Images of collected cervical cell samples from pap smear test acquired from the Herlev dataset

**Pre-processing:** The images were converted to gray so that their texture features could be analyzed by the texture analysis tool gray co-occurrence matrix (GLCM).

**Segmentation:** The nucleus of cervical cells shows the largest change in morphology when precancerous. Therefore, nucleus of each cervical cell was then masked using a border detection method to effectively identify our regions of interest. This automated making method isolates the nucleus by placing a mask over pixel values that are not defined within the threshold.

**Texture Analysis:** The GLCM function was then introduced to evaluate the texture feature of the nuclei present in the image. This function gave us the ability to measure the contrast, correlation, energy and homogeneity of the nuclei.

- **Contrast:** Measures the local variations in the gray co-occurrence matrix
- **Correlation:** Measures the joint probability occurrence of the specified pixel pairs
- **Energy:** Provides the sum of squared elements int the GLCM. Also known as uniformity
- **Homogeneity:** Measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal

**Results**

Simulated pap smears were generated to emulate the environment that a cytologist sees when screening the collected cells from a pap smear test. The simulated environments mirror samples that consist of normal cervical cells and approximately 0%, 5%, 10%, or 20% of precancerous cells. Our algorithm used GLCM to perform a texture-based image analysis on the morphology of the nuclei present in each simulation.

### Part I: GLCM feature ENERGY is a discriminating feature

<table>
<thead>
<tr>
<th>Normal_Abnormal</th>
<th>Normal</th>
<th>Abnormal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contrast</td>
<td>0.99688324</td>
<td>0.974129672</td>
</tr>
<tr>
<td>Correlation</td>
<td>0.834401332</td>
<td>0.966954834</td>
</tr>
<tr>
<td>Energy</td>
<td>0.960262506</td>
<td>0.999271039</td>
</tr>
<tr>
<td>Homogeneity</td>
<td>0.015628774</td>
<td>0.005365792</td>
</tr>
</tbody>
</table>

The number of cells in each layout range from 4 cells to 324 to account for instances of low cell count and the extreme. In each layout, normal cervical cells are simulated in an environment that contains a single precancerous cell. When the algorithm evaluated the texture features of the nuclei within each simulated sample it still showed that there are measureable discrepancies in GLCM texture values when a single precancerous cell is present.

### Part II: ENERGY feature is able to discriminate a pap smear slide with abnormal cells.

Group of cells in a pap smear slides from the Byrnel Herlev Database were evaluated to see if the GLCM trends seen in the simulated environments were accurate. The environments consisted of normal cervical cells in the presence of approximately 0%, 20%, 85%, and 100% precancerous cells. Energy showed the most significant change in value when the number of abnormal cells present increased.

### Results Cont.

#### GLCM Analysis of 36 Cervical Cells | Normal_Abnormal

<table>
<thead>
<tr>
<th>Energy</th>
<th>Correlation</th>
<th>Contrast</th>
<th>Homogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.96</td>
<td>0.89</td>
<td>0.82</td>
<td>0.97</td>
</tr>
<tr>
<td>0.78</td>
<td>0.84</td>
<td>0.83</td>
<td>0.97</td>
</tr>
<tr>
<td>0.68</td>
<td>0.85</td>
<td>0.84</td>
<td>0.97</td>
</tr>
</tbody>
</table>

When analyzing the simulated images it was shown that there was a measurable difference in the texture values of abnormal and normal cervical cell nuclei. After being processed by the algorithm, the entropy of the simulated samples increased significantly as we added more abnormal cells to the layout. Correlation had an inverse relationship with entropy and went down when we introduced additional precancerous cells. The homogeneity also showed a small decrease under these conditions. The contrast increased slightly but was not significant enough to be considered a determining factors. This observation was also consistent through the analysis of one precancerous cell amongst an increasing number of normal cells. This trend was also apparent when testing the Byrnel Herlev dataset. We observed that the contrast, correlation, entropy and homogeneity are all affected when abnormal cells are present in a collected pap smear sample.

**Conclusion**

In this study our algorithm incorporated a texture image analysis system that utilized GLCM to analyze simulated pap smear samples as well as Byrnel pap smear images from the Herlev dataset. The results proved that the morphological features present in the nuclei of precancerous and normal cervical cells can be differentiated, quantified, and evaluated. This method of automated screening can reduce the human error present in current screening producers and increase efficiency in accurately detecting the presence of abnormal cervical cancer cells while the disease is in its preliminary stages.

**References**