Detecting Prominent Textural Features for Cervical Cancer Lesions Using Outlier Detection and Removal Techniques

1 S.Pradeep Kumar Kenny, 2 S.Allwin
1 Dept. Of Computer Science and Engineering, Infant Jesus College of Engineering and Technology, Tuticorin, India
2 Dept. of Computer Science, Infant Jesus College of Engineering and Technology, Tuticorin, India

1 kenny_isles@yahoo.co.in, 2 allwinstephen@gmail.com

ABSTRACT
Cervical cyto images have developed profound interest among research enthusiasts in the recent past. The reason for this is not only because of its killer reputation but also due to the diverse challenging unsolved problem it possess. Cervical cancer like all other cancer develops through various stages before it actually causes potential harm. These stages can be detected through visual information such as the shape, texture etc., by a pathologist. However there aren’t so many pathologists around to do the job leaving the alternative choice to an automated system to detect these features. However since each image has different properties, finding a profound set of features which can be used to detect cancer for a particular stage of cancer is still a problem. In an automated system detecting the stage of cancer using a feature for a particular stage of cancer is not going to work. A combination of features can work but finding the right combination is a challenge. Too many features in the combination will drastically slow the system and too little will affect the accuracy of the system. To strike a balance between the two a unique technique using outlier detection and removal is used to isolate the best feature that can be used to detect a particular stage of cancer.

Keywords: Cervical Cytology, Cervical Cancer, Outlier detection, Outlier removal, features, classification

1. INTRODUCTION
The cervix is the conception channel that links the vagina to the uterus. Cervical cancer forms in the cervix of a women's reproductive system. Cervical cancer is credited with the dubious record of being the second most common cause of deaths among women worldwide. The main reason for this is that cervical cancer like all other cancer show symptoms only at a later stage. The best way to foretell if a person is likely to have cervical cancer is to have the person undergo periodic screening tests. The most commonly used screening method is the Pap smear test in which a cotton swab or a stick is used to swipe the cervix region to obtain a sample. This then is examined under a microscope by a specialist who with his knowledge tells whether the cell is affected or not. The problem with this approach is that it is time consuming and it is not feasible for pathologist to screen billions of women around the world. A computerized framework which could automate the above discussed manual process would prove to be better alternative.

For a pathologist to detect the stage of the cancer he relies vastly on his knowledge of the shape, texture and other properties of the cells to detect the cancer. All cancers develop through a series of stage before becoming invasive. The human body transcend from being a single cell at conception to over a trillions cells at adulthood through a process called growth which is governed by the mitosis process. The mitosis process is in charge of cell division and development of different parts in the human body. Each part of the human body stops growing at a predetermined stage which implies that the mitosis process should also terminate at the same time. However in a few cases the mitosis process has a tendency to proceed in spite of the fact that the development of a certain body part has stopped. This results in a circumstance where the isolated cells don't have new space to occupy and hence the cells start invading the cytoplasm region. At this stage it is difficult to detect the cancer. Once the cells find no more room they clump over one another causing tumours to form. This is the onset of the precancerous stage where cancers are normally detected. The accumulation of cells continues and forms the cancer stage where the cancer has spread to neighbouring regions and is very difficult to cure beyond this point.

Pathologists examine a tissue sample under the microscope. Every image is a collection of separate objects, having varied properties such as colour, texture etc., In short we can say an image has different contents. The objects present in the image tell what image it is. Using these properties we can define what the image is. For e.g:- if an image whose shadow is falling of the east side, it can be easily concluded that the sun is on the west side. These properties of an image can be calculated.
using various mathematical formulas which would be discussed later in this paper.

In this paper we have detected several textural features for a predefined cervical cancer dataset. After this we have clustered the obtained results and removed the outliers to refine a combinational set of features for feature detection. This paper has been organized into IV sections. Section II discusses how the features are extracted and classified by removing the outliers. And Section III shows the results. Conclusions are made in Section IV.

2. FEATURE SET FOR EXTRACTION OF CERVICAL CYTO IMAGES

The proposed method is an effective method which requires certain sets as shown in Fig.1

<table>
<thead>
<tr>
<th>Input Image</th>
<th>Pre-Processing &amp; Image Segmentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outlier Detection &amp; Removal</td>
<td>Feature Extraction</td>
</tr>
</tbody>
</table>

Fig 1: Block Diagram

2.1 Input Image

A tissue sample is obtained from the cervix and is placed on a slide which is converted into a digital image. The input image is normally represented in RGB colour model as shown Fig.2 (a)

(a) Colour Image (b) Y Component (Greyscale) Image

However for processing we need the image to be greyscale and hence the following equation is employed.

\[ Y = (0.299 \times R) + (0.587 \times G) + (0.114 \times B) \]  \hspace{1cm} (1)

where the grey value equivalent of the colour pixel is represented by Y, the Red component is represented by R, the Green component is represented by G and the Blue component is represented by B in the colour image. The resultant image is shown in Fig 2(b).

2.2 Pre-Processing & Image Segmentation

The Input Image is not a strong image. It contains variations in illumination, brightness, and distorted regions etc., which need to be strengthened so that features can be extracted more accurately. Also the region of interest which in this case is the cell nucleolus has to be separated for the entire image. There are numerous methods proposed for isolating these regions of interest [1]-[9]. Anyone of these methods can be used for isolating the regions of interest.

The pre-processing technique employed here first converts the RGB colour model image into a Gray scale colour model. It then filters the noise with the help of a median filter and then uses a set of morphological operators like Morphological reconstruction, gradient, final gradient, marker extraction to further strengthen the image. At the end we end up getting an image as shown in fig 3 after applying a watershed algorithm[10]. The algorithm for this approach is shown below.

**Algorithm:**

Step1: Convert image from RGB to Greyscale Colour Model  
Step 2: Pre-process the image using median filter  
Step 3: Apply Morphological operators to strengthen the image  
Step 4: Apply Markers  
Step 5: Apply watershed to the image and extract the objects in it.

Now that the ROI is isolated from the rest of the image and we can extract the features better. Cervical cancer like all other cancers develops through a series of stages. Here in this paper the data set being used to test the hypothesis is the Herlev data set, which consists of 917 single pap smear images[11]. Cyto technicians and doctors have manually classified each cell into one of the 7 classes, namely normal superficial, normal intermediate, normal columnar, light dysplastic, moderate dysplastic, light dysplastic and carcinoma in situ. The first three correspond to normal classes and the last four classes correspond to abnormal classes as show in fig 4. The pre-classification helps to serve as the gold standard for the entire work.

Sample Images of Herlev Dataset

**Normal Cells**

Superficial Squamous

Intermediate Squamous
3. FEATURE EXTRACTION

The segmented image contains a frozen time stamp with it. We all know that we see an object as a result of light bouncing off its surface. Different surfaces bounce back the light at different intensities providing distinct patterns. A rough surface appears rough because the light from its surface is uneven, hence a mixture of dark and light areas appears. Using these patterns we can conclude the physical properties of object surfaces.

There are several textural features available but in this paper we are taking into consideration the grey level co-occurrence matrix (GLCM) features, Haralick, gradient and Tamura based features.

GLCM is one of the simplest ways to describe the features in a given image. It makes use of statistical moments obtained from the intensity moments of the intensity histogram of an image. This is shown in the following equation

\[ P(i,j) = \sum_{x=1}^{N} \sum_{y=1}^{N} \begin{cases} 1, & \text{if } I(x,y) = i \text{ and } I(x+\Delta_x, y+\Delta_y) = j \\ 0, & \text{Otherwise} \end{cases} \]  

\[ (2) \]

Where \( i \) and \( j \) are the image intensity values, \( x \) and \( y \) are the spatial positions in an image and the offset \( (\Delta_x, \Delta_y) \) specifies the distance between the pixel-of-interest and its neighbour.

In 1973 Haralick introduced 14 statistical features which are generated by calculating the features for each one of the co-occurrence matrix obtained by using the directions 0\(^{\circ}\), 45\(^{\circ}\), 90\(^{\circ}\) and 135\(^{\circ}\). Then by averaging these four values. The Symbol \( \Delta \) representing the distance parameter, can be selected as one or higher. In general \( \Delta \) can be set to 1 as the distance parameter. Where \( E = 2^N \) is the extinction image after intensity conversion.

A 3x3 gradient approximation is used. This is an approximation of \( \left( \frac{\partial I}{\partial x}, \frac{\partial I}{\partial y} \right) \). The Gradient features can be considered as a quantification of the velocity of grey values. The measuring field is eroded by a 3x3 square[12]. These values are then plotted using the following equations.

\[ h = \text{histogram } (X) \]  

\[ h(v) = \text{frequency of pixel value of } v \]  

Tamura Image is a notion where we calculate a value for the three features at each pixel and treat these as a spatial joint coarseness-contrast-directionality (CND) distribution

In this paper we have employed around 34 features for our analysis. The features implemented are listed in table I [12]-[16].

<table>
<thead>
<tr>
<th>Name of the feature</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>[ M1 = \frac{1}{M^2} \sum_{v} v^2 h(v) - (\sum_{v} v h(v))^2 ]</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>[ M2 = \frac{1}{M^2} \sum_{v} v^2 h(v) - (M1)^2 ]</td>
</tr>
<tr>
<td>Median</td>
<td>[ \text{MED} = \text{median } (X) ]</td>
</tr>
<tr>
<td>Entropy</td>
<td>[ -\sum_{v} h(v) \log_{2} (h(v)) \text{ entropy of } h ]</td>
</tr>
<tr>
<td>Nucleocytoplasmic Ratio (NCR)</td>
<td></td>
</tr>
<tr>
<td>-----------------------------</td>
<td>--</td>
</tr>
<tr>
<td><strong>Sum Variance</strong></td>
<td>[ \sum_{i=0}^{N-1} (i - \text{Sum Ent})^2 P_{x+y} (i) ]</td>
</tr>
<tr>
<td><strong>Hyperchromasia</strong></td>
<td>[ f - f \times b ]</td>
</tr>
<tr>
<td><strong>Sum</strong></td>
<td>[ M0 =</td>
</tr>
<tr>
<td><strong>Contrast</strong></td>
<td>[ \sum_{i=-L}^{L-1} \left( \sum_{j=-L}^{L-1} P (i,j) \right)^2 ]</td>
</tr>
<tr>
<td><strong>Skewness</strong></td>
<td>[ M3 = \frac{1}{(M2)^2} \left( \sum v^3 h(v) - 3 \sum v^2 h(v) + 2 \sum v h(v) \right) ]</td>
</tr>
<tr>
<td><strong>Kurtosis</strong></td>
<td>[ M4 = \frac{1}{(M2)^2} \left( \sum v^4 h(v) - 4 \sum v^3 h(v) + 6 \sum v^2 h(v)^2 - 3 \sum v^2 h(v) \right) ]</td>
</tr>
</tbody>
</table>
| **Max** | \[ \max\{x(g)\} \]
| class of largest non - zero frequency of \( h \) |
| **Autocorrelation** | \[ \sum \sum (i,j) P (i,j) \] |
| **Correlation** | \[ \frac{\sum \sum (i,j) P (i,j) - \mu_x \mu_y}{\sigma_x \sigma_y} \] |
| **Cluster Prominence** | \[ \sum \sum (t + j - \mu_x - \mu_y)^3 P (t,j) \] |
| **Cluster shade** | \[ \sum \sum (t + j - \mu_x - \mu_y)^2 P (t,j) \] |
| **Dissimilarity** | \[ \sum \sum \|t-j\| P (t,j) \] |
| **Energy** | \[ \sum \sum (P (t,j))^2 \] |
| **Entropy of GLCM** | \[ \sum \sum P (t,j) \log (P (t,j)) \] |
| **Homogeneity** | \[ \sum \sum \frac{1}{1 + (t-j)^2} P (t,j) \] |
| **Maximum Probability** | \[ \max P (t,j) \] |
| **Sum of average** | \[ \sum \sum P_{x+y} (i) \] |
| **Sum of square variance** | \[ \sum \sum (i - \mu)^2 P (t,j) \] |
| **Sum of entropy** | \[ \text{Sum Ent} = \sum \sum P_{x+y} (i) \log (P_{x+y} (i)) \] |
| **Difference variance** | \[ \text{variance of } P_{x+y} (i) \] |
### Difference Entropy

\[
D(E) = - \sum_{x} p(x) \log p(x)
\]

### Inverse difference

\[
ID = \sum_{x} \frac{1}{x}
\]

### Inverse difference normalized

\[
ID_n = \sum_{x} \frac{1}{1 + \left(x - \mu_x\right)^2}
\]

### Inverse difference moment normalized

\[
ID_m = \sum_{x} \frac{1}{1 + \left(x - \mu_x\right)^2}
\]

### Information Measurement of Correlation 1 & 2

\[
F12 = \frac{HXY - HXY1}{\max(HX, HY)}
\]

\[
F13 = \left(1 - \exp[-2.0(HXY2 - HXY1)]\right)^{1/2}
\]

### Coarseness

Step 1:-

For every point \(x\), calculate differences between the not overlapping neighborhoods on opposite sides of the point in horizontal and vertical direction:

\[
D(x) = \{D(x, h), D(x, v)\}
\]

Step 2:-

At each point \(x\), select the size leading to the highest difference value:

\[
S(x) = \max_{D(x)} \max_{D(x, h)} \max_{D(x, v)}
\]

Step 3:-

Finally take the average over \(2S\) as a coarseness measure for the image:

\[
F_{co} = \frac{1}{N_c} \sum_{n=1}^{N_c} \sum_{m=1}^{N_c} D(n, m)
\]

### Directionality

\[
D = \frac{N_c}{N^2} \sum_{i=1}^{N_c} \sum_{j=1}^{N_c} D(i, j)
\]

### Tamura feature Contrast

\[
F_{T} = \frac{1}{N_1 N_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \frac{N_{ij} - \mu_1 \mu_2}{\sigma_1 \sigma_2}
\]

Where

- \(N_o\) is the number of pixels in the cell nucleolus region.
- \(C_c\) is the number of pixels in the cytoplasmic region.
- \(f\) is the grayscale image.
- \(b\) is the structuring element.
- \(\varepsilon\) is the opening operation.
- \(O\) is the total number of elements in the ROI.
- \(h(V)\) is the frequency of pixel value \(V\).
- \(N_h\) is the total number of elements in the ROI.
- \(\mu\) is the mean.
- \(\sigma\) is the standard deviation.
- \(X\) is the set of all pixels of \(X\) in \(O\) which is the class of largest non-zero frequency of \(h\).
- \(p(h)\) is the \(i, j\)th entry in the normalized graytone spatial dependence matrix.
- \(N_{i,j}\) is the number of distinct gray levels in the quantized image.
- \(\mu_1\) and \(\mu_2\) are the means and standard deviations of \(p_h\) and \(p_f\).
quantized number of gray levels.

\[ m_4 = \frac{1}{N-1} \sum \left( x_i - \bar{x} \right)^4 \] is the fourth moment about the mean \( \bar{x} \).

\[ \sigma^2 \] is the variance of the gray values present in the image and \( \sigma \) has to be experimentally determined to be \( \frac{2^5}{3} \).

4. OUTLIER DETECTION & REMOVAL

The above mentioned features are extracted for all the images present in the Herlev dataset and are listed separately for each feature. The Grubbs test is used to detect outliers from normal distribution. The data that are been tested are the minimum and maximum values. The outcome of this test is a probability which indicates whether the data belongs to the core population. If the data being investigated has some other asymmetric distribution then these test will give false results.

The test is based on the samples difference of mean and the most extreme data considering the standard deviation. The test detects outliers one by one with varied probabilities from a given dataset with assumed normal distribution. The test is based on the following formulas [17]

\[ T_{\text{max}} = \frac{X_{\text{max}} - \bar{X}}{\sigma} \]

\[ T_{\text{min}} = \frac{X_{\text{min}} - \bar{X}}{\sigma} \]  

Where

- \( X_{\text{max}} \) or \( X_{\text{min}} \) is the suspected single outlier (max or min)
- \( \sigma \) is the standard deviation of the whole data set
- \( \bar{X} \) is the mean

Once the outliers are deleted the total percentage of images present are calculated using the following formula

\[ \% \text{ of correct detection} = \frac{\Delta n}{\Delta s} \]  

Where

- \( \Delta n \) is the total images present in a stage after deletion of outliers
- \( \Delta s \) is the total images present in a the given dataset for a particular stage

Using the above formulas the values of features are plotted as shown in fig 6. The features which have close proximity tend to cluster together as seen in the figure. Each dot in the graph is an image of the Superficial Squamous stage. Here the Nucleo cytoplasmic value is computed and plotted. By doing this we can now be assured as to how much it can successfully can detect for this stage of cancer. Similarly the outlier plots for other stage of cancer for this feature are plotted. And each one will show the outliers. The one that has the densest cluster and the least amount of outliers is the stage where this feature can work well. Using this hypothesis for each stage each feature is plotted as such and the resultant values are extracted. This is done by calculating the percentage of features left after outlier detection.

5. EXPERIMENTAL RESULTS

The above discussed methodologies and techniques were implemented on all the images across all stages and the features were extracted. A sample of segmented images belonging to the carcinoma in situ stage is shown in Fig 7

![Fig 5: Nucleo cytoplasmic ratio feature for Superficial Squamous stage](image)

![Fig 6: Segmented images of Carcinoma in situ stage](image)

The outliers for each feature is extracted as shown in fig 8 and the resultant percentage of detection is shown in table II

6. CONCLUSION

The proposed algorithm extracts features from the dataset that are the most prominent for a particular stage of cancer by using the outlier detection and removal techniques. Using the values depicted in table II any novice can be able to interpret these values. In future
these results can be further refined by making use of ranking techniques to further concentrate the combination of features that can detect the cancer effectively.

Fig 7: Scatter plots for various stages (a) Superficial Squamous (b) Intermediate Squamous (c) Columnar (d) Mild Dysplasia (e) Moderate Dysplasia (f) Severe Dysplasia (g) Carcinoma in Situ

Table 2: Comparative analysis and percentage of accuracy

<table>
<thead>
<tr>
<th></th>
<th>F</th>
<th>S1</th>
<th>S2</th>
<th>S3</th>
<th>S4</th>
<th>S5</th>
<th>S6</th>
<th>S7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>90.54</td>
<td>75.71</td>
<td>84.69</td>
<td>86.26</td>
<td>82.88</td>
<td>88.83</td>
<td>74.67</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>82.43</td>
<td>78.57</td>
<td>84.69</td>
<td>73.08</td>
<td>87.67</td>
<td>81.73</td>
<td>78</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>85.14</td>
<td>84.29</td>
<td>83.67</td>
<td>85.71</td>
<td>81.51</td>
<td>86.8</td>
<td>83.33</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>74.32</td>
<td>74.29</td>
<td>79.59</td>
<td>78.02</td>
<td>77.4</td>
<td>88.83</td>
<td>83.33</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>81.08</td>
<td>72.86</td>
<td>87.76</td>
<td>90.66</td>
<td>84.25</td>
<td>84.77</td>
<td>84</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>81.08</td>
<td>72.86</td>
<td>81.63</td>
<td>89.01</td>
<td>83.56</td>
<td>87.82</td>
<td>83.33</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>79.73</td>
<td>72.86</td>
<td>74.49</td>
<td>87.91</td>
<td>74.66</td>
<td>83.25</td>
<td>79.33</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>83.78</td>
<td>72.86</td>
<td>87.76</td>
<td>84.62</td>
<td>87.67</td>
<td>82.74</td>
<td>91.33</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>81.08</td>
<td>87.14</td>
<td>89.8</td>
<td>88.46</td>
<td>78.77</td>
<td>79.7</td>
<td>84.67</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>87.84</td>
<td>50</td>
<td>84.69</td>
<td>53.3</td>
<td>87.67</td>
<td>85.79</td>
<td>93.33</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>72.97</td>
<td>88.57</td>
<td>96.94</td>
<td>74.18</td>
<td>87.67</td>
<td>82.19</td>
<td>82</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>85.14</td>
<td>81.43</td>
<td>81.63</td>
<td>84.07</td>
<td>85.62</td>
<td>87.82</td>
<td>97.33</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>89.19</td>
<td>80</td>
<td>77.55</td>
<td>91.21</td>
<td>87.67</td>
<td>77.16</td>
<td>92.67</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>90.54</td>
<td>92.86</td>
<td>87.76</td>
<td>79.67</td>
<td>88.36</td>
<td>90.36</td>
<td>85.33</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>87.84</td>
<td>84.29</td>
<td>82.65</td>
<td>89.56</td>
<td>83.56</td>
<td>71.07</td>
<td>74.67</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>91.89</td>
<td>92.86</td>
<td>80.61</td>
<td>82.42</td>
<td>87.67</td>
<td>86.8</td>
<td>86.67</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>90.54</td>
<td>92.86</td>
<td>87.76</td>
<td>79.67</td>
<td>88.36</td>
<td>90.36</td>
<td>85.33</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>72.97</td>
<td>77.14</td>
<td>85.71</td>
<td>78.02</td>
<td>85.62</td>
<td>79.19</td>
<td>82</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>85.14</td>
<td>84.29</td>
<td>80.61</td>
<td>89.01</td>
<td>74.66</td>
<td>80.71</td>
<td>85.33</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>86.49</td>
<td>87.14</td>
<td>80.61</td>
<td>86.81</td>
<td>85.62</td>
<td>82.74</td>
<td>78</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>85.14</td>
<td>82.86</td>
<td>79.59</td>
<td>86.26</td>
<td>78.77</td>
<td>77.66</td>
<td>72</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>72.97</td>
<td>91.43</td>
<td>83.67</td>
<td>80.22</td>
<td>89.73</td>
<td>86.29</td>
<td>80</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>81.08</td>
<td>94.29</td>
<td>83.67</td>
<td>71.98</td>
<td>87.67</td>
<td>82.74</td>
<td>81.33</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>91.89</td>
<td>94.29</td>
<td>80.61</td>
<td>82.42</td>
<td>82.19</td>
<td>85.79</td>
<td>80.67</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>75.68</td>
<td>82.86</td>
<td>80.61</td>
<td>66.48</td>
<td>89.73</td>
<td>79.7</td>
<td>86.67</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>90.54</td>
<td>90</td>
<td>83.67</td>
<td>85.71</td>
<td>79.45</td>
<td>77.66</td>
<td>84.67</td>
<td></td>
</tr>
<tr>
<td></td>
<td>27</td>
<td>28</td>
<td>29</td>
<td>30</td>
<td>31</td>
<td>32</td>
<td>33</td>
<td>34</td>
</tr>
<tr>
<td>---</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>27</td>
<td>71.62</td>
<td>85.71</td>
<td>85.71</td>
<td>87.36</td>
<td>87.67</td>
<td>77.16</td>
<td>75.33</td>
<td>75.33</td>
</tr>
<tr>
<td>28</td>
<td>85.14</td>
<td>87.14</td>
<td>80.61</td>
<td>81.32</td>
<td>82.19</td>
<td>83.25</td>
<td>82.22</td>
<td>82.33</td>
</tr>
<tr>
<td>29</td>
<td>83.78</td>
<td>87.14</td>
<td>75.51</td>
<td>80.61</td>
<td>81.32</td>
<td>82.19</td>
<td>83.25</td>
<td>84.33</td>
</tr>
<tr>
<td>30</td>
<td>77.03</td>
<td>82.86</td>
<td>87.76</td>
<td>88.46</td>
<td>82.19</td>
<td>83.25</td>
<td>84.33</td>
<td>85.33</td>
</tr>
<tr>
<td>31</td>
<td>79.73</td>
<td>82.86</td>
<td>87.76</td>
<td>88.46</td>
<td>82.19</td>
<td>83.25</td>
<td>84.33</td>
<td>85.33</td>
</tr>
<tr>
<td>32</td>
<td>87.84</td>
<td>90.00</td>
<td>89.80</td>
<td>90.54</td>
<td>89.77</td>
<td>89.77</td>
<td>89.77</td>
<td>89.77</td>
</tr>
<tr>
<td>33</td>
<td>72.97</td>
<td>81.43</td>
<td>88.65</td>
<td>78.02</td>
<td>84.25</td>
<td>88.32</td>
<td>92.33</td>
<td>92.33</td>
</tr>
<tr>
<td>34</td>
<td>74.32</td>
<td>71.43</td>
<td>82.65</td>
<td>78.57</td>
<td>83.56</td>
<td>84.77</td>
<td>85.33</td>
<td>85.33</td>
</tr>
</tbody>
</table>

Where

Abnormal Cells Normal Cells
S1: Carcinoma in Situ S5: Columnar
S2: Mild Dysplasia S6: Intermediate Squamous
S3: Moderate Dysplasia S7: Superficial Squamous
S4: Severe Dysplasia F Features

Note: The order of the stages has been changed to help in computation. Features 1 to 34 represent the same order as in Table I

REFERENCES
[14] Prof. Dr.-Ing.H.Ney., Features for Image retrieval, December 2003

AUTHOR PROFILES

S. Pradeep Kumar Kenny received his M.Tech degree in Computer and Information Technology from Manonmaniam Sundaranar University, Tirunelveli. Currently he is a research scholar in the Department of Centre for Information Technology & Engineering, Manonmaniam Sundaranar University Tirunelveli and a Assistant Professor in the Department of Computer Science and Engineering in Infant Jesus College of Engineering and Technology, Tuticorin. His research interest include Content based Image Retrieval, and fuzzy logic.
Dr. S. Allwin received his PhD in Computer Science from Manonmaniam Sundaranar University, Tirunelveli and M.E degree in Computer Science from University of Madras. Presently he is working as a Professor cum Director in the Department of Computer Science and Engineering, Infant Jesus College of Engineering, Tiruchirapalli. His research interest are primarily in the area of Bio-Medical Image Processing.