

# Pap Smear Images Classification for Early Detection of Cervical Cancer

Ayubu Hassan Mbagu  
Tianjin University of Technology and Education  
Dagu Nan Road, Hexi District  
Tianjin, China

Pei ZhiJun  
Tianjin University of Technology and Education  
Dagu Nan Road, Hexi District  
Tianjin, China

## ABSTRACT

In this presents the analyses of the Pap smear cervical cell images for cervical screening and detection. Initially preprocessed the cell images to remove unwanted noises, Followed by extraction of the cell from the background to obtain the cytoplasm and nucleus of the cell which is the region of interest. It is the only parts of the cell which can be used to differentiate normal cell from abnormal one. 20 salient features were extracted for training of support vector machine. SVM-RFE is used for features selection; the RFE algorithm removes unimportant features based on backward sequential selection by iteratively deleting one feature at a time, resulting in suboptimal combination of  $r(r < n)$  features with best predictive performance. Normalisation of the data was done to the feature so as to obtain a uniform distinct range – the feature values were normalised to 0 and 1. We conducted an experiment for evaluation of classifier; we adopted 10 cross validation and left one out for the testing of the classifier. Radial basis function and polynomial kernel function were used. In the first phase, we varied the value of gamma for training and evaluation, and the best classification accuracy was obtained at 9; where accuracy was 95.33% and sensitivity was 93.21% the performance of the classifier was increased as the value of gamma increased. In the second phase, we used polynomial kernel function, showing better performance of 97.02% accuracy and 95.50% of the sensitivity of the classifier. The classifier gives promising results with average accuracy of 92.961%, sensitivity 90.833% and specificity 80.39%.

## General Terms

Support Vector Machine, Polynomial Kernel Function, Radial Basis Function.

## Keywords

Pap smear, cervical cancer, Radial basis function, Polynomial kernel, Support Vector Machine, SVM-RFE

## 1. INTRODUCTION

Cervical cancer is the second most common cancer affecting women; it kills more than 288,000 women each year worldwide. At least 80% of cervical cancer deaths occur in developing countries, with most occurring in the poorest regions such as South Asia, Sub-Saharan Africa and parts of Latin America [1]. However, at the same time, it is one of the most preventable and treatable cancers. Since the most common form of cervical cancer starts with pre-cancerous changes and develops very slowly, up to 90% of cervical cancer may be prevented if the cell changes are detected and treated early. The early detection of cervical cancer is important both for the sake of the patients concerned and to optimise the use of medical resources [2].

The early detection is achieved by Pap screening of the cervical cells. The Pap test or screening is the most popular

and successful in preventing cervical cancer. The death rate has been reduced in developed countries since they started using the Pap test. One of the limitations of the Pap test is that it needs to be examined by humans. So an accurate analysis of the hundreds of thousands of cells in each sample is not always possible due to technical and human errors. It is a difficult and demanding job [3].

Although cervical cancer can be prevented, Pap smear images must be evaluated properly. Detection errors can be due to inappropriate smear thickness which causes cell overlapping or unwanted particles in the smear. Also, diagnosis is done by a cytotechnologist and a cytopathologist may be wrong if the number of cancerous cells is small or he or she is insufficiently experienced.

## 2. RELATED WORKS

Several automatic and semi automatic systems to diagnose Pap smear have been published by researchers using different approaches and algorithms. Also, several Master's thesis projects concerning the classification of Pap smear cells have been carried out by researchers. They found different results in terms of accuracy, specificity and sensitivity of their designed classifiers. Some used only features extracted from the nucleus to classify cells into normal and abnormal, while others used both features extracted from nucleus and cytoplasm as the features from cytoplasm are also useful for distinguishing normal cells from abnormal cells.

**Martin (2003)** in his thesis project performed classification on an old and new database; in the new database features were extracted, the classification with Fuzzy C-means and Gustafson-Kessel clustering were applied and errors were measured with K-fold cross validation. A supervised clustering procedure was proposed and compared with unsupervised clustering. Results showed supervised clustering performed better. For the old data false-positive and false-negative errors were  $< 5\%$ ; supervised Fuzzy C-means clustering performed best with a false positive of 2.02% and false-negative of 1.38%, while in the new database, unacceptable errors  $> 5\%$  were obtained. In terms of robustness, Martin's analysis showed Fuzzy C-means clustering had a high robustness to noise in the features, whereas Gustafson-Kessel showed a much smaller robustness. He concluded that Fuzzy C-means gave the best classification results and had the best robustness against noise. **Norup (2005)** proposed different methods for classification of Pap smear images, which were global inductive models – KNN, WKNN and NNH – and advanced method, which was Neuro-Fuzzy Inference for transductive Reasoning (NFI). He also constructed Nearest Class Center of gravity (NCC), where the constructed one showed the best results of all test classifiers with an overall error (OE%) of 5.1%. False-negative/false-positive was roughly equally sized. The advanced NFI performs worse than the simple NCC method; the achieved

error was 5.1%/5.7%. N. **Sarbotova (2013)** extracted 19 features from each Pap smear image, by using FeedForward Artificial Neural Network for the classification of cells into two classes – normal and abnormal. The result obtained was 79% of correct classification. She suggested that a large dataset would be needed in order to obtain better results. FI/NCC on pap-smear and when she tested the methods in Iris Benchmark the results was 4.5%/3.6% NFI/NCC.

**Genctav et al. (2012)** proposed unsupervised segmentation and classification of cervical cells. The classification was posed as a grouping problem by ranking the cells based on their features' characteristics modelling abnormality degrees. The proposed procedure constructed a tree using hierarchical clustering and then arranged the cells in a linear order by using an optimal leaf algorithm that maximises the adjacent leaves without any requirement of training examples or parameter adjustment. The test performance using two data sets showed the effectiveness of their proposed approach for the images as having inconsistent staining, poor contrast and overlapping cells. They did not present the performance of the proposed approach in percentage but still gave promising results.

**Plissiti et al. (2012)** presented a framework for classification of cervical cells into normal and abnormal depending only on the features extracted from the nucleus and ignoring the cytoplasm area, since they believed that the nucleus area is the only distinguishable area in the complex Pap smear images if the images are overlapping to a high degree. Two unsupervised classifiers were used in their work, which are spectral clustering and Fuzzy C-means. Only seven features extracted from the nucleus were used in their experiments. The result showed that the spectral clustering had an improvement of 12.16% in Isomap. When they used K-PCA (Polynomial Kernel) the Fuzzy C-means showed the classification rate of 11.17%. The best classification was obtained with K-PCA (Gaussian kernel), which was 90.58%.

### 3. PROPOSED METHODOLOGY

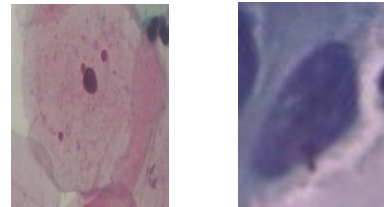
#### 3.1 Data set

Experiments are based on Pap smear new database benchmark from Herlev University Hospital. The database consists of 917 images of cropped single cell Pap smear images, which were collected by the department of pathology at Herlev University Hospital and the Department of Automation at the Technical University of Denmark. The pictures were obtained by skilled cyto-technicians using a microscope connected to a frame grabber. The pictures were taken with a resolution of 0.201µm/pixel and the average image size is 150x140 pixels. The cyto-technicians classified all pictures into diagnostic categories. Every picture was classified by two different cyto-technicians to be sure of the actual class [7].

The new database is organised as follows [2]

1. Normal – Superficial squamous epithelial, 74 cells.
2. Normal – Intermediate squamous epithelial, 70 cells
3. Normal – Columnar epithelial, 98 cells.
4. Abnormal – Mild squamous non-keratinizing dysplasia, 182 cells.

5. Abnormal – Moderate squamous non-keratinizing dysplasia, 146 cells.
6. Abnormal – Severe squamous non-keratinising dysplasia, 197 cells.
7. Abnormal – Squamous cell carcinoma in situ intermediate, 150 cells.



**Fig 1: Normal and Abnormal Pap smear cervical cell images respectively**

#### 3.2 Preprocessing

Preprocessing step is needed for background extractions and definitions of the region of interest in the image. First, the RGB images are transformed into grey scale images. Then, the images pass through a median filter to remove the unwanted noise, whereby the grey level of every pixel is replaced by the median of the intensity levels of the pixel neighbourhood while preserving edge sharpness. Linear contrast enhancement is applied to increase the contrast of the cervical cell image, since the Pap smear images suffer from inconsistent staining and poor contrast. Linear contrast algorithm spreads out the distribution of the grey level over the whole possible range of the histogram (0 to 255). Significantly, increasing their contrast, whereby the biological changes of the nucleus and cytoplasm can be clearly seen, also reduces the influence of the background on the cervical cell [8]. Otsu's algorithm is used in segmenting cell from background. The method is optimum in the sense that it maximises between class variance, a well known measure used in statistical discriminate analysis [9].

#### 3.3 Segmentations

Image segmentation is the problem of partitioning an image into meaningful parts, often consisting of an object and background. Segmentation can aid the medical staff in browsing through such large images by highlighting objects of particular importance. In addition, segmentation can give output models of organs, tumours and other structures for further analysis, quantification or simulation. In this thesis we use canny edge detector algorithm for the edge detection and boundary of the cytoplasm and nucleus, since the quality of the lines with regard to continuity, thinness and straightness is superior in canny image. Also, it is the only algorithm capable of finding the best contours while eliminating all the edges associated with grey level matter in the original image [9]. The edge detection on a cell is important, since the shape and size of the cytoplasm and nucleus provide information about illness [3].

#### 3.4 Features extraction

Feature extraction can be seen as a special kind of data reduction, the goal of which is to find the subset of informative variable based on image data. It deals with converting information to a format that is usable for the classifier algorithm; for example, a picture cannot easily be fed direct into a classifier algorithm, and instead special

characteristic are extracted from the pictures [2]. Feature extraction and multivariate data projection constitute an important issue in pattern recognition and exploratory data analysis, whereby feature extraction can avoid the curse of dimensionality and improve the generalisation ability of the classifiers and reduce the computational requirement of pattern classification [10]. The objective is to extract those features that preserve class separability well.

Many researchers focus on the features of the nucleus for the classification of a cell into normal and abnormal. We know that cytoplasm also has important features which help to distinguish between normal and abnormal cells.

In this thesis 20 features were extracted and used

**Table 1: Features extracted from cervical cell images**

Nucleus features	Cytoplasm features
1. Nucleus area	1. Cytoplasm area
2. Nucleus brightness	2. Cytoplasm brightness
3. Nucleus longest diameter	3. Cytoplasm shortest diameter
4. Nucleus shortest diameter	4. Cytoplasm longest diameter
5. Nucleus elongation	5. Cytoplasm elongation
6. Nucleus roundness	6. Cytoplasm roundness
7. Nucleus perimeter	7. Cytoplasm perimeter
8. Nucleus relative position	8. Maxima in cytoplasm
9. Maxima in nucleus	9. Minima in cytoplasm
10. Minima in nucleus	
11 Nucleus/Cytoplasm ratio	

### 3.5 Features selection

Feature selection is the process of selecting an optimum subset of features from the enormous set of potentially useful features which may be available in a given problem domain. It is often performed implicitly by the human designer of a classifier selecting the features which appear to him or her to be of most potential use. In medical image analysis, the selection of precise features from among various image modalities is the toughest and most challenging task. Some features are predominant in some categories of medical images. Therefore, finding the best features for unique images is a very crucial task in medical image processing. By comparing the various features of cancerous images with the normal images the severity and the stage of cancer in a Pap smear image can be detected [11]. A wide range of features can be computed and these can be used for building the classifier. However, increasing the number of features to build a classifier tends to increase the mis-classification error. In addition, the prediction variability increases and the classifier becomes very sensitive to the outliers.

The importance of feature selections is described below:

1. Using a small feature set may improve classification accuracy by eliminating noise inducing features.
2. A small feature set should be more generalisable for unseen data. If training data are in short supply, the use of a small number of features may reduce risk of over fitting the parameters of a classifier to the training data.
3. The use of a small feature set raises the credibility of the estimated performance of the classifier.
4. Once the best features of a given classifier have been identified, the time and computation required to measure features may be reduced, which in turn may reduce the cost/running time of the system.
5. If the development of the classifier involves a search or learning algorithm, reducing the number of features also reduces the search space that needs to be explored by the learning algorithm. This, therefore, may reduce the time needed to learn in a sufficiently accurate classification function.

The algorithms of the feature selection can often be classified into three approaches which are filter, wrapper and embedded methods [12]. In this thesis only filter and wrapper will be considered.

#### 3.5.1 Filter

The filter method eliminates unfavourable features, but reserves salient features by taking advantage of low computation complexity. Measures such as information entropy, dependence and consistency are generally used to select salient features before classification [12].

The entropy is defined as

$$E^k = \sum \sum S_{ij}^k \log S_{ij}^k + (1 - S_{ij}^k) \log (1 - S_{ij}^k) \quad (1)$$

Where  $S_{ij}^k$  is the similarity measure between any two samples  $i$  and  $j$  of feature  $k$  and is defined as  $S_{ij}^k = e^{-\alpha d_{ij}^k}$  with  $d_{ij}^k$  indicating the Euclidean distance of feature  $k$  between sample  $x_i$  and  $x_j$ .

As evident in the definition, the greater the Euclidean distance, the smaller the similarity between two samples.

#### 3.5.2 Wrappers

The wrapper method assesses individual subsets of features in a recursive way by considering their predictive efficiency to a given classifier. For a vector space with  $n$  features, the RFE algorithm removes unimportant features based on backward sequential selection by iteratively deleting one feature at a time, resulting in a suboptimal combination of  $r (r < n)$  features with best predictive performance [12]. The feature extraction techniques utilising Support Vector Machine based on recursive feature elimination (RFE) was first proposed by Guyon [13]. It returns a ranking of the features of a classification problem by training an SVM with a linear kernel and removing the feature with the smallest ranking score.

The criterion used by the SVM-RFE is the weight magnitude, which is related to each features support to discrimination function. The recursive feature elimination algorithm, recursively identifies the features with the smallest weights in magnitude and removes them. Multiple SVM-RFE [14] is another feature selection method that uses backward feature elimination procedure similar to SVM-RFE. For SVM-RFE it

starts with all features by deleting a feature repeatedly until  $r$  features are left, which leads to the largest margin separating two classes. Weight magnitude which is inversely proportional to the margin is usually used as the ranking criterion in determining importance of individual features. The eliminated feature  $p$  is the one that minimises the variation of the weight [12]. However, instead of using a single linear SVM at each recursive step, multiple linear SVMs are trained on the sub-samples of the training data obtained from multiple runs of k-fold cross validation. The feature ranking score is then computed from the statistical analysis of the weight vectors of the multiple linear SVMs. Similar to the SVM-RFE method, the feature with the smallest ranking score is omitted at each step. Since no classification algorithm is perfect, each has its positive and negative sides. By using wrapper approach fitness is obtained that better reflects the performance of the features on the actual classifier algorithm [2].

$$\|W_{-p}\|^2 = \sum_{i,j=0}^N \lambda_i \lambda_j Y_i Y_j k(x_i^T x_j) \quad (2)$$

### 3.6 Classification

The purpose of any cervical screening system is to determine whether or not a sample contains any evidence of cancer (or precancerous lesions). There are some differences in how the classification process is approached.

#### 3.6.1 Support Vector Machine

A Support Vector Machine (SVM) is a classification model that finds an optimal separating hyperplane that discriminates two classes. An SVM is a linear discriminator; however, it can perform non-linear discriminations as this is a kernel method [15]. A special property of SVM is that it can simultaneously minimise the empirical classification error and maximise the geometric margin of a classifier. It is a powerful methodology for solving problems in nonlinear classification, function estimation and density estimation, leading to many applications including image interpretation, data mining, biometric authentication, biotechnological investigation and other electrical applications [12]. The Support Vector Machine has become one of the most popular classification methods in medical applications [14]. The main objective of SVM is to orient the separating of the hyperplane to a direction that maximises the distance between the support vectors of each class and the hyperplane [14].

##### 3.6.1.1 Theory of linear separable binary classification

In this study Pap smear images were categorized into two classes normal and abnormal, this kind of situation is described as classification problems. Classification of two class data is called binary classification and can be described as follows.

Assume that we have set of labeled objects denoted by the order pairs  $(x_i, y_i)$ ,  $i = 1 \dots n$  where  $x_i \in \mathbb{R}^d$  are known as feature vectors and  $y_i \in \{-1, +1\}$  are class labels. The classification task is to generate the rule that assigns any new object (point)  $x$  to one of the classes.

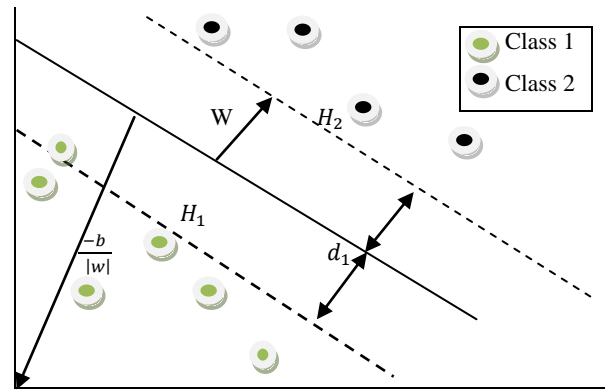
The support vector Machine has become one of the most popular classification methods in medical applications.

Assume that the data is linear separable, so a line can be drawn in a graph of  $x_1$  vs.  $x_2$  just having only two features,

likewise a hyperplane can be drawn on a graph of  $x_1 \ x_2 \dots x_D$  when  $D > 2$ . That hyperplane is given by

$$w^T x + b = 0 \quad (3)$$

Where  $w$  is the normal vector of the hyperplane, and the term  $\frac{b}{\|w\|}$  is the perpendicular distance between the hyperplane and the origin. The Support Vectors in essence are those objects closet to the separation hyperplane. The main objective of Support Vector Machine (SVM) is to orient the separating the hyperplane to a direction that maximizes the distance between support vectors of each class and the hyperplane. The figure below shows an illustration of two linear separable classes of data by hyperplane [16, 17].



**Fig 2: Illustration on how two classes can be separated by hyperplane**

Figure 2 depicts that implementing a support vector machines (SVM) can be cut down to selecting the appropriate values of the variables  $w$  and  $b$ . Hence the training data for two classes can be described as follows:

$$w^T \cdot x_i + b \geq +1 \quad (4)$$

For  $y_i = +1$

$$w^T \cdot x_i + b \leq -1 \quad (5)$$

For  $y_i = -1$

The generalization of these formulas can be given as follows:

$$y_i(w^T \cdot x_i + b) - 1 \geq 0 \quad \forall_i \quad (6)$$

The two planes containing the support vectors for the classes 1 and 2 are named as  $H_1$  and  $H_2$  respectively, these two planes can be computed as follows:

$$w^T \cdot x_i + b = +1 \quad (7)$$

For  $H_1$

$$w^T \cdot x_i + b = -1 \quad (8)$$

For  $H_2$

The variable  $d_1$  and  $d_2$  are the perpendicular distances of separating hyperplanes to  $H_1$  and  $H_2$  respectively and are known as SVM margins. The SVM aims to orient the hyperplane in a way that it maximizes the margin. This margin is equal to  $\frac{1}{\|w\|}$ . To maximize this margin it is necessary to minimize  $\|w\|$ , the maximization of the margin of separation between binary classes is equivalent to minimizing the Euclidean norm of weight vector  $w$  [17]

### 3.6.2 Normalisation

Normalisation is a very important step for any algorithm to perform well. As the feature values range significantly among different features the data set was normalised to have feature values between 0 and 1. Simply, normalisation is done so that the values are uniformly distributed within a distinct range. Normalisation is given as follows:

$$D = \frac{C - \min}{\max - \min} \quad (9)$$

Where

D = value after normalisation,  $0 \leq D \leq 1$

C = original value

Max and min is the maximum value and minimum value respectively.

## 3.7 Experiment setup

In this thesis various kernel functions were included for designing and evaluating the SVM classifier. Radial basis function (RBF) and polynomial function were used for mapping the original sample space into new Euclidian space, whereby a linear classifier was designed. Data were randomly divided into training set and test set. The SVM classifier was trained with salient features selected and tested with the cell image in the test set. We conducted an experiment with different values of Gaussian kernel parameter varying from 0.5 to 9 and polynomial kernel degree was set at 3 and 5. The regularisation parameter cost function C varied from 10 to 50. We adopted 10 cross validation to train the model with the best performance using training set, which was followed by testing phase to verify the performance of the classifiers using the test set. We also adopted leave-one-out method for validation of the classifier; thus, each iteration 9 fold was used for training and the rest was used for validation during test phase. We repeated the procedure several times by varying the values of parameter for better performance of the classifier. Each time the values of parameters were varied, we used individual fold for validation. This classification is for two classes only, means normal and abnormal. We used LIBSVM software as in [14].

### 3.7.1 Evaluation of classifier

The performance of the SVM classifier was evaluated by calculating its accuracy, sensitivity and specificity after varying the kernel parameters:

$$Accuracy = \frac{TP+TN}{TP+TN+FN+FP} \quad (10)$$

$$Specificity = \frac{TN}{TN+FP} \quad (11)$$

$$Sensitivity = \frac{TP}{TP+FN} \quad (12)$$

Where

Accuracy is the overall performance of the classifier.

Specificity measures the proportion of normal cells correctly identified as such by classification algorithm.

Sensitivity measures the proportion of the abnormal cells correctly characterised as such by classification algorithm.

TP is the number of cells which are abnormal and truly classified as such.

TN is the number of normal cells which are truly classified as such.

FP is the number of the normal cells wrongly classified as abnormal.

FN is the number of abnormal cells wrongly classified as normal.

## 3.8 Results

The results obtained in two phase experiments of Support Vector Machine classifier in two kernels functions which are Radial basis and Polynomial kernel function.

**Table 2: Results obtained from SVM classifier**

SVM	Sensitivity	Specificity	Accuracy
Gamma=0.5	85.08%	95.03%	89.05%
Gamma = 3	86.12%	96.10%	90.23%
Gamma = 6	88.35%	96.48%	92.36%
Gamma = 9	93.21%	96.76%	95.33%
Polynomial d= 3	95.65%	98.01%	93.78%
Polynomial d = 5	96.59%	99.00%	97.02%

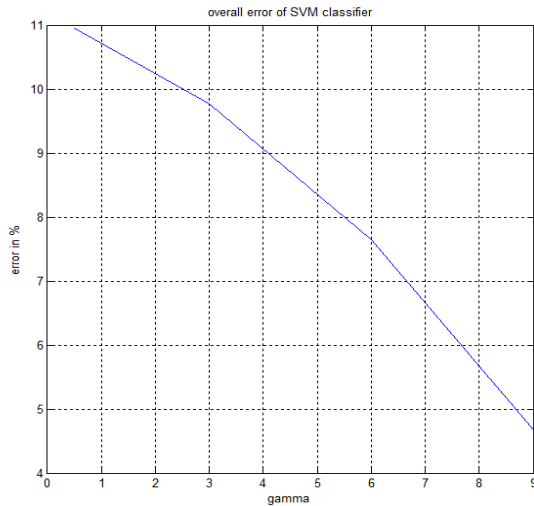


Fig 3: Overall error of designed SVM classifier

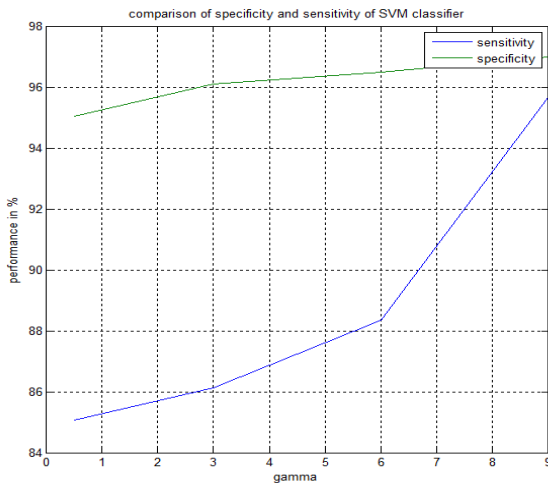


Fig 4: Comparison of designed SVM classifier in term of specificity and sensitivity

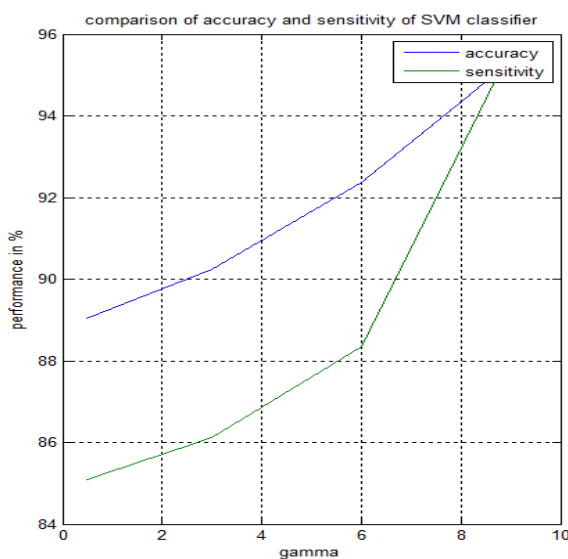


Fig 5: comparison of designed SVM classifier in term of accuracy and sensitivity

### 3.9 Discussion

This research described the analysis of the Pap smear cell images for the diagnosis and detection of cervical cancer at early stage. Segmentation algorithms were used to help during extraction of the salient features from the cell. Since the output of the segmentation gave us only region of interest which made easy for us to do analysis. Feature selection played an important role in our work, since it reduces the computation time and save the memory space. In both phase of experiment LIBSVM software was used for the training and testing of SVM classifier integrated with MATLAB. This software needs feature scaling so as to have all features in acceptable range which can be  $[-1 +1]$  or  $[0 1]$ .

The experiment was done at various value of gamma and various value of error were obtained, for example figure 3 depict various errors obtained during experiment at low value of gamma parameter the error was so high which was 10.95% and at  $\gamma = 5$  the value of error was 9.2% and at  $\gamma = 9$  the error obtained was 4.67% this show that error will be minimized as the large value of  $\gamma$  will be selected. Thus the best performance of the SVM classifier at Radial Basis function will be achieved at large value of gamma.

Figure 4 shows the designed SVM classifier its ability to detect is higher since at very low value of  $\gamma = 0.5$  the classifier specificity is high about 95.03%, then the performance of the classifier increased slightly as the gamma increased that meaning to say specificity of the classifier is slightly affected by increase of the gamma value. While the sensitivity increased gradually with the increase of the value of gamma from  $\gamma=0.5$  to  $\gamma=9$ .

Figure 5 shows the comparison of the accuracy and sensitivity of the classifier starts at  $\gamma = 0.5$  while the performance was 89.05%, the accuracy of the classifier increased slightly as the the value of gamma increased also the sensitivity gradually increased as the gamma increase

For polynomial kernel function two values of the parameter degree d were tested showing the improvement of the classifier; the value of degree selected was 3 and 5 and obtained accuracy increased from 93.78% to 97.02%, specificity from 98.01% to 99.00% and sensitivity from 95.65% to 96.59%. These results show that as the value of the RBF and Polynomial kernel functions parameter increase, this affects directly the performance of the classifier. Selection of best kernel parameter is also a great challenge on designing of the multiclass Support Vector Machine based on classification accuracy, sensitivity and specificity.

With the comparison of the results obtained by [12] SVM based two clusters with seven selected salient features was accuracy (98.83%), sensitivity (91.40%) and specificity (99.90%). Plissit et al. obtained the best classification by using K-PCA with Gaussian kernel, which was 90.58% with features extracted from the nucleus and ignoring the features from the cytoplasm.

The performance of the designed SVM classifier is low compared with the results obtained by [12] and [7], but we used 20 salient features extracted from cytoplasm and nucleus of the cell compared with [12], who used only seven features extracted from the cell, while plastillis was used on features extracted from nucleus only. Martin and Norup in their study had the same 20 features used here, but in classification they used different algorithm; they achieved low overall errors meaning that their classification presents promising results. Sabortova used 19 features extracted from nucleus and

cytoplasm of the cell and classification was done by feed forward artificial neural networks. The achieved performance of the classification was much less than what we obtained.

#### 4. CONCLUSIONS

In this work the performance of the SVM classifier is promising one this is due to the results obtained from SVM which deemed to be best accuracy performance is 95.33% at highest value of gamma (i.e  $\gamma = 9$ ). These results were obtained during testing of the classifier in Gaussian kernel function. When we trained and tested the classifier by using polynomial kernel, the performance of the classifier was much better and higher compared with the Gaussian kernel. However, the best accuracy with polynomial kernel was 97.02% at polynomial degree value of  $d=5$ . From the graphs shows that the performance of the SVM classifier rose as the value of the kernel function parameter increased.

In future work authors are interested on trying many classification algorithms and different feature extraction and selection algorithms which will result in better performance in terms of accuracy, sensitivity and specificity in both single cell and overlapped cells images. Also to train and test classifiers for classification of the cells into four classes and seven classes for this database images.

#### 5. ACKNOWLEDGMENTS

I would like to thank my supervisor Prof. Pei ZhiJun for his support and editing this paper. The study is also supported by the Research Foundation of Tianjin University of Technology and Education under Grant No KJY1312

#### 6. REFERENCES

- [1] International Agency for Research on Cancer, World Health Organization, available [online] Available from <http://screening.iarc.fr/cervicalindex.php> [accessed 10 September 2014]
- [2] Eric Martin (2003) Pap-smear classification, Technical University of Denmark-DTU, Master's Thesis.
- [3] M. Mohideen Fatima alias Niraimathi, V. Seenivasagam (2012) A hybrid image segmentation of cervical cells by bi-group enhancement and scan line filling. International Journal of Computer Science and Information Technology and Security, 2 (2), 368–375.
- [4] Jonus Norup (2005) Classification of Pap-smear data by transductive neuro-fuzzy methods, Technical University of Denmark-DTU, Master's Thesis.
- [5] Hana Sarbortova (2013) Detection of cervical cancer in Pap smear images, University of Wisconsin-Madison, Final Project Report
- [6] Ash Genctav, Selim Aksoy, Sevgen Onder (2012) Unsupervised segmentation and classification of cervical cell images. *Pattern Recognition*, 45, 4151–4168
- [7] Marina E. Plissiti, Christophoros Nikou (2012) Cervical cell classification based exclusively on nucleus, ICIAR2 2012, part II LNCS, pp. 483–470.
- [8] Nor Ashidi Mat Isa, Nazahah Mustafa, Kamal Zuhairi Zamli, Mohd Yusoff Mashor (2006) Improvement of Contrast Enhancement Technique for cervical cell of Pap smear images by reducing the effects of unwanted background information, IEEE.
- [9] Rafael C. Gonzalez, Richard E. Woods (2010) Digital image processing. 3<sup>rd</sup> ed. Beijing, Pearson Education Asia Ltd.
- [10] Jianchang Mao, Anil K. Jain (1995) Artificial neural networks for feature extraction and multivariate data projection. *IEEE Transactions on Neural Networks*, 6 (2), pp.296-317.
- [11] Edwin Jayasingh, S. Allwin (2013) Detection of cancer in pap smear cytological images using bag of texture features. *IOS Journal of Computer Engineering (IOS-JCE)*, 11 (1), pp. 01–07.
- [12] Yung-Fu Chen, Po-Chi Huang, Ker-Cheng Lin, Hsuan-Hung Lin, Li-En Wang, Chung-chuan Cheng, Tsung-Po Chen, Yung-Kuan Chan, John Y. Chiang (2013) Semiautomatic segmentation and classification of Pap smear cells. *IEEE Journal of Biomedical and Health Informatics*, pp 1–15.
- [13] I. Guyon, J. Weston, S. Barnhill, V. Vapnik (2002) Gene selection for cancer classification using Support Vector Machines. *Machine Learning* 46, pp. 389–422.
- [14] Ramin Moshavegh, Babak Ehteshami Bejnordi (2013) Chromatin pattern analysis of cell nuclei for improved cervical cancer detection, Master's Thesis in Biomedical Engineering, Department of Signal and systems, Chalmers University of Technology, Sweden.
- [15] Luz Helena Camargo Casallas (2012) Classification of squamous cell cervical cytology, Master's Thesis, Faculty of Medicine – Engineering Faculty, Universidad Nacional de Colombia, Bogota D.C Colombia
- [16] S. Theodoridis and K. Koutroumbas (2009) Pattern Recognition, 4<sup>th</sup> ed. China Machine Press.
- [17] Simon Haykin (2009) Neural networks and learning machines, 3<sup>rd</sup> ed. China Machine press, Pearson education Asia Ltd.