IOSR Journal of Dental and Medical Sciences (IOSR-JDMS)
e-ISSN: 2279-0853, p-ISSN: 2279-0861.Volume 14, Issue 2 Ver. VI (Feb. 2015), PP 85-94
www.iosrjournals.org

Image Morphometry of Routine Slides For Cancer Diagnosis

Prasad.P.H [1], Sheeba.V.S [2], Vineetha Nandakumar [3], Jyothi.C.R [4]
[1] *Additional Professor, Pathology, Government Medical College, Thrissur, Kerala, India 680596*
[2] *Professor of Electronics and Comm. Engg, Govt. Engg. College, Thrissur, Kerala, India 680009*
[3] *M Tech Student, Electronics and Comm. Engg, Govt. Engg. College, Thrissur, Kerala, India 680009.*
[4] *Associate Professor, Pathology, Government Medical College, Thrissur, Kerala, India 680596*

Abstract: *The first changes of malignant transformation of cells occur at the DNA level, which later manifest as change in DNA content. Microscopic diagnosis of cancer by assessment of DNA content of the nucleus is highly subjective, minor changes of DNA content may be missed by visual examination. By Quantitative Pathology the diagnosis can be made more precise.Feulgen is the nuclear stain used in most of the studies on image morphometry which requires special staining procedure. In the present study we have carried out the morphometric analysis on haematoxylin stained cytology slides used for routine reporting.The microscopic images captured using a digital camera were analyzed to find out the area, total optical density, hue, saturation and perimeter of the nucleus. The mean and standard deviations of these parameters, for a set of known benign and malignant cases were used to train Support vector machine (SVM) classifier. This classifier was then used to test data from a series of unknown cases and the results were compared with the final pathological diagnosis of these cases to check the efficiency of the method. The SVM classifier was trained using data from 25 benign and 25 malignant cases. The classifier was then used to analyze 34 cases, the nature of which was unknown to the person who analyzed the cases. 100 % efficiency was obtained in differentiating benign and malignant cases using the classifier.*

Keywords: *Image morphometry, DNA Ploidy, Nuclear Area, total optical density, TOD, SVM, cancer diagnosis, quantitative pathology*

I. Introduction

Cancer is the second most common cause of death in nearly all Countries. The incidence of cancer is increasing all over the world. It is a well accepted fact that the cure rate of cancer is higher if an early diagnosis is made. Cancer is a dreaded disease irrespective of whether it affects poor or affluent people. The diagnosis of cancer in a patient turns his life upside down. Everybody fears cancer because of the pain, disfigurement, financial burden and the agonizing death that ultimately follows. In spite of the tremendous advances that has occurred in the diagnosis and treatment of cancer, the outcome for advanced cancer is grave. The only hope for a better cure and survival is an early diagnosis.

Normal cells of the body contain 23 pairs of chromosomes represented as 2N. A small proportion of cells will be having 4N number of chromosomes just before mitotic division. No other variations are permitted by the cell cycle regulatory mechanisms in healthy cells. If any changes to these chromosomes occurs the cell cycle will be stopped and cells will be given adequate time for correcting these mistakes. If the mistakes are uncorrectable the cell will be destroyed by apoptosis. When the cells turn malignant, these regulatory mechanisms are ineffective and various abnormalities in the form of quantitative differences in DNA content occur. Many developments have occurred in the detection of different prognostic markers on the cancer cells. These markers will predict the specific treatment appropriate for the case and the likely prognosis of the patient. The role of these investigations comes only after establishing the diagnosis of malignancy. In spite of the developments in medical science, the gold standard for diagnosis of malignancy is still microscopic examination of stained cells or tissue by experts who look for various features of malignancy in these cells.In malignant cells, due to ineffective nuclear regulatory mechanisms, the uniformity of the DNA content of cells will be lost. Each cell might have a random number of chromosomes. When more chromatin is present in the nucleus it will take up more stain. This imparts darker color and increased area of the nucleus. This is known as anueploid state. It is considered as a hallmark of malignancy in most of the tumors. In stained cells this manifests as variation in parameters like nuclear size, total optical density, perimeter, etc. Pathologists diagnose malignancies by visually estimating these parameters by looking through the microscope, which is highly subjective. Subtle variations may be missed and these variations cannot be quantified. Due to visual illusions, some features may not be detected by the eyes and some features may be misinterpreted. The diagnosis of cancer can be made more precise by digital image analysis and estimating DNA Ploidy.

DNA ploidy analysis is a useful tool in diagnosis and predicting the prognosis of cancers. Many studies have shown it as an independent prognostic marker.[1-6]

The early study on image analysis of nuclear morphology appeared in 1990's when Danque PO *et al* [7] compared DNA ploidy estimate of 12 solid tumors by six methods. They observed that Image analysis of touch preparations detected most of the tetraploid and multiple aneuploid peaks. In their study, Flow cytometry occasionally did not detect tetraploid and multiple aneuploid peaks. In 1993, C Chaplin *et al* [8] correlated data from nuclear morphometric analysis of carcinoma breast imprints with features like Ki67 immunostaining, AgNOR and DNA content of the tumors. DNA ploidy analysis of hepatic neoplasm was done by Li Jun Mi *et al* in 1994 [9] and found that aneuploid peaks were much increased in carcinomas.

Recently, comparison of DNA ploidy and biomarker expression in paraffin embedded tissue sections was carried out by Stijn JHM Fleskens *et al* in 2010[10]. They analysed DNA ploidy in 22 paraffin embedded oral premalignancies. The disadvantage of using paraffin sections for DNA ploidy is that it may not be accurate because oblique sectioning or sections from the tips of nuclei may give erroneous results. To avoid such mistakes, we used aspiration cytology smears which represent the entire nuclei.[11, 12]

Jeffrey S. Ross *et al* [13] presented image based DNA ploidy analysis, which permitted easy identification of malignant cells and avoided unwanted cells from the study. The analysis could be made even when the cells were in small numbers unlike flow cytometry. In this study the image morphometry was done using direct reading of image from the microscope. The background staining or dense cytoplasmic staining may alter the results if the readings are taken directly from the slide. In our study the images were captured and nuclei were separated from background using edge detection and thresholding, hence the accuracy could be increased. Emily G. Barr Fritcher *et al* [14] studied the role of routine cytology, quantitative nuclear morphometry by digital image analysis, and genetic alterations by fluorescence in situ hybridization for detecting pancreatobiliary tract malignancy. They found a strong link between the visual assessment of cells by routine cytology examination and quantifiable nuclear morphometric features. In their study only 50 abnormal nuclei were analyzed from each case. In our study more than 100 cells were analysed per case.

For the automation of cancer detection process, different Machine Learning Classifiers such as Linear Discriminant Analysis (LDA), K-Nearest Neighbor (KNN), Artificial Neural Network (ANN) etc. were used by several authors [15-18]. Support Vector Machine algorithm is a popular tool for machine learning tasks and SVM-based approaches are able to significantly outperform competing methods in many applications. Support Vector Machine classifier was proposed by many researchers for Breast cancer diagnosis [19-21]. But data they had considered for classification was either the dataset taken from internet or mammograms/ultrasound images and not cell images. In this paper, we incorporate the DNA ploidy analysis with machine learning to develop an automated cancer detection technique which can assist the pathologists in doubtful cases. Support Vector Machine learning was used for the classification of Benign and Malignant tissues based on the data obtained from the DNA ploidy analysis of 84 cases. The nuclear images of papanicoloau stained cytological smears of benign and malignant breast lesions were captured by a high quality digital camera. The images are preprocessed and segmented before extracting the features. The data collected from the images were then fed to Support Vector Machine which is trained to perform the classification of data into Malignant or Benign. The **SVM** was then used to test a set of cases, the actual diagnosis of which was unknown to the person analyzing these cases. The paper is organized as follows. Section 2 discusses the methods for image acquisition, preprocessing, segmentation and feature extraction. Section 3 describes details of classification using SVM. The results are discussed in section 4 and concluding remarks are given in section 5

II. Materials And Methods

There are five steps in image morphometry as shown in Fig 1, which includes image capturing, preprocessing, segmentation, feature extraction and classification.

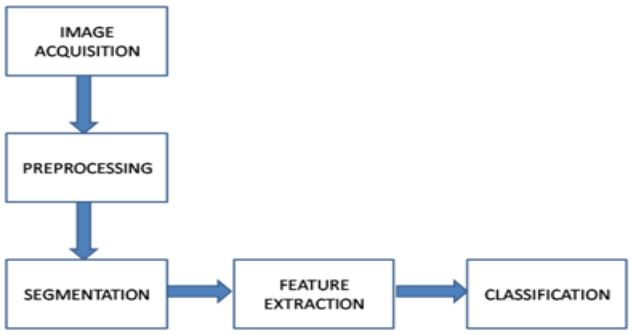

Fig:1 Flowchart of the method

2.1 Image aquisition

All the cell images used in the study were obtained by Fine Needle Aspiration Cytology (FNAC) of various breast lesions. The aspiration was done using 21G needle and 10 cc syringe. The aspirated material was immediately transferred onto clean glass slides and quickly fixed in 85% isopropyl alcohol for about 15 minutes. Extreme care was taken to avoid drying and smearing artifacts. The slides were then stained by haematoxylin solution and differentiated in acid alcohol, followed by OG6 and EA 36 stains (Papanicolaou method) and finally mounted in DPX. The slides were then examined using microscope and images were captured from representative areas. Only optimally stained cells in a clean background were photographed. Areas showing drying or smearing artifacts were avoided. Similarly, areas where morphology of cells was masked by blood or necrotic debris were avoided. Canon EOS 550D SLR camera was used for capturing images of the highest quality. The camera lens was removed and a microscope adapter fitted with 10X lens was used to attach the camera to the microscope. This step was to avoid the autofocusing and auto zooming actions of digital camera lens and make sure that all images were captured at the same magnification. All the images were captured using high power objective of the microscope (40X). No further zooming in or out of the images was done after capturing. Fixation and smearing artifacts can cause false variations in nuclear size and shape. In order to avoid this, optimally fixed and stained areas were selected in each slide and photographed. Same lighting, exposure time and magnification settings were used to capture all images used in this study.

2.2 Preprocessing of the Images

In Papanicolaou stained slides, the nucleus of the cells will be deep blue in color and the cytoplasm will be reddish in color with some background staining of the proteinaceous material. If the nuclear DNA is to be analyzed, only the stained nuclei should be present in the image. Other components are not needed and it may mask the nuclear features. In order to separate nucleus from cytoplasm and background staining, intensity thresholding was done.

Since the images used in this study had only one type of dark object (nucleus) in a lightly stained area (cytoplasm), Global Intensity Thresholding was used. A single threshold value is applicable for one image. But a slightly different threshold value may be needed for other images depending upon the background staining intensity which was found out by trial and error.

2.3 Segmentation

In image analysis, specific diagnostic information is extracted from the image. For this, scene segmentation is applied to locate the objects of interest; for example the outline of a nucleus. In order to determine the location of the nuclei in the tissue, a segmentation method based on the Chan-Vese algorithm for Active Contours was used. The basic idea in active contour models is to evolve a curve, subject to constraints from a given image, in order to detect objects in that image [22]. The curve starts from outside the object to be detected and moves towards the interior normal and stops at the boundary of the object.

This method was based on the minimization of the energy associated to the contour as a sum of internal and external energy. Let C denote the evolving curve in a bounded open subset of R^2. Assume that the image I is formed by two regions of approximately piecewise-constant intensities, of distinct values I_1 and I_2. Also assume that the object to be detected is represented by the region with the value I_1 and its boundary denoted by C_0. Then $I \approx I_1$ inside the object [or inside (C_0)], and $I \approx I_2$ outside the object [or outside(C_0)]. In active contour model the following fitting term is minimized

$$F_1(C) + F_1(C) = \int_{inside(C)} |I(x,y) - c_1|^2 dxdy + \int_{outside(C)} |I(x,y) - c_2|^2 dxdy \quad (1)$$

where C is any other variable curve, and the constants c_1 ,c_2 depending on C, are the averages of I inside and outside C respectively. If the curve C is outside the object, then $F_1(C) > 0$ and $F_2(C) \approx 0$ If the curve C is inside the object, then $F_2(C) > 0$ and $F_1(C) \approx 0$. If the curve is both inside and outside the object, then $F_1(C) > 0$ and $F_2(C) > 0$. The fitting energy is minimized, if the curve C is on the boundary of the object.

Finally the energy functional which is to be minimized, is formed by adding some regularizing terms, like the length of the curve C, and (or) the area of the region inside C, as given below

$$F(c_1, c_2, C) = \mu.Length(C) + \nu.Area(inside(C))$$
$$+ \lambda_1 \int_{inside(C)} |I(x,y) - c_1|^2 dxdy + \lambda_2 \int_{outside(C)} |I(x,y) - c_2|^2 dxdy \quad (2)$$

where $\mu \geq 0, \nu \geq 0, \lambda_1, \lambda_2 > 0$ are fixed parameters. In this model the parameters are assumed as $\lambda_1 = \lambda_2 = 1, \nu = 0$

The segmentation results in a binary image in which all the nuclei are marked in white color as shown in Fig. 2. This resultant image can be used as a reference for feature extraction from the original RGB and gray scale images.

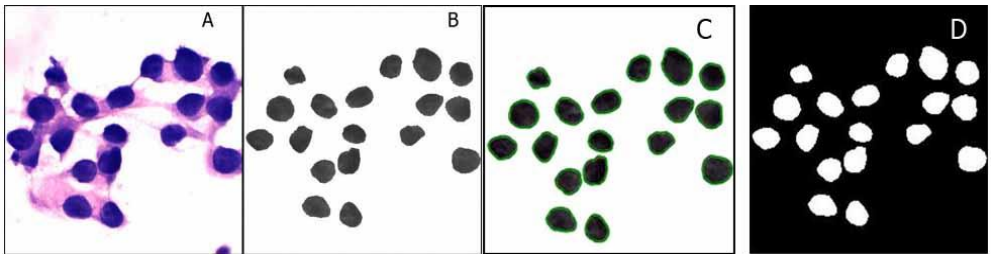

Fig.2 (A) Sample Input; (B) Result of Intensity Thresholding (C) Segmented Image; (D) Segmented Binary Image

2.4 Feature Extraction

After separating nuclei from the cell images, various features needed for differentiating benign and malignant cells were extracted by analyzing the nuclear images. The features considered in this paper are Area, Perimeter, Total optical density (TOD), Hue and Saturation.

Area: In the digital image, the nuclear area is represented by a group of pixels. The area is calculated by counting the number of pixels inside the boundary of the nucleus. This represents the size of the nucleus. When the chromatin content varies the area of a nucleus also varies. In benign cells the chromatin content will be uniform and so only minimum variation of area is expected.

Perimeter: The Perimeter of the nucleus is calculated by counting the number of boundary pixels at the periphery of nuclear image. This is also directly proportional to the area of the nucleus. When there is a marked variation of contour of the nucleus, there will be an exaggeration of the variance of the perimeter.

TOD: The optical density of the pixel represents the chromatin content of that particular area. Since the nuclear chromatin distribution is nonuniform, the optical density of each pixel will be different. In order to get a true representation of the chromatin content of a nucleus, one has to take the sum of the optical density of each and every pixel in the nucleus. This is known as the total optical density (TOD).

TOD is calculated after converting the RBG nuclear image to a grayscale image. The value of each pixel varies from 0 to 255. The value of each pixel represents the intensity of staining of the nucleus at that particular area which is directly proportional to the DNA content of the cell in that area. Optical Density = zero represents full bright area or an area of no DNA content and Optical Density= 255 denotes a full black area or an area of maximum DNA content [11]. When the optical density values of each pixel in a nucleus are summed up, it gives the Total Optical Density (TOD). This directly represents the total DNA content of the nucleus.

Hue represents the color information of the image. It is independent of the intensity which varies from image to image depending upon the intensity of light and period of exposure of the camera. These errors can be eliminated by noting the hue value of the image.

Saturation represents the richness or purity of color.

Use of the RGB model is problematic because the information of interest, i.e., the color of the stain (determined by the absorption characteristics), is mixed with variations in the amount of stain. A widely used procedure to extract the chromatic (color) information from the RGB data is the hue-saturation-intensity (HSI) model. The RGB to HSI transform decouples the intensity information from the color information [23].

On an average more than 100 cells were analyzed in all the cases. In each of these nuclei five features were estimated. The mean and variance of these five features in each case was calculated and analyzed. Histograms of the area and TOD of nuclei from each case was plotted and compared. Quantification of these parameters helps us to differentiate benign and malignant nuclei, hence to detect malignancy.

III. Classification

The automated identification of malignancy and benignity of the test series of images is done using **Support Vector Machine (SVM).** It is a supervised learning model with associated learning algorithms that analyze data and recognize patterns. When a set of training examples is given, each marked as belonging to either benign or malignant categories, an SVM training algorithm builds a model that assigns new examples into benign or malignant category, making it a binary linear classifier.

Support Vector Machine is a kind of large-margin classifier: a vector space based machine learning

method where the goal is to find a decision boundary between the two classes. The data points closest to the separating hyperplane are called as support vectors. The decision function is fully specified by these Support Vectors which are actually a subset of the training data. The distance from this decision surface to the support vectors, is termed as the margin. Hence the decision criterion is defined as "maximizing the margin" [19, 24-26]. An SVM model is a representation of the inputs as points in space. Let vector $x \in R^n$ denote a pattern to be classified, and scalar y denotes its class label (i.e., $\{y \in \pm 1\}$). Besides, let $\{(x_i, y_i), i = 1,2,\cdots,m\}$ denote a given set of m training examples. The goal is to construct a classifier that can correctly classify an input pattern *x* that is not necessarily from the training set. The decision hyperplane for linear SVM is defined in terms of a normal vector *w* (perpendicular to the hyperplane) and an intercept term *b*, as given below.

$$f(x) = w^T x + b = 0$$

For a given training set, there may exist many hyperplanes that separate the two classes, the SVM classifier is based on the hyperplane that maximizes the separating margin between the two classes (Fig 3). Mathematically this hyperplane can be found by minimizing the following cost function [19].

$$J(w) = \frac{1}{2}\|w\|^2 \text{ subject to constraints } y_i(w^T x_i + b) \geq 1, \quad i = 1,2,\cdots,m \tag{3}$$

In practice the training data may not be completely separable by a hyperplane. In such case, slack variables, denoted by ξ_i , can be introduced to relax the separability constraints in (3) as given below

$$y_i(w^T x_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1,2,\cdots,m \tag{4}$$

The cost function can be modified as

$$J(w) = \frac{1}{2}\|w\|^2 + \Upsilon \sum_{i=1}^{m} \xi_i \tag{5}$$

where Υ is a user-defined, positive, regularization parameter

In the more general case in which the data points are not linearly separable in the input space, a nonlinear transformation is used to map the data vector *x* into a high-dimensional space (called feature space) prior to applying the linear maximum – margin classifier. A nonlinear operator $\phi(.)$ is used to map the input pattern into a higher dimensional space ***H***. The nonlinear SVM classifier so obtained can be defined as,

$$f(x) = w^T \phi(x) + b \tag{6}$$

which is linear in terms of the transformed data $\phi(x)$, but nonlinear in terms of original data $x \in R^n$. The nonlinear mapping function $\phi(.)$ never appears explicitly in the training process or in the determination of the decision hyper plane. Instead, it introduces implicitly through the Kernel Function K (. , .)

$$K(x,z) = \phi(x)^T \phi(z) \tag{7}$$

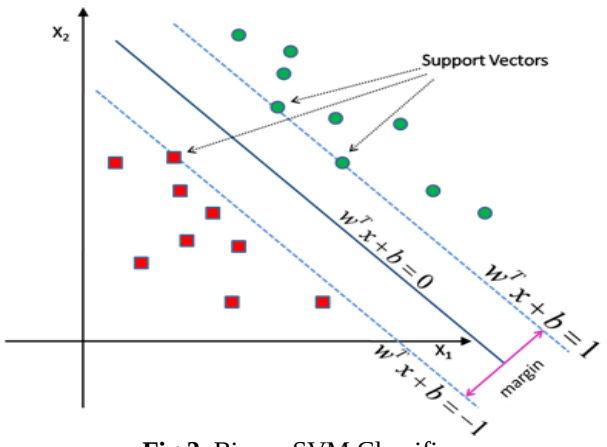

Fig.3 Binary SVM Classifier.

In terms of this kernel function, the SVM optimization problem can be formulated as:

$$\text{maximize} \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} y_i y_j \alpha_i \alpha_j K(x_i, x_j)$$

$$\text{Subject to} \quad \sum_{i=1}^{m} \alpha_i y_i = 0 \quad \text{and} \quad 0 \le \alpha_i \le \Upsilon, \ \textit{for all } i = 1, 2, \cdots, m \tag{8}$$

where α_i are the Lagrangian multipliers associated with each nonlinearity constraint of the optimization problem. x_i with nonzero α_i will be the support vectors. The kernel function in an SVM plays the central role of implicitly mapping the input vector (through an inner product) into a high-dimensional feature space, in which better separability is achieved. When a kernel function is selected, it is necessary to verify that it is associated with the inner product of some nonlinear mapping. In this paper, we consider two kernel functions; polynomial kernels and Gaussian RBF kernels. A dataset of 50 cases (25 Benign and 25 Malignant) was used for training the SVM and a dataset of 34 cases was used for testing. The results are discussed in next section.

IV. Results and Discussion

A total of 84 cases were analyzed in the study. Of these, 34 were benign and 50 were malignant. All malignant cases were from female patients. Of the 34 benign cases, 33 were from females and one from a male with gynecomastia. The age range of benign cases was from 14 to 68 years. The mean age of benign cases was 31.32 years. The cytological diagnosis of these cases were either fibroadenoma or fibroadenosis.

The age range of malignant cases was 31 to 90 years (Fig:4). The mean age of patients was 53.58 years. Cytological diagnosis of 49 cases was duct carcinoma and one case was lobular carcinoma which was confirmed by biopsy. Low to high grade carcinomas were included in the study.

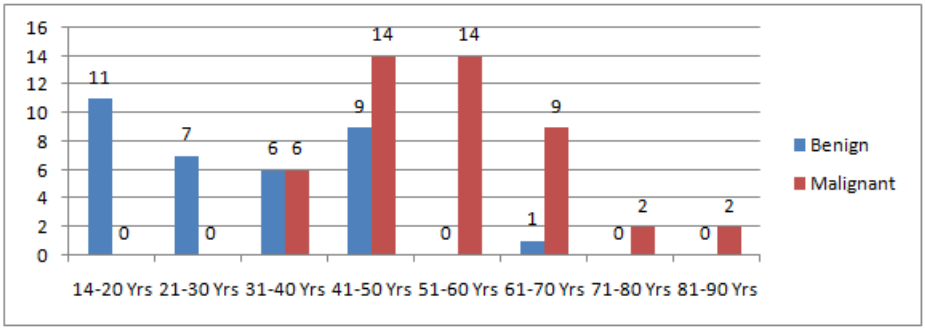

Fig:4 Age range of patients included in the study.

The area of the cell nuclei was determined by counting the total number of pixels in the nuclei. The variation in the area was less in benign cases compared to malignant cases. The mean area of the nuclei of benign cells varied from 366.8293 to 707.1111 pixels while in malignant cases it varied from 840.8205 to 2314.11 pixels. This was due to the marked variation in nuclear size in malignant cases. This is demonstrated in Fig.5.

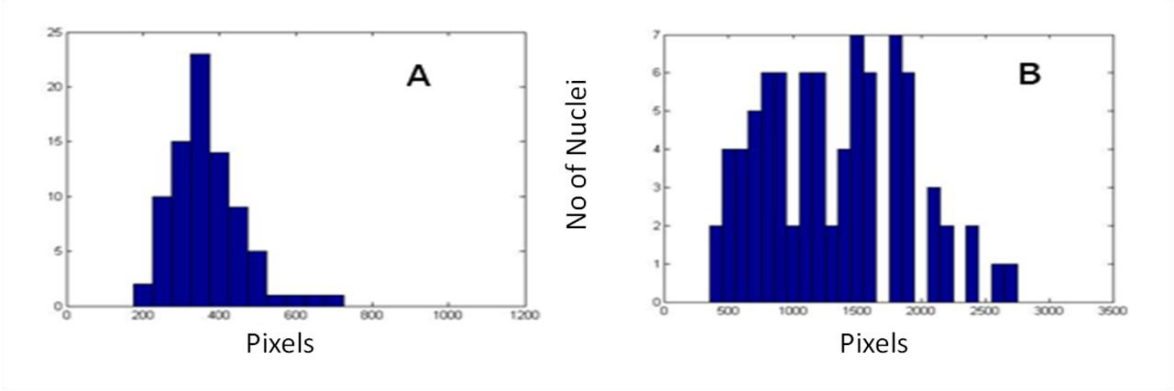

Fig:5 Histogram showing area of a typical (A) Benign and (B) Malignant case.

The comparison of mean nuclear area of 25 benign and malignant cases are shown in Fig. 6

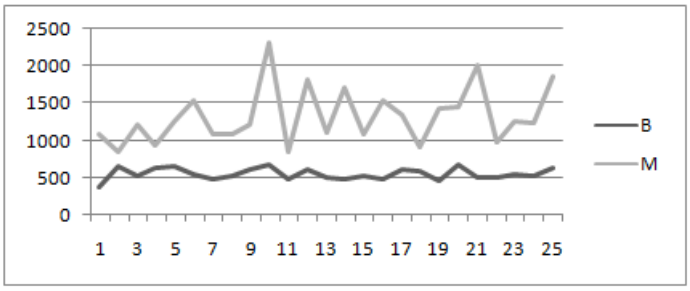

Fig: 6 The variation in mean nuclear area among 25 benign and 25 malignant cases.

Fig. 7 shows the variation in standard deviation of nuclear area among 25 benign and malignant cases

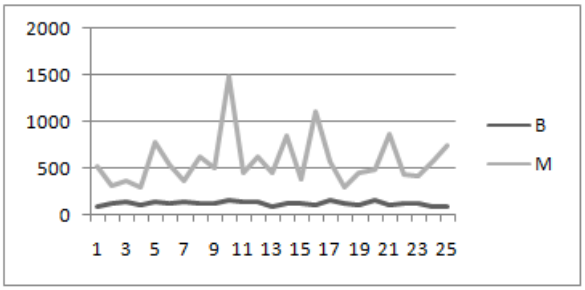

Fig: 7 The variation in SD of nuclear area among 25 benign and malignant cases.

Fig:8 shows variation in the total optical density of a typical benign and malignant case. Only minimal variation was seen in benign cases while marked variation with multiple peaks were seen in malignant cases, which is due to marked variation in DNA content of malignant nuclei.

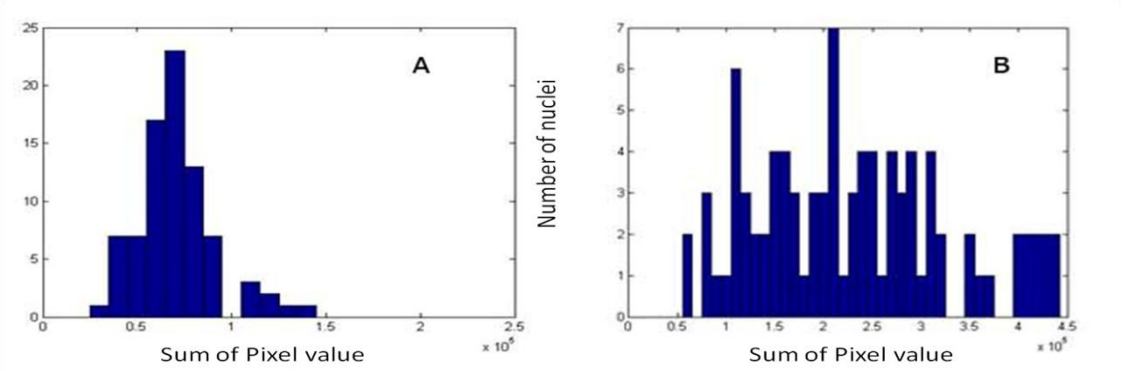

Fig: 8 The Total Optical Density (TOD) of a typical (A)benign and (B)malignant case.

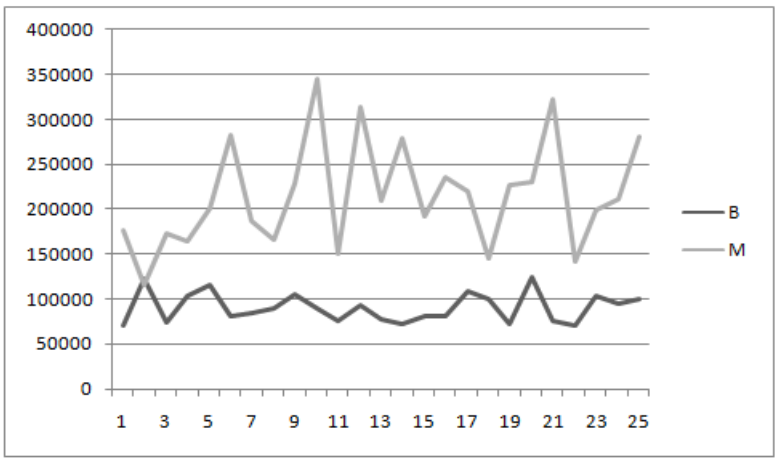

Fig: 9 Variation in the mean TOD of 25 benign and malignant cases.

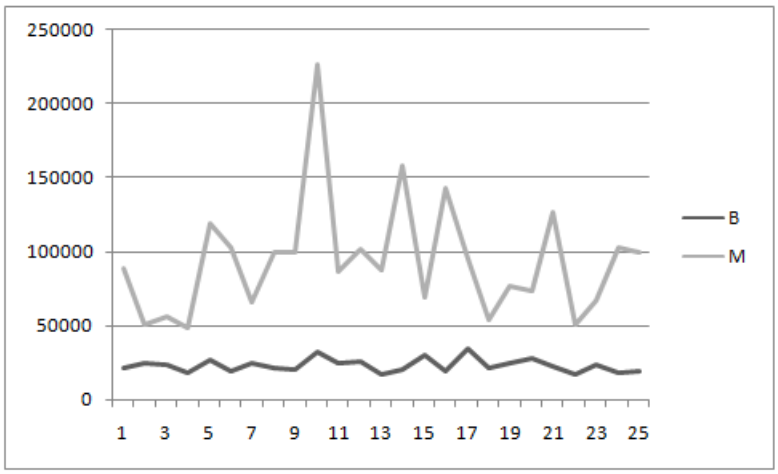

Fig: 10 Variation in standard deviation of TOD of nuclei of 25 benign and malignant cases.

Similar results were obtained for other parameters like mean and standard deviations of Hue, Saturation and Perimeter.

A dataset of 50 cases [25 Benign and 25 Malignant] was used for training the SVM. On an average more than 100 cells are analyzed in all the cases. In each of these nuclei, the mean and variance of the five features i.e., area, TOD, perimeter, hue and saturation were estimated. These ten parameters were given as input vector to SVM. Table 1 shows a sample data from 10 benign and malignant cases. During the training phase, tuning of Kernel parameters and slack penalty coefficient was required. SVM was trained and tested using Polynomial Kernel Functions of degree 3, 7 and 10, Gaussian Radial Basis Kernel with $\sigma = 0.5,1$ and 2.

A dataset of 34 cases was used in the testing phase. The actual nature of these cases, whether benign or malignant, was totally unknown to the person who analyzed the data. A Performance evaluation with different parameters was done by analyzing the number of True Positive [TP], True Negative [TN], False Positive [FP] and False Negative [FN] results (Benign representing Positive class and Malignant representing Negative class), and then calculating the efficiency of SVM classifier as given below.

$$Efficiency = \frac{True\ observations}{Total\ number\ of\ testdata}$$

Observations are given in Table 2. From the Model selection results, it is seen that Gaussian RBF Kernel Function with Υ =1 or 10 and σ =0.5, σ =1 and σ =2 gives the highest efficiency with the test data cases used. With the appropriate selection of data samples for training, we could get an accuracy of 100% on the test data.

	Mean Area	SD Area	Mean TOD	SD TOD	Mean Hue	SD Hue	Mean Sat	SD Sat	Mean Peri	SD Peri
Benign	366.829	95.454	70856.720	20865.738	273.558	76.984	215.373	64.085	78.362	9.946
Benign	655.292	122.897	122426.990	24356.545	466.700	88.986	385.478	75.334	106.205	11.754
Benign	521.984	133.211	73512.394	23206.875	413.651	114.615	193.633	69.271	92.539	12.971
Benign	636.163	105.363	104335.327	18525.342	229.064	117.151	96.580	29.321	103.048	8.571
Benign	654.849	141.819	116931.288	26537.175	450.658	188.997	107.078	34.187	105.465	11.660
Benign	554.568	123.915	80172.574	18863.127	380.618	86.682	331.982	76.936	96.757	10.971
Benign	483.234	135.547	84541.245	24593.898	360.606	104.248	185.630	47.655	90.279	13.071
Benign	528.273	119.480	89572.161	21414.164	385.328	90.899	273.252	54.753	94.879	11.351
Benign	613.906	118.973	105151.846	20774.723	434.140	87.349	350.960	71.390	102.970	10.881
Benign	675.544	154.241	90126.791	32469.313	444.874	101.573	260.746	87.030	106.937	13.217
Malignant	1076.857	536.127	176500.482	89128.616	749.540	371.798	693.442	348.660	133.234	36.529
Malignant	840.821	321.998	115819.628	51462.100	594.447	227.971	486.673	196.000	120.494	23.529
Malignant	1208.077	376.805	172529.628	56222.334	862.170	299.911	672.211	224.862	144.339	21.930
Malignant	932.818	311.372	165025.027	48773.074	721.705	234.052	272.336	81.276	126.205	21.288
Malignant	1259.942	784.783	200494.863	118862.364	858.318	528.393	455.538	283.677	144.225	45.664
Malignant	1533.691	552.319	282161.911	102972.667	1028.152	369.123	479.564	175.133	162.077	32.226
Malignant	1084.194	371.639	186291.065	66301.191	785.726	272.275	577.223	186.589	135.495	25.916
Malignant	1086.496	637.742	166211.256	99251.837	762.219	446.050	625.065	376.775	133.036	39.066
Malignant	1206.788	517.527	228395.404	99433.234	878.333	391.949	579.784	235.384	143.427	30.341
Malignant	2314.110	1492.717	344483.978	226629.101	1617.513	1044.310	1267.952	826.484	193.039	66.817

Table :1 Sample date from 10 benign and malignant cases.

Kernal	Υ	TP	TN	FP	FN	Efficiency (%)
Polynomial Order=3	0.1	8	24	0	2	94.18
	1	9	24	0	1	97.06
	10	9	24	0	1	97.06
Polynomial Order=7	0.1	9	24	0	1	97.06
	1	9	24	0	1	97.06
	10	9	24	0	1	97.06
Polynomial Order=10	0.1	9	17	7	1	76.47
	1	9	17	7	1	76.47
	10	9	17	7	1	76.47
Gaussian RBF σ = 0.5	0.1	10	24	0	0	100
	1	10	24	0	0	100
	10	10	24	0	0	100
Gaussian RBF σ = 1	0.1	10	24	0	0	100
	1	10	24	0	0	100
	10	10	24	0	0	100
Gaussian RBF σ = 2	0.1	10	24	0	0	100
	1	10	24	0	0	100
	10	10	24	0	0	100

Table 2 Results of SVM Model selection and classification.

V. Conclusion

In this paper nuclear DNA ploidy analysis was done from nuclear images and computer based diagnosis of cancer was made. This was compared with pathological diagnosis made by Pathologist. Ten parameters of the cell nuclei, i.e. mean and standard deviations of nuclear area, total optical density, hue, saturation and perimeter of nuclei were analyzed. A SVM Classifier was successfully trained to distinguish the malignant cells from the benign ones. In the present study, we have used Haematoxylin as a nuclear stain, cases were selected from the routine slides used for reporting. Feulgen is the nuclear stain used in most of the studies on image morphometry. Here we have proved that image morphometric analysis can be successfully applied on haematoxylin stained slides used for routine reporting, to enhance the diagnostic accuracy, even though unlike feulgen the binding of haematoxylin is not stoichiometric.

Acknowledgements

The project was funded by State Board of Medical Research, Govt. of Kerala, India, through Institutional Research Committee, Govt Medical College Thrissur

References

[1]. António E Pinto, Filipa Areia, Teresa Pereira, Paula Cardoso, Mariana Aparício, Giovani L Silva, Mónica C Ferreira, and Saudade André**, "**Clinical relevance of the reappraisal of negative hormone receptor expression in breast cancer", SpringerPlus 2013,2:375 Published online Aug 9, 2013. doi: 10.1186/2193-1801-2-375 (2013).

[2]. Ermiah E, Abdalla F, Buhmeida A, Alshrad M, Salem N, Pyrhönen S, Collan Y., **"**Prognostic significance of DNA image cytometry in Libyan breast cancer", Oncology.;83(3):165-76. doi: 10.1159/000339788. Epub 2012 Aug 15(2012)

[3]. Song T, Lee JW, Kim HJ, Kim MK, Choi CH, Kim TJ, Bae DS, Kim BG., "Prognostic significance of DNA ploidy in stage I endometrial cancer", Gynecologic Oncology, Volume 122, Issue 1, July 2011, Pages 79-82(2011) doi: 10.1016/j.ygyno.2011.03.017.

[4]. Bagwell CB, Clark GM, Spyratos F, Chassevent A, Bendahl PO, Stål O, Killander D, Jourdan ML, Romain S, Hunsberger B, Wright S, Baldetorp B.. "DNA and cell cycle analysis as prognostic indicators in breast tumors revisited".Clin Lab Med.;21:875-895 (2001).

[5]. G Bradley, EW Odell, S Raphael, JHo,LWLe, S Benchimol and S Kamel-Reid, "Abnormal DNA content in oral epithelial dysplasia is associated with increased risk of progression to carcinoma", British Journal of Cancer 103, 1432 – 1442, (2010). doi:10.1038/sj.bjc.6605905

[6]. Baron TH, Harewood GC, Rumalla A, Pochron NL, Stadheim LM, Gores GJ, Therneau TM, De Groen PC, Sebo TJ, Salomao DR, Kipp BR., "A Prospective Comparison of Digital Image Analysis and Routine Cytology for the Identification of Malignancy in Biliary Tract Strictures", Clinical Gastroenterology And Hepatology;2:214–21(2004). DOI: 10.1016/S1542-3565(04)00006-0

[7]. Danque PO, Chen HB, Patil J, Jagirdar J, Orsatti G, Paronetto F., "Image analysis versus flow cytometry for DNA ploidy quantitation of solid tumors: a comparison of six methods of sample preparation", Mod Pathol. May;6(3):270-5 (1993)

[8]. Charpin C, Andrac L, Devictor B, Habib M, Lavaut M, Allasia C, Bonnier P, Piana L., "Digital image-analysis of nuclear morphometry, dna-ploidy and AgNORs in breast-carcinoma cell imprints", International Journal of Oncology, November , Volume 3 Number 5 Pages: 949-956 (1993).

[9]. Mi LJ, Patil J, Hornbuckle WE, Cote PJ, Gerin JL, Tennant BC, Paronetto F., "DNA ploidy analysis of hepatic preneoplastic and neoplastic lesions in woodchucks experimentally infected with woodchuck hepatitis virus", Hepatology, Volume 20, Issue 1, pages 21–29, July (1994).

[10]. Fleskens SJ, Takes RP, Otte-Höller I, van Doesburg L, Smeets A, Speel EJ, Slootweg PJ, van der Laak JA., "Simultaneous assessment of DNA ploidy and biomarker expression in paraffin-embedded tissue sections", Histopathology, Volume 57, Issue 1, pages 14–26, July (2010). doi: 10.1111/j.1365-2559.2010.03599.x.

[11]. L. G. Koss, D. Thompson, and P. H. Bartels, Diagnostic Cytology and Its Histopathologic Bases. 5th Edition, Lippincott Williams and Wilkins, 2006.]

[12]. Marluce Bibbo and David Wilbur, Comprehensive Cytopathology , 3rd edition, ISBN: 978-1-4160-4208-2, Elsevier Health Sciences, 2008]

[13]. Ross JS, Linette GP, Stec J, Ross MS, Anwar S, Boguniewicz A., "DNA Ploidy and Cell Cycle Analysis in Breast Cancer", Am J Clin Pathol;120(Suppl 1):S72-S84 (2003). DOI: 10.1309/QD096UGF70T5H46G

[14]. Barr Fritcher EG, Kipp BR, Slezak JM, Moreno-Luna LE, Gores GJ, Levy MJ, Roberts LR, Halling KC, Sebo TJ, "Correlating Routine Cytology, Correlating routine cytology, quantitative nuclear morphometry by digital image analysis, and genetic alterations by fluorescence in situ hybridization to assess the sensitivity of cytology for detecting pancreatobiliary tract malignancy", Am J Clin Pathol. 2007 Aug;128(2):272-9.

[15]. Raouf. N.G. Naguib and H. Sakim, "DNA ploidy and cell cycle distribution of breast cancer aspirate cells measured by image cytometry and analyzed by ANN for their prognostic significance,"IEEE Transactions on Information Technology in Biomedicine,vol. 3, pp. 61–69, March 1999. DOI:10.1109/4233.748976

[16]. Baolin Wu, Tom Abbott, David Fishman, Walter McMurray, Gil Mor,Kathryn Stone,David Ward,Kenneth Williams and Hongyu Zhao,"Comparison of statistical methods for classification of ovarian cancer using mass spectrometry data", Oxford Journal on Bioinformatics (2003) 19 (13): 1636-1643. doi: 10.1093/bioinformatics/btg210

[17]. Liyang Wei; Yongyi Yang; Nishikawa, R.M.; Yulei Jiang, "A study on several Machine-learning methods for classification of Malignant and benign clustered microcalcifications," IEEE Transactions on Medical Imaging, vol.24, no.3, pp.371,380, March 2005 doi: 10.1109/TMI.2004.842457

[18]. Li Rong; Sun Yuan, "Diagnosis of Breast Tumor Using SVM-KNN Classifier," Second WRI Global Congress on Intelligent Systems (GCIS), vol.3, no., pp.95,97, 16-17 Dec. 2010 doi: 10.1109/GCIS.2010.278

[19]. El-Naqa, I; Yongyi Yang; Wernick, M.N.; Galatsanos, N.P.; Nishikawa, R.M., "A support vector machine approach for detection of microcalcifications," IEEE Transactions on Medical Imaging, vol.21, no.12, pp.1552,1563, Dec. 2002 doi: 10.1109/TMI.2002.806569

[20]. Ruey-Feng Chang, Wen-Jie Wu, Woo Kyung Moon, Yi-Hong Chou, Dar-Ren Chen, "Support Vector Machines for Diagnosis of Breast Tumors on US Images",Academic Radiology,Volume 10, Issue 2, February 2003, Pages 189–197, DOI: 10.1016/S1076-6332(03)80044-2

[21]. Mehmet Fatih Akay, "Support vector machines combined with feature selection for breast cancer diagnosis", Expert Systems with Applications: An International Journal, Volume 36 Issue 2, March, 2009, Pages 3240-3247 , doi:10.1016/j.eswa.2008.01.009

[22]. Chan, T.F.; Vese, L.A, "Active contours without edges," IEEE Transactions on Image Processing , vol.10, no.2, pp.266-277, Feb 2001 doi: 10.1109/83.902291

[23]. Jeroen A W M van der Laak, Automated Identification of Cell and Tissue Components in Pathology, ISBN 90-1234567-0, Department of Pathology,University Medical Center St. Radboud University Nijmegen, The Netherlands. (2001)

[24]. C. Cortes and V. Vapnik, "Support-vector networks," Machine Learning,20, 273-297 (1995), Kluwer Academic Publishers, Boston.

[25]. Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze, An Introduction to Information Retrieval, Cambridge University Press Cambridge, England, Online edition, 2009 Cambridge UP

[26]. Vineetha Nandakumar, Prasad P.H., Sheeba V.S., "A Support Vector Machine Approach for Detection of Malignancy Using DNA Ploidy Analysis," Fourth IEEE International Conference on Advances in Computing and Communications (ICACC) 2014, vol., no., pp.138-142, 27-29 Aug. 2014 doi: 10.1109/ICACC.2014.39.