Pathology & Oncology Research (2019) 25:777–790
https://doi.org/10.1007/s12253-019-00582-8

ORIGINAL ARTICLE

A Novel Approach of Mathematical Theory of Shape and Neuro-Fuzzy Based Diagnostic Analysis of Cervical Cancer

Subrata Kar[1] · Dwijesh Dutta Majumder[2]

Received: 13 February 2018 / Accepted: 15 January 2019 / Published online: 6 February 2019

Abstract

This study aims to detect the abnormal growth of tissue in cervix region for diagnosis of cervical cancer using Pap test of patients. The proposed methodology classifies cervical cancer for pattern recognition either benign or malignant stages using shape and neuro-fuzzy based diagnostic model. In this experiment, firstly the authors segment Pap smear images of cervical cells using fuzzy c-means clustering algorithm and shape theory to classify them according to the presence of abnormality of the cells. Secondly the features extraction process is performed in the part of nucleus and cytoplasm on the squamous and glandular cells and the authors used input variables such as cytoplasm area (CA), cytoplasm circularity (CC), nucleus area (NA), nucleus circularity (NC), nucleus-cytoplasm ratio (NCR), and maximum nucleus brightness (MNB) in fuzzy tools and used fuzzy rules to evaluate the cervical cancer risk status as an output variable. The proposed neuro-fuzzy network system was developed for early detection of cervical cancer. A neural network was trained with 15-Pap image datasets where Levenberg–Marquardt(LM) a feed-forward back-propagation algorithm was used to get the status of the cervical cancer. Out of 15 samples database, 11 data set for training, 2 data set for validation and 2 data set for test were used in the ANN classification system. The presented fuzzy expert system(FES) successfully identified the presence of cervical cancer in the Pap smear images using the extracted features and the use of neuro-fuzzy system(NFS) for the identification of cervical cancer at the early stages and achieve a satisfactory performance with 100% accuracy.

Keywords Cervical cancer · Pap smear images segmentation · Features extraction · Shape theory · Neuro-fuzzy classification system

Introduction

In our study Mathematical Shape theory and Neuro-Fuzzy based methodology that are used in screening test of Pap smear images for the abnormal growth of tissue in the cervical cells at the early stages. In our previous work, we have made a discussion that nanoscale (1×10^{-9} to 100×10^{-9}m)-sized particles are of great importance because of their biophysical and chemical properties at atomic /molecular scale level as reported by Majumder [1, 2].This comparative investigative study has been done in understanding the measurement of benignancy and malignancy based work by Kar and Majumder [3]. Physicians are able to detect the abnormal growth of tissue in a CT scan and MR images of the brain in the very early precancerous stages. Several shape-based features like shape distance (SD) and shape similarity measure (SSM) of brain of the above-mentioned patterns are extracted from CT scan and MR images for classification results. We have experimented with the development of a medical expert system for diagnosis and treatment using soft computing approaches such as the fuzzy inference system (FIS) and ANN for decision-making in oncology in breast cancer, brain cancer and cervical cancer.. We have developed a new approach to decision-making in oncology which has been successfully applied to the early diagnosis and therapy planning of breast cancer and brain cancer. In this study, we propose to apply the same approach to the early detection and treatment planning of cervical cancer. Pap smear

✉ Subrata Kar
subrata.bacchu@gmail.com

Dwijesh Dutta Majumder
drdduttamajumder@gmail.com

[1] Department of Mathematics, Dumkal Institute of Engineering & Technology, Murshidabad, West Bengal Pin-742406, India

[2] Department of Electronics and Communication Sciences Unit, Indian Statistical Institute, Kolkata, West Bengal Pin-700108, India

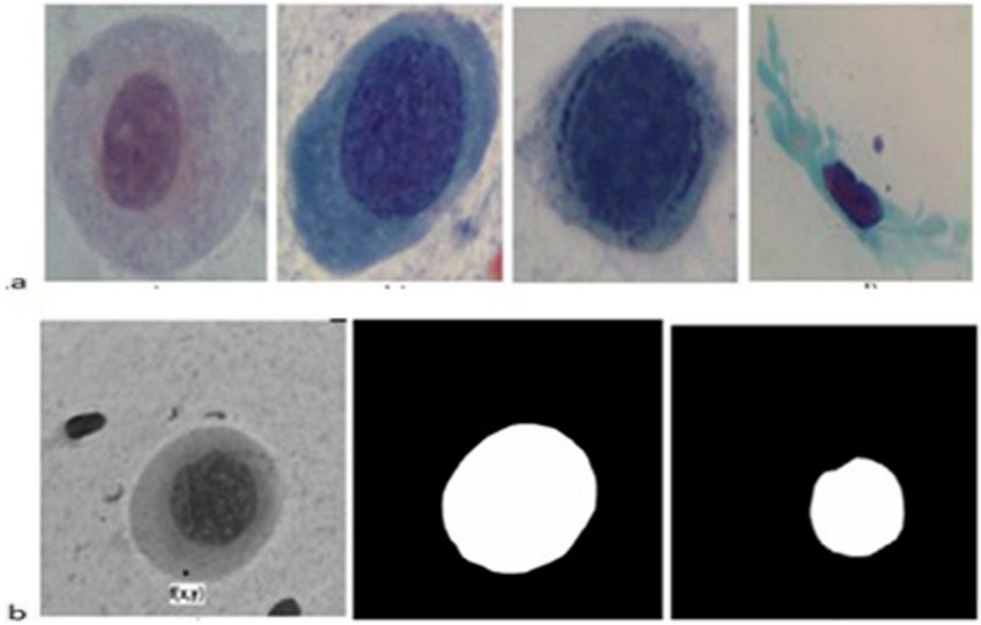

Fig. 1 **a** Basal cells: Normal cell, Moderate dysplasia (CIN 2), Severe dysplasia (CIN3) and CIS (Carcinoma-in-situ) **b** Image segmentation of a cervical cell: original cell, cytoplasm and nucleus

test of a patient is very important medical diagnosis to detect pre-cancerous changes in the cervical cells and findings the important risk factor for cervical cancer caused by HPV infection [4]. In the first stage, we have made a segmentation procedure of Pap smear images using fuzzy c-means clustering algorithm [5–7] with the help of two types of random number generators corresponding membership matrix for each pixel.

In this study, three validity measures such as partition coefficient(PC), partition entropy(PE) and compactness and separation function(SC) are used for getting good clustering results. Mathematical theory of shape are used here to classify them according to the presence of abnormality in the structural behaviuor of the cells [8]. Secondly, we extracted features such as cytoplasm area (CA), cytoplasm circularity (CC), nucleus area (NA), nucleus circularity (NC), nucleus-cytoplasm ratio (NCR), and maximum nucleus brightness (MNB) which are taken from the part of nucleus and cytoplasm on the squamous cells and glandular cells [9–11].These extracted features are considered as input variables and fed it into fuzzy tools and using fuzzy rules to evaluate the cervical risk status as output variable. In the third stage, the extracted features datasets are classified using neuro-fuzzy classification system and we successfully recognized the abnormal cells such as squamous intraepithelial neoplasia (SIL) and differentiated them from

Table 1 Data description

Pap images /Extracted features	Cytoplasm area (CA)	Cytoplasm circularity (CC)	Nucleus area (NA)	Nucleus circularity (NC)	Maximum nucleus brightness (MNB)	Nucleus-cytoplasm ratio (NCR)
Pap 1	680,029	0.7300	17,795.2	0.8520	102.6	38.976
Pap 2	50,446	0.7866	50,861.3	0.8630	90.33	10.723
Pap 3	700,000	0.8	120,000	0.7	95.2	32.4
Pap 4	450,000	0.5400	40,295	0.856	92	30
Pap 5	505,045	0.52	42,430	0.827	94	33
Pap 6	258,753	0.6480	112,275	0.8340	107.2	2.2290
Pap 7	194,935	0.7225	59,347.6	0.8400	90.75	3.3854
Pap 8	580,560	0.56	45,525	0.83	93	35
Pap 9	375,000	0.5	65,000	0.5	90	25
Pap 10	178,631	0.2487	40,367.2	0.6980	77.75	4.7378
Pap 11	620,450	0.62	47,425	0.84	96.13	36
Pap 12	650,700	0.65	48,456	0.85	95.12	34.2
Pap 13	560,402	0.58	44,350	0.812	92	32
Pap 14	640,720	0.74	54,350	0.83	94	34
Pap 15	630,420	0.72	52,450	0.82	96	35

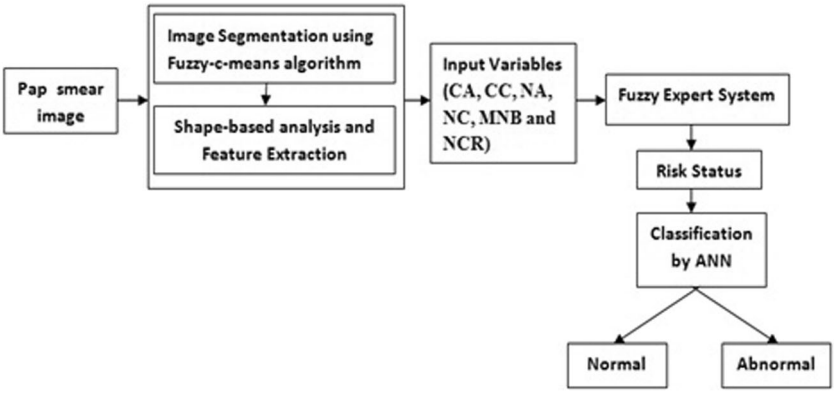

Fig. 2 Block diagram of Pap smear microscopic images classification process by using neuro-fuzzy system

the normal epithelial cells. We have taken a 15-sample database for the proposed work and experiment fed into Neuro-Fuzzy System for classification of cervical cancer. We have shown the early diagnosis of cervical cancer using a neuro-fuzzy system(NFS) where LM is a feed forward back propagation learning algorithm used to train the NN for a satisfactory performance results.

Materials and Methods

The methodology is divided into three phases. Firstly cervical cells are segmented using FCM clustering algorithm. The segmented images of the cervical cells are analyzed with the help of shape theory to classify them as normal or abnormal behaviour of the cells. Secondly the features extraction process is performed in the part of nucleus and cytoplasm area on the squamous cells and glandular cells. Thirdly, the extracted features like CA, CC, NA, NC,NCR, and MNB are classified using neuro-fuzzy system as benign or malignant in nature (Fig. 1).

Data Description

For screening, out of 78 cervical cell images [12], we have taken 15 sample database from the website Leica Microsystems http://www.leica-microsystems.com/ [10] for the proposed work and experiment for classification of cervical cancer (Table 1).

Fuzzy C-Means Algorithm in Pap Smear Image Segmentation

In Pap smear test, cervical cells are collected from the surface of the cervix using a brush or spatula and are smeared to into a slide. The slide is then stained Papaniclolaou method. The staining makes it possible to observe the characteristics of the cells under a microscope [11]. A stained slide when looked upon under a microscope shows the cell nuclei and the cytoplasm along with the background. The cytologists look for any cellular change in the slides. The factors that indicate any abnormality are the change in the shape and size of the cell nuclei, increasing nucleus-cytoplasm area ratio (N/C ratio), and increasing chromatin content of the nucleus etc. We

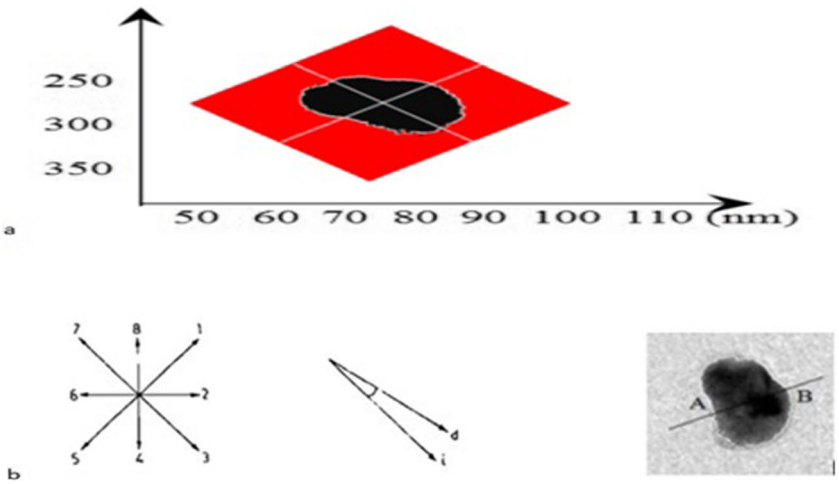

Fig. 3 **a** Contour with major axis and two intersecting points with the contour **b** Directional codes, Angle between d and i is (i-d)45^0 and Two reference points A and B

Table 2 Fuzzy sets of input variables

Input field	Range	Fuzzy set
Cytoplasm area	50,000–250,000	Very Small
	150,000–450,000	Small
	300,000–700,000	Medium
	500,000–800,000	High
Cytoplasm circularity	0–0.4	Very Small
	0.2–0.6	Small
	0.4–0.9	Medium
	0.7–1	High
Nucleus area	10,000–30,000	Very Small
	20,000–40,000	Small
	30,000–60,000	Medium
	50,000–120,000	High
Nucleus circularity	0–0.5	Very Small
	0.3–0.7	Small
	0.5–0.9	Medium
	0.8–1	High
Maximum nucleus brightness	60–80	Very Low
	70–100	Low
	85–105	Medium
	95–120	High
Nucleus cytoplasm ratio	0–10	Very Low
	5–25	Low
	15–45	Medium
	30–50	High

mounted a high resolution camera on microscope through an adapter. The 400x magnified microscopic images are shot by the camera and passes into computer memory for analysis. We consider colour Pap smear images in RGB colour [13] channel for analysis. An image is nothing but a collection of pixels and each pixel having a particular value for the three colour channels red (R), green (G) and blue (B). Hence to find the different areas inside the image, we consider the image as a set of data having pixel values R, G and B. This dataset (x_1,x_2,x_3....., x_n) is classified using Fuzzy c-mean (FCM) clustering algorithm to distinguish the different regions inside image namely nucleus and cytoplasm [14]. To initiate the clustering process we generate random numbers corresponding to each of the R, G and B value. Three kinds of random number generator are used, general random number generator, chaos function and chaos function with timestamp. These random numbers constitute the membership value (μ) of the pixels. The μ value of the pixels are compared with that of the cluster center and classified accordingly. The clustering process segments the images namely into three classes: cytoplasm, nucleus and the background.

Features Extraction of Pap Smear Image

Our main work focuses on the segmentation of the cell nuclei and the cytoplasm of the microscopic Pap smear images of uterine cervix and classification of normal or abnormal cells such as Cervical intra-epithelial neoplasia (CIN) and Carcinoma in Situ (CIS) is based on Neuro-Fuzzy system [15]. Fuzzy c-means algorithm used to segment the images and generalized shape theory is applied to classify the images. The extracted features from the part of nucleus and cytoplasm of the Squamous cells (exocervix) and Glandular cells (endocervix) are cytoplasm area (CA), cytoplasm circularity (CC), nucleus area (NA), nucleus circularity (NC),nucleus-cytoplasm ratio (NCR), and maximum nucleus brightness (MNB). Fifteen samples are taken and fed it into Neuro-Fuzzy classification system [8, 12] (Fig. 2).

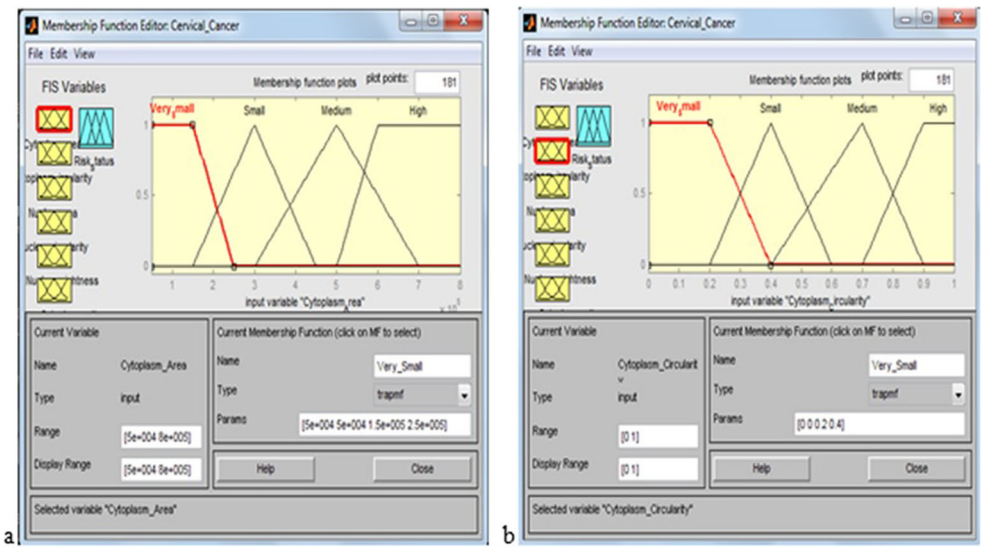

Fig. 4 Membership for the input variables **a** cytoplasm area **b** cytoplasm circularity

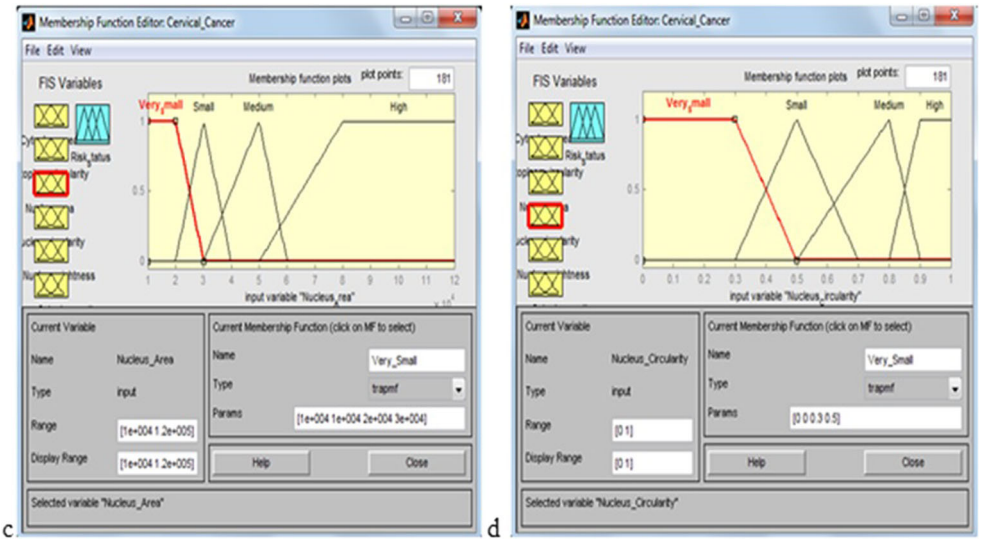

Fig. 5 Membership for the input variables **c**. nucleus area **d** nucleus circularity

Shape Based Analysis of Pap Smear Image

The perception of shape has been used for pattern recognition, computer vision, shape analysis, and image registration. Here we shall consider shape analysis and shape based similarity measures based on Dutta Majumder's generalized shape theory, shape distance and shape metric approach [5, 16]. The Generalized shape theory uses co-ordinate transformation of landmark points of the region of interest (ROI) in the respective images. Whereas shape metrics and shape distance uses degrees of match between the corresponding shapes of the two nucleus images.

Generalized Shape Theory

In Generalized Shape theory we align on the basis of some invariant landmark points on the boundary of the region of interest (ROI) by means of translation and/or scaling and/or rotation. We consider a geometrical figure X in R^k space containing N control points that can be represented by X: N× K matrix. Now two shapes, X and X′ are of the same shape if they are related by the following rigid body transformation equation $X' = \beta X\Gamma + I_N\nu\Gamma$ where Γ: K × K is a rotation matrix and $|\Gamma| = 1$, $I_N = N \times 1$ of one, ν= K ×1 is a translation vector, β: isotropic scaling factor and > =0. It is possible to formulate an approximate co-ordinate transformation for mapping between two sets of landmarks in a least square sense using Taylor series expansion. For two sets of landmark points (x_m, y_m) and (x_m', y_m'), m = 1,2,…,n, one set can be expressed in terms of other as follows:

$$x' = q_0 + q_1x + q_2y + q_3x^2 + q_4xy + q_5y^2 + \quad (1)$$

$$y' = r_0 + r_1x + r_2y + r_3x^2 + r_4xy + r_5y^2 + \quad (2)$$

The linear components of the Eqs. (1) and (2) are related to the affine part and the higher order terms q_3, q_4, q_5, r_3, r_4, r_5 are related to the non-affine (non-linear) part of the transformation equations. So, we can say that affine and projective

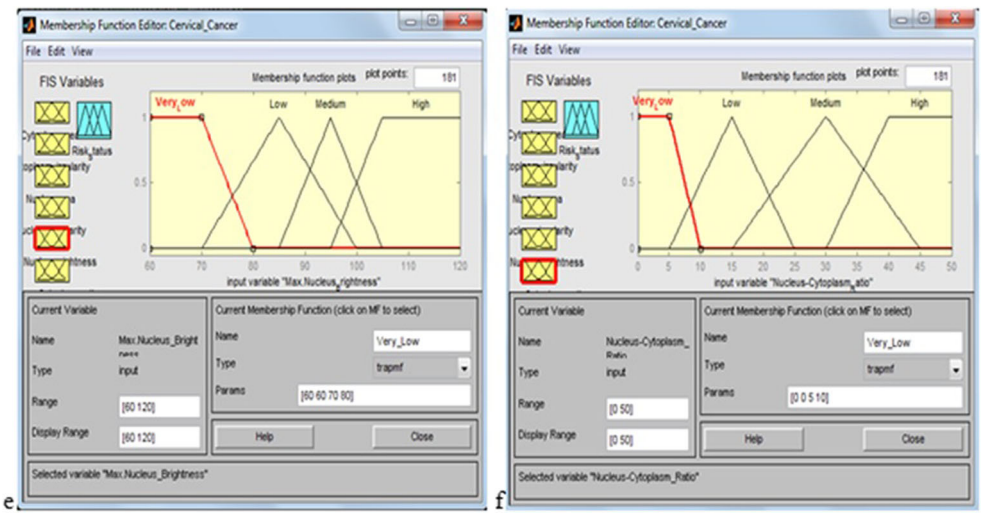

Fig. 6 Membership for the input variables **e**. maximum nucleus brightness **f** nucleus cytoplasm ratio

Table 3 Fuzzy set and membership function of output variable "Risk status of cervical cancer"

Output field	Range	Fuzzy set
Risk status of cervical cancer	0–60	Low
	50–100	High

transformations are two special cases of generalized shape theory. In affine transformation, one needs at least 3 landmark points as it has 3 DOF (Degree Of Freedom), whereas projective transformation needs at least 4 control points. Our study has shown that even in the case of change in structural (shape) properties, the algorithm is general enough to capture the changes. With the change, re-registration may be necessary with different set of invariant points.

Shape Metric and Shape Distance in Cancer Cell Image Registration

We present a theoretic approach to define shape and shape distance in our endeavor to apply the concept in cytological image registration. The shape of an object is defined as a subset X in R^2 if (i) X is closed and bounded. (ii) Interior of X is non-empty and connected. (iii) Closure property holds on interior of X. An object, Y in R^2 has the same shape as X in R^2 if it preserves translation, rotation and scaling invariance. In terms of set, these three transformations can be represented as follows:

$$\text{Translation} : Y = \{(x + a), (y + b) : x, y \in X\},$$

where a, b are real numbers.

$$\text{Rotation} : Y = \{P_1(\alpha).P_2\ (\beta)\ X\}$$

where P_1 & P_2 are rotation around x and y axes respectively, $\alpha.\ \beta\ \in[0,2\pi)$

$$\text{Scaling} : Y = \{(kx, ky) : x, y \in X\},$$

where k is a + ve real number.

Each of translation, dilation and rotation defines an equivalence relation on F. If R is an equivalence relation on F and objects A,B,C ∈F then: Reflexivity: (A, A) ∈ R, Symmetry: if (A, B) ∈ R ⇒ (B, A) ∈R, Transitivity: if (A, B) ∈ R ⇒ (B, C) ∈R then (A, C) ∈R hold under each of translation, rotation and dilation of shape transformation.

Shape Distance Measure

Distance d_1 between shape X and Y in F is defined as: $d_1(X,Y) = m_2[(X\text{-}Y)\cup(Y\text{-}X)]$, where m_2 is Lebesgue measure in R^2 and d_1 satisfies all the following metric properties of algebra [1, 17].

i) $d_1(X,Y) \geq 0$
ii) $d_1(X,Y) = 0$ if and only if $X = Y$
iii) $d_1(X,Y) = d_1(Y,X)$ and
iv) $d_1(X,Y) + d_1(Y,Z) \geq d_1(X,Z)$

For shape similarity measure we consider two cells are of same shape if and only if one of there image is translation, dilation and rotation of other. In order to register two cells images, it is necessary to normalize the form of images of interest in terms of position, size and orientation. After normalization, shape distance is measured in terms of volume of mismatch. The shape is described on the basis of its structural features using certain chain codes with respect to one reference point. Reference points are obtained from the intersection points of the major axis with the contour of

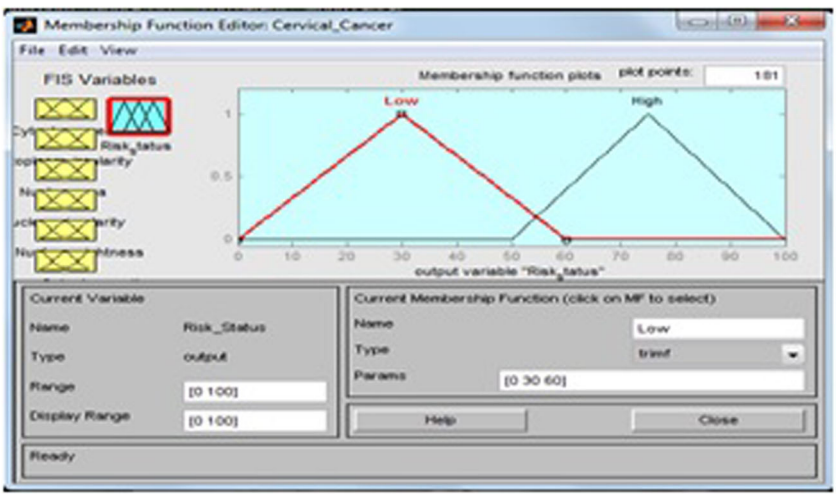

Fig. 7 Membership function for Risk status of cervical cancer

Table 4 Rule base of the system

Rule no.	Cytoplasm area (CA)	Cytoplasm circularity (CC)	Nucleus area (NA)	Nucleus circularity (NC)	Maximum nucleus brightness (MNB)	Nucleus-cytoplasm ratio (NCR)	Cervical cancer risk status
Rule1	Very Small	Very Small	Very Small	Very Small	Very Low	Very Low	Low
Rule2	Very Small	Very Small	Very Small	Very Small	Very Low	Very Low	High
Rule23	Medium	High	High	High	High	High	Low
Rule29	High	Medium	Medium	Medium	Medium	Medium	Low
Rule32	High	High	High	High	High	High	High

the region, which is invariant under translation, rotation, and dilation of the region and the major axis is unique [11].

Shape Similarity Measure

The degree of matching between shapes is measured in terms of shape distance function [7]. The shape theory as discussed by Dutta Majumder et al. states that, two objects are of same shape if and only if one is translation, dilation and rotation of other. In order to match two images, it is necessary to normalize the form of images of interest in terms of position, size and orientation. After normalization, shape distance is measured in terms of volume of mismatch. The shape is described on the basis of its structural features using certain chain code with respect to one reference point. Reference points are obtained from intersection points of the unique major axis with the contour of the region, which is invariant under translation, rotation, and dilation of the region. The shape in terms of chain code is an equivalence class in the space R^K and satisfies the equivalence class relation with another region B in R^K. Convex polygonal approximations of objects in terms of chain code are aligned after normalizing with respect to translation, rotation and scaling. The centroid of an object A, C(A), is equivariant under all of the above mentioned transformations, i.e. C(f(A)) = f(C(A)) for all **f** in F, and F is closed under composition with translation. The centroid (x_g, y_g) of the contour is given by ***n*** number of points

$$\text{as}: x_g = \frac{1}{n}\sum_{j=1}^{n} x_j \text{ and } y_g = \frac{1}{n}\sum_{j=1}^{n} y_j$$

and in polar coordinate the shortest perpendicular distance from the point(x_j, y_j) to a straight line $y = x\tan\alpha + c$ is the absolute value of $c\cos\alpha - r_j\sin(\alpha - \alpha_j)$. The least square method is applied to the boundary points to find the axis of the region which makes an angle α to the ***x*** axis and passes through the

Table 5 Tested values of the system

Pap images / Extracted features	Cytoplasm area (CA)	Cytoplasm circularity (CC)	Nucleus area (NA)	Nucleus circularity (NC)	Maximum nucleus brightness (MNB)	Nucleus-cytoplasm ratio (NCR)	Cervical cancer risk status (%)
Pap 1	680,029	0.7300	17,795.2	0.8520	102.6	38.976	50
Pap 2	50,446	0.7866	50,861.3	0.8630	90.33	10.723	50
Pap 3	700,000	0.8	120,000	0.7	95.2	32.4	50
Pap 4	450,000	0.5400	40,295	0.856	92	30	50.4
Pap 5	505,045	0.52	42,430	0.827	94	33	50.3
Pap 6	258,753	0.6480	112,275	0.8340	107.2	2.2290	50
Pap 7	194,935	0.7225	59,347.6	0.8400	90.75	3.3854	50
Pap 8	580,560	0.56	45,525	0.83	93	35	50.4
Pap 9	375,000	0.5	65,000	0.5	90	25	50
Pap 10	178,631	0.2487	40,367.2	0.6980	77.75	4.7378	50
Pap 11	620,450	0.62	47,425	0.84	96.13	36	50.4
Pap 12	650,700	0.65	48,456	0.85	95.12	34.2	50.4
Pap 13	560,402	0.58	44,350	0.812	92	32	50.4
Pap 14	640,720	0.74	54,350	0.83	94	34	50.4
Pap 15	630,420	0.72	52,450	0.82	96	35	50.4

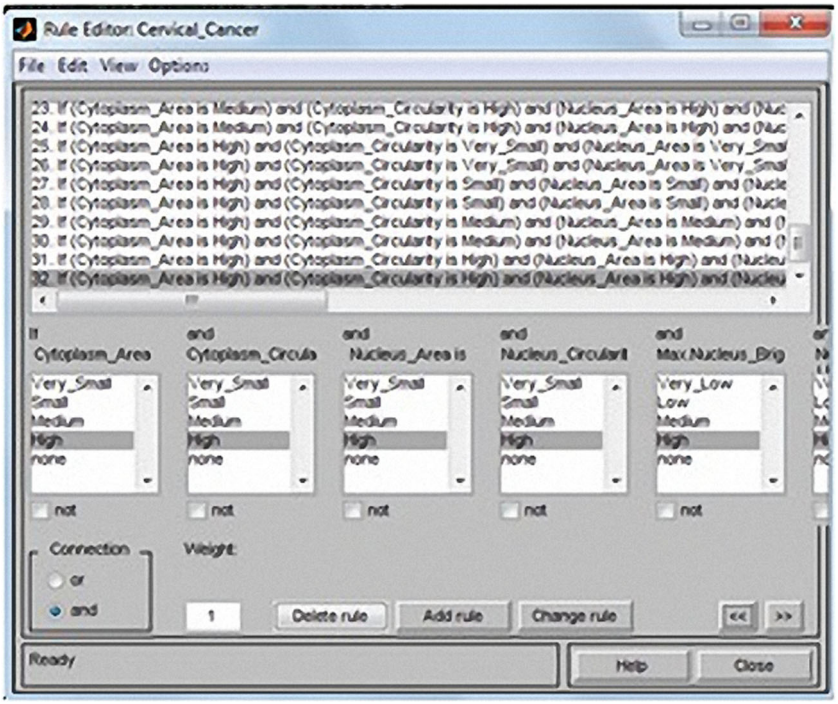

Fig. 8 Developed fuzzy rules

centroid (x_g, y_g). In polar coordinate, the square error is non-negative real number.

$$f(\alpha, c) = \frac{1}{n}\sum_{j=1}^{n}\left(c\cos\alpha + r_j\sin\left(\alpha - \alpha_j\right)\right)^2$$

where $\alpha \in [0, \pi)$ and r_j should be minimum for required conditions. The slope of axis (α) is obtained from the best linear fit solution by minimizing f (α, c) passing through (x_g,y_g). and making an angle α with $\boldsymbol{x}$ axis. The slope satisfies the relation

$$tan2\alpha = \frac{2\sum_{j=1}^{n}\left(x_j - x_g\right)\left(y_j - y_g\right)}{\sum_{j=1}^{n}\left(x_j - x_g\right)^2 - \sum_{j=1}^{n}\left(y_j - y_g\right)^2},$$

$$\cos2\alpha\left\{\sum_{j=1}^{n}\left(x_j - x_g\right)^2 - \left(y_j - y_g\right)^2\right\} + 2\sin2\alpha\sum_{j=1}^{n}\left(x_j - x_g\right)\left(y_j - y_g\right) \geq 0$$

Feature Extraction of the Boundary of the Region of Interest

To extract the feature of the boundary of the region of interest(ROI) it is helpful to represent the closed contour with a set of direction codes. The direction code may be taken among "n" selected points on the contour, which has same distance between any two consecutive points. The direction d makes an angle(i-d) 45°with direction i, where real number d ∈ 1 to 8 and i = (1,2,..8) (Fig. 3).

Let $d_m = (d_{ij})^{n}{}_{j=1}$ where m = A, B, be the contour starting from each reference point A and B and denoted by d_A and d_B

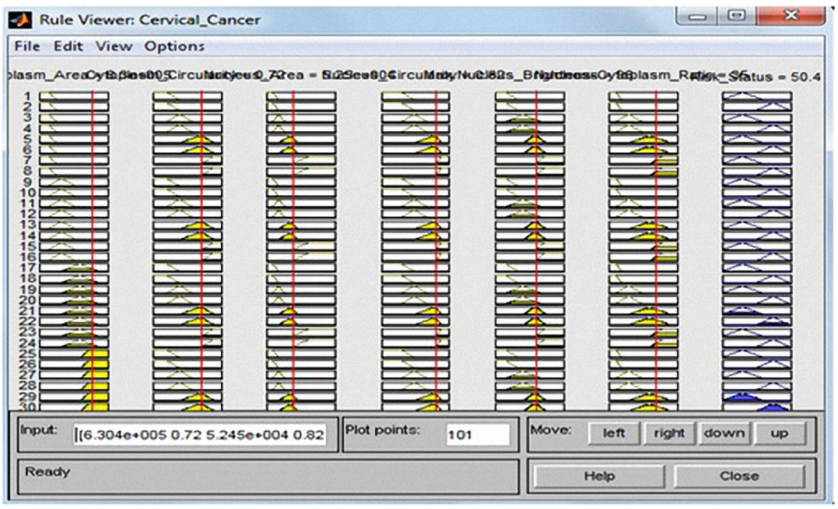

Fig. 9 Calculation of the value cervical cancer risk(CCR) for the values CA = 630,420, CC = 0.72, NA = 52,450, NC = 0.82, MNB = 93, NCR = 35

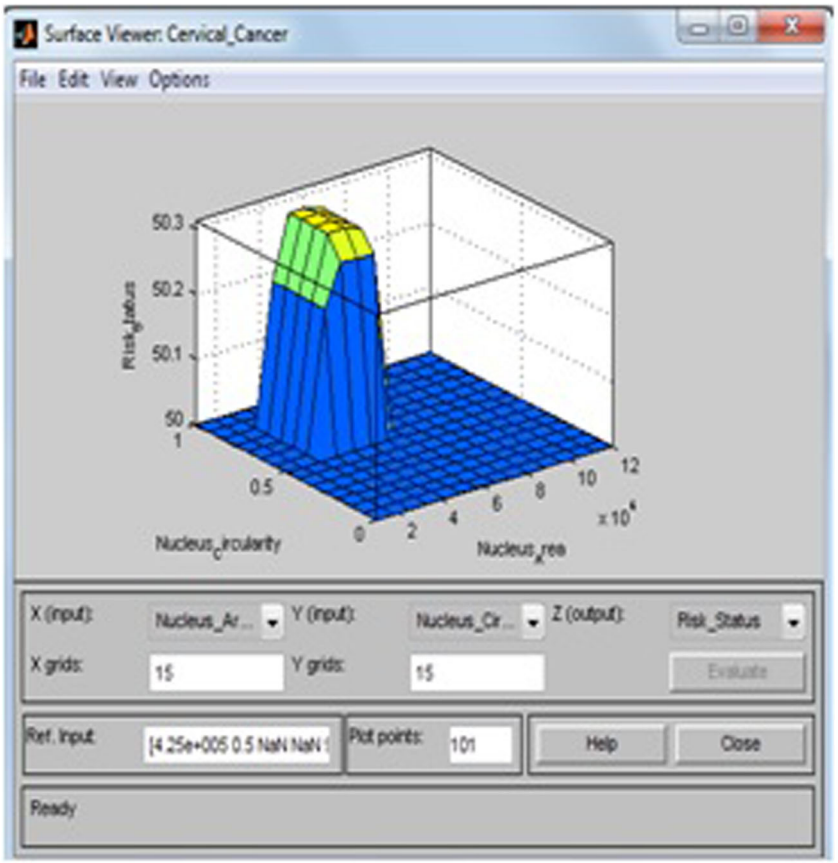

Fig. 10 Surface viewer of nucleus area and nucleus circularity

respectively. If d_2 is a rotation of d_1 then $d_2 = d_1 + \gamma$ for any real number γ [16]. For all j we can write $d_{2\,j} = d_{1\,j} + \gamma \ \forall\, j$. The distance function D, in terms of the direction code between the contour of interest and the model is defined as $D(d_1,d_2) = \sum_{j=1}^{n} \min\left((d_{1j} - d_{2j}), 8 - (d_{1j} - d_{2j})\right)$, The normalized value of D is D/n and the shape similarity measure between the two shapes is given by [18], $\mu = 1 - \frac{D}{n}$, where $0 \le \mu \le 1$ and $0 \le D < 1$, smaller value of D indicates higher degree of Similarity [5, 16].

Classification of Cervical Cancer

In this section we propose the shape theoretic and neuro-fuzzy based approach for pattern classification of cervical cancer.

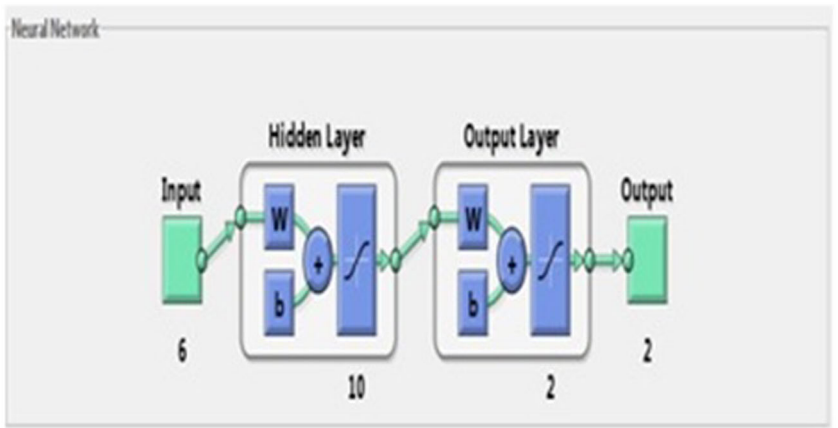

Fig. 11 Neural network model

Fuzzy Inference System

Fuzziness is relevant to complexity because the nature of a system when its complexity exceeds a certain threshold it becomes impractical or computationally infeasible to make precise assertions about it. Fuzzy control is based on fuzzy logic – a logical system which is much closer in spirit to human thinking and natural language than traditional logical systems. The fuzzy logic controller (FLC) based on fuzzy logic provides a means of converting a linguistic control strategy based on expert knowledge into an automatic control strategy [19]. The Fuzzy Inference System (FIS) has a vital role in the medical field to provide medical assistance to the radiologist to diagnose the abnormality in the medical images [20]. Decision Support Systems through information technology become a part of classification of cervical cancer.

Input Variables

Extracted features from the part of nucleus and cytoplasm of Squamous and Glandular cells of the cervix are cytoplasm area (CA), cytoplasm circularity (CC), nucleus area (NA), nucleus circularity (NC), nucleus-cytoplasm ratio (NCR), and maximum nucleus brightness (MNB) considered as input variables. Membership functions of them are trapezoidal and triangular fuzzy numbers. Fuzzy sets, Range sets and Membership functions of the input variables are identified in Table 2 and Figs. 4, 5 and 6.

For example, the membership functions of the Fuzzy Sets of the Cytoplasm circularity input variable are given below and the remaining membership functions will follow the same.

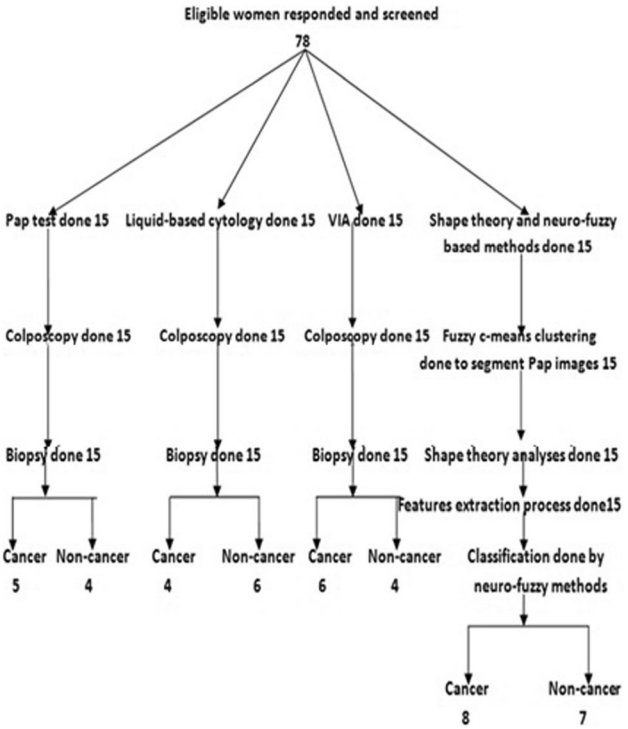

Fig. 12 Cervical cancer screening by Pap test, Liquid-based cytology, VIA and Shape theory and neuro-fuzzy based methodology along with outcome of confirmatory diagnosis

$$\mu_{\text{V.Small}}(x) = \begin{cases} 1 & x <= 0.2 \\ (0.4-x)/0.2 & 0.2 < x < 0.4 \\ 0 & x > 0.4 \end{cases} \quad \mu_{\text{Small}}(x) = \begin{cases} 0 & x <= 0.2 \\ (x-0.2)/0.2 & 0.2 < x < 0.4 \\ 1 & x = 0.4 \\ (0.6-x)/0.2 & 0.4 < x < 0.6 \\ 0 & x >= 0.6 \end{cases}$$

$$\mu_{\text{Medium}}(x) = \begin{cases} 0 & x <= 0.4 \\ (x-0.4)/0.3 & 0.4 < x < 0.7 \\ 1 & x = 0.7 \\ (0.9-x)/0.2 & 0.7 < x < 0.9 \\ 0 & x >= 0.9 \end{cases} \quad \mu_{\text{High}}(x) = \begin{cases} 0 & x <= 0.7 \\ (x-0.7)/0.2 & 0.7 < x < 0.9 \\ 1 & x >= 0.9 \end{cases}$$

Output Variable

Risk Status of Cervical Cancer The Risk Status of 'cervical cancer' is to classify the benignancy or malignancy of cervical tumor lesions of patients. Membership function of them are triangular fuzzy numbers. Fuzzy sets, Range sets and Membership functions of 'Risk Status' of cervical cancer are identified in Table 3 and Fig. 7.

Table 6 Pap smear images confusion matrix

Actual	Predicted	
	Cancerous (Positive)	Non-cancerous (Negative)
Cancerous(Positive)	8 (TP)	0 (FP)
Non-cancerous(Negative)	0 (FN)	7 (TN)

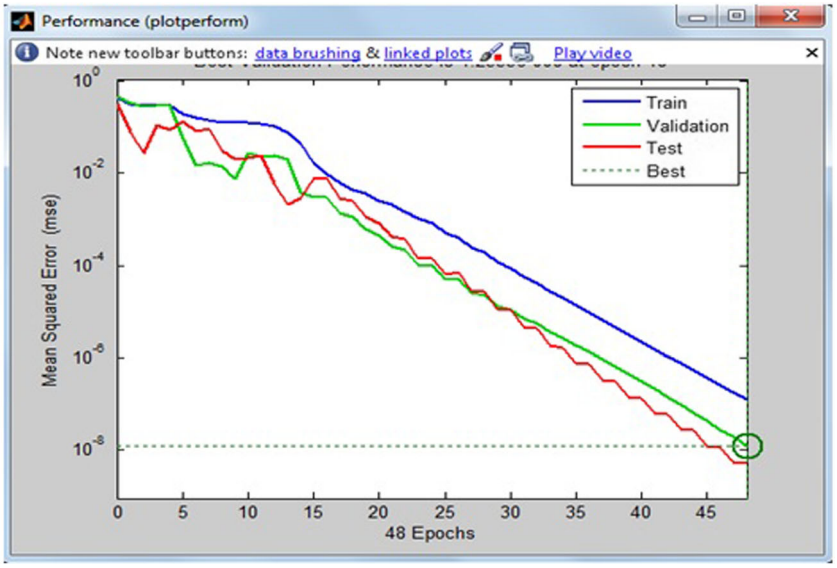

Fig. 13 Results from neural network

Fuzzy Rule Base

The Fuzzy rule base consists of 15 rules that determine the Risk status (Low Risk, Medium Risk, High Risk) by evaluation of the input variables CA, CC, NA,NC,NCR, and MNB and it plays an important role in further treatment decision. The Rule base of the system is shown in Table 4.

Defuzzification

In this stage, truth degrees (α) of the rules are determined for the each rule by aid of the min and then by taking max between working rules. For example, for CA = 630,420, CC = 0.72, NA = 52,450, NC = 0.82, MNB = 96, NCR = 35 the Rule 4 and Rule 30 will be fired and we will obtain: α_4 = min(Very Small CA, Small CC, Small NA, Small NC, Low MNB, Low NCR) = min(0, 0, 0, 0, 0.267, 0) =0, α_{30} = min(High CA, Medium CC, Medium NA, Medium NC, Medium MNB, Medium NCR) = min(1, 0.9, 0.755, 0.8, 0.9,0.67) = 0.67. From Mamdani max-min inference we will obtain the membership function of our system as max(α_4, α_{30}) = max(0, 0.67) = 0.67 that means Very High Cervical cancer risk. Then we can calculate the crisp output. The crisp value of the CCR is

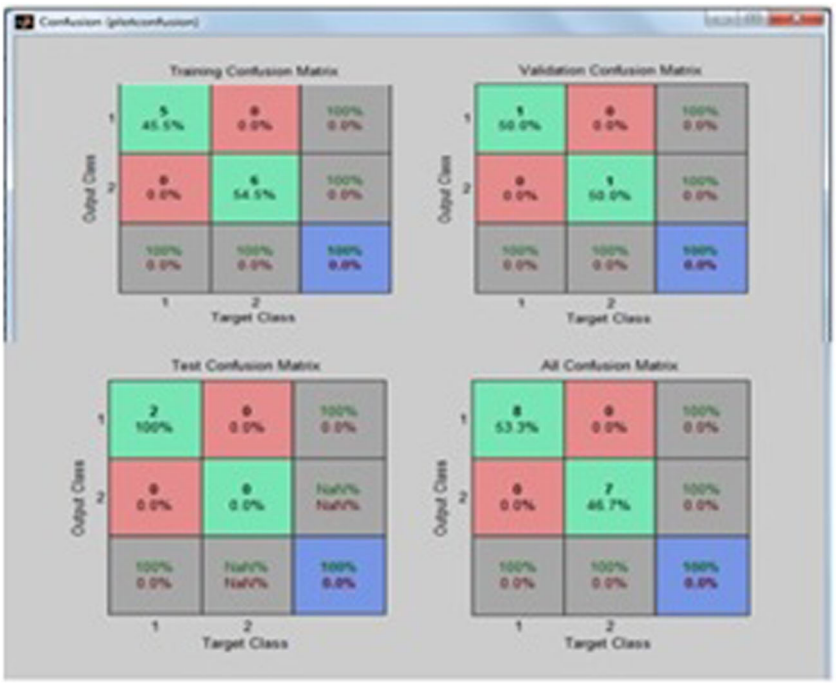

Fig. 14 Confusion matrix of the neural network

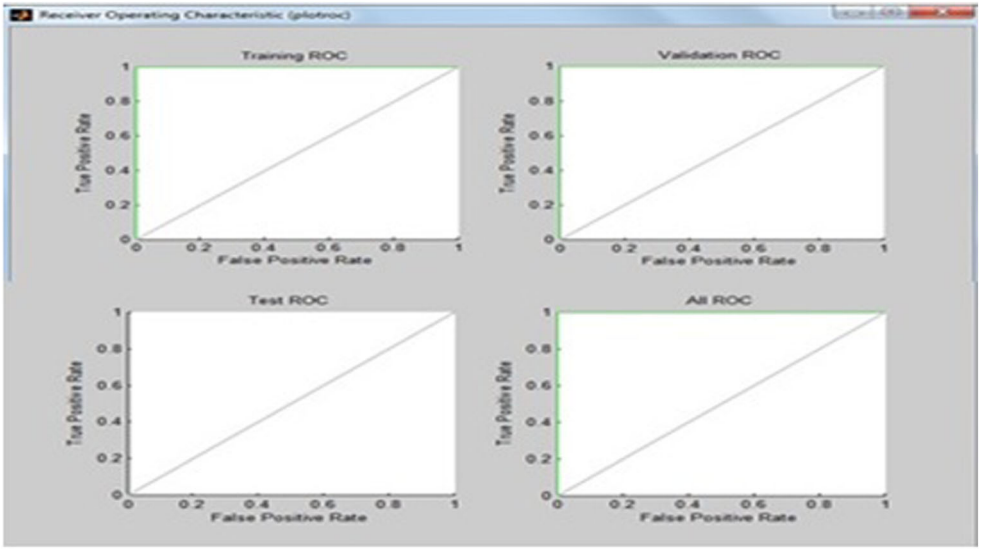

Fig. 15 ROC curve

calculated by the method center of gravity defuzzifier by the formula:

$$z^* = \frac{\int z\mu_{middle}(z)dz}{\int \mu_{middle}(z)dz}$$

The tested values of the system are shown in Table 5. Mamdani inference engine is used for rule evaluation purpose. The proposed fuzzy expert system uses 32 rules in designing fuzzy rule base as shown in Fig. 8.

The rules have been developed using *if-then* method. Figure 9 shows a rule view of this proposed design. For example, from the Fig. 8, if cytoplasm area(CA) is high, cytoplasm circularity(CC) is high, nucleus area(NA) is high, nucleus circularity is high, maximum nucleus brightness(MMB) is high and nucleus -cytoplasm ratio is high then output risk is high. Using these rules, the result risk in term of percentage (%) has been computed. Surface views of this design are presented in Fig. 10. As also seen from the Fig. 9 the value of Cervical cancer risk (CCR) = 50.4 This means that the patient has the cervical cancer with a possibility 50.4%. Because this is a quite high percentage and it's tendency goes to malignancy stage and doctor has to decide a biopsy.

Neuro-Fuzzy Classifier and Back-Propagation Learning Algorithm

A neuro-fuzzy classifier is used to diagnose the patients cervical cancer disease at early stage. The ANN is a network that interconnect all nodes. The input of a specific node is the weighted sum of the output of all the nodes to which it is connected. The output value of a node is a non-linear function which is known as 'transfer function' or it is called as ' sigmoid function' $f(x) = \frac{1}{1+e^{-x}}$ of its input values. The nodes in the input layer receive the data and transfer them to nodes in the first hidden layer through the weighted links. The data which are processed in the first hidden layer of the network the result is transformed to the nodes in the next layer. The nodes in the last layer provide the network's output. The j-th node in a hidden layer processes the incoming data x_i by (i) calculating the weighted sum and adding a "bias" term θ_j as: $net_j = \sum_{i=1}^{m} x_i w_{ij} + \theta_j$,(j = i,2,.....n), where w_{ij} is the weight of the link from the j-th input node to the i-th hidden node. (ii) transforming the net_j through the sigmoid function $f(x) = \frac{1}{1+e^{-x}}$ and (iii) transferring the result to nodes in the next layer. We used here the feed forward neural network where the

Table 7 Confusion matrices of various screening methods for detecting cervical cancer

Actual	Predicted	
	Cancerous	Non-cancerous
Cancerous	5 (TP)	2 (FP)
Non-cancerous	4 (FN)	4 (TN)
Confusion matrix of Pap test model		
Actual	Predicted	
	Cancerous	Non-cancerous
Cancerous	4 (TP)	2 (FP)
Non-cancerous	3 (FN)	6 (TN)
Confusion matrix of liquid-based cytology model		
Actual	Predicted	
	Cancerous	Non-cancerous
Cancerous	6 (TP)	2 (FP)
Non-cancerous	3 (FN)	4 (TN)
Confusion matrix of VIA model		

Table 8 Performance of various screening modalities for detecting cervical cancer

Test characteristic	Pap test (%)	Liquid-based cytology (%)	VIA %)	Shape theory and neuro-fuzzy-based methodology (%)
Sensitivity(Se)	55.5	57.1	66.6	100
Specificity(Sp)	66.6	75	66.6	100
Positive predictive value(Ppv)	71.42	67	75	100
Negative predictive value(Npv)	50	67	57.14	100

Pap test Papanicolaou test, *VIA* Visual inspection of cervix using acetic acid

network connections are allowed only between the nodes in one layer and those in the next layer. Back propagation algorithm is the most widely used learning algorithm to train multilayer feed forward network and applied for applications like character recognition, image processing, pattern classification, medical diagnosis etc. We used here the Levenberg-Marquardt (LM) feed forward back- propagation learning algorithm which is used to train the Neural Network for the diagnosis of cervical cancer. The Training of these networks consists in finding a function between a set of input values and a set of output values. This function is accomplished by adjusting the value of the weights w_{ij}, using a learning algorithm. After the weights are adjusted on the training set, their value is fixed and the ANN's are used to classify unknown input images. In the Levenberg-Marquardt (LM) back-propagation learning algorithm we have to minimize the error term defined as $E_p = \frac{1}{2}\sum_j \left(t_{pj} - o_{pj}\right)^2$ where the index p corresponds to one input vector and the vectors t_p and o_p are the target and observed output vectors corresponding to the input vector p respectively. The proposed neural network with 6 number of inputs is simulated with a set of 10 numbers of hidden neurons and the network is evaluated with a learning function LM in Fig. 11 which shows a model of the neural network with 10 hidden neurons (Fig. 12).

Results

We have developed an FES and NFS for the early detection of cervical cancer from the ROI of cervical tumor lesions data of patients, where the NFS shows its ability to distinguish the risk factors of various patients in a more precise way. The experimental results show that six input features like cytoplasm area, cytoplasm circularity, nucleus area, nucleus circularity, nucleus– cytoplasm ratio and maximum nucleus brightness of cervical cells in 15 Pap smear images based neural network is giving 100% classification rate. The confusion matrix of Pap smear images presented in the Table 6.The performance measures are calculated for six input features of cervical cells in Pap smear images based NN are shown in Figs. 13, 14 and 15. The proposed network was trained with 15 cases Pap smear images data sets. These 15 cases are fed to the ANN with 6 input neurons, one hidden layer of 10 neurons and 2 output neurons. Out of 15 samples database, 11 data set for training, 2 data set for validation and 2 data set for test were used in the ANN classification system. From the Pap smear images confusion matrix the number of output of data sets of True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN) are 8,0,7, 0 respectively. The Sensitivity, Specificity and Accuracy are equal to 100% each.

A Comparative Study of Cervical Cancer Screening Methods

A number of different tools or methods have been developed and investigated over the years as alternative screening tests for cervical cancer such as Pap smear test, Liquid-based cytology, Visual inspection of the cervix with acetic acid(VIA), Shape theory and neuro-fuzzy based methodology [4]. They have served extensively for screening of cervical cancer. The diagnostic analysis of cervical cancer at early stages of different methods are summarized and reviewed in detail in the following Fig. 12, Tables 7 and 8. The experimental results have proved that shape theory and neuro-fuzzy based methodology is a superior detection method of cervical cancer compared to the current well-established methods [21].

Discussions

The proposed method used Fuzzy c-means clustering algorithm approach with enhanced initialization procedure to segment Pap smear images of cervical cells. The segmented images of the cervical cells were analyzed with the help of a generalised mathematical theory of shape to classify them according to the presence of abnormality in the morphological behaviour of the cells. Features extraction process is performed in the part of nucleus and cytoplasm on the squamous cells and glandular cells of the cervix. A neuro-fuzzy system(NFS) is developed where the extracted features such as cytoplasm area, cytoplasm circularity, nucleus area, nucleus circularity, nucleus– cytoplasm ratio and maximum nucleus

brightness considered as input variables are classified successfully to indicate malignant or benign in nature.

In conclusion the method of cervical cancer diagnosis presented in this paper is a successful model in assisting the doctors in the screening treatment for early detection of cervical cancer. This paper presented Fuzzy c-means clustering algorithm in Pap smear images by which cervical cells are segmented and analyzed with the help of shape theory to classify them accordingly to the presence of abnormality in the morphological behaviour of the cells. Secondly the features extraction process is performed in the part of nucleus and cytoplasm on the Squamous cells (exocervix) and Glandular cells (endocervix). In the third stage, the extracted features data are classified using neuro-fuzzy system and we successfully recognized the abnormal cells such as squamous intraepithelial neoplasia (SIL) and differentiated them from the normal epithelial cells. The results we have obtained from the clinical findings and texture features analysis of Pap smear images the most important enhancement is to establish a direct relationship between the actual cell dimensions and the pixel dimension of the digital images. The morphological information of the cells is used for the pattern classification system. The experimental results have proved that shape theory and neuro-fuzzy based methodology is a superior detection method of cervical cancer compared to the current well-established methods. We intend to extend these methodologies for early detection of cancer in other regions such as ovarian region and human liver.

Acknowledgements We acknowledge with thanks the help and a cooperation we received from our colleges at Dumkal Institute of Engineering & Technology, Murshidabad, West Bengal, India and Indian Statistical Institute, Kolkata, India.

Funding This study was funded by us.

Compliance with Ethical Standards

Conflict of Interest Statement The authors declare that they have no conflict of interest.

Research Involving Human Participants and/or Animals This article does not contain any studies with human participants or animals performed by any of the authors.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Jantzen J, Dounias, G (2006) Analysis of PAP smear image data. Proc. of Nature Inspired Smart Information System, NISIS 2006, pp 1–10
2. Majumder DD, Bhattacharya M (1998a) A new shape based technique for classification and registration: application to multimodal medical images. Int J of Image Processing and Communication 4(3–4):45–70
3. Sreedevi MT, Usha BS, Sandya S (2012) Pap smear image based detection of cervical cancer. Int J Comput Appl 45(20):0975–8887
4. Shah Sunny S et al (2016) Current technologies and recent developments for screening of HPV-associated cervical and orpharyngeal cancers. Cancers 8(85):1–27
5. Banarjee DK, Parui SK, Majumder DD (1994) A shape metric for 3-D objects. Indian Journal Pure Appl Math 25(1&2):95–111
6. Majumder DD (2015) Physics, chemistry and biology of nanoscale bio-technology-need for a unified framework for nanoarchitectures, DNA structures and nano-computing. EKHOMANJARI: a magazine on Science and Culture, BKC College, Kolkata, pp 13–32
7. Parui SK, Majumder DD (1982) A new definition of shape similarity. Pattern Recogn Lett 1:87–120
8. Plissiti ME, Nikou C, Charchanti A (2010) Automated detection of cell nuclei in pap smear images using morphological reconstruction and clustering. IEEE Trans Inf Technol Biomed 15(2):233–241
9. Genctav A, Aksoy S, Onde S (2012) Unsupervised segmentation and classification of cervical cell images. Pattern Recogn 45:4151–4168
10. Leica Microsystems, http://www.leica-microsystems.com/
11. Mahanta LB, Nath CK, Karan S, Nath DC, Majumder DD, Sharma JD (2011) Fuzzy mathematical and shape theoretic approach to cervical cell classification. Int J Comput Appl (0975–8887), 36, (6):1–5
12. Gallegos-Funes FJ, Gomez-Mayora ME, Lopez-Bonilla JL, Cruz-Santiago R (2009) Rank M- type radial basis function (RMRBF) neural network for pap smear microscopic image classification. Apeiron 16(4):542–554
13. Rosidi B, Jalil N, Pista NM, Ismail LH, Mengko ETL (2011) Classification of cervical cells based on labeled colour intensity distribution. International Journal of Biology and Biomedical Engineering 5(4):230–238
14. Talukdar J, Nath CK, Talukdar PH (2013) Fuzzy clustering based image segmentation of pap smear images of cervical cancer cell using FCM algorithm. Int. Journal of Engineering and Innovative Technology(IJEIT) 3(1):460–462
15. Plissiti ME, Nikou C, Charchant A (2011) Combining shape, texture and intensity features for cell nuclei extraction in pap smear images. Pattern Recogn Lett 32(6):838–853
16. Majumder DD (1995) A study on a mathematical theory of shapes in relation to pattern recognition and computer vision. Indian Journal of Theoretical Physics 43(4):19–30
17. Kar S, Majumder DD (2017) A mathematical theory of shape and neuro-fuzzy methodology- based diagnostic analysis: a comparative study on early detection and treatment planning of brain cancer. Int J Clin Oncol 22(4):667–681
18. Byriel J (1999) Neuro-fuzzy classification of cells in Cervical smears, M.Sc Thesis, Technical University of Denmark, 1999
19. Majumder DD, Majumdar KK (2004) Complexity analysis, uncertainty management, and fuzzy dynamical systems: a cybernetic approach. Kybernetes 33(7):1143–1184
20. Majumder DD, Bhattacharya M (2000) Cybernetic approach to medical technology: application to cancer screening and other diagnostics. Kybernetes 29:871–895
21. Wen C, Ming-Yu L, Horng-Jyh H, Yen-Chem H, Chien-Hung C, Kun-His T, Chi-Hung H (2009) Automatic segmentation of abnormal cell nuclei from microscopic image analysis for cervical cancer screening. 2009 IEEE 3rd International Conference on Nano/Molecular Medicine and Engineering, tainan, 18–21 Oct 2009