A New Approach to Microcalcification Detection Using Fuzzy Soft Set Approach

SREEDEVI S
Research Scholar, University of Kerala,
Thiruvananthapuram, Kerala
Sathyansree123@gmail.com

ELIZABETH SHERLY
Professor
Indian Institute of Information Technology
and Management, Kerala
sherly@iiitmk.ac.in

Abstract

This paper presents a new computer aided detection method for identifying malignant images in digital mammograms using fuzzy soft set theory approach. Fuzzy soft set theory is a mathematical model based on parameterization concept for solving uncertainties and we have implemented a more efficient decision making method using fuzzy soft aggregation operator. This method helps the radiologists to classify normal and malignant images from a set of mammogram images and also gives information regarding the intensity of malignancy in each abnormal image. After the preprocessing, images are segmented using hierarchical fuzzy c-means clustering with features incorporated. Among the 14 features, 9 important features are selected for fuzzy soft set implementation. Experiments are performed on publicly available MIAS dataset and the results are evaluated with the help of a radiologist. This algorithm shows a recognition accuracy of 93.97% in the detection of malignant images.

Keywords: Digital Mammography, fuzzy soft set theory, fuzzy c-means, NL-means, fuzzy soft aggregation operator

1. Introduction

Breast cancer is one of the most common diseases in the present world and is an important cause of death among women over the age of 40 [1]. The presence of microcalcification clusters in mammographic x-ray image has been considered as an important indication of breast cancer, and its detection at its earlier stage is important to prevent and treat the disease [2]. Due to the subjective or varying decision criteria, it is very difficult for a human radiologist to make correct decisions in microcalcification detection in digitized mammograms. But the performance can be increased if there exist an efficient method to help the radiologist in the detection process. For this purpose computer-aided diagnosis (CAD) systems are being developed for improving the performance of the radiologist in the detection or classification of abnormalities present in digital mammogram images. Using CAD systems, radiologists can identify the breast cancer at an early stage and make their decisions more accurately with minimum time and cost.

Various methods have been developed for microcalcification detection and are widely used for diagnosing breast cancer at its earlier stage. Spiesberger[3] developed a computer-aided mammographic screening algorithm in which he used a decision tree to characterize the candidates. The characterization is based on the brightness, compactness, and statistical measures. Then the presence of microcalcifications is measured based on a value called cross-correlation coefficient and a threshold value which is set as 0.65. If the cross-correlation coefficient is greater than the threshold value then it can be declared that microcalcification is present. Davies et al. [4], [5][6] segment clustered microcalcifications based on a local threshold value. Cheng et al [7] proposed a method to detect microcalcification using fuzzy logic. Maji et al [8] describes the application of soft set theory to a decision making problem using rough sets.

This paper proposes a new automatic method for classifying malignant images in a set of mammograms using fuzzy soft set method. This method provides an effective method for solving problem of knowledge representation in an uncertain and imprecise environment. From the recent studies it is clear that new techniques can be developed to assist radiologists for making correct decisions using quantitative method derived from a mathematical model for mammogram [9]. The proposed method helps the radiologist to identify the malignant images from a set of images and also helps to understand the intensity of malignancy in each abnormal image for further processing. The proposed approach consists of the following five major steps: noise removal, contrast enhancement, pectoral muscle removal, segmentation, feature extraction, feature selection, fuzzification, and fuzzy soft set implementation.

2. Theoretical background of Fuzzy Soft Set Theory(fs-sets)

Traditional tools for modeling, computing and reasoning are not applicable for some real life problems that involve data which are not clearly defined, uncertain and not crisp. Such complicated real life problems can be solved with the use of mathematical principles based on uncertainty and imprecision. To handle such uncertainties in an effective way, a number of theories have been proposed by various researchers. Some important and commonly used theories are fuzzy set theory [10], vague sets [11], rough set theory [12] etc. These theories may be considered as efficient mathematical tools for dealing with uncertainties and imprecision embedded in a system. But all these theories have their own limitations which is the lack of the parameterization tool associated with these theories. To overcome the above limitations, Molodtsov[13] put forward the concept of soft set theory as a mathematical tool for dealing with uncertainties. Maji et al. [14] introduced the concept of fuzzy soft sets in decision making problems, which is a hybrid model of fuzzy sets and soft sets.

The idea of fuzzy logic was introduced by Professor L. A. Zadeh and it was considered as the most appropriate theory, for dealing with uncertainties. Assume that we have a fuzzy set A, and if an element x is a member of this fuzzy set A, this mapping can be denoted as

$$\mu_A(x)\in[0,1],\left(A=\left(x,\mu_A(x)\mid x\in X\right)\right) \tag{1}$$

In this work, let U denote the initial universal set, E be a set of parameters, $P(U)$ is the power set of U and $A\subseteq E$.

A pair (F, E) is called a soft set over a given universal set U, if and only if F is a mapping of a set of parameters E into the power set of U. That is $F:E\to P(U)$.Clearly; a soft set over U is a parameterized family of subsets of a given universe U. In soft sets, the parameter sets and the approximate functions are crisp. But in $fs-sets$, while the parameter sets are crisp, the approximate functions are fuzzy subset of U. Let Γ_A,Γ_B , $\Gamma_C,\ldots$ etc refers to $fs-sets$ and $\gamma_A,\gamma_B,\gamma_C,\ldots\ldots.,etc$ refers to their fuzzy approximate functions[15].

Definition 2.1

An fs-set Γ_A over U is a set defined by a function γ_A representing a mapping $\gamma_A:\mathrm{E}\to\mathrm{F}(U)$ such that γ_A(x) = ϕ if x ∉ A.

Here, γ_A is called fuzzy approximate function of the fs-set Γ_A, and the value $\gamma_A(x)$ is a set called x-element of the $fs-sets$ for all $x\in E$. Thus, an $fs-sets$ Γ_A over U can be represented by the set of ordered pairs

$$\Gamma_A=\left\{\left(x,\gamma_A(x)\right):x\in E,\gamma_A(x)\in F(U)\right\}$$

Where, $F(U)$ is the set of all fuzzy sets over U. The set of all $fs-sets$ over U will be represented by $FS(U)$.

Here we develop fuzzy soft aggregation $(fs-aggregation)$ operator that generates an aggregate fuzzy set from $fs-sets$ and its cardinal set. fs-aggregation operator is an operation by which several approximate functions of an fs-set are combined to produce a single fuzzy set which is the aggregate fuzzy set of fs-set. We can select the single crisp alternative from this set for getting better selection.

Definition 2.2

Let $\Gamma_A\in FS(U)$, $U=\{u_1,u_,\ldots\ldots,u_n\}$, $E=\{x_1,x_2,\ldots\ldots,x_m\}$ and $A\subseteq E$, then

Γ_A can be presented by the following table,

Γ_A	x_1	x_2	....	x_m
u_1	$\mu_{\gamma A(x_1)}(u_1)$	$\mu_{\gamma A(x_2)}(u_1)$	....	$\mu_{\gamma A(x_m)}(u_1)$
u_2	$\mu_{\gamma A(x_1)}(u_2)$	$\mu_{\gamma A(x_2)}(u_2)$	....	$\mu_{\gamma A(x_m)}(u_2)$
.				
.				
.u_n	$\mu_{\gamma A(x_1)}(u_n)$	$\mu_{\gamma A(x_2)}(u_n)$	....	$\mu_{\gamma A(x_m)}(u_n)$

Where $\mu_{\gamma A(x)}$ is the membership function of γ_A .

Definition2.3

Let $\Gamma_A \in FS(U)$. Then the cardinal set of Γ_A denoted by $c\Gamma_A$ and is defined by

$$c\Gamma A = \{\mu_c \Gamma_A(x)/x : x \in E\},$$

is a fuzzy set over E. The membership function $\mu_c \Gamma_A$ of $c\Gamma_A$ is given by

$$\mu_c \Gamma_A : E \to [0,1], \mu_c \Gamma_A(x) = \frac{|\gamma_A(x)|}{|U|}$$

Where $|U|$ is the cardinality of universal set U, and $|\gamma_A(x)|$ is the scalar cardinality of fuzzy set $\gamma_A(x)$. The set of all cardinal sets of the fs-sets over U will be denoted by $cFS(U)$, that is $cFS(U) \subseteq F(E)$.

Definition2.4

Let $\Gamma_A \in FS(U)$ and $c\Gamma_A \in cFS(U)$. Assume that $E = \{x_1, x_2, ...x_m\}$ and $A \subseteq E$, then $c\Gamma_A$ can be presented by the following table.

E	x_1	x_2	..........x_m
$\mu_c\Gamma_A$	$\mu_c\Gamma_A(x_1)$	$\mu_c\Gamma_A(x_2)$	$\mu_c\Gamma_A(x_m)$

Definition2.5

Let $\Gamma_A \in FS(U)$ and $c\Gamma_A \in cFS(U)$ Then fs-aggregation operator, denoted by FS_{agg}, is defined by

$$FS_{agg} : cFS(U) \times FS(U) \to F(U),\ FS_{agg}(c\Gamma_A, \Gamma_A) = \Gamma_A^*$$

Where $\Gamma_A^* = \{\mu\Gamma_A^*(u)/u : u \in U\}$ is a fuzzy set over U. Γ_A^* is called the aggregate fuzzy set of the fs – set Γ_A. The membership function $\mu\Gamma_A^*$ of Γ_A^* is defined as follows:

$$\mu\Gamma_A^* : U \to [0,1], \mu\Gamma_A^*(u) = \frac{1}{|E|} \sum_{x \in E} \mu_c \Gamma_A(x) \mu_{\gamma A(x)}(u)$$

Where $|E|$ represent the cardinality of E

Definition2.6

Let $\Gamma_A \in FS(U)$ and Γ_A^* be its aggregate fuzzy set. U={u_1, u_2........u_n}be the universal set, then the Γ_A^* can be represented by the following table.

Γ_A	$\mu\Gamma_A^*$
U_1	$\mu\Gamma_A^*(u_1)$
u_2	$\mu\Gamma_A^*(u_2)$
.	
.	
u_n	$\mu\Gamma_A^*(u_n)$

Theoream1

Let $\Gamma_A \in FS(U)$ and . If $M\Gamma_A, M_c\Gamma_A,$ and

Let $\Gamma_A \in FS(U)$ and $A \subseteq E$. If M_{Γ_A} , $M_{c\Gamma_A}$ and $M_{\Gamma_A^*}$ are representation matrices of $\Gamma_A, c\Gamma_A$ and Γ_A^*, respectively, then

$$|E| \times M_{\Gamma_A^*} = M_{\Gamma_A} \times M_{c\Gamma_A}^T$$

Where $M_{c\Gamma_A}^T$ is the transpose of $M_{c\Gamma_A}$ and $|E|$ represent the cardinality of E.

3. Proposed method

Before implementing fuzzy soft set algorithm, the original mammogram image has to be passed through a set of process called pre-processing, segmentation, feature extraction and feature selection. In our preprocessing, the first step is, removing the noise present in the image using Modified Robust outlyingness ratio with NL-means filter to get original like image. The next step is removing unwanted background objects like labels and artifacts present in the background of the image. Then a contrast enhancement process is used to sharpen the image features such as edges, boundaries or contrast and make the images more clear and useful for further analysis. In the next step for removing pectoral muscle, instead of applying different methods for left and right breast separately, a general method is designed for detecting and removing pectoral muscle in right MLO [16]. After preprocessing, the region of interest that contains micro calcification clusters are segmented from the background region using hierarchical fuzzy c means clustering incorporating with a feature vector containing four statistical features mean, standard deviation, skewness and Kurtosis extracted from the preprocessed image.. Then adaptive h-dome transformation and a threshold is used to segment the microclacification clusters from the segmented ROI.

Here we use Gray Level Co-occurrence Matrix (GLCM) to extract texture features from mammogram image for four spatial orientations horizontal, left diagonal, vertical and right diagonal corresponding to0^0, 45^0, 90^0, and 135^0 and five pixel distance (d = 1, 2, 3,4 and 5). TableI shows the accuracy obtained from the evaluation of co-occurance matrix extracted from four spatial orientations and five pixel distances 1, 2,3,4,5. Since 0^0 directions with 3 pixel distance give better result, 14 features are extracted in 0^0 directions with 3 pixel distance.

The extracted features are Contrast, Correlation, Energy, homogeneity, entropy, Max.prob, Dissimilarity, Inverse Difference Moment (IDM), Mean, Skewness, Shade, Variance, Autocorrelation and Difference entropy. Here the selected parameters are Contrast, Correlation, Energy, homogeneity, Max.prob, Dissimilarity, Mean, autocorrelation and entropy. Information Gain evaluation method with Ranker search is used to select 9 features out of 14 features.

Implement fuzzy soft set algorithm on the above extracted and selected features. A fuzzy π function is used to transform the crisp value of the attributes in to fuzzy values [17]. Feature selection is an important factor in the case of fuzzy soft set theory because unless the features selected are apt for a particular problem, it will not give accurate result.

Table 1. Pixel displacement, accuracy relation at different directions

Pixel Distance	Accuracy at 0^0	Accuracy at 45^0	Accuracy at 90^0	Accuracy at 135^0
One	78	74	61	60
Two	86	74	62	60
Three	94	81	66	63
Four	94	87	70	67
Five	94	91	70	75

4. Fuzzy soft Implementation algorithm

In this work, $U = \{m_1, m_2 ... m_n\}$ are the original mammogram images and $U' = \{pm_1, pm_2 ... pm_n\}$ is a set of processed images. That is $pm_1, pm_2 ... pm_n$ are microcalcification clusters obtained from segmented regions using h-dome transform. Each pm_i has undergone the process of feature extraction and then feature selection. The extracted features are stored in the set $E = \{e_1, e_2, ..., e_m\}$ and the selected

Features are in the set $E' = \{e'_1, e'_2, ..., e'_k\}$, $E' \subseteq E$.

Where n, m, k represents the number .of original images, number of extracted features and number of selected features respectively.

Step1: For all $e'_{i,j}$ construct the fuzzy value, $\mu_A(x) \in [0,1]$ using the fuzzy π function given as follows. $\forall i = 1,2,....,n, j = 1,2,....,k$

$$F(x; a, b, c) = \begin{Bmatrix} 0 & x \le a \\ 2[(x-a)/(c-a)]^2 & a \le x \le b \\ 2[(x-c)/(c-a)]^2 & b \le x \le c \\ 1 & x \ge c \end{Bmatrix} \quad (2)$$

$$\pi(x; p, q) = \begin{Bmatrix} F(x; (q-p), (q-p)/2, q) & if\ x \le q \\ 1 - F(x; q, (q+p)/2, (q+p)) & Otherwise \end{Bmatrix} \quad (3)$$

Where, x is the crisp value of the attribute; q is the cross over point which is calculated by the following equation

$$q=\frac{2\left[\left(\max\left(e_i^{'}\right)-\min\left(e_i^{'}\right)\right)+\left(mean\left(e_i^{'}\right)-median\left(e_i^{'}\right)\right)\right]}{sd\left(e_i^{'}\right)} \qquad (4)$$

p is the fuzzy band width defined as

$$p = max\{(q-K),(N-q)\} \qquad (5)$$

K is the mean value of the attribute $E_i^{'}$ and N is the maximum value of the attribute $E_i^{'}$. The π function in Eq. (3) is used to fuzzify the crisp value of the attributes.

Step 2: Construct fuzzy soft set Γ_A over $U^{'}$

Create a table for $\forall i = 1, 2, ..., n$, $j = 1, 2, ..., m$ specified as in definition 2.2. put fuzzy values for the selected features to the corresponding entries in this table and put the remaining entries as zero.

Step3: Find the cardinal set $c\Gamma_A$ of Γ_A

As specified in definition2.3 &2.4, find $c\Gamma_A$. It is calculated by dividing the scalar cardinality of fuzzy set by the cardinality of universe. Out of 14 features 8 features are selected and the cardinal for 8 attributes is calculated.

step4. Find the aggregate fuzzy set Γ_A^* of Γ_A.

Using the definition2.5, 2.6 and theorem1, find Γ_A^*.

Step5. Separate malignant images based on the following conditions on $\mu\Gamma_A^*$

If $\mu_A^* \geq 0.5$ then it is a malignant image otherwise it is normal.

Step 6. Stop

5. Experimental results and discussion

The fuzzy soft set algorithm is implemented on a combination of 100 normal, 100 malignant images from the publicly available MIAS dataset and validated with real data collected from RCC, Trivandrum. The selected 9 features are taken as the subset $E^{'}$ of E. Using Eq. (2)-Eq. (5), the crisp values of the parameters are converted to the corresponding fuzzy value. The following Results are based on a sample of 5 normal, and 5 malignant images from MIAS database. mdb077- mdb301 are normal images and mdb184 – mdb117 are malignant images. The combined membership grade of normal and malignant images is represented in Table.1 and its graphical representation is given in Figure 1

Table 1.Membershipgrade of fuzzy c means segmented images

Sr..no	Image	Membership Grade
1	mdb077	0.178
2	mdb216	0.193
3	mdb262	0.113
4	mdb300	0.181
5	mdb301	0.179
6	Mdb184	0.894
7	mdb158	0.801
8	mdb095	0.613
9	mdb102	0.69
10	mdb117	0.792

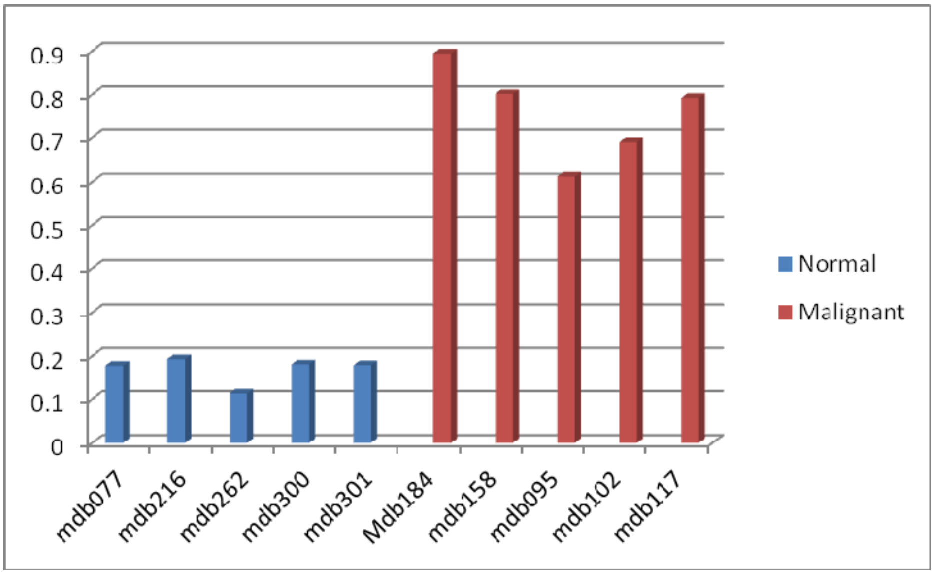

Fig 1. Graphical representation of membership grade

The ascending order of membership grade of malignant images shows the intensity of malignancy in each image and is shown in Table 2 and its graphical representation is shown in Fig.2. From the table, the intensity of malignancy is high in mdb184 and then mdb158 and so on. The time and memory used for the execution of this algorithm is given in Table3. The algorithm is implemented using jquery

Table 2. .Membershipgrade of malignant images

Sr..no	Image	Membership Grade
1	Mdb184	0.894
2	mdb158	0.801
3	mdb117	0.792
4	mdb102	0.69
5	mdb095	0.613

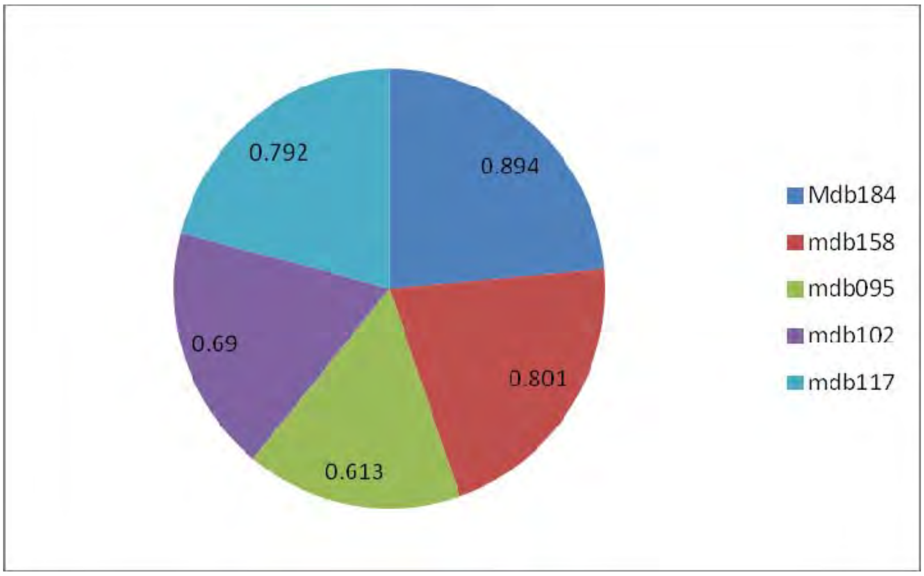

Figure2. Level of malignancy in the image

Table3. .execution time and memory usage of the algorithm

No of normal	No of malignant	memory usage (in bytes)	Time taken (in ms)
20	20	158	41.21
40	40	184	69.61
60	60	224	98.15
80	80	276	138.11
100	100	298	155.16

Table 4. Comparison of popular feature extraction and classification methods

Author	Method Used	Accuracy(%)
Rangayyan et al., 1997	**Region based edge profile acutance measure**	**92**
Alolfe et al., 2009	**Forward stepwise linear regression method with a combination of SVM and LDA classifier**	**90**
Proposed Method	**Gray level co-occurance Matrix with $fs-set$ t method**	**93.97**

6. Conclusion

The fuzzy soft sets theory ($fs-sets$) is an extension of soft set theory in which mapping is done from parameter to the crisp subset of universe. But in real word situations, where we have to deal with parameters with fuzzy characteristics, the soft set method is more complicated. But in fuzzy soft set the extension is done in such a manner that the fuzzy membership is used to describe parameter approximate elements of parameters.

From the above result, we can separate malignant and normal images from a set of images by specifying the condition that the membership grade of normal images lies in the range 0.000-0.299, and malignant lies above 0.500. Besides the malignancy detection, level of malignancy can also be identified by this method. This classification will help the radiologist to analyze a bulk of images and to concentrate only on malignant images by avoiding the normal images. The hierarchical fuzzy c means clustering incorporating with a vector contains four statistical features are utilized to get better segmentation result. The above algorithm is implemented on a combination of 100 normal, and 100 malignant images and the result shows 93.97% of accuracy on analyzing these images.

References

[1] Oliver, A; Freixenet, J; Marti,R; Pont, J; Perez, E; Denton, E R E : Zwiggelaar, R.(2008) : A novel breast tissue density classification methodology, IEEE Transactions on Information Technology in Biomedicine, 12(.1), pp. 55-65.

[2] Eddaoudi, F;.Regragui, F; K.Laraki. (2006): Characterization of normal mammograms based on the Statistical Features, Proc. Second International Symposium on Communications, Control and Signal Processing (ISCCSP).

[3] Spiesberger, W.(1979): Mammogram inspection by computer, IEEE Trans. Biomed. Eng., 26, pp. 213–219.

[4] Davies, D. H ; Dance, D. R. (1990) : Automatic computer detection of clustered calcifications in digital mammograms, Phys. Med., Biol., 35(8), pp. 1111–1118.

[5] Davies, D. H ; Dance, D. R ; Jones, C. H.(1990): Automatic detection of clusters of calcifications, SPIE Medical Imaging IV: Image Processing, 1233, pp. 185–191.

[6] Davies, D. H ; Dance, D. R ; Jones, C. H. (1989): Automatic detection of microcalcifications in digital mammograms using local area thresholding techniques,SPIE Medical Imaging III: Image Processing, 1092, pp. 153–157.

[7] Heng-Da Cheng; Yui Man Lui ; Rita I. Freimanis. (1998): A Novel Approach to Microcalcification Detection Using Fuzzy Logic Technique, IEEE Trans. Med. Imag., 17(3).

[8] Maji, P.K ;Roy, A. R; Biswas, R.(2002): An application of soft sets in a decision making problem, Comput. Math. Appl, 44, pp. 1077–1083.

[9] Guliato, D; Rangayyan, R.M; Carvalho, J.D; Santiago, S. A..(2006): Spiculation-Preserving polygonal Modeling of Contours of Breast Tumors", in Proceedings of the 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS), New York, pp. 2791-2794.

[10] Zadeh, L. A. (1965): Fuzzy Sets, Information and Control, 8, pp. 338-353.

[11] Gau W; Buehrer, D.J.(1993): Vague sets, IEEE Trans. System Man Cybernet, 23(2), pp. 610–614.

[12] Pawlak, Z.(1982): Rough sets, International Journal of Inform. Comput. Science, l(11), pp. 341–356.

[13] Mlolodtsov, D. A. (1999): Soft set theory-first results, Computers and and Mathematics with Applications, 37(4), pp. 19-31.

[14] Maji, P. K.; Biswas R; Roy, A. R. (2001): Fuzzy Soft Sets, Journal of Fuzzy Mathematics, 9(3), pp. 589-602.

[15] Cagman, N; .Enginoglu, S; Citak, F.(2011): Fuzzy soft set theory and applications, International journal of Fuzzy Systems, .8(.3), pp. 137-147.
[16] Sreedevi S; Elizabeth Sherly, (2015): A Novel Approach for Removal of pectoral muscles in Digital Mammogram, Proceedings of ICICT-2014, Elsevier Procedia Computer Science, 46, pp. 1724 – 1731.
[17] Pal, S. K; Majumder, D. K. D, (1986): Fuzzy Mathematical Approach to Pattern Recognition. New York: Wiley.