

How to cite this paper:

Saima Anwar Lashari, Rosziati Ibrahim, & Norhalina Senan. (2017). FUSSCYIER: Mammogram images classification based on similarity measure fuzzy soft set in Zulikha, J. & N. H. Zakaria (Eds.), Proceedings of the 6th International Conference of Computing & Informatics (pp 56-61). Sintok: School of Computing.

## FUSSCYIER: MAMMOGRAM IMAGES CLASSIFICATION BASED ON SIMILARITY MEASURE FUZZY SOFT SET

**Saima Anwar Lashari, Rosziati Ibrahim\*, and Norhalina Senan**

*Universiti Tun Hussein Onn Malaysia  
{saima, rosziati, halina }@uthm.edu.my*

**ABSTRACT.** Automatic digital mammograms reading become highly enviable, as the number of mammograms to be examined by physician increases enormously. It is premised that the computer aided diagnosis system is mandatory to assist physicians/radiologists to achieve high efficiency and productivity. To handle uncertainties of medical images, fuzzy soft set theory has been merely scrutinized, even though the choice of convenient parameterization makes fuzzy soft set suitable and feasible for decision making applications. Therefore, this study investigates the practicability of fuzzy soft set for classification of digital mammogram images to increase the classification accuracy while lower the classifier complexity. The proposed method FussCyier involves three phases namely: pre-processing, training and testing. Results of the research indicated that proposed method gives high classification performance with wavelet de-noise filter Sym8 with the accuracy 75.64%, recall 84.67% and CPU time 0.0026 seconds.

**Keywords:** mammogram images, computer aided diagnosis system, fuzzy soft set

### COMPUTER AIDED DIAGNOSIS SYSTEM: AN OVERVIEW

Computer aided diagnosis system (CAD) simply represents an important application with the ability to recognize image processing that can assist medical practitioners in enhancing diagnostic decisions (Sharma & Khanna, 2015). In general, CAD system comprises of a set of pattern recognition algorithms, primarily to assist radiologists in detecting potentially diseased lesions. The introduction of CAD as a diagnostic technology became imperative to rectify the problem; it also ensures preciseness in the interpretation of clinical images. The momentum of the CAD market was highly pronounced since 1998, when R2 Technology established the first CAD license from the U.S. FDA for industrial application named Image Checker (Tang *et al.*, 2009).

In the realm of breast cancer detection, digital mammography is a standard tool for the early detection of breast cancer and it is still widely used all over the world. The process is easy and has a few side effects (Lee & Chen, 2015). In general, it depends on the correct interpretation of mammograms by a radiologist. Because of the subtlety and variation of the breast, errors can be common. However, because of the limitations of the human visual system, it is complicated for radiologists to present equally precise and reliable evaluation of mammogram images (Al-Najdawi *et al.*, 2015). Thus, automatic digital mammograms reading become highly enviable and the computer aided diagnosis systems becomes a key requirement to assist the physicians/radiologists to attain high productivity and effectiveness (Otoom *et al.*,

2015). Therefore, the nature of mammogram images which inherit uncertainty leads to move towards soft set theory which can handle uncertainties that occurs in real world problems. Thus, automatic digital mammograms reading turn out to be extremely enviable, thus, computer aided diagnosis (CAD) systems are required to assist the physicians/radiologists in detecting subtle lesions and reducing the probability of the risk of failure in distinguishing abnormalities (Fenton *et al.*, 2013). In other words, CAD in screening mammographic images is considered as an immediate available opinion for radiologists in identifying high suspicious regions of malignancy (Howell *et al.*, 2014).

### SIMILARITY MEASURE FUZZY SOFT SET

Fuzzy set was initiated by Zadeh to permit elements to belong to a set in a gradual rather than an immediate way (Zadeh, 1965). Subsequently, growth of several data mining applications is based on this simple concept and nowadays, it is basically impossible to encounter any problem where applications and products are not based on fuzzy sets. Besides, to measure similarity among two objects is a basic phase for several data mining tasks for instance classification and clustering. Similarity measure enumerate the diverse patterns, signals, images or sets are alike at what extend (Handaga *et al.*, 2012). Baccour *et al.*, (2014) present properties of fuzzy similarities from the literature and discuss their validation to the common existing properties.

Similarity measure between fuzzy sets is plenteous, it is premised that numerical evaluations between fuzzy similarities measure (FSMs) are important to show experimental differences between them (Lashari *et al.*, 2015). Hence, studies on the similarity measure between fuzzy soft sets are scarce in the literature, despite the increasing volume of mammogram images classification. Lately, fuzzy set theory brings methods to handle uncertainty, such as Mushrif *et al.*, (2006) offered a method for natural textures classification based on soft set theory. All extracted features were real numbers. The proposed method successfully classified natural textures. The proposed algorithm had very low computational complexity when compared with Bayes classification technique. Handaga *et al.*, (2012) presented a method based on similarity measure between two fuzzy soft sets which deals with real numbers. The proposed method did modification in the classification phase and replace the classification function with similarity measuring function between two fuzzy soft sets to increased classification accuracy. Yet, the proposed method had high algorithm complexity (Lashari *et al.*, 2016). Below is the example to illustrate the proposed method, how it works with real numbers.

**Example:** Consider the following example, where  $U=\{x_1, x_2, x_3, x_4\}$  and  $E=\{e_1, e_2, e_3\}$ . Let two generalized fuzzy soft set over the parameterized universe  $(U,E)$

$$F_p = \left[ \begin{array}{ccc|c} 0.2 & 0.5 & 0.9 & 1.0 \\ 0.1 & 0.2 & 0.6 & 0.5 \\ 0.2 & 0.4 & 0.7 & 0.9 \end{array} \middle| \begin{array}{c} 0.6 \\ 0.8 \\ 0.4 \end{array} \right]$$

and

$$G_\delta = \left[ \begin{array}{ccc|c} 0.4 & 0.3 & 0.2 & 0.9 \\ 0.5 & 0.5 & 0.2 & 0.1 \\ 0.4 & 0.4 & 0.2 & 0.1 \end{array} \middle| \begin{array}{c} 0.5 \\ 0.7 \\ 0.9 \end{array} \right]$$

Here

$$m(\rho, \delta) = 1 - \sum |\rho_i - \delta_i|$$

$$m(\rho, \delta) = 1 - 0.1 + 0.1 + 0.5$$

$$m(\rho, \delta) = 0.3$$

$$S(F, G) = \frac{1}{n} \sum_{i=1}^n \left( 1 - \left| \mu_F^i - \mu_G^i \right| \right)$$

$$m_1(F, G) = \frac{1}{4} \left( (1 - |0.2 - 0.4|) + (1 - |0.5 - 0.3|) + (1 - |0.9 - 0.2|) + (1 - |1 - 0.9|) \right)$$

$$m_1(F, G) = \frac{1}{4} (.8 + .8 + .3 + .7 + .1)$$

$$m_1(F, G) = \frac{2.7}{4} = 0.675$$

$$m_2(F, G) = \frac{1}{4} \left( (1 - |0.1 - 0.5|) + (1 - |0.2 - 0.5|) + (1 - |0.6 - 0.2|) + (1 - |0.5 - 0.1|) \right)$$

$$m_2(F, G) = \frac{1}{4} (.6 + .7 + .6 + .1)$$

$$m_2(F, G) = \frac{2}{4} = 0.5$$

$$m_3(F, G) = \frac{1}{4} \left( (1 - |0.2 - 0.4|) + (1 - |0.4 - 0.4|) + (1 - |0.7 - 0.2|) + (1 - |0.9 - 0.1|) \right)$$

$$m_3(F, G) = \frac{1}{4} (.8 + 1 + .5 + .2)$$

c

$M_1(\underline{F}, \underline{G}) \cong 0.675$ ;  $M_2(\underline{F}, \underline{G}) \cong 0.5$ ;  $M_3(\underline{F}, \underline{G}) \cong 0.625$ . Thus  $\max [M_i(\underline{F}, \underline{G})] \cong 0.675$

Hence the similarity between the two GFSS  $F_\rho$  and  $G_\delta$  will be  $S(F_\rho, G_\delta) = M_i(\underline{F}, \underline{G}) \times m(\rho, \sigma) = 0.675 \times 0.3 = 0.20$  for universal fuzzy soft set where  $\rho = \sigma = 1$  and  $m(\rho, \sigma) = 1$ , then similarity  $S(F_\rho, G_\delta) = 0.675$ .

## PROPOSED METHOD

This section illustrates the proposed method FussCyier which consists of three phases, pre-processing, training and testing phase as shown in Figure 1. Each phase contains its different steps and delivers useful results to be used in the next phase. Dataset was obtained from the Mammographic Image Analysis Society (MIAS) (Suckling *et al.*, 1994). There are hundred and twelve images (63 benign images and 51 malignant images). The wavelet de-noising filter with hard and soft threshold functions have been applied for de-noising images to get better image quality. Later, six statistical features were extracted from region of interest (ROI) of the mammogram images (Lashari *et al.*, 2016). Afterwards, each dataset divided into two parts: 70% for training and 30% for testing and data were selected randomly for every experiment.

<i>Pre-Processing phase</i>	
1. De-noised images using wavelet hard and soft threshold functions	
2. Feature normalization to obtain a feature vector	
<i>Training phase</i>	
1. Given $N$ samples from the data $w$	
2. Calculate the cluster center vector $E_w$ $i = 1, 2, \dots, N$ using equation below	
$E_w = \frac{1}{N} \sum_{i=1}^N E_{wi} \quad (1)$	
3. Obtain Fuzzy soft set model $(F_w, E)$ , is a cluster center vector for each class $w$ having $D$ features	
4. Repeat the steps 2 and 3 for all $W$ classes	
<i>Testing phase</i>	
1. Obtain the unknown class data	
2. Attain a fuzzy soft set model for unidentified class data $(\tilde{G}, E)$ , compute similarity measure based on distance between $(\tilde{G}, E)$ and $(F_w, E)$ for each $w$	
$S(F, G) = \frac{1}{n} \sum_{i=1}^n \left( 1 - \left  \mu_F^i - \mu_G^i \right  \right) \quad (2)$	
3. Allocate the unknown data to class $w$ if distance measure is maximum	
$w = \arg \left[ \max_{w=1}^W S(G, F_w) \right] \quad (3)$	

**Figure 1: Proposed method for mammogram images classification**

Table 1 demonstrates wavelet de-noising filter Daub3 (Level 1) gives classification accuracy 75.64% (hard threshold), recall 84.67% with CPU time 0.0032 seconds whereas, wavelet de-noising Sym8 (Level 1) carried out accuracy 75.64% (soft threshold), recall 84.67% with CPU time 0.0026 seconds.

**Table 1: Performance Analysis of FussCyier**

Wavelet de-noising filters with decomposition levels		Accuracy (%)	Recall	CPU Time
Daub3 (Level 1)	Hard threshold	75.64	84.67	0.0032
	Soft threshold	74.17	86.67	0.0033
Daub3 (Level 4)	Hard threshold	65.61	77.33	0.0030
	Soft threshold	73.70	82.67	0.0027
Daub3 (Level 8)	Hard threshold	71.87	76.00	0.0028
	Soft threshold	74.08	82.00	0.0029
Sym8 (Level 1)	Hard threshold	75.64	84.67	0.0026
	Soft threshold	74.04	85.33	0.0032
Sym8 (Level 4)	Hard threshold	75.64	84.67	0.0026
	Soft threshold	74.19	84.00	0.0031
Sym8(Level 8)	Hard threshold	68.20	73.33	0.0028
	Soft threshold	70.49	80.00	0.0028

Results of the experimental setups indicate that soft threshold delivers better classification rate than hard threshold. Soft thresholding provides visually pleasing image and decreases the unforeseen sharp variations which arises in hard thresholding. Henceforth, soft thresholding is

preferred over hard thresholding (Lashari *et al.*, 2016). To appraise and validate the performance of FussCyier, with existing state of the art classifiers namely, neural network (NN) and Bayesian. FussCyier provides accuracy 75.64% (with de-noise filter) and accuracy 66.49% (without de-noise filter) which is comparatively better than other reported techniques such as NN where accuracy 56.3% (with de-noise filter) and accuracy 63.6% (without de-noise filter) and whereas Bayesian offered classification accuracy 57.5% (with de-noise filter) and classification accuracy 63.1% (without de-noise filter) (Naveed *et al.*, 2012).

## CONCLUSION

In the presented work, the problem of mammogram images classification has been thoroughly investigated. The concept of distance similarity measure fuzzy soft set theory for mammogram images FussCyier is introduced. Different experiments were carried out to evaluate the performance of the FussCyier, the acquired result illustrates that the FussCyier performs relatively better than existing classifiers, thus providing a rather new picture of mammogram images classification. So far, contemporary studies support and concur as a matter of fact that CAD technology has a great positive impact on early breast cancer detection and as well also improved the performance of radiologists to reduce variation within radiologists. The goal of this research was to show the feasibility of fuzzy soft set to classify mammogram images. This could be the first step towards developing a classification system for detection of tumour in mammogram images. However, many interesting scopes and topics are left behind. These could form the basis for researcher to study those scopes and topics further, which may pose unseen challenges.

## ACKNOWLEDGMENTS

This research is supported by Research, Innovation, Commercialization, Consultancy, Universiti Tun Hussein Onn Malaysia.

## REFERENCES

- Al-Najdawi, N., Biltawi, M., & Tedmori, S. (2015). Mammogram image visual enhancement, mass segmentation and classification. *Applied Soft Computing*, 35, 175-185.
- Baccour, L., Alimi, A. M., & John, R. I. (2014). Some Notes on Fuzzy Similarity Measures and Application to Classification of Shapes, Recognition of Arabic Sentences and Mosaic. *IAENG International Journal of Computer Science*, 41(2), 81-90.
- Begum, S. A., & Devi, O. M. (2011).: Fuzzy algorithms for pattern recognition in medical diagnosis. *Assam University Journal of Science and Technology*, 7(2), 1-12.
- Handaga, B., Herawan, T., & Deris, M. M.: FSSC: An Algorithm for Classifying Numerical Data Using Fuzzy Soft Set Theory. *International Journal of Fuzzy System Applications (IJFSA)*, 2(4), 29-46 (2012).
- Jaffar, M. A., Ahmed, B., Hussain, A., Naveed, N., Jabeen, F., & Mirza, A. M. (2009). Multi domain Features based Classification of Mammogram Images using SVM and MLP. In *Innovative Computing, Information and Control (ICICIC), 2009 Fourth International Conference on* (pp. 1301-1304). IEEE
- Lashari S. A. and Ibrahim R. (2015). Performance Comparison of Selected Classification Algorithms Based on Fuzzy Soft Set for Medical Data. In *Advanced Computer and Communication Engineering Technology* (pp. 813-820), Springer International Publishing
- Lashari, S. A., Ibrahim, R., Senan, N., Yanto, I. T. R., & Herawan, T. (2016). Application of Wavelet De-noising Filters in Mammogram Images Classification Using Fuzzy Soft Set. In *International Conference on Soft Computing and Data Mining* (pp. 529-537). Springer, Cham
- Lee, H., & Chen, Y. P. P. (2015). Image based computer aided diagnosis system for cancer detection. *Expert Systems with Applications*, 42(12), 5356-5365.
- Majumdar, P., & Samanta, S. K. (2010). Generalised fuzzy soft sets. *Computers & Mathematics with Applications*, 59(4), 1425-1432

- Molodtsov, D. (1999). Soft set theory—first results *Computer and mathematics with applications*. Volume 37, issue 4-5, pages19-31.
- Mushrif, M. M., Sengupta, S. & Ray, A. K.: (2006). Texture classification using a novel soft set theory based classification algorithm. In: *LNCS*, Volume 3851, pp.246-254, Springer, Heidelberg
- Naveed N., Hussain A., Jaffar, M. A. and Choi T. S. (2012). Quantum and impulse noise filtering from breast mammogram images. *Computer methods and programs in biomedicine*, 108(3), 1062-1069.
- Otoom, A. F., Abdallah, E. E., & Hammad, M. (2015). Breast Cancer Classification: Comparative Performance Analysis of Image Shape-Based Features and Microarray Gene Expression Data. *International Journal of Bio-Science & Bio-Technology*, 7(2).
- Sharma, S., & Khanna, P. (2015). Computer-aided diagnosis of malignant mammograms using Zernike moments and SVM. *Journal of digital imaging*, 28(1), 77-90
- Steimann, F. (2001).: On the use and usefulness of fuzzy sets in medical AI. *Artificial Intelligence in Medicine*, 21(1), 131-137
- Suckling, J., Parker, J., Dance, D., Astley, S., Hutt, I., Boggis, C., & Taylor, P. (1994). The mammographic image analysis society digital mammogram database. In *Excerpta Medica. International Congress Series* (Vol. 1069, pp. 375-378).
- Tang, J., Rangayyan, R. M., Xu, J., El Naqa, I., & Yang, Y. (2009). Computer-aided detection and diagnosis of breast cancer with mammography: recent advances. *Information Technology in Biomedicine, IEEE Transactions on*, 13(2), 236-251.
- Theodoridis, S., Pikrakis, A., Koutroumbas, K., & Cavouras, D. (2010). *Introduction to Pattern Recognition: A Matlab Approach: A Matlab Approach*. Academic Press.
- Zadeh, L. A. (1965). Fuzzy sets. *Information and control*, 8(3), 338-353.