

# Investigations on Impact of Feature Normalization Techniques on Classifier's Performance in Breast Tumor Classification

Bikesh Kumar Singh  
Assistant Professor  
Department of Biomedical  
Engineering  
N.I.T. Raipur

Kesari Verma  
Assistant Professor  
Department of Computer  
Applications  
N.I.T. Raipur

A. S. Thoke  
Professor  
Department of Electrical  
Engineering  
N.I.T. Raipur

## ABSTRACT

Feature extraction and feature normalization is an important preprocessing technique, usually employed before classification. Feature normalization is a useful step to restrict the values of all features within predetermined ranges. However, appropriate choice of normalization technique and normalization range is an important issue, since, applying normalization on the input could change the structure of data and thereby affecting the outcome of multivariate analysis and calibration used in data mining and pattern recognition problems. This paper investigates and evaluates some popular feature normalization techniques and studies their impact on performance of classifier with application to breast tumor classification using ultrasound images. For evaluating the feature normalization techniques, back-propagation artificial neural network [BPANN] and support vector machine [SVM] classifier models are used. Results show that that normalization of features has significant effect on the classification accuracy.

## General Terms

Pattern Recognition, Medical Image Processing.

## Keywords

Feature extraction, feature normalization, classifier's performance, breast tumor classification.

## 1. INTRODUCTION

Data normalization is a preprocessing technique usually employed before feature selection and classification. Complex real time pattern recognition systems use features that are generated by many different feature extraction algorithms with different kinds of sources. These features may have different dynamic ranges. Popular distance measures, for example the Euclidean distance, implicitly assign more weighting to features with large ranges than those with small ranges. Feature normalization is thus required to approximately equalize ranges of the features and make them have approximately the same effect in the computation of similarity [1]. Further in complex real time pattern recognition problems with huge numbers of features with large dynamic ranges, normalization may improve the performance of classifier model. However, appropriate choice of normalization technique and normalization range is an important issue, since, applying normalization on the input could change the structure of data and thereby affecting the outcome of multivariate analysis and calibration used in data mining and pattern recognition problems. Some comparative studies of various normalization techniques for different applications are reported in [2-6]. There is no universally

defined rule for normalizing datasets and thus the choice of a particular normalization rule is largely left to the discretion of the user [4]. In this study, the effectiveness of four popular normalization techniques in classifying breast tumors using ultrasound images is investigated using two different classifier models based on back-propagation artificial neural network [BPANN] and support vector machine [SVM]. Experiments were performed on MATLAB® software platform. The rest of the paper is organized as follows. Section 2 presents overview of popular normalization techniques investigated in this study. In section 3, methodology adopted is explained followed by results and discussions in section 4. Finally, conclusion and future scopes are presented in section 5.

## 2. BRIEF OVERVIEW OF FEATURE NORMALIZATION TECHNIQUES

Let  $x_i$  and  $\hat{x}_i$  denote the original and normalized feature value respectively and further it is assumed that the specific feature contain 'N' values. Various normalization techniques investigated in this study are:

### 2.1 Z – Score Normalization

A very common technique to normalize the features to zero mean and unit variance is Z- score normalization [6]. It is a linear technique in which, initially, mean ( $\bar{x}$ ) and standard deviation ( $\sigma$ ) of the specific feature values are computed using:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad (1)$$

and, 
$$\sigma^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2 \quad (2)$$

The normalized feature is then given by:

$$\hat{x}_i = \frac{x_i - \bar{x}}{\sigma} \quad (3)$$

### 2.2 Min – Max Normalization

This technique performs a linear transformation on the original data. For mapping a value, of an attribute  $x_i$  from range  $[\min(x_i), \max(x_i)]$  to a new range  $[\min x_{new}, \max x_{new}]$ , the normalized feature is given by:

$$\hat{x}_i = \frac{x_i - \min(x_i)}{\max(x_i) - \min(x_i)} (\max x_{new} - \min x_{new}) + \min x_{new} \quad (4)$$

The advantage of Min-Max normalization is that it preserves the relationships among the original data values [7]. In this study  $\max x_{new} = 1$  and  $\min x_{new} = -1$  is used.

### 2.3 Linear Scaling to Unit Range

This is also a linear transformation technique to normalize data in range [0, 1]. Given a lower bound  $\min(x_i)$  and upper bound  $\max(x_i)$  of an attribute  $x_i$ , the normalized value is given by:

$$\hat{x}_i = \frac{x_i - \min(x_i)}{\max(x_i) - \min(x_i)} \quad (5)$$

Linear scaling to unit range is special case of min-max normalization in which  $\max x_{new} = 1$  and  $\min x_{new} = 0$ .

### 2.4 Softmax Scaling

In addition to linear scaling, non linear normalization techniques may be used in cases where data are not evenly distributed around the mean [8]. In such cases the transformations based on nonlinear (i.e., exponential, logarithmic, sigmoid etc.) functions can be used to map the data within specified intervals. One such popular technique is so called *softmax* scaling which squashes the data values nonlinearly in the interval [0, 1]. The normalized feature is given by:

$$\hat{x}_i = \frac{1}{1 + e^{-y}} \quad (6)$$

where  $y = \frac{x_i - \bar{x}}{r\sigma}$  and  $r$  is a user defined parameter. It is noted

in the above equation that for small values of  $y$  i.e., for values of  $x_i$  closer to mean,  $y$  is an approximately linear function. Values away from mean are squashed exponentially.

## 3. MATERIALS AND METHODS

### 3.1 Application of Feature Normalization

In this study, the application considered for evaluating the performance of feature normalization techniques is breast tumor classification using ultrasound images. Breast Cancer is the second leading cause for death of women due to cancer. In past few years, it is one of the major health issues, as its incidence is increased in recent years. Since, the causes of breast cancer are still unknown, its early detection plays key role in reducing death rates. Apart from early detection, accurate and reliable diagnosis in distinguishing benign and malignant tumor is extremely important to save human life. Computer aided diagnosis (CAD) has been evolved as a clinically proven tool to assist the physicians for detection and diagnosis of breast cancer. Feature extraction, feature normalization and classification are important step in CAD systems.

### 3.2 Breast Ultrasound Dataset

The dataset consisting of 178 cases, of which 88 benign and 90 malignant cases are use in this study. The images obtained

were first converted in to gray scale images possessing 256 gray levels, frame count equal to one and 90 dpi horizontal and vertical resolution. Since some of the images acquired were of variable size, hence all the images were resized to  $400 \times 400$  (0.16 Mega pixels) before filtering.

### 3.3 Preprocessing and Feature Extraction

Breast ultrasound images suffer from intrinsic artifact called speckle resulting in low resolution, poor contrast and blurry edges. Diagnostic features such as cysts and masses may be small and ill-defined with respect to surrounding tissues due to presence of speckle noise. A lot of despeckle methods have been developed by researchers. An extensive comparative study of state-of-art speckle reduction techniques is carried out in [9]. On the basis of results in [9], we employ a wavelet based despeckle filter due to its effectiveness in providing smoothening while preserving edges, boundaries and other sharp details. After speckle reduction, region of interest (ROI) was extracted manually under the guidance of expert Radiologist. This is illustrated in figure 1. After extracting ROI, total of 457 texture and shape features were extracted from breast ultrasound images. The extracted features are presented in table 1. The details of these features can be found in references mentioned therewith.

### 3.4 Feature Normalization

The features extracted were normalized using four normalization techniques discussed in section 2. To analyze the effectiveness of four normalization techniques i.e. z-score, min-max, linear scaling and softmax, the normalized features were applied as input to two different classifier models based on BPANN and SVM. This is illustrated in figure 2. The BPANN and SVM are the most popular and widely used classifier in pattern recognition and classification. In BPANN model, adaptive gradient descent learning algorithm was used due to its low storage requirements and assured convergence. The learning rate and performance goal were fixed at 0.01 and 0 respectively while mean square error (MSE) was used as performance function. On the other hand, in SVM model linear kernel function was used in experiments. Further for both the classifier models 67% samples were used for training phase and 33% samples were used for testing phase. The classification accuracy of each classifier model was recorded. The classification accuracy is a useful performance measure used to evaluate the overall efficiency of classifier model. It is defined as:

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP} \times 100 \quad (7)$$

where TP, TN, FP, FN are true positive rate, true negative rate, false positive rate and false negative rate respectively. Accuracy is total percentage of test samples correctly classified by the classifier model. TP is the total number of correctly classified malignant samples, TN is the total number of correctly classified benign samples, FP is total number of benign samples wrongly classified as malignant samples and FN is total number of malignant samples wrongly classified as benign samples. In order to calculate TP, TN, FP and FN, confusion matrix is usually employed. All the experiments were performed in MATLAB®, version R2012a, software platform.

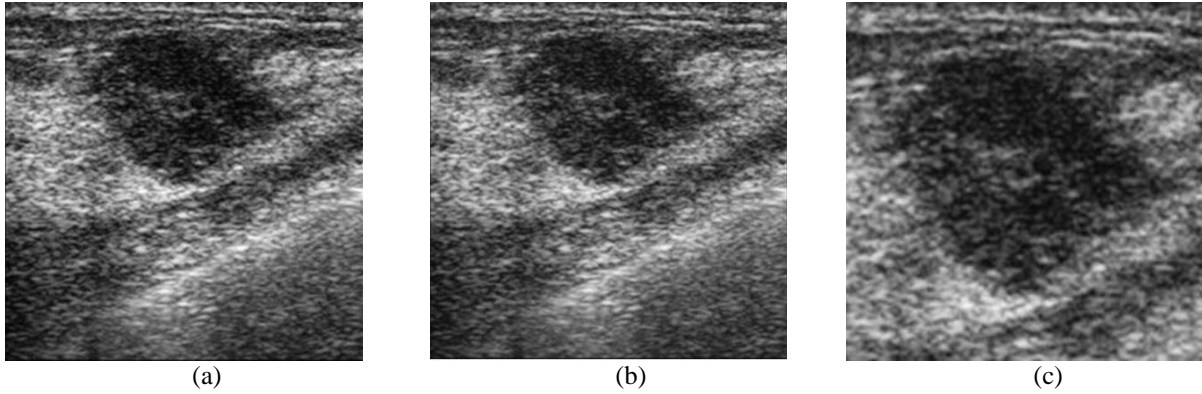


Fig. 1: (a) Original noisy image (b) Wavelet filtered image (c) Cropped ROI

Table 1. Summary of texture and shape features used in classification of breast tumor [10-23]

Category of feature	Number of features	Name of features
First order statistics	8	mean ( $m$ ), variance ( $\mu_2$ ), standard deviation ( $\sigma$ ), skewness ( $\mu_3$ ), kurtosis ( $\mu_4$ ), and measure of smoothness ( $R$ ), measure of uniformity ( $U$ ), and entropy ( $e$ ) (F[1] to F[8])
Haralick texture features	26	Mean and range of following features were calculated: Angular second moment, contrast, correlation, sum of squares, inverse difference moment, sum average, sum variance, sum entropy, entropy, difference variance, difference entropy, information measures of correlation-1 and information measures of correlation-2 (F[9] to F[34])
Gray level difference statistics	4	Contrast, angular second moment, entropy, and mean (F[35] to F[38])
Neighbourhood gray tone difference features	5	Coarseness, contrast, busyness, complexity, strength (F[39] to F[43])
Statistical feature matrix	4	Coarseness, contrast, periodicity, roughness (F[44] to F[47])
Laws texture features	6	LL, EE, SS, LE, ES, LS (F[48] to F[53])
Fractal based texture features	4	Hurst coefficient $H(k)$ for $k=1,2,3,4$ . (F[54] to F[57])
Spectral texture features	379	199 radial features ( $S(r)$ ) (F[58] to F[256]) and 180 angular features ( $S(\theta)$ ) (F[257] to F[436])
Run length texture	11	SRE, LRE, GLNU, RLNU, RPC, LGRE, HGRE, SRLGE, SRHGE, LRHGE, LRLGE (F[437] to F[447])
Regional features	3	Area, perimeter and perimeter <sup>2</sup> /area (F[448] to F[450])
Moment invariants	7	$\phi_7$ to $\phi_7$ (F[451] to F[457])

#### 4. RESULTS AND DISCUSSIONS

This section presents results of comparative analysis of various data normalization techniques discussed in section 2 using the proposed approach. Table 2 shows the results of the two classifier models obtained with original and normalized features. It is observed that normalization of data has significant effect on the classification accuracy. Normalization may have positive as well as negative impact on classifier performance. The performance of normalization techniques is also illustrated in figure 3. The classification accuracy of BPANN and SVM are found to be 80.35% and 76.89% respectively when original features are used. It is found that, for BPANN as well as SVM, the classification accuracy obtained with linearly scaled normalized data is

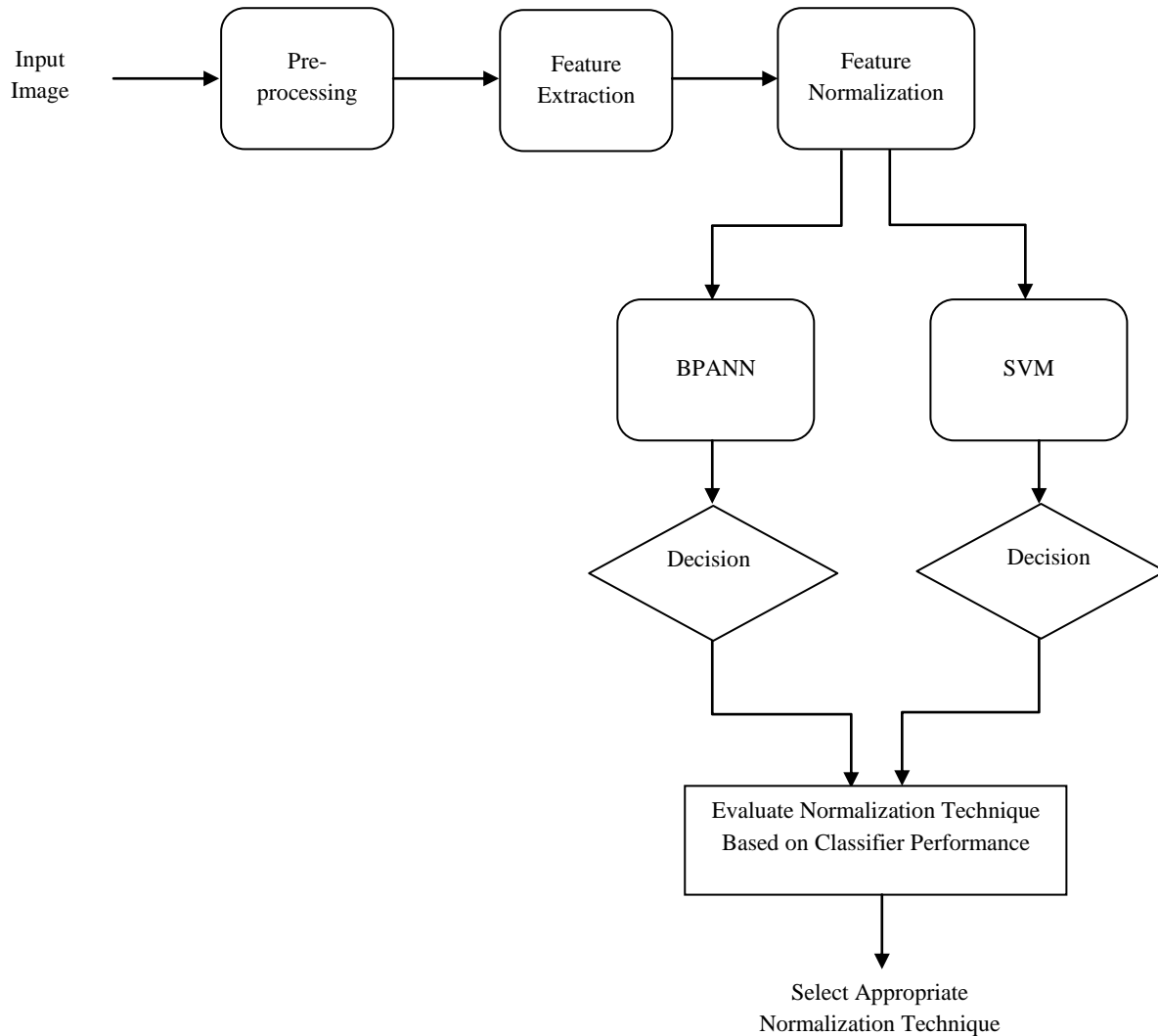
highest i.e. 80.35% and 77.89% respectively. The performance of softmax normalization is close to that obtained by original data. The classification accuracy obtained with z-score and min-max normalized data is even less than that of original data. The z-score normalization achieves smallest classification accuracy of 55.2% using BPANN. This shows that normalization may have negative impact on classifiers performance. Based on the results of this study it is found that normalization using linear scaling outperformed other normalization techniques followed by softmax scaling. Furthermore, BPNN outperform SVM for most of the normalization techniques. Thus, linearly scaled normalized features would be best choice for application considered in

this paper. It is also found that BPANN outperform SVM for most of the normalization techniques.

## 5. CONCLUSIONS

Appropriate choice of normalization technique for equalizing the dynamic range of extracted features is of important concern in pattern recognition and computer aided diagnosis. In this paper, four popular data normalization techniques were evaluated using BPANN and SVM based classifier models. The effect of normalizing features on classifier's performance for classifying malignant and benign tumors using breast ultrasound images is studied. In our application, linearly scaled features outperformed other

normalization methods followed by softmax normalization. We conclude that, normalization method should be chosen carefully and on the basis of experiments, so as to avoid its negative influence on classifiers performance. Further it cannot be claimed that linear scaling normalization will be useful for every application. Researchers should select appropriate feature normalization technique based on their own discretion and quantifiable results.



**Fig. 2: Proposed approach for selection of appropriate normalization technique based on classifier performance**

**Table 2. Classification accuracy of classifier models with original and normalized data set**

Classifier model	Original data	Normalization technique			
		Z-score	Min-max	Linear scaling	Softmax
BPANN	80.35%	55.2%	72.4%	80.35%	80%
SVM	76.89%	75.9%	73.1%	77.89%	76.8%

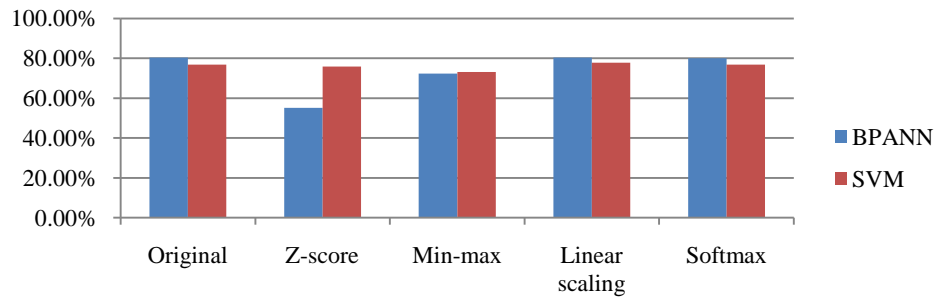


Fig. 3: Comparison of different normalization techniques based on classifier performance

## 6. REFERENCES

- [1] Salama, M. A., Hassanien A. E. and Fahmy A. A. 2010. Reducing the influence of normalization on data classification. In Proceedings of International Conference on Computer Information Systems and Industrial Management Applications, 609 – 613.
- [2] Aksoy S. and Haralick R. Feature normalization and likelihood-based similarity measures for image retrieval. Pattern Recognit. Lett., Special Issue on Image and Video Retrieval, 2000.
- [3] Manikandan G., Sairam N., Sharmili S. and Venkatakrishnan S. 2003. Achieving privacy in data mining using normalization. Indian Journal of Science and Technology, 6, 4268-4272.
- [4] Vaishali R. Patel and Rupa G. Mehta. 2011. Impact of outlier removal and normalization approach in modified k-means clustering algorithm. International Journal of Computer Science Issues, 8, 331-336.
- [5] Saranya C. and Manikandan G. 2013. A study on normalization techniques for privacy preserving data mining. International Journal of Engineering and Technology, 5, 2701-2704.
- [6] Jayalakshmi T. and Santhakumaran A. 2011. Statistical normalization and back propagation for classification. International Journal of Computer Theory and Engineering, 3, 89-93.
- [7] Manikandan, G., Sairam, N., Sharmili S. and Venkatakrishnan S. 2013. Achieving privacy in data mining using normalization, Indian Journal of Science and Technology, 6, 4268-4272.
- [8] Theodoridis, S. and Koutroumbas, K. 2009. Pattern Recognition, 4th edition, Elsevier.
- [9] Singh B. K., Verma K. and Thoke A. S. 2015. Objective and Optical Evaluation of Despeckle Filters in Breast Ultrasound Images, IETE Technical Review, 15 pages.
- [10] Haralick R. M., Shanmugam K. and Dinstein I. 1973 Textural features for image classification, IEEE Transactions on Syst., Man & Cybernetics, SMC-6, 610-621.
- [11] Srinivasan, G. N. and Shobha G. 2008. Statistical texture analysis, Proceedings of World Academy of Science, Engineering and Technology, 36, 1264-1269.
- [12] Gonzalez, R. C. and Woods R. E. 2010. Digital Image Processing Using MATLAB, 2nd edition, Pearson Prentice Hall.
- [13] Haralick R. M. 1979. Statistical and structural approaches to texture, Proc. IEEE, 67, 786-804.
- [14] Weszka, J. S., Dyer, C. R. and Rosenfeld, A. 1976. A comparative study of texture measures for terrain classification, IEEE Transactions on Systems, Man and Cybernetics, SMC-6, 269-285.
- [15] Wu, C. M., Chen, Y. C. and Hsieh, K. S. 1992. Texture features for classification of ultrasonic liver images, IEEE Transactions on Medical Imaging, 11, 141-152.
- [16] Amadasun, M. and King R. 1989. Texture features corresponding to textural properties, IEEE Transactions on Systems, Man and Cybernetics, 19, 1264-1274.
- [17] Stoitsis, J., Golemati, S. and Nikita, K. S. 2006. A modular software system to assist interpretation of medical images—Application to vascular ultrasound images, IEEE Transactions on Instrumentation and Measurement, 55, 1944-1952.
- [18] Wu C. M. and Chen Y. C. 1992. Statistical feature matrix for texture analysis, CVGIP: Graphical Models and Image Processing, 54, 407-419.
- [19] Laws, K. I. 1980. Rapid texture identification, Image Processing for Missile Guidance, 238, 376-380.
- [20] Stromberg W. D. and Farr T. G. 1986. A Fourier based textural feature extraction procedure, IEEE Transactions on Geoscience and Remote Sensing, GE-24.
- [21] Singh B. K. 2011. Mammographic image enhancement, classification and retrieval using color, statistical and spectral Analysis, International Journal of Computer Applications, 10, 18-23.
- [22] Marshkole, N., Singh B. K. and Thoke, A. S. 2011. Texture and Shape based Classification of Brain Tumors using Linear Vector Quantization, International Journal of Computer Applications 30(11), 21-23.
- [23] Singh, B. K., Yadav, A. and Singh, S. 2011. ANN based Classifier System for Digital Mammographic Images International Journal of Computer Applications, 35(13), 39-42.