A mammographic images classification technique via the Gaussian Radial Basis Kernel ELM and KPCA

[1]Bacha Sawssen bachasawcen@gmail.com
[2]Taouali Okba taoualiok@gmail.com
[1]Liouane Noureeddine Noureddine.liouane@enim.rnu.tn

[1] Laboratory of Automatic, Signal and Image Processing, National Engineering School of Monastir,University of Monastir Tunisia
[2]Department of Computer Engineering, Faculty of Computers and Information Technology, University of Tabuk, Tabuk, Saudi Arabia

Abstract: Computer Assisted Diagnosis (CAD) and Artificial Intelligence (AI) are hot topics in the field of medical imaging. Recently, many methods have been proposed. In this research work, a novel mammography image classifier based on Kernel Extreme Learning Machine (KELM) and Kernel Principal Component Analysis (KPCA) is presented. The proposed algorithm, called KELM-KPCA, aims to detect breast tumors by classifying mammographic images. The system first used discrete Tchebichef transform (DTT) to extract features from the images. After normalization of the feature vectors, the kernel principal component analysis (KPCA) is applied to reduce the dimensionality of the features. The reduced characteristics were then subjected to classification by KELM. The k-factor cross-validation strategy was used to improve the generalization of the proposed algorithm. The Mammographic Image Analysis Society (MIAS) dataset is used in this work. The simulation results were compared to the existing algorithms and it was observed that the proposed work outperforms other algorithms. Work on the same dataset in terms of accuracy, F-score, sensitivity and specificity.

***Keywords*— Mammography, CAD, classification, machine learning, KPCA**

I. INTRODUCTION

Cancer is a problem of extreme significance with social and financial implications to the public health. Different kinds of cancers have been already reported in literature and can be classified by the type of cells that are initially affected. Breast cancer is one of the most common cancer along with lung and bronchus cancer, prostate cancer, colon cancer, and pancreatic cancer among others [1].

Early detection of breast cancer can be achieved by mammography techniques which allow the visualization of tissue structure of the breast. Computer assisted diagnosis (CAD) of mammograms attempts to help radiologists providing an automatic procedure to detect possible cancers in mammograms [2-11].

The utilization of data science and machine learning approaches in medical fields proves to be prolific as such approaches may be considered of great assistance in the decision making process of medical practitioners. With an unfortunate increasing trend of breast cancer cases, comes also a big deal of data which is of significant use in furthering clinical and medical research, and much more to the application of data science and machine learning in the aforementioned domain.

The presented work aims at developing a CAD model which can classify the mammograms as normal or abnormal, and further, benign or malignant accurately. In the present model, a contrast limited adaptive histogram equalization (CLAHE) is used for image preprocessing and the discrete Tchebichef transform (DTT) to extract the characteristics of the ROIs. In order to reduce the size of the feature vector and simplify the classification, a core PCA (KPCA) is used, which reduces the number of redendants features. Extensive experiments are performed on one widely used mammogram dataset, namely, Mammographic Image Analysis Society (MIAS) [12] to justify the usefulness of the proposed work. Further, the performance of the proposed scheme is compared with that of the state-of-the-art approaches.

The structure of this paper is as follows. Section 2 deals with the proposed scheme, where extraction of features, dimensionality reduction is discussed in detail. Classification algorithms used are presented in section 3. We illustrate the results of our work in section 4. Finally, section 5 presents a conclusion of this study as well as a general discussion of the results obtained.

II. METHODOLOGY

A. Dataset used

The MIAS database of Mammographic Image Analysis Society, London, UK, is used in the present work. The database consists of 322 mammogram images of left and right breasts with mediolateral oblique (MLO) view.

All the images are in gray-scale format having the dimension of 1024×1024 pixels. Out of 322 images, 207 images are of normal (N) category and 115 images are of abnormal (A) category.

TABLE 1. NUMBER OF TRAINING AND TESTING IMAGES

	Total number of images		
	Normal	Abnormal	
		Malignant	Benign
Training set (80%)	165	42	49
Testing set (20%)	42	11	13
Total	**207**	**53**	**62**

The abnormal category is further divided into six classes: calcification,well-defined / circumscribed masses, spiculated masses, architectural distortion, asymmetry and other ill-defined masses. In the present work, all the six types of abnormalities are considered as an abnormal category for classification problem.

B. Dataset Preprocessing

Pre-processing is considered to be a vital step before performing any further modules in a CAD system.

The fundamental improvement needed in mammography is an increase in contrast. The contrast between malignant tissue and normal dense tissue may be present on a mammogram, but below the threshold of human perception. Several methods for improving image contrast have been proposed in the literature.

Hence, in the present work, initially, we apply, manually, a cropping operation to the original mammograms which contains background, muscle, and the label (fig 1). This information can be seen as noise in the process of classification. The desired ROIs have a size of 127 * 127 pixels.

In case of abnormal (malignant or benign) mammograms, the cropping is done using the given ground truth information about the position and radius of the abnormal regions. However, for normal mammograms, cropping is done on any arbitrary location to get the Region of Interest (ROI).

Pre-processing of the ROIs is considered to be a vital step before performing any further modules in a CAD system. As some of the images collected from the datasets are of low contrast, so it is required to enhance the contrast of such images. Hence, in the present work, contrast limited adaptive histogram equalization (CLAHE) [13] is utilized to improve the quality of the low-contrast images.

CLAHE has produced good results on medical images. This method is formulated based on dividing the image to several non-overlapping regions of almost equal sizes. This method applies histogram equalization to a contextual region. Each pixel of original image is in the center of the contextual region.

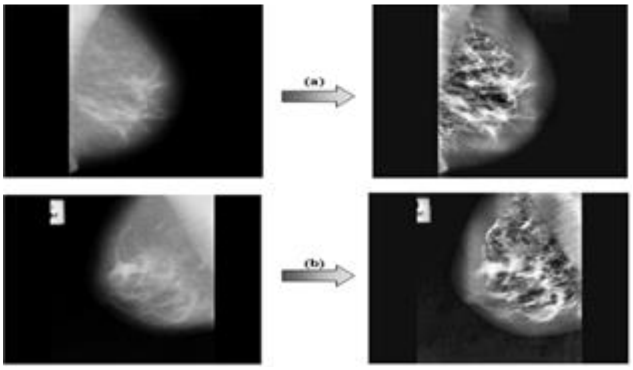

Fig.1. Enhancement results for mammogram image, First row: normal (a), malignant(with cancer) (b). Second row: Results for the corresponding enhanced Images

The original histogram is clipped and the clipped pixels are redistributed to each gray level. The new histogram is different with ordinary histogram, because each pixel intensity is limited to a user-selectable maximum. So CLAHE can limit the noise.

C. Features extraction and dimensionality reduction

- **Discrete Tchebichef Transform (DTT)**

Image moments have been exploited largely in several computer vision and related applications, such as pattern recognition. Moments are the scalar values utilized to represent a function (image), indicating important properties of the image. The moments are either categorized as non-orthogonal or orthogonal based on the nature of the corresponding basis functions. Zernike and Legendre moments use continuous orthogonal polynomials as basic functions. Since the Zernike and Legendre polynomials are defined only inside the unit circle, the computation of those moments requires a coordinate transformation and suitable approximation of the continuous moment integrals. Discrete orthogonal moments such as the tchebichef moments [14] are directly defined in the image coordinate space and preserve the property of orthogonality in a moment set. Discrete tchebichef moments (DTM) are thus expected to perform better than continuous moments, particularly in applications requiring independent shape characteristics. DTM are determined by projecting the input image on to a set of tchebichef polynomials.

- **Kernel Principal Component Analysis (KPCA)**

Kernel principal component analysis (KPCA) is an extension of principal component analysis [15]. It is normally applied for dimensionality reduction of features.

The classical PCA method performs well only on linear processes, a nonlinear PCA technique called kernel PCA (KPCA), has been developed by [16] and widely used to model various nonlinear processes [17,18]. KPCA maps the original inputs into a high dimensional feature space using a kernel method. In this work, we used KPCA [19] method that reduced the size of the image.

III. CLASSIFICATION ALGORITHMS

A. ELM

Extreme learning machine (ELM) is a type of single hidden layer feed-forward network (SLFN) proposed by **[**20**]**. In ELM, the used structure is a typical SLFN structure which consists of an input layer, a hidden layer, and an output layer such as: For training samples $X = \{x_i, y_i\}_{i=1,\ldots N}$ where $x_i = [x_{i1}, x_{i2}, \ldots, x_{in}]^T \in R^n$, $y_i = [y_{i1}, y_{i2}, \ldots, y_{im}]^T \in R^m\}_{i=1,\ldots,N}$ where N is the number of instances, d is the dimension, m is the number of output nodes. For regression problems m = 1, while for classification problems m is the number of categories, classes or labels. The output function of ELM for SLFNs with L hidden layer nodes is given by:

$$f(x_i) = y_i = \sum_{j=1}^{k} \beta_j \, h(w_j . x_i + b_j) \qquad i = 1, \ldots, N \quad (1)$$

Where $w_j = [w_{j1}, w_{j2}, \ldots, w_{jn}]$ is the weight vector connecting the input neurons to the j[th] hidden neuron, $\beta_j = [\beta_{j1}, \beta_{j2}, \ldots, \beta_{jm}]^T$ is the weight vector connecting the output neuron to the j[th] hidden neuron, k is the number of hidden neurons in the hidden layer, b_j is the bias of the j[th] hidden neuron, $w_j . x_i$ indicates the inner product of w_j and x_i, and h is a sigmoid function.

Equation (1) can be written in a matrix format as:

$H\beta = Y$ Where H is the hidden layer output matrix of the neural network:

$$\boldsymbol{H} = \begin{bmatrix} h(w_1 . x_1 + b_1) & \cdots & h(w_k . x_1 + b_k) \\ \vdots & \ddots & \vdots \\ h(w_1 . x_N + b_1) & \cdots & h(w_k . x_N + b_k) \end{bmatrix}_{\boldsymbol{N \times k}} \qquad (2)$$

With $\beta = \begin{pmatrix} \beta_1^T \\ \vdots \\ \beta_k^T \end{pmatrix}_{k \times m}$ $\beta = H^+ Y$ and $Y = \begin{pmatrix} y_1^T \\ \vdots \\ y_k^T \end{pmatrix}_{N \times m}$

Where H^+ represents the Moore-Penrose pseudo-inverse [21] of H such as: $H^+ = (H^T . H)^{-1} . H^T$

So the output of ELM is:

$$f(x) = h(x)\beta = h(x)\left(\frac{I}{C} + H^T H\right)^{-1} H^T Y \qquad (3)$$

B. Kernel Extreme Learning Machine (KELM)

As for ELM with kernels, it obtains a better regression and classification accuracy by introducing kernel. As proposed in [22**]**, if $h(x)$ is unknown, i.e., an implicit function, one can apply the Mercer's conditions on ELM, and define a kernel matrix for ELM that takes the form:

$$K_{ELM} = HH^T\colon\ K_{ELM}(i,j) = h(x_i).h(x_j) = K(x_i, x_j) \quad (4)$$

In KELM $H = [h(x_1)^T \ \ldots \ h(x_N)^T]^T$ represents hidden layer output matrix which maps data x_i from the input space to the hidden layer feature space and it is irrelevant to target value y_i and number of output nodes m. The kernel matrix $K_{ELM} = HH^T$ is related only to input data x_i and number of training samples N, for regression, binary classification and multi class classification. Then, the output function of ELM classifier (Eq 3) can be written compactly as:

$$f(x) = \left(\begin{bmatrix} K(x, x_1) \\ \vdots \\ K(x, x_N) \end{bmatrix}^{T}_{2} \left(\frac{I}{C} + \mathrm{K}_{ELM}\right)^{-1} Y \right) \qquad (5)$$

TABLE 2 TWO COMMON KERNELS WITH THEIR FORMULA AND PARAMETERS

Name	Formula and parameters
Radial Basis Function Kernel	$K_\sigma(x,y) = exp\left(-\frac{1}{2}\frac{\|x-y\|^2}{\sigma^2}\right)$ σ: parameter of Rbf kernel
Wavelet Kernel	$K_w(x,y) = cos\left(w_1 \frac{\|x-y\|}{w_2}\right) exp\left(-\frac{\|x-y\|^2}{w_3}\right)$ w_1, w_2, w_3 **:** Parameters of wavelet kernel

IV. EXPEREMENTAL RESUTS AND ANALYSIS

A. Statistical analysis

In order to show the effectiveness of our approach, objective evaluation criteria are measured: sensitivity, specificity and accuracy. They are described below. The definition of TP, FN, FP and TN are illustrated in Table 3.

TABLE 3: DEFINITION OF TP, FN, FP AND TN

	Total population	Predicted class	
		Abnormal	Normal
True class	Abnormal	TP	FN
	Normal	FP	TN

Where TP and TN respectively mean True Positive and True Negative, which determine the number of abnormal (cancerous) cases classified as such (TP) and the number of normal (non-cancerous) cases classified as such (TN). FP and FN respectively denote False Positive and False Negative, which represent the classification of abnormal cases in normal cases (FN) and normal cases in abnormal cases (FP). The specificity (Sp) of a classifier measures the ability to correctly detect normal cases, while the definition of the sensitivity (Se) is given as the rate of True positive instances and False Negative instances that have been classified as True Positive. This measure is used in the medical field as it gives knowledge about the number of cases that are correctly identified either as malignant or as benign. It is the ability of the model to find all the relevant cases in the dataset.

TABLE 4: EVALUATION CRITERIA

Indices	Calculation Equations
Sensitivity	$\frac{TP}{TP+FN}$
Specificity	$\frac{TN}{TN+FP}$
Accuracy	$\frac{TP+TN}{TP+FP+TN+FN}$
F_score	$2\times\frac{TP}{TP+FP}\times\frac{TP}{TP+FN}\div\left(\frac{TP}{TP+FP}+\frac{TP}{TP+FN}\right)$

B. Results and discussion

The proposed CAD model is experimented with the standard reference dataset MIAS. A total of 310 images have been collected from MIAS. The collected images are first classified as normal or abnormal, then benign or malignant using the proposed CAD system. The performance of the proposed model is evaluated according to different performance measures, namely sensitivity, specificity and success rate (ACC). Before the feature extraction module, ROIs are segmented from unnecessary background regions using trimming. Using the truth information on the ground regarding the coordinates of anomalies in the images, 127 × 127 size ROIs are generated. After cropping, the ROIs are pretreated with CLAHE to improve the contrast. Then, the DTT technique is applied to the extracted ROIs to obtain the characteristic matrix. By applying DTT, we obtain a matrix of characteristics of size s × F, where s and F respectively indicate the number of ROI and the number of characteristics generated. In this work, 252 numbers of entities (F) are generated from DTT, which is quite important. Thus, in order to reduce the size of the feature vector and simplify the classification, KPCA is used, which reduces the number of features by preserving 95% of the variance of the original data. The reduced functionality is transmitted to the four ML algorithms to classify mammograms as normal or abnormal, followed by benign or malignant after initialization of the classifier parameters (table 5).

The performance of KPCA is evaluated in term of the threshold (T). The threshold (T) is an important parameter in KPCA and it represents the variance explained approximately of the total variance of the data. We compared the results at T of 60, 70, 80, 90, 95, and 99%, respectively (fig 2,3).

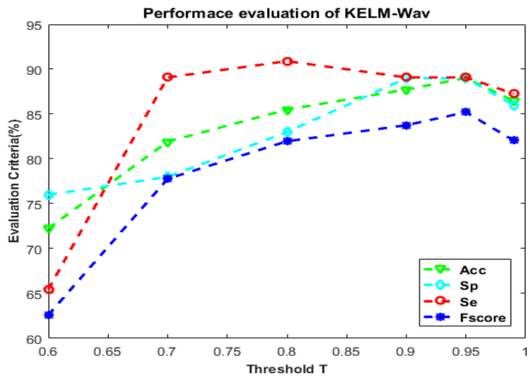

Figure 2: Performance evaluation of $KELM_{WAV}$

The best result was obtained when we used 95% of the variance explained approximately of the total variance of the data (threshold T) using $KELM_{RBF}$ and $KELM_{WAV}$ classifier.

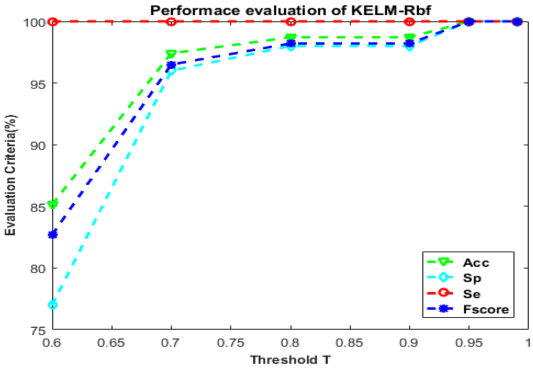

Fig.3: Performance evaluation of $KELM_{RBF}$

TABLE5. CONFIGURATION OF THE CLASSIFIERS USED TO PERFORM THE TESTS.

Classifier	parameters
ELM	Activation function: hard limit . Number of neurons in the hidden layer =250
$KELM_{RBF}$	Parameter of RBF kernel $\sigma = 0.7$
$KELM_{WAV}$	Parameters of wavelet kernel: $w_1 = 0.35, w_2 = 0.1, w_3 = 2$

In order to show the effectiveness of our approach, objective evaluation criteria are measured: Accuracy, sensitivity, specificity, F_score. They are described in table 4. The definition of TP, FN, FP and TN are illustrated in Table 3. For all the configurations, we performed tests for N/A classification (table 6) and M/B classification (table 7). Tables 6 and 7 show the results of accuracy, specificity and F_score of different classifiers obtained for N/A and M/B classification, respectively. Based on the tables above, it was verified that when using $KELM_{RBF}$ and $KELM_{WAV}$ we obtained better results. On the other hand, ELM showed to be more efficient in the cases N/A and M/B classification category using MIAS mammography database.

TABLE 6 CLASSIFICATION RESULTS OF DIFFERENT CLASSIFIERS FOR N/A CLASSIFICATION

Methods	Evaluation Criteria				
	Acc (%)	Sp	Se	F_score	Test-time (s)
ELM	82.26	0.72	1	0.5	0.0016
$KELM_{RBF}$	100	1	1	1	0.004
$KELM_{WAV}$	99.35	0.99	1	0.99	0.009

Based on the table above, it was verified that when using RBF kernel of KELM ($KELM_{RBF}$) we obtained better results of N/A and M/B classification, which were 100% of accuracy, 100% of specificity,100% of sensitivity and a F_score of 100%, for the N/A and M/B classification.

On the other hand, $KELM_{WAV}$ classifier showed to be more efficient in the case of the N / A classification where it gave 99.35% of accuracy, 99% of specificity, 100% of sensitivity and an F_score of 99% than in the case of the M / B classification where it has given 98.18% accuracy, 96% specificity, 100% sensitivity and an F_score of 98%.

TABLE 7 CLASSIFICATION RESULTS OF DIFFERENT CLASSIFIERS FOR M/B CLASSIFICATION

Methods	Evaluation Criteria				
	Acc (%)	Sp	Se	F_score	Test time (s)
ELM	81.8	0.67	1	0.74	0.0008
$KELM_{RBF}$	100	1	1	1	0.004
$KELM_{WAV}$	98.18	0.96	1	0.98	0.009

TABLE 8 COMPARISON OF THE PROPOSED CAD WITH OTHER CLASSIFIERS

References	Database used	Classification Method	Accuracy (%)
[24]	MIAS	SVM	96.5
[25]	MIAS	SVM	92.85
[26]	DDSM	IGWO-ELM	99.5
Proposed methodology	**MIAS**	**ELM**	**82.26 (N/A)**
			81.18 (M/B)
		$KELM_{RBF}$	**100 (N/A)**
			100 (M/B)
		$KELM_{WAV}$	**99.35 (N/A)**
			99.18 (M/B)

V. CONCLUSION AND FEATURE WORKS

In this work, we proposed an efficient method for the detection of breast cancer in mammographic images. The classification rate was obtained using Kernel PCA and KELM. The confusion matrix obtained by classification gives, 100% as classification rate, 100% as sensitivity rate and 100% as specificity rate on the MIAS mammographic image dataset. From the above results, we can conclude that the introduction of Kernel ELM as a feature extraction technique gives higher rates of precision, sensitivity and specificity than other popular methods used in recent literature. However, we believe that increasing the dataset would provide more representative results for pre-diagnosis. So, in future work we will develop a method that can be used to achieve higher results for a large mammographic dataset, the technique of deep learning can also be explored.

References

[1] 2017. Cancer Statistics. (Mar 2017). https://www.cancer.gov/about-cancer/ understanding/statistics

[2] Rampun A, Scotney BW, Morrow PJ, Wang H, Winder J (2018) Breast density classification using local quinary patterns with various neighbourhood topologies. J Imaging 4(1):14

[3] Wichakam et al. (2016) Wichakam I, Vateekul P, Learning K, Convnets CN. Combining deep convolutional networks and svms for mass detection on digital mammograms. 2016 8th international conference on knowledge and smart technology (KST); 2016. Pp. 239–244.

[4] Thawkar S, Ingolikar R (2018) Classification of masses in digital mammograms using firefly based optimization. Int J Image Graphics and Signal Process 10(2):25

[5] Toprak, A. (2018). Extreme Learning Machine (ELM)-Based Classification of Benign and Malignant Cells in Breast Cancer. *Medical science monitor: international medical journal of experimental and clinical research*, *24*, 6537.

[6] Soumya Hundekar and Saritha Chakrasali, "An Identification of Breast Cancer Disease by using ANN using Contourlet Transform", International Journal of Scientific Research and Development, Vol. 6, No. 11, pp. 1-8, 2019.

[7] Soumya Hundekar and Saritha Chakrasali, "An Observation and Categorization of Breast Cancer utilizing Support Vector and Artificial Neural Networks using Discrete Wavelet Transform", ICTACT Journal on Soft Computing, Vol. 9, No. 2, pp. 1851-1855, 2019.

[8] Yap, M. H., Pons, G., Martí, J., Ganau, S., Sentís, M., Zwiggelaar, R., ... & Martí, R. (2017). Automated breast ultrasound lesions detection using convolutional neural networks. *IEEE journal of biomedical and health informatics*, *22*(4), 1218-1226.

[9] Wang, Z., Yu, G., Kang, Y., Zhao, Y., & Qu, Q. (2014). Breast tumor detection in digital mammography based on extreme learning machine. *Neurocomputing*, *128*, 175-184.

[10] Xie W, Li Y, Ma Y (2016) Breast mass classification in digital mammography based on extreme learning machine. Neurocomputing 173:930–941

[11] Qiu, Y., Wang, Y., Yan, S., Tan, M., Cheng, S., Liu, H., & Zheng, B. (2016, March). An initial investigation on developing a new method to predict short-term breast cancer risk based on deep learning technology. In *Medical Imaging 2016: Computer-Aided Diagnosis* (Vol. 9785, p. 978521). International Society for Optics and Photonics.

[12] Suckling J, The MiniMIAS database, Mammographic Image Analysis Society—MIAS, www.wiau.man.ac.uk/services /MIAS/MIAScom.html.

[13] Reza, A. M. (2004). Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement. Journal of VLSI signal processing systems for signal, image and video technology, 38(1), 35-44.

[14] Marcos JV, Crist´obal G (2013) Texture classification using discrete Tchebichef moments. JOSA A 30(8):1580

[15] Wang, Q. (2012). Kernel principal component analysis and its applications in face recognition and active shape models. *arXiv preprint arXiv:1207.3538*.

[16] Schölkopf, B., Smola, A., & Müller, K. R. (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural computation*, *10*(5), 1299-1319.

[17] Jaffel I, Taouali O, Harkat MF, Messaoud H (2018) Fault detection and isolation in nonlinear systems with partial Reduced Kernel Principal Component Analysis method. Trans Inst Meas Control 40(4):1289–1296.

[18] Fazai R, Taouali O, Harkat MF, Bouguila N (2016) A new fault detection method for nonlinear process monitoring. Int J Adv Manuf Technol 87(9-12):3425–3436.

[19] Taouali O, Jaffel I, Lahdhiri H, Harkat MF, Messaoud H (2016)New fault detection method based on reduced kernel principal component analysis (RKPCA). Int J Adv Manuf Technol 85(5- 8):1547–1552

[20] Huang, G.B., Zhu, Q.Y., Siew, C.K.: Extreme learning machine: theory and applications. Neurocomputing 70(1–3), 489–501 (2006).

[21] Barata J, Carlos A, Hussein M (2012) The Moore–Penrose Pseudoinverse: A tutorial review of the theory. Braz J Phys 42(1-2) :146–165

[22] Huang, G. B., Zhou, H., Ding, X., & Zhang, R. (2012). Extreme learning machine for regression and multiclass classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, *42*(2), 513-529

[23] C. Cortes and V. Vapnik. 1995. Support-vector Networks. Machine Learning 20.3 (1995), 273–297. https://doi.org/10.1007/BF00994018

[24] Hariraj, V., Wan, K., & Zunaidi, I. (2017). An Efficient Data Mining Approaches for Breast Cancer Detection and Segmentation in Mammogram.

[25] Bhateja, V., Gautam, A., Tiwari, A., Satapathy, S. C., Nhu, N. G., & Le, D. N. (2018). Haralick Features-Based Classification of Mammograms Using SVM. In Information Systems Design and Intelligent Applications (pp. 787-795). Springer, Singapore.

[26] Mohanty, F., Rup, S., Dash, B., Majhi, B., & Swamy, M. N. S. (2019). A computer-aided diagnosis system using Tchebichef features and improved grey wolf optimized extreme learning machine. *Applied Intelligence*, *49*(3), 983-1001.