Measurement 187 (2022) 110233

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

A novel machine learning approach for breast cancer diagnosis

Bacha Sawssen[a], Taouali Okba[a,b,*]

[a] National Engineering School of Monastir,University of Monastir, Tunisia
[b] Department of Computer Engineering, Faculty of Computers and Information Technology, University of Tabuk, Tabuk, Saudi Arabia

ARTICLE INFO

Keywords:
Breast cancer
Machine learning
Classification
Deferential evolution
Optimization

ABSTRACT

Breast cancer disease is a major public health problem among women worldwide. This article proposes an expert system for the diagnosis of breast cancer disease based on an evolutionary algorithm known as Differential Evolution (DE) of a Radial-Based Function Kernel Extreme Learning Machines (RBF-KELM). In the structure of the RBF-KELM, there are two adjustable parameters of the RBF-kernel which are the penalty parameter C and the RBF-kernel's parameter (σ). These parameters play a major role in the efficiency of RBF-KELM. In this study, the optimal values of these parameters have been obtained using a differential evolution (DE) algorithm. To validate the effectiveness of the suggested approach, DE-RBF-KELM was examined on the two datasets: The Mammographic Image Analysis Society (MIAS) and the wisconsin breast cancer database (WBCD) and the results were satisfactory compared to conventional approaches.

1. Introduction

An automated medical diagnosis, which combines advanced machine learning algorithms and the latest medical approaches [1], has become a very important interdisciplinary technology, producing accurate and non-invasive diagnostics. In recent years, different approaches have been developed for the effective diagnosis and treatment of different diseases such as cancer. Cancer is an extremely important problem with social and financial implications for public health. Different types of cancer have already been reported in the literature and can be classified according to the type of affected cells such as breast cancer.

In recent decades, the frequency of breast cancer has been increasing steadily [2]. Nevertheless, the mortality scenario from this disease appears to have stayed stable thanks to the early detection of breast cancer [3], which is a key to improving the prognosis. However, the diagnosis made does not always bring precise results. Thus, early detection of breast cancer can be performed using computerized analysis programs for better diagnosis. These computer programs are known as computer-aided detection/diagnostic systems (CAD).

The design of a computer-aided detection system for breast cancer summarized in key modules such as the extraction, the selection, and the classification of functionalities that are the subject of rigorous research by the community [4].

The most crucial factor for an automatic diagnostic system is the classifier which is the heart of the system. The reliable classifier must diagnose the disease with the greatest possible accuracy. Different classifiers have been used for automatic detection of breast cancer disease in the literature, such as k nearest neighbor (k-NN) [5], artificial neural network (ANN) [6] and support vector machine (SVM) [7,8].

In recent years, artificial neural networks benefit a lot from randomly generated parameters, not only in the learning speed but also in the generalization performance. Among them, Extreme Learning Machine (ELM) is popular for its fast learning speed, real-time processing capability, and is well recognized by the research community [9-11].

Recently, ELM has been widely used in classification, regression problems, and detection of cell morphology of mammography image features [12] thanks to its properties of rapid-learning ability.

ELM randomly selected all hidden node parameters from generalized hidden single layer feedforward networks (SLFN) and analytically determines the output weights of SLFNs. Although the output weights are calculated analytically, there does not exist a rule for determining the optimal number of hidden neurons and the activation function. ELM may not provide high classification performance due to the above-mentioned cases. To overcome these limitations, a Kernel Extreme Learning Machine (KELM) [13] is proposed to solve these problems [14,15]. The kernel type can be in the form of a polynomial kernel (Poly-KELM), a Kernel with a Radial Base (RBF-KELM), or a Wavelet Kernel (Wav-KELM) (Table 2).

RBF-KELM uses two important parameters, namely the control parameter C and and the RBF-kernel's parameter(sigma (σ)) to have the

* Corresponding author at: National Engineering School of Monastir,University of Monastir, Tunisia.
E-mail address: taoualiok@gmail.com (O. Taouali).

https://doi.org/10.1016/j.measurement.2021.110233
Received 6 February 2021; Received in revised form 11 September 2021; Accepted 24 September 2021
Available online 1 October 2021

Nomenclature

CAD	Computer-Aided Diagnosis
ELM	Extreme Learning Machine
KELM	Kernel Extreme Learning Machine
RBF-KELM	Radial Base Function-kernel ELM
Poly-KELM	Polynomial-kernel ELM
Wav-KELM	Wavelet-kernel ELM
SLFN	Single Layer Feed-forward Network
PCA	Principal component analysis
KPCA	Kernel Principal component analysis
DE	Differential Evolution
SVM	Support vector machines
k-NN	K-Nearest Neighbors
ANN	Artificial Neural network
WBCD	Wisconsin Breast Cancer Diagnostic dataset
ROI	Region Of Interest
DTT	Discrete Tchebychef Transform
NP	Number of a Population
CR	Crossover Rate
F	mutation scale Factor
N/A	Normal or Abnormal
M/B	Malignant tumor or Benign
MIAS	Mammographic Image Analysis Society database
WBCD	Wisconsin Breast Cancer Database
TNR	True Negative Rate
TPR	True Positive Rate
FNR	False Negative Rate
FPR	False Positive Rate

final output weights. These parameters play a major role in the efficiency of the RBF-KELM. Consequently, the values of these two parameters must be carefully adjusted according to the problem solved. In this study, the optimal values of these two parameters have been obtained using a differential evolution (DE) algorithm.

The presented work proposes an expert diagnostic system for breast cancer disease build on the RBF-KELM classifier and a differential evolution algorithm called DE-RBF-KELM. The competitivenessof the suggested DE-RBF-KELM method is assessed by static methods such as classification accuracy, F-score and the sensitivity and specificity analysis. In addition, the efficiency of the suggested designis compared to that of advanced approaches.

This article is structured as follows. Section 2 presents the suggested methodology and basic details required as a DE-RBF-KELM project. Section 3 presents the classification results that demonstrate the effectiveness and speed of the suggested approach. Finally, Section 4 presents the final remarks as well as the possibilities for future improvements.

2. Materials and methods

2.1. Datasets used

The first dataset is the MIAS database [16], 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 size 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).

The abnormal category is further divided into six classes: calcification, spiculated masses, architectural distortion, asymmetry and other ill-defined masses. In the present work, all the six types of abnormalities are arranged in the abnormal category for a classification problem.

The second dataset is the Wisconsin Breast Cancer Diagnostic (WBCD) dataset. In particular, much work has been published on Wisconsin Breast Cancer Diagnostic (WBCD), see for example [3,17].

This dataset contains 569 different instances of which 63% of the total cases are benign and 37% are malignant. All attributes are calculated from a scanned image of a fine needle aspirate (FNA) of patient breast tissue. All cell nuclei in breast tissue are described by ten true-valued characteristics, and for all of these characteristics the mean, standard error, and "worst" (mean of the 3 most important values) are determined. In total, 30 attributes for all images were obtained [18].

Table 1
Number of training and validation 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

Table 2
Formulas and parameters of the used kernel types.

Name	Rule and parameters
Poly-KELM	$K_{ELM} = K_P(x,y) = (1 + \langle x,y\rangle)^P$ P: Degree of polynomial kernel
RBF-KELM	$K_{ELM} = K_\sigma(x,y) = exp\left(-\frac{1}{2}\frac{\|x-y\|^2}{\sigma^2}\right)$ σ: Parameter of RBF-Kernel
Wav-KELM	$K_{ELM} = K_w(x,y) = cos\left(w_1\frac{\|x-y\|}{w_2}\right)exp\left(-\frac{\|x-y\|^2}{w_3}\right)$ w_1, w_2, w_3: Parameters of wavelet kernel

2.2. Methodology adopted

The strategy used in this research is to select the best characteristics using the Kernel Principal Component Analysis (KPCA) algorithm to obtain better breast cancer prediction results using the proposed DE-RBF-KELM machine learning algorithms. The experimental results are validated on two data sets:

The first is the public MIAS database, which is a collection of digital mammographic images. In general, these images include various types of imaging noise, pectoral muscle, and artifacts that are undesirable regions in digital mammography. Therefore, these undesirable regions must be removed to extract the true Region of Interest (ROI) on which the most suitable features are extracted. Then Contrast Limited Adaptive histogram equalization (CLAHE) [19] is used to improve the contrast of these images. Then the discrete Chebyshev transform is used to extract the DTMs from each given image.

The second dataset is the Wisconsin Breast Cancer Diagnostic (WBCD) dataset. This dataset contains 569 different instances. All these instances are described by ten true-value characteristics, and for all these characteristics, the mean, the standard error, and the «worst» (mean of the three most important values) are determined. A total of 30 attributes for all images were obtained. In this case, it is not necessary to use either histogram equalization or discrete Chebyshev analysis.

2.2.1. Pre-processingof the MIAS dataset

The MIAS dataset is a collection of digital mammographic images. Typically these images include various types of imaging noise, pectoral

muscle, and artifacts which are unwanted regions in digital mammograms. Therefore, these unwanted regions must be removed to extract the real region of interest (ROI) on which the most adequate features are extracted. this step is considered to be a vital step before running other modules in a CAD system.

This step recapitulates in two parts, in the first one, a cropping operation on the original mammograms was done to remove the background, the muscle, and the label which can be seen as noise which decreases the classification performances. The desired ROIs have a size of 127 * 127 pixels. The MIAS dataset contains ground truth information given on the position and radius of the abnormal regions in each image. This information is used for manually cropping abnormal images that contain tumors. However, normal images are cropped to any arbitrary location to get ROI. As some of the images selected from the datasets are of low contrast, so it is required to enhance the contrast of such images. Hence, Therefore, in the second part of this step, CLAHE is used to enhance the quality of the low-contrast ROIs.

Fig. 1 shows a case of an abnormal mammogram in the MIAS database called "mdb117″ as well as the corresponding ROI and its enhancement result using CLAHE.

2.2.2. Feature extraction based on DCT

In recent years, mathematical moments were widely used in several computer vision areas of research such as objects and shape descriptors, pattern recognition, template matching, and image analysis. These moments are scalar quantities used to characterize an image, reflecting significant attributes of it. They provide the most salient characteristics and precise features of representation capability for any given image [20]. The literature distinguishes two kinds of orthogonal moments: the continuous and the discrete. However, discrete moments provide better performances than continuous [21]. Several types of mathematical discrete moments have been developed and used as objects feature extraction since they have less redundancy and represent the object more faithfully than the continuous moments such as Zernike [22] or Legendre [23].

The present study, focuses on a certain category of discrete mathematical moments called discrete Chebyshev moments.. These moments are obtained by a Discrete Chebyshev Transform (DCT) and they are widely used for objects recognition and classification [24]. DTM are determined by projecting the input image on to a a collection of Chebyshev polynomials as follows:

For a given image $f(x,y)$ of size $N \times N$ and value $x$with in range $[0, N-1]$, a collection of Chebyshev polynomials $t_n(x); n = 0,1,\cdots N-1$, is defined through the use of the following recurrence relation:

$$t_n(x) = \frac{(2n-1)t_1(x)t_{n-1}(x) - (n-1)(1-\frac{(n-1)^2}{N^2}t_{n-2}(x))}{n} \quad (1)$$

where $t_0(x) = 1$ and $t_1(x) = \frac{2x+1-N}{N}$

Set $\{t_i\}$ has a squared norm as follows:

$$\rho(n,N) = \sum \{t_i(x)\}^2 \quad (2)$$

The Chebyshev moments $T_{mn}(m,n = 0,1,\cdots,S-1)$ can be represented by:

$$T_{mn} = \frac{1}{\rho(m,S)\rho(n,S)} \sum_{x=0}^{N-1}\sum_{y=0}^{N-1} t_m(x)t_n(y)f(x,y) \quad (3)$$

Along spatial coordinates, the repetitive variation patterns are represented by the image texture. Because of such a repetitive property and using a set of Chebyshev polynomials, the patterns can be captured by correlating an image, so that a larger magnitude of *Tmn*will bring to a higher correlation. The $F(k)(k = 0,1,...2N-2)$ feature vector can measure such a correlation while considering the image directions expressed by:$F(k) = \sum_{m+n=k}|T_{mn}|$.

Feature $F(k)$ gives data about the texture properties which is able to be viewed like a texture signature.

2.2.3. Dimensionality reduction with KPCA

The most common technique which is widely used to downsize the dimensionality of features is the principal component analysis (PCA) [25]. Despite the proven effectiveness of the PCA technique, it is build on the linearity of the system. To address the above issue, KPCA method has been developed in the literature [26]. KPCA can be presented in two steps: the first step is to project the input data onto the high dimensional space through a nonlinear mapping function, and the second step is to implement PCA in that feature space.

KPCA has been widely used to model various nonlinear processes [27–29]. In this work, KPCA method is used that reduced the size of the image as following:

$X = [x_1, x_2, \cdots, x_n]^T$ is the data matrix up grade to unit variance and zero mean. The mapping of sample in to high dimensional space can be written as:

$$\varphi : E \subset \mathbb{R}^m \rightarrow F \subset \mathbb{R}^H; \qquad x \rightarrow \varphi(x) \quad and \quad C_\varphi = \frac{1}{N}\sum_{i=1}^{N} \varphi_i \varphi_i^{'} = \frac{1}{n-1} X^T X \quad (4)$$

where $x \in E \subset \mathbb{R}^m$ is a data vector and $\boldsymbol{C}_\varphi$ is the covariance matrix in F

Where $\varphi = \varphi(x_i) \in \mathbb{R}^N$and $X = [\varphi(x_1), \varphi(x_2), \cdots, \varphi_N]^T \in \mathbb{R}^{N*h}$. The KPCA reference model (The principal components of the mapped data $\varphi(x_1), \varphi(x_2), \cdots, \varphi(x_N)$) is computed by solving the eigenvalue decomposition of the covariance matrix C_φ in high dimensional space H as follows:

$$\lambda_j \mu_j = C_\varphi \mu_j \quad with \quad j = 1, \cdots, h \quad (6)$$

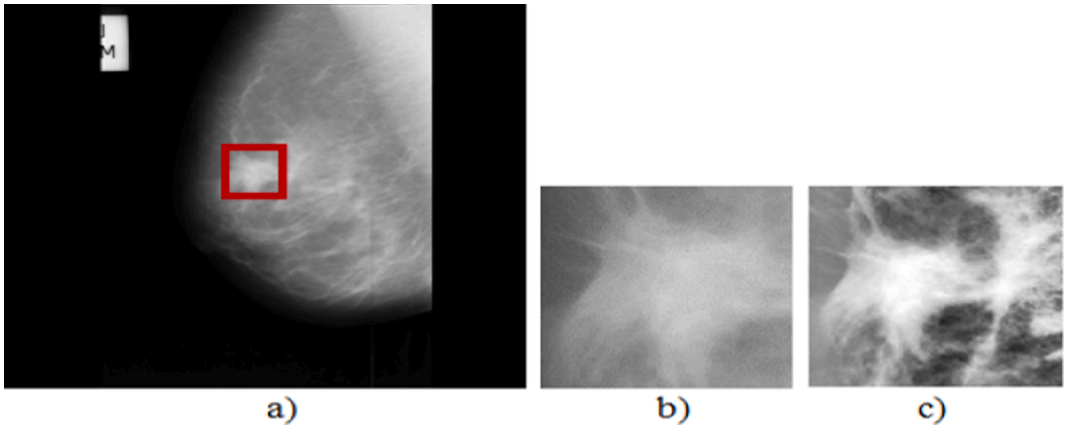

Fig. 1. Enhancement results for mammogram image (Case "mdb117″): (a) Original mammography, (b) the corresponding ROI (c) the corresponding enhanced ROI.

where μ_j is the j[th] eigenvector of C_φ corresponding to eigenvalueλ_j.

For$\lambda_j \neq 0$, there exist coefficients $\alpha_{ij} i = 1, \cdots, N$ such all eigenvectors $\mu_j$can be considered a linear combination of $[\varphi(x_1), \varphi(x_2), \cdots, \varphi_N]$ and can be expressed by:

$$\mu_j = \sum_{i=1}^{N} \alpha_{ij} \varphi(x_i) \quad j = 1, \cdots, n \tag{7}$$

However, in practice, the mapping function φ is not defined and then the covariance matrix C_φ in high dimensional spacecannot be calculated implicitly. Thus, instead of solving eigenvalue problem directly on C_φ, the kernel trick is applied. The inner product given in (4) may be calculated by a kernel function $\boldsymbol{K}(.,.)$ that satisfies Mercer's theorem [30] as follows:

$$\langle \varphi(x), \varphi(x^{'}) \rangle_H = K(x, x^{'}) \quad \forall x, x^{'} \in \mathbb{R}^m \tag{8}$$

Let us define a Gram matrix $K \in \mathbb{R}^{N \times N}$ associated to a function k as:

$$K = \begin{bmatrix} K(x_1, x_1) & \cdots & K(x_1, x_N) \\ \vdots & \ddots & \vdots \\ K(x_N, x_1) & \cdots & K(x_N, x_N) \end{bmatrix} \in \mathbb{R}^{N \times N} \tag{9}$$

Applying the Gram matrix may reduce the problem of eigenvalue decomposition of C_φ. Hence, eigendecomposition of the Gram matrix K is equivalent to performing PCA in $\mathbb{R}^H$.

Many kernel functions have been proposed in literature such as: Linear Kernel, Polynomial kernel, a kernel with a radial base (RBF-kernel). RBF-kernel is used in the proposed methodology.

2.2.4. Classification

ELM is a type of single hidden layer feed-forward network (SLFN). ELM was originally proposed for training SLFNs and was then extended for training the generalized SLFNs where the hidden layer needs not be neuron alike.

In ELM, the used structure is a typical SLFN structure which consists of an input layer, a hidden layer, and an output layer (see Fig. 2). Each neuron is linked using a weighted connection called *weight* (w). Other parameters are also used, mainly the *bias* (b) and the activation function (g) which calculates the output(y). The transfer function *f* can be logarithmic, linear, Tang hyperbolic, radial basis, or sigmoid functions. The network is described using the triplet (w, b, g).

The ELM learning objective is to get output weight, where

$$H\beta = Y \Rightarrow \beta = H^{+} Y \tag{10}$$

H is the hidden output matrix. H^{+} is the Moore-Penrose generalized inverse [31] of H. To solve H^{+}, ELM can be introduced as the following. The hidden layer output matrix H with L hidden neurons can be computed as:

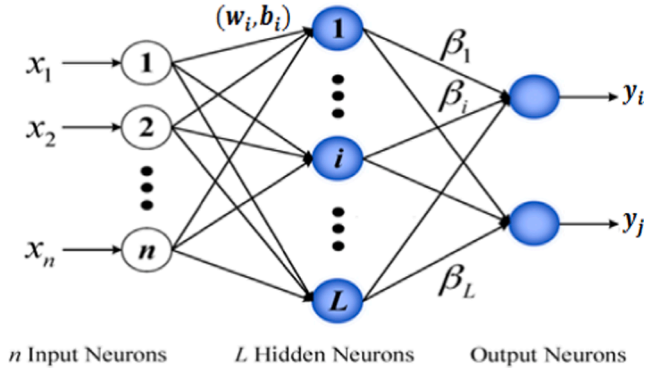

Fig. 2. Structure of ELM.

$$H = \begin{bmatrix} g(w_1 . x_1 + b_1) & \cdots & g(w_L . x_1 + b_L) \\ \vdots & \ddots & \vdots \\ g(w_1 . x_N + b_1) & \cdots & g(w_L . x_N + b_L) \end{bmatrix}_{N \times L} \tag{11}$$

where $X = [x_1, x_2, \cdots, x_n] \in R^{n \times N}$ is the input data of N samples, $Y = [y_1, y_2, \cdots, y_n] \in R^{N \times c}$ is the input label, $w = [w_1, w_2, \cdots, w_L] \in R^{L \times c}$is the input weight, $\beta = [\beta_1, \beta_2, \cdots, \beta_L] \in R^{L \times C}$ is the hidden layer biases, $g(\hat{A}\cdot)$ is the activation function. With the hidden layer output matrix H, ELM can be formulated as follows:

$$\min_{\beta \in R^{L \times C}} \frac{1}{2} \|\beta\|^2 + \frac{C}{2} \sum_{i=1}^{N} \|\varepsilon_i\|^2 \tag{12}$$

$$s.t \quad h(x_i)\beta = y_i - \varepsilon_i, i = 1, 2, \cdots, N \Leftrightarrow H\beta = Y - \varepsilon \tag{13}$$

where $\boldsymbol{\varepsilon} = [\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \cdots, \boldsymbol{\varepsilon}_N]$ denotes the prediction error matrix to the training data, and C is a penalty constant on the training errors.

ELM has some inevitable problems, such that some hidden nodes can receive an input weight very close to 0, commonly called dead nodes because of the random selection of the input weights. This phenomen leads to the minimal effect of these nodes and eventually affects the output accuracy so that some high dimensional dot product operations will appear in the training process. This problem is not unique but also for nonlinear samples, the linear weighted mapping method often has inevitable errors, which makes it possible to downsize the training precision.

KELM increases the robustness of ELM by transforming linearly non-separable data in a small space into linearly separable data. As for ELM with kernels, it obtains a better regression and classification accuracy by introducing kernel, if $h(x)$ is unknown, i.e., an implicit function, one can apply the Mercer's conditions on ELM, and define a for ELM that takes the form $K_{ELM} = HH^T$.

With:

$$K_{ELM}(i,j) = h(x_i)h(x_j) = K(x_i, x_j) \tag{14}$$

So, the output function of the kernel ELM (KELM) is:

$$f(x) = h(x)\beta = \begin{pmatrix} K(x, x_1) \\ \vdots \\ K(x, x_N) \end{pmatrix} \left(\frac{I}{C} + \mathrm{K}_{ELM} \right)^{-1} Y \tag{15}$$

Now there are two implementations of ELM: 1) ELM and 2) KELM. The difference between ELM and KELM lies in that the output function of ELM is obtained using the matrix H, while the output function of KELM is computed from K_{ELM}. H is the hidden layer output matrix defined in (Eq. (11)) and is dependent on the network architecture. However, K_{ELM} is independent on the network architecture and can be chosen as any kernel function from Table 2. Inthe proposed methodology, and to implement KELM, three common kernels are used such as: The kernel with a radial base (RBF-KELM), the polynomial-kernel (Poly-KELM) and the wavelet-kernel (Wav-KELM).

In the implementation of ELM, it is necessary to solve the coefficient β defined in (Eq. (10)). While in KELM, it is necessary to solve the coefficient $\left(\frac{I}{C} + \mathrm{K}_{ELM} \right)^{-1} Y$.

2.3. Proposed DE_RBF_KELM

2.3.1. Model selection based on DE

In evolutionary algorithms, differential evolution (DE) [32] is a method which optimizes a problem of improvement of a candidate solution compared to a quality measure given iteratively. Such methods are commonly known as metaheuristics as they make few or no assumptions about the problem being optimized and can search very large spaces of candidate solutions.

Instead of recombining the solutions under conditions imposed by a probabilistic scheme, and unlike other evolutionary algorithms, the DE algorithm generates new individuals by disturbing the solutions with a difference in scale between two randomly selected population vectors [33]. This increases the heterogeneity in the populations.

In addition, if the offspring outperform its parent, DE adopts a one-to-one reproduction logic that allows the replacement of only one individual. To evolve the population, DE generally uses crossbreeding, mutation and selection operators like the GA genetic algorithm, but the definitions of these operators are quite different.

Additionally, DE considers mutation to be the most important operation as crossover is the main evolutionary strategy in GA.

Differential evolution has also been shown to be more efficient and robust than other evolutionary techniques such as AG and particle swarm optimization [34]. It converges faster, especially for difficult cases.

The DE-based model selection flowchart in RBF-KELM is shown in Fig. 3. The initial step consists of reading the data set and randomly initializing the hyperparameters, then the learning precision is calculated. Then this precision used by the DE tuning algorithm in updating the hyperparameters. When the terminal condition is satisfied, the routine stops and displays the best values of the hyperparameter.

The parameters illustrated in this flowchart are as follows: the total number of individuals in a population (NP), the crossover rate (CR), the mutation scale factor (F), the search space limits (range), and the cost function that can be evaluated to solve the problem (f).

2.3.2. DE based tuning algorithm

The tuning algorithm is used to select the hyperparameters of RBF-KELM in DE, each population is made up of individuals, chosen in the feasible space of solutions to the problem. Each individual represents the pair of positive hyperparameters (C, σ). The objective function, which guides the direction of the overall optimization process is the precision of the formed RBF-KELM. The key operations in this process are mutation, crossing and selection.

(a) **Initialization step**

For each individual i of population$P$, a se tof multidimensional vectors $x_i^P = (C_i^x, \sigma_i^x)$ is defined, where$i \in [1, NP]$, NP being the population size. The individuals of the initial population,C_i^xandγ_i^x, are generated randomly using

$$\begin{cases} C_i^x = C_{min} + rand[0,1] \times (C_{max} - C_{min}) \\ \sigma_i^x = \sigma_{min} + rand[0,1] \times (\sigma_{max} - \sigma_{min}) \end{cases} \quad (16)$$

(b) **Mutation step**

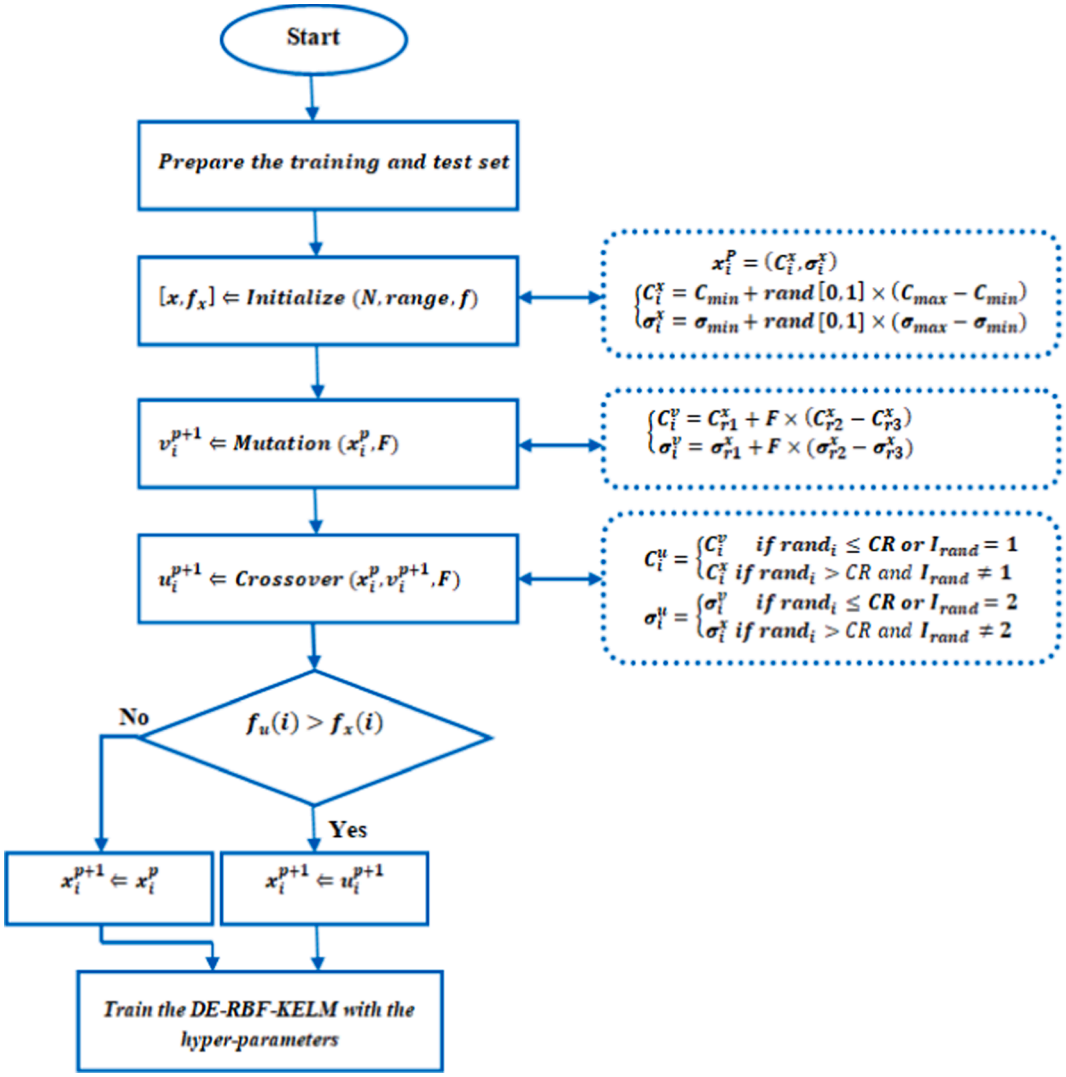

Fig. 3. Flow-diagram of DE based model selection method.

There are three populations that build on the existing population x_{r1}^{P}, x_{r2}^{P} and x_{r3}^{P} are firstchosen randomly. Individuals $v_i^P \triangleq (C_i^v, \sigma_i^v)$ in the population P aftermutationare then generated according to

$$\begin{cases} C_i^v = C_{r1}^x + F \times \left(C_{r2}^x - C_{r3}^x\right) \\ \sigma_i^v = \sigma_{r1}^x + F \times \left(\sigma_{r2}^x - \sigma_{r3}^x\right) \end{cases} \tag{17}$$

F is a control parameter called the step size, which is user-defined generally taken in the range of [0,5].

(c) **Crossover step**

Crossover step can increase the diversity of the population, in which individuals $u_i^P \triangleq (C_i^u, \sigma_i^u)$ are subject to the following rules.

$$C_i^u = \begin{cases} C_i^v & if \ \ rand_i \le CR \ \ or \ \ I_{rand} = 1 \\ C_i^x & if \ \ rand_i > CR \ \ and \ \ I_{rand} \ne 1 \end{cases} \tag{18}$$

$$\sigma_i^u = \begin{cases} \sigma_i^v & if \ \ rand_i \le CR \ \ or \ \ I_{rand} = 2 \\ \sigma_i^x & if \ \ rand_i > CR \ \ and \ \ I_{rand} \ne 2 \end{cases} \tag{19}$$

C_i^u and σ_i^u are the individuals in the population$P$ after crossover. I_{rand} is either 1 or 2 and the quality $rand_i$ is chosen from [0,1], randomlychosen once for each individual. The crossover rate CR, is also a user-defined control parameter lying in the range [0,1] (Table 3). Crossover ensures that each individual is partly similar to a mutatedindividual.

(d) Selection step

The selection step is to select individuals for the next generation. This is done by comparing the values of objective function between the corresponding individuals of the population after mutation and crossing and the parental population P. Thus,

$$x_i^{P+1} = \begin{cases} u_i^{P+1} iff \left(u_i^{P+1}\right) \ge f\left(x_i^P\right) \\ x_i^P otherwise \end{cases} \tag{20}$$

The most certain achievement of the selection step is that all individuals in the new population are better or at least as good as those in the earlier one. Fig. 4 presents the flow-diagram of the developed CAD system.

3. Results and discussion

3.1. Experimental evaluation

The proposed architecture is approved on the two standard reference datasets MIAS and WBCD. MIAS dataset is widely used in mammography analysis and it's freely available. Another reason is various cases of MIAS are labeled by an expert radiologist that build on the experience and biopsy. The mammogram of MIAS is selected from the United Kingdom National Breast Screening Program. Every image is 1024 × 1024 pixels. The selected images are first classified as N/A, then M/B using the proposed classification technique DE-RBF-KELM.

Before the feature extraction module, ROIs are isolated from unnecessary background regions using trimming. Using the truth information on the ground regarding the coordinates of anomalies in the images, 127 × 127 sizes ROIs are generated. After cropping, the ROIs are pretreated with CLAHE to enhancethe contrast. Next, the Discrete Tchebichev Transformation (DTT) technique is used to produce the characteristic matrix of extracted ROIs. The application of the DTT, gives us a matrix of characteristics of size s × f, where s and F respectively indicate the total number of ROI and the total number of characteristics generated. Finally, 252 numbers of entities (f) are generated from DTT, which is quite important.

Table 3
Configuration of the developedclassifier.

Parameter name	Parameter value
Crossover rate (***CR***)	0.5
Control parameter (C)	lying in the range [1,10]
Mutation scale factor (F)	0.5
Sigma (σ)	lying in the range [10e−5, 10e−2]

The second dataset used in this application is The Wisconsin Breast Cancer Database (WBCD) dataset who has been widely used in research experiments. Features are computed from a digitized image of a fine needle aspirate (FNA) of a breast mass. They describe the propertiesof the cell nuclei present in the image. The inputs to the suggested CAD-model are the feature values computed from digitized image of FNA, and the outputs from the model are the two identified classification diagnosis classes benign (B) or malignant (M).

Thus, KPCA was used to downsize the overall size of the features by preserving 95% of the variance of the original data to simplify the classification. The reduced functionality is transmitted to the proposed classification system to classify mammograms as normal or abnormal (N/A), followed by malignant tumor or benign (M/B) for MIAS dataset and M/B for WBCD dataset.

The efficiency of the suggested modelis evaluated according to different performance measures, namely sensitivity, specificity, F-score, success rate (Acc) and the area under the receiver operating characteristic (ROC curve) called AUC. An ROC curve plots the TP and FP values for different classification thresholds. Decreasing the value of the classification cutoff allows more items to be classified as positive, which increases the false positives and true positives rate. AUC stands for area under the ROC curve. This value measures the entire two-dimensional area under the entire ROC curve (by integral calculations) from (0,0) to (1,1).

These different performance measures are calculated using to the entities TNR, FPR, FNR, and (Table 4).

In this Table, TNR and TPR respectively mean True Negative Rate and True Positive Rate, which determine how many normal (non-cancerous) cases classified as (TNR) and how many abnormal ones (cancerous) classified as (TPR), and where FNR and FPR respectively denote False Negative Rate and False Positive Rate, which represent classifying normal cases in abnormal ones (FPR) and abnormal cases in normal ones (FNR). The specificity (Sp) of a classifier is to measure how normal cases can be correctly detected, while the sensitivity (Se) is defined as a frequency of TPR and FNR instances categorized as TPR. The measure used in the medical field provides information on cases correctly identified as benign (B) or malignant (M). In view of the fact, all the relevant cases can be obtained from the dataset by the model.

$$Sensitivity \ \ (Se) = \frac{TPR}{TPR + FNR} \tag{21}$$

$$Specificity \ \ (Sp) = \frac{TNR}{TNR + FPR} \tag{22}$$

$$\text{Accuracy} \ \ (\text{Acc}) = \frac{\text{TPR} + \text{TNR}}{\text{TPR} + \text{FPR} + \text{TNR} + \text{FNR}} \tag{23}$$

$$F_{Score} = 2 \times \frac{\text{TPR}}{\text{TPR} + \text{FPR}} \times \frac{\text{TPR}}{\text{TPR} + \text{FNR}} \tilde{\text{A}} \cdot \left(\frac{\text{TPR}}{\text{TPR} + \text{FPR}} + \frac{\text{TPR}}{\text{TPR} + \text{FNR}}\right) \tag{24}$$

3.2. Results

In these experimental studies, an expert diagnostic system for breast cancer disease that builds on the DE-RBF-KELM method is introduced. The efficiency of the suggested method is also evaluated by classification accuracy, sensitivity, specificity analysis, F-score, the area under the curve AUC, and the computational cost. The suggested optimization

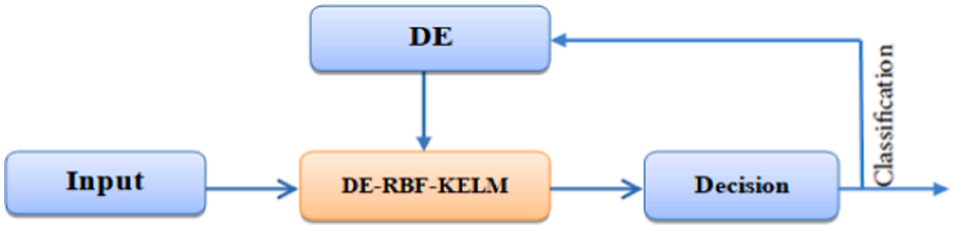

Fig. 4. The flow-diagram of the developed CAD system.

Table 4
Definitions of TNR, FPR, FNR and TPR.

Total population		Predicted class	
		Abnormal	Normal
True class	Abnormal	TPR	FNR
	Normal	FPR	TNR

method is used to search for the optimal values of parameters of the RBF-Kernel function (c and sigma).

To validate the efficiency of the suggested DE_RBF_KELM algorithm, experiments on the two databases MIAS and WBCD were carried out to compare the DE_RBF_KELM algorithm with the SVM algorithm, Poly_-KELM, RBF_KELM and Wav_KELM for breast cancer diagnosis. Table 3 shows the different parameters of the suggested algorithm who are determined using the differential evolution algorithm DE.

The experiment was conducted on professional windows 10 with an Intel (R) Core (TM) processor i7-7700 CPU @ 3.60 GHz 3.60 GHz and total memory space of 64 GB.

- **Experiment I: Validation on the MIAS dataset**

The comparison of the results using the methods SVM, Poly-KELM, Wav-KELM, and RBF-KELM using the MIAS database for N/A and M/B classification are shown in Fig. 5.

According to Fig. 5, for the N/A classification, the suggested DE-RBFKELM method showed a training accuracy and a test accuracy of 100%, which is the best performance compared to the RBFKELM, Poly-KELM, Wav-KELM, and SVM, which gave an accuracy of 98.06%, 58.02%, 98.71% and 95.56% for training respectively and 98.71%, 73.55%, 94.83% and 93.55% for the test respectively. Although the proposed DE-RBF-KELM method improved theefficiency of N/A classification. However, the training and test time of the suggested DE-RBF-KELM method is extremely long compared to the basic RBF-KELM method used in this part which lasted 4.04 ms for the training and 2.07 ms for the test compared to those of RBF-KELM which lasted 4 ms for the training and 0.7 ms for the test.

The statistical evaluation criteria for the N/A classification (see Fig. 5) are calculated at the same time using TPR, TNR, FPR, and FNR, respectively. In this study, the suggested DE-RBF-KELM methodshows the best performance such as 100% of specificity, sensitivity, and F-score compared to the RBF-KELM, Poly-KELM, Wav-KELM and SVM who gave a sensitivity of 100%, 78.18%, 100% and 81.81% respectively and a specificity of 98%, 71%, 92% and 100% respectively and F-score values of 98.21%, 67.72%, 85.22% and 90% respectively.

The statistical evaluation criteria for the M/B classification (see Fig. 5). In this study, the suggested DE-RBF-KELM method shows the best performance such as 100% of specificity, sensitivity, and F-score compared to the RBF-KELM, Poly-KELM, Wav-KELM and SVM who gave a sensitivity of 100%, 86.76%, 100% and 80% respectively and a specificity of 96%, 32%, 64% and 100% respectively and F-score values of 98.36%, 71.23%, 86.95% and 88.89% respectively.

The M/B classification results also showed the superiority of our proposed method compared to the other four classification algorithms (see Fig. 5) which gave a success rate of 100% for testing compared to RBF-KELM, Poly-KELM, Wav -KELM and SVM which gave a test Acc of 98.18%, 72.7%, 94.54% and 90.9% respectively.

However, the computational cost of the suggested DE-RBF-KELM methodis extremely short which is 1.07 ms for the training phase and 0.74 ms for the testing phase.

Comparison results using the proposed DE-RBF-KELM method and the classic RBF-KELM, Poly-KELM and Wav-KELM classifier using the same MIAS datastare summurized in Tables 5 and Fig. 5. In these classic classifiers, each of Polynomial, Wavelet and base radial (RBF) kernel type are respectively used. Readers can find more detailed information on these kernel functions [35,36]. As this table shows, the best result of the suggested DE-RBF-KELM method is 100% using optimal Kernel-function parameter values C and sigma (σ).

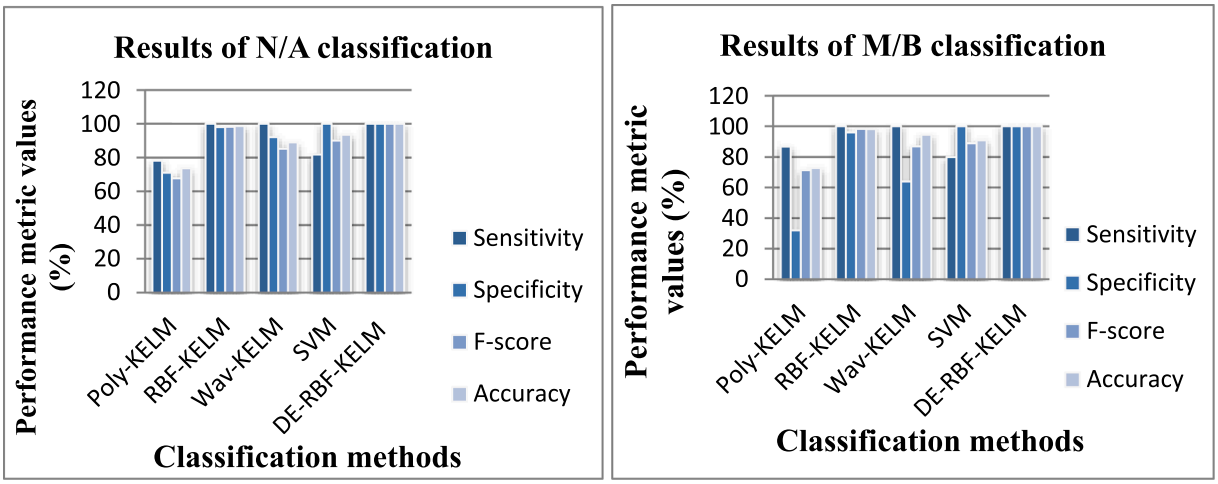

Fig. 5. Bar chart of classification results for N/A classification and M/B classification using MIAS dataset.

Table 5
Computational cost of different methods for N/A and M/B classification.

	N/A classification		M/B classification	
	Train_Time (ms)	Test_Time (ms)	Train_Time (ms)	Test_Time (ms)
SVM	26.24	11.63	12.4	11.5
Poly-KELM	5.2	1.2	4	0.4
RBF-KELM	4	0.7	3.5	0.42
Wav-KELM	5.8	1.2	3.7	0.6
DE-RBF-KELM (Proposed)	4.04	2.07	1.07	0.74

- **Experiment II: Validation on the WBCD dataset**

The same methods as described above were applied to the WBCD dataset. Fig. 6 lists each algorithm's final classification performances. It is shown in the Fig. 6 that in comparison with the algorithms RBF_KELM, Poly_KELM, Wav_KEL and SVM, the proposed DE_RBF_KELM algorithm achieves the highest classification accuracy on WBCD dataset.

In this study, the suggestd DE-RBF-KELM method shows the best performance (see Table 6 and Fig. 6) such as 91.34% of specificity, 91.01% of sensitivity, and 92.83% of F-score value compared to the RBF-KELM, Poly-KELM, Wav-KELM, and SVM, who gave a sensitivity of 94.83%, 93.82%, 100%, 85.91% respectively and a specificity of 84.61%, 76.92%, 80.85%, 95.23% respectively and F-score values of 92.81%, 90.51%, 91.26 %, and 91.04% respectively.

The DE_RBF_KELM algorithm outperforms the RBF_KELM algorithm and obtains the best performance. Tables 6 and Fig. 6 reveal that better overall classification accuracy on breast cancer data sets can be obtained by means of searching for optimal kernel parameters.

To demonstrate the superiority of the suggested model over some of the recently CAD systems, a detailed comparison is summurized in Table 7. This table shows the validity of the suggested DE-RBF-KELM method, so the performances are also better compared to previous studies using different data sets. The DE-RBF-KELM method proposed in this study shows better diagnostic aid.

3.3. Discussion

Detection and diagnosis of breast cancer at an early stage helps in reducing the fatality rate to a greater extent. In the present work, an perfected CAD system has been proposed for breast cancer classification in digital mammograms. Initially, CLAHE is used to enhance the low-contrast images. Then, DCT and KPCA are employed to extract and reduce the features. The reduced feature set is then passed through Poly-KELM, RBF-KELM, Wav-KELM, and SVM classifiers to detect breast cancer in mammograms. The proposed model experimented on the two popular data sets MIAS and WBCD.

It was observed that RBF-KELM achieves the highest performance

Table 6
Classification results for WBCD daset.

	TPR	FPR	TNR	FNR
SVM	61	2	40	10
Poly-KELM	167	24	80	11
RBF-KELM	168	16	92	10
Wav-KELM	141	27	114	0
DE-RBF-KELM (proposed)	162	9	95	16

Table 7
Comparison of Acc and AUC of proposed methodology with other existing approaches.

Reference	Classification technique	Dataset used	Performance metric/s value/s
(AlFayez et al, 2020) [37]	MLP	Public (DMR-IR)	Acc = 80.04 %
(Viswanath et al, 2019) [38]	SVM	Raw Sample Images	Acc = 84.84 %
(Beham et al, 2019) [39]	KNN	149 Mammographic images from the pixel scan hub, Trichy	Acc = 71.14 %
(Saini et al, 2020) [40]	VGG16	BreakHis	Acc = 86%
(Agnes et al, 2019) [41]	MA-CNN	MIAS	AUC = 0.99
(Shen et al, 2019) [42]	CNN	DDSM	AUC = 0.91
Proposed model	**DE_RBF_KELM**	**MIAS**	**Normal/ Abnormal (N/ A) Acc = 100%, AUC = 1 Malignant/ Benign (M/B) Acc = 100%, AUC = 1**
	DE_RBF_KELM	**WBCD**	**Acc = 91.13% AUC = 0.95**

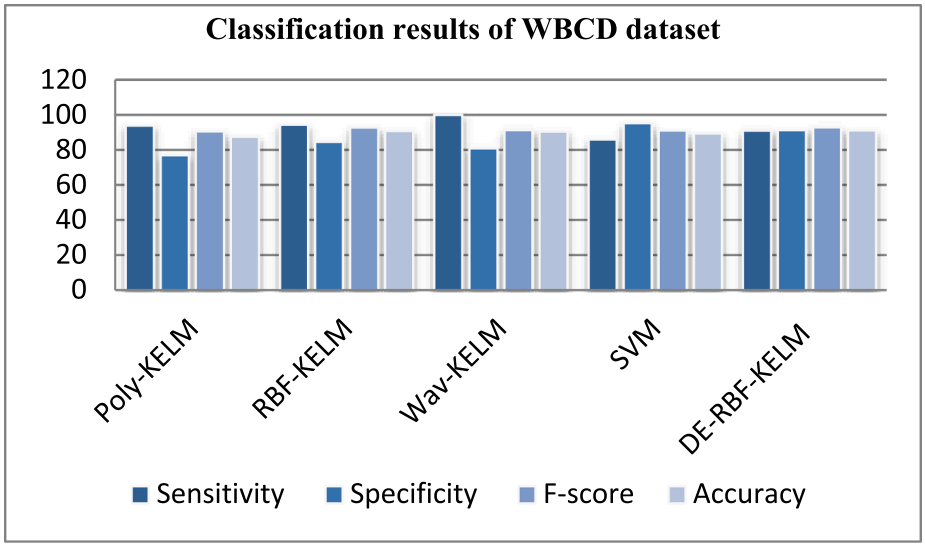

Fig. 6. Bar chart of classification results for WBCD dataset

also at a classification rate of 98.71% for the N/A classification and 98.18% for the M/B classification for the MIAS dataset and 90.78% for the WBCD dataset.

To improve classification performance, the optimization of the c and sigma parameters of the RBF-KELM classifier with the differential evolution algorithm DE is proposed in this work.

The proposed DE-RBF-KELM method has given better performances than the Poly-KELM, RBF-KELM, Wav-KELM, and SVM algorithms. The confusion matrix obtained by the DE-RBF-KELM method gives a 100%, 91.13% classification rate, a 100%, 91.01% sensitivity rate, a 100%, 91.34% specificity rate as well as an F-score of 100%, 92.83% on the MIAS and WBCD data sets respectively. The area under the roc curve is also calculated for the two databases as it was equal to 1 for N/A and M/B classification from the MIAS dataset and 0.95 for the WBCD dataset.

The results obtained show that the introduction of DE for the RBF-KELM parameters's optimization as classification techniques gives higher rates of precision, specificity, and sensitivity than other classification methods.

Furthermore, the competitiveness of the suggested model has been compared with nine recent schemes and it has been noticed that the suggested model achieves improved results over the competent schemes. The high success rate with respect to the accuracy of the suggestedtechnique helps radiologists to make an accurate diagnosis decision to reduce unnecessary biopsies.

3.4. Limitations

Several limitations should be considered when interpreting our findings on MIAS dataset as the calculation time which is a little high in the N/A classification compared to the M/B classification. However, our study was multi-institutional and included all screening mammograms meeting the requirements. For the WBCD dataset, although our data included the important patient-level characteristics of age, breast density, and first or subsequent screening, the breast cancer risk assessment data for analysis doesn't completed. Otherwise, In the absence of lymph node abnormalities detectable on clinical examination or imaging, the guidelines provide for the dissection of the first axillary draining lymph nodes during surgery. It is not always possible to arrive at surgery without diagnostic doubts, and machine learning algorithms can support clinical decisions.

4. Conclusion and future works

This paper presents a novel approach of mammogram images classification method using the differential evolution algorithm for the diagnosis of breast cancer. Preprocessing techniques are used to remove noise and artifacts, suppress background, and enhance contrast to obtain ROIs. Then, DTT and KPCA are employed to extract and reduce the features. Finally, the reduced features are applied in Poly-KELM, RBF-KELM, Wav-KELM, SVM and the proposed DE-RBF-KELM classifiers for mammogram images classification. From the given results, it can be concluded that the introduction of the differential evolution (DE) algorithm as a technique for optimizing the parameters of the RBF_KELM classifier gives a classification rate, a specificity rate, and a sensitivity rate higher than other recently used popular state-of-the-art methods. In the future direction of research, there are two points that could further improve the results of this study. First, the prototype of the method must be implemented so that in the future portable devices can be developed for the automatic classification of breast cancer using mammography images. RBF-KELM and CNN could be conducted.

CRediT authorship contribution statement

Bacha Sawssen: Methodology, Software, Investigation, Data curation, Formal analysis. **Taouali Okba:** Conceptualization, Validation, Writing – original draft, Writing – review & editing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] M. Avanzo, M. Porzio, L. Lorenzon, L. Milan, R. Sghedoni, G. Russo, G. Mettivier, Artificial intelligence applications in medical imaging: a review of the medical physics research in Italy, Physica Med. 83 (2021) 221–241.

[2] R.A. Smith, K.S. Andrews, D. Brooks, S.A. Fedewa, D. Manassaram-Baptiste, D. Saslow, R.C. Wender, Cancer screening in the United States, 2017: a review of current American Cancer Society guidelines and current issues in cancer screening, CA Cancer J. Clin. 67 (2) (2017) 100–121.

[3] B. Zheng, S.W. Yoon, S.S. Lam, Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms, Expert Syst. Appl. 41 (4) (2014) 1476–1482.

[4] K. Wan, I. Zunaidi, An efficient data mining approaches for breast cancer detection and segmentation in mammogram, J. Adv. Res. Dyn. Control Syst. (2017).

[5] V. Nandagopal, S. Geeitha, K.V. Kumar, J. Anbarasi, Feasible analysis of gene expression–a computational based classification for breast cancer, Measurement 140 (2019) 120–125.

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

[7] R. Vijayarajeswari, P. Parthasarathy, S. Vivekanandan, A.A. Basha, Classification of mammogram for early detection of breast cancer using SVM classifier and Hough transform, Measurement 146 (2019) 800–805.

[8] M. Abdar, V. Makarenkov, CWV-BANN-SVM ensemble learning classifier for an accurate diagnosis of breast cancer, Measurement 146 (2019) 557–570.

[9] J. Cao, T. Chen, J. Fan, Landmark recognition with compact BoW histogram and ensemble ELM, Multimedia Tools Appl. 75 (5) (2016) 2839–2857.

[10] A. Iosifidis, A. Tefas, I. Pitas, Graph embedded extreme learning machine, IEEE Trans. Cybern. 46 (1) (2015) 311–324.

[11] S. Li, Z.H. You, H. Guo, X. Luo, Z.Q. Zhao, Inverse-free extreme learning machine with optimal information updating, IEEE Trans. Cybern. 46 (5) (2015) 1229–1241.

[12] A. Toprak, Extreme learning machine (ELM)-based classification of benign and malignant cells in breast cancer, Med. Sci. Monitor: Int. Med. J. Exp. Clin. Res. 24 (2018) 6537.

[13] G.B. Huang, H. Zhou, X. Ding, R. Zhang, Extreme learning machine for regression and multiclass classification, IEEE Trans. Syst. Man Cybernet. Part B (Cybernet.) 42 (2) (2012) 513–529.

[14] S. Chen, C. Gu, C. Lin, Y. Wang, M.A. Hariri-Ardebili, Prediction, monitoring, and interpretation of dam leakage flow via adaptative kernel extreme learning machine, Measurement 166 (2020) 108161.

[15] Y. Zhou, B. Sun, W. Sun, A tool condition monitoring method based on two-layer angle kernel extreme learning machine and binary differential evolution for milling, Measurement 166 (2020) 108186.

[16] Suckling, J., Parker, J., Dance, D., Astley, S., Hutt, I., Boggis, C., Ricketts, I., et al., 2015. Base de données MIAS (Mammographic Image Analysis Society) v1.21 [Dataset]. https://www.repository.cam.ac.uk/handle/1810/250394.

[17] M. Patrício, J. Pereira, J. Crisóstomo, P. Matafome, M. Gomes, R. Seiça, F. Caramelo, Using Resistin, glucose, age and BMI to predict the presence of breast cancer, BMC Cancer 18 (1) (2018) 29.

[18] UCI Machine Learning Repository: Breast Cancer Wisconsin (Diagnostic) Data Set, Retrieved 15 Mar 2012, from UCI Machine Learning Repository: Breast Cancer Wisconsin (Diagnostic) Data set: http://archive.ics.uci.edu/ml/datasets/Breast?Cancer?Wisconsin?(Diagnostic), 2012.

[19] S.M. Pizer, R.E. Johnston, J.P. Ericksen, B.C. Yankaskas, K.E. Muller, Contrast-limited adaptive histogram equalization: speed and effectiveness, in: Proceedings of the First Conference on Visualization in Biomedical Computing, IEEE, 1990, pp. 337–345.

[20] P. Kaur, H.S. Pannu, A.K. Malhi, Comprehensive study of continuous orthogonal moments—a systematic review, ACM Comput. Surv. (CSUR) 52 (4) (2019) 1–30.

[21] H. Shu, H. Zhang, B. Chen, P. Haigron, L. Luo, Fast computation of Tchebichef moments for binary and grayscale images, IEEE Trans. Image Process. 19 (12) (2010) 3171–3180.

[22] A. Aggarwal, C. Singh, Zernike moments-based Gurumukhi character recognition, Appl. Artif. Intell. 30 (5) (2016) 429–444.

[23] K.M. Hosny, G.A. Papakostas, D.E. Koulouriotis, Accurate reconstruction of noisy medical images using orthogonal moments, in: 2013 18th International Conference on Digital Signal Processing (DSP), IEEE, 2013, pp. 1–6.

[24] A. Hmimid, M. Sayyouri, H. Qjidaa, Fast computation of separable two-dimensional discrete invariant moments for image classification, Pattern Recogn. 48 (2) (2015) 509–521.

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

[26] B. Schölkopf, A. Smola, K.R. Müller, Nonlinear component analysis as a kernel eigenvalue problem, Neural Comput. 10 (5) (1998) 1299–1319.

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

[28] R. Fazai, O. Taouali, M.F. Harkat, N. Bouguila, A new fault detection method for nonlinear process monitoring, Int. J. Adv. Manuf. Technol. 87 (9–12) (2016) 3425–3436.
[29] O. Taouali, I. Jaffel, H. Lahdhiri, M.F. Harkat, H. Messaoud, New fault detection method based on reduced kernel principal component analysis (RKPCA), Int. J. Adv. Manuf. Technol. 85 (5–8) (2016) 1547–1552.
[30] J. Mercer, Functions of positive and negative type and their connection with the theory of integral equations. Philosophical transactions of the royal society of London. Series A, containing papers of a mathematical or physical character, Philos. Trinsdict. Rogyal Soc. 209 (1909) 415–446.
[31] J. Barata, A. Carlos, M. Hussein, The moore-penrose pseudoinverse: a tutorial review of the theory, Braz. J. Phys. 42 (1–2) (2012) 146–165.
[32] R. Storn, K. Price, Differential evolutiona simple and efficient heuristic for global optimization over continuous spaces, J. Global Optim. 11 (4) (1997) 341–359.
[33] F. Neri, V. Tirronen, Recent advances in differential evolution: a survey and experimental analysis, Artif. Intell. Rev. 33 (1) (2010) 61–106.
[34] J. Vesterstroem, R. Thomsen, A comparative study of differential evolution, particle swarm optimization, and evolutionary algorithms on numerical benchmark problems, in: Proc. Sixth Congress on Evolutionary Computation (CEC-2004), vol. 2, 1980–1987. IEEE Press, Portland, OR, 2004.
[35] S. Ding, Y. Zhang, X. Xu, L. Bao, A novel extreme learning machine based on hybrid kernel function, JCP 8 (8) (2013) 2110–2117.
[36] W. Huang, N. Li, Z. Lin, G.B. Huang, W. Zong, J. Zhou, Y. Duan, Liver tumor detection and segmentation using kernel-based extreme learning machine, in: 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2013, pp. 3662–3665.
[37] F. AlFayez, M.W.A. El-Soud, T. Gaber, thermogram Breast Cancer Detection: a comparative study of two machine learning techniques, Appl. Sci. 10 (2) (2020) 551.
[38] H. Viswanath, L. Guachi-Guachi, S.P. Thirumuruganandham, EasyChair Preprint Breast Cancer Detection Using Image Processing Techniques and Classification Algorithms Breast Cancer Detection Using Image Processing Techniques and Classification Algorithms, 2019.
[39] M.P. Beham, R. Tamilselvi, S.M. Roomi, A. Nagaraj, Accurate classification of cancer in mammogram images, in: Innovations in electronics and communication engineering, Springer, Singapore, 2019, pp. 71–77.
[40] M. Saini, S. Susan, Deep transfer with minority data augmentation for imbalanced breast cancer dataset, Appl. Soft Comput. 97 (2020) 106759.
[41] S.A. Agnes, J. Anitha, S.I.A. Pandian, J.D. Peter, Classification of mammogram images using multiscale all convolutional neural network (MA-CNN), J. Med. Syst. 44 (1) (2020) 1–9.
[42] L. Shen, L.R. Margolies, J.H. Rothstein, E. Fluder, R. McBride, W. Sieh, Deep learning to improve breast cancer detection on screening mammography, Sci. Rep. 9 (1) (2019) 1–12.