# Classification of Breast Cancer Using Neutrosophic Techniques and Deep Neural Network

**Warda M. Shaban** ( warda.mohammed2010@yahoo.com )
Nile Higher Institute for Engineering and Technology

**Research Article**

**Keywords:** Breast classification, Artificial Intelligence, Feature Selection, Neutrosophic Techniques, Deep Neural Network.

**Posted Date:** October 26th, 2021

**DOI:** https://doi.org/10.21203/rs.3.rs-771965/v1

# Classification of breast cancer using Neutrosophic Techniques and Deep Neural Network

**Warda M. Shaban**

*Nile Higher Institute for Engineering and Technology, Mansoura, Egypt.*

## *Abstract*

*Breast cancer is one of the most common types of cancer that affects women globally and it is the primary cause of death. Early detection of breast cancer is a vital process that can facilitate appropriate treatment, stop the progression of cancer cells, and reduce morbidity and mortality. Artificial Intelligence (AI) and Machine Learning (ML) are the most popular methods that can be used to detect and classify breast cancer accurately. In this paper, a new strategy for classifying breast cancer using Neutrosophic Techniques (NTs) and machine learning techniques is introduced, which is called Breast Cancer Classification Strategy ($BC^2S$). The proposed $BC^2S$ consists of two phases, which are; Data Preprocessing Phase ($DP^2$) and Breast Cancer Classification Phase ($BC^2P$). The main aim of the data preprocessing phase is to; (i) extract features from mammogram images and then remove the outlier items, (ii) select the most effective and informative features from those extracted features using new feature selection method called Efficient Ant Colony Optimization (EACO), and (iii) convert the selected features from classical domain into neutrosophic domain using NTs to give accurate classification through the next classification phase $BC^2P$. The proposed classification model uses Deep Neural Network (DNN) to determine whether the patient is normal or infected with benign or malignant cancer. According to experimental results, the proposed strategy outperforms other competitors in terms of accuracy, precision, recall, and F-measure.*

***Keywords:*** *Breast classification, Artificial Intelligence, Feature Selection, Neutrosophic Techniques, Deep Neural Network.*

## 1. Introduction

Breast cancer is the most dangerous type of cancer that affects women. According to the American Cancer Society (ACS), new cases for 2021 will be 284,200 in the United States and the estimated number of death will be 44,130 [1]. Every year, at least 1.67 million cases of invasive breast cancer are detected worldwide, and about 522,000 deaths, so breast cancer represents a grave healthcare problem [2]. Hence, detection and classification of breast cancer at early stage is a vital process to prevent the spread of cancer cells as well as helping patients to take the treatment, thus saving many lives. Various medical imaging techniques are available to detect breast cancer [3]. These techniques are Magnetic Resonance Images (MRI), Positron Emission Tomography (PET), Thermography (Thermal Imaging) Ultrasound imaging (Sonography), and Mammography as illustrated in figure 1.

Magnetic Resonance Images (MRI) are employed for diagnosing breast cancer. MRI images are performed using radio waves and magnetic fields to produce extremely defined cross sections images [4]. Although it is the best choice for women who have high risk factor for breast cancer, it is not suitable for patients with

metallic devices such as bullets, shrapnel, and surgical clips. Also, MRI machine makes some people uncomfortable [4]. For a PET scan, a radioactive substance called Fluoro Deoxy Glucose (FDG) is injected into the arm vein to separate the tissues. The most radioactive sugar is consumed by the harmful tissues, which are the most active cells. Due to an abnormal increase in glucose metabolism, PET has the ability to detect malignant pathology. But, there are many disadvantages of PET scan such as being very expensive, slow, and not accurate enough [5].

Thermography (Thermal Imaging) is an infrared scan in which an image is performed by mapping the variation of temperature over the breasts. Although thermography is painless, fast, safe, and non-invasive, there is no evidence to support its effectiveness. [6]. For ultrasound imaging (Sonography), it utilizes sound waves for the inner assessment of the body part. The main advantages of ultrasound imaging are the absence of radiation and it is a painless technique. However, it cannot completely cover the breasts and also has poor resolution [7].

Mammography is an X-ray examination. It is the most popular way used to screen breast cancer, however, it suffers from; (i) a false-negative mammogram looks normal even though breast cancer is present, (ii) a false-positive mammogram looks abnormal even though there is no cancer in the breast, and (iii) periods of waiting when additional examinations are required. Therefore, it is important to improve the sensitivity and specificity of any imaging technique used for detecting breast cancer by reading them more than once by different radiologists [8]. Double reading leads to high costs for patients and needs experienced radiologists [8]. Hence, another assist to accurately detect breast cancer should be used such as the use of artificial intelligence. Table 1 shows a brief comparison between various imaging techniques.

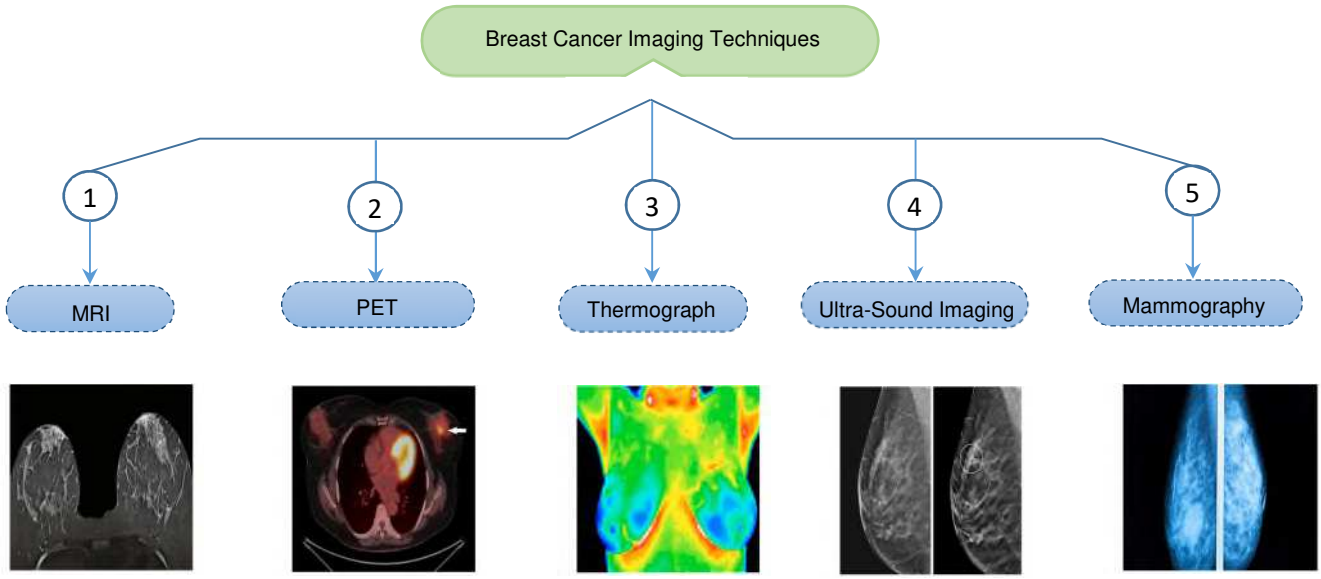


Figure 1, Different Breast Cancer imaging techniques.

Artificial intelligence (AI) has proven to be an efficient tool in healthcare applications. AI techniques are becoming faster, smaller, and more robust over time. These developments are considered the engine of AI to solve complex problems in many fields especially in the medical field [9]. Hence, AI is the best choice for

radiologists and pathologists for disease detection. ML or AI uses the concept of programming to learn how to perform a task automatically. By following this concept, it becomes more efficient and accurate over time. Actually, all ML techniques work in the same way. First, they receive the input training data to build a mathematical model. This mathematical model is based on the input training data. Then, use this mathematical model to solve the problem at hand [10]. Several techniques based on AI and ML have been developed to detect and classify breast cancer [11-15]. However, they cannot introduce the required accuracy and they are more complex.

Table 1: A brief comparison between various imaging techniques.

| Imaging Techniques | Procedure | Advantages | Disadvantages |
|---|---|---|---|
| Magnetic Resonance (MRI) Imaging | MRI images are performed using radio waves and magnetic fields. | ▪ It is suitable for women who have high risk factor for breast cancer. | ▪ It is not suitable for patients with metallic devices.<br>▪ The MRI machine makes some people uncomfortable. |
| Positron Emission Tomography (PET) | In PET scan, fluoro Deoxy Glucose (FDG) is infused into an arm vein to separate between tissues. | ▪ Able to detect malignant pathology. | ▪ High expensive, slow, and not accurate enough. |
| Thermography ( Thermal Imaging) | Thermography is an infrared scan maps the variation in temperature over the breasts and afterward, the image is formed. | ▪ Safe, fast, painless and non-invasive. | ▪ No evidence to prove its effectiveness. |
| Ultrasound imaging (Sonography) | In ultra-Sound imaging, the images were performed using sound waves. | ▪ Absence of radiation<br>▪ Safe and painless. | ▪ Cannot completely cover the breasts<br>▪ Poor resolution. |
| Mammography | Mammography is an X-ray examination for breast cancer. | ▪ It doesn't prevent breast cancer, but they can save lives by finding breast cancer as early as possible. | ▪ It suffers from false positive and false negative. |

Neutrosophy is a branch of philosophy which is an extension of fuzzy sets [16]. Neutrosophy is applied in many applications to solve the problem of uncertainty and indeterminacy. Actually, there are many approaches to neutrosophy includes neutrosophic set theory, neutrosophic logic, neutrosophic probability, and neutrosophic statistics [16,17]. These approaches have proven their superiority in the medical field [18-20]. Neutrosophy set (NS) is considered as a generalization of the classic set, interval valued fuzzy set, fuzzy set, interval valued intuitionistic fuzzy set, and the intuitionistic fuzzy set where each element is defined by a degree of membership, indeterminacy, and non-membership [16,17]. Actually, using neutrosophic techniques provides more information than fuzzy classifiers.

This paper focuses on introducing a new Breast Cancer Classification Strategy ($BC^2S$) to detect and classify breast cancer based on neutrosophic techniques using mammogram images. The proposed strategy consists of two phases, which are; i) Data Preprocessing Phase ($DP^2$), and ii) Breast Cancer Classification Phase ($BC^2P$). Through $DP^2$, historical breast cancer patient's data are collected. Then, four main processes are performed which are; feature extraction, outlier rejection, feature selection, and applying neutrosophic techniques to transform the selected features into neutrosophic domain. Firstly, the features are extracted from the input mammogram images using Grey Level Co-occurrence matrix (GLCM). Then, hasted data that has a very exponential behavior compared to the others should be rejected. Moreover, to promote the classification accuracy and reduce time complexity, the redundant features should be removed and select only the most

effective and significant features using Efficient Ant Colony Optimization (EACO). EACO is a new proposed feature selection method that combines between Binary Ant Colony Optimization (BACO) and crossover strategy. In fact, using BACO can provide accurate subset selection, but it falls into the local optimum. Hence, crossover strategy is applied to solve the problem of limited search space by producing new ant. Consequently, EACO can select the most effective subset of features as; (i) it can provide accurate selection by using BACO as a wrapper method, and (ii) it can solve the problem of limited search space of Ant Colony Optimization (ACO) by the using the crossover strategy. Finally, neutrosophic techniques are applied to convert those selected features from classical domain into neutrosophic domain by extracting three components which are; membership degree (or T), indeterminacy degree (or I), and non-membership degree (or F) for each selected feature. Then, the most robust component of these three components which is membership component is used to provide better classification.

During the second phase (e.g. $BC^2P$), the most robust and efficient component from these three components for each feature namely membership component is used to train the proposed classification model. In this paper, Deep Neural Network (DNN) is used to detect and classify breast cancer patients. DNN is an artificial neural network with multiple hidden layers between input layer and output layer. DNN is an effective tool used in many applications especially in the medical field. Hence DNN can detect and classify breast cancer patients accurately as; (i) it depends on the most robust neutrosophic component for each selected feature, (ii) it has the ability to work with insufficient knowledge, and (iii) it not only follows an algorithm, but it can also predict a solution to a task and draw conclusions based on previous experience. Experimental results have shown that the proposed model outperforms other competitors as it provides the best classification accuracy.

The rest of the paper is organized as follows; Section 2 provides the previous efforts about breast cancer patients' classification. Section 3 focuses on the proposed Breast Cancer Classification Strategy ($BC^2S$). Experimental results are introduced in section 4. Finally, conclusions are presented in section 5.

## 2. Related Work

During this section, the previous efforts to classify breast cancer patients were discussed. In [21], an automatic breast cancer classification method was proposed. The proposed method depended on deep feature fusion and enhanced routing called Feature Extraction BreaKHis CapsNet (FE-BKCapsNet). In fact, the proposed FE-BKCapsNet consists of two networks, called Convolutional Neural Network (CNN) and Capsule Network (CapsNet). CNN was used to highlight semantics, while Capsule Network (CapsNet) focused on detailed information about the position and posture. The proposed FE-BKCapsNet passes through two steps. Firstly, convolution features and capsule features were extracted simultaneously. These features were combined into new capsules to create more discriminative features. Then, the routing coefficients were optimized indirectly and adaptively by adjusting the loss function and incorporating the routing process into the overall optimization process. Experimental results in [21] demonstrated that the proposed model achieved the highest classification accuracy in detecting malignant breast cancer.

As illustrated in [22], a novel Deep Learning Framework (DLF) to detect and classify breast cancer in early stage was developed. DLF based on the concept of transfer learning. The proposed DLF was implemented in two steps. In the first step, the image features were extracted using different pre-trained CNN architectures,

namely; GoogLeNet, Visual Geometry Group Network (VGGNet), and Residual Networks (ResNet). Then, these extracted features were used to feed a fully connected layer to classify malignant and benign cells using average pooling classification. Results in [22] proven that the proposed DLF outperforms all other deep learning architectures for detecting and classifying breast tumors in terms of accuracy.

As presented in [23], a new Deep Learning (DL) based model was proposed to distinguish between cancer and normal cases using mammogram images. The proposed DL model uses Convolutional Neural Network (CNN) architecture based on simplified feature learning and fine-tuned classifier model. Actually, the proposed DL model aims to, (i) evaluate the applicability of different feature learning models, and (ii) enhance the learning ability of the DL models using various CNN architectures for diagnosing of breast cancer accurately. The experimental results in [23] proven the efficiency of the proposed DL model in terms of accuracy, sensitivity, specificity, and precision in which it provides about 92.84%, 95.30%, and 96.72% respectively.

In [24], a new method for detecting the immunohistochemical response of breast cancer was proposed. The proposed method works by finding heterogeneous thermal patterns in the target area. Firstly, ResNet-50 pre-trained model was used to extract deep high dimensional features from low-rank thermal matrix approximation using sparse principal component analysis. Then, a trained Sparse Autoencoder (SA) was used to reduce the dimensions of these features to 16 latent space thermal features. Finally, Random Forest (RF) was used to perform the classification process using these features. The experimental results in [24] shown that the proposed model (ResNet-50-SARF) was performed well.

As deposited in [25], a Patch-Based Classifier using Convolutional Neural Network (PBC-CNN) was proposed. PBC-CNN was used for classifying breast cancer into 4-classes which are; normal, benign, in situ and invasive carcinoma. Actually, the proposed PBC-CNN was implemented in two different modes, One Patch in One Decision (OPOD) and All Patches in One Decision (APOD). OPOD mode was used to predict the class label of each patch. While, in APOD mode, for each extracted patch, the class label was extracted as done in OPOD. Then, the final decision about the class label of the image was taken by using a majority voting scheme. According to results in [25], the proposed OPOD mode achieved a patch-wise classification accuracy of 77.4% for 4 and 84.7% for 2 classes respectively, while APOD technique achieved an image-wise classification accuracy of 90% for 4-class and 92.5% for 2-class classification respectively.

In [26], an improved Random Forest (RF) based Rule Extraction (RFRE) method was developed for breast cancer diagnosis. RFRE was used to perform accurate and interpretable classification rules from a decision tree ensemble. Firstly, RF was used to build a large number of decision trees models to generate a large number of decision rules. Then, a rule extraction approach was developed to separate the decision rules from trained trees. At last, an improved multi-objective evolutionary algorithm was used to search for an optimal base predictor in which the constituent rule set is the best trade-off between accuracy and interpretability.

As illustrated in [27], a new method using pre-trained Deep Residual Network (DRN) model and Support Vector Machine (SVM) was proposed. DRN-SVM was used to detect and classify benign and malignant breast tumors using ultrasound images. The proposed DRN-SVM was proposed to reduce the physician effort and improve the classification performance. Firstly, pre-trained DRN was used to extract image features from the convolutional layer of the trained network. Then, SVM with a sequential minimal optimization solver was

utilized to classify these extracted features. The proposed DRN-SVM was applied to 2099 unlabeled 2D breast ultrasound images. Results in [27] proven the potential applicability of the proposed approach to detect and classify benign and malignant breast tumors.

As presented in [28], a Support Vector Machine (SVM) based diagnosis system for breast tumor detection was developed. The proposed diagnosed system called Principle Component Analysis (PCA) based Differential Evolution (DE) algorithm and SVM (PCA-DESVM). The proposed PCA-DESVM system consists of three main steps. In the first step, the redundant information was removed from the original data and informative patterns were extracted using PCA. Then, the optimal values for SVM parameter were tuned using DE algorithm. At last, SVM was used to differentiate between the incoming tumors. In the above literature, the proposed breast cancer classification techniques advantages and disadvantages have been summarized in table 2. All these previous proposed methods used various techniques to diagnosis breast cancer at early stages.

Table 2: a brief comparison of the current breast cancer classification methods.

| Classification Methods | Description | Advantages | Disadvantages |
|---|---|---|---|
| Feature Extraction BreaKHis CapsNet (FE-BKCapsNet) [21] | FE-BKCapsNet consists of two networks called CNN and Capsule Network (CapsNet). CNN was used to extract deep features, while CapsNet was used to extract capsule features. | CNN-CapsNet has the ability to detect malignant tumors from breast cancer efficiently. | It is very time-consuming to classify breast tumor based on capsule features and convolution features. |
| Deep Learning Framework (DLF) [22] | DLF is used to detect and classify breast cancer in early stage. In DLF, features are extracted using three different CNN architectures then; these features are used to feed a fully connected layer for classification. | Transfer learning allows the training of data with fewer datasets and requires less calculation costs. | One of the biggest limitations to transfer learning is the problem of overfitting. |
| Deep Learning (DL) model [23] | DL model is used to distinguish between cancer and normal cases on mammogram using CNN architecture. | CNN has the ability to classify breast cancer accurately. | This method is slow. |
| The proposed ResNet-50 based on a Sparse Autoencoder and Random Forest (ResNet-50-SARF)[24] | ResNet-50-SARF is a new method that was used to Detect Vasodilation as a Potential Diagnostic Biomarker in Breast Cancer Using Deep Learning-Driven Thermomics. | This method addressed one of the biggest challenges in high-dimensional deep feature selection, which selected the best representative deep thermomics from high-dimensional features extracted from a pre-trained deep neural networks model. | This method cannot provide the optimal accuracy. |
| Patch-Based Classifier using Convolutional Neural Network (PBC-CNN) [25] | PBC-CNN is used for automatic classification of breast cancer into 4 classes. It consists of two mode; OPOD and APOD. | OPOD mode cannot give the optimal accuracy. | This method is slow. |
| Improved Random Forest (RF) based Rule Extraction (RFRE) [26] | RFRE is used to perform accurate and interpretable classification rules from a decision tree ensemble. RFRE consists of IRF that was used was used to create the abundant decision rules and | Data normalization is not required as it uses a rule-based approach. | RFRE cannot give the optimal accuracy. |

| | | | |
|---|---|---|---|
| | RE that was developed to separate the decision rules from trained trees. | | |
| Deep Residual Network model and Support Vector Machine (DRN-SVM) [27] | DRN-SVM is used to detect and classify benign and malignant ultrasound breast tumors images. DRN-SVM consists of DRN that was used for feature extraction and SVM that was used to classify extracted features. | SVM is more effective in high dimensional spaces. | SVM requires long training time for large datasets |
| Principle Component Analysis (PCA) based Differential Evolution (DE) algorithm and SVM (PCA-DESVM) [28] | PCA-DESVM was used for breast tumor detection. Firstly, PCA was used to extract informative patterns from original data then; DE was used to tune SVM parameter that was used to differentiate between the incoming tumors. | PCA-DESVM has the ability to classify breast tumor accurately. | The proposed PCA-DESVM is slow |

## 3. The proposed Breast Cancer Classification Strategy ($BC^2S$)

Breast cancer is the most dangerous disease that women face. As a result, automatic medical diagnosis has become extremely important, particularly when a quick, correct, and accurate decision is required. [29-31]. In this section, the proposed Breast Cancer Classification Strategy ($BC^2S$) will be explained in detail. The main target of the automatic system is to reduce human errors. Human errors resulted from many reasons such as; low experience of radiologists, or errors resulted from image interference, noise, blurring, etc. Consequently, using an automatic system will help to avoid these errors as well as reduce costs and reduce time, thus, saving many patients [32]. As shown in figure 2, the proposed $BC^2S$ is composed of two phases. These phases are; Data Preprocessing Phase ($DP^2$) and Breast Cancer Classification Phase ($BC^2P$). The main aim of $DP^2$ is to extract a set of features, reject outlier items, and eliminate redundant features. Then, neutrosophic technique is applied to the filtered features to provide efficient data for the next phase. In $BC^2P$, a robust, fast, and accurate classification process is performed using Deep Neural Network (DNN).

### 3.1.Data Preprocessing Phase ($DP^2$)

The main objective of the Data Preprocessing Phase ($DP^2$) is to clean and filter the patient data. To achieve this goal, $DP^2$ consists of three stages. These stages are; feature extraction stage, feature selection stage, and neutrosophic stage. The details of each stage will be explained in the following subsections.

#### 3.1.1. Feature Extraction Stage

Feature Extraction (FE) is the process of transforming the input image into a set of features. The main aim of FE is to decrease the number of features in a dataset by producing new features from the existing features and then removing the original features [33]. In fact, FE process is performed before using the classification model. Extracting features from mammograms helps the classification model to perform well and thus, make the right decisions [34,35]. Recently, there are several techniques used for feature extraction. These techniques are; Gabor filter, co-occurrence matrix, wavelet transform based-features, etc. [33]. Actually, Gray Level Co-

occurrence Matrix (GLCM) is the most popular, robust method used to convert the input image into a set of features [34,35].

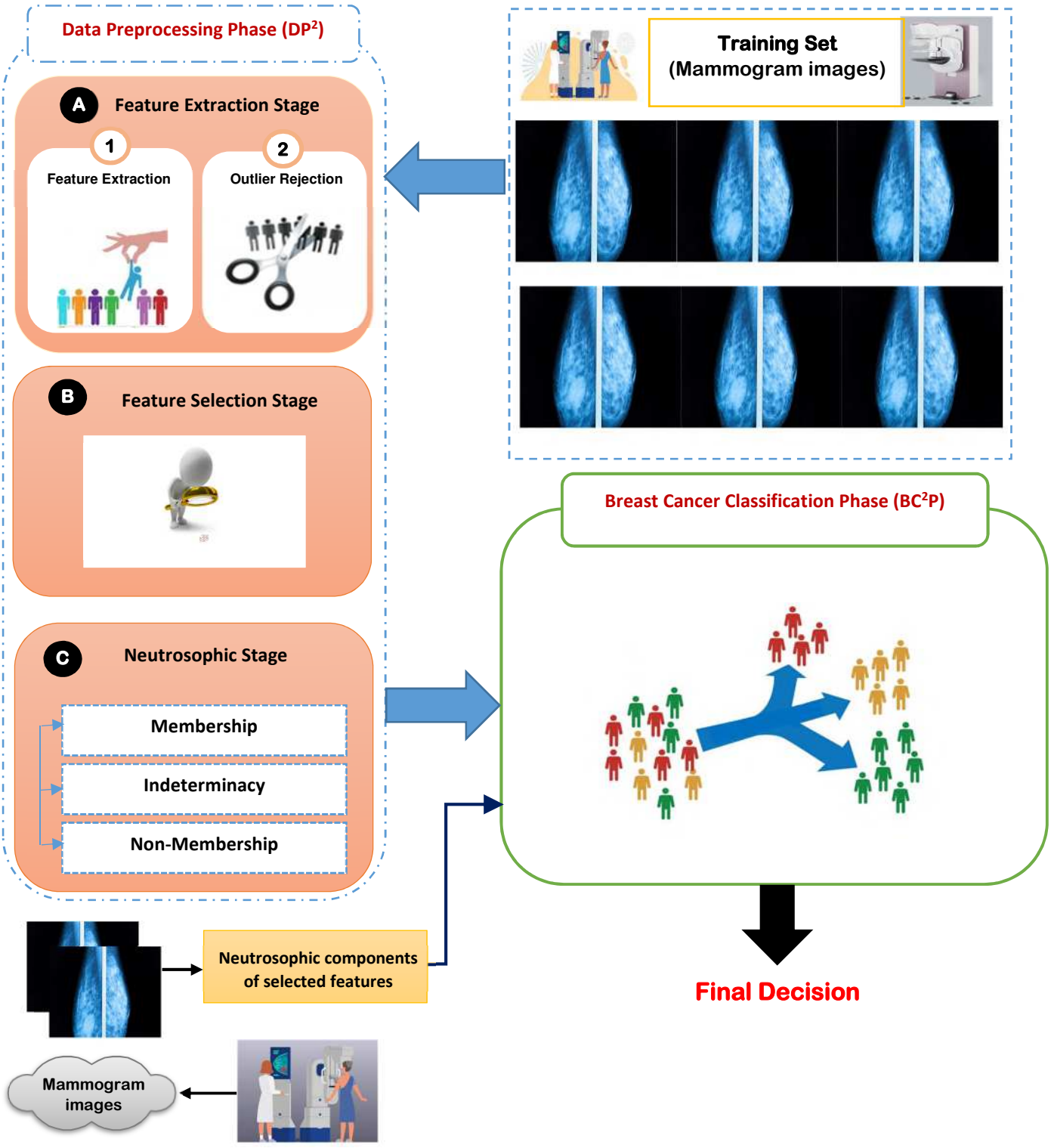


Figure 2, the proposed BC²S

These features are; contrast, homogeneity, correlation, entropy, energy, cluster shade, cluster prominence, autocorrelation, maximum probability, sum of squares (variance),…etc. [34,35].

After performing feature extraction process, outlier items should be rejected before starting to train the used classifier. This is known as outlier rejection. Hence, outlier rejection is the process of removing or eliminating hasted data that has very exponential behavior when compared to others [36-39]. In fact, it is difficult to design a perfect classification model in the presence of these outliers. Therefore, eliminating those outliers can enhance the performance of the classification model by using outlier methods. Actually, outlier methods can be classified into two categories which are; classic outlier approach and spatial outlier approach [37,38]. In this paper, Genetic Algorithm (GA) has been used as an outlier rejection method to eliminate hasted data from the input to promote the classifier's performance [40].

### 3.1.2. Feature Selection Stage

After the feature extraction stage is completed, removing redundant features from those extracted features is a vital task. The reason is that the existence of redundant features can reduce the classifier's performance, increase the training time, and make the system more complex. Thus, removing the redundant features and selecting the most effective and informative ones leads to provide a more robust, efficient, and cost-effective classification model [41-43]. In this paper, a new feature selection method is introduced. The proposed method called Efficient Ant Colony Optimization (EACO). EACO is a wrapper feature selection method that is used to select the most informative features based on a specific evaluation matrices. EACO is a new technique that can precisely select the optimal subset of features that involves the most effective, significant, and informative features for breast cancer classification. Unlike classic method that searches from a single point, EACO deals with large search spaces.

EACO is based on a metaheuristic algorithm which is called Ant Colony Optimization (ACO). ACO was proposed by Dorigo et al. and it was inspired by the behaviors of real ant colonies [44]. It was motivated to solve the problem of traveling salesman by finding the shortest path, similar to the behavior of ants to find the food source [45,46]. Ants start from the nest and move along the same path by following another one. The reason is that every ant deposits a chemical substance called pheromone during its passage and the other ants follow the path that has a high concentration of pheromone. ACO was designed to solve problems in continuous search spaces. In fact, feature selection is an optimization problem that occurs in binary search spaces. Hence, ACO is converted to Binary ACO (BACO) to solve binary optimization problems (e.g. feature selection). Consequently, BACO is a modified version of ACO by assuming that each feature node has two sub-nodes, "1" for selecting the feature and "0" for deselecting the feature, as shown in figure 3.

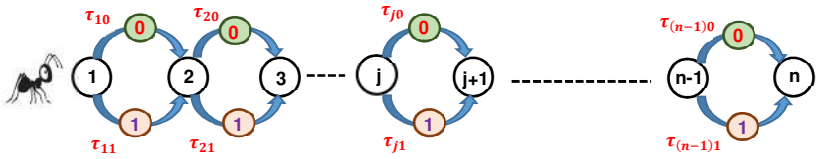


Figure 3, BACO graph.

As shown in figure 3, the graph's edges are made up of "0" and "1" values, which represent the state candidates for each bit. At each iteration, an ant visits all nodes in order to construct a candidate solution. Ants start to search for the optimal path by moving from node *1* to *n* on the graph. Initially, the pheromone concentration is equal in both routes of '1' and '0'. Assume that the pheromone value on the "*i"* bit to state "*j"* is represented by $\tau_{i,j}$, where *j*$\epsilon$ *{0, 1}*. Ant "*A*" determines the direction based on the concentration of pheromone in the different routes. The probability of movement is calculated by using (1) and (2) [47]:

$$P_{i,j}^{A}(0)(t) = \frac{[\tau_{i,j}(0)]^{\alpha} * [\eta_{i,j}(0)]^{\beta}}{[\tau_{i,j}(0)]^{\alpha} * [\eta_{i,j}(0)]^{\beta} + [\tau_{i,j}(1)]^{\alpha} * [\eta_{i,j}(1)]^{\beta}} \qquad (1)$$

$$P_{i,j}^{A}(1)(t) = 1 - P_{i,j}^{A}(0)(t) \qquad (2)$$

Where $P_{i,j}^{A}(t)$ is the probability of movement of ant "*A"* from bit "*i*" to state "*j*" at time t. Additionally, $\tau_{i,j}(0)$ is the pheromone density of the side which the "*j*" is "0" and $\tau_{i,j}(1)$ is the pheromone density of the side which "*j*" is "1". α is the relative importance of the pheromone; α> 0. β is the relative importance of the visibility; β> 0. $\eta_{i,j}(0)$, $\eta_{i,j}(1)$ are the visibility density of the side which the "*j*" is "0" and "*j*" is "1" respectively and $\eta_{i,j} = {}^{1}/_{d_{i,j}}$; $d_{i,j}$ is the Euclidian distance. According to the probability of movement, each ant chooses their path based on the concentration of pheromones on both sides. After each ant has completed its tour, it has a candidate solution in which its length is equal to the number of extracted features. Thus, each solution is represented in the form of a binary vector, where "1" donated that the feature is selected and "0" donates that the feature is eliminated or deselected. Table 3 shows an example of an ant solution, assuming that the number of features "*n=12"* in n-dimensional feature space (n = no. of extracted features). Thus, the ant solution can be represented as; *s= {f1, f2, f3, f4,…, f12}*.

Table 3, the presentation of a binary solution structure.

| $f_1$ | $f_2$ | $f_3$ | $f_4$ | $f_5$ | $f_6$ | $f_7$ | $f_8$ | $f_9$ | $f_{10}$ | $f_{11}$ | $f_{12}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 0 |

Hence, implementing EACO as a feature selection method necessitates a number of critical steps, as illustrated in Figure 4. In figure 4, after each ant builds its own solution, all the ants' solutions are represented in S. Then, to measure the degree of each ant $A_i$, the fitness function of EACO is implemented based on the accuracy index of the classifier. In fact, the fitness function represents the accuracy of a classifier used such as Naïve Bayes (NB) classifier to determine which features are most effective for classifying breast cancer. The fitness function of each ant can be calculated by using (3):

$$Fit(A_i) = Accuracy\ (A_i) \qquad (3)$$

Where *Accuracy* $(A_i)$ is the classification accuracy according to a subset of features in $i^{th}$ ant. The algorithm searches for the best ant with the highest *Fit* $(A_i)$. Although BACO can accurately select the most effective features, it falls into local optimum, which causes premature convergence. Consequently, to solve this

problem, BACO combined with crossover strategy to enhance the exploration of the search space. Combining BACO with crossover strategy leads to create a new solution as described in (4):

$$A_{low-best}(t) = crossover\big(A_{lowest}(t), A_{best}(t)\big) \quad (4)$$

Where $crossover()$ is a process that performs the crossover-based process on the loser ant. $A_{lowest}(t)$ is the loser ant at iteration *t*, and $A_{best}(t)$ is the winner (best) ant at iteration *t*. After evaluating all the ants' solutions, the crossover process takes place, where the best ant and the lowest ant according to their fitness function are selected to perform a pairwise competition and the lowest ants are updated according to the best ones. Hence, this technique includes more exploration and improvements on the lowest ant by creating similar random opportunities either to choose a number of features from the best ant or to keep the existing features according to equation (5):

$$A^i_{low-best}(t) = \begin{cases} A^i_{best}(t) & if \quad rand(i) = 1 \\ A^i_{lowest}(t) & if \quad rand(i) = 0 \end{cases} \quad (5)$$

Where $A^i_{best}(t)$ is the $i^{th}$ feature in the best ant at iteration *t*, and $A^i_{low}(t)$ is the $i^{th}$ feature in the lowest ant at iteration *t.* $rand(i)$ is the generated random value (0 or 1)**.** After updating the lowest ant, the new ant $A^i_{low-best}(t)$ is evaluated using equation (3). Then, the best ant in all swarm $S_{Global}$ ($S_G$) is selected according to (6):

$$S_G(t) = \begin{cases} A_{best}(t) & if\ Fit(A_{best}(t)) > Fit(A_{low-best}(t)) \\ A_{low-best}(t) & otherwise \end{cases} \quad (6)$$

Where $S_G(t)$ is the global solution which represents the best solution in the swarm at iteration *t.* $A_{best}(t)$ is the winner (best) ant at iteration *t*. Additionally, $A_{low-best}(t)$ is the new generated ant from lowest and best ants. *Fit* ($A_{best}(t)$) represents the fitness value of the best ant at iteration *t*, and $Fit(A_{low-best}(t))$ represents the fitness value of the new generated ant at iteration *t*. Figure 5, shows an illustrated example of crossover strategy.

After selecting the best ant in the swarm, the pheromone density is updated. Actually, by the time, the pheromone density will evaporate. Suppose that $\tau_{i,j}(t+1)$ is the intensity at the next iteration *(t+1),* then the pheromone density is updated according to (7) & (8) [45]:

$$\tau_{js}(0)(t+1) = (1-\rho)\tau_{i,j}(0)(t) + \Delta\, \tau^{best}_{i,j} \quad (7)$$

$$\tau_{js}(1)(t+1) = (1-\rho)\tau_{i,j}(1)(t) + \Delta\, \tau^{best}_{i,j} \quad (8)$$

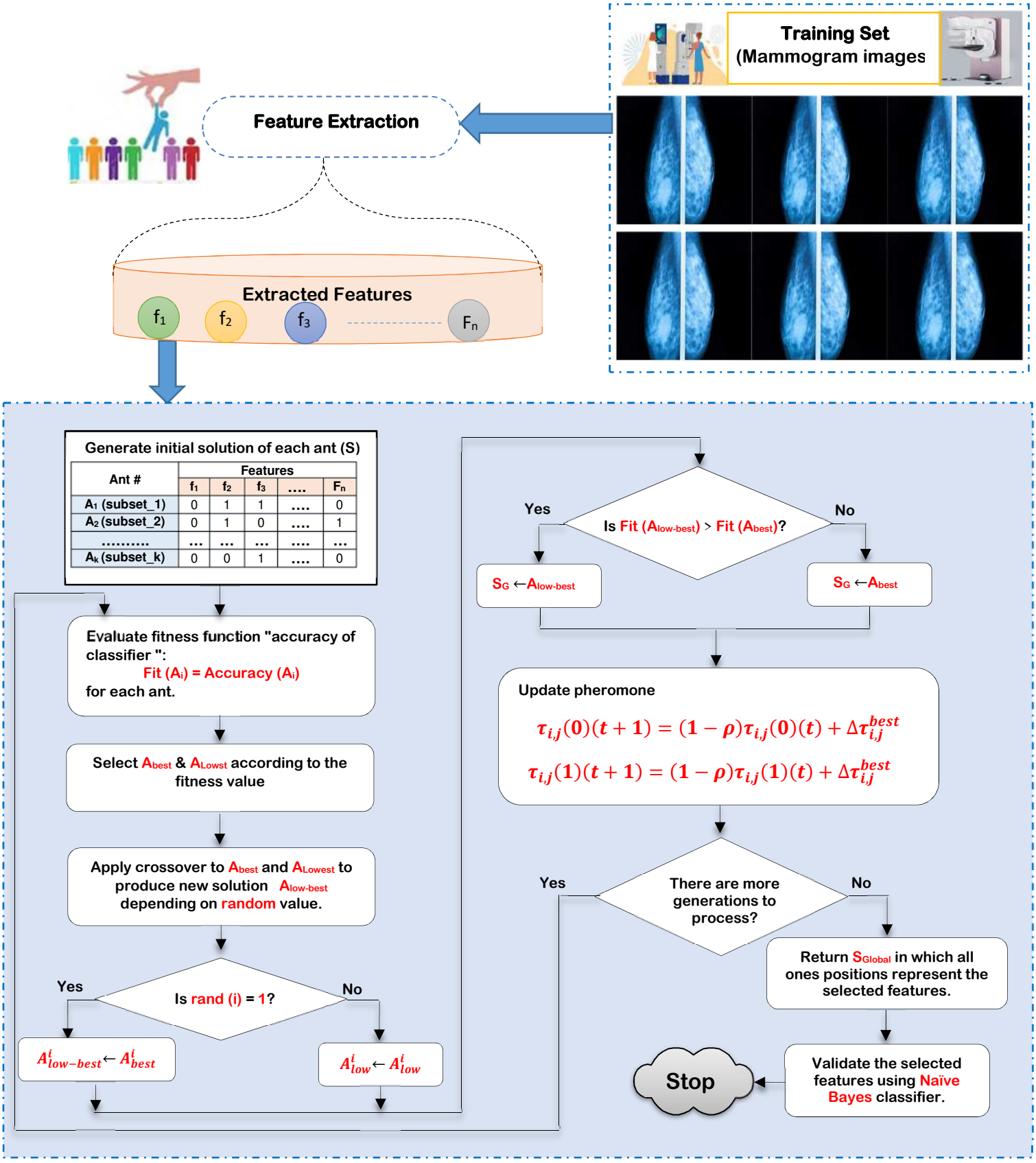


Figure 4, The sequential steps of EACO method.

Where $\tau_{i,j}(t+1)$ is the intensity at the next iteration *(t+1)*, $\rho$ is the evaporation rate; $\rho \in [0,1]$. Additionally, $\Delta\,\tau_{i,j}^{best}$ is the incremental pheromone which is defined as; $\Delta\,\tau_{i,j}^{best} = \frac{1}{f(S^{best})}$; $f(S^{best})$ is the best fitness value, if value "*j*" is selected for *i*-bit of the built solution by ant and zero otherwise. According to equation (7) and equation (8), the pheromone density for the next iteration *(t+1)* is updated only for the best ant. These processes are continued until the termination conditions is satisfied. At last, the best ant in the swarm $S_{Global}$ with the highest fitness value according to the accuracy index of NB classifier is the output and the algorithm terminates where all features donated "1" represent the most significant features for classifying breast cancer. After applying EACO algorithm on the breast's datasets, seven features will be selected as the most significant features. These features are; contrast, homogeneity, correlation, entropy, energy, cluster prominence, and shade.

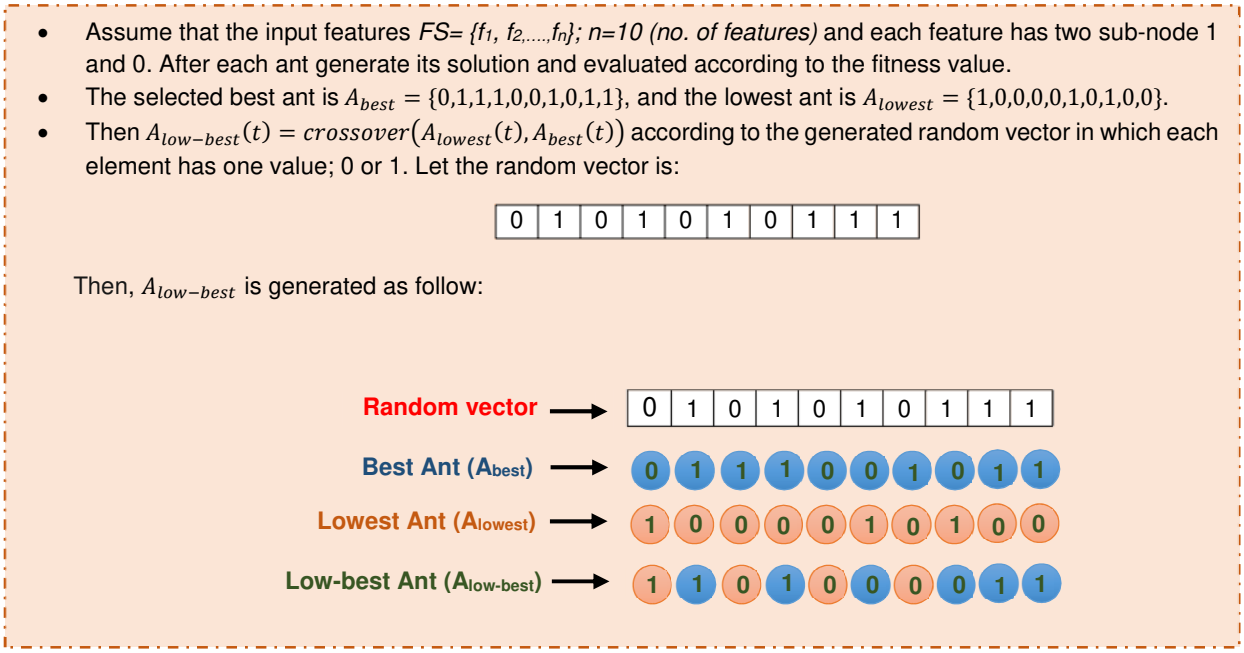


Figure 5, An illustrated example of crossover strategy.

To implement EACO, suppose that there are '*n*' dimensional Feature Space; *FS= {f1, f2,…fn}*. Additionally, the input training data of '*x*' patients can be expressed as; *M= {S1, S2,… Sx}* and the testing data of '*v*' patients can be expressed as; *H= {B1, B2,…, Bv}*. Each item of $S_i \in M$ and $B_j \in H$ is expressed as an ordered set of '*n*' features; $S_i(f_1, f_2, \ldots, f_n)=[f_{1i}, f_{2i}, .., f_{ni}]$and $B_j(f_1, f_2, .., f_n]=[f_{1i}, f_{2i}, .., f_{ni}]$. Hence, each item $S_i$ and $B_j$ can be expressed in '*n*' dimensional features space. In our case, it is very important to reduce '*n*' dimensions or remove redundant features from breast's dataset to overcome overfitting problem and promote the classification model performance. Algorithm 1 illustrated the sequential steps of EACO method.

## Feature Selection using EACO Algorithm

***Inputs:***

*K=No. of ants in swarm "swarm size"*
*A=$A_1$….$A_K$; group of ants in swarm.*
*TRD= (M, FS); Training dataset.*
*TED= (H, FS); Testing dataset.*
*n=|FS|; No. of features in training and testing dataset.*
*t= number of iterations.*
*$\tau_o$ = initial pheromone level for each node state.*
*Initiate ρ; evaporation rate.*
*Initiate α; the relative importance of the pheromone value*
*Initiate β; the relative importance of the heuristic information.*

***Output:***

*O*= the optimum ant of the whole swarm ($S_{Global}$).

***Steps:***

// Construct a complete path for each ant.

**For each ant $A_i \in A$**
  **For each bit**
    $$P_{i,j}^{A}(0)(t) = \frac{[\tau_{i,j}(0)]^{\alpha} * [\eta_{i,j}(0)]^{\beta}}{[\tau_{i,j}(0)]^{\alpha} * [\eta_{i,j}(0)]^{\beta} + [\tau_{i,j}(1)]^{\alpha} * [\eta_{i,j}(1)]^{\beta}}$$
    $$P_{i,j}^{A}(1)(t) = 1 - P_{i,j}^{A}(0)(t)$$
  **End For**
**End For**

// Evaluate fitness value of each ant.

**For every $A_i \in A$**
  Fit ($A_i$)=Accuracy ($A_i$)
**End for**
Select the two ant $A_{best}$, $A_{lowest}$ according to Fit ($A_i$).

// applying crossover strategy using "random value".

Apply crossover to $\mathbf{A_{best}}$ and $\mathbf{A_{lowest}}$ to produce new ant solution $A_{low\text{-}best}$ according to random value
**For every $A_{best}^{i}$, $A_{lowest}^{i}$**
  $$A_{low-best}^{i} = \begin{cases} A_{best}^{i} & if\ rand(i) = 1 \\ A_{lowest}^{i} & Else \end{cases}$$
**End for**

// Update the optimum solution of the whole swarm ($S_{Global}$).

**For every $A_i \in A$**
  $$S_{Global} = S_G = \begin{cases} A_{best} & if(\ Fit(A_{best}) > Fit(A_{low-best})) \\ A_{low-best} & Else \end{cases}$$
**End for**

| Algorithm Parameters | |
|---|---|
| K | No. of ants in swarm "swarm size |
| A | Group of ants in swarm; A=$A_1$,….$A_K$. |
| TRD | Training dataset contents of training items M and its features FS; TRD= (M, FS). |
| TED | Testing dataset contents of testing items H and its features FS; TED= (H, FS). |
| n | No. of features in training and testing dataset; n=\|FS\|. |
| t | Number of iterations. |
| $\tau_o$ | Initial pheromone level for each node state. |
| ρ | Evaporation rate; ρ∈ [0,1]. |
| α | The relative importance of the pheromone value; α>0. |
| β | The relative importance of the heuristic information; β>0. |
| rand | Uniformly distributed random number; rand ∈ [0,1]. |
| O | The optimum ant of the whole swarm ($S_{Global}$). |
| A | Group of ants in swarm; A=$A_1$, $A_2$,…., $A_K$. |
| $A_i$ | The $i^{th}$ ant in the swarm. |
| Accuracy($A_i$) | The classification accuracy according to the selected features in $i^{th}$ ant. |
| Fit($A_i$) | The fitness value of $A_i$ ant. |
| $P_{i,j}^{A}$ | The probability of movement from the bit "*i*" to the state "*j*" at time *t*, $j \in [0,1]$. |
| $\tau_{i,j}(0)$ | The pheromone density of the side which the "*j*"is "0" |
| $\tau_{i,j}(1)$ | The pheromone density of the side which the "*j*"is "1" |
| $\eta_{i,j}(0)$ | The visibility density of the side which the "*j*"is"0" |
| $\eta_{i,j}(1)$ | The visibility density of the side which the "*j*"is"1" |
| $A_{best}$ | The best solution in the swarm |
| $S_{Global}$ | The optimum solution of the whole swarm; $S_G$. |
| $A_{lowest}$ | The worst solution in the swarm. |
| $A_{low\text{-}best}$ | The updated worst solution |
| $\Delta\tau_{i,j}^{best}$ | The incremental pheromone; $\Delta\tau_{i,j}^{best} = \frac{1}{f(S^{best})}, f(S^{best})$; *best fitness value* |
| $\tau_{i,j}(t+1)$ | The intensity at time *(t + 1)* |

**For the best ant in whole swarm; $A_{best} \in A$**
  $$\tau_{i,j}(0)(t+1) = (1-\rho)\tau_{i,j}(0)(t) + \Delta\tau_{i,j}^{best}$$
  $$\tau_{i,j}(1)(t+1) = (1-\rho)\tau_{i,j}(1)(t) + \Delta\tau_{i,j}^{best}$$
**End for**

**if (there are more generations to process) then**
  *Go to step 1.*
**Else**
  *Return $S_{Global}$ in O, where all ones bits in this path represents the best selected features.*
**End if**
Validated the selected features using NB classifier.

Algorithm 1: Efficient Ant Colony Optimization (EACO).

### 3.1.3. Neutrosophic Stage

In 1980, F. Smarandache introduced the neutrosophy theory, which is a branch of philosophy [48,49]. Neutrosophic technique (NT) uses neutrosophic sets and the principles of neutrosophic logic. Neutrosophic set (NS) is a generalization of the fuzzy set theory, intuitionistic fuzzy set, paraconsistent set, dialetheist set,

paradox set, and tautological set. NS is an enhanced mathematical model that is applied to solve the problem of uncertain and ambiguous data. It includes a simple neutrosophic rule-based method in the form of; *IF A* AND *B* THEN *Y* to solve the problem rather than trying to design a system that is mathematically similar to the fuzzy classifier [48-50]. Thus, NT is similar to fuzzy technique. By using fuzzy system, the implementation of neutrosophic inference system was based on Mamdani's fuzzy inference method [51]. Figure 6, shows the block diagram of neutrosophic system using fuzzy logic toolbox in MATLAB. As shown in figure 6, the Fuzzy Inference System (FIS) was designed through three components. These components are; Neutrosophic truth component, Neutrosophic indeterminacy component, and Neutrosophic falsehood component.

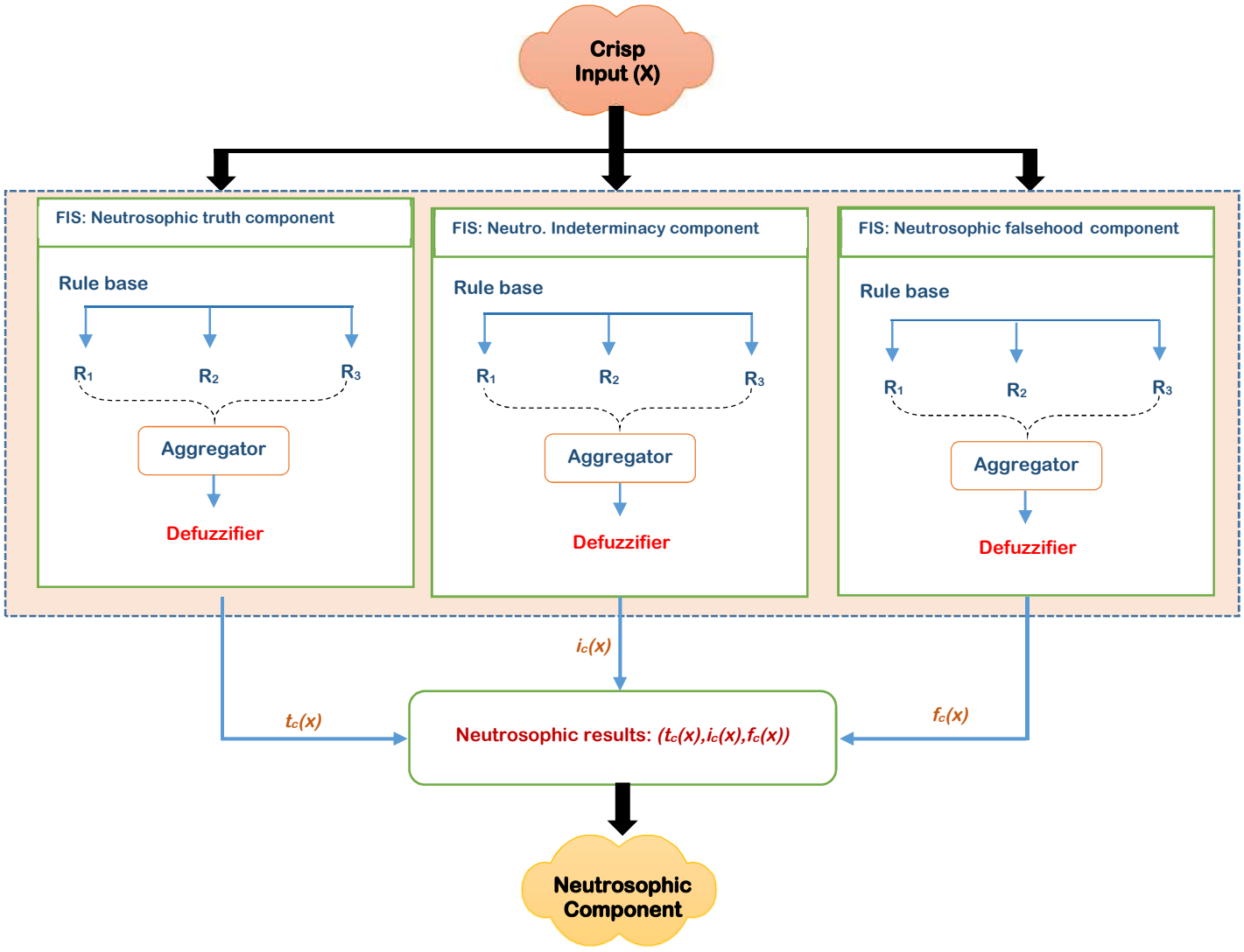


Figure 6.Block diagram for a Neutrosophic components

Suppose that Y be a universe of discourse and *A* be a set included in *Y*, which consists of bright pixels. The image is transformed into neutrosophic domain by representing it using three distinctive membership components (*T*, *I*, and *F*) where *T* represents the truth scale, *F* indicates the scale of false, and *I* defines the scale of intermediate. All of these components are independent of one another and their values are between 0

and 1. According to this definition, the neutrosophic image pixels set $P_{NS}$ *are* represented by these components *T, I, F*. Hence, the image pixel *P* in neutrosophic domain is defined as $P_{NS}$*( T, I, F)* and belongs to *A* as follows; *t%* of bright pixel set is true membership function, *i%* of bright pixel set is indeterminacy membership function, and *f%* is a falsity-membership function in the set, where *t, i,* and *f* are varies in *T*, *I*, and *F* respectively. Then, to convert any image from the classical domain into neutrosophic domain, each pixel in the image is represented by three membership components (e.g. T, I, F). Consequently, the pixel *P(m,n)* can be converted into neutrosophic domain ($P_{ND}$) using the following equation[52]:

$$P_{ND}(m,n) = \{T(m,n), I(m,n), F(m,n)\} \qquad (9)$$

Where $P_{ND}(m,n)$is the image pixels in neutrosophic domain, $T(m,n)$ is the probability that the pixel *P(m,n)* belongs to white object, $I(m,n)$ is the probability that the pixel *P(m,n)* belongs to indeterminate set, and $F(m,n)$ is the probability that the pixel *P(m,n)* belongs to background (non-white group). The probability that the pixel *P(m,n)* belongs to white object $T(m,n)$ can be determined by using (10) [53]

$$T(m,n) = \frac{\bar{g}_{(m,n)} - \bar{g}_{min}}{\bar{g}_{max} - \bar{g}_{min}} \qquad (10)$$

Where $\bar{g}_{(m,n)}$is the local mean window size pixels. The probability that the pixel *P(m,n)* belongs to indeterminate set $I(m,n)$ can be calculated using (11)[ 53]:

$$I(m,n) = 1 - \frac{H_o(m,n) - H_o}{H_{o_{max}} - H_{o_{min}}} \qquad (11)$$

Where $H_o(m,n)$ is the homogeneity value of *T* at *(m,n)* that can be determined by eq. (12) [53]:

$$H_o(m,n) = \left| g_{(m,n)} - \bar{g}_{(m,n)} \right| \qquad (12)$$

Finally, the probability that the pixel *P(m,n)* belongs to background (non-white group) $F(m,n)$ can be calculated using eq.(13)[ 53]:

$$F(m,n) = 1 - T(m,n) \qquad (13)$$

In our proposed model, after extracting GLCM features and selecting the most significant features, these selected features are transformed into neutrosophic components as shown in figure 7. Hence, each feature is represented by three neutrosophic sets, *T, I, F*. Thus, we can make a better choice and rank all the alternatives according to those three functions. As mentioned before, the selected features are homogeneity, contrast, correlation, entropy, energy, cluster prominence, and shade. Those features are transformed from classical domain into neutrosophic domain as follows:

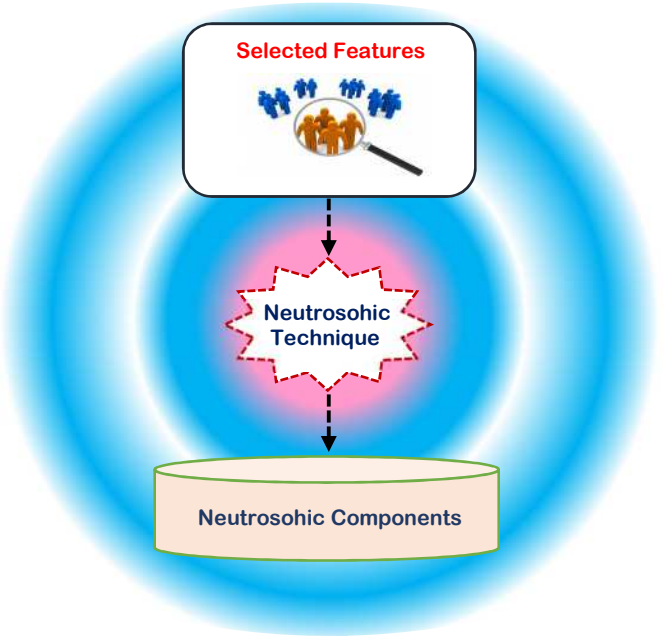


Figure 7, Transforming GLCM features into neutrosophic domain.

### A) Neutrosophic Homogeneity Feature

Image homogeneity measures the similarity of the elements distribution in the gray level co-occurrence matrix which can be calculated by using (14):

$$Homogenity = \sum_{m=0}^{K-1}\sum_{n=0}^{K-1}\frac{1}{1+(m-n)^2}.P(m,n) \qquad (14)$$

Where $P(m,n)$ refers to element *m, n* of GLCM, K is the number of image gray levels. *m, n* donate the number of rows and columns. To convert the homogeneity image into neutrosophic domain, it is represented by three neutrosophic sets, *T, I, F* which are determined using the following equations:

$$HOM_{(T)} = \sum_{m=0}^{K-1}\sum_{n=0}^{K-1}\frac{1}{1+(m-n)^2}.P_T(m,n) \qquad (15)$$

$$HOM_{(I)} = \sum_{m=0}^{K-1}\sum_{n=0}^{K-1}\frac{1}{1+(m-n)^2}.P_I(m,n) \qquad (16)$$

$$HOM_{(F)} = \sum_{m=0}^{K-1}\sum_{n=0}^{K-1}\frac{1}{1+(m-n)^2}.P_F(m,n) \qquad (17)$$

Where $HOM_{(T)}$ is the neutrosophic homogeneity truth component, $HOM_{(I)}$ is the neutrosophic homogeneity indeterminacy component, and $HOM_{(F)}$ is the neutrosophic homogeneity falsehood component. The same is applied for the rest of the selected features. Table 4 shows the transformation of some of the selected features from classical domain into neutrosophic domain.

Table 4: shows the transformation of contrast, correlation, entropy, and energy into neutrosophic components.

| Features | Description | Classical domain | Neutrosophic Domain |
|---|---|---|---|
| Contrast (CON) | CON calculates the intensity difference between a pixel and its neighbors. | $CON = \sum_{m=0}^{K-1} k^2 \sum_{m=0}^{K-1} P(m,n)$<br>Where :<br>$k = \lvert m-n \rvert$ | ▪ $CON_{(T)} = \sum_{m=0}^{K-1} k^2 \sum_{m=0}^{K-1} P_T(m,n)$<br>▪ $CON_{(I)} = \sum_{m=0}^{K-1} k^2 \sum_{m=0}^{K-1} P_I(m,n)$<br>▪ $CON_{(F)} = \sum_{m=0}^{K-1} k^2 \sum_{m=0}^{K-1} P_F(m,n)$ |
| Correlation (COR) | COR computes the joint probability occurrence of the specified pixel pairs. | $COR = \sum_{m=0}^{K-1} \sum_{n=0}^{K-1} \frac{(m-\mu)(n-\mu)}{\sigma^2}.P(m,n)$<br>Where:<br>$\mu = \sum_{m,n=0}^{K-1} m.P(m.n)$<br>$\boldsymbol{\sigma} = \sum_{m,n=0}^{K-1} (i-\mu)^2.P(m,n)$ | ▪ $COR_{(T)} = \sum_{m=0}^{K-1} \sum_{n=0}^{K-1} \frac{(m-\mu)(n-\mu)}{\sigma^2}.P_{(T)}(m,n)$<br>▪ $COR_{(I)} = \sum_{m=0}^{K-1} \sum_{n=0}^{K-1} \frac{(m-\mu)(n-\mu)}{\sigma^2}.P_{(I)}(m,n)$<br>▪ $COR_{(F)} = \sum_{m=0}^{K-1} \sum_{n=0}^{K-1} \frac{(m-\mu)(n-\mu)}{\sigma^2}.P_{(F)}(m,n)$ |
| Entropy (ENT) | ENT is a measure of randomness that defines the amount of information. | $ENT = \sum_{m=0}^{K-1} \sum_{n=0}^{K-1} -\ln(P(m,n).P(m,n)$ | ▪ $ENT_{(T)} = \sum_{m=0}^{K-1} \sum_{n=0}^{K-1} -\ln(P_{(T)}(m,n).P_{(T)}(m,n)$<br>▪ $ENT_{(I)} = \sum_{m=0}^{K-1} \sum_{n=0}^{K-1} -\ln(P_{(I)}(m,n).P_{(I)}(m,n)$<br>▪ $ENT_{(F)} = \sum_{m=0}^{K-1} \sum_{n=0}^{K-1} -\ln(P_{(F)}(m,n).P_{(F)}(m,n)$ |
| Energy (ENG) | ENG provides the sum of squared elements in the GLCM. Also known as uniformity or the angular second moment. | $ENG = \sum_{m=0}^{K-1} \sum_{n=0}^{K-1} (P(m.n))^2$ | ▪ $ENG_{(T)} = \sum_{m=0}^{K-1} \sum_{n=0}^{K-1} (P_{(T)}(m.n))^2$<br>▪ $ENG_{(I)} = \sum_{m=0}^{K-1} \sum_{n=0}^{K-1} (P_{(I)}(m.n))^2$<br>▪ $ENG_{(F)} = \sum_{m=0}^{K-1} \sum_{n=0}^{K-1} (P_{(F)}(m.n))^2$ |

### 3.2.Breast Cancer Classification Phase (BC$^2$P)

The second and final phase in the proposed strategy is Breast Cancer Classification Phase (BC$^2$P) in which the final decision is taken to classify the input case as a normal case or infected with benign or malignant breast cancer. Hence, the main objective of the classification process is to determine which category the new case will be included in. The classification process consists of two phases; training phase and testing phase. In the training phase, the classifier is learned on a set of the most efficient and robust membership component from the three components (e.g. T, I, F) that are extracted from a set of selected features which are extracted from the breast cancer image. In the testing phases, the learned classifier is used to discriminate between data that was not necessarily encountered during the training phase. In this paper, due to its performance and efficiency, Deep Neural Network (DNN) was used to perform the classification process.

DNN is an artificial neural network with multiple hidden layers between the input layer and the output layer. It is a powerful machine learning method used for various applications [54]. DNN is a forward neural network where the input flows along with weights and bias directly from the input layer to the output layer through a number of hidden layers. The basic DNN parameters are; the number of nodes in input and output layers, bias, learning rate, initial weights, number of hidden layers, number of nodes in every hidden layer, and the termination conditions. Actually, the number of hidden layers should be more than two layers and they are

activated using different activation functions such as Relu, Sigmoid, etc. The number of neurons in the output layer is proportional to the number of target classes. Thus, for binary class, only one neuron is used in the output layer, so the number of neurons is equal to the number of target classes [54]. Figure 8, shows the architecture of DNN model.

In this paper, the most robust component of the three neutrosophic components that are extracted from the selected features is used to train DNN classifier for breast cancer classification. DNN classifier has one input layer with *n* input nodes. Additionally, three hidden layers are used with 10, 20, and 10 nodes, and an output layer with three nodes. Although, using more hidden layers makes the model more complex, it provides the best results. In the DNN model, the bias value is set to be 1 and the learning rate is 0.15, which is the default value. Additionally, the weights of the nodes can be randomly generated. Actually, after every epoch, the weights are updated by the network during back propagation by calculating the error rate. Also, the termination condition is met by either reaching the number of epochs or achieving the result expected from the learning model. In fact, the output is predicted as a normal patient, or a patient infected with benign or malignant breast cancer, so softmax activation function is the best choice to train the classification model. Since it is a multi-classification process, softmax is the right choice.

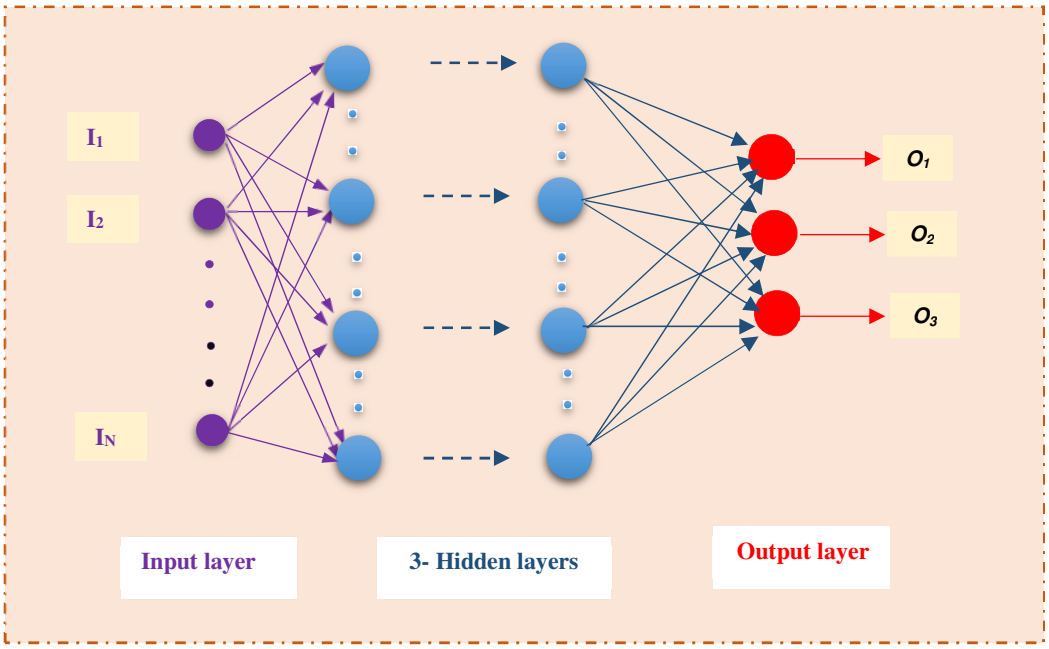


Figure 8, The architecture of DNN model.

## 4. Experimental results

In this section, the proposed Breast Cancer Classification Strategy ($BC^2S$) will be evaluated. It is composed of two phases, which are, (i) Data Preprocessing Phase ($DP^2$), and (ii) Breast Cancer Classification Phase (BC2P). In $DP^2$, the features are extracted from mammogram images using GLCM and the outlier items are rejected. Then, the most informative and significant features are selected from those extracted features using the proposed EACO method. Finally, the selected features are transformed into neutrosophic domain to

provide the best classification. During BC$^2$P, the most efficient and robust component of the extracted three components is used to train DNN model to make the correct decision. Our implementation is based on MIAS mammogram dataset [55,56]. MIAS is an internet dataset, which has been used to reproduce the results introduced in this paper. MIAS dataset is a public data that has been generated on Kaggle. It contains 322 cases, of which 209 are normal cases, 61 have benign breast cancer, and 52 have malignant breast cancer. It has been divided into two categories, namely; training set and testing set. The training set is used to train the model. While the testing set is used for measuring the performance of the proposed model. Hence, the number of training and testing patients are 226 (70%) and 96 (30%) respectively as illustrated in table 5. Table 6 shows the applied parameters with the corresponding used values.

Table 5: MIAS database image selection for training and testing.

| | Normal | Benign | Malignant | Total |
|---|---|---|---|---|
| Training | 140 | 46 | 40 | 226 |
| Testing | 69 | 15 | 12 | 96 |
| Total | 209 | 61 | 52 | 322 |

Table 6: The parameters applied with the corresponding used values.

| Parameter | Description | Applied value |
|---|---|---|
| n | No. of extracted features | 22 |
| $\tau_o$ | Initial Pheromone for each state | 0.2 |
| ρ | Evaporation rate | 0.2 |
| α | The relative importance of the pheromone value | 2 |
| β | The relative importance of the heuristic information | 1 |

### 4.1.Evaluation Matrices

During the next experiments, accuracy, error, precision, and recall are calculated. Then, additional criteria will be measured to clear the application result, which are; F-measure, macro-average, and micro-average. Confusion matrix is applied to calculate the values of these measurements as introduced in table 7. Noticeably, various formulas are used as a summarization of the confusion matrix as depicted in table 8.

Table 7: Confusion matrix.

| | | Predicted Label | |
|---|---|---|---|
| | | **Positive** | **Negative** |
| **Known Label** | **Positive** | True Positive (TP) | False Negative (FN) |
| | **Negative** | False Positive (FP) | True Negative (TN) |

Table 8: Confusion matrix formulas.

| Measure | Formula | Intuitive Meaning |
|---|---|---|
| Precision (P) | TP / (TP + FP) | The percentage of positive predictions those are correct. |
| Recall / Sensitivity(R) | TP / (TP + FN) | The percentage of positive labeled instances that were predicted as positive. |
| Accuracy(A) | (TP + TN) / (TP + TN + FP + FN) | The percentage of predictions those are correct. |
| Error(E) | 1-Accuracy | The percentage of predictions those are incorrect. |
| Macro-average | $\sum_{i=1}^{c} P_i / c$ "for Precision" | The average of the precision and recall of the system on different c classes. |
| | $\sum_{i=1}^{c} R_i / c$ "for Recall" | |
| Micro-average | (TP1 + TP2) / (TP1 + TP2 + FP1 + FP2) "for precision" | the summation up to the individual true positives, false positives, and false negatives of the system for different classes and the apply them to get the statistics |
| | (TP1 + TP2) / (TP1 + TP2 + FN1 + FN2) "for Recall" | |
| F-measure | 2*PR/(P+R) | The weighted harmonic mean of Precision and Recall |

### 4.2.Testing the proposed Breast Cancer Classification Strategy (BC$^2$S)

In this subsection, it is the time to test the proposed BC$^2$S. To prove the effectiveness of our proposed strategy, it is compared against some of the recently used classification methods as presented in table 2. Those recent

methods are FE-BKCapsNet [21], DLF [22], ResNet-50-SARF [24], PBC-CNN [25], and RFRE [26]. All capabilities proposed are used in our BC$^2$S, hence, EACO is used for feature selection, and the proposed DNN model is employed for classification. The experimental results are presented in figures (9→17). As shown in figures (9→12), our strategy achieves the highest values of accuracy, precision, and recall. On the other hand, it achieves a minimum error rate as compared to other competitors. Moreover, it introduces the highest values of macro-average precision, macro-average recall, micro-average precision, micro-average recall, and F-measure as illustrated in figures (13→17). This proves the effectiveness of the proposed strategy BC$^2$S for classifying breast cancer patients.

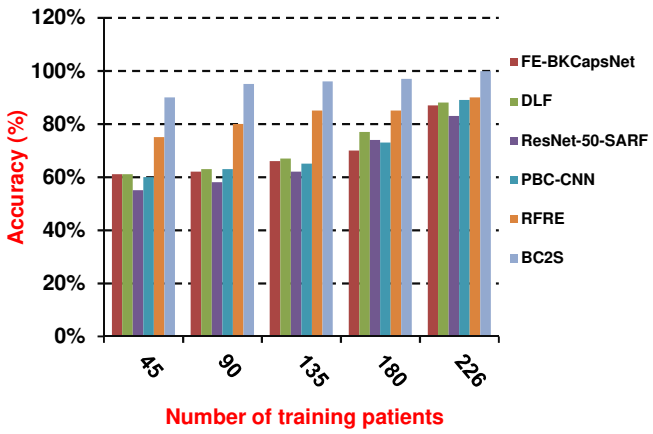


Figure 9, Accuracy of the different classification techniques.

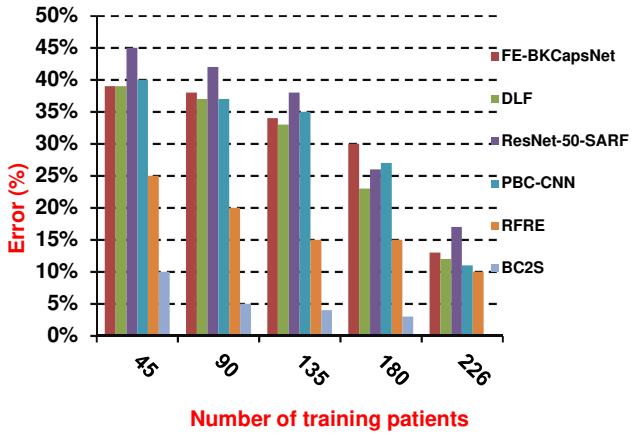


Figure 10, Error of the different classification techniques.

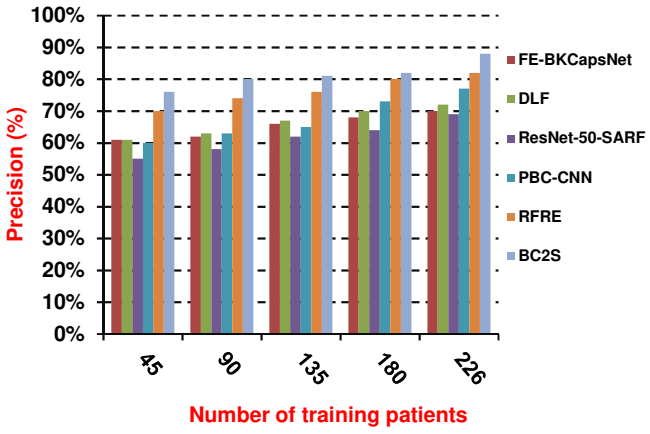


Figure 11, Precision of the different classification techniques

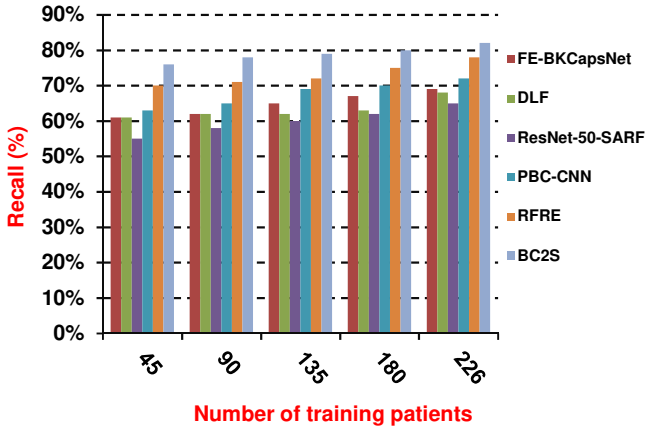


Figure 12, Recall of the different classification techniques

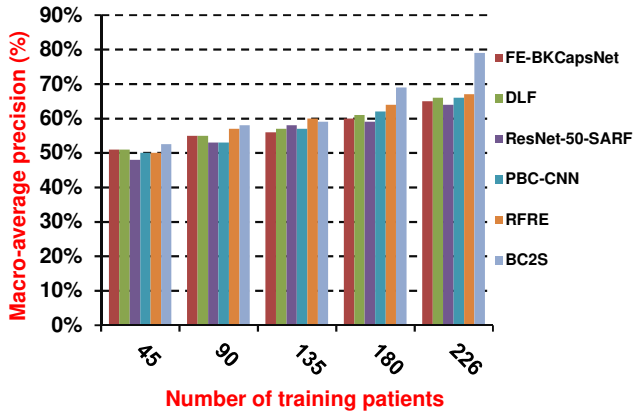


Figure 13, Macro-average precision of the different classification techniques.

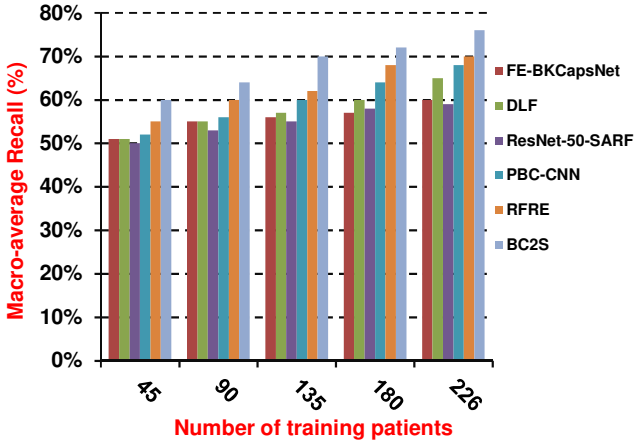


Figure 14, Macro-average recall of the different classification techniques.

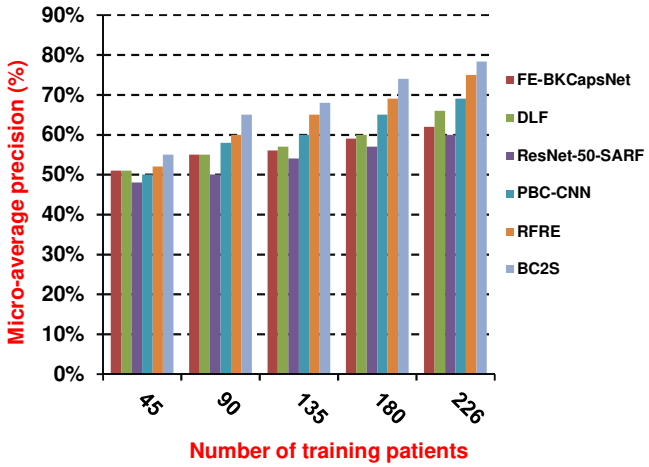


Figure 15, Micro-average precision of the different classification techniques.

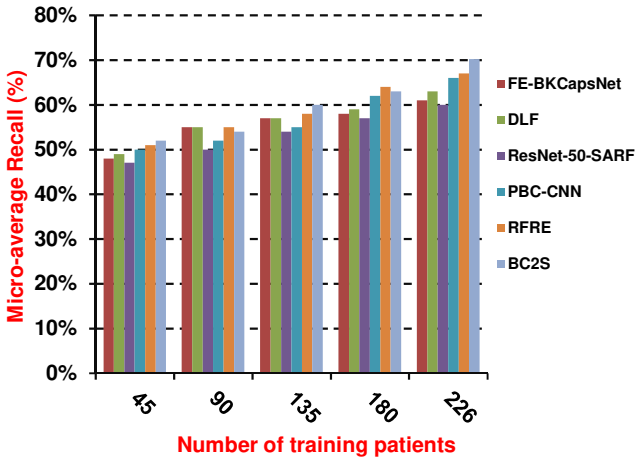


Figure 16, Micro-average recall of the different classification techniques

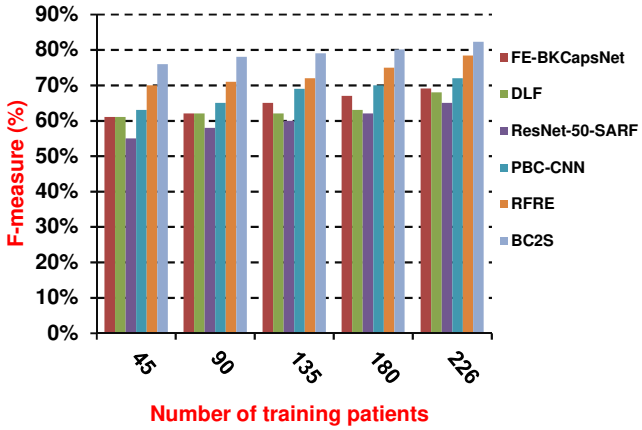


Figure 17, F-measure of the different classification techniques.

As shown in figures (9→17), it is concluded that the performance of all methods is enhanced by increasing the number of training patients. Figure 9, presents the accuracy for a different number of training patients. When training patients=226, the accuracies are 0.87, 0.88, 0.83, 0.89, 0.90, and 0.999 for FE-BKCapsNet, DLF, ResNet-50-SARF, PBC-CNN, RFRE, and $BC^2S$ respectively. While the error values are 0.13, 0.12, 0.17, 0.11, 0.1, and 0.01 respectively, as shown in figure 10. According to figure 9 and figure 10, the proposed $BC^2S$ provides the best performance. The reason is that the proposed $BC^2S$ gives accurate classification for infected patients based on the most efficient and robust neutrosophic component, namely, membership component that are extracted from the most effective and informative features which are selected using the proposed EACO method. Hence, the proposed strategy could help to classify breast cancer patients with a minimum error rate. Results in figure 11 and figure 12, when training patients=226, the precision values are 0.7, 0.72, 0.69, 0.77, 0.82, and 0.88 for FE-BKCapsNet, DLF, ResNet-50-SARF, PBC-CNN, RFRE, and $BC^2S$ respectively. Additionally, when training patients= 226, the recall values are 0.69, 0.68, 0.65, 0.72, 0.78, and 0.82 for FE-BKCapsNet, DLF, ResNet-5-SARF, PBC-CNN, RFRE, and $BC^2S$ respectively.

The results in figures (13→16) show that $BC^2S$ is better than for FE-BKCapsNet, DLF, ResNet-50-SARF, PBC-CNN, and RFRE. When training patients=226, $BC^2S$ provides the highest macro-average precision with a value equal to 0.795, while ResNet-50-SARF introduces the lowest value. Additionally, $BC^2S$ provides the highest macro-average recall with value of 0.76, while it is 0.59 for ResNet-50-SARF. When training patients =226, the micro-average precision values are 0.62, 0.66, 0.6, 0.69, 0.75, and 0.78 for FE-BKCapsNet, DLF, ResNet-50-SARF, PBC-CNN, RFRE, and $BC^2S$ respectively. At the same order, the micro-average recall values of FE-BKCapsNet, DLF, ResNet-50-SARF, PBC-CNN, RFRE, and $BC^2S$ are 0.61, 0.63, 0.6, 0.66, 0.67, and 0.702. Moreover, the proposed $BC^2S$ provides the highest F-measure with a value equal to 0.82, while the lowest value reached to 0.65 provided by ResNet-50-SARF as shown in figure 17. Hence, the results of the proposed strategy shown in figures (9→17), prove the effectiveness of the proposed strategy that outperforms other competitors in terms of accuracy, precision, recall, and F-measure. $BC^2S$ is more simple, robust, efficient, and flexible for classifying breast cancer patients.

## 5. Conclusions

Nowadays, breast cancer is one of the most common diseases affecting women. It is the second cause of death. According to World Health Organization (WHO) and American Society Center (ASC), early detection of breast cancer is a vital process that provides a great opportunity to make the right decision to select the suitable treatment protocol, consequently, saving many lives. In this paper, a new strategy for detecting and classifying breast cancer was introduced. The proposed strategy called Breast Cancer Classification Strategy ($BC^2S$) which is composed of two phases, which are; Data Preprocessing Phase ($DP^2$) and Breast Cancer Classification Phase ($BC^2P$). In $DP^2$, the features are extracted from mammogram images, then, rejected the hasted data and selected the most informative features using Efficient Ant Colony Optimization (EACO) method. EACO is an improved method that combines BACO and crossover strategy. Although the EACO method accurately identifies the most useful features, it suffers from time complexity. But the process of selecting the most important features is only done once. Finally, the selected features are transformed into neutrosophic domain for more accurate classification. During $BC^2P$, the most efficient neutrosophic component is used to feed the

DNN model to classify patients into normal patients, patients infected with benign breast cancer, or patients infected with malignant breast cancer. Experimental results prove the effectiveness of the proposed strategy as it outperforms other competitors in terms of accuracy, error, precision, recall, and F-measure.

## Compliance with ethical standards

## Conflict of interest

The authors declare that they have no conflict of interest.

## Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

## 6. References


[1] American Cancer Society, https://www.cancer.org/.

[2] B. Moloney, D. O'Loughlin, S. Abd Elwahab, et. al., " *Breast Cancer Detection—A Synopsis of Conventional Modalities and the Potential Role of Microwave Imaging,* " diagnostics, Multidisciplinary Digital Publishing Institute (MDPI), Volume 10, Issue 2, https://doi.org/10.3390/diagnostics10020103, 2020, PP. 1-13.

[3] G. Meenalochini, and S. Ramkumar, "*Survey of machine learning algorithms for breast cancer detection using mammogram images,*" Materials Today: Proceedings, Elsevier, Volume 37, https://doi.org/10.1016/j.matpr.2020.08.543, 20210, PP. 2738-2743.

[4] S. Heller, and L. Moy, " *Breast MRI Screening: Benefits and Limitations*" Current Breast Cancer Reports, Springer, Volume 8, https://doi.org/10.1007/s12609-016-0230-7, 2016, PP. 248-257.

[5] A. Emami, H. Ghadiri, P. Ghafarian, et. al., "Performance *evaluation of developed dedicated breast PET scanner and improvement of the spatial resolution by wobbling: a Monte Carlo study,"* Japanese Journal of Radiology, Springer, Volume 38, https://doi.org/10.1007/s11604-020-00966-w, 2020, PP. 790–799.

[6] D. Singh, and A. Singh, "*Role of image thermography in early breast cancer detection- Past, present and future*, " Computer Methods and Programs in Biomedicine, Elsevier, Volume 183, https://doi.org/10.1016/j.cmpb.2019.105074, 2020, PP. 1-9.

[7] R. Guo, G. Lu, B. Qin, et. al., "*Ultrasound Imaging Technologies for Breast Cancer Detection and Management: A Review,*" Ultrasound in Medicine & Biology, Elsevier, Volume 44, Issue 1, 2018, PP. 37-70.

[8] S. Ramadan, "*Methods Used in Computer-Aided Diagnosis for Breast Cancer Detection Using Mammograms: A Review*," Journal of healthcare engineering, Hindawi, Volume 2020, https://doi.org/10.1155/2020/9162464, 2020, PP. 1-21.

[9] Z. Ahmed, Kh. Mohamed, S. Zeeshan, et. al., "*Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine,*" Database, Volume 2020, https://doi.org/10.1093/database/baaa010, 2020, PP. 1-35.

[10] N. Mansour, A. Saleh, M. Badawy, et. al., "*Accurate detection of Covid-19 patients based on Feature Correlated Naïve Bayes (FCNB) classification strategy*," Journal of Ambient Intelligence and Humanized Computing, Springer, http://doi.org/10.1007/s12652-020-02883-2, 2020, PP. 1-34.

[11] Z. Mushtaq, A. Yaqub, Sh. Sani, et. al., "*Effective K-nearest neighbor classifications for Wisconsin breast cancer data sets*," Journal of the Chinese Institute of Engineers, Volume 43, Issue 1, 2020, PP. 1-14

[12] N. Zakaria, Z. Shah, and S. Kasim, "*Protein Structure Prediction Using Robust Principal Component Analysis and Support Vector Machine*", International Journal of Data Science, Volume 1, Issue 1, 2020, PP. 14-17.

[13] R. Vijayarajeswari, P. Parthasarathy, S. Vivekanandan, et. al., "Classification *of mammogram for early detection of breast cancer using SVM classifier and Hough transform*," Measurement, Elsevier, Volume 146, https://doi.org/10.1016/j.measurement.2019.05.083, 2019, PP. 800-805.

[14] A. Assiri, S. Nazir, and S. Velastin, "*Breast Tumor Classification Using an Ensemble Machine Learning Method*," Journal of Imaging, Multidisciplinary Digital Publishing Institute (MDPI), Volume 6, Issue 6, 2020, PP. 1-13

[15] J. Quist, L. Taylor, J. Staaf, et. al., "*Random Forest Modelling of High-Dimensional Mixed-Type Data for Breast Cancer Classification*," Cancers, Multidisciplinary Digital Publishing Institute (MDPI), Volume 13, Issue 5, 2021, PP. 1-15.

[16] S. Das, B. Roy, M. Kar, et. al., "*Neutrosophic fuzzy set and its application in decision making,*" Journal of Ambient Intelligence and Humanized Computing, Springer, Volume 11, https://doi.org/10.1007/s12652-020-01808-3, 2020, PP. 5017–5029.

[17] R. Tan, and W. Zhang, "*Decision-making method based on new entropy and refined single-valued neutrosophic sets and its application in typhoon disaster assessment,"* Applied Intelligence, Springer, Volume 51, https://doi.org/10.1007/s10489-020-01706-3, 2021, PP. 283–307.

[18] G. Nguyen, L. Son, A. Ashour, et. al., "*A survey of the state-of-the-arts on neutrosophic sets in biomedical diagnoses,*" International Journal of Machine Learning and Cybernetics, Springer, Volume 10, https://doi.org/10.1007/s13042-017-0691-7, 2019, PP. 1-13.

[19] J. Chai, G. Selvachandran, F. Smarandache, et. al., "*New similarity measures for single-valued neutrosophic sets with applications in pattern recognition and medical diagnosis problems*," Complex & Intelligent Systems, Springer, Volume 7, https://doi.org/10.1007/s40747-020-00220-w, 2021, PP. 703-723.

[20] M. Aslam, O. Arif, and R. Sherwani, et. al., "*New Diagnosis Test under the Neutrosophic Statistics: An Application to Diabetic Patients,"* BioMed Research International, Hindawi, Volume 2020, https://doi.org/10.1155/2020/2086185, 2020, PP. 1-7.

[21] P. Wang, J. Wang, Y. Li, et. al., "*Automatic classification of breast cancer histopathological images based on deep feature fusion and enhanced routing*," Biomedical Signal Processing and Control, Elsevier, Volume 65, https://doi.org/10.1016/j.bspc.2020.10234, 2021, PP. 1-8.

[22] S. Khan, N. Islam, Z. Jan, et. al., " *A novel deep learning based framework for the detection and classification of breast cancer using transfer learning*," Pattern Recognition Letters, Elsevier, Volume 125, https://doi.org/10.1016/j.patrec.2019.03.022, 2019, PP. 1-9.

[23] G. Altan, "*Deep Learning-based Mammogram Classification for Breast Cancer,* " International Journal of Intelligent Systems and Applications in Engineering (IJISAE), Volume 8, Issue 4, 2020, PP. 171-176.

[24] B. Yousefi, H. Akbari, and X. Maldague, "*Detecting Vasodilation as Potential Diagnostic Biomarker in Breast Cancer Using Deep Learning-Driven Thermomics,*" Biosensors, Multidisciplinary Digital Publishing Institute (MDPI), Volume 10, Issue 11, 2020, PP. 1-18.

[25] K. Roy, D. Banik, D. Bhattacharjee, et. al., " *Patch-based system for Classification of Breast Histology images using deep learning,* " Computerized Medical Imaging and Graphics, Elsevier, Volume 71, https://doi.org/10.1016/j.compmedimag.2018.11.003, 2019, PP. 90-103.

[26] S. Wang, Y. Wang, D. Wang, et. al., "*An improved random forest-based rule extraction method for breast cancer diagnosis*," Applied Soft Computing, Elsevier, Volume 86, https://doi.org/10.1016/j.asoc.2019.105941, 2020, PP. 1-35.

[27] W. Shia, and D. Chen, "*Classification of malignant tumors in breast ultrasound using a pretrained deep residual network model and support vector machine*, " Computerized Medical Imaging and Graphics, Elsevier, Volume 87, https://doi.org/10.1016/j, 2021, PP. 1-7.

[28] L. Yang, and Z. Xu, " *Feature extraction by PCA and diagnosis of breast tumors using SVM with DE-based parameter tuning*, " International Journal of Machine Learning and Cybernetics, Springer, Volume 10, DOI 10.1007/s13042-017-0741-1, 2019, PP. 591–601.

[29] W. Shaban, A. Rabie, A. Saleh, et. al., "*Detecting COVID-19 patients based on fuzzy inference engine and Deep Neural Network*," Applied Soft Computing, Elsevier, Volume 99, https://doi.org/10.1016/j.asoc.2020.106906, 2021, PP. 1-19.

[30] W. Shaban, A. Rabie, A. Saleh, et. al., "A *new COVID-19 Patients Detection Strategy (CPDS) based on hybrid feature selection and enhanced KNN classifier*," Knowledge- Based Systems, Elsevier, Volume 205, https://doi.org/10.1016/j.knosys.2020.106270 , 2020, PP. 1-18.

[31] W. Shaban, A. Rabie, A. Saleh, et. al., "*Accurate detection of COVID-19 patients based on distance biased Naïve Bayes (DBNB) classification strategy*," Pattern Recognition, Elsevier, Volume 119, http://doi.org/10.1016/j.patcog.2021.108110 , 2021, PP. 1-15.

[32] K. Shukla, A. Tiwari, S. Sharma, et. al., "*Classification of histopathological images of breast cancerous and non cancerous cells based on morphological features,*" Biomedical Pharmacy Journal, Volume 10, Issue 1, 2017, PP. 353-366.

[33] Y. Xiao, J. Wu, Z. Lin, et. al., "*Breast Cancer Diagnosis Using an Unsupervised Feature Extraction Algorithm Based on Deep Learning*," *Proceeding in 37th Chinese Control Conference (CCC),* doi: 10.23919/ChiCC.2018.8483140, 2018, PP. 9428-9433.

[34] G.Kanagaraj, and P. Kumar, "*Pulmonary Tumor Detection by virtue of GLCM*," Journal of Scientific & Industrial Research, Volume 79, 2020, PP. 132-134.

[35] A. Zotin, Y. Hamad, K. Simonov, et. al., "*Lung boundary detection for chest X-ray images classification based on GLCM and probabilistic neural network*," Procedia Computer Science, Elsevier, Volume 159, https://doi.org/10.1016/j.procs.2019.09.314, 2019, PP. 1439-1448.

[36] K. Lee, and E. Johnson, "*Robust Outlier-Adaptive Filtering for Vision-Aided Inertial Navigation,"* Sensors, Multidisciplinary Digital Publishing Institute (MDPI), Volume 20, https://doi.org/10.3390/s20072036, 2020, PP. 1-24.

[37] A. Rabie, S. Ali, A. Saleh, et. al., "*A new outlier rejection methodology for supporting load forecasting in smart grids based on big data,"* Cluster Computing, Springer, Volume 23, https://doi.org/10.1007/s10586-019-02942-0, 2020, PP. 509–535.

[38] A. Rabie, S. Ali, A. Saleh, et. al., " *A fog based load forecasting strategy based on multi-ensemble classification for smart girds*," Journal of Ambient Intelligence and Humanized Computing, Springer, Volume 11, Issue 1, 2020, PP. 209-236.

[39] A. Rabie, S. Ali, A. Saleh, et. al., "*A new outlier rejection methodology for supporting load forecasting in smart grids based on big data*," Cluster Computing, Springer, Volume 23, https://doi.org/10.1007/s10586-019-02942-0, 2020, PP. 509–535.

[40] A. Duraj, and L. Chomatek, "*Outlier Detection Using the Multiobjective Genetic Algorithm,"* Journal of Applied Computer Science, Volume 25, Issue 2, 2017, PP. 29-42.

[41] G. Dhiman, D. Oliva, A. Kaur, et. al., "*BEPO: A novel binary emperor penguin optimizer for automatic feature selection,*" Knowledge- Based Systems, Elsevier, Volume 211, https://doi.org/10.1016/j.knosys.2020.106560, 2021, PP. 1-13.

[42] S. Shah, H. Shabbir, S. Rehman, et. al., "*A Comparative Study of Feature Selection Approaches: 2016-2020*," International Journal of Scientific & Engineering Research, Volume 11, Issue 2, 2020, PP. 469-478.

[43] B. Sathiyabhama, S. Kumar, J. Jayanthi, et. al., "*A novel feature selection framework based on grey wolf optimizer for mammogram image analysis*," Neural Computing and Applications, Springer, https://doi.org/10.1007/s00521-021-06099-z, 2021, PP. 1-20.

[44] M. Dorigo, M. Birattari, and T. Stutzle, "*Ant Colony Optimization: artificial ants as a computational intelligence technique*," IEEE Computational Intelligence, Volume 11, 2006, PP. 28-39.

[45] Z. Mnaban, F. Tab, and C. Salavati, "*Fast unsupervised feature selection based on the improved binary ant system and mutation strategy*," Neural Computing and Application, Springer, Volume 31, https://doi.org/10.1007/s00521-018-03991-z, 2019, PP. 1-20.

**[**46] M. Paniri, M. Dowlatshahi, and H. pour, "*Ant-TD: Ant colony optimization plus temporal difference reinforcement learning for multi-label feature selection,"* Swarm and Evolutionary Computation, Elsevier, Volume 64, https://doi.org/10.1016/j.swevo.2021.100892, 2021.

[47] Y. Wana, M. Wang, Z. Ye, et. al., "*A feature selection method based on modified binary coded ant colony Optimization Algorithm,"* Applied Soft Computing, Elsevier, Volume 49, https://doi.org/10.1016/j.asoc.2016.08.011, 2016, PP. 248-258.

[48] A. Abd El-Khalek, A. Khalil, M. Abou-elsoud, et. al., "*A Robust Machine Learning Algorithm for Cosmic Galaxy Images Classification Using Neutrosophic Score Features,"* Neutrosophic Sets and Systems, Volume 42, 2021, PP. 79-101.

[49] B. Amin, A. Salama, I. El-Henawy, et. al., "*Intelligent Neutrosophic Diagnostic System for Cardiotocography Data,"* Computational Intelligence and Neuroscience, Hindawi, https://doi.org/10.1155/2021/6656770, 2021, PP. 1-12.

[50] F. Fahmi, F. Apriyulida, I. Nasution, et. al., "A*utomatic Detection of Brain Tumor on Computed Tomography Images for Patients in the Intensive Care Unit,"* Journal of Healthcare Engineering, Hindawi, Volume 2020, https://doi.org/10.1155/2020/2483285, 2020, PP. 1-13.

[51] I. Yasser, A. Twakol, A. AbdEl-Khalek et. al., "*COVID-X: Novel Health-Fog Framework Based on Neutrosophic Classifier for Confrontation Covid-19*," Neutrosophic Sets and Systems, Volume 53, https://digitalrepository.unm.edu/nss_journal/vol35/iss1/1, 2020, PP. 1-21.

[52] M. Nasef, F. Eid, and A. Sauber, "*Skeletal scintigraphy image enhancement based neutrosophic sets and salp swarm algorithm*," Artificial Intelligence in Medicine, Elsevier, Volume 109, https://doi.org/10.1016/j.artmed.2020.101953, 2020, PP. 1-10

[53] S. Wady, R. Yousif, and H. Hasan, "*A Novel Intelligent System for Brain Tumor Diagnosis Based on a Composite Neutrosophic-Slantlet Transform Domain for Statistical Texture Feature Extraction*," BioMed Research International, Hindawi, Volume 2020, https://doi.org/10.1155/2020/8125392, 2020, PP. 1-21.

[54] S. M, P. Maddikunta, P. M., et. al., "*An effective feature engineering for DNN using hybrid PCA-GWO for intrusion detection in IoMT architecture,*" Computer Communications, Elsevier, Volume 160, https://doi.org/10.1016/j.comcom.2020.05.048, 2020, PP. 139–149.

[55] https://www.kaggle.com/kmader/mias-mammography?select=Info.txt

[56] J. Melekoodappattu, P. Subbian, and M. Queen, "*Detection and classification of breast cancer from digital mammograms using hybrid extreme learning machine classifier*," International Journal of Imaging Systems and Technology, Wiley Online Library, Volume 31, https://doi.org/10.1002/ima.22484, 2021, PP. 909-920.