Two way Threshold Based Intelligent Water Drops Feature Selection Algorithm for Accurate Detection of Breast Cancer

Dhruba Jyoti Kalita (mestop12@gmail.com)
Gaya College of Engineering, Gaya
Vibhav Prakash Singh
Motilal Nehru National Institute of Technology
Vinay Kumar
National Institute of Technology Jamshedpur

Research Article

Keywords: CAD System, Mammography, Texture features, SVM, Meta-heuristic Optimization

Posted Date: June 18th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-613900/v1

Version of Record: A version of this preprint was published at Soft Computing on November 23rd, 2021. See the published version at https://doi.org/10.1007/s00500-021-06498-3.

List of variables

- Variables related to feature extraction:

  g_i: The grey scale value of fneighbourhoood pixel.

  g_c: The grey scale value of the center pixel.

  P: Connectivity from the neighbourhood pixels.

  R: Neighbourhood radius for Nequally spaced pixels.

- Variables used in IWD algorithm:

  T^{IWD}: The complete solution.

  T^{IB} : Iteration best solution.

  $N_{Features}$: Number of final features.

  N_{IWD}: Number of Water Drops.

  (a_v, b_v, c_v): Variables to update the velocity of the water drops.

  (a_s, b_s, c_s): Variables to update the soil of the local path.

  $MaxIter$: Maximum number of iterations.

  $initSoil$: Initial value of the local soil.

  $initVel$: Initial velocity associated with each of the water drop.

  $V_c^{(IWD_r)}$: Feature list visited by each water drop r.

  $initVel^{(IWD_r)}$: Velocity of the water drop r.

  $soil^{(IWD_r)}$: Soil associated with the water drop r.

  ρ_n: Local soil updating parameter.

  ρ_{IWD}: Global soil updating parameter.

  ε_s: Parameter to prevent zero division.

- Variables related to thresholding:

  N_f: Total Number of features.

  T_{dist}: Threshold Distance.

Two-way Threshold based Intelligent Water Drops Feature Selection Algorithm for Accurate Detection of Breast Cancer

[a] Dhruba Jyoti Kalita, [b]Vibhav Prakash Singh, [c] Vinay Kumar

[a b c] Department of Computer Science & Engineering

[a] Gaya College of Engineering, Gaya, India

[b] Motilal Nehru National Institute of Technology, Allahabad, Prayagraj, India
[c] National Institute of Technology, Jamshedpur, India
[a] mestop12@gmail.com, [b] vibhav@mnnit.ac.in, [c] vkumar.cse@nitjsr.ac.in

Abstract: Breast cancer is one of the common reasons for deaths of women over the globe. It has been found that a Computer- Aided Diagnosis (CAD) system can be designed using X-ray mammograms for early-stage detection of breast cancer, which can decrease the death rate to a large extend. This paper work proposes a novel 2-way threshold based Intelligent water drops (IWD) algorithm for feature selection to design an effective and efficient CAD system that can detect breast cancer in early stage. This approach first extracts the Local Binary Patterns (LBP) in wavelet domain from mammograms and then apply our introduced 2-way threshold based (IWD) algorithm to extract most important subset of features from the extracted features set. 2-way thresholding is a technique to find a lower bound (LB) and an upper bound (UB) on the number of features to be selected in the optimal subset. So, using these threshold values IWD is capable of producing multiple optimal subsets of features rather than producing a single optimal subset of features. The best subset among the above subsets is then used train and deploy Support Vector Machine (SVM) to classify new mammograms. The results have shown that the proposed model outperforms many of the existing CAD systems. Further we have compared our introduced feature selection technique with other meta heuristic features selection techniques such as Ant Colony Optimization (ACO), Particle Swarm Optimization (PSO), Simulated Annealing (SA), Genetic Algorithm (GA), Gravitational Search Algorithm (GSA), Inclined Planes System Optimization (IPO) and Grey Wolf Optimization Algorithm (GWO) and found that it outperforms the others. The accuracy, precision, recall, specificity and F1-score of our proposed framework are measured as 99%, 98.7% ,98.123%, 96.2% and 98.4% respectively.

Keywords: CAD System, Mammography, Texture features, SVM, Meta-heuristic Optimization.

1. Introduction:

Breast cancer is one of the vital categories of cancer, which causes a huge number of deaths of women around the globe. According to “Breast Cancer Research Foundation (BCRF)”, around

2.3 million new cases of breast cancer were recorded in the year 2020 [1]. Between the year 2008 and 2012 the detected breast cancer cases increased by 20% as well as the mortality rate by 14% [1]. One of the main reasons for increase in these rates is unhealthy lifestyle in most of the urban and economically developed countries. With the involvement of the modern technologies, though we cannot reduce the increasing rate of breast cancer cases, but we can decrease the mortality rate. By detecting breast cancer in early stage, one can take the necessary actions to prevent its further growth. One of such detections techniques is CAD.

CAD systems [2] [3] [4] are automated systems which assist doctors to interpret the medical images. They can be treated as interdisciplinary technology which combines components of artificial intelligence and computer vision with radiological and pathology image processing. For the diagnosis of breast cancer, the CAD systems need to be trained with mammograms. Mammograms are the images generated by Mammography [5] [6]. It uses a low-dose X-ray system to look inside the internal tissues and parts of the breast. In this kind of system certain part of the body is exposed to some amount of ionizing radiation to get images of inside of the body. Later physicians can consult and check those images for further diagnosis.

Designing of a CAD system consists of various steps such as pre-processing, Segmentation, Feature Extraction, Feature Selection and Designing efficient classifier [7-20]. In the pre-processing phase removal of artefacts, noise etc. from the mammograms need to be done. Segmentation is basically the process of finding Region of Interests (ROI). The Region of Interest is the phase where all detected regions are analysed for special characteristics. In feature extraction, the features from the mammograms get extracted in the form of vectors. After feature extraction classifier modelling and validation of this classifier is done so that newly coming mammograms or test samples can be classified properly.

There have been various CAD systems proposed in recent years that have their own advantages and disadvantages. **Table 1** is depicting various CAD systems along with their limitations

Table 1: Various CAD systems

Authors/Years	**Proposed Methods/Objectives**	**Limitations**
Lia, Y., Chena, H., Yangb, Y., Yanga, N. (2013) [21]	Proposed a Kalman Filter based CAD system that uses homogeneous texture and high intensity deviation for the identification of the edge of the pectoral muscle.	In spite of a proper segmentation process, the accuracy rate achieved is very less (90% for mini-MIAS)
Anitha, J., Peter, J.D.(2012) [22]	proposed a CAD system that uses an automated morphological operation-based segmentation for finding the suspicious masses in the breast.	The authenticity of the proposed work has been proved only against GLCM based classification, which restricts it to be more generalized. On the other hand, the accuracy detection rate has achieved as 95% only with 44 mass mammograms.

Soulami, K. B., Saidi, M. N., & Tamtaoui, A. (2016) [23]	Proposed a CAD system that uses- i) SVM as a base classifier ii) Entropy Thresholding for pectoral muscle removal iii) PSO for ROI extraction iv) GLCM to extract shape and the texture features.	The heavy weighted processing in the segmentation and feature extraction phases makes the designing of the CAD system slow. Accuracy of classification is also very less (83.3 %)
S. Wang, R.V. Rao, P. Chen, Y. Zhang, A. Liu, L. Wei, (2017) [24]	Proposed a CAD system that uses-i) weighted type fractional Fourier Transform based feature extraction technique ii) PCA to reduce the size of the extracted feature vector iii) JAYA-FNN for detection.	Accuracy of the classification is very less (93 %), in spite using. On the other hand, they could have used some other advanced classifier such as SVM
A.M. Anter, A.E. Hassenian (2016) [25]	proposed a CAD system that uses – i) RG for segementation ii) GLCM for feature extraction iii) k-NN for classification	Accuracy of the classification is very less (94%). The proposed CAD system gives best performance only for k =1, in k-NN. The initial breast profile segmentation algorithm used here is not much effective.
N.F. Abubacker, A. Azman, S. Doraisamy, M.A.A. Murad (2017) [26]	Proposed an associative classifier-based fuzzy neural network integrated CAD system for mammogram classification	The classifier, presented in this system takes association rules as input, involve creation and training using Fuzzy Neural Network which makes in slow in the training process. On the other hand, the accuracy rate achieved is also 95%
Singh, V. P., Srivastava, S., & Srivastava, R. (2017) [25]	Proposed a CAD system that uses Random Forest as a base classifier and center symmetric –LBP (CS-LBP) features in wavelet domain.	Less Accuracy (97.3%)
Chandy, D. A., Christinal, A. H., Theodore, A. J., & Selvan, S. E. (2017) [28]	Proposed a Content based mammogram retrieval system that uses Neighbourhood feature selection method.	The overall efficiency and effectiveness of the CAD system has not been discussed. The results and analysis have been done only with respect to the feature selection method

Luo, S. T., & Cheng, B. W. (2012) [29]	Proposed a CAD system or predictive system for the diagnosis of breast masses using Forward Selection (FS) and Backward Selection (BS) feature selection methods	Even though FS and BS are well known feature selection techniques, there exists many other modern metaheuristic feature selection techniques that are much more efficient and effective. Authors have not provided any kind of comparisons between their proposed feature selection techniques and those metaheuristic feature selection techniques.
Mohanty, F., Rup, S., Dash, B., Majhi, B., & Swamy, M. N. S.(2019) [30]	Proposed a CAD system that uses a wrapper-based feature selection technique to find the optimal features.	Wrapper based feature selection is expensive in nature. There exists many other metaheuristic optimization techniquesc that could have been used for the purpose of feature selection
Chaieb, R., & Kalti, K. (2019) [31]	Proposed a CAD system that uses Gray-Level Run –Length Matrix features as the relevant features	The phase of feature extraction is slow in this proposed CAD system.

So, one can build an effective and efficient CAD system by altering techniques in either one or more phases (pre-processing, segmentation, feature extraction, feature selection and designing efficient classifier), while designing. From the literature, it has been found that in most of the works, main focus remains on the in pre-processing, segmentation, feature extraction and classification phases. Only few researchers have put emphasis on the feature selection in the post analysis phase or features selection has been applied in a very refinement level for mammogram classification. On the other hand, in spite of using most demanding and appealing techniques in all the phases only few researchers could have achieved a proper performance of a CAD system which can identify abnormalities in mammograms. It also demands more resources to train the model, if we consider all the features extracted. Feature selection basically happens after the feature extraction from the mammograms while designing a CAD system. If feature extraction gives n number of features, we try to find m dominating features from those n features, where $m<n$, in feature selection. There can be found many feature selection techniques available in literature in different problem domains, but only few of them have been applied to design a CAD system. In the next paragraph we have put some insight on various feature selection techniques. So, to remove the above-mentioned gap we have tried to build a CAD system by keeping very basic techniques for all the phases except the feature selection. For the feature selection we have introduced a 2-way threshold based IWD algorithm. This algorithm finds subsets of features from the extracted LBP [18] feature in wavelet domain. The results have shown that our proposed CAD system out performs many of the existing state-of-art works in the literature.

In literature there are many feature selection techniques available. All these techniques are broadly categorized into four basic categories- filter approach [32][33], wrapper approach [34][44], embedded approach [35] and hybrid approach [69]. On the other hand, many meta-heuristic optimization algorithms, have also been successfully applied for feature selection in mammogram classification. Few of them are ACO [36][37], PSO [38][39][46], SA [40] and GA [41][42]. In **Table 2**, a list of works that combine metaheuristic techniques with mammogram classification have presented along with their limitations.

Table 2. Various works in the literature that combines Meta-heuristics

Authors/Years	**Proposed Methods/Objectives**	**Limitations**
Khosravi, M. H., & Bagherzadeh, P. (2019) [43]	Proposed a feature selection technique based on IWD, by incorporating a suitable objective function. In this work the authors have considered the feature selection problem as a multi objective problem and tried to reduce the number of features without any prior knowledge on optimal number of features.	In case of the UCI medical datasets such as Liver and Parkinsons the proposed model has achieved accuracy as 80% and 85% respectively.
Alirezanejad, M., Enayatifar, R., Motameni, H., & Nematzadeh, H.(2020) [44]	Proposed a feature selection method for medical data sets called Xvariance and Mutual Congestion for selecting the features. The proposed methods are heuristic approach of features selection.	The ranking method and the forward selection methods used in the proposed work sometimes makes the fetaure selection techniques un-reliable. On the other hand Mutual congestion is reliable only for high dimensional datasets and Xvariance for low dimensional datasets
Shuaib, M., Adebayo, O. S., Osho, O., Idris, I., Alhassan, J. K., & Rana, N.(2019) [45]	Proposed a method to classify an email as a spam and non-spam by extracting silent features in the email corpus using Whale optimization algorithm. The authors have used rotation forest algorithm for classification.	This technique is specific to the email corpus only. It may fail with other types of data
Sahiner, B., Chan, H. P., Petrick, N., Helvie, M. A., & Goodsitt, M. M. (1998) [47]	Proposed a method to design high-sensitivity classifiers using GA based feature selection technique and applied their work to CAD system. They have used Linear Discriminant Analysis (LDA) for classification.	The authors have considered only Sensitivity and Specificity for the performance measurement. No other metrics have been used for this purpose.

Zheng, B., Chang, Y. H., Wang, X. H., Good, W. F., & Gur, D. (1999) [48]	Investigated GA based feature selection technique for computerized mass dtection in mammograms. For classification they have used Bayesian Belief Network.	While doing the investigation authors have not consider other metaheuristic optimization techniques such as PSO, ACO, SA etc. On the other hand , Additon of any features to the optimal subset decrease the performance drastically
Shaikh, T. A., & Ali, R. (2020) [49]	Proposed an intelligent heathcare system that uses Harmony Search (HS) and Simulated Annaeling (SA) combinely for precise and accurate malignancy. The classifier that has been used for their proposed CAD system is SVM (with kernel) and to proof the effieciency of their proposed framework they have applied it on local mammographic dataset.	The authors have considered only the basic performance measures of the classfier to show the efficiency of the proposed CAD system. They have not compared the results with any other metaheuristic optimization feature selection techniques. The use of a single meta-heuristic feature selection technique may further decrease the computational time.
Dheeba, J., Singh, N. A., & Selvi, S. T. (2014) [50]	Proposed a CAD system that uses Particle Warm Optimized Neural Network (PSOWNN)	Even though the area under Receiver Operating Characteristic (ROC) curve is measured as 96% for this proposed work, the sensitivity and specificity are found to be near 92%
Mohebian, M. R., Marateb, H. R., Mansourian, M., Mañanas, M. A., & Mokarian, F. (2017) [51]	Proposed an ensamble method for learning mammografic dataset that uses PSO to refine the the extracted features.	The minimum sensitivity, specificity, precision and accuracy of the proposed method are found to be 77%, 93%, 95% and 85% respectively.
Punitha, S., Amuthan, A., & Joseph, K. S. (2018) [52]	Proposed a CAD system that uses DragonFly Optimization (DFO) as a region growing technique. This technique generated the initial seed points and thresholds optimally. The authors have used a Feed Forward Neural Network (FFNN) for classification that can be trained using Back Propagation Algorithm	No congergence analysis has been provided for the proposed work. The efficiency of the framework has been measured only in terms of sensitivity and specificity.
Sambandam, R. K., & Jayaraman, S. (2018) [53]	Proposed a Self-Adaptive Dragon Fly Optimization (DFO) technique for multilevel segmentation of digital images. The proposed technique has been applied to the medical image too.	The analysis is specific to the segmentation of the digital image only.

Maleki, F., Nooshyar, M., & Fatin, G. Z. (2014). [54]	Proposed a threshold algorithm based on the Harmony Search Algorithm (HSA) for CAD system.	The analysis is specific to the thresholding only.
Xue, B., Zhang, M., & Browne, W. N. (2015) [55]	Proposed a variation of PSO for feature selection that emplyes various initialization and apprising mechanisms.	The accuracy attained for this proposed work is 94 %.
Chougrad, H., Zouaki, H., & Alheyane, O. (2020) [56]	Proposed a joint learning of the functions using classification for multi-label image. One of the the benchmark datasets, the authors have used is- MIAS.	This proposed work involves lot of processing. The analysis of time complexity of the proposed frame work could have been another factor, that would have been considered by the authors.
Aličković, E., & Subasi, A. (2017) [57]	Proposed a CAD system that combines GA based feature selection technique and Random Forest (RF) for diagnosis of Breast Cancer. For analysis the authors have applied their proposed work to Wisconsin Breast Cancer datasets.	The proposed work considers only the "Accuracy Rate" as performace metric. No, other metrics have been considered for performance evaluation.
Turabieh, H., & Muhanna, M. (2016) [58]	Proposed a CAD system that combines GA based feature selection technique and Adaptive Neuro Fuzzy Interface System (ANFIS)	The Proposed framework achieves only 71 % accuracy rate.

It has been found that, almost all the metaheuristic feature selection techniques produce a single subset of features for which a global optimum value of the objective function can be obtained. Each time we run such an algorithm; the total number of features selected in the subset remains almost same. But, through experiments we have found that, by controlling the number of features to be selected within the subset, we can further enhance the performance of such an algorithm. So, in spite of running those algorithms to produce a single subset (with fixed number of features), we can run them to generate multiple subsets with different number of features within them. To do so, we need to run them for a range assigned to the *parameter,* that represents the total number of features in a subset. In this paper work, we have introduced a mechanism called 2-way thresholding to produce multiple subsets of features from the extracted features set. The reason behind this is to make IWD flexible enough to find more accurate result.

IWD is a meta-heuristic optimization algorithm proposed by Shah-Hosseini [59] in the year 2009. It is a population-based algorithm based on the natural phenomenon of flow of the water drops while finding their way to the river. IWD is best suited for finding minimal cost path. While finding the minimal cost path, it uses the previous experience. This algorithm operates

on a graph (N, E) , where N is the set of nodes and E is the set of edges which can be generated from the problem in hand. A set of IWDs needs to be initialized on the graph so that a simulation for movement of water drops can be created. At each iteration all IWDs achieve their solutions by traversing the nodes of the graph to get a complete solution (i.e., path) T^{IWD}. At the end of each iteration an iteration-best solution T^{IB} is obtained with the assessment through a quality function.

The IWD that we have used for feature selection is associated with a controlling parameter called Final number of parameters denoted by $N_{Features}$. In spite of putting restriction on the number of important features IWD can select from the extracted feature set, it would be much better if we would make it flexible. The controlling parameter mentioned above has been used to fulfil this purpose. To set the range for this control parameter we have introduced a mechanism of 2-way thresholding which find a *lower bound (LB)* and *upper bound (UB)* on the features of the dataset. IWD with this concept of thresholding for optimal feature subsets selection improves the convergence rate as well as the detection performances of the CAD system. So, the primary contributions of our propose frame work can be listed as below-

- Designed a CAD system in which we first extracted the LBP features from the mammograms by decomposing them using Discrete Wavelet Transform (DWT). We have decomposed the images through the multi-resolution for making texture visualization clearer. Further, we have extracted computationally light weighted local binary patterns from each level. This feature takes the advantages of the gradient-based feature and holds the characteristics such as tolerance against illumination changes and robustness against monotonic grey-level changes.
- Introduced a 2-Way Threshold based IWD feature selection mechanism to select most dominating set of features from the extracted features set. The 2-Way Thresholding mechanism first finds a LB and UB on the number of dominating features to be selected, that makes the IWD algorithm capable of finding a number of subsets of dominating features. IWD algorithm without 2-Way Thresholding can only find a rigid subset of dominating features and as a result the variations of the performance of the CAD systems with respect to the increase and decrease of the total number of features in the optimal subset cannot be captured properly.
- Used one of the subsets from the selected subsets of dominating features to train SVM.

The rest of the paper has been divided into three sections. "**Section 2**" contains elaborative details of the proposed framework and the methods used to develop the proposed framework. In this section LBP feature selection in wavelet domain, IWD algorithm and 2- way thresholding has been explained. "**Section 3**" explains and depicts all the results and analysis of the experiments that we carried out to proof our claims. Finally, the "**Section 4**" concludes the proposed work giving some hints on its scope in the near future.

2. Proposed Methods based on 2-Way Threshold based IWD:

The CAD system proposed in this work consists of two basic phases. The first phase is the pre-phase and the second one is the post-phase. The pre-phase consists of all the trivial steps of

image processing such as pre-processing and feature extraction. The basic contribution of this paper work can be found in the post-phase of the model where we have introduced a 2-way threshold based IWD. This phase takes the dataset or features set generated by the pre-phase as input that contains a total of 1024 features and apply the algorithm. A basic model of our proposed CAD system has been shown in **Fig 1**.

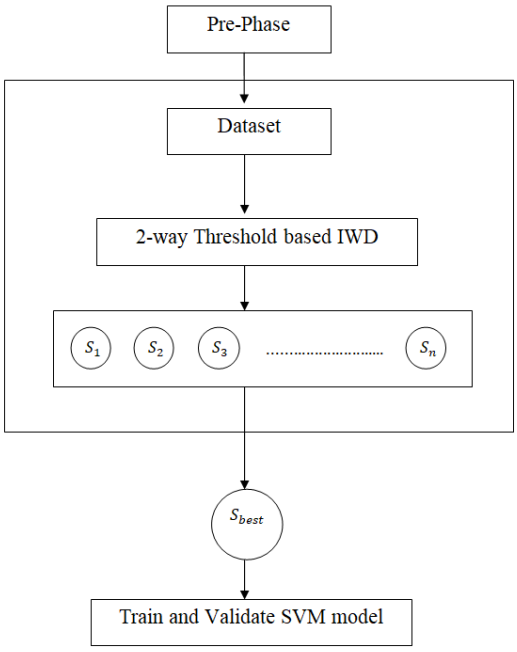
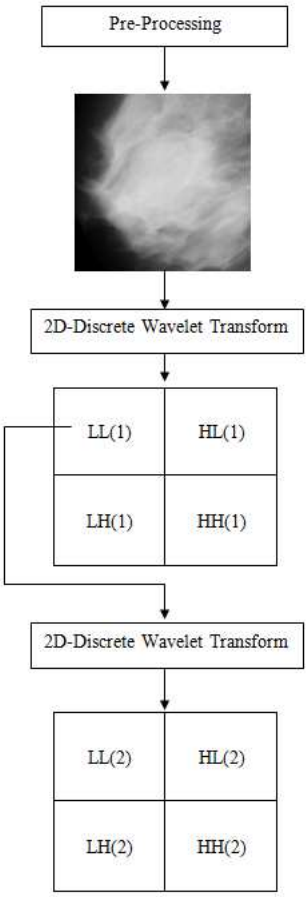


Fig 1: General model of the proposed CAD system

Further the post analysis comprises of designing a classification model that classifies mammograms more accurately. The base classifier, we have used here is Support Vector Machine (SVM). In the first step, we trained SVM classifier by fetching the dataset or features set generated after pre-processing. In this step we did not go through any overhead of making SVM fine tune by adding any optimization technique nor by fetching selective features from the dataset. We then tried to find the accuracy of the model using $k - foldcorssvalidation$ method for various values of k.

2.1. Feature Extraction for the CAD system.

In this work, we have extracted Wavelet based LBP Features (W-LBP) from each mammogram by decomposing them using Discrete Wavelet Transform (DWT). This decomposition was done up to two levels. **Fig 2(a)** is depicting the process of wavelet decomposition.



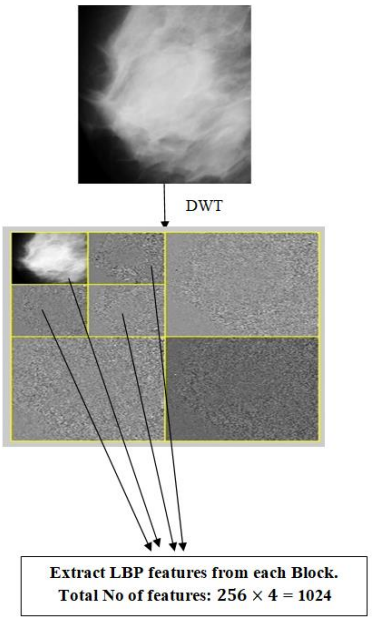


Fig 2(a): Wavelet Decomposition **Fig 2(b):** Wavelet based LBP feature extraction

2D-DWT decomposes the ROIs of the mammograms into four sub-bands LL (1) (Low-Low), LH (1) (Low-High), HL (1) (High-Low) and HH (1) (High-High) respectively in different resolution levels. While doing so, it preserves the low and high frequency details of the images. Among all these four bands LL (1) can be considered as the finest version of the original image fetched as an input. The LL (1) sub-band further decomposed using 2D-DWT to get four second level sub bands and we extracted LBP features from each sub band as shown in **Fig 2(b).** Since the dimensions of LBP features are 256 the total number of features, we extracted is 1024 (256 × 4).

LBP features are grey-scale local texture features. They are computationally light weight features derived from local neighbourhood of each pixel in the image. LBP operator can be defined mathematically as below:

$$LBP_{P,R} = \sum_{i=0}^{P-1} 2^i S(g_i - g_c)$$

$$S(x) = \begin{cases} 0, if x < 0 \\ 1, else \end{cases}$$

where

g_i: The grey scale value of neighbourhood pixel

g_c: The grey scale value centre pixel.

P: Connectivity from neighbourhood pixels.

R: Neighbourhood radius for N equally spaced pixels.

2.2. Intelligent Water Drops Algorithm (IWD)

IWD is an efficient population-based nature-inspired meta-heuristic optimization algorithm proposed by Shah-Hosseini [59] in the year 2009. This algorithm is based on the observation of the movement of the water drops while finding their way to the river, lakes or seas. This algorithm leads to a solution based on the previous experiences i.e., the solutions obtained in the previous iterations. The algorithm is best suited for finding minimal cost path. From the observation it has been found that water drops tend to move through a path with less soil. The algorithm tries to remove soil from the components of best solution so that other water drops get attracted to the path of the solution.

The given problem that needs to be solved using IWD can be represented as a graph (N, E) , where N is the set of nodes and E is the set of edges. To make the algorithm operate, a set of IWDs needs to be initialized. At each iteration, all IWDs achieve their solutions by traversing the nodes of the graph to get a complete solution (i.e., path) T^{IWD}. At the end of each iteration, an iteration-best solution T^{IB} can be obtained with the assessment through a quality function. The steps to perform this algorithm can be summed up as follows-

Phase1: Initialization this phase is responsible for initialization of the static and dynamic parameters of the process. In this phase the given problem is also converted to a graph representation. The static parameters are as follows-

N_{IWD} : Number of water drops.

(a_v, b_v, c_v): Variables to update the velocity of the water drops.

(a_s, b_s, c_s): Variables to update the soil of the local path.

$MaxIter$: Maximum number of iterations.

$initSoil$: Initial value of the local soil.

The dynamic parameters of the algorithm get initialized at the start of the iteration and get updated during the search process. The dynamic parameters are as follows-

$V_c^{IWD_r}$:Feature list visited by each water drop r

$initVel^{IWD_r}$: Velocity of the water drop r

$Soil^{IWD_r}$: Soil of the water drop r

In this phase a complete graph representation $G = (N, E)$ of the given problem is produced. In this representation Ndenotes the set of nodes (Features of the given problem) and E denotes the set of edges. The algorithm distributes all the water drops randomly on the nodes of the graph.

Phase 2: Building Solution this phase is responsible for building solutions for all water drops in a single iteration. To do so this phase goes through two steps-

i) **Edge Selection:** In this step, the water drop r on a feature i uses a probability function to choose the next un-visited feature j.The probability functions is shown below-

$$P_i^{IWD_r^{(j)}} = \frac{f(soil(i,j))}{\sum_{l=V_c} IWD_r^{f(soil(i,l))}} \quad (1)$$

In the above probability function $f(soil(i,j))$ gives the inverse value of the soil between nodes i and j and $soil(i,j)$ gives the amount of soil on the local path between nodes i and j

ii) **Rules of updating**: As a water drop r moves from node i to j the values of the velocity and soil of the water drop gets updated. These values at time $(t+1)$ have been shown below-

$$vel^{IWD_r}(t+1) = vel^{IWD_r}(t) + \frac{a_v}{b_v + c_v soil(i,j)} \quad (2)$$

The soil that gets removed from the local path carried by the water drop r can be represented as-

$$\Delta soil(i,j) = \frac{a_s}{b_s + c_s time\left(i,j; vel^{IWD_r}(t+1)\right)} \quad (3)$$

In the equation (3) $time\left(i,j; vel^{IWD_r}(t+1)\right)$ is the time required by the water drop r to move from node i to node j .

$$time\left(i,j; vel^{IWD_r}(t+1)\right) = \frac{HUD(i,j)}{vel^{IWD_r}(t+1)} \quad (4)$$

In the equation (4) HUD is problem dependent which is called as *Heuristic Undesirability.*

Based on the equation (5) and (6), the soil on the path between node i and node j , as well as, soil carried by each water drop can be updated

$$soil(i,j) = (1-\rho_n).soil(i,j) - \rho_n.\Delta soil(i,j) \quad (5)$$
$$soil^{IWD} = soil^{IWD} + \Delta soil(i,j) \quad (6)$$

Phase 3: Rules for reconstruction the iteration's best solution T^{IB} can be found out from the set of solutions obtained by every IWD and it can be computed as follows-

$$T^{IB} = arg(min)q(T^{IWD})\forall T^{IWD} \quad (7)$$

In the equation (7), T^{IB} is a best solution containing least number of features among all $IWDs$. After getting this solution T^{IB}, the path's soil that forms the best solution of the current path gets updated as below-

$$soil(i,j) = (1+\rho_{IWD}).soil(i,j) - \rho_{IWD}\frac{1}{q(T^{IWD})} \quad (8)$$

Phase 4: Condition for Termination The *phase 2* and *phase 3* keeps getting executed until the maximum number of iterations is reached. It can be shown below-

$$T^{TB} = \begin{cases} T^{TB} \; if \; q(T^{TB}) \geq q(T^{IB}) \\ T^{IB}, otherwise \end{cases} \quad (9)$$

The dynamic parameters need to be set to default values at the start of the next iteration.

2.2.1. Heuristic

In case of our problem, heuristic can be defined as a function which gets evaluated for various choices to decide the alternatives to be followed. This paper work, the proposed feature selection algorithm used this function during the feature selection phase i.e., the evaluation of the function determines the next feature to be selected in a subset and the most optimal subset to be selected by the quality function.

In case of IWD, two metrics – the soil content on the path and HUD *(Heuristic Undesirability)* decides the path to be followed or next feature to be selected during the searching of the solution. The soil content $Soil(i,j)$ represents the content of soil on the path that connects node i and node j. The relationship between the $Soil(i,j)$ and probability of the selection of the path can be represented as:

$$Soil(i,j) \propto \frac{1}{Probability of selection of the path}$$

The heuristic undesirability $HUD(i,j)$ can be defined as the undesirability of selecting node j after selecting node i into the set of solutions.

After going through various literature on CAD system designing, we have found that most of the feature selection techniques used for in CAD systems are based on single-objective optimization approach. So, in this proposed work we have kept our approach as single-objective approach, but, in future, we are really willing to extend our proposed work to use powerful multi-objective optimization approach [75]. In this work, we have chosen cross validation error rate as the HUDfunction.

$$F\big(D(t_n)\big) = Error$$

The relationship between the $HUD(i,j)$ and probability of selecting the node j can be represented as follows:

$$HUD(i,j) \propto \frac{1}{Probabilityofselectionofnodej}$$

If the value of $HUD(i,j)$ decreases, the probability of selecting node j increases and if the value of $HUD(i,j)$ increases the probability of selecting node j decrease. This is due to the fact that, the heuristic chosen in our case is the cross-validation error rate

The quality function to decide the iteration's best solution in this paper work has been defined as:

$$q(T^{IWD}) = SL\,(T^{IWD})/DR\,(T^{IWD})$$

Where

$SL(T^{IWD})$: Number of features in the subset selected by the IWD

$DR(T^{IWD})$: Detection rate of SVM for the IWD solution.

So, the iteration best solution is the one that gives the minimum value of the quality function which can be represented as:

$$T^{IB} = arg(min)q(T^{IWD})\forall T^{IWD}$$

Algorithm 1:

Input: Complete graph $G(N,E)$ of features.
Output: Optimal subset of features (T^{TB})
Initialization of the static parameters
While $Number\_Of\_Iterations \leq n_{ITER}$:

i) Initialization of the dynamic parameters
ii) Distribution of n_{IWD} number of IWDs on the nodes of the graph

iii) Add the source node just visited to the visited node list $V_C^{IWD_r}$

For each IWD:

i) Select the next node according to $P_i^{IWD_r}$

ii) Append next node to the visited list.

iii) Move drop IWD to the next selected node

Update the values of

i) Velocity, $vel^{IWD_r}(t+1)$

ii) Soil carried by the IWD on path from node i to node j , $\Delta soil(i,j)$

iii) Soil of the , $soil^{IWD}$.

iv) Soil of the path, $soil(i,j)$

End For

For each IWD solutions:

i) Calculate Subset Length, $SL(T^{IWD})$

ii) Calculate Detection Rate of SVM, $DR(T^{IWD})$

End For

Iteration's best solution $T^{IB} = arg(min)q(T^{IWD})\forall T^{IWD}$.

Update the soil value of the path followed by $IWDs$of iteration's best solution T^{IB}

if $q(T^{TB}) \geq q(T^{IB})$, then $T^{TB} = T^{IB}$

End While

2.3. Concept of Thresholding:

In this paper work, we have introduced a mechanism to produce multiple subsets of features from the extracted features set. The reason behind this is to make IWD flexible enough to find more accurate result. It has been found that, almost all the metaheuristic feature selection techniques produce a single subset of features for which a global optimum value of the objective function can be obtained. Each time we run such an algorithm; the total number of features selected in the subset remains almost same. But, through experiments we have found that, by controlling the number of features to be selected within the subset, we can further enhance the performance of such an algorithm. So, in spite of running those algorithms to produce a single subset (with fixed number of features), we can run them to generate multiple subsets with different number of features within them. To do so, we need to run them for a range assigned to the *parameter,* that represents the total number of features in a subset. This phenomenon can be explained more clearly with the help of **Fig 3.** In the figure each circle is representing a subset of features selected by a metaheuristic feature selection technique and the radius of each circle r_1, r_2, r_3, r_4 and r_5 are representing the total number of features in each subset where $r_1 < r_2 < r_3 < r_4 < r_5$. If one of these subsets, is the optimum subset produce by the algorithm, then we can easily get multiple subsets by incrementing or decrementing the radius. This also encourages us to find a lower threshold value and an upper threshold value for the radius.

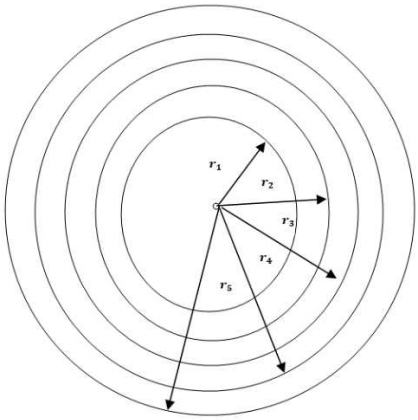


Fig 3. Subset of features with varying number of features. Number of features are represented by radius, r_n, of each circle.

The 2-way thresholding is a process of finding *Lower Bound (LB)* and an *Upper Bound (UB)* on the total number of features, $N_{Featues}$ to be selected in the subset by IWD. To find the *LB* and *UB* in this work, we first performed an exhaustive search by making the total number of $\boldsymbol{N_{Features}}$ in a subset vary from 1 to 1024. While doing so we followed complete randomization without using IWD i.e., we let our algorithm pick up features without any quality function or objective function evaluation. We evaluated the Cross-validation error rate for each of this subset and recorded the results as shown in **Algorithm 2.**

Algorithm 2:

Input: Total number of features(N_f), Dataset, Threshold Distance (T_{dist})
Output: Lower Bound(LB), Upper Bound(UB)
$E[N_f]$: Global array of size N_f to store the Error rates.
$F = FALSE$ // Flag Variable
While$i \leq N_f$:
    $D \leftarrow$Dataset with i random features.
    $Error \leftarrow Cross\_Validation\_Error(D)$.
    $Accuracy \leftarrow 100 - Error$
    $E[i] \leftarrow Accuracy$.
    IF $i \geq 2$:
        For $j \leftarrow i - 1$ to 1:
            IF $E[j]! = E[i]$:
                $LB \leftarrow i - 1$
                $F \leftarrow TRUE$
                Break
            End IF
        End For
    End IF
    $UB \rightarrow LB + T_{dist}$
    IF $F = True$:

Break.
End IF
$i \leftarrow i + 1$
End While

A case study to understand the *LB* and *UB* selection mechanism has been depicted in the **Fig 4.** As the total number of features in our dataset after LBP features extraction is 1024, we initialized N_f with a value 1024 ($N_f = 1024$. We performed the iterations as explained in the Algorithm 2 and found that for the index value 60 and 61, the difference $E[60] - E[61]$ is not equal to zero. So, our LB is 60 now. To find the UB we added a Threshold distance T_{dist} which is an input to the algorithm. In our case we added a threshold distance $T_{dist} = 40$. After adding the T_{dist} we dicided UB as 100. It is up to the implementation what value one can choose for T_{dist}. In this case study we have found that for all index value less that 60, the $Accuracy$ is 95 $\%$.

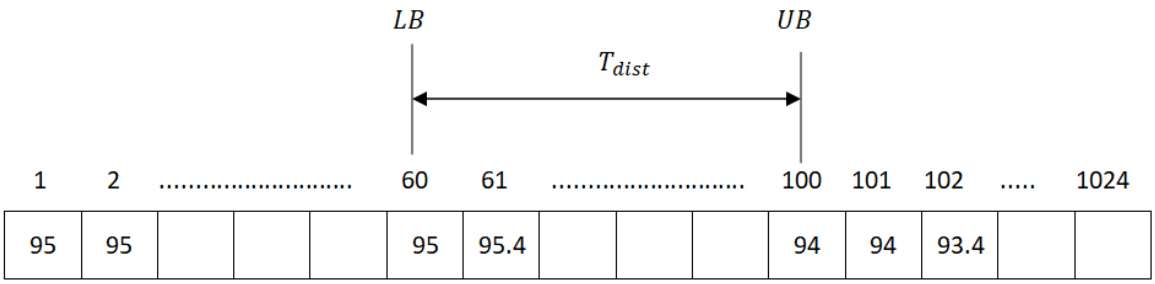


Fig 4: Finding the threshold values without IWD

In this process of selecting *LB* and *UB*, low accuracy rates had been encountered as no algorithm has been applied to select *i* features. We found that by applying IWD for this randomization we can increase the Accuracy rate to a significant amount. A case study for this randomization has been shown in **Fig 5**

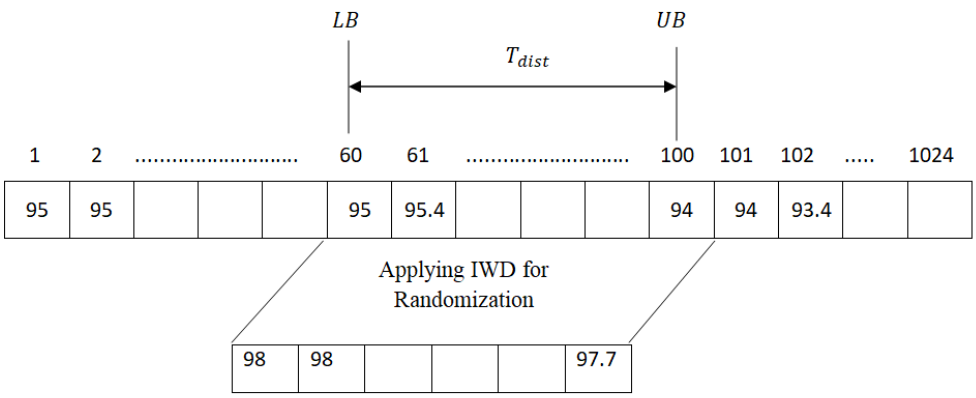


Fig 5: Applying IWD for selectin the features

Fig 6 is depicting the whole proposed feature selection frame work in the form of a flow chart.

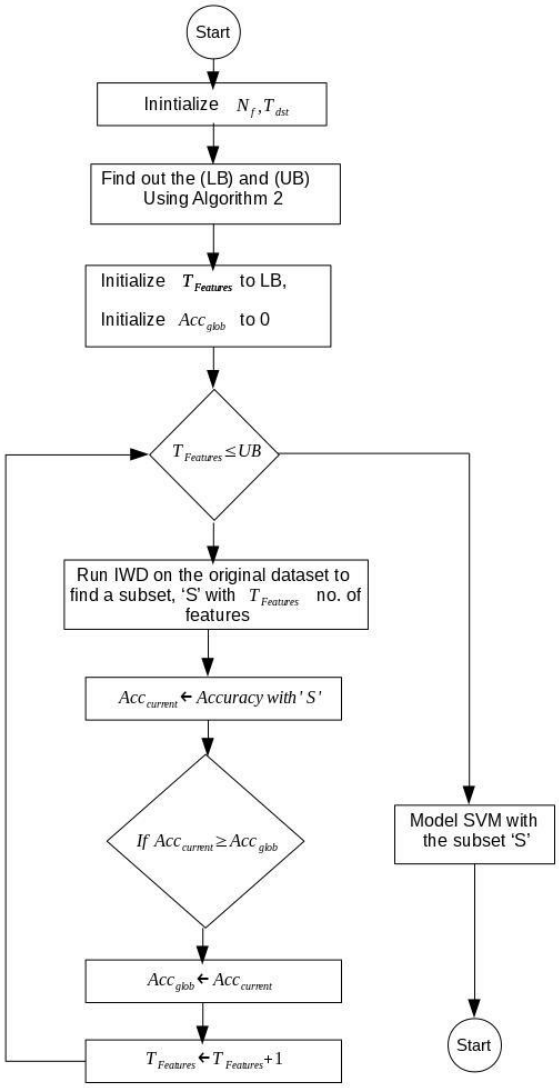


Fig 6. Flow chart for the proposed framework

Algorithm 3:

Input: Lower Bound (LB), Upper Bound (UB)

Output: An array of Accuracy Measure for various features $A[\ ]$.

$A[UB - LB]$: Global array to store the Accuracies

$i \leftarrow LB$

$j \leftarrow 0$

While $i \leq UB$:

$D \leftarrow IWD(i)$ // select i random features using IWD

Error $\leftarrow Cross\_Validation\_Error(D)$.

$Accuracy \leftarrow 100 - Error$
$A[j] \leftarrow Accuracy.$
$i \leftarrow i + 1$
$j \leftarrow j + 1$
End While

3. Result Analysis and Discussion:

Python 3.8 is the basic environment, that we have used for the implementation of our proposed CAD system. This environment gives flexibility of integrating various packages or modules to accomplish a particular task. For our task the packages or modules that have been used are shown in the **Table1**.

Table 3: Python Packages used for the implementation

Package-Name with version	**Description**
Scikit-learn-0.21.1	Module for classification, Regression and Clustering
Pandas – 1.0.3	Module for Data Analysis
Numpy-1.17.0	Module for scientific computing
Scipy-1.3.1	Module for Scientific Computing
Matplotlib-3.1.1	Module for visualization

The dataset that has been used to perform our experiments is mini-MIAS dataset. This dataset contains 322 digital mammograms each of size 1024 × 1024 pixels. After applying the basic feature extraction techniques on the mammographic images of the MIAS dataset, the features had been recorded in a .csv file. For one particular image we have recorded 1024 features with the class of that sample attached to it. Like this, we recorded a total of 327 samples. **Fig 7** is depicting a hit map that has been generated to show the co-linearity of few of the attributes.

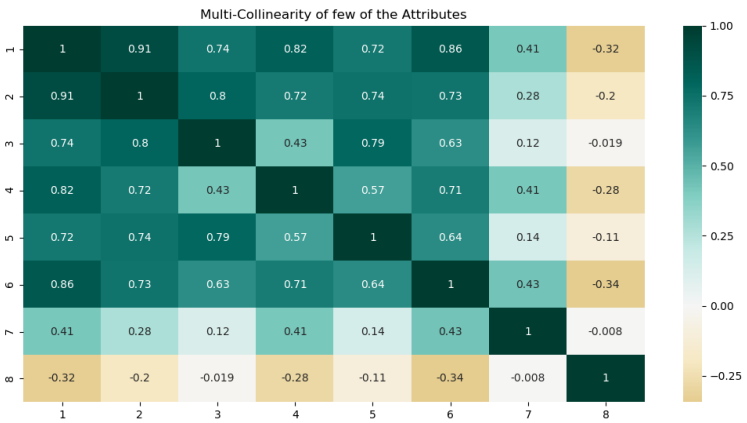


Fig 7: Co-Linearity of Attributes in the datase

For the depiction of proper result and analysis, in this paper work we have followed a step-by-step implementation model. In the first step, we have performed the experiments to proof the superiority of IWD algorithm, over other metaheuristic feature selection approaches such as ACO, PSO, SA, GA, GSA, IPO and GWO for our defined problem. To do this, we have done a comparative analysis of these algorithms by comparing their performances. We have evaluated the statistics of the minima obtained in each of the 1000 runs of all the algorithms for our defined objective function. These statistics are the mean value, the standard deviation, the minimum value (best) and maximum value (worst)**. Fig 8(a) – 8(h)** are depicting the results obtained from this evaluation.

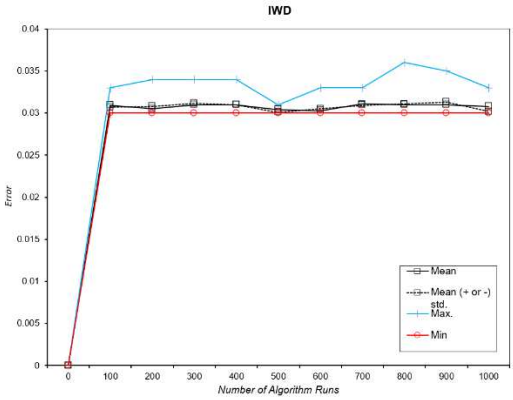


Fig 8(a)

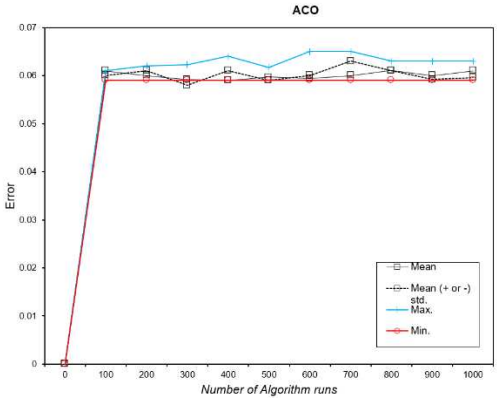


Fig 8(b)

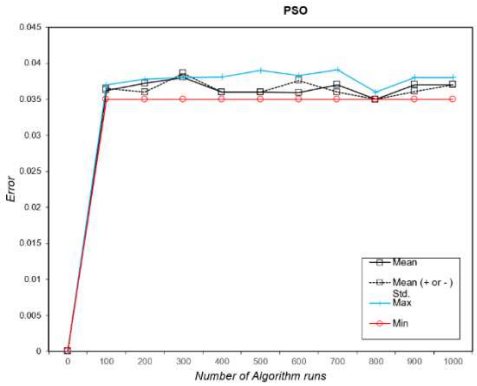


Fig 8(c)

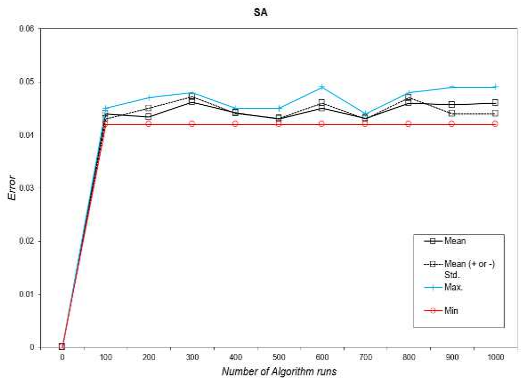
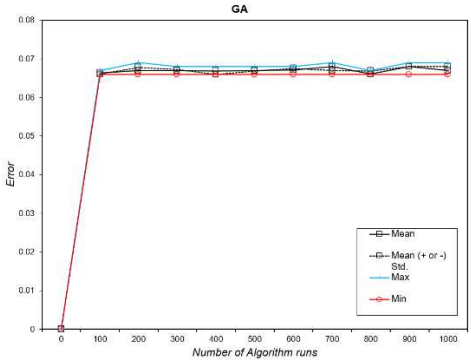


Fig 8(d)



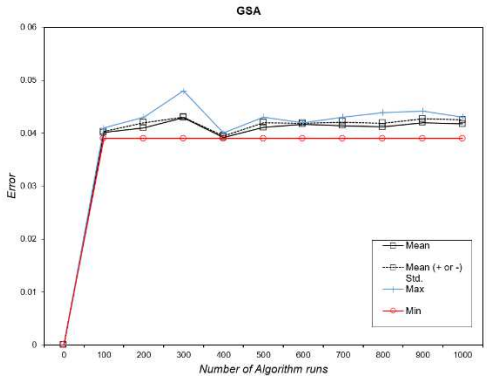


Fig 8(e) Fig 8(f)

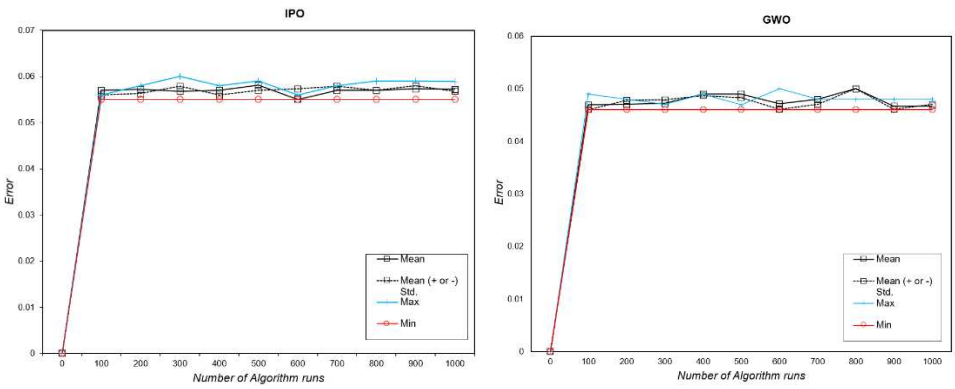


Fig 8(g) Fig 8(h)

Fig 8(a)-8(h). Statistical Analysis for algorithms IWD, ACO, PSO, SA, GA, GSA, IPO and GWO respectively.

From the above results, it is very much clear that, IWD works much better than the other algorithms, but the IWD, we have considered for the above evaluation is the IWD without any integration of 2-way thresholding. So, in the next step we have evaluated the same statistics for IWD with 2-way thresholding. The results have shown further improvements in the performance of IWD algorithm. **Fig 9** is depicting the results of this evaluation.

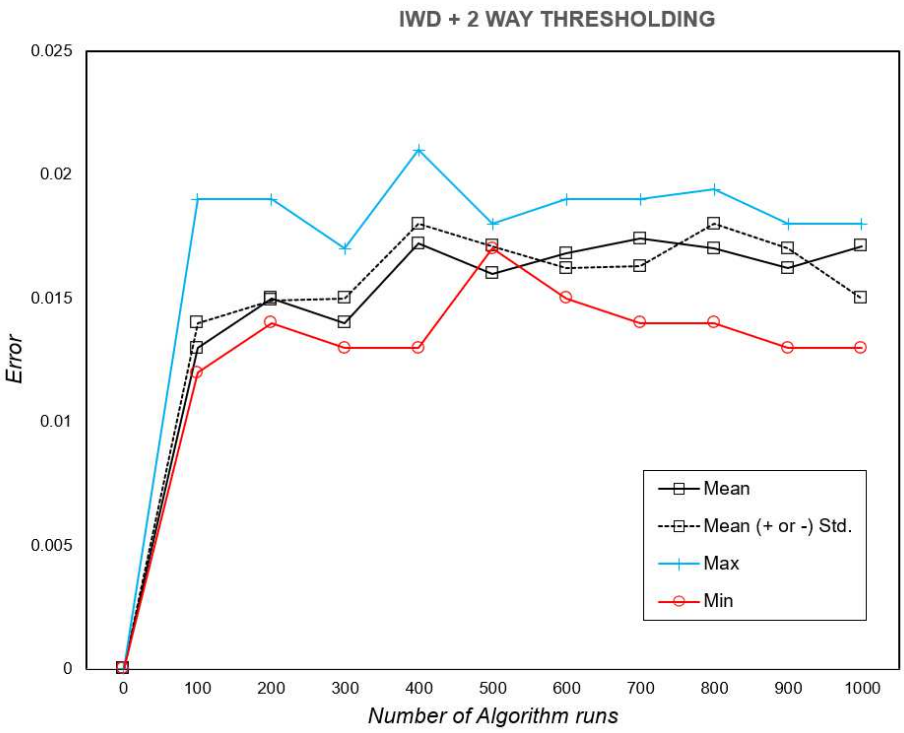


Fig 9. Statistical Analysis for IWD with 2-way thresholding.

The reason behind applying 2-way thresholding to IWD is that, there may exists a better subset near to the optimal subset found by the algorithm, with the total number of features greater or less than the total number of features in the optimal subset. This holds true for all the algorithms we have mentioned in this paper work. The variations in the error rate with respect to the total number of features in the subsets for all the above-mentioned algorithms have been depicted from **Fig 10(a)** to **10(h).**

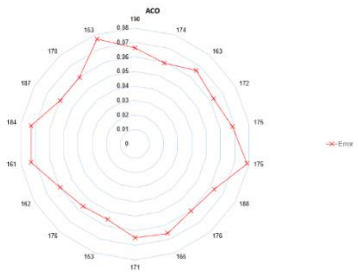


Fig 10(a)

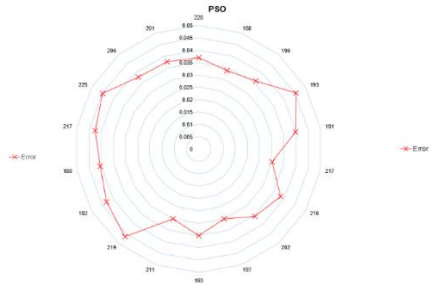


Fig 10(b)

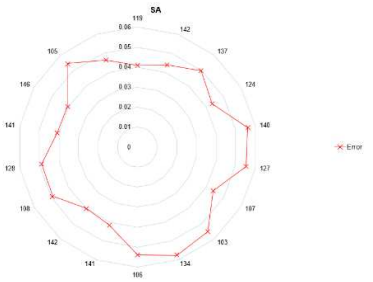


Fig 10(c)

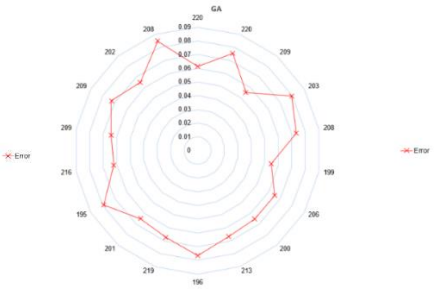


Fig 10(d)

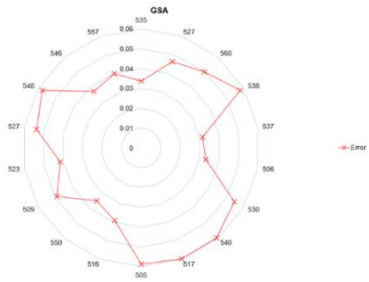


Fig 10(e)

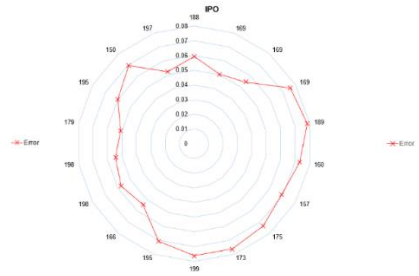
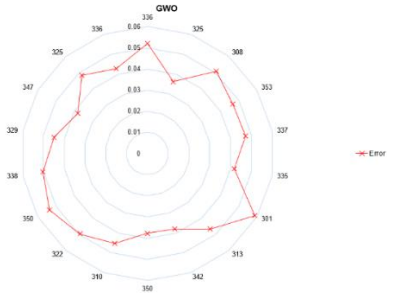


Fig 10(f)



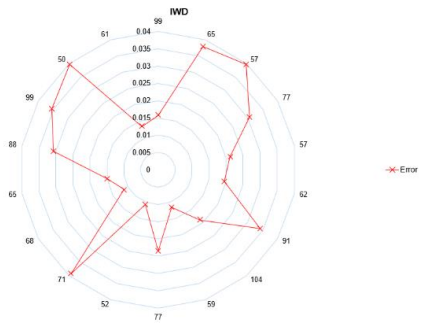


Fig 10(g) Fig 10(h)

Fig 10(a)- 10(h). variations in the error rate with respect to the total number of features in the optimal subsets for ACO, PSO, SA, GA, GSA, IPO, GWO and IWD respectively

So, from the above results it is also clear that IWD algorithm works much better than the other algorithms even with the variations in the values for $N_{Features}$. For each run of the algorithm, with different total number of features in the selected subset, IWD generates less error rate (Objective function value) compared to the other algorithms.

To consider these variations in the error rate with respect to $N_{Features}$, 2-way threshold based IWD first finds a lower bound and an upper bound for $N_{Features}$. To find out these bounds it runs the **Algorithm 2.** The result obtained by running this algorithm has been shown in the **Table 4**.

Table 4: Global Accuracy Rate without using IWD

S.No.	Number of Random Features	Global Accuracy Rate
1	**60**	**95**
2	61	95.4
3	62	95.4
4	63	95.4
5	64	95
6	64	95.7
7	66	95
8	67	95.8
9	68	95.8
10	69	95.4
11	70	95
12	71	95
13	72	95
14	73	95
15	74	95
16	75	95
17	76	95
18	77	95

19	78	95
20	79	94
21	80	94
22	81	94.1
23	82	94.1
24	83	94
25	84	94
26	85	94
27	86	94
28	87	94
29	88	94
30	89	94
31	90	94
32	91	94
33	92	94
34	93	94
35	94	94
36	95	94
37	96	94
38	97	94
39	98	94
40	99	94
41	**100**	**94**
42	101	94
43	102	93.9
44	103	92

From the above evaluation, we found *LB* as 60 and *UB* as 100, because below 60 the global accuracy rate remains at 95 and beyond 100 the global accuracy rate decreases gradually. Now we ran IWD algorithm for each value in this range and obtained multiple subsets of features with $N_{Features}$ varies from 60 to 100. Finally, we keep the best feature subset for validation of our classifier which is SVM. Few of the best subsets obtained during this run of the algorithm has been shown in the **Table 5.**

Table 5: The subsets generated by IWD within the range [60,100]

Subsets	$N_{Features}$	**Features selected**	**Global Accuracy in %**
S_1	60	[613, 314, 264, 652, 4, 4, 0, 12, 669, 949, 479, 1, 885, 9, 346, 2, 815, 8, 334, 8, 7, 853, 163, 12, 504, 0, 70, 0, 4, 775, 4, 12, 3, 72, 3, 486, 355, 312, 11, 2, 11, 658, 11, 0, 9, 0, 536, 5, 6, 1012, 639, 7, 9, 10, 245, 12, 7, 6, 121, 4, 6, 555, 545, 12, 12, 1, 12, 13, 4, 14, 15, 16, 575, 762, 833, 17, 3, 54, 9, 18, 19, 20, 535, 21, 22, 23, 24, 87, 0, 208]	98

S_2	65	[621, 9, 12, 4, 429, 4, 993, 9, 849, 9, 12, 661, 0, 176, 638, 1, 3, 4, 3, 2, 5, 6, 310, 7, 9, 8, 632, 273, 10, 123, 11, 13, 344, 66, 915, 662, 433, 5, 14, 6, 793, 15, 419, 9, 283, 16, 323, 792, 183, 402, 17, 0, 699, 18, 373, 19, 677, 129, 20, 249, 6, 835, 137, 353, 21, 362, 4, 22, 6, 9, 12, 23, 150, 24, 60, 12, 405, 25, 9, 625, 12, 26, 4, 0, 9, 27, 11, 28]	98
S_3	70	[1001, 5, 12, 687, 870, 0, 64, 965, 762, 669, 15, 487, 1006, 290, 1, 182, 9, 76, 9, 474, 0, 392, 473, 2, 201, 3, 6, 409, 5, 855, 5, 985, 701, 597, 719, 32, 6, 117, 4, 112, 7, 272, 5, 7, 123, 304, 59, 4, 11, 69, 34, 977, 800, 99, 111, 8, 294, 6, 561, 573, 48, 0, 300, 1013, 1, 47, 975, 10, 9, 0, 5, 710, 12, 7, 1017, 12, 930, 13, 530, 14, 291, 16, 606, 0, 980, 17, 18, 3, 721]	99
S_4	75	[132, 11, 495, 11, 0, 1, 2, 12, 141, 639, 769, 11, 12, 885, 223, 138, 699, 0, 3, 359, 9, 4, 12, 181, 12, 712, 5, 7, 588, 5, 207, 6, 0, 4, 6, 7, 8, 9, 5, 0, 11, 12, 448, 11, 1013, 9, 5, 696, 0, 301, 0, 10, 154, 6, 386, 190, 461, 0, 12, 13, 275, 4, 0, 245, 1, 14, 237, 15, 0, 108, 762, 0, 6, 533, 16, 12, 52, 23, 333, 17, 990, 603, 632, 18, 19, 20, 479, 21, 0, 6, 22, 24, 25, 26, 1, 6, 534, 170, 561, 4, 27, 56, 9, 28, 837, 29, 12, 779, 9, 12, 12, 30, 12, 692, 12, 984, 0, 6, 9, 31, 32]	99
S_5	80	[319, 0, 546, 27, 208, 1, 726, 238, 2, 440, 273, 3, 4, 5, 1, 584, 942, 34, 6, 7, 8, 12, 9, 12, 12, 6, 10, 66, 11, 732, 12, 13, 0, 345, 0, 190, 14, 0, 0, 92, 5, 5, 15, 377, 12, 392, 45, 12, 3, 16, 17, 3, 129, 9, 102, 12, 18, 573, 9, 753, 19, 119, 5, 855, 20, 992, 0, 6, 12, 21, 22, 838, 4, 23, 9, 7, 0, 12, 3, 199, 12, 9, 24, 25, 1004, 26, 28, 9, 646, 836, 560, 148, 29, 510, 30, 31, 779, 4, 135, 4, 32, 82, 405, 112, 154, 12, 33, 35, 762, 725, 36, 37, 95, 708]	98
S_6	85	[458, 184, 596, 0, 477, 0, 6, 1, 40, 2, 252, 270, 92, 3, 365, 822, 987, 0, 0, 6, 0, 5, 4, 379, 0, 5, 698, 7, 8, 0, 480, 9, 12, 272, 501, 12, 0, 584, 10, 543, 676, 11, 9, 758, 12, 2, 13, 0, 12, 14, 225, 5, 15, 172, 103, 9, 12, 12, 16, 17, 196, 893, 18, 19, 0, 2, 20, 619, 21, 672, 340, 1, 11, 9, 91, 0, 251, 231, 22, 0, 12, 5, 0, 23, 24, 0, 12, 6, 0, 5, 25, 26, 269, 4, 421, 446, 72, 279, 27, 28, 29, 656, 30, 191, 31, 32, 33, 883, 34, 464, 381, 12, 633, 35, 12, 12, 12, 36, 5, 11, 617, 0, 37, 908, 38, 0, 262, 734]	99
S_7	90	[58, 0, 7, 9, 480, 673, 0, 778, 0, 9, 272, 6, 1, 5, 2, 6, 847, 6, 3, 9, 7, 12, 12, 12, 339, 12, 12, 36, 741, 4, 696, 5, 675, 3, 147, 1, 8, 215, 9, 55, 10, 9, 11, 347, 194, 12, 344, 166, 620, 892, 6, 703, 356, 13, 24, 4, 14, 139, 37, 15, 1020, 16, 63, 17, 18, 398, 12, 9, 4, 6, 303, 12, 19, 12, 0, 773, 20, 21, 3, 422, 22, 219, 82, 5, 193, 4, 23, 258, 0, 12, 25, 136, 432, 32, 26, 12, 208, 6, 306, 27, 28, 5, 296, 29, 30, 246, 728, 31, 33, 155, 5, 34, 685, 609, 35, 104, 66, 160, 172, 0, 38, 9, 39, 515, 40, 0, 8, 616, 0, 12, 291]	98

S_8	95	[802, 0, 1, 6, 2, 9, 4, 698, 9, 6, 208, 326, 3, 4, 181, 5, 520, 466, 5, 32, 6, 440, 716, 564, 9, 7, 9, 334, 8, 314, 46, 10, 1006, 6, 12, 11, 9, 530, 678, 9, 12, 13, 14, 15, 1, 3, 22, 3, 16, 604, 3, 929, 5, 227, 351, 17, 264, 201, 18, 19, 20, 379, 21, 719, 758, 583, 393, 0, 23, 6, 116, 4, 526, 493, 24, 12, 25, 578, 390, 431, 234, 176, 12, 11, 1, 586, 9, 9, 463, 26, 27, 28, 29, 30, 81, 6, 9, 39, 31, 33, 800, 420, 668, 34, 272, 35, 36, 911, 4, 9, 508, 825, 124, 845, 37, 872, 38, 9, 402, 40, 41, 0, 11, 871, 0, 0, 67, 224, 7, 107]	97
S_9	100	[1022, 408, 633, 244, 0, 9, 919, 5, 4, 5, 12, 100, 12, 9, 0, 9, 0, 9, 276, 5, 168, 1, 1, 920, 9, 908, 312, 4, 2, 4, 6, 935, 0, 7, 385, 7, 7, 547, 3, 63, 6, 8, 10, 11, 301, 386, 13, 145, 48, 29, 107, 940, 4, 487, 9, 6, 177, 14, 15, 16, 193, 148, 3, 2, 12, 894, 239, 212, 0, 269, 17, 1, 58, 461, 18, 20, 59, 19, 8, 3, 203, 21, 982, 22, 23, 0, 1, 739, 428, 6, 837, 902, 24, 0, 255, 727, 2, 707, 6, 6, 25, 61, 26, 27, 602, 28, 30, 31, 72, 32, 33, 938, 34, 8, 12, 35, 6, 2, 36, 37, 4, 3, 94, 38, 12, 9, 2, 5, 39, 667, 6, 178, 12, 40, 402, 229, 41, 42, 272, 5, 9, 43, 93, 44, 12, 9, 426, 366, 45]	98
S_{10}	105	[872, 0, 899, 0, 498, 12, 7, 12, 521, 1, 2, 6, 3, 329, 5, 67, 5, 9, 4, 9, 12, 651, 390, 4, 353, 6, 7, 7, 8, 10, 661, 4, 0, 11, 908, 722, 13, 438, 419, 14, 701, 7, 0, 177, 6, 15, 414, 5, 16, 972, 200, 99, 3, 346, 17, 437, 18, 19, 20, 21, 721, 875, 12, 22, 333, 4, 194, 520, 6, 87, 6, 9, 23, 24, 0, 25, 619, 12, 12, 839, 8, 26, 892, 242, 12, 376, 27, 28, 29, 0, 775, 30, 1, 210, 902, 9, 601, 31, 32, 282, 12, 698, 6, 876, 900, 33, 480, 34, 752, 5, 0, 2, 161, 231, 9, 632, 7, 447, 35, 299, 0, 36, 499, 148, 39, 37, 818, 38, 40, 765, 2, 41, 918, 699, 42, 717, 3, 230, 43, 227, 12, 528, 45, 159, 44]	96

Among all the above selected subsets of features S_3 or S_4 or S_6 can be used to train the SVM for post analysis i.e., for classification. When new samples come, based on these features SVM can then classify the mammograms. The precision-recall curves of SVM while trained with all the above subsets have been shown in the Fig 9

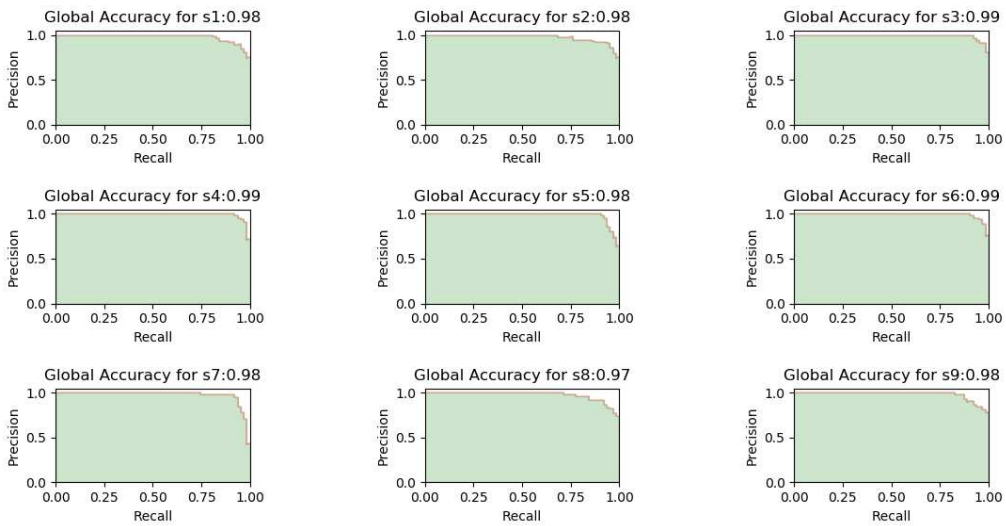


Fig 11: Precision –Recall rate curves for each of the subsets generated

Now with respect to the number of features in the final subset selected and the objective function value generated for this subset, we can compare the performance of the mentioned algorithms. It has been observed that IWD wit 2-way thresholding out performs all the other algorithms. **Table 6** is depicting the total number of features in the final subset and the global accuracy rate for this subset.

Table 6: Results obtained by running various meta-heuristic feature selection techniques

Feature Selection Algorithm	Number of features selected	Global accuracy (Based on the quality function defined
IWD with thresholding	70	0.99
IWD Without thresholding	88	0.967
ACO	161	0.941
PSO	210	0.965
SA	123	0.958
GA	207	0.934
GSA	543	0.961
IPO	178	0.945
GWO	339	0.954

To do the convergence analysis of our proposed 2-way threshold based IWD algorithm, we have generated convergence curves for 9 runs of this algorithm on the extracted features set. In each of these runs, we found that the algorithm converged properly without any fail. **Fig 12** is depicting the convergence curves for IWD with 2-way thresholding.

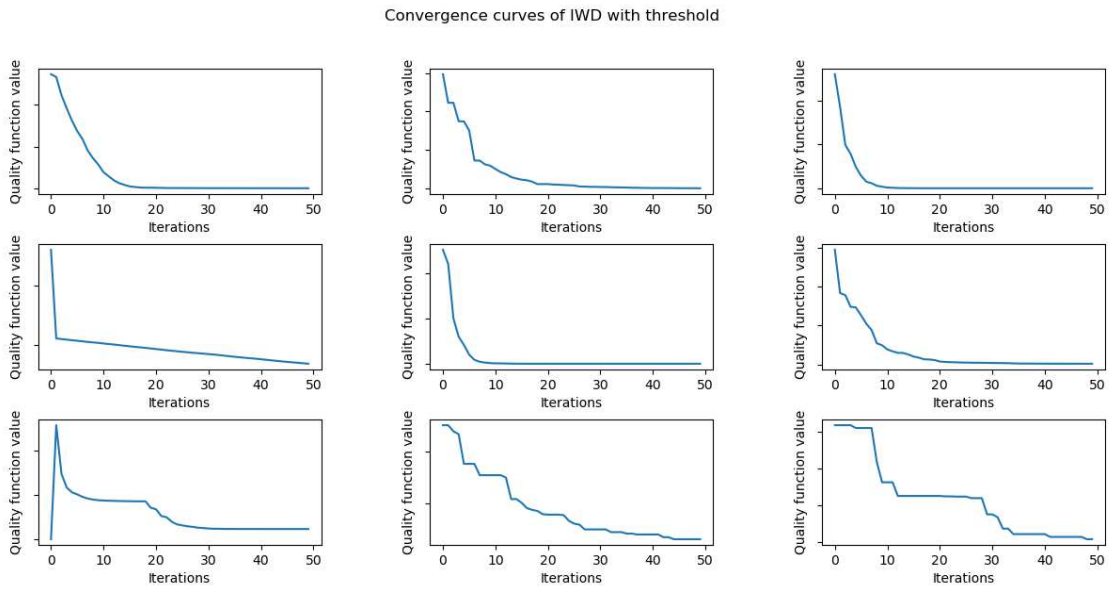


Fig 12. Convergence of IWD with thresholding

Further, we have also compared the time of convergence of our propose algorithm with the time of convergence of ACO, PSO, SA, GA, GSA, IPO and GWO. We found that our algorithm

converges faster that all the mentioned metaheuristic feature selection techniques. **Fig 13.** is depicting the convergence rate of each of the algorithm.

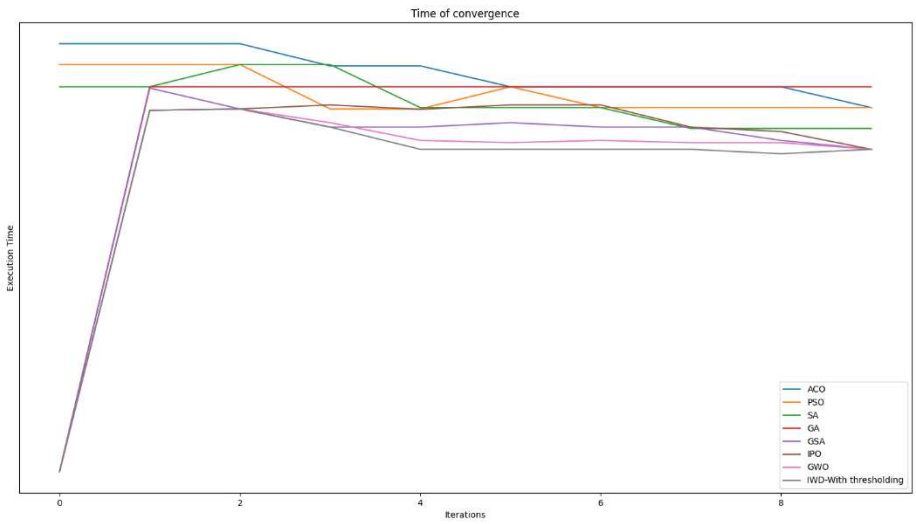


Fig 13. Convergence rates of various metaheuristic feature selection techniques

Few of the parameters used in our algorithm have been chosen through experiments and few of them have been chosen based on their standard experimental values found in various literature. The interpretation of each of the parameters and how each of these parameters has assigned that particular value have been explained below-

- $\boldsymbol{N_{Features}}$**:** (This parameter can be considered as the controlling parameter of our two-way thresholding-based algorithm. It can take a value within the *Lower Bound (LB)* and the *Upper Bound (UB)* which have been decided using **Algorithm 3.** In our case the *LB* and *UB* have been found as 60 and 100.
- $\boldsymbol{n_{IWD}}$: This parameter is to define the number of intelligent water drops for our algorithm. The value of this parameter can vary from problem to problem. We have found that for a value greater than 200 for this parameter the fitness values of our fitness function do not indicates any improvements and that is why we have chosen a number between 100 and 200 for this parameter.
- $\boldsymbol{n_{ITER}}$**:** The number of iterations for our algorithm has been chosen as 100 for all the runs with $T_{Features}$ values set within LB and UB. Through various run of IWD for our defined problem we have found that beyond number of iterations, $n_{ITRER} = 70$ **,** the algorithm gets converged and the values of the fitness function almost remain constant.
- $\boldsymbol{a_v, b_v, c_v}$**:** These are velocity parameters used to update the velocity of the water drops. After going through various applications of IWD algorithm found in literature [70][71][72][73][74] we decided to set the values for these parameters as $a_v = 1, b_v = 0.01$and $c_v = 1$. Running our proposed algorithm with other random values of these

parameters we found that the algorithm gives highest performance for the above set values.

- $\boldsymbol{a_s, b_s, c_s}$: These are soil parameters used to update the soil associated with the water drops. After going through various applications of IWD algorithm found in literature [70][71][72][73][74] we decided to set the values for these parameters as $a_s = 1, b_s = 0.01$and $c_s = 1$. Running our proposed algorithm with other random values of these parameters we found that the algorithm gives highest performance for the above set of values.
- $\boldsymbol{\epsilon_s, \rho_{IWD}}$ and $\boldsymbol{\rho_n}$: $\boldsymbol{\epsilon_s}$ is a constant parameter which needs to be a small positive number to prevent division by zero in the function $f(.)$ used in the algorithm. It has been suggested that the value of this parameter should be 0.01 in [70][71][72][73][74]. ρ_n is the local soil updating parameter whose value should be a small positive number less than 1 [70][71][72][73][74]. So, the value of this parameter has been chosen as 0.9. The global soil updating parameter ρ_{IWD} on the other hand has been chosen as given in [72]
- $\boldsymbol{initVel, initSoil}$: The constant $initVel$ is the initial velocity associated with each of the water drops. We set the value of this parameter to 4 as suggested in [70][71][72][73][74]. On the other hand, the constant represents the initial soil associated with each path between every two nodes i and j such that $soil(i,j) = initSoil$. This parameter can be chosen as any random value as suggested in [70][71][72][73][74]. We found that for $initSoil = 1000$ our algorithm gives maximum performance.

Table 7: Setting of the Parameters

Types	Parameters	Values
Static Parameters	n_{IWD}	200
	a_v, b_v, c_v	1,0.01, 1
	a_s, b_s, c_s	2, 0.01, 1
	$initSoil$	1000
	n_{ITER}	50
	$\epsilon_s, \rho_{IWD}, \rho_n$	0.01,0.9,0.9
Dynamic Parameters	$V_c^{IWD_r}$	Empty List
	$initVel$	4
	$soil^{IWD_r}$	0

In the second step, we used the optimal subsets obtained from each of the algorithms to validate the classifier. To measure the performance of the classification model we have used four fundamental metrics and they are True Positive (TP) , False Positive(FP), True Negative (TN) and False Negative (FN). From these metrics, we have evaluated the following performance measures.

- Accuracy.
- Precision.
- Recall or Sensitivity.

- Specificity.
- F1 – Score.

The above performance measures have been calculated as shown in eq. (10), eq. (11), eq. (12), eq. (13) and eq. (14)

$$Accuracy = \frac{TN+TP}{TN+TP+FN+FP} \quad (10)$$

$$Precision = \frac{TP}{TP+FN}. \quad (11)$$

$$Recall\ or\ Sensitivity = \frac{TP}{TP+FN} \quad (12)$$

$$Specificity = \frac{TN}{TN+FP} \quad (13)$$

$$F1 - Score = \frac{2*Precision*Recll}{Precision+Recall} \quad (14)$$

We have evaluated the performance measures for 5 most used classifiers such as SVM, Naïve Bayes (NB), k- Nearest Neighbours, Decision Tree (DT) and Random Forest (RF). From the results we have found that combination of 2-way thresholding with SVM yields the best results in terms of the above performance measures**. Table 8(a)** to **8(e)** are depicting the results the evaluations.

Table 8(a). Performance measures for SVM combined with various metaheuristic feature selection Techniques

Classifier + Feature Selection Tech.	Accuracy	Precision	Recall	Specificity	F1-Score
SVM + ACO	94.1	94	94.5	93.4	94.2
SVM + PSO	96.5	94.3	94.6	94	94.4
SVM + SA	95.8	94.2	96.3	94.2	95.2
SVM + GA	93.4	92.9	93	91.6	92.9
SVM + GSA	96.1	95.1	95	94.8	95.0
SVM + IPO	94.5	93	94	92.7	93.4
SVM + GWO	95.4	95	95.3	95.5	95.1
SVM + IWD	97.9	95.499	97.3	95.3	96.3
SVM+ IWD+2-WAY THRESHOLD	**99**	**98.7**	**98.123**	**96.2**	**98.4**

Table 8(b). Performance measures for NB classifier combined with various metaheuristic feature selection Techniques

Classifier + Feature Selection Tech.	Accuracy	Precision	Recall	Specificity	F1-Score
NB+ ACO	94.4	94.0	93.8	93.3	93.8
NB+ PSO	96.2	93.2	93.1	90.2	93.1
NB+ SA	95.1	95.8	93.7	93.5	94.7
NB+ GA	95.3	94.5	93.3	92.9	93.8

NB+ GSA	94.8	95.4	93.6	93.8	94.4
NB+ IPO	95.9	95.5	94.2	96.1	94.8
NB+ GWO	95.1	93.3	94.0	95.1	93.6
NB+ IWD	97	97.49	94.5	93.5	95.9
NB+ IWD+2-WAY THRESHOLD	98	97.33	95.59	95.3	96.4

Table 8(c). Performance measures for k-NN classifier combined with various metaheuristic feature selection Techniques

Classifier + Feature Selection Tech.	Accuracy	Precision	Recall	Specificity	F1-Score
k-NN + ACO	92.1	95.4	93.5	91.7	94.4
k-NN+ PSO	93.6	93.8	95.6	92.7	94.6
k-NN+ SA	94.6	93.5	95.3	92.4	94.3
k-NN+ GA	92.9	95.1	93.1	93.6	94.0
k-NN+ GSA	92.2	93.9	96.5	95.3	95.1
k-NN+ IPO	91.6	96.4	94.3	92.2	95.3
k-NN+ GWO	94.7	93.1	93.8	92.6	93.4
k-NN+ IWD	95.23	96.67	97.3	94.1	96.9
k-NN+ IWD+2-WAY THRESHOLD	**97.5**	**96**	**96.01**	**94.2**	**96.0**

Table 8(d). Performance measures for DT classifier combined with various metaheuristic feature selection Techniques

Classifier +Feature Selection Tech.	Accuracy	Precision	Recall	Specificity	F1-Score
DT+ ACO	92.1	90.3	91.5	89.4	90.8
DT+ PSO	92.2	92.6	90.6	90.2	91.5
DT+ SA	91.8	90.4	91.3	90.4	90.8
DT+ GA	93.4	89.0	90	92.2	89.4
DT+ GSA	93.1	91.5	90.4	91.4	90.9
DT+ IPO	91.5	90.5	91.2	87.4	90.8
DT+ GWO	92.4	92.7	92.3	91.3	92.4
DT+ IWD	93.9	91.4	92.6	90.0	91.9
DT+ IWD+2-WAY THRESHOLD	**95.5**	**93.2**	**94.7**	**92.1**	**93.9**

Table 8(e). Performance measures for RF classifier combined with various metaheuristic feature selection Techniques

Classifier + Feature Selection Tech.	Accuracy	Precision	Recall	Specificity	F1-Score
RF+ ACO	94.4	88.9	93.8	92.1	91.2
RF+ PSO	93.8	90.5	94.7	89.4	92.5
RF+ SA	92.8	91.2	93.3	90.1	92.2
RF+ GA	91.6	92.6	93	91.2	92.7
RF+ GSA	93.1	91.7	93.1	91.8	92.3

RF+ IPO	94.7	93.6	93.2	92.3	93.3
RF+ GWO	92.4	90.2	92.3	90.0	91.2
RF+ IWD	94.6	93.499	94	92.4	93.7
SVM+ IWD+2-WAY THRESHOLD	**95**	**93.7**	**94.823**	**92.8**	**94.2**

From the results depicted in the **Table 8(a)-8(e),** we found *SVM combined with 2-way thresholding-based feature selection technique* gives the best result in terms of the performance measures. We recorded the accuracy rate of NB classifier with IWD (with thresholding) as 98% which is more than the accuracy rate of NB classifier IWD (without thresholding). The recorded accuracy rate for the latter case is 97 %. On the other hand, we have recorded the accuracy rate of K-NN classifier with IWD (with thresholding) and IWD (without thresholding) as 97.5% and 95.23 % respectively. Similarly, the measured accuracy rates for DT with IWD (with thresholding), DT with IWD (without thresholding), RF with IWD (with thresholding) and RF with IWD (without thresholding) are 93.9%, 95.5%, 94.6% and 95% respectively. So, we found that the combination of 2-way threshold based IWD with SVM yields the best results. The recorded accuracy rate for SVM with IWD (with thresholding) and IWD (without thresholding) are 99% and 97.9% respectively. On the other hand, it is very much clear form the results that SVM combined with 2-way threshold best IWD not only increase the accuracy, but also other performance metrics such as – Precision, Recall or Sensitivity, Specificity and F1-Score.

Further, we have also evaluated the *Area Under Curves (AUC)* for combinations of various meta-heuristic feature selection techniques with various classifiers. **Fig 14(a)** to **Fig 14(e)** is depicting the *Receiver Operating Characteristic (ROC)* curves for this evaluation.

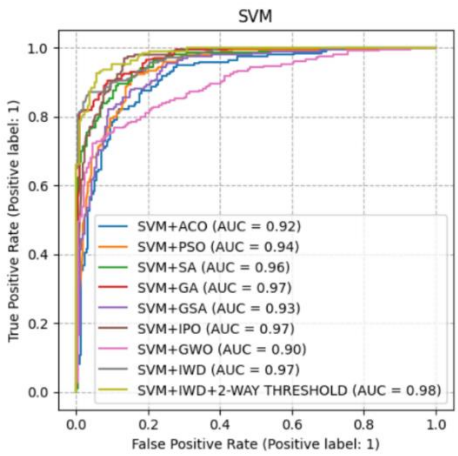


Fig 14(a)

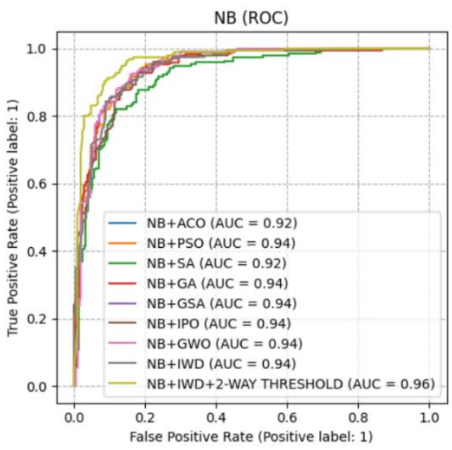


Fig 14(b)

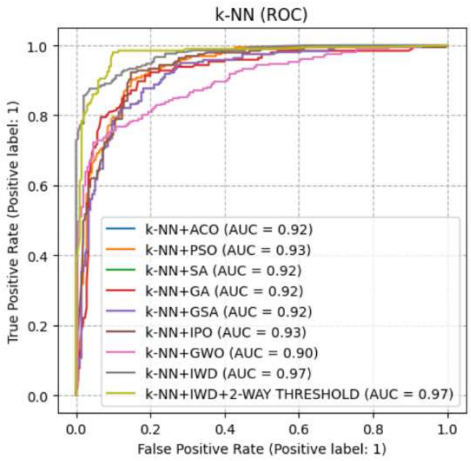


Fig 14(c)

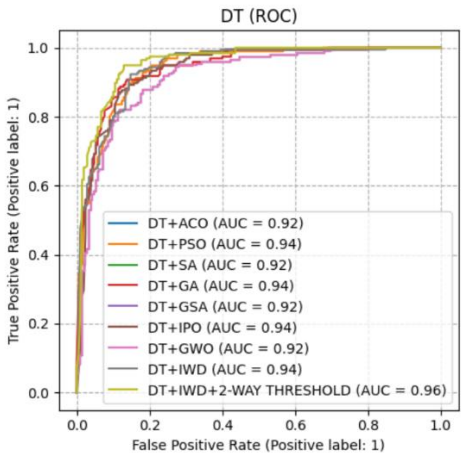


Fig 14(d)

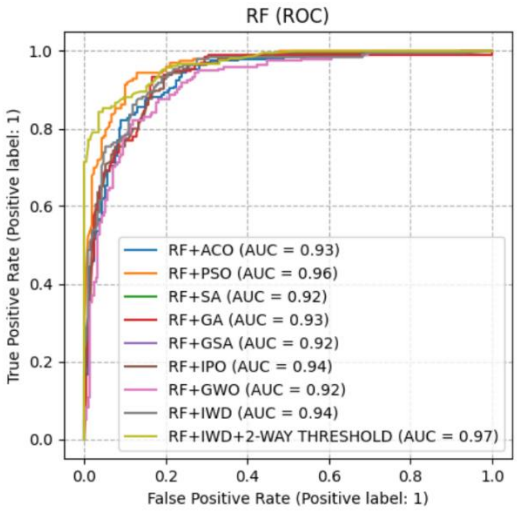


Fig 14(e)

Fig 14(a)-14(e). ROC curves for combinations of various meta-heuristic feature selection techniques with various classifiers.

To, justify the performance of our CAD system we have also provided a comparative analysis with various other CAD systems along with their accuracy in the **Table 9**.

Table 9: Comparing with various other CAD Systems

State-of-art Work	Various techniques used	performance measures	Datasets used
Proposed Work	Pre-processing+ LBP features in wavelet domain+ 2-way threshold based IWD +SVM	**Accuracy:99%, Precison:98.7%, Recall:98.123%,**	MIAS

Yanfeng et al. [21]	Homogeneous texture and high intensity deviation for pectoral muscle edge detection + Kalman filter to refine the roughness of the identified edge	Accuracy: 90%,	MIAS
Khaoula et. al [23]	Entropy Thresholding + PSO + Fourier Transform+ GLCM+SVM	Accuracy: 83.7%	
Wang et al. [24]	WFRFT+PCA+(Jaya-FNN)	Accuracy: 92.27%	MIAS
Anter et al. [25]	RG+GLCM+k-NN	Accuracy:97 %	MIAS
Abubacker et al. [26]	GLCM+GARM+ACFNN	Accuracy: 95.11 %	DDSM
Singh et al. [27]	Pre-processing+ CS-LBP features in Wavelet Domain+SVM-RFE based feature selection+ Random Forest classification	Accuracy: 97.25%	MIAS
Thawkar and Ingolikar [60]	Biogeography based Optimization (BBO) + ANFIS	Accuracy: 98.92%	DDSM
Ancy et al. [61]	Graph cut Segmentation + GLCM+SVM	Accuracy: 98.11%	MIAS
Thawkar and Ingolikar [62]	Firefly Algorithm (FFA) + Artificial Neutral Network (ANN)	Accuracy: 95.23%	DDSM
Dhahbi et al.[63]	Discrete curvelet transform + t test ranking for feature selection +k-NN classifier	Accuracy : 91.27% for MIAS database	MIAS and DDSM
Jona et al.[64]	Gray level co-occurrence matrix based statistical features + SVM classifier	Accuracy: 94.0%	MIAS
G¨orgel et al.[65]	Stationary wavelet transform (SWT)+SVM Classifier	Accuracy rate: 96.0%	MIAS
Bhosle et al. [66]	Adaboost + KNN-RBFSVM	Accuracy: 96.87%	DDSM
Jiao et al. [67]	Parasitic metric learning	Accuracy: 97.4 % (DDSM), 97.4 % (MIAS)	DDSM,MIAS
Mabrouk et al. [68]	Automatic ANN	Accuracy: 97%	MIAS

From the above experiments, it has been found that our proposed CAD System integrated with 2-way threshold based IWD for feature selection generates an accuracy of 99% in less convergence time.

4. Conclusion and Future Work:

This work proposed a CAD system which integrates an effective feature selection technique in the post analysis phase. This feature selection technique is based on a Meta heuristic optimization Algorithm IWD. To make IWD more flexible we have also introduced a concept of thresholding. Using this thresholding IWD is capable of finding a set of subsets of features from the dataset in spite of finding a single rigid subset. Even though we have applied this model on mammogram classification, this model can be used for other real world classification problem as well. In medical domain this model will be suitable for datasets with large number of features or attributes. Further this model can also be used for applications such as- detection

of tumor, detection of polyps in the colon and lung cancer etc. We have compared our proposed CAD system with many of the existing systems and results have shown that our system outperforms the other. Further we have compared the feature selection technique used in this paper work with other metaheuristic approaches such as- ACO, PSO, Simulated Annealing and GA. The results have shown that our introduce features selection technique outperforms the others.

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 by any authors

Acknowledgement:

This work is financially supported by a granted project **1-573645901** from "**Collaborative Research Scheme (TEQIP-III)**" under "**All India council of Technical Education**", New Delhi, India.

References:

[1] Statistical data for breast cancer, from IARC (International Agency for Research on Cancer) of WHO (World Health Organization). Available in: http://www.iarc.fr/en/media-centre/pr/2013/pdfs/pr223_E.pdf. Accessed on February 2013

[2] "Computer –aided Diagnosis: Tipping Point of Digital Pathology". Digital Pathology Association. 27 April 2017.

[3] Baker, J. A.; Rosen, E. L.; Lo, J. Y.; et al. (2003). "Computer-Aided Detection (CAD) in Screening Mammography: Sensitivity of Commercial CAD Systems for Detecting Architectural Distortion". *American Journal of Roentgenology*. **181** (4): 1083–1088.

[4] Suzuki, K.; Li, F.; Sone, S.; Doi, K. (2005). "Computer-aided diagnostic scheme for distinction between benign and malignant nodules in thoracic low-dose CT by use of massive training artificial neural network". *IEEE Transactions on Medical Imaging*. **24**(9): 1138–1150.

[5] Pisano, E. D., & Yaffe, M. J. (2005). Digital mammography. *Radiology*, *234*(2), 353-362.

[6] Gøtzsche, P. C., & Jørgensen, K. J. (2013). Screening for breast cancer with mammography. *Cochrane database of systematic reviews*, (6).

[7] Bandyopadhyay, S. K. (2010). Pre-processing of mammogram images. International Journal of Engineering Science and Technology, 2(11), 6753-6758.

[8] Mirzaalian, H., Ahmadzadeh, M. R., Sadri, S., & Jafari, M. (2007, May). Pre-processing Algorithms on Digital Mammograms. In MVA (pp. 118-121).

[9] Ibrahim, N. S. A., Soliman, N. F., Abdallah, M., & El-Samie, F. E. A. (2016, December). An algorithm for pre-processing and segmentation of mammogram images. In 2016 11th International Conference on Computer Engineering & Systems (ICCES) (pp. 187-190). IEEE.

[10] Comer, M. L., Liu, S., & Delp, E. J. (1996, June). Statistical segmentation of mammograms. In *Proceedings of the 3rd International Workshop on Digital Mammography* (pp. 475-478).

[11] Kupinski, M. A., & Giger, M. L. (1998). Automated seeded lesion segmentation on digital mammograms. *IEEE Transactions on medical imaging*, *17*(4), 510-517.

[12] Lu, W., Dou, R., & Zhang, G. (2013, October). A new method for extracting region of interest in mammograms. In *2013 IEEE International Conference on Medical Imaging Physics and Engineering* (pp. 228-230). IEEE.

[13] Hong, B. W., & Sohn, B. S. (2009). Segmentation of regions of interest in mammograms in a topographic approach. *IEEE Transactions on Information Technology in Biomedicine*, *14*(1), 129-139.

[14] Mustra, M., Grgic, M., & Rangayyan, R. M. (2016). Review of recent advances in segmentation of the breast boundary and the pectoral muscle in mammograms. *Medical & biological engineering & computing*, *54*(7), 1003-1024.

[15] Singh, V. P., Srivastava, A., Kulshreshtha, D., Chaudhary, A., & Srivastava, R. (2016). Mammogram classification using selected GLCM features and random forest classifier. *International Journal of Computer Science and Information Security*, *14*(6), 82.

[16] Beura, S., Majhi, B., & Dash, R. (2015). Mammogram classification using two dimensional discrete wavelet transform and gray-level co-occurrence matrix for detection of breast cancer. *Neurocomputing*, *154*, 1-14.

[17] Heinlein, P., Drexl, J., & Schneider, W. (2003). Integrated wavelets for enhancement of microcalcifications in digital mammography. *IEEE Transactions on medical imaging*, *22*(3), 402-413.

[18] Du, L., You, X., Xu, H., Gao, Z., & Tang, Y. (2010, August). Wavelet domain local binary pattern features for writer identification. In *2010 20th International Conference on Pattern Recognition* (pp. 3691-3694). IEEE.

[19] Zhang, X., Zhang, Y., Han, E. Y., Jacobs, N., Han, Q., Wang, X., & Liu, J. (2017, November). Whole mammogram image classification with convolutional neural networks. In *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 700-704). IEEE.

[20] Liu, X., & Tang, J. (2013). Mass classification in mammograms using selected geometry and texture features, and a new SVM-based feature selection method. *IEEE Systems Journal*, *8*(3), 910-920.

[21] Lia, Y., Chena, H., Yangb, Y., Yanga, N.: Pectoral muscle segmentation in mammograms based on homogenous texture and intensity deviation. Pattern Recogn. **46**(3), 681–691 (2013)

[22] Anitha, J., Peter, J.D.: A wavelet based morphological mass detection and classification in mammograms. In: International Conference on Machine Vision and Image Processing (MVIP), pp. 25–28 (2012)

[23] Soulami, K. B., Saidi, M. N., & Tamtaoui, A. (2016, May). A cad system for the detection of abnormalities in the mammograms using the metaheuristic algorithm particle swarm optimization (pso). In *International Symposium on Ubiquitous Networking* (pp. 505-517). Springer, Singapore.

[24] S. Wang, R.V. Rao, P. Chen, Y. Zhang, A. Liu, L. Wei, Abnormal breast detection in mammogram images by feed-forward neural network trained by jaya algorithm, Fund. Inform. 151 (1–4) (2017) 191–211.

[25] A.M. Anter, A.E. Hassenian, Computer aided diagnosis system for mammogram abnormality, in: Medical Imaging in Clinical Applications, Springer, 2016, pp. 175–191.

[26] N.F. Abubacker, A. Azman, S. Doraisamy, M.A.A. Murad, An integrated method of associative classification and neuro-fuzzy approach for effective mammographic classification, Neural Comput. Appl. 28 (12) (2017) 3967–3980.

[27] Singh, V. P., Srivastava, S., & Srivastava, R. (2017). Effective mammogram classification based on center symmetric-LBP features in wavelet domain using random forests. *Technology and Health Care*, *25*(4), 709-727.

[28] Chandy, D. A., Christinal, A. H., Theodore, A. J., & Selvan, S. E. (2017). Neighbourhood search feature selection method for content-based mammogram retrieval. *Medical & biological engineering & computing*, *55*(3), 493-505.

[29] Luo, S. T., & Cheng, B. W. (2012). Diagnosing breast masses in digital mammography using feature selection and ensemble methods. *Journal of medical systems*, *36*(2), 569-577.

[30] Mohanty, F., Rup, S., Dash, B., Majhi, B., & Swamy, M. N. S. (2019). Mammogram classification using contourlet features with forest optimization-based feature selection approach. *Multimedia Tools and Applications*, *78*(10), 12805-12834.

[31] Chaieb, R., & Kalti, K. (2019). Feature subset selection for classification of malignant and benign breast masses in digital mammography. *Pattern Analysis and Applications*, *22*(3), 803-829.

[32] Xing, E. P., Jordan, M. I., & Karp, R. M. (2001, June). Feature selection for high-dimensional genomic microarray data. In *Icml* (Vol. 1, pp. 601-608).

[33] Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of machine learning research*, *3*(Mar), 1157-1182.

[34] Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection. *Artificial intelligence*, *97*(1-2), 273-324.

[35] Blum, A. L., & Langley, P. (1997). Selection of relevant features and examples in machine learning. *Artificial intelligence*, *97*(1-2), 245-271.

[36] Kashef, S., & Nezamabadi-pour, H. (2015). An advanced ACO algorithm for feature subset selection. *Neurocomputing*, *147*, 271-279.

[37] Kanan, H. R., Faez, K., & Taheri, S. M. (2007, July). Feature selection using ant colony optimization (ACO): a new method and comparative study in the application of face recognition system. In *Industrial Conference on Data Mining* (pp. 63-76). Springer, Berlin, Heidelberg.

[38] "Feature selection using PSO-SVM." *International Journal of Computer Science* (2007).

[39] Chuang, L. Y., Chang, H. W., Tu, C. J., & Yang, C. H. (2008). Improved binary PSO for feature selection using gene expression data. *Computational Biology and Chemistry*, *32*(1), 29-38.

[40] Mafarja, M. M., & Mirjalili, S. (2017). Hybrid whale optimization algorithm with simulated annealing for feature selection. *Neurocomputing*, *260*, 302-312.

[41] Tan, F., Fu, X., Zhang, Y., & Bourgeois, A. G. (2008). A genetic algorithm-based method for feature subset selection. *Soft Computing*, *12*(2), 111-120.

[42] Babatunde, O. H., Armstrong, L., Leng, J., & Diepeveen, D. (2014). A genetic algorithm-based feature selection.

[43] Khosravi, M. H., & Bagherzadeh, P. (2019). A new method for feature selection based on intelligent water drops. *Applied Intelligence*, *49*(3), 1172-1184.

[44] Alirezanejad, M., Enayatifar, R., Motameni, H., & Nematzadeh, H. (2020). Heuristic filter feature selection methods for medical datasets. *Genomics*, *112*(2), 1173-1181.

[45] Shuaib, M., Adebayo, O. S., Osho, O., Idris, I., Alhassan, J. K., & Rana, N. (2019). Whale optimization algorithm-based email spam feature selection method using rotation forest algorithm for classification. *SN Applied Sciences*, *1*(5), 390.

[46]Mohammadi, A., Mohammadi, M., & Zahiri, S. H. (2018). Design of optimal CMOS ring oscillator using an intelligent optimization tool. *Soft Computing*, *22*(24), 8151-8166.

[47] Sahiner, B., Chan, H. P., Petrick, N., Helvie, M. A., & Goodsitt, M. M. (1998). Design of a high-sensitivity classifier based on a genetic algorithm: application to computer-aided diagnosis. Physics in Medicine & Biology, 43(10), 2853.

[48] Zheng, B., Chang, Y. H., Wang, X. H., Good, W. F., & Gur, D. (1999). Feature selection for computerized mass detection in digitized mammograms by using a genetic algorithm. Academic radiology, 6(6), 327-332.

[49] Shaikh, T. A., & Ali, R. (2020). An intelligent healthcare system for optimized breast cancer diagnosis using harmony search and simulated annealing (HS-SA) algorithm. Informatics in Medicine Unlocked, 21, 100408.

[50] Dheeba, J., Singh, N. A., & Selvi, S. T. (2014). Computer-aided detection of breast cancer on mammograms: A swarm intelligence optimized wavelet neural network approach. Journal of biomedical informatics, 49, 45-52.

[51] Mohebian, M. R., Marateb, H. R., Mansourian, M., Mañanas, M. A., & Mokarian, F. (2017). A hybrid computer-aided-diagnosis system for prediction of breast cancer recurrence (HPBCR) using optimized ensemble learning. Computational and structural biotechnology journal, 15, 75-85.

[52] Punitha, S., Amuthan, A., & Joseph, K. S. (2018). Benign and malignant breast cancer segmentation using optimized region growing technique. Future Computing and Informatics Journal, 3(2), 348-358.

[53] Sambandam, R. K., & Jayaraman, S. (2018). Self-adaptive dragonfly based optimal thresholding for multilevel segmentation of digital images. Journal of King Saud University-Computer and Information Sciences, 30(4), 449-461.

[54] Maleki, F., Nooshyar, M., & Fatin, G. Z. (2014). Breast cancer segmentation in digital mammograms based on harmony search optimization. Tech. J. Eng. Appl. Sci, 4(4), 477-484.

[55] Xue, B., Zhang, M., & Browne, W. N. (2014). Particle swarm optimisation for feature selection in classification: Novel initialisation and updating mechanisms. Applied soft computing, 18, 261-276.

[56] Chougrad, H., Zouaki, H., & Alheyane, O. (2020). Multi-label transfer learning for the early diagnosis of breast cancer. Neurocomputing, 392, 168-180.

[57] Aličković, E., & Subasi, A. (2017). Breast cancer diagnosis using GA feature selection and Rotation Forest. Neural Computing and Applications, 28(4), 753-763.

[58] Turabieh, H., & Muhanna, M. (2016). GA-based feature selection with ANFIS approach to breast cancer recurrence. International Journal of Computer Science Issues (IJCSI), 13(1), 36.

[59] Shah-Hosseini, H. (2009). The intelligent water drops algorithm: a nature-inspired swarm-based optimization algorithm. *International Journal of Bio-inspired computation*, *1*(1-2), 71-79.

[60] S. Thawkar, R. Ingolikar, Classification of masses in digital mammograms using biogeography-based optimization technique, J. King Saud Univ.-Comput. Inf. Sci. (2018).

[61] C. Ancy, L.S. Nair, Tumour classification in graph-cut segmented mammograms using GLCM features-fed SVM, in: Intelligent Engineering Informatics, Springer, 2018, pp. 197–208

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

[63] Dhahbi S, Barhoumi W, Zagrouba E. Breast cancer diagnosis in digitized mammograms using curvelet moments. Computers in biology and medicine. 2015 Sep 1; 64:79-90. DOI: http://dx.doi.org/10.1016/j.compbiomed.2015.06.012

[64] Jona J, Nagaveni N. A hybrid swarm optimization approach for feature set reduction in digital mammograms. WSEAS Transactions on Information Science and Applications. 2012;9:340-9.

[65] Görgel P, Sertbas A, Ucan ON. Mammographical mass detection and classification using local seed region growing–spherical wavelet transform (lsrg–swt) hybrid scheme. Computers in biology and medicine. 2013 Jul 1;43(6):765-74. DOI: http://dx.doi.org/10.1016/j.compbiomed.2013.03.008

[66] U. Bhosle, J. Deshmukh, Mammogram classification using AdaBoost with RBFSVM and Hybrid KNN–RBFSVM as base estimator by adaptively adjusting γ and C value, Int. J. Inf. Technol. 11 (4) (2019) 719–726.

[67] Z. Jiao, X. Gao, Y. Wang, J. Li, A parasitic metric learning net for breast mass classification based on mammography, Pattern Recognit. 75 (2018) 292–301

[68] M.S. Mabrouk, H.M. Afify, S.Y. Marzouk, Fully automated computeraided diagnosis system for micro calcifications cancer based on improved mammographic image techniques, Ain Shams Eng. J. (2019)

[69] Deng, W., Xu, J., Song, Y., & Zhao, H. (2020). An effective improved co-evolution ant colony optimisation algorithm with multi-strategies and its application. *International Journal of Bio-Inspired Computation*, *16*(3), 158-170.

[70] Hosseini, H. S. (2007, September). Problem solving by intelligent water drops. In 2007 IEEE congress on evolutionary computation (pp. 3226-3231). IEEE.

[71] Shah-Hosseini, H. (2010). Intelligent water drops algorithm A new optimization method for solving the vehicle routing problem. In IEEE International Conference on Systems, Man and Cybernetics (SMC) (pp. 4142-4146).

[72] Alijla, B. O., Peng, L. C., Khader, A. T., & Al-Betar, M. A. (2013, March). Intelligent water drops algorithm for rough set feature selection. In Asian Conference on Intelligent Information and Database Systems (pp. 356-365). Springer, Berlin, Heidelberg.

[73] Alijla, B. O., Wong, L. P., Lim, C. P., Khader, A. T., & Al-Betar, M. A. (2014). A modified intelligent water drops algorithm and its application to optimization problems. Expert Systems with Applications, 41(15), 6555-6569.

[74] Shah-Hosseini, H. (2009). The intelligent water drops algorithm: a nature-inspired swarm-based optimization algorithm. International Journal of Bio-inspired computation, 1(1-2), 71-79.

[75] Shahraki, N. S., & Zahiri, S. H. (2020, September). Multi-objective optimization algorithms in analog active filter design. In 2020 8th Iranian Joint Congress on Fuzzy and intelligent Systems (CFIS) (pp. 105-109). IEEE.