

Heterogeneous Ensemble Pruning based on Bee Algorithm for Mammogram Classification

Ashwaq Qasem¹, Shahnorbanun Sahran³
Centre of Artificial Intelligence,
Faculty of Information Science and Technology,
Universiti Kebangsaan Malaysia,
43650 Bangi, Selangor, Malaysia.

Siti Norul Huda Sheikh Abdullah²
Center for Cyber Security,
Faculty of Information Science and Technology, Universiti
Kebangsaan Malaysia, 43650 Bangi, Selangor, Malaysia

Dheeb Albashish⁴
Computer Science Department, Prince Abdullah Bin Ghazi
Faculty of Information Technology,
Al-Balqa Applied University, Jordan

Rizuana Iqbal Hussain⁵, Shantini Arasaratnam⁶
Department of Radiology,
Universiti Kebangsaan Malaysia Medical Centre,
56000, Cheras, Kuala Lumpur, Malaysia⁵
Department of Radiology, General Hospital of Kuala
Lumpur, Malaysia⁶

Abstract—In mammogram, masses are primary indication of breast cancer; and it is necessary to classify them as malignant or benign. In this classification task, Computer Aided Diagnostic (CAD) system by using ensemble learning is able to assist radiologists to have better diagnosis of mammogram images. Ensemble learning consists of two steps, generating multiple base classifiers and then combining them together. However, combining all base classifier in the ensemble model increases the computational cost and time. Therefore, ensemble pruning is an important step in ensemble learning to select the ensemble's members. Due to huge subsets of combination in the ensemble, selecting the proper ensemble subset is desirable. The objective of this paper is to select the optimal ensemble subset by using Bee Algorithm (BA). A pool of different classifier models such as Support vector machine, k-nearest neighbour and linear discriminant analysis classifiers have been trained using different samples of training data and 12 groups of features. Then, by using this pool of classifier models, BA was used to exploit and explore random uniform combination subsets of the trained classifiers. As a result, the best subset will be selected as the optimal ensemble. The mammogram image dataset that was used to test the model has been collected from Hospital Kuala Lumpur (HKL) and consists of 263 benign and malignant masses. The proposed method gives 77 % of Area Under Curve(AUC), 83% of accuracy, 93% of specificity and 60% of sensitivity.

Keywords—Ensemble learning; ensemble pruning; bee algorithm; mammogram; breast cancer

I. INTRODUCTION

Breast cancer is the most widely recognized dangerous cancer among ladies and the second driving reason for death [1]. According to the study from the World Health Organization, breast cancer is the most diagnosed cancer after lung cancer (10.9% of cancers in men and women is breast cancer), and the second driving reason for death [1], i.e., in 2008, 458 503 women worldwide died because of breast cancer [2].

Although no effective ways to prevent breast cancer exist, the early discovery of breast cancer is deemed significant for the decrease of associated death. Therefore, considerable effort is focused on cancer diagnosis at early stages. One of the most useful tools in early detection and diagnosis of breast cancer is the imaging technologies and mammography [3].

Mammography is the popular technique designed to image the breast [3]. In mammography, the most critical symptoms of breast cancer are masses. Mass type diagnosis (benign or malignant) in mammography images is a challenging issue for radiologists.

CAD helps in diagnosis and identification of masses. Typical CAD system consists of three stages as shown in Fig. 1; Pre-processing, which includes image segmentation, feature extraction, and classification into normal, benign, and malignant. In image segmentation stage, a mammogram will be segmented to extract Region of Interest (ROI). ROI maybe extracted manually [4]–[9] or it can be extracted automatically using any segmentation method [10]–[14].

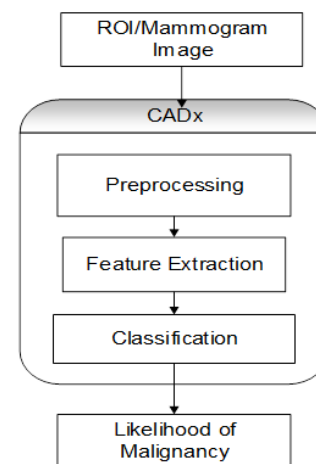


Fig. 1. General Steps of CADx.

The second stage is the feature extraction, which plays an important role for achieving high performance in the classification stage. This can be only achieved by extracting the suitable features that describe whether the suspicious region is benign or malignant. One of the most popular types of extracted features is the texture features.

In classification stage, the appropriate classifier model is trained using the training samples with the extracted features and then used to predict the class of the unseen pattern. Classifier design in the classification stage of CAD system is one of the key steps to get higher performance. Most of researchers have been focus in using a single classifier such as the Support Vector Machine (SVM) [6], [7], [15], Multi-Layer Perceptron (MLP) [5], [7], Linear Discriminant Analysis (LDA) [4], [7], Decision Tree (DT) [6], and K-Nearest Neighbour (KNN) [6]. However, due to the heterogeneity in the mammogram, single classifiers make errors on different samples. Thus, the ensemble learning in machine learning is used to improve the diagnosis of the mammogram. It has shown its ability in solving different classification problems as histopathology image grading [16]–[18], intrusion detection system [19] and breast cancer detection and diagnosis [20], [21].

Ensemble learning has become important because of its ability to improve the performance of single classifier system in theory and practice [22]. Generally, ensemble model consists of a group of independent trained base classifiers that are aggregated together in order to classify new samples. Fusing the outputs of all these base classifiers to get a final ensemble output is based on the aggregation rule. Using the appropriate base classifiers helps to improve the overall

ensemble results. Fig. 2 is showing the different between single classifier model and ensemble learning.

Combining all base classifiers in the ensemble model will increase the computational cost and time. Using subset of the ensemble can outperform the whole ensemble. Therefore, selecting the member of ensemble subset (ensemble pruning) is an important step.

Although, choosing the member of ensemble is an important issue, the challenge here is how to decide the criteria for selecting ensemble members [23]. This means, to have a successful ensemble, balancing between diversity and accuracy of the ensemble members should be achieved. To choose the best ensemble, optimization methods like genetic algorithm, artificial bee colony have been implemented [19], [24]–[27]. Fig 3 is showing ensemble selection using GA.

Bee Algorithm (BA) was firstly proposed by Pham et al. [28]. It is a population based algorithm inspired by natural behaviour of food searching of honeybee. Generally, BA has three important steps; initialization, local search, and global search. At the beginning, scout bees are distributed in random uniform in the search space starting to explore it. Then, recruited bees start exploiting and searching in the most promising areas identified by the scout bees. BA was applied and showed a good performance to problems in many fields, such as multi-level thresholding [29].

In this study, optimization of ensemble learning is emphasized. Two criteria of ensemble are used; 1) The generation of multiple heterogeneous base classifier, so that diversity among them is ensured, and 2) The use of BA optimization method in pruning ensemble.

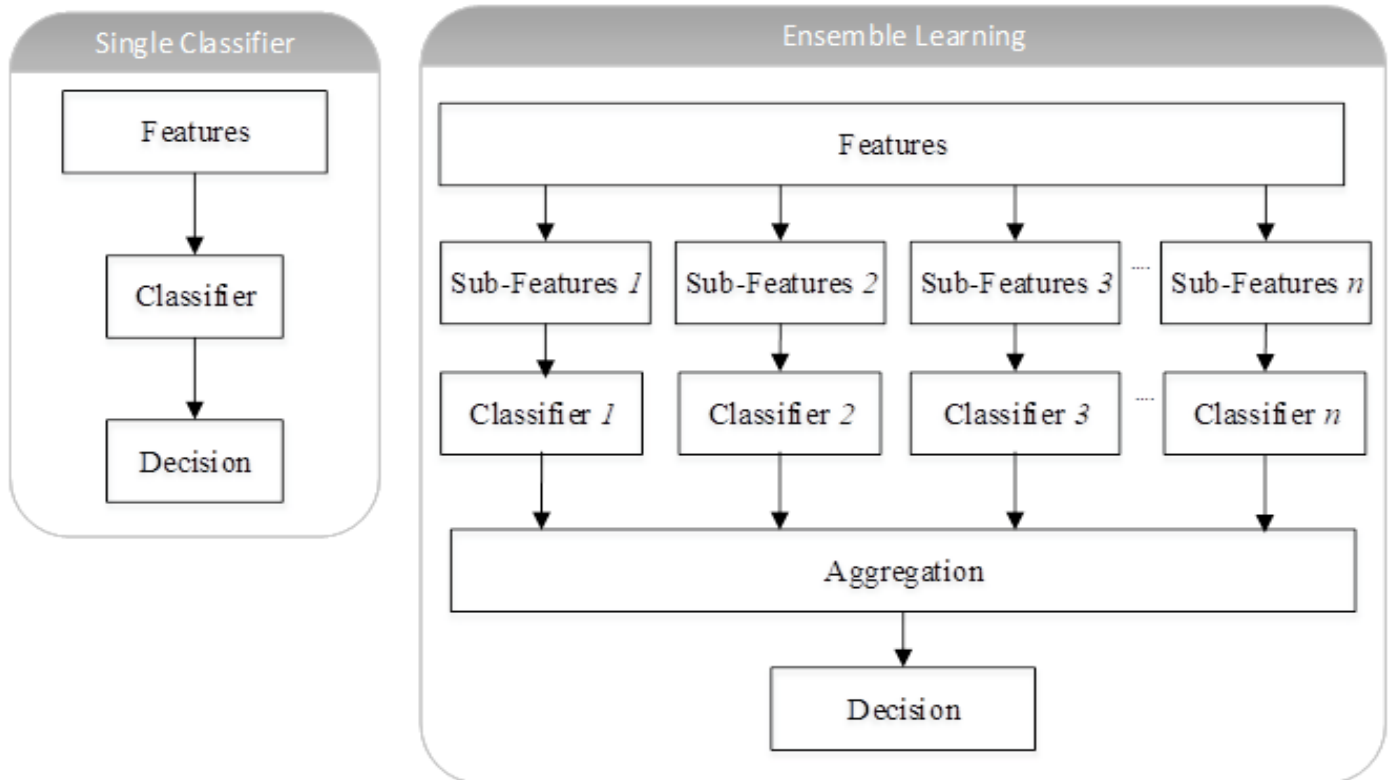


Fig. 2. Ensemble Learning Vs Single Classifier.

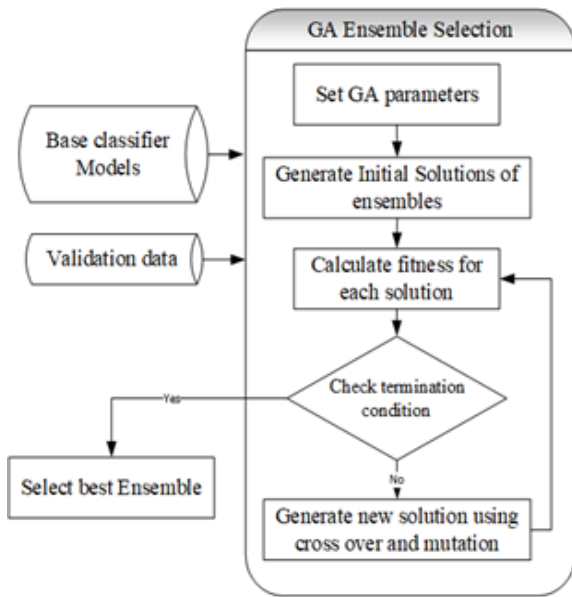


Fig. 3. Ensemble Selection using GA.

II. MATERIAL AND METHOD

In this paper an ensemble pruning method using bee algorithm for breast cancer classification is proposed. First of all, the mammogram Region of Interest (ROI), where the mass is in its centre, is manually cropped and resized into 512×512. Different features extracted from each ROI. Then, about twelve categories of features has been extracted as shown in Table I.

The extracted features were used to build the ensemble model. Our ensemble classifier system consists of three phases; pool generation, ensemble pruning, and ensemble testing. The key here is to select the optimal number of ensemble committee members by using optimization method. BA is used in order to select the best number of classifier that used to build an ensemble that ensure to achieve the better performance

TABLE I. FEATURES GROUPS

| | Features | No. of Features |
|----|---|-----------------|
| 1 | Gray Level Co-Occurrence Matrix (GLCM) | 38 |
| 2 | histogram of oriented gradients (HOG) | 81 |
| 3 | Local Binary Pattern (LBP) | 256 |
| 4 | First Order Statistics (FOS) | 5 |
| 5 | Haralick Spatial Gray Level Dependence Matrices (SGLDM) | 25 |
| 6 | Gray Level Difference Statistics (GLDS) | 4 |
| 7 | Neighborhood Gray Tone Difference Matrix (NGTDM) | 5 |
| 8 | Statistical Feature Matrix (SFM) | 4 |
| 9 | Laws Texture Energy Measures (TEM) | 6 |
| 10 | Fractal Dimension Texture Analysis (FDTA) | 4 |
| 11 | Fourier Power Spectrum (FPS) | 2 |
| 12 | Shape | 2 |

A. Pool Generation

Firstly, best base classifier is generated using all groups of features. Unlike Choi et al. [21], who depended on single base classifier in the generation process, multi-based classifier was considered. Fig. 4 shows how to generate multiple based classifier.

In details, dataset, after extracting features, is divided into three sets; 1) training set, which is used in generation phase; 2) validation set, which is used in the selection phase, and; 3) testing set, which is used to evaluate the selected ensemble.

In the generation phase, feature groups, as shown in table I, were used to build classifier model using different based classifiers. SVM, KNN, and LDA are considered here as base classifiers. SVM has been chosen due to its robustness in handling overfitting when there is no balancing between the training sample and features. KNN which is considered as a lazy classifier, is used here because it is simplicity and low bias, while LDA is good for correlation.

Assuming we have k group of features and n base classifiers then $k \times n$ classifier models are built. Then the best model is selected at each iteration based on the data sampling process. The outcomes of this phase are M best classifier models which will be the input of the pruning phase.

Let F represents the features pool $F = \{f_i, i = 1, \dots, k\}$, C represents the classifier pool $C = \{c_i, c = 1 \dots n\}$, T represents the training samples $T = \{(x_i, l_i)\}_{i=1}^N$, where class label is $l_i \in \{0,1\}$ and N is the number of training sample, and M is the number of boosting rounds.

Based on Freund and Schapire [30], the initial distribution $D_0(x_i) = 1/N$, whereas for each m -th round, the distribution $D_m(x_i)$ on the training sample x_i can be calculated as follow:

$$D_m(x_i) = \frac{w_{m,i}}{\sum_{i=1}^N w_{m,i}} \text{ for } i = 1 \dots N \quad (1)$$

where $w_{m,i}$ is the weight for the i -th training sample at the m -th boosting round.

We apply random resampling with Uniformed training data for obtaining different subsample in each M round.

Let $C_{m,k,n}$ is the n -th base classifier c_n trained with k -th feature f_k and m -th resample subset. So for each m round we are going to select the best base classifier C_m which gives the most accurate result among the weighted training set as follow:

$$C_m = \min_{h_{m,k,n}} \epsilon_{h_{m,k,n}} \quad (2)$$

$$\epsilon_{C_{m,k,n}} = \sum_{i=1}^N D_m(x_i) |C_{m,k,n}(x_i) - l_i| \quad (3)$$

where, D is a weighted distribution, N number of training samples, l is the training label and C is the classifier output.

The idea here is that different classifier trained by different features will make the mass that cannot be recognized by one representation easily detected by another. Using multiple features with different base classifiers increases ensemble diversity between ensemble members and avoid making coincident errors.

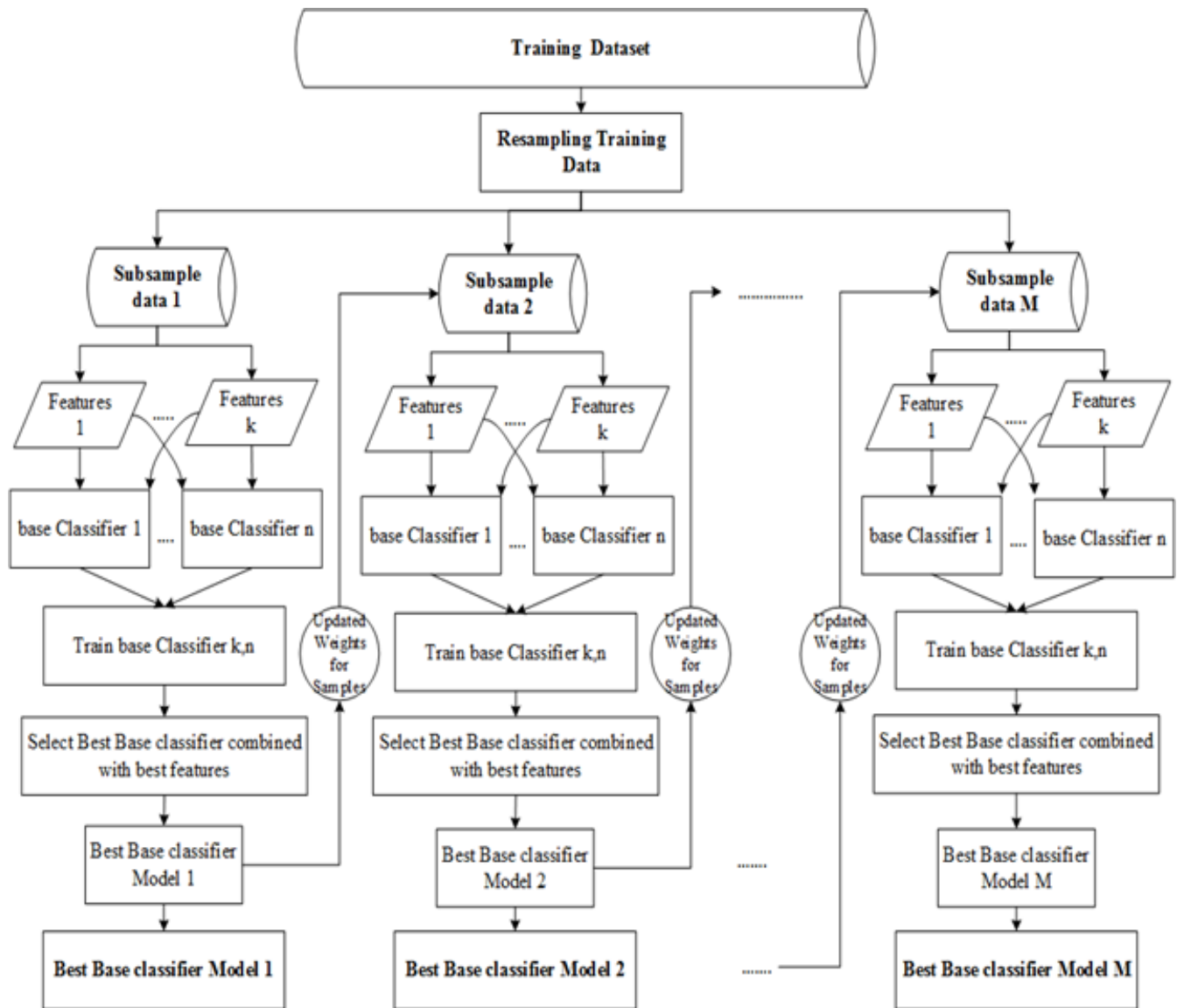


Fig. 4. Adapted Ensemble Generation.

$$\beta_t = \varepsilon_t / (1 - \varepsilon_t) \quad (4)$$

where ε is the error of the best classifier.

$$a_t = \frac{1/\beta_t}{\sum_{t=1}^M 1/\beta_t} \quad (5)$$

B. Ensemble Pruning (Selection)

In the selection phase, M best classifier models will be used to build the ensemble using the validation dataset. BA was used for selecting models to be included in the ensemble. The output of this phase is the best ensemble solution and it will be evaluated using the testing data set in the testing phase.

Fig. 5 illustrates the combination the generated based classifiers to build the ensemble models and to select the optimal ones.

Assuming that there are n trained based classifier in the pool, therefore, the search space is 2^n . Corresponding to that, BA was used.

1) BA for ensemble pruning

The first step in BA is the initialization via randomly uniform. This initialization returns continuous values between $[0,1]$. To convert it to discrete, thresholding is applied. For each scout bee, initialization as an ensemble randomly created by combining various number of based classifier. This random combination is done based on the discretization process for the random generated numbers. Fig. 6 shows the steps of BA. For example, initial position $ipos$ is created by generating n random numbers in between $[0,1]$, where n is the pool size. Then, discretization is applied using threshold $0 < Th < 1$ using equation 6:

$$DePos_i = \begin{cases} 0 & ipos < Th \\ 1 & ipos \geq Th \end{cases} \quad (6)$$

where $DePos_i$ is the descritized position of i based classifier, $i \in [1,2,3, \dots, n]$, $ipos$ is the initial position and Th is the threshold value.

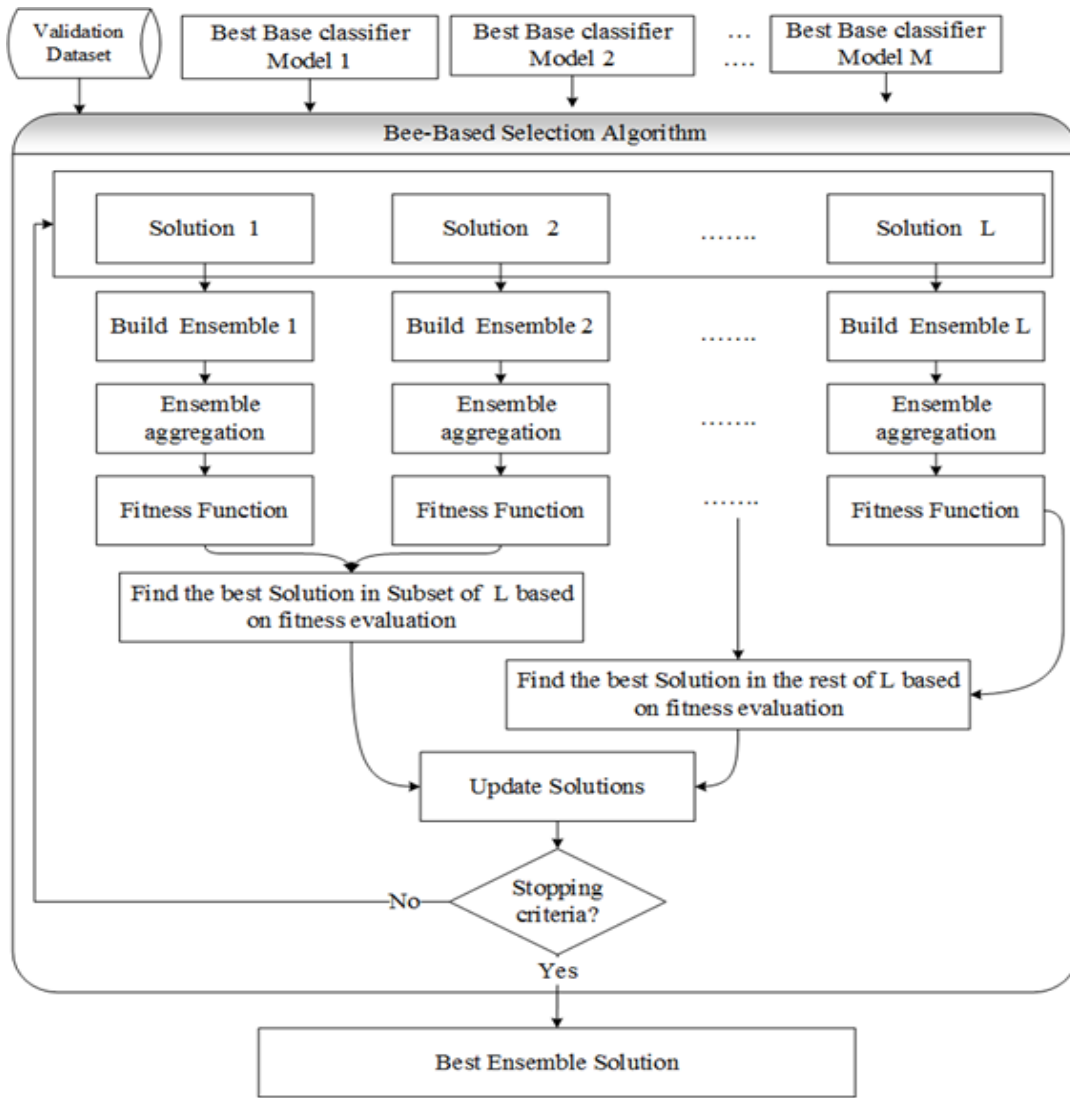


Fig. 5. The Proposed Ensemble Pruning using BA.

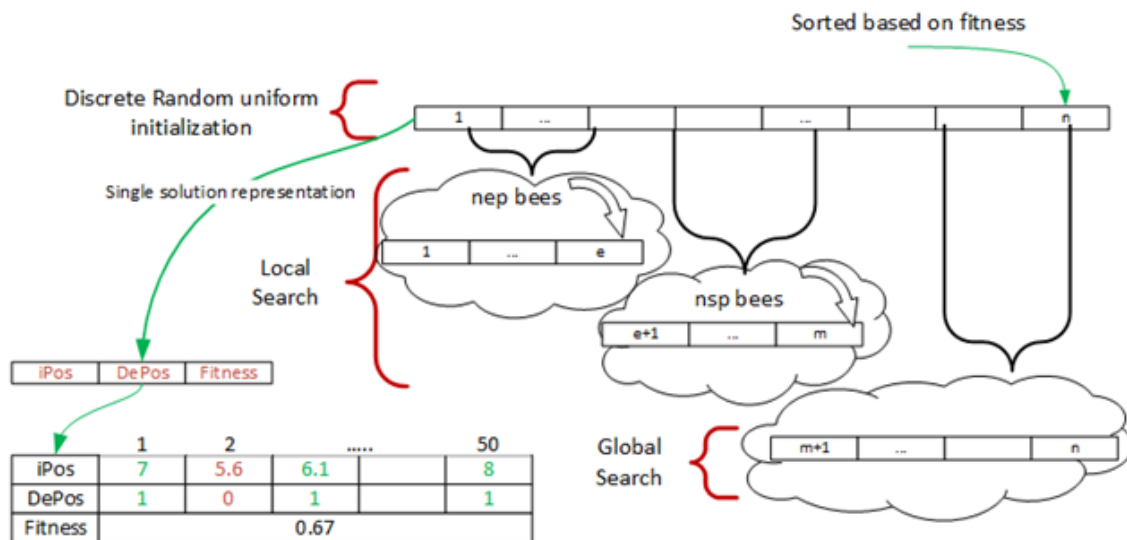


Fig. 6. The Proposed Ensemble Pruning Process using BA, where N is Number of Scout Bees, M is the Selected Sites, E is the Elite Sites, Nsp is Recruited Bees for Selected Sites and Nep is Recruited Bees for Elite Sites.

Eventually, The ensemble is created based on the $DePos$ vector. For each based classifier in th pool C_i

$$C_i \rightarrow \begin{cases} \text{Part of ensmeble} & DePos_i = 1 \\ \text{Out of ensmeble} & DePos_i = 0 \end{cases} \quad (7)$$

So that the initial ensemble for each scout bee can be define as below

$$\text{Initial Ensemble} = \{C_i\}_{i=1}^M \quad (8)$$

where $DePos_{i=1}^M = 1$, M is a variable number representing the number of ensemble members.

After initial ensemble is created, fitness function of each ensemble is calculated based on the performance of the ensemble. The performance of the ensemble is considered by both AUC and diversity of the ensemble. As stated in literature, there is no unique measure for diversity, in this work, calculation of ensemble fitness is done based on Choi at el. [21] equation as follow:

$$\text{Ensemble finness} = \overline{AUC} - \lambda \overline{Div} \quad (9)$$

where λ is the controlling variable, \overline{AUC} and \overline{Div} are calculated as:

$$\overline{AUC} = (AUC - AUC_{min}) / (AUC_{max} - AUC_{min}) \quad (10)$$

$$\overline{Div} = (Div - Div_{min}) / (Div_{max} - Div_{min}) \quad (11)$$

where, AUC_{min} and AUC_{max} are the minimum and maximum AUC of ensemble members, Div_{min} and Div_{max} are the minimum and maximum diversity between pairs members in the ensemble. AUC is calculated as:

$$AUC = AUC_{ens} + (AUC_{ens} - AUC_{mean}) \quad (12)$$

where, AUC_{ens} is the ensemble AUC based on Majority Voting (MV) aggregation and,

$$AUC_{mean} = \frac{1}{M} \sum_{i=1}^M AUC_i \quad (13)$$

where, M is the number of selected based-classifier in the ensemble. AUC_i the value of AUC obtained using i -th classifier.

Div is calculated based on three diversity pair wised measures, Q statistic, correlation coefficient and double-fault as follow:

$$Div = Q_{avg} + \rho_{avg} + DF_{avg} \quad (14)$$

where, Q_{avg} , ρ_{avg} , and DF_{avg} are the average values of the statistic, correlation coefficient and double-fault respectively and could be calculated as:

$$(Q, \rho, DF)_{avg} = \frac{2}{M(M-1)} \sum_{i=1}^{M-1} \sum_{j=i+1}^M (Q, \rho, DF)_{i,j} \quad (15)$$

TABLE II. CONTINGENCY TABLE

| | C_i correct | C_i wrong |
|---------------|---------------|-------------|
| C_j correct | a | b |
| C_j wrong | c | d |

The Q, ρ and DF are calculated between two paired classifiers C_i , and C_j based on Table II. The contingency table is used when calculating diversity between two classifiers i, j for the same testing data. The values a, b, c and d have different meaning where a is the number of instance in the testing data that correctly classified by both classifier i, j , while d is the number of instance in the testing data that are incorrectly classified by both classifier i, j ; c is the number of instance in the testing data that are correctly classified by classifier i and misclassified by classifier j , and b is the number of instance in the testing data the misclassified by classifier i and correctly classified by classifier j .

The Q statistic for two classifiers C_i and C_j is:

$$Q_{t,m} = \frac{\bar{a}\bar{d} - \bar{b}\bar{c}}{\bar{a}\bar{d} + \bar{b}\bar{c}} \quad (16)$$

where, \bar{a} is the propability that both classifiers classify the instance correctly. b, \bar{c} and \bar{d} can be defined in the same way:

$$\bar{a} = a/N, \bar{b} = b/N, \bar{c} = c/N, \bar{d} = d/N,$$

where, N , is number of testing instance.

While, the correlation of two classifiers C_i and C_j is:

$$\rho_{t,m} = \frac{\bar{a}\bar{d} - \bar{b}\bar{c}}{\sqrt{(\bar{a}+\bar{b})(\bar{c}+\bar{d})(\bar{a}+\bar{c})(\bar{b}+\bar{d})}} \quad (17)$$

The double-fault measures the misclassified instances by both classifiers C_i, C_j . It is calculated as below:

$$DF_{t,m} = \bar{d} \quad (18)$$

After initialization of BA, local and global searches are started. The output of BA is the best k ensembles where k is the number of BA generation.

C. Ensemble Testing

After the selection process is finished, multiple solution is presented. The final solution is identified as the best among the best. This final ensemble will be tested using the test dataset. The aggregation of ensembles is done using MV.

III. RESULT AND DISCUSSION

A. Dataset Description

The dataset used here is a self-collected dataset from Hospital Kuala Lumpur (KHL). It is consisting of 236 mammogram ROI images. All images have either benign or malignant mass. The expert radiologist has gone through all images to evaluate them and identify the correct place of mass in each.

B. Experiment Setup

The experiment was run 10 times to see the stability of the model. In each run, double cross validation is applied to the data with $k=15$. In the generation stage, the pool size (M) was set to 50. So that 50 base classifiers are generated. Sampling with $r=30$ is applied to the training data.

The based classifier parameters for both single classifier approach and ensemble approach were similar. First, SVM tested by LibSVM toolbox with radial basis function kernel.

C and γ was calculated using grid search with range $[2^{-5}, 2^{10}]$ [31]. Then, KNN classifier was used with the default $k = 1$. Regulation used in LDA with γ interval $[0,1]$.

For BA pruning ensemble, parameters set to their default and most used values; 10 iterations, 30 scout bees, 15 selected site, 12 elite sit, 15 recruited bees for selected sites and 30 recruited bees for elite sites. The threshold value (Th) for discretization of BA initialization was set to 6. GA parameters are set as A. Ekbal and S. Saha [27] with population size of 50 and generation is set to 100. Same fitness function used for GA and BA based on performance and diversity of the ensemble as equation (9) with controlling variable $\lambda = 0.25$ as mentioned in Choi et al. [21]. MV is used in ensemble aggregation.

C. Evaluation Matrices

To evaluate the proposed method, confusion matrix was calculated in order to calculate the accuracy, specificity and sensitivity.

Sensitivity is the ability of the model for detecting malignant masses, and it can be calculated based on the following equation:

$$Sensitivity(SE) = \frac{TP}{TP+FN} \times 100 \tag{19}$$

where TP is the number of corrected malignant diagnostics instance and FN is the malignant instance that was wrongly diagnosed as benign.

Specificity is representing the ability of the model to identify the benign masses, and it is calculated as:

$$Specificity(SP) = \frac{TN}{TN+FP} \times 100 \tag{20}$$

where, TN is the number of benign instance that classified correctly and FP is the the number of benign instance that was misclassified.

The accuracy of the model is describing its ability of detecting both benign and malignant.

$$Accuracy(AC) = \frac{TP+TN}{TP+FP+TN+FN} \times 100 \tag{21}$$

On the other hand, AUC calculated the area under Receiver operating characteristic ROC curve, which draws the relation between sensitivity rate and specificity rate.

D. Result of Proposed Heterogeneous BA Ensemble Pruning

The objective of this experiment is to show the output of the proposed ensemble. Table III shows the average result with standard deviation of 10 run of the proposed BA heterogeneous ensemble pruning using different aggregation method.

TABLE III. RESULT OF PROPOSED BA ENSEMBLE PRUNING FOR ONE RUN USING DIFFERENT AGGREGATION METHOD

| | MV | Product | Max | Sum | Min | Average |
|-----|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| AUC | 0.79 ±0.13 | 0.79 ±0.08 | 0.79 ±0.8 | 0.80 ±0.12 | 0.79 ±0.08 | 0.80 ±0.12 |
| AC | 0.85 ±0.09 | 0.74 ±0.09 | 0.74 ±0.09 | 0.85 ±0.09 | 0.74 ±0.09 | 0.85 ±0.09 |
| SP | 0.94 ±0.05 | 0.66 ±0.13 | 0.66 ±0.13 | 0.94 ±0.06 | 0.66 ±0.13 | 0.94 ±0.06 |
| SE | 0.64 ±0.25 | 0.92 ±0.09 | 0.92 ±0.09 | 0.66 ±0.24 | 0.92 ±0.09 | 0.66 ±0.24 |

TABLE IV. EXAMPLE RESULT OF THE ENSEMBLE MEMBER AUC, CLASSIFIER AND FEATURE GROUP OF THE PROPOSED BA ENSEMBLE PRUNING OF ONE CROSS VALIDATION ITERATION

| | Ensemble Members | | | | | | | | | | | | | | |
|------------|------------------|-------|-------|-------|-------|-------|-------|-------|-------|--------|--------|--------|--------|--------|--------|
| | Mem 1 | Mem 2 | Mem 3 | Mem 4 | Mem 5 | Mem 6 | Mem 7 | Mem 8 | Mem 9 | Mem 10 | Mem 11 | Mem 12 | Mem 13 | Mem 14 | Mem 15 |
| Single AUC | 0.86 | 0.90 | 0.90 | 0.95 | 0.90 | 0.80 | 0.90 | 0.90 | 0.90 | 1.00 | 0.90 | 0.90 | 0.90 | 0.90 | 0.90 |
| classifier | LDA | LDA | SVM | KNN | LDA | LDA | LDA | SVM | KNN | KNN | LDA | LDA | LDA | LDA | SVM |
| Feature | SFM | GLD S | LBP | LBP | TEM | GLC M | TEM | LBP | GLD S | GLCM | GLCM | TEM | GLDS | GLCM | LBP |

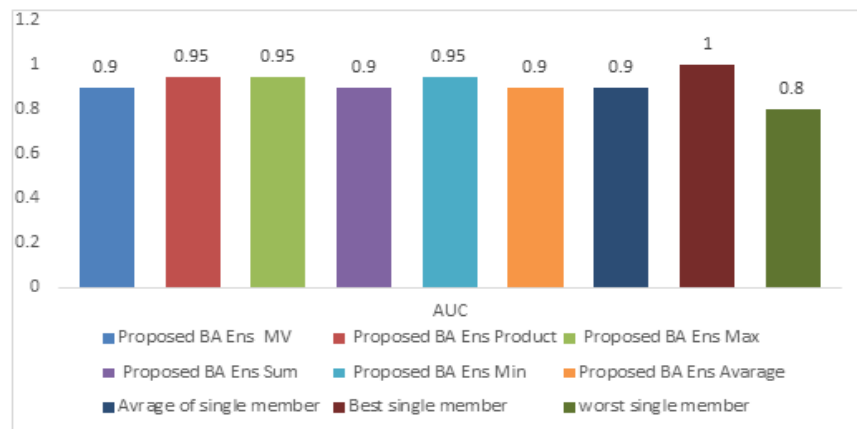


Fig. 7. Comparison between AUC of Proposed BA Ensemble and its Members AUC of a One Cross Validation Iteration.

Table IV is showing an example of an ensemble member after applying the proposed BA pruning method.

We can notice in Table IV the ensemble contains various classifiers as well as features groups. It might be noticed that there is a repetition in term of similar classifier using the same feature group. However, each classifier was actually trained by different training samples. The subsampling steps applied in the generation process guaranteed that all the trained classifiers inside the pool are different from each other's in terms of classifier model or feature group or training sample. This variation will maintain the diversity of the ensemble members.

Moreover, Fig. 7 is comparing with the AUC of the proposed ensemble framework using the 6 aggregation method compared to the AUC of ensemble members; best, average, and worst. The performance of the proposed BA ensemble can outperform the average of its members.

As this is heterogeneous ensemble, Fig. 8 shows comparison between the proposed ensemble frameworks with single classifier framework. The single classifier framework was built by using the three based classifiers used in the ensemble; SVM, KNN, and LDA. For the single classifier, all features used in the ensemble were combined together. The proposed BA ensemble outperforms the three single classifiers in terms of average AUC and AC.

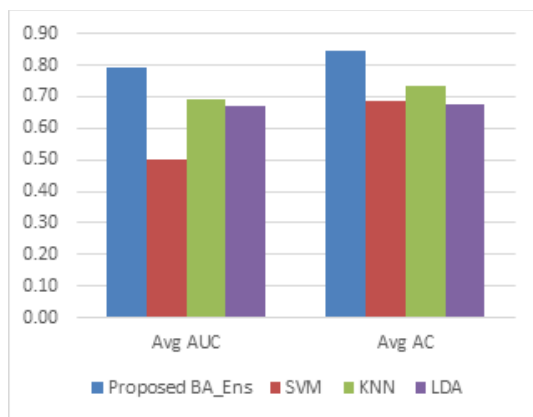


Fig. 8. Comparison between the Result of the Proposed BA Ensemble Pruning Framework and Single Classifier System.

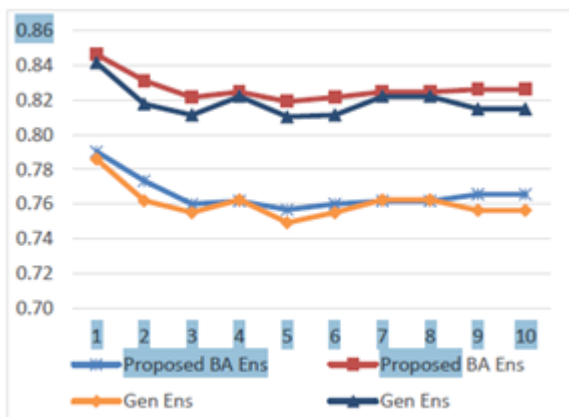


Fig. 9. Comparison of 10 Runs between Proposed Ensemble Pruning using BA and Ensemble Pruning using GA.

Furthermore, to show the effectiveness of using BA ensemble pruning, comparison between the proposed framework and ensemble pruning using genetic algorithm (GA) was performed. Fig. 9 shows that using BA outperformed GA in selection ensemble members. T-Test analysis result showed that the using of BA in ensemble pruning is statically significant than GA with $P < 0.00037$.

IV. CONCLUSIONS

In this paper, a novel ensemble pruning method based on BA optimization algorithm is presented to select the optimal ensemble base classifiers. This method assists the radiologists for diagnosing the malignity and benignity of the masses. The ROI was manually cropped. Ensemble generation process was done using three different classifiers and twelve group of features extracted directly from the ROI. Thresholding was applied to discretize BA initialization. AUC of the ensemble and diversity between ensemble's members were used to calculate the fitness function of BA. Different aggregation functions were used to test the result of the proposed method. The result showed that the proposed framework is stable. And comparing to single classifier frameworks and GA in ensemble pruning, using BA in ensemble pruning enhanced the quality of the ensemble learning.

ACKNOWLEDGMENT

We would like to thank the Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, for providing the facilities and financial support under "ETP-2013-053 diagnostic services nexus for breast cancer". We would also like to acknowledge Human Life Advancement Foundation (HLAF) for the financial support to the main researcher. We have obtained Ethics approval with number "UKM 1.5.3.5/244/FTSM-002-2015" entitled "Diagnostic services nexus for breast cancer" from UKM Medical Center, Malaysia, for collecting and conducting experiments on breast cancer patient records. Here, we would like and acknowledge Prof Fuad Ismail from Department of Radiotherapy and Oncology, UKM medical centre and Prof Norlia Abdullah from Department of Surgery, UKM medical centre for their guidance in preparing and applying ethical proposal. Finally, we got an ethical approval from Medical Research and Ethics Committee entitles of "Diagnostic services nexus for breast cancer" with registration No "NMRR-15-21800-27949(IRR)" to allow us collecting mammogram data from HKL.

REFERENCES

- [1] L. A. Torre, F. Bray, R. L. Siegel, J. Ferlay, J. Lortet-tieulent, and A. Jemal, "Global Cancer Statistics, 2012," *CA a cancer J. Clin.*, vol. 65, no. 2, pp. 87–108, 2015.
- [2] X. Liu and J. Tang, "Mass classification in mammograms using selected geometry and texture features, and a new SVM-based feature selection method," *IEEE Syst. J.*, vol. 8, no. 3, pp. 910–920, 2014.
- [3] D. L. Monticciolo, M. S. Newell, R. E. Hendrick, M. A. Helvie, L. Moy, B. Monsees, D. B. Kopans, P. R. Eby, and E. A. Sickles, "Breast Cancer Screening for Average-Risk Women: Recommendations From the ACR Commission on Breast Imaging," *J. Am. Coll. Radiol.*, vol. 14, no. 9, pp. 1137–1143, 2017.
- [4] R. Rabidas, A. Midya, J. Chakraborty, and W. Arif, "A Study of Different Texture Features Based on Local Operator for Benign-malignant Mass Classification," *Procedia Comput. Sci.*, vol. 93, no. September, pp. 389–395, 2016.

- [5] A. Tahmasbi, F. Saki, and S. B. Shokouhi, "Classification of benign and malignant masses based on Zernike moments," *Comput. Biol. Med.*, vol. 41, no. 8, pp. 726–735, 2011.
- [6] M. Z. Do Nascimento, A. S. Martins, L. A. Neves, R. P. Ramos, E. L. Flores, and G. A. Carrijo, "Classification of masses in mammographic image using wavelet domain features and polynomial classifier," *Expert Syst. Appl.*, vol. 40, no. 15, pp. 6213–6221, 2013.
- [7] M. Abdel-Nasser, H. A. Rashwan, D. Puig, and A. Moreno, "Analysis of tissue abnormality and breast density in mammographic images using a uniform local directional pattern," *Expert Syst. Appl.*, vol. 42, no. 24, pp. 9499–9511, 2015.
- [8] Y. a S. Duarte, M. Z. Nascimento, and D. L. L. Oliveira, "Classification of mammographic lesion based in Completed Local Binary Pattern and using multiresolution representation," *J. Phys. Conf. Ser.*, vol. 490, p. 012127, 2014.
- [9] N. Azizi, Y. Tlili-guiassa, and N. Zemmal, "A Computer-Aided Diagnosis System for Breast Cancer Combining Features Complementarily and New Scheme of SVM Classifiers Fusion," *Int. J. Multimed. Ubiquitous Eng.*, vol. 8, no. 4, pp. 45–58, 2013.
- [10] B. W. Hong and B. S. Sohn, "Segmentation of regions of interest in mammograms in a topographic approach," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 1, pp. 129–139, 2010.
- [11] A. Rojas Domínguez and A. K. Nandi, "Detection of masses in mammograms via statistically based enhancement, multilevel-thresholding segmentation, and region selection," *Comput. Med. Imaging Graph.*, vol. 32, no. 4, pp. 304–315, 2008.
- [12] W. Xie, Y. Li, and Y. Ma, "Breast mass classification in digital mammography based on extreme learning machine," *Neurocomputing*, 2015.
- [13] D. Saraswathia and E. Srinivasan, "A high- sensitivity computer-aided system for detecting microcalcifications in digital mammograms using curvelet fractal texture features," *Comput. Methods Biomech. Biomed. Eng. Imaging Vis.*, no. September, 2016.
- [14] A. Qasem, S. Norul, H. Sheikh, and F. I. Shahnorbanun Sahran, Tengku Siti Meriam Tengku Wook, Rizwana Iqbal Hussain, Norlia Abdullah, "Breast Cancer Mass Localization Based on Machine Learning," in 2014 IEEE 10th International Colloquium on Signal Processing & its Applications (CSPA), 2014, pp. 31–36.
- [15] A. Qasem, S. Norul, H. Sheikh, S. Sahran, and F. Ismail, "An Accurate Rejection Model for False Positive Reduction of Mass Localisation in Mammogram," *Pertanika J. Sci. Technol.*, vol. 25, no. S6, pp. 49–62, 2017.
- [16] D. Albashish, S. Sahran, A. Abdullah, N. A. Shukor, and S. Pauzi, "Ensemble Learning of Tissue Components for Prostate Histopathology Image Grading," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 6, no. 6, pp. 1134–1140, 2016.
- [17] D. Albashish, S. Sahran, and A. Abdullah, "Lumen-Nuclei Ensemble Machine Learning System for Diagnosing Prostate Cancer in Histopathology Images," *Pertanika J. Sci. Technol.*, vol. 25, pp. 39–48, 2017.
- [18] S. Sahran, D. Albashish, A. Abdullah, N. A. Shukor, and S. Hayati Md Pauzi, "Absolute cosine-based SVM-RFE feature selection method for prostate histopathological grading," *Artificial Intelligence in Medicine*, vol. 87, pp. 78–90, 2018.
- [19] A. A. Aburomman and M. Bin Ibne Reaz, "A novel SVM-kNN-PSO ensemble method for intrusion detection system," *Appl. Soft Comput. J.*, vol. 38, pp. 360–372, 2016.
- [20] Y. Zhang, N. Tomuro, J. Furst, and D. S. Raicu, "Building an ensemble system for diagnosing masses in mammograms," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 7, no. 2, pp. 323–329, 2012.
- [21] J. Y. Choi, D. H. Kim, K. N. Plataniotis, and Y. M. Ro, "Classifier ensemble generation and selection with multiple feature representations for classification applications in computer-aided detection and diagnosis on mammography," *Expert Syst. Appl.*, vol. 46, pp. 106–121, 2016.
- [22] N. V. Chawla and J. Sylvester, "Exploiting Diversity in Ensembles: Improving the Performance on Unbalanced Datasets," *Int. Work. Mult. Classif. Syst. Springer Berlin Heidelberg*, vol. 4472, pp. 397–406, 2007.
- [23] H. Parvin, M. Mirnabibaboli, and H. Alinejad-Rokny, "Proposing a classifier ensemble framework based on classifier selection and decision tree," *Eng. Appl. Artif. Intell.*, vol. 37, pp. 34–42, 2015.
- [24] A. Peimankar, S. J. Weddell, T. Jalal, and A. C. Laphorn, "Ensemble Classifier Selection Using Multi-Objective PSO for Fault Diagnosis of Power Transformers," *Evol. Comput.*, vol. IEEE Congr, pp. 3622–3629, 2016.
- [25] E. Parhizkar and M. Abadi, "BeeOWA: A novel approach based on ABC algorithm and induced OWA operators for constructing one-class classifier ensembles," *Neurocomputing*, vol. 166, pp. 367–381, 2015.
- [26] L. Shi, L. Xi, X. Ma, M. Weng, and X. Hu, "A novel ensemble algorithm for biomedical classification based on Ant Colony Optimization," *Appl. Soft Comput. J.*, vol. 11, no. 8, pp. 5674–5683, 2011.
- [27] A. Ekbal and S. Saha, "Classifier Ensemble Selection Using Genetic Algorithm for Named Entity Recognition," *Res. Lang. Comput.*, vol. 8, no. 1, pp. 73–99, 2010.
- [28] D. T. Pham, A. Ghanbarzadeh, E. Koç, S. Otri, S. Rahim, and M. Zaidi, "The Bees Algorithm - A Novel Tool for Complex Optimisation Problems," *Intell. Prod. Mach. Syst. - 2nd I*PROMS Virtual Int. Conf. 3-14 July 2006*, no. August 2015, pp. 454–459, 2006.
- [29] W. A. Hussein, S. Sahran, and S. N. H. S. Abdullah, "A fast scheme for multilevel thresholding based on a modified Bees Algorithm," *Knowledge-Based Syst.*, vol. 101, pp. 114–134, 2016.
- [30] Y. Freund and R. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Comput. Syst. Sci.*, vol. 55, pp. 119–139, 1997.
- [31] C.-J. L. Chang, Chih-Chung, "LIBSVM: a library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, p. 27, 2011.