

- [Register](#)
- [Current](#)
- [Login](#)
- [Archives](#)
- [Announcements](#)
- [About](#)
  - [About the Journal](#)
  - [Submissions](#)
  - [Editorial Team](#)
  - [Privacy Statement](#)
  - [Contact](#)

[Search](#)

- [Register](#)
- [Login](#)

1. [Home](#) /
2. [Archives](#) /

### 3. **Vol 6 No 1 (2021): FUOYE Journal of Engineering and Technology Vol. 6 Iss. 1 (March 2021 issue)**



FUOYE Journal of Engineering and Technology Vol. 6 Iss. 1 (March 2021 issue) Official date of Publication: 31st March 2021  
Published: 2021-03-23

## Articles

- [Effect of Tillage Direction on Phosphorus Loss from Sediment Yield and Surface Runoff](#)  
Temitayo A Ewemoje, Ibrahim A Kutu
  - [PDF](#)
- [Effect of Pretreating Poultry Wastes with Cocoa Pod Husk Ash and Cassava Peel Ash on Biogas Production](#)  
Adewole A Aderinlewo, Obukomena O Akivie, Olayemi J Adeosun, Peter O Omotainse
  - [PDF](#)
- [Changes in Proximate Composition, Vitamin C and  \$\beta\$ -Carotene Contents of Oven Dried Pawpaw \(Carica papaya\) Fruit as Influenced by Pretreatment Methods](#)  
Lydia A Adepoju, Zinash D Osunde, Kehinde J Falua
  - [PDF](#)
- [Empirical Assessment of Missing Link in Agricultural Technology Development and Transfer Nexus in Nigeria](#)  
Sunday I Ogunjimi, Toyin F Ojo, Ibikunle E Obabire, Olajumoke O Alabi
  - [PDF](#)
- [An Improved Vehicle Registration and Licensing System](#)  
Abass O Adisa, Safiriyu I Eludiora
  - [PDF](#)
- [Nature-Inspired Meta-heuristic Optimization Algorithms for Breast Cancer Diagnostic Model: A Comparative Study](#)  
Tinuke O Oladele, Babatunde J Olorunsola, Taye O Aro, Hakeem B Akande, Oyenike A Olukiran
  - [PDF](#)
- [A Review on the Detection of Offensive Content in Social Media Platforms](#)  
Solomon Akinboro, Oluwadamilola Adebuseye, Akintoye Onamade
  - [PDF](#)
- [Rain Attenuation Prediction Models in Microwave and Millimeter Bands for Satellite Communication System: A Review](#)  
Hammed O Busari, Olaosebikan A Fakolujo
  - [PDF](#)
- [Effect of Clusters in Energy-Efficient Cooperative White Space Detection in a Cognitive Radio System](#)  
Samson I Ojo, Zachaeus K Adeyemo, Damilare O Akande, Rebecca O Omowaiye
  - [PDF](#)
- [Estimation of attenuation due to rain within Ka and Ku bands in Oyo State of Nigeria](#)  
Hammed O Busari, Olaosebikan A Fakolujo
  - [PDF](#)
- [Design and Construction of a Low-Cost Automatic Single-Phase Transfer Switch](#)  
Abiodun E Amoran, Gerald K Ijamaru, Olaitan Akinsanmi, Watson Agbramu
  - [PDF](#)
- [Adaptive Power Control with Selection Combiner over Nakagami-m Fading Channel](#)  
Robert O Abolade, O O Oyedokun, Samson I Ojo, Zachaeus K Adeyemo, I A Ojerinde
  - [PDF](#)

- [Investigating Possibilities of fuelling NIRR-1 with Low Enriched Uranium Silicide-Aluminum Dispersion Fuels](#)  
Abdulhameed Salawu, Asuku S Shuaibu, Memunat Tairu, Medinat O Olabode, Bright O Agada, Temitope J Adegoke  
◦ [PDF](#)
- [Characterization of Gosa Municipal Solid Wastes at Abuja, Nigeria](#)  
Kamoru Seidu, Ibrahim D Muhammad, Idris I Ozigis  
◦ [PDF](#)
- [Effect of Elevated Temperatures on SAC305 Solder Alloy Thermal Interface Material in a Microelectronic Assembly](#)  
Mathias Ekpu  
◦ [PDF](#)
- [Performance of Some Biocoal Briquettes Produced from Mixture of Sawdust and varying Nigeria Coals as Composite Domestic Fuel](#)  
Joshua O Adekunle, Bayode J Olorunfemi, Samuel B Adejuyigbe  
◦ [PDF](#)
- [Influence of Particle Size and Particle Range Distribution on the Microstructure of Al-Mg-Si/PKSA Composite for Amateur Solid Rocket Chamber](#)  
Elijah O Oyedeki, Muhammed Dauda, Shehu A Yaro, Malik Abdulwahab  
◦ [PDF](#)
- [The Development and Characterisation of Recycled Polyethylene Reinforced Particulate Cockle Shell for Automobile Application](#)  
Johnson O Agunsoye, Chioma I Madueke, Adeola A Bamigbaiye, Kenneth C Agbakansi, Alice Oladipo  
◦ [PDF](#)
- [Analysis of Processing Hydrothermal Liquefaction of Microalgae Reaction Product with and without Dichloromethane Solvent for Biocrude Recovery](#)  
Blessing E Eboibi  
◦ [PDF](#)
- [An Improved Data-Driven Model for the Prediction of Minimum Transport Condition for Sand Transport in Multiphase Flow Systems](#)  
Adegboyega B Ehinmowo, Oluwadamilare O Ariyo, Oluwatosin A Ohiro, Olawale T Fajemidupe, Kazeem K Salam  
◦ [PDF](#)
- [Fitting Rating Curves to Selected Streams in Southeast Nigeria](#)  
Andy O Ibeje  
◦ [PDF](#)
- [Assessment of Spatio-temporal levels of pollution parameters due to dumpsite leachate in Lagos, Southwest Nigeria](#)  
Geoffrey E Nwosu, Adedayo A Badejo, Adebola A Adekunle, Olayemi J Adeosun  
◦ [PDF](#)
- [Experimental and Statistical Study on Black Cotton Soil Modified with Cement-Iron Ore Tailings](#)  
Paul Yohanna, Ianna M Kanyi, Roland K Etim, Oshioname A Ebere, Kolawole J Osinubi  
◦ [PDF](#)

Information

- [For Readers](#)
- [For Authors](#)
- [For Librarians](#)

**FUOYE Journal of Engineering and Technology**



# Nature-Inspired Meta-heuristic Optimization Algorithms for Breast Cancer Diagnostic Model: A Comparative Study

\*<sup>1</sup>Oladele T. Omolewa, <sup>1</sup>Olorunsola B. Julius, <sup>2</sup>Aro T. Oladele, <sup>3</sup>Akande H. Babalola and <sup>2</sup>Olukiran O. Adunni

<sup>1</sup>Department of Computer Science, University of Ilorin, Ilorin, Nigeria

<sup>2</sup>Department of Mathematical and Computing Sciences, KolaDaisi University Ibadan, Oyo State, Nigeria

<sup>3</sup>Department of Telecommunication Science, University of Ilorin, Ilorin, Nigeria

{tinuoladele|juliusbabatunde140|taiwo774|hakeembabs|olukiran.oyenike}@gmail.com

Received: 07-JAN-2021; Reviewed: 02-MAR-2021; Accepted: 12-MAR-2021

<http://dx.doi.org/10.46792/fuoyejt.v6i1.598>

**Abstract-** The selection of features is used to obtain a subset of features by the removal of irrelevant features with no or less predictive output. Meta-heuristic algorithms are appropriate for the selection of features because feature subset representation is direct and the evaluation is easily accomplished. This paper performed a comparative study on the impact of meta-heuristic optimization algorithms on breast cancer diagnosis using Ant Colony Optimization (ACO) and Particle Swarm Optimization (PSO). The two feature selection algorithms were used to obtain the relevant attributes from the Wisconsin breast cancer (original) dataset. The selected attributes were passed to seven learning algorithms: Support Vector Machine (SVM), Decision Tree (C4.5), Naïve Bayes (NB), K Nearest Neighbour (KNN), Neural Network (NN), Logistic Regression (LR), and Random Forest (RF). The diagnostic model was evaluated based on accuracy, precision, recall, and F1-measure. Experimental evaluation showed that the highest accuracy of 97.1388% was obtained in both PSO and ACO using RF classifier, the highest precision value of 0.9720 was recorded in ACO using RF classifier, the highest recall value of 0.9750 was achieved in PSO using RF classifier, the highest F1-measure value of 0.9700 was obtained in PSO using SVM, the highest kappa statistic of 0.9370 was obtained in both PSO and ACO using RF and the lowest time of 0s that was taken to build a model was recorded in PSO using KNN and NB, and also in ACO using KNN. The paper concluded that the breast cancer diagnostic model using PSO and ACO with different learning algorithms revealed that the accuracy of RF outperformed other algorithms. Also, it was shown that ACO produced better precision using RF compared with PSO and PSO gave better recall using RF compared with ACO, PSO recorded an efficient F1-measure using SVM. The best time used to build a model was obtained in PSO for KNN and NB, and ACO with KNN.

**Keywords-** Breast cancer, Data mining, Diagnosis, Feature selection, Meta-heuristic

## 1 INTRODUCTION

Cancer of the breast is a sickness in which cells in the breast region grow out of control. The type of breast cancer is determined by which cells in the breast turn into cancer. Breast cancer happens in females and not often in males. Indications of breast cancer include breast lump, discharge of blood from the nipple and variations in the shape or texture of the breast or nipple. Breast cancer diagnosis and prognosis pose a major challenge to researchers in the medical (Padmapriya, 2014). The medical experts are faced with various problems in the diagnosis of some diseases in which breast cancer is included. Several issues that often influence breast cancer prediction include: the existing predictive system is expensive and time-consuming, inadequate understanding of symptoms, risk factor, lack of quality diagnostic measure for the patient and no adequate information to properly predict the case of breast cancer in the society.

In the data mining approach for the diagnosis of diseases, data almost always contains more features needed or sometimes irrelevant features to build a predictive model (Sheena, Krishan & Gulshan, 2016). The use of irrelevant features in building a model may result in more memory space used during the training process and classification phase. The selection of features has been identified to be a dynamic area of research in machine learning and data mining techniques (Kursa, 2010). Meta-heuristic optimization techniques, among the feature selection techniques have been proved to be very effective in computation compared with other techniques like the filter, wrapper, and hybrid methods (Kumari & Swarnkar, 2011).

The selection of features is a general problem common to large datasets. To make a diagnostic model discriminative, it is necessary to employ nature-inspired meta-heuristic optimization algorithms or Swarm intelligence optimization such as Genetic algorithms (GA), Elephant Algorithm (EA), Lion Optimization Algorithm (LA), Simulated Annealing (SA), Firefly Algorithm (FA), Gravitational Search Algorithm (GSA), Ant Colony Optimization (ACO), and so on (Adi & Aldasht, 2018). These methods can be effective for this problem, which require less amount of computation time and memory. PSO is a computational approach that optimizes a problem by iteratively trying to improve a candidate solution with regard to a given measure of quality by iteratively trying to improve a candidate (Gao, Li, & Chen, 2019). Also, the ACO is a population-based metaheuristic algorithm that can be used to determine approximate solutions to difficult optimization problems (Vivek & Shetty, 2015).

This paper performed a comparative study using two meta-heuristic optimization techniques: ACO and PSO for feature selection on breast cancer dataset to show their impacts during breast cancer diagnosis. The remaining aspects of the paper are arranged as follows: related work, methodology, results and discussion, and conclusion.

## 2 RELATED WORK

In data mining, several studies have been carried out on breast cancer diagnosis. Singh, Choudhury, and De (2020) applied data mining to produce association rules, statistical and nature-inspired computational techniques for breast cancer diagnosis. Evolutionary algorithms were employed to obtain the best classifiers. The analysis was done in three phases, which included the training of complete dataset and creation of new tested data, ten-fold, 98% training and 2% tested data) were applied to

\*Corresponding Author

identify the type of disease using nine attributes of the breast cancer dataset. The result showed that a method was strategized to reduce the error of estimation in breast cancer data.

Punitha, Amuthan and Joseph (2019) developed effective breast cancer diagnosis. The study combined an improved Monarchy Butterfly Optimization Technique (IABC-EMBOT) and Artificial Bee Colony. The variation of Monarchy Butterfly Optimization that enhanced the degree of exploration applying the exploitation rate of the searching space was considered. Also, the concentration was on the removal of some ABC scheme problems by improving the probability of the process for search diversification via phenomenal updates simplified through the operator of a dynamic and adaptive butterfly. The model was evaluated using the Wisconsin dataset, an accuracy of 97.53% was obtained.

Darwish, Sayed and Hassanien (2018) presented a dual-step model for breast cancer diagnosis. Feature selection was achieved by the application of moth flame optimization, grey wolf optimizer, whale optimization and flower pollination. The selected features were passed to three classifiers: SVM, K-NN, and C4.5. The evaluation of each algorithm was achieved using different parameters. The experimental analysis with Wisconsin prognosis breast cancer (WPBC) and Wisconsin breast cancer diagnosis datasets showed the effectiveness of the model for feature selection of data and breast cancer classification.

Yeh and Chan (2017) considered swarm intelligence techniques for feature selection. An application to identify Regions of Interest (ROIs) on mammograms as normal tissue, benign or malignant was introduced. A linear polynomial kernel function of SVM was applied for classification. The ROIs contained 69 benign cases, 54 malignant masses, and 68 normal tissues. A total of 277 features were obtained from each region. GA, SA, ACO, and PSO were trained and tested employing 80% for training and 20% for testing. Results revealed that GA and PSO performed better than SA and ACO in feature selection. A combination of these techniques performed better than other techniques in obtaining the best features for classification of breast cancer.

Mazen, AbulSeoud and Gody (2016) applied genetic algorithm and firefly algorithm to enhance the weights between layers and biases of the neural network to reduce the fitness function. The model was evaluated with the Wisconsin Breast Cancer dataset for efficiency. Comparison of the model outputs with the other nature-inspired computational techniques showed that the developed genetic algorithm-based firefly algorithm recorded the lowest mean squared error of 0.0014 compared with firefly which recorded 0.002, 0.003 obtained in biogeography-based optimization, ACO recorded 0.0135 and PSO gave 0.035.

### 3 METHODOLOGY

#### 3.1 BREAST CANCER DIAGNOSTIC MODEL

The developed breast cancer diagnostic model was implemented in Weka 3.9.4 data mining environment. This environment was used due to its efficient results representation visualization when compared with other

data analysis software. Data from Wisconsin Breast cancer dataset was converted into arff format and then loaded into WEKA environment. The selection of relevant features was done using nature-inspired algorithms: ACO and PSO to rank the features according to their relevance. Seven learning algorithms: C4.5, SVM, KNN, NN, LR, RF and NB were used for classification. The framework of the Breast Cancer Diagnostic Model is shown in Figure 1.

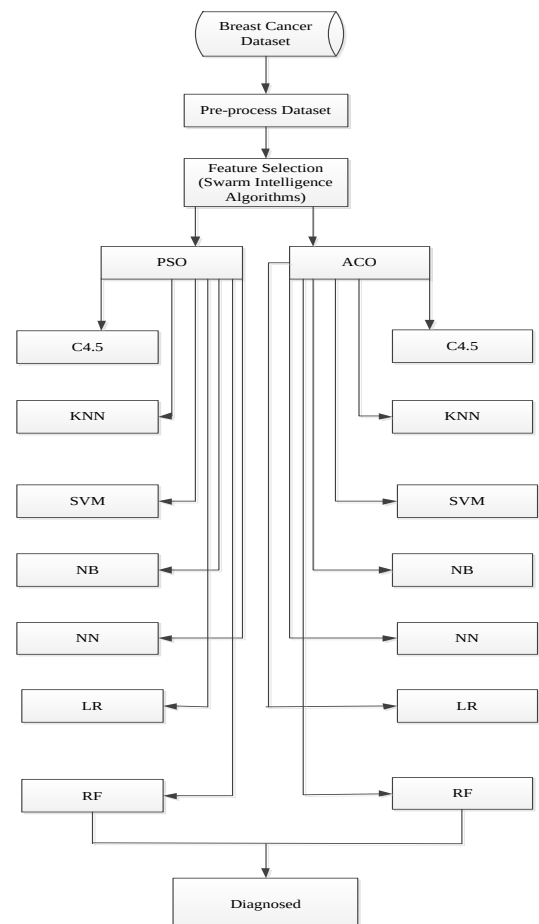


Fig. 1: Framework of the Breast Cancer Diagnostic Model

#### 3.2 ACQUISITION OF BREAST CANCER DATASET

This study used the Wisconsin breast cancer dataset to evaluate the developed breast cancer diagnosis model using data mining classification techniques. The details of the attributes can be found in the Wisconsin Breast Cancer Dataset (WDBC) dataset.

#### 3.3 NATURE-INSPIRED OPTIMIZATION ALGORITHMS

The breast cancer diagnosis model used different meta-heuristic algorithms to perform the selection of the most relevant and discriminant features before the classification. Two nature-inspired optimization algorithms: PSO and ACO were employed. Figure 2 and 3 describe the steps of the nature-inspired optimization algorithms.

Step 1: Start  
 Step 2: Procedure ACO-meta-heuristics  
 Step 3: Schedule Activities  
 Step 4: Construct ant solutions  
 Step 5: Update pheromones  
 Step 6: Daemon actions  
 Step 7: End

Fig. 2: ACO Algorithm

Step 1: Start  
 Step 2: Initialize group of Particles  
 Step 3: Evaluate pBest for each particle  
 Step 4: If current position is better than pbest then Update pBest Else assign pBest to gBest, where pBest is the current fitness value and gbest is the global best  
 Step 5: Compute velocity  
 Step 6: Update particle position  
 Step 7: Target reach Else Goto Step 3  
 Step 8: End

Fig. 3: PSO Algorithm

## 4. RESULTS AND DISCUSSION

### 4.1 RESULTS OF PERFORMANCE EVALUATION METRICS

Several parameters for evaluation which include accuracy, precision, recall, f-measure, kappa-statistic, and time taken to build a model were used to show the effectiveness of the breast cancer diagnostic model.

Table 1. Without Feature Selection Algorithm

| Classifier | Accuracy (%) | Precision | Recall | F-Measure | Kappa-Statistic | Time-Taken (s) |
|------------|--------------|-----------|--------|-----------|-----------------|----------------|
| C4.5       | 94.5637      | 0.9460    | 0.9460 | 0.9460    | 0.8799          | 0.11           |
| KNN        | 95.1359      | 0.9510    | 0.9510 | 0.9510    | 0.8919          | 0              |
| SVM        | 96.7096      | 0.9670    | 0.9670 | 0.9670    | 0.9274          | 0.11           |
| NB         | 95.9943      | 0.9620    | 0.9600 | 0.9600    | 0.9127          | 0.02           |
| NN         | 95.8512      | 0.9590    | 0.9590 | 0.9590    | 0.9086          | 1.22           |
| LR         | 96.5665      | 0.9660    | 0.9660 | 0.9660    | 0.9240          | 0.44           |
| RF         | 96.7096      | 0.9680    | 0.9670 | 0.9670    | 0.9278          | 0.47           |

Table 2. Feature Selection Using PSO Algorithm

| Classifier | Accuracy (%) | Precision | Recall | F-Measure | Kappa-Statistic | Time-Taken (s) |
|------------|--------------|-----------|--------|-----------|-----------------|----------------|
| C4.5       | 94.5637      | 0.9460    | 0.9460 | 0.9460    | 0.8799          | 0.02           |
| KNN        | 95.1359      | 0.9510    | 0.9510 | 0.9510    | 0.8919          | 0              |
| SVM        | 96.9957      | 0.9700    | 0.9700 | 0.9700    | 0.9337          | 0.02           |
| NB         | 95.9943      | 0.9620    | 0.9600 | 0.9600    | 0.9127          | 0              |
| NN         | 95.2790      | 0.9530    | 0.9530 | 0.9530    | 0.8958          | 0.88           |
| LR         | 96.5665      | 0.9660    | 0.9660 | 0.9660    | 0.9240          | 0.05           |
| RF         | 97.1388      | 0.9440    | 0.9750 | 0.9590    | 0.9370          | 0.11           |

Table 3. Feature Selection Using ACO Algorithm

| Classifier | Accuracy (%) | Precision | Recall | F-Measure | Kappa-Statistic | Time-Taken (s) |
|------------|--------------|-----------|--------|-----------|-----------------|----------------|
| C4.5       | 94.5637      | 0.9460    | 0.9460 | 0.9460    | 0.8799          | 0.02           |
| KNN        | 95.1359      | 0.9510    | 0.9510 | 0.9510    | 0.8919          | 0              |
| SVM        | 96.9957      | 0.9700    | 0.9700 | 0.9700    | 0.9337          | 0.38           |
| NB         | 95.9943      | 0.9620    | 0.9600 | 0.9600    | 0.9127          | 0.02           |
| NN         | 95.2790      | 0.9530    | 0.9530 | 0.9530    | 0.8958          | 0.88           |
| LR         | 96.5665      | 0.9660    | 0.9660 | 0.9660    | 0.9240          | 0.03           |
| RF         | 97.1388      | 0.9720    | 0.9710 | 0.9710    | 0.9370          | 0.13           |

In Table 1, the highest accuracy of 96.7096% was obtained in the RF classifier and SVM, the highest precision value of 0.9680 was recorded in RF classifier, the highest recall value of 0.9670 was recorded in SVM and RF classifier, the highest kappa statistics of 0.9278 was obtained in RF classifier. The lowest time of 0s which was taken to build the model was achieved in the KNN classifier.

In Table 2, the highest accuracy of 97.1388% was obtained in the RF classifier, the highest precision value of 0.9700 in the SVM classifier, the highest recall value of 0.9750 was recorded in RF classifier, the highest F1-measure value of 0.9700 was obtained in SVM classifier, the highest kappa statistics of 0.9370 was obtained in RF classifier. The lowest time of 0s which was taken to build a model was recorded in KNN and NB classifier.

In Table 3, the highest accuracy of 97.1388% was obtained in RF classifier, the highest precision value of 0.9720 in RF classifier, the highest recall value of 0.9710 was recorded in RF classifier, the highest F1-measure value of 0.9370 was recorded in RF classifier, the highest kappa statistics of 0.9370 was obtained in RF classifier. The lowest time of 0s which was taken to build a model was achieved in KNN.

## 5 CONCLUSION

The selection of features has been identified to be one of the major problems with large datasets. In data mining, the application of irrelevant features in building a model may lead to more memory space used during the training and classification phase. To make classification faster and more accurate, it is necessary to select the subset of discriminative features. This paper conducted a performance analysis to investigate the effect of two meta-heuristic based attribute selection techniques on breast cancer disease diagnosis. The results showed the effectiveness of the attribute selection algorithms with machine learning algorithms employed for classification. It was revealed that the nature-inspired meta-heuristic optimisation algorithms employed using ACO and PSO performed effectively better than when no attributes selection was used in the selection of relevant attributes from the Wisconsin breast cancer dataset.

## REFERENCES

- Adi, S. I., & Aldasht, M. (2018). American Journal of Computer Parallel Evolutionary Algorithms for Feature Selection in High Dimensional Datasets. *American Journal of Computer Science and Engineering Survey*, 6(1), 13–21.
- Darwish, A., Sayed, G. I., & Hassanien, A. E. (2018). Meta-Heuristic Optimization Algorithms Based Feature Selection for Clinical Breast Cancer Diagnosis. *Journal of the Egyptian Mathematical Society*, 3(26), 321–336. <https://doi.org/10.21608/JOMES.2018.2673.1023>
- Gao, D., Li, X., & Chen, H. (2019). Application of Improved Particle Swarm Optimization in Vehicle Crashworthiness. *Mathematical Problems in Engineering*. <https://doi.org/10.1155/2019/8164609>
- Kumari, B., & Swarnkar, T. (2011). Filter versus Wrapper Feature Subset Selection in Large Dimensionality Microarray: A Review. *International Journal of Computer Science and Information Technologies*, 2(3), 1048–1053.
- Kursa, M. B. (2010). Feature Selection with the Boruta Package. *Journal of Statistical Software*, 36(11), 1–13.
- Mazen, F., AbulSeoud, R. A., & M. Gody, A. (2016). Genetic

- Algorithm and Firefly Algorithm in a Hybrid Approach for Breast Cancer Diagnosis. *International Journal of Computer Trends and Technology*, 32(1), 62–68. <https://doi.org/10.14445/22312803/ijctt-v32p111>
- Padmapriya, B. (2014). A survey on breast cancer analysis using data mining techniques. *IEEE International Conference on Computational Intelligence and Computing Research*, 4–5. <https://doi.org/10.1109/ICCIC.2014.7238530>
- Punitha, S., Amuthan, A., & Joseph, K. S. (2019). Enhanced Monarchy Butterfly Optimization Technique for effective breast cancer diagnosis. *Journal of Medical Systems*, 43(206), 1–14. <https://doi.org/10.1007/s10916-019-1348-8>
- Sheena, Krishan, K., & Gulshan, K. (2016). Analysis of Feature Selection Techniques: A Data Mining Approach. *International Conference on Engineering & Technology*, 17–21.
- Singh, D., Choudhury, J. P., & De, M. (2020). A Comparative Study of Meta-Heuristic Model to Assess the Type of Breast Cancer Disease. *IETE Journal of Research*, 0(0), 1–12. <https://doi.org/10.1080/03772063.2020.1775139>
- Vivek, G., & Shetty, V. (2015). Survey on Swarm Intelligence Based Optimization Technique for Image Compression. *International Journal of Innovative Research in Computer and Communication Engineering*, 3(2), 1058–1063.
- Yeh, J., & Chan, S. (2017). Population-Based Metaheuristic Approaches for Feature Selection on Mammograms. *2017 IEEE International Conference on Agents (ICA)*, 140–144.