CONFERENCE PROCEEDINGS, JANUARY 2025
Published by the Faculty of Science (FSC), FULafia
Print ISSN: 2354–3388 DOI: https://doi.org/10.62050/fscp2024.430

A Joint Optimization Scheme for Enhanced Breast Cancer Diagnosis Using Particle Swarm Optimization (PSO) and Binary Particle Swarm Optimization (BPSO)

Y. E. Ahmed✉, I. M. Abdullahi, D. Maliki & M. A. Akogbe
Department of Computer Engineering, Federal University of Technology, Minna, Nigeria
✉omaregbako@gmail.com

Abstract: One of the leading diseases globally is cancer and breast cancer is not exempted. The objective of the WHO Global Breast Cancer Initiative (GBCI) is to reduce global breast cancer mortality by 2.5% per year, thereby averting 2.5 million breast cancer deaths globally between 2020 and 2040. The three pillars toward achieving these objectives are: health promotion for early detection; timely diagnosis; and comprehensive breast cancer management. In this study we propose an early and comprehensive detection technique in combating breast cancer diagnosis by combining the strength of both PSO (Particle Swarm Optimization) and BPSO (Binary Particle Swarm Optimization) to achieve optimal solution. The results obtained indicated the superiority of the Hybrid PSO-BPSO model in detection over an existing solution by achieving an accuracy of 98.82% on both the WBCD and WDBC datasets.

Keyword: Breast cancer, algorithm, optimization, particle swarm optimization

Introduction

The World Health Organization (WHO) states in March 2022 that an estimated 2.3 million women were diagnosed with breast cancer, resulting in 670,000 casualties globally. Breast cancer is a form of cancerous tumor that originates in the breast tissue. While it can affect both men and women, it is significantly more common in women [1]. Stages of the cancer typically range from 0 to IV, with stage (0) signifying non-invasive disease and stage (IV) representing advanced cancer that has metastasized to other parts of the body which might be life threatening. Fabisiewicz *et al.* highlighted that the significance of early breast cancer detection rests in its ability to have a significant influence on patient well-being and state of mind [2]. The chances of survival are high if the breast cancer is detected at early stage of the malignant tumor, thus minimizes the risk, and need for intensive therapies, and often allows for breast-conserving surgery.

Artificial Intelligence (AI) has revolutionized breast cancer detection, significantly improving the accuracy and efficiency of diagnoses [3]. This advancement is primarily driven by the development and application of machine learning algorithms, such as decision trees, random forests, and support vector machines, which analyze mammograms, clinical data, and genetic markers to identify cancerous tissues, predict outcomes, and facilitate diagnosis. Furthermore, the advent of Deep Learning (DL) and Convolutional Neural Networks (CNNs) has notably enhanced the ability to detect breast cancer through medical imaging, with CNNs automatically extracting critical features from mammograms to identify abnormalities with high precision [4].

A thorough examination of several critical concepts was conducted in the course of this research, with a particular focus on breast cancer, including its complexities, classifications, and diagnostic difficulties. Jain *et al.* proposes a hybrid model designed for gene selection and cancer classification through the analysis of DNA microarray data [5]. The model combines the Correlation-based Feature Selection (CFS), a multivariate filter method, with an enhanced Binary Particle Swarm Optimization (BPSO) algorithm and a Naïve-Bayes classifier. The authors assert that their model exhibits notable classification accuracy and can identify a concise subset of prognostic genes applicable to various cancer types.

Houssein *et al.* also worked on a cutting-edge deep learning framework for diagnosing breast cancer, utilizing the Improved Marine Predators Algorithm (IMPA) to fine-tune the Convolutional Neural Network's (CNN) hyper-parameters [6]. A significant advancement is the fusion of IMPA with transfer learning to boost the CNN’s accuracy in diagnosis. The methodology unfolds across four stages: preprocessing of data, optimization of hyper-parameters, the learning process, and evaluation of performance [7].

Materials and Methods

In this research we propose an innovative hybridization of Binary Particle Swarm Optimization (BPSO) and Particle Swarm Optimization (PSO) and classification of the tumor through artificial neural network classifier. The proposed technique relied on a supervised learning data of patients with breast cancer diagnosis that were stored in UCIrvine Machine Learning Repository. A summary of these datasets is provided in Table 1.

Table 1: Summary of the UCI dataset

Parameter	WBCD (Original)	WDBC (Diagnostic)
Number of Attributes	11	32
Number of Instances	699	569
Number of Classes	2	2

The WBCD (Original) dataset is divided into two class attributes: Benign (B) and Malignant (M). It consists of 699 instances, with 458 benign samples and 241 malignant samples. Similarly, the WDBC (Diagnostic) dataset comprises of 569 instances with no missing values. The non-predictive attribute in this dataset will be the patient ID number [8, 9]. The existing data were then normalized through the linear method in such a way that numerical values could be expressed in terms of 0 and 1 binary sets. The mathematical expression for the data normalization is given below.

$$normD = \frac{d - min\ D}{max\ D - min\ D}(n\,max\ D - n\,min\ D) + n\,min\ D \qquad (1)$$

Where, D denotes the dataset to be normalized which comprises of patient samples in rows and diagnosis features in columns. The $normD$ represents the normalized variables, $d$illustrate the value of a particular sample and feature outcome, while $minD$ and $maxD$ denotes the minimum and maximum value of a particular feature.

The hybrid model

In this research a Hybrid (BPSO/PSO) model was developed using MATLAB (R2020a), and 70% of the data is used for the training, whereas the remaining 30% is used for both testing and validation equally. The training process involves the model learning and identifying patterns within the data through supervised learning, where training samples' features and output classes serve as inputs. Using the PSO metaheuristic optimization algorithm, the optimal hyper-parameters of the model were determined, focusing on the number of neurons in the fully connected layer and the learning rate for diagnosing the breast cancer through adjusting the weights and the bias [10]. An incorrect selection of these parameters could result in over-fitting and subpar model performance. In feature selection the BPSO algorithm is utilized, each feature is either included or excluded which is depicted by binary values of 1 and 0 respectively in the solution vector. Each particle in the swarm represents a potential solution, with its length equivalent to the number of features. After optimization, the trained model underwent testing on unseen data to evaluate its generalization capability. Fig. 1 below depicts the architectural diagram of the proposed method with various components and the flow of each process.

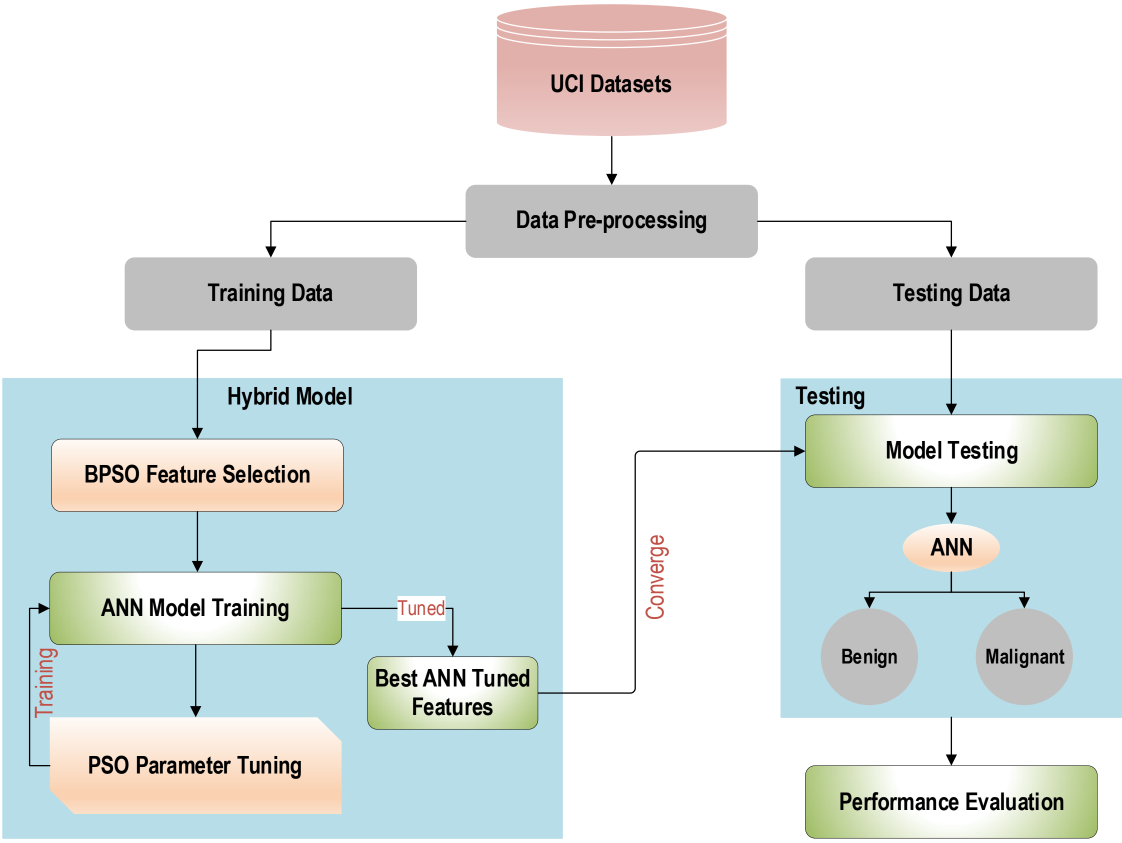

Figure 1: Conceptual diagram of the hybrid proposed method

Table 2 depicts the hybrid model simulation parameters which indicate number of iterations in which the model undergoes in order to find the average performance metrics. Table 3 depicts the Particle Swarm Optimization parameters that were used to tune the selected features in order to avoid the algorithm stuck in local optimal and reduces the risk of hyper-parameter before neural network training.

Table 2: Simulation parameters

Parameter	Value
Population Size (N)	10
Number of Iteration ($f(x)$)	3
Inertia Weight (ω)	2
Min Inertia Weight (ω_{min})	0.9
Max Inertia Weight (ω_{max})	0.4
Cognitive Component (c_1)	1.0
Social Component (c_2)	2.0

Table 3: PSO tuning parameters

Parameter	Value
Swarm Size (N)	5
Number of Iteration ($f(x)$)	1
Inertia Weight (ω)	1
Min Inertia Weight (ω_{min})	0.9
Max Inertia Weight (ω_{max})	0.4
Cognitive Component (c_1)	1.5
Social Component (c_2)	2.0

Results and Discussion

The primary focus is on evaluating the effectiveness of the feature selection process and the overall performance of the classification model. The Fig. 2 below shows the convergence curve of the hybrid (BPSO/PSO) optimization process. The curve tracks the global best fitness value achieved by the algorithm across multiple iterations, with a clear downward trend indicating continuous improvement.

i. The Y-axis represents the global best fitness value, which is a measure of the optimization objective. The values range approximately from 0.20185 to 0.2014, representing a slight but consistent decrease over the iterations.
ii. The X-axis denotes the number of iterations, with the optimization process running for a total of 3 iterations.

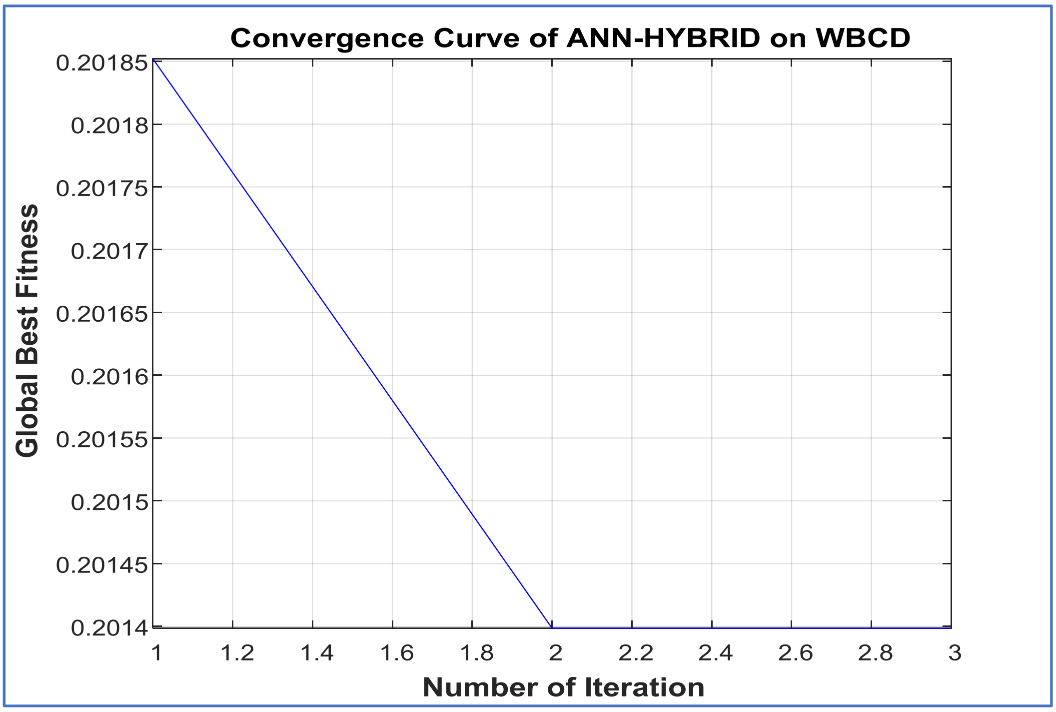

Figure 2: Convergence curve of the hybrid model on WBCD dataset

The curve demonstrates that the optimization algorithm effectively minimizes the fitness value, converging to an optimal solution as the number of iterations increases. This behavior confirms the robustness of the hybrid (BPSO/PSO) approach in finding an optimal set of features and model parameters.

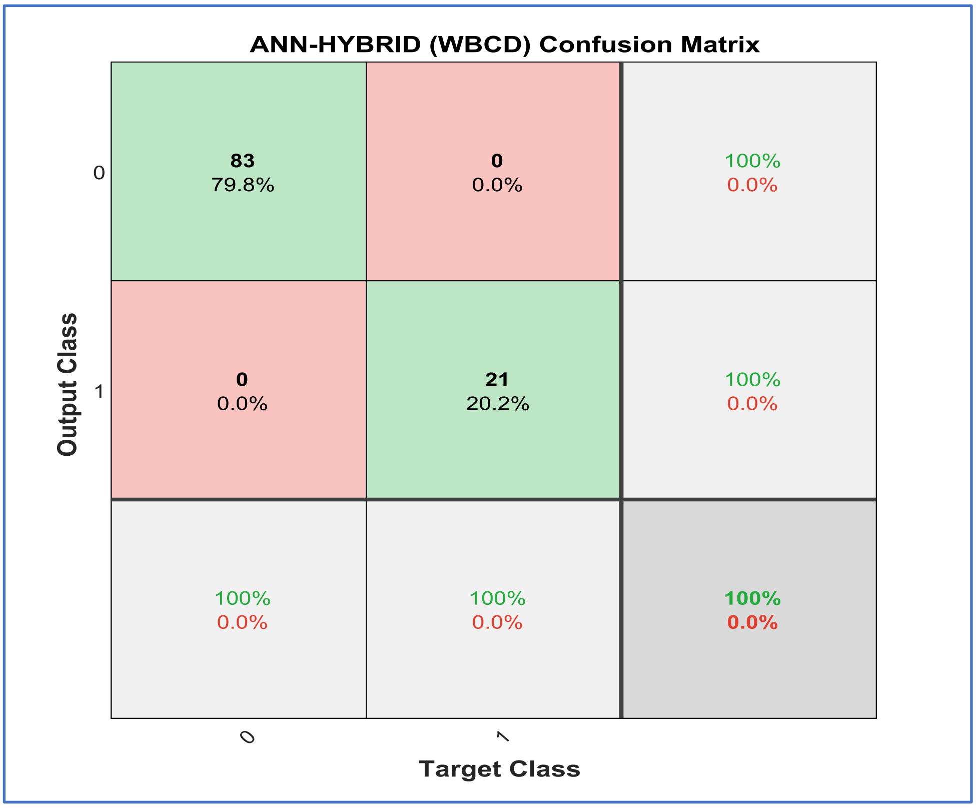

Figure 3: Confusion matrix of the hybrid model on WBCD dataset

The confusion matrix indicates that the model achieved a perfect classification on the test set, with no errors in prediction. This is reflected in the high accuracy rates of 79.8% for benign cases and 20.2% for malignant cases, which suggests that the model is highly effective in distinguishing between the two classes.

i. TP (21 samples): These are instances where the model correctly identified malignant cases.
ii. TN (83 samples): These are outcomes where the model correctly identified benign cases.
iii. FP (0 sample): There are no occurrences where malignant cases were incorrectly classified as benign.
iv. FN (0 sample): There are no occurrences where benign cases were incorrectly classified as malignant.

The Hybrid (BPSO/BSO) optimization technique was utilized to select the most intricate features from the WBCD dataset that enhance the accurate classification of breast cancer characteristics. As shown in Fig. 3, the model exhibits robust performance in classifying instances. The model specificity, which reflects the model's ability to correctly identify benign cases, reaches 100%. Additionally, the sensitivity (recall) for detecting malignant tumors is also 100%, showcasing the model's ability to correctly identify all malignant cases without errors. This emphasizes the model's reliability in clinical applications. Furthermore, the precision for predicting malignant tumors is 100%. Table 4 presents the optimal features selected by the model during training, chosen for their critical contribution to enhancing classification accuracy and reducing error rates.

Table 4: Optimal predicted features on WBCD dataset

Parameter	Feature Identification Number (FID)
Clump Thickness	1
Uniformity Cell Size	2
Marginal Adhesion	4
Single Epithelial Cell Size	5
Mitoses	9

The artificial neural network configuration was plotted using the predicted five (5) cancer features listed in Table 4, as the network inputs in the input layer, a single hidden layer consisting of 22 neurons, and one output layer. This structure is designed to enable the model to effectively grasp data patterns while avoiding overfitting.
Figure 4 depicts the network structure of the predicted model by the hybrid optimization algorithm which indicates a fully connected feed forward neural network, with corresponding network parameters.

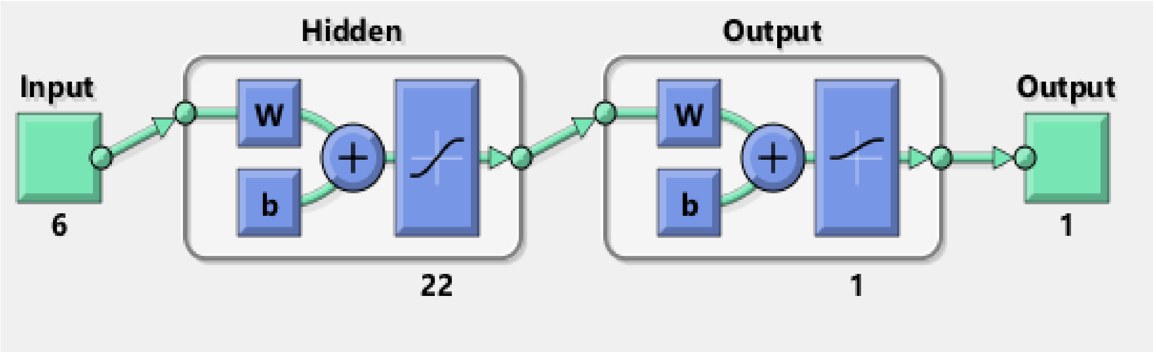

Figure 4: Neural network architecture of the predicted model

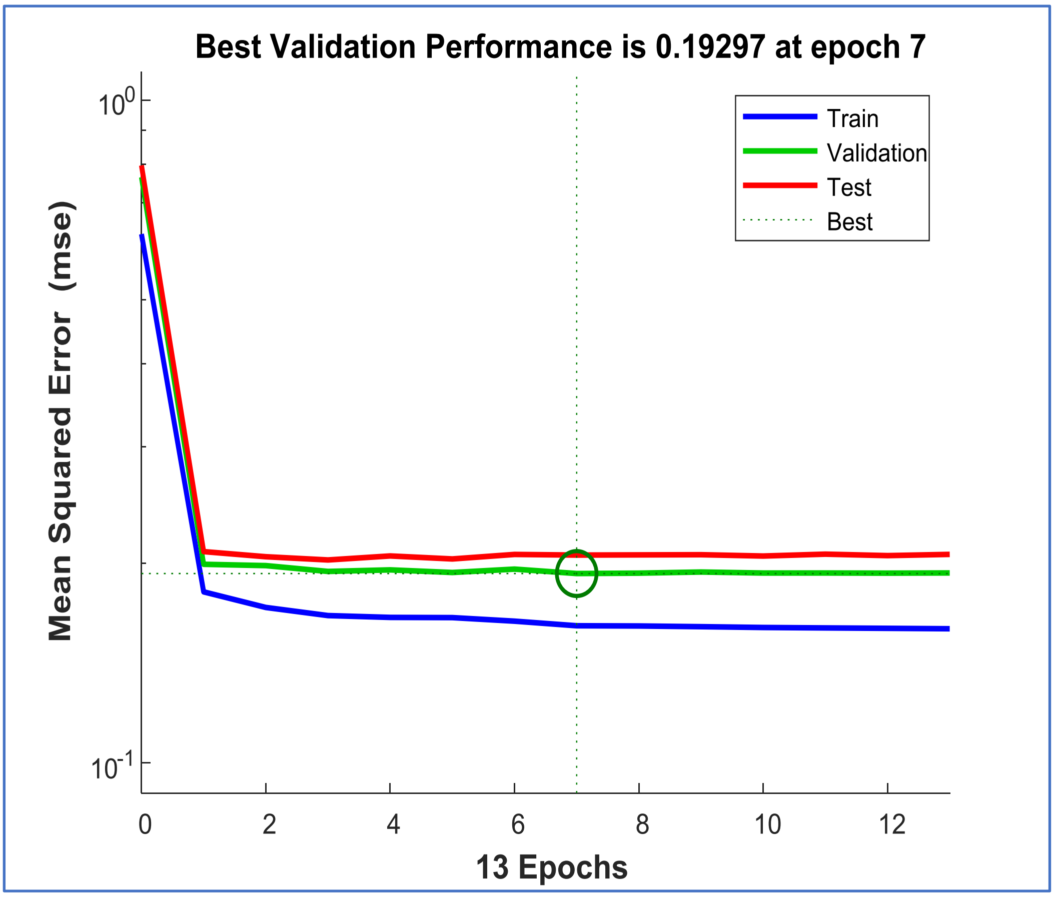

Figure 5: Validation performance on the predicted features

Figure 5 illustrates the model's performance of the model during training, showing significant improvement in the initial epochs. As the MSE decreases rapidly during the early epochs and then stabilizes, indicating the model is learning the training data effectively, with no major overfitting or underfitting.

Conclusion

The algorithm's performance and classification were assessed based on convergence curve, accuracy, specificity, recall, precision, and f1 score. The simulation results indicate that the hybrid (BPSO/PSO) approach outperformed other innovative methods, including ensemble techniques and correlation-based feature selection. While these alternative methods have also demonstrated robust performance, the hybridization of the model provides a balanced and consistent accuracy across different clinical datasets. The hybridization indicates the importance of hyper-parameter tuning before neural network training which in turns eliminates the parameter overfitting.

References

[1] Eriksson, M., Czene, K., Strand, F., Zackrisson, S., Lindholm, P., Lång, K., Förnvik, D., Sartor, H., Mavaddat, N., Easton, D. & Hall, P. (2020). Identification of women at high risk of breast cancer who need supplemental screening. *Radiology*, 297(2), 327–333. https://doi.org/10.1148/radiol.2020201620

[2] Fabisiewicz, A., Kober, P., Koptyra, M., Brewczynska, E., Sienkiewicz-Kozlowska, R., Pienkowski, T. & Siedlecki, J. A. (2005). Detection of circulating cancer cells in peripheral blood as a prognostic factor in early breast cancer. *Breast Cancer Research*, 7(2), P5.01. https://doi.org/10.1186/bcr1182

[3] McKinney, S. M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafian, H., Back, T., Chesus, M., Corrado, G. S., Darzi, A., Etemadi, M., Garcia-Vicente, F., Gilbert, F. J., Halling-Brown, M., Hassabis, D., Jansen, S., Karthikesalingam, A., Kelly, C. J., King, D. & Shetty, S. (2020). International evaluation of an AI system for breast cancer screening. *Nature*, 577(7788), 89–94. https://doi.org/10.1038/s41586-019-1799-6

[4] Mohemmed, A. W., Zhang, M. & Browne, W. N. (2010). Particle swarm optimisation for outlier detection. *Proceedings of the 12th Annual Conference on Genetic and Evolutionary Computation*, pp. 83–84.

[5] Jain, I., Jain, V. K. & Jain, R. (2018). Correlation feature selection based improved-Binary Particle Swarm Optimization for gene selection and cancer classification. *Applied Soft Computing*, 62, 203–215. https://doi.org/10.1016/J.ASOC.2017.09.038

[6] Houssein, E. H., Emam, M. M. & Ali, A. A. (2022). An optimized deep learning architecture for breast cancer diagnosis based on improved marine predators' algorithm. *Neural Computing and Applications*, 34(20), 18015–18033. https://doi.org/10.1007/s00521-022-07445-5

[7] Habib, G. & Qureshi, S. (2022). Optimization and acceleration of convolutional neural networks: A survey. *Journal of King Saud University - Computer and Information Sciences*, 34(7), 4244–4268. https://doi.org/10.1016/J.JKSUCI.2020.10.004

[8] Wolberg, William, Mangasarian, Olvi, Street, Nick & Street, W. (1993). Breast Cancer Wisconsin (Diagnostic). UCI Machine Learning Repository. https://doi.org/https://doi.org/10.24432/C5DW2B

[9] Wisudawati, L. M., Madenda, S., Wibowo, E. P. & Abdullah, A. A. (2020). Feature extraction optimization with combination 2D-discrete wavelet transform and gray level co-occurrence matrix for classifying normal and abnormal breast tumors. *Modern Applied Science*, 14(5), 51–62.

[10] Qiao, J., Wang, G., Yang, Z., Luo, X., Chen, J., Li, K. & Liu, P. (2024). A hybrid particle swarm optimization algorithm for solving engineering problem. *Scientific Reports*, 14(1), 8357.