Breast Cancer Classification Using Seahorse Swarm Optimization

Tanya Dixit[1, *], Arunima Jaiswal[1] and Manaswini De[1]

[1] *Department of Computer Science and Engineering, Indira Gandhi Delhi Technical University for Women, Kashmere Gate, Delhi - 110006*

Abstract

Globally, breast cancer ranks as the most widespread form of cancer among women. Machine Learning and Deep Learning approaches provide more effective means for detecting and managing this condition when compared to conventional detection methods. Advanced deep learning techniques, including LSTM (Long Short-Term Memory), Gate Recurrent Units (GRU), and Deep Belief Networks (DBN) have been used for classification of cancer. In this paper, the publicly available breast cancer dataset namely Wisconsin dataset is employed to investigate the efficacies of these deep learning techniques for classification of breast cancer. Further, the network architecture parameters are tuned for achieving better results using one of the latest swarm intelligence technique namely Sea Horse Optimization. Success rate of 95.61%, 96.49% and 98.24% respectively have been achieved by the proposed SHO-LSTM, SHO-GRU and SHO-DBN models when applied to the Wisconsin dataset.

Keywords

Breast Cancer Classification, Long Short-term Memory Network, Deep Belief Network, Gated Recurrent Unit, Sea horse Optimization[1]

1. Introduction

Breast cancer happens to be the most common cancer type among women globally. According to the Global Cancer Observatory (GLOBOCAN 2020) report [1], India ranked third worldwide in terms of cancer cases and it has been predicted that the number of cases would be increasing drastically in future. Therefore, studies on prediction of breast cancer have great importance of prevention and control of this disease. Various studies have been carried out to highlight the present scenario, challenges and cancer awareness among women in India [2-4]. Leveraging Machine Learning (ML) and Deep Learning (DL) techniques for breast cancer classification not only enhances the faster detection at an early stage with improved accuracy but also contributes to better patient care by reducing the subjectivity of human interpretation of medical information. Authors of this paper have earlier studied the impact of reduction of features on breast cancer classification data by employing classical and quantum machine learning algorithms using the Wisconsin dataset [5]. Further, the effect tuning of hyper-parameters on deep learning models such as Long Short-Term Memory (LSTM) networks, Gate Recurrent Units (GRU) networks and Deep Belief Networks (DBN) has been investigated in this paper.

Various deep learning techniques e.g., MLP, RNN, LSTM, GRU and DBN play crucial roles in breast cancer detection by capturing context and features using sequential data. LSTM and GRU are two variants of Recurrent Neural Networks (RNN) that use gating mechanism to selectively update information overtime, with GRU being simpler and faster than LSTM [6]. DBNs, a form of artificial neural networks, are adept at unsupervised feature learning. They consist of layers of Restricted Boltzmann Machines (RBMs), which are generative models capable of learning valuable features from raw input data without supervision [7].

In Machine Learning, selection of optimal values of various parameters of a model is termed as hyper-parameter tuning. Hyper-parameters are basically the configuration settings that control the learning process of a model, such as learning rate, the number of neural network layers, model architecture, batch size, activation functions, etc. specific to the model employed. Various strategies such as grid search and random search, meta-heuristic optimization, swarm optimization, Bayesian Optimization, etc. are generally used as optimization technique in hyper-parameter tuning. Among these approaches, swarm intelligence plays a significant role in achieving optimal solutions and reducing detection time for efficient control of disease. Sea-horse optimization (SHO) is a new swarm intelligence-based meta-heuristic approach rooted in swarm intelligence, drawing inspiration from the captivating behaviors observed in sea horses in their natural environment [8]. In this study, the detection of breast cancer using deep learning models namely LSTM, GRU and DBN has been attempted and the results of the optimized SHO-LSTM, SHO-GRU and SHO-DBN are compared when applied to the Wisconsin dataset for breast cancer classification.

The structure of this paper is organized as follows: Section 2 offers a concise literature review covering diverse techniques employed for breast cancer detection using ML, DL and hyper-parameter optimization. Section 3 delves into details regarding the dataset employed in this study. Methodology is elucidated in Section 4, followed by a discussion of the results in Section 5. Lastly, conclusions based on the findings are drawn in Section 6.

2. Related Work

A brief overview of a few relevant research works carried out in the field for detection of breast cancer using various techniques has been presented in this section.

The authors of [6], proposed a stacked GRU-LSTM-BRNN deep learning model using Recurrent Neural Networks (RNNs) to classify patients' health records as benign or malignant for breast cancer diagnosis using the Wisconsin Breast Cancer Dataset. The paper compares the three baseline modes: RNN, stacked LSTM and stacked GRU using metrics like accuracy, MSE and Cohen-Kappa score. The results reported by the proposed model outperform the baseline models on all metrics, achieving 97.34% accuracy, 0.97 F1-score, 0.03 MSE and 0.94 Cohen-Kappa score.

In [8], the researchers proposed a swarm intelligence-based meta-heuristic technique called the Sea-horse optimizer (SHO). SHO replicates diverse movement patterns and the probabilistic predation mechanism observed in sea horses. The performance of SHO has been assessed across 23 established functions and the CEC2014 benchmark function. Experimental findings showcase SHO as a proficient optimizer capable of effectively addressing constraint problems. The authors of [9] introduced a novel Chaotic Sea Horse Optimization with DL models (CSHODL-PDC) for classification of pneumonia on CXR images. CSHODL-PDC makes use of NASNet Large model and Fuzzy Deep Neural Network. The proposed model achieved the maximum accuracy of 99.22%, 98.96% precision and recall of 99.22%.

The researchers of [10] proposed a Particle Swarm Optimization (PSO) optimized Multilayer Perceptron Neural Network (MLP). This model is compared with other machine learning models like K-Nearest Neighbors, Decision Tree and Naïve Bayes and shows higher accuracy, sensitivity and specificity. In [11], a stacked GRU (SGRU) for Deep Transfer Learning is used for breast cancer classification along Chaotic Sparrow Search Algorithm (CSSA) for hyper-parameter optimization. The proposed model achieves the accuracy of 98.61% compared to existing models when applied on a benchmark image dataset.

The paper [12] aims to overcome the limitations of Back Propagation Learning Algorithm for RNNs, such as slow convergence, local minima and long term dependencies by using the R programming language to implement proposed models and compare them with standard RNN and LSTM. It employs four distinct meta-heuristic algorithms – Harmony Search, Ant Lion Optimization, Sine Cosine and Grey Wolf Optimizer – to train the LSTM model for classification tasks using real and medical time-series datasets, including the Breast Cancer Wisconsin and Epileptic Seizure Recognition datasets. The proposed models have been reported to achieve higher accuracy rates than the standard ones on both data sets.

The study outline in [13] presents an arithmetic optimization algorithm combined with deep-learning-based histopathological breast cancer classification (AOADL-HBCC), which comprises four sequential steps: noise elimination and contrast enhancement, feature extraction utilizing AOA and SqueezeNet, feature selection employing DBN, and classification utilizing the Adamax Optimizer. The results show that the proposed model achieves the highest accuracy of 96.77% on the 100x dataset and 96.4% on 200x dataset,

outperforming the other models. The study in [7] introduces a hybrid model that individually trains Random Forest (RF), MLP, and DBN on the Wisconsin Breast Cancer dataset. These models are then integrated using a weighted average method to achieve final classification. The proposed model achieves 96.5% accuracy against individual accuracies of 93.9%, 91.3% and 97.5% respectively.

The paper [14] proposes an Enhanced Sea Horse Optimization (ESHO) combined with sine-cosine and Tent Choatic Mapping to adaptively tune the ResNet-50 parameters and optimize its performance on two agricultural image datasets: jade fungus and corn diseases. The ResNet-50 model optimized by ESHO achieves an accuracy of 96.7% for corn disease image recognition and 96.4% for jade fungus image recognition. In [15], a novel hybrid PSO-SHO algorithm combining the advantages of PSO and SHO is proposed. The proposed approach strives to minimize the real power losses and the voltage deviation of the power system by optimizing the generator voltages, transformer tap settings and reactive power compensators.

3. Data

The dataset that is employed is the Wisconsin Breast Cancer Dataset (WBC) [16]. The WBC dataset consists of 569 samples of breast cancer patients with a distribution of 357 benign and 212 malignant cases. 30 characteristic features are quantified from digitized images of breast masses. They are numerical measures of attributes like cell nuclei, means, standard errors and worst values of texture, perimeter, radius, smoothness, compactness, area, concavity, symmetry, concave points and fractal dimension. The classifying labels are denoted by 0 (benign) and 1 (malignant) under the attribute 'diagnosis'.
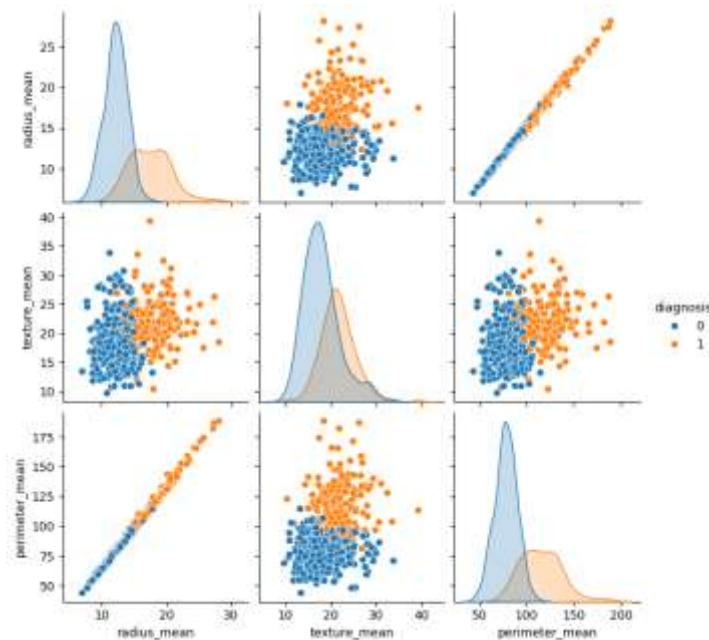

Figure 1. Data Visualization

3.1. Data Pre-Processing

The data preprocessing involves cleaning, scaling features, and encoding labels for the target variables. Subsequently, it is portioned into training and testing sets, with 80% allocated for training data and 20% for testing data. Further, reshaping is used of the training and testing feature arrays to add an additional dimension representing number of channels. LSTMs and GRUs expect input data in a specific 3D shape, typically represented as (batch_size, timesteps, and features). For deep learning models, particularly those using RNNs, input data should be reshaped to meet the input requirements of the model [17].

4. Proposed Method

In this paper, Gate Recurrent Unit Networks (GRU), Long Short-Term Memory (LSTM) and Deep Belief Networks (DBN) have been used to predict breast cancer. Further, we have designed these models with optimized parameters for our classification problem. In order to achieve this, we have used the Seahorse Optimization algorithm to optimize the hyper-parameters and achieve better accuracy. The focus of optimization is on various parameters involved in these neural networks, like learning rate, filter numbers, neurons, epochs etc. The fitness function optimizes based on the classification accuracy. The figure 2 demonstrates a flowchart of our proposed method.
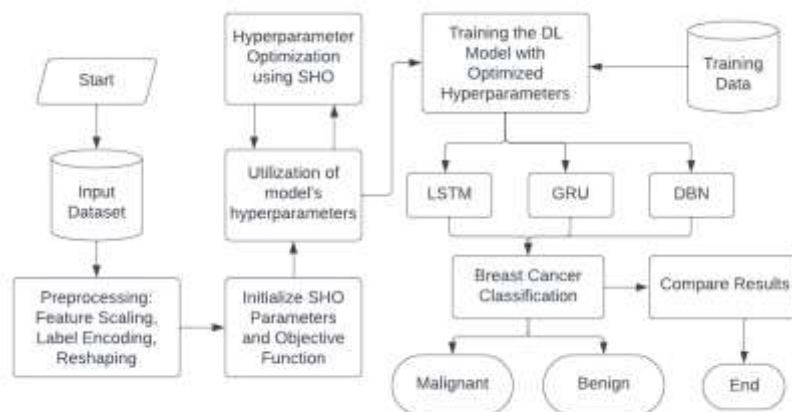

Figure 2. Proposed Method Flowchart

The process begins with by selecting a deep learning mode, such as GRU, LSTM, or DBN. The input data is preprocessed. Seahorse Optimizer's (SHO) parameters like epochs, population size are initialized along with the Objective function that optimizes based on the model's accuracy. SHO then generates hyper-parameters for the chosen model, and the respective model is trained with these hyper-parameters. These steps are repeated till the maximum number of iterations is completed. Overtime, SHO produces the highest global accuracy, and we subsequently assess the model's performance using various traditional performance metrics based on the parameters associated with this accuracy. The models and optimizer's involved are discussed in detail in the coming sections.

4.1. Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) is a deep learning model that addresses the challenges of learning long-term dependencies [18], which traditional recurrent neural networks struggle with. They are specialized neural networks designed to handle sequential data and learn long-term dependencies. LSTMs extend the basic RNN cell [19]. The basic RNN cell takes input at each step, computes a hidden state based on this input and its previous hidden state. The output from the cell helps in training and prediction. Instead of this simple cell, an LSTM cell contains three key components: (1) Input Gate: This gate controls how the amount of new information enters the cell, (2) Forget Gate: Determines which information from the previous hidden state needs to be forgotten, (3) Output Gate: Sets the output based on the current input and hidden state. Additional to the hidden state (similar to RNN's hidden state), LSTMs have a cell state which is essentially a separate memory component that stores long-term information. Cell state is updated with the combined information of input, forget and output gates.

4.2. Gate Recurrent Units (GRU)

Gated Recurrent Units (GRUs) stem from RNNs, incorporating a gating mechanism initially introduced in [20]. Similar to LSTMs, GRUs employ gating mechanism to selectively incorporate or forget certain features, albeit lacking an output gate, thereby resulting in fewer parameters compared to LSTMs. GRU also finds use in processing sequential data such as text, speech and time-series data. GRU can control the flow of information from previous activation state while computing the new activation state [6]. Compared to LSTM, GRU has a superior convergence rate as it has lesser number of parameters and can outperform LSTM models [18]. The GRU comprises two gating mechanism: the reset gate and the update gate. The reset gate regulates the extent to which the previous hidden state is forgotten, while the update gate determines the amount of new input required to update the hidden state.

4.3. Deep Belief Networks (DBN)

Deep Belief Network (DBN) is a generative graphical model, composed of multiple layers of latent variables or more popularly known as "hidden units". They have connections between the layers but no between units within each layer [19]. Consider visualizing these systems as intricate, multi-layer networks with each layer processing information from the preceding one, progressively constructing a sophisticated comprehension of the entire dataset. DBNs are built by layering simple, unsupervised networks such as Restricted Boltzmann Machines (RBMs) or auto encoders. The configuration of the output layer in a DBN is contingent upon the specific task at hand. For instance, in a classification task involving k classes, the DBN would utilize k SoftMax units, each dedicated to one class [7]. The hidden layer of each sub-network serves as the visible layer for the next one in sequence. The training process involves contrastive divergence applied layer by layer, from the lowest visible layer, which serves as the training set [21].

4.4. Hyper-parameter Optimization of Deep Learning Models

Hyper-parameters are external configuration variables set by programmers to operate model training. They are parameters that define the details of learning process. Examples of hyper-parameter optimization encompass various instances such as learning rate (which regulates the magnitude of steps taken during gradient descent optimization), batch size (which dictates the quantity of training examples utilized in each iteration of gradient descent), the number of hidden layers, activation functions (like ReLU, sigmoid, tanh), dropout rate (indicating the fraction of neurons within dense layers, among others. Hyper-parameter optimization (HPO) is the process of selecting optimal values for a machine/deep learning model's hyper-parameters. HPO can be seen as the last step of model design and the initial step of neural network training [22]. It finds a tuple of hyper-parameters that gives an optimal model with enhanced accuracy or prediction. Over the years many techniques have been employed for hyper-parameter optimization such as Grid Search, Random Search, Bayesian Optimization, Genetic Algorithms and even swarm based techniques like Particle Swarm Optimization and Ant Colony Optimization [23]. For the optimization of DL models proposed in this study we have taken the hyper-parameters as represented in table 1, along with the new Seahorse Optimization Algorithm.

Table 1. Hyper-parameters optimized using SHO

Model	Hyper-parameters
LSTM	Filters, neurons, batch-size, epochs
GRU	Filters, neurons, batch-size, epochs
DBN	Hidden layers, learning rate, epochs

4.5. Seahorse Optimization Algorithm

Introduced in 2022, the Seahorse Optimization (SHO) algorithm represents a novel swarm-based meta-heuristic optimization approach [8]. SHO replicates the natural movement, hunting and breeding patterns observed in seahorses. Seahorse movement behavior encompasses two scenarios: (1) Spiral Movement of the hippocampus in conjunction with the ocean vortex and (2) Brownian motion of the hippocampus amidst the waves. Further, predatory behavior consists of the following two situations: success and failure. The breeding behavior of seahorse is described by random mating and offspring inherit traits of both parents. An equal mix of male and female seahorses is taken in the population. To enhance the balance of the SHO algorithm, global strategies are applied to motion behavior and local strategies are applied to predation behavior.

The following equations (1) and (2) denote the spiral and Brownian movement behaviors of seahorses respectively. SHO utilizes Lévy flight to emulate the spiral movement observed in seahorses, which aids in preventing SHO from becoming trapped in local optima. In the spiral movement, the three dimensional vector of coordinates (x, y, z) is denoted by x, y and z. Regarding the Brownian equation l denotes the constant coefficient and β_t represents the coefficient for the motion random walk.

$$X_{new}^i(t+1) = X_i(t) + Levy(\lambda)\big(\big(X_{elite}(t) - X_i(t) \times x \times y \times z + X_{elite}(t)\big)\big) \tag{1}$$

$$X_{new}^1(t+1) = X_i(t) + rand \times l \times \beta_t \times \big(X_i(t) - \beta_t \times X_{elite}\big) \tag{2}$$

Equation (3) is the mathematical representation of predation, with r_2 representing a random number generated by SHO to distinguish between success and failure scenarios. If r_2 exceeds 0.1, the predation by the seahorse is deemed successful; otherwise, it results in failure. α denotes the step size of the seahorse's movement in pursuit of prey.

$$X_{new}^2(t+1) = f(x) = \begin{cases} \alpha \times \big(X_{elite} - rand \times X_{new}^1(t)\big) + (1-\alpha) \times X_{elite}, & r_2 > 0.1 \\ (1-\alpha) \times \big(X_{new}^1(t) - rand \times X_{elite}\big) + \alpha \times X_{new}^1(t), & r_2 \le 0.1 \end{cases} \tag{3}$$

Equation (4) and (5) denote the mathematical equations used to calculate parent seahorses. X_{sort}^2 is fitness values of the population in ascending order. (6) denotes the mathematical equation representing an offspring. r_3 is a random number between [0, 1].

$$father = X_{sort}^2(1:pop/2) \tag{4}$$

$$mother = X_{sort}^2(pop/2 + 1:pop) \tag{5}$$

$$X_i^{offspring} = r_3 X_i^{father} + (1-r_3)X_i^{mother} \tag{6}$$

The algorithm starts with a randomly generated population of seahorses, with each seahorse representing a potential solution. It uses a normal distribution to decide among the two movement behaviors of seahorses. It mimics the high success rate of seahorses in hunting to enhance exploitation capabilities. It draws inspiration from the breeding behavior of seahorses to generate new solutions, hoping to improve upon the current best solution. If the hunt is successful, the seahorse (problem) moves towards the prey (best solution) otherwise search space exploitation continues. The algorithm followed by SHO is explained in the following flowchart:
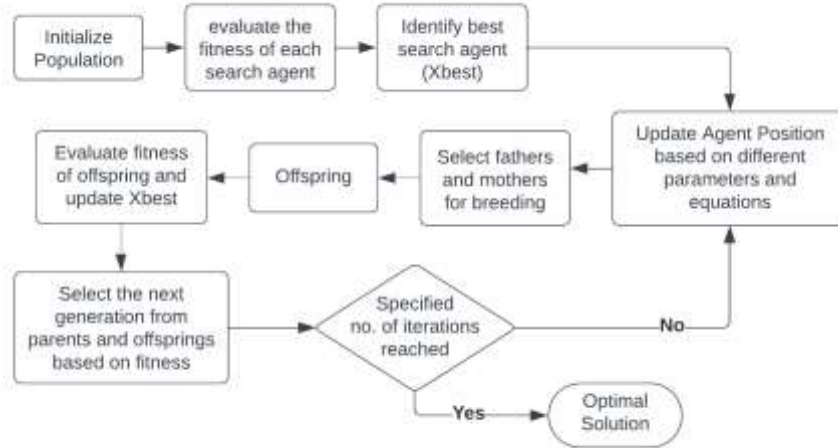

Figure 2. Flowchart illustrating the working of SHO algorithm

The SHO algorithm maintains a balance between exploration (diversification) and exploitation (intensification) to mitigate the risk of getting stuck in local optima and to efficiently locate the global optima. This algorithm has been applied to various engineering design problems, and shows promising results.

5. Results

Various Performance metrics namely accuracy, F1-score, recall, precision and specificity have been used for comparison of results between standard LSTM, GRU and DBN and their SHO optimized counterparts and these parameters are given in the following equations (7)-(11).

$$Accuracy = \frac{TN + TP}{TP + TN + FP + FN} x100 \tag{7}$$

$$Recall = \frac{TP}{TP + FN} \times 100 \tag{8}$$

$$Specificity = \frac{TN}{TN + FP} \tag{9}$$

$$F1\ Score = \frac{2 \times precision \times recall}{precision + recall} \tag{10}$$

$$Precision = \frac{TP}{FP + TP} \tag{11}$$

Accuracy is the ratio of correctly classified instances to the aggregate instances. TP (True Positive), TN (True Negative), FP (False Positive) and FN (False Negative) are the number of correctly and incorrectly classified positive and negative cases. Sensitivity or recall is the fraction of positive cases that the classifier correctly identifies, whereas specificity is the fraction of negative cases that a classifier correctly identifies. Precision is the proportion of true positives out of all predicted positives. F1-score is a measure of performance that combines precision and recall as it is the harmonic mean of the two metrics.

The results obtained from DL models with and without SHO optimization are presented in the following Table 2 and Table 3 respectively.

Table 2. Results obtained without Optimization on DL Models

Model	Accuracy	F1-Score	Recall	Precision	Specificity
LSTM	92.10	91.08	97.87	85.18	88.05
GRU	93.85	92.13	87.23	97.61	98.50
DBN	92.11	90.11	87.23	93.18	95.52

Table 1. Results obtained with Seahorse Hyper-parameter Optimization

Model	Accuracy	F1-Score	Recall	Precision	Specificity
SHO-LSTM	95.61	94.73	95.74	93.75	95.52
SHO-GRU	96.49	95.91	100.00	92.15	94.02
SHO-DBN	98.25	97.87	97.87	97.87	98.51

Before optimization, the LSTM and GRU models achieved accuracies of 92.10% and 93.85%, respectively, while the DBN model achieved an accuracy of 92.11%. However after the hyper-parameter optimization with SHO, the performances of all models significantly improved. The SHO-LSTM achieved an accuracy of 95.61%, SHO-GRU demonstrated an accuracy of 96.49% and SHO-DBN acquired an accuracy of 98.25%.

Furthermore, the SHO-DBN model consistently outperformed the other models across all metrics, demonstrating its effectiveness in breast cancer detection. Notably, the SHO-GRU model achieved perfect recall (100%), indicating its ability to correctly identify all positive cases of breast cancer. GRU demonstrated better performance than LSTM across all metrics with or without optimization, indicating its greater efficiency due to lesser number of model parameters. The graphical comparisons of performance metrics for the un-optimized and optimized DL models are presented in Figure 4(a), 4(b), and 4(c).
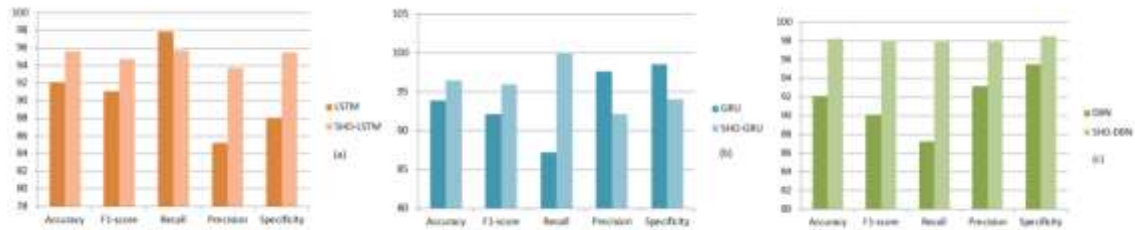

Figure 3. Performance metrics (a) LSTM, (b) GRU and (c) DBN

6. Conclusion

Machine learning and deep learning techniques offer more efficient ways to detect and manage breast cancer compared to traditional methods. Advanced DL models, namely LSTM, GRU and DBN have been successfully applied to Wisconsin dataset for breast cancer classification. Additionally, network architecture parameters are fine-tuned using the state-of-the-art swarm intelligence technique known as Sea Horse Optimization. In this study, we explored the effectiveness of these deep learning techniques for achieving improved performances when applied to two different breast cancer datasets. Specifically, the SHO-DBN model emerges as the most effective model for this task, with the highest accuracy, precision, specificity and F1-Score. These findings underscore the potential of optimized deep learning models as valuable tools in the early detection and management of breast cancer.

References

[1] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, F. Bray, "Global cancer statistics 2020:GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries", *CA: A Cancer Journal for Clinicians*, vol. 71, no. 3, pp. 209-249, 2021.

[2] A A. Gupta, K. Shridhar, P. K. Dhillon, "A review of breast cancer awareness among women in India: Cancer literate or awareness deficit?", *European J. Cancer* , vol. 51, no. 14, pp. 2058-2066, 2015.

[3] R. Mehrotra, K. Yadav, Breast cancer in India: "Present scenario and the challenges ahead", *World journal of clinical oncology*, vol. 13.3, pp. 209-218, 2022, doi:10.5306/wjco.v13.i3.209.

[4] K. Sathishkumar, M. Chaturvedi, P. Das, S. Stephen, P. Mathur, "Cancer incidence estimates for 2022 & projection for 2025: Result from National Cancer Registry Programme", *The Indian journal of medical research*, vol. 156, no. 4 & 5, pp. 598-607, 2022, doi:10.4103/ijmr.ijmr_1821_22.

[5] M. De, A. Jaiswal, T. Dixit, "Comparative analysis of classical and quantum machine learning for breast cancer classification", *International Conference on Computational Intelligence and Mathematical Applications*, 2023.

[6] S. Dutta, J. K. Mandal, T. H. Kim, S. K. Bandyopadhyay, "Breast Cancer Prediction Using Stacked GRU-LSTM-BRNN", *Applied Computer Systems*, vol. 25, no. 2, pp. 163-171, 2020, doi: 10.2478/acss-2020-0018.

[7] S. Yamani, Z. H. Choudhury, "Integrating Random Forest, MLP and DBN in a Hybrid Ensemble Model for Accurate Breast Cancer Detection", *International Journal of Innovative Science and Research Technology*, vol 8, no. 7, pp. 1556-1564, 2023, doi:10.31219 /osf.io/sdjqf.

[8] S. Zhao, T. Zhang, S. Ma, M. Wang, "Sea-horse optimizer: a novel nature-inspired meta-heuristic for global optimization problems", *Applied Intelligence*, vol. 53, no. 10, pp. 11833-11860, 2023, doi:10.1007/s10489-022-03994-3.

[9] V. Parthasarathy, S. Saravanan, "Chaotic Sea Horse Optimization with Deep Learning Model for lung disease pneumonia detection and classification on chest X-ray images", *Multimedia Tools and Applications*, 2024, doi: 10.1007/s11042-024-18301-0.

[10] M. Alimardani, M. Almasi, "Investigating the application of particle swarm optimization algorithm in the neural network to increase the accuracy of breast cancer prediction", *International Journal of Computer Trends and Technology*, vol. 68, no. 4, pp. 65-72, 2020, doi: 10.14445/22312803/IJCTT-V68I4P112.

[11] K. Shankar, A. K. Dutta, S. Kumar, G. P. Joshi, I. C. Doo, "Chaotic Sparrow Search Algorithm with Deep Transfer Learning Enabled Breast Cancer Classification on Histopathological Image", *Cancers*, vol. 14, no. 11, 2022, p. 2770, doi: 10.3390/cancers14112770.

[12] T. A. Rashid, P. Fattah, D. K. Awla, "Using accuracy measure for improving the training of LSTM with metaheuristic algorithms", *Procedia Computer Science*, vol. 140, pp. 324-333, 2018, doi:10.1016/j.procs.2018.10.307.

[13] M. Obayya, S. Alkhalaf, F. Alrowais, S. Alshahrani, A. Alzahrani, A. Alzahrani, "Hyperparameter Optimizer with Deep Learning-Based Decision-Support Systems for

Histopathological Breast Cancer Diagnosis", *Cancers*, vol. 15, p. 885, 2023, doi: 10.3390/cancers15030885.

[14] Z. Li, S. Qu, Y. Xu, X. Hao, N. Lin, "Enhanced Sea Horse Optimization Algorithm for Hyperparameter Optimization of Agricultural Image Revolution", *Mathematics*, vol. 12, no. 3, p. 368, 2024, doi:10.3390/math12030368.

[15] H. M. Hasanien, I. Alsaleh, M. Tostado-Véliz, M. Zhang, A. Alateeq, F. Jurado, A. Alassaf, "Hybrid particle swarm and sea horse optimization algorithm-based optimal reactive power dispatch of power systems comprising electric vehicles", *Energy*, vol. 286, pp. 129583, 2024, doi:10.1016/j.energy.2023.129583.

[16] "Diagnostic Wisconsin Breast Cancer Database", *UCI Machine Learning Repository*. Available: https://archive.ics.uci.edu/dataset/17/breast+cancer+wisconsin+diagnostic.

[17] M. Chhetri, S. Kumar, P. P. Roy, B.-G. Kim, Deep BLSTM-GRU "Model for Monthly Rainfall Prediction: A Case Study of Simtokha", Bhutan*, Remote Sensing*, vol. 12, no. 19, p. 3174, 2020, doi:10.3390/rs12193174.

[18] A. Sherstinsky, "Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network", *Physica D: Nonlinear Phenomena*, vol. 404, p. 132306, 2020, doi: 10.1016/j.physd.2019.132306.

[19] G. E. Hinton, "Deep belief networks". In *E. N. Zalta (Ed.), Scholarpedia*, vol. 4.5, p. 5947, 2009, doi:10.4249/scholarpedia.5947. Available: http://www.scholarpedia.org/article/Deep_belief_networks.

[20] J. Chung, C. Gulcehre, K. Cho, Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling", *arXiv preprint*, arXiv: 1412.3555v1, 2014, doi: 10.48550/arXiv.1412.3555.

[21] Y. Bengio, P. Lamblin, D. Popovici, H. Larochelle, "Greedy layer-wise training of deep networks". In B. Schölkopf, J. Platt, T. Hoffman (Eds.), *Advances in Neural Information Processing Systems*, MIT Press, vol. 19, pp. 153-160, 2006. Available: https://deeplearning.cs.cmu.edu/ pdfs/Bengio_et_al_NIPS_2006.pdf.

[22] T. Yu, H. Zhu, "Hyper-Parameter Optimization: A Review of Algorithms and Applications", 2020. Available: https://arxiv.org/abs/2003.05689.

[23] W.-C. Yeh, Y.-P. Lin, Y.-C. Liang, C.-M. Lai, X.-Z. Gao, "Simplified Swarm Optimisation for the Hyperparameters of a Convolutional Neural Network", 2023. Available: https://arxiv.org/ftp/arxiv/papers/2103/2103.03995.pdf.

[24] H. Alahmer, A. Alahmer, M. I. Alamayreh, M. Alrbai, R. Al-Rbaihat, A. Al-Manea, R. Alkhazaleh, "Optimal water addition in emulsion diesel fuel using machine learning and sea-horse optimizer to minimize exhaust pollutants from diesel engine", *Atmosphere*, vol. 14, p. 449, 2023, doi: 10.3390/atmos14030449.