

Impact of Genetic Algorithm for the Diagnosis of Breast Cancer: Literature Review

Abebe Alemu Balcha¹, Samuel Alemu Woldie²

¹Department of Computer Science and Engineering, Adama Science and Technology University, Adama, Ethiopia

²Department of Computer Science and Technology, Kotebe University of Education, Addis Ababa, Ethiopia

Email: abesoftware@gmail.com, biosinfotechdb@gmail.com

How to cite this paper: Balcha, A.A. and Woldie, S.A. (2023) Impact of Genetic Algorithm for the Diagnosis of Breast Cancer: Literature Review. *Advances in Infectious Diseases*, 13, 41-46.

<https://doi.org/10.4236/aid.2023.131005>

Received: November 27, 2022

Accepted: February 24, 2023

Published: February 27, 2023

Copyright © 2023 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

In recent research from the total number of new cancer cases in Africa about 29.46% and in Ethiopia 31.85% are breast cancer cases. 25.84% of all cancer related death is from breast cancer. One of the challenges in the treatment of breast cancer is early detection. Researchers agreed that, improving the preventive mechanism of breast cancer is an early predicting and detecting model. Research efforts are continuing to present different solution approaches using advanced techniques of Artificial intelligence (AI), Machine learning (ML), Deep Learning (DL), and Computational Intelligence as well. A genetic algorithm is a hyper-parameter optimization algorithm that belongs to the class of evolutionary algorithms. Genetic Algorithm (GA) is used for complex search spaces for search and optimization. This reviewed literature paper shows the positive effect of GA in the diagnosis of breast cancer on AI algorithms.

Keywords

Genetic Algorithm, Breast Cancer, Feature Classification, Optimization

1. Introduction

Cancer is a broad term for a range of diseases that are caused by the uncontrolled spread of a body's cells. These cells ultimately form a tumor in the body and are possibly to invade surrounding tissue or spread throughout the body [1] [2].

Machine learning consists of a wide range of algorithms that are programmed to solve problems based on data, [3], often by identifying patterns [4] that are invisible to humans. In order to achieve actionable accuracy, these algorithms are improved either by optimizing the parameters of the machine learning or by

reducing irrelevant data by feature selection. Genetic algorithms (GAs) are one class of metaheuristic algorithms that were inspired by biological genetic mechanisms to choose optimal solutions [1]. GAs can be applied either as the base classifier, or as an optimized for the parameters of base classifiers [5], or as a feature selector on the data [1].

In order to make cancer-related healthcare optimized and accessible to all, it is important that early detection and noninvasive testing for various types of cancer are widespread and accurate. Since accuracy and efficiency are incredibly important features for any cancer detection process, there is a need for a highly accurate optimization function for parameters and features. Because of this, the metaheuristic GA is a popular field of research for cancer detection and prediction-based algorithms.

This paper presents a literature review of the impact of genetic algorithms in the detection and prediction of cancer. The various research studies are organized based on the function of the utilized GA. The final result of the paper will deduce on the impact of GA in the accuracy of the diagnosis of BC for early detection and prediction.

2. Genetic Algorithm

The genetic algorithm is a hyper-parameter optimization algorithm. The genetic algorithm is a kind of heuristic algorithm inspired by the theory of evolution. It is widely used in search problems and optimization problems [6] [7], by performing biological heuristic operators such as selection, mutation, and crossover [8].

In Paper [2] to verify the performance of GA on the transfer CNN tasks, three datasets (Dataset 1, 2 and 3) are tested. The genetic operations show a significant improvement in the average accuracy on all the given datasets. The accuracy in the first generation is barely better than a random choice. While after the system converged to the best individual achieved an accuracy of 97%. Around the 14th generation, the system converged and gives the average recognition accuracies at 93%, 90%, and 87% of the three datasets, respectively. The average recognition accuracy is updated from 76% to 88% by generation [2]. The best individual gives a fairly high accuracy in the first generation. But it still can be proved that the GA is more efficient than a random search for diagnosing Breast Cancer. The Genetic algorithm works to optimize neural networks by adjusting the weights and the bias. In order to give more accuracy in a neural network, use a hybrid genetic algorithm.

In the paper [9], the performance comparison of various deep neural network models architecture found by GA and PSO on validation set. GA was implemented to provide the optimal number of hidden layer and hidden nodes. The experiment was done with Neural Network of having 6 hidden layers and 297 hidden nodes each of with ReLu activation function. And then re-implement GA on the first result, got a significant result. The model in the prediction of BRCA1/BRCA2 is pathogenicity with the precision of 99.65% and AUC of 98.65. GA showed improvement with a precision of 98.96%, [9].

The ideal scenario of applying GA is if there is not much data available and the theory on the problem is not remarkably high. On the other side, CNN and DL could perform better when working with a large amount of data [9] [10]. Use the global search capability of the GA to evolve the CNN weights for the histopathological breast image classification problem [11], train CNN model using the BreakHis dataset images as input and three different optimization approaches. In research paper [11], GA-based classifier performs almost as powerfully as the Adam optimizer, with a negligible difference. The batch size is equal to 32, the other two algorithms have the best accuracy, and when the batch size is equal to 128, GA has scored best accuracy. Paper [8], uses search method based on the principles of natural selection and genetics [8]. Likewise, combined Genetic Algorithm and SVM has performed a high rank and selected the best ones in a few generations [12]. Second, the author used another GA to perform feature selection in the ensemble method, which is worked in conjunction with the SVM. In this method, the small set of features becomes best in classifying the data concerning breast cancer diagnosis at the stage.

In addition, using GA in combination with other algorithms enhances the accuracy of results in the case of a small data set. The main advantages of hybridized GA with other methods, [7] [12] [13] are better solution quality, better efficiency, a guarantee of feasible solutions, and optimized control parameters. Hybridization of intelligent techniques for an effective predictive model is essential as reported, [11]. The accuracy of the GA and ANN hybrid model has greater than the single Back propagation neural network (BPANN). Using principles of global optimization, GANN performed well.

In the research paper, [12], is extracted a series of image features from 146, 29 malignant and 117 cases were benign. The decision tree based EG2 algorithm was used for classification. GA is used to select the best Multi Classifier System (MCS), which the author used to create multiple random subsets of attributes (ensemble), which consists of a set of the independently trained classifier. In this work, the ensemble has several sets ranging from 1, 3, 5 and 7 fold cross-validation, the crossover performed at two points. The GA iteration stops as the criterion was the moment when 100 new populations were generating without increasing. The paperwork, [12] registered an accuracy of 91.09%, the sensitivity of 80.60%, and the specificity of 94.82% was achieved, with Ensemble $V = 7$.

In the research paper [13], the algorithm C4.5 has an accuracy of 91.2% with the combination of K-Means clustering ($K = 2$) and combining with GA as a feature selection accuracy become 94.824%. These it improved by 3.596% of accuracy in the diagnosis of breast cancer.

In the paper of [12] [14] [15], the genetic algorithm-based weighted average method, can be implemented in the prediction of multiple models. The comparison has been done between Particle swarm optimization (PSO), Differential Evolution (DE) and Genetic algorithm (GA). The genetic algorithm outperforms weighted average methods. The other comparison has been done between the classical ensemble method and GA based weighted average method and deduced

that GA based weighted average method outperforms.

Paper [16], propose a methodology for classifying the breast as normal and abnormal. It contributes to the development of automatic segmentation of RoI. To reduce the number of features and increase the performance classifier, the Authors used GA. The results obtained were: 98% accuracy, 97% sensitivity, and 100% specificity using the Artificial Neural Networks (ANN) classifier.

3. Research Methodology

Research papers collected from different publishers in the years 2018 to 2022. The research applied to the mammogram image type, Databases from Medline (Ovid), Embase (Ovid); Web of Science, and the Cochrane Database of Systematic Reviews (CENTRAL). The topics focused on image classification and analysis using Artificial intelligence, Machine learning, and Deep Learning for the prediction and detection of Breast cancer of those hybridized with genetic algorithms in effect to compare the result. GA uses for feature selection to optimize the performance of the accuracy of the diagnosis. Papers organize with AI algorithms having the best accuracy, sensitivity, and specificity. In addition, selection criteria are highly applicable in public research, WDBC, and Kaggle. The target features of the breast cancer image and the stage of the tumor are taken as the methods of the review for this paper.

4. Result

Algorithm C4.5 combining with GA has highest accuracy comparing with the combination of K-Means in the diagnosis of breast cancer, [13]. There is around 3.6% of improvement in the accuracy of feature selection when combining with GA. GA runs to optimize the result of neural networks to classify Breast cancer. GA used for feature selection to choose best fitness value for diagnosis with the WDBC data set from UCI ML repository, [15].

The performance in **Table 1** shows an improvement using GA in run two with different layers.

Because of the best fitness selection criterion, the error rate decrease when using GA with ANN [15] to diagnose breast cancer. GA-NN technique has shown in **Table 2** a proper precision in cancer diagnosis.

Table 1. Metaheuristic performances on run 1 and run 2 [9].

Models	Precision	Sensitivity	Specificity	Accuracy	AUC
GA (run 1)	0.9897	0.9822	0.9899	0.9861	0.9848
GA (run 2)	0.9896	0.9879	0.9901	0.9892	0.9869

Table 2. The test result technique performance, [15].

Algorithm	TEST DATA	CANCER DATA	MSE	Accuracy
NN	200	182	4.8E-6	0.91
GANN	200	188	5.4E-12	0.94

5. Discussion

In these research papers, we noticed that even though there is a big change in the diagnosis of breast cancer, still it is the second cancer disease registered to death. AI has a noble impact on the enhancement of performance in the diagnosis of breast cancer. However, it has a gap to utilize and address in medical imaging analysis. First, deep learning needs big data set for pre-training the image, and the second one is to create a new algorithm, used by a lesser data set for training the data. The diagnosis performance depends largely upon the volume of raw data and the quality of training. Thus quality DL requires high-quality image data as input. The advantage of the genetic algorithm remains the ability to achieve the optimal solution from a limited population.

Based on the current data analysis and research, radiologists miss 15% to 35% of breast cancer from mammography image data. In addition, most research papers focus mainly on the accuracy metrics to evaluate the performance of detecting breast cancer, instead of using the confusion matrix. The future has to focus on AUC and F-Score matrix to evaluate the performance of the training and to detect the first stage of breast cancer.

Besides this, there has to be a highly accurate algorithm to predict early, whether the tumour is cancer or not. Further research is needed to enhance the performance of current AI, ML, DL, and ANN technologies with a smaller dataset. From this paper, we noticed that the feature selection can use GA to improve the accuracy of the diagnosis. And hence we recommend that in the algorithm of ML and DL hybridized GA is a crucial point.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Aalaei, S., Shahraki, H., Rowhanimanesh, A. and Eslami, S. (2016) Feature Selection Using Genetic Algorithm for Breast Cancer Diagnosis: Experiment on Three Different Datasets. *Iranian Journal of Basic Medical Sciences*, **19**, 476-482.
- [2] Li, C., Jiang, J., Zhao, Y., Li, R., Wang, E., Zhang, X. and Zhao, K. (2021) Genetic Algorithm Based Hyper-Parameters Optimization for Transfer Convolutional Neural Network. In: Tiwari, R., Ed., *International Conference on Advanced Algorithms and Neural Networks (AANN2022)*, Vol. 12285, SPIE, Bellingham. <https://doi.org/10.1117/12.2637170>
- [3] AttyaLafta, H., KdhimAyooob, N. and Hussein, A.A. (2017) Breast Cancer Diagnosis Using Genetic Algorithm for Training Feed Forward Back Propagation. *2017 Annual Conference on New Trends in Information & Communications Technology Applications (NTICT)*, Baghdad, 7-9 March 2017, 144-149. <https://doi.org/10.1109/NTICT.2017.7976101>
- [4] Bhandari, A., Tripathy, B.K., Jawad, K., Bhatia, S., Rahmani, M.K.I. and Mashat, A. (2022) Cancer Detection and Prediction Using Genetic Algorithms. *Computational Intelligence and Neuroscience*, **2022**, Article ID: 1871841.

- <https://doi.org/10.1155/2022/1871841>
- [5] Dou, Y. and Meng, W. (2021) An Optimization Algorithm for Computer-Aided Diagnosis of Breast Cancer Based on Support Vector Machine. *Frontiers in Bioengineering and Biotechnology*, **9**, Article 698390. <https://doi.org/10.3389/fbioe.2021.698390>
- [6] Ghosh, S. and Bhattacharya, S. (2020) A Data-Driven Understanding of COVID-19 Dynamics Using Sequential Genetic Algorithm Based Probabilistic Cellular Automata. *Applied Soft Computing*, **96**, Article ID: 106692. <https://doi.org/10.1016/j.asoc.2020.106692>
- [7] Sari, M. and Tuna, C. (2018) Prediction of Pathological Subjects Using Genetic Algorithm. *Computational and Mathematical Methods in Medicine*, **2018**, Article ID: 6154025. <https://doi.org/10.1155/2018/6154025>
- [8] Zeebaree, D., Haron, H., Abdulazeez, A.M. and Zeebaree, S.R.M. (2017) Combination of K-Means Clustering with Genetic Algorithm: A Review. *International Journal of Applied Engineering Research*, **12**, 14238-14245. <http://www.ripublication.com>
- [9] Pellegrino, E., *et al.* (2022) Deep Learning Architecture Optimization with Metaheuristic Algorithms for Predicting BRCA1/BRCA2 Pathogenicity NGS Analysis. *BioMedInformatics*, **2**, 244-267. <https://doi.org/10.3390/biomedinformatics2020016>
- [10] Elite Data Science (2022) Dimensionality Reduction Algorithms: Strengths and Weaknesses. <https://elitedatasience.com/dimensionality-reduction-algorithms>
- [11] Sharma, D.K., Hota, H.S., Brown, K. and Handa, R. (2022) Integration of Genetic Algorithm with Artificial Neural Network for Stock Market Forecasting. *International Journal of Systems Assurance Engineering and Management*, **13**, 828-841. <https://doi.org/10.1007/s13198-021-01209-5>
- [12] Resmini, R., Silva, L., Araujo, A.S., Medeiros, P., Muchaluat-Saade, D. and Conci, A. (2021) Combining Genetic Algorithms and SVM for Breast Cancer Diagnosis Using Infrared Thermography. *Sensors*, **21**, Article No. 4802. <https://doi.org/10.3390/s21144802>
- [13] Andoyo, F.A. and Arifudin, R. (2021) Optimization of Classification Accuracy Using K-Means and Genetic Algorithm by Integrating C4.5 Algorithm for Diagnosis Breast Cancer Disease. *Journal of Advances in Information Systems and Technology*, **3**, 1-8. <https://doi.org/10.15294/jaist.v3i1.49011>
- [14] Alalayah, K.M.A., Almasani, S.A.M. and Qaid, W.A.A. (2018) Breast Cancer Diagnosis based on Genetic Algorithms and Neural Networks. *International Journal of Computer Applications*, **180**, 42-44. <https://doi.org/10.5120/ijca2018916605>
- [15] Chauhan, P. and Swami, A. (2018) Breast Cancer Prediction Using Genetic Algorithm Based Ensemble Approach. 2018 *9th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Bengaluru, 10-12 July 2018, 1-8. <https://doi.org/10.1109/ICCCNT.2018.8493927>
- [16] Sánchez-Ruiz, D., Olmos-Pineda, I. and Olvera-López, J. (2020) Automatic Region of Interest Segmentation for Breast Thermogram Image Classification. *Pattern Recognition Letters*, **135**, 72-81. <https://doi.org/10.1016/j.patrec.2020.03.025>