



Breast Cancer Prediction using Modified Support Vector Machine and Grey Wolf Optimization Technique

Mr.K.Madhavan, Ph.D., Research Scholar (Part-Time) Department Of Computer Science,
Bharathiar University, Coimbatore , Tamilnadu -641046
E mail -madankmadan3@gmail.com

Dr. R. Manicka Chezian , Associate Professor P.G. Department of Computer Science,
NGM College, Pollachi.

Abstract

Breast cancer has been classified as one of the leading causes of mortality among women in different parts of the world. However, it is well established that early identification and precise diagnosis of the illness assures the patient's long survival. The ability to distinguish between benign and malignant tumours at the right time is crucial in the diagnosis of breast cancer. An attempt is made in this work to develop a novel prediction system that explores the full potential of support vector machine (SVM) using a modified grey wolf optimization (GWO) technique.

The proposed meta-heuristic algorithm named as Modified Grey Wolf Optimization (MGWO) and Support Vector Machine (SVM) technique (MGWO-SVM) has been developed for extracting valuable information through appropriate selection of relevant features in the popular breast cancer datasets like wisconsin diagnostic breast cancer database (WDBC) for early-stage detection of the disease.

Keywords: *Breast cancer detection, Grey wolf optimization, SVM algorithm.*

1. Introduction

In recent years, computational diagnostic tools and artificial intelligence approaches have provided automated procedures for objective evaluations by employing quantitative metrics and machine learning techniques (Maglogiannis et al. 2009). Research interest in machine learning (ML) has attained peak levels in the recent decade. This renewed importance is accelerated by less expensive computing power and memory. Thus, large amounts of data can be efficiently stored, processed, and analysed. Machine learning plays a crucial role in a variety of crucial applications, including data mining, natural language processing, image recognition, expert systems, and prediction.

Cancer is the most hazardous illness in the world, with breast cancer being the most harmful for women (Khuriwal & Mishra, 2018). Fatality among women is increasing every year because of breast cancer. Detecting breast cancer manually requires a significant amount of time and is difficult for the physician to classify. Therefore, the detection of cancer by a variety of automated diagnostic approaches is essential.

In several disease prevention research studies, strategies for early cancer diagnosis are extensively addressed in an effort to combat the exponentially increasing cancer incidence. Commonly utilised pre-diagnostic procedures include yearly mammography (Ades et al.,

2014), gene diagnosis (Jazaeri et al 2002), clinical diagnosis (Sotiriou et al., 2003), etc. Moreover, with the advancement of biomedical and information technology over the years, numerous prognostic factors related to breast cancer have been identified, allowing numerous researchers to develop more sophisticated early detection models based on various data-driven prediction methodologies, such as support vector machines (SVMs), logistic regression (LR), multilayer perceptrons (MLPs), and decision trees (DTs)

2. Background

Rojas-Domínguez et al. (2017) have analysed the different algorithms like Boltzmann-UMDA, Firefly algorithm, Fruit-fly optimization algorithm, particle Swarm optimization, and Bat algorithm to determine their optimal performance using measures like efficacy, generality, efficiency, and complexity. According to experimental findings on 15 medical diagnostic issues, estimation of distribution algorithms (EDAs) has provided optimal results. The optimization process was guided by a unique performance indicator that has the potential to increase the generalizability of the solutions while preserving their efficacy. Wang et al. (2018) have proposed an ensemble learning approach for breast cancer detection based on support vector machines (SVM). The proposed method has minimised the diagnostic variance and boosted diagnosis accuracy in order to address the restriction of individual model performance. Based on the suggested Weighted Area Under the Receiver Operating Characteristic Curve Ensemble (WAUCE) methodology, twelve distinct SVMs were hybridised. The Wisconsin Breast Cancer, Wisconsin Diagnostic Breast Cancer, and Surveillance, Epidemiology, and End Results (SEER) programme breast cancer datasets have all been examined to assess the efficacy of the proposed model. The experimental results have demonstrated that the WAUCE model outperforms five different ensemble processes and two widely used ensemble models, namely adaptive boosting and bagging classification tree, in terms of accuracy and variance for the detection of breast cancer. Comparing the proposed WAUCE model to the best single SVM model on the SEER dataset, the latter has improved the accuracy by 33.34% and decreased variance by 97.89%. Khuriwal and Mishra (2018) have applied an adaptive ensemble voting approach for diagnosing breast cancer using Wisconsin Breast Cancer database. The study showed how ANN and logistic algorithms have outperformed ensemble machine learning techniques for identifying breast cancer, even when the number of variables were reduced. The results showed that the ANN technique using the logistic algorithm achieved 98.50 percent accuracy compared to another machine learning algorithm. Yin et al. (2019) have presented a unique SVM parameter optimization approach based on the advanced whale optimization algorithm (AWOA), which is an enhanced whale of algorithm (WOA) with an external archiving strategy. On the basis of AWOA, a novel framework for SVM parameter optimization was constructed. To illustrate the efficacy of the proposed strategy, six representative data sets were selected to assess the impact of the SVM classification problem. Experimental data have indicated that AWOA can achieve higher accuracy and better convergence than the three conventional parameter optimization approaches (WOA, PSO, and DE). Using a hybrid optimization-based Support Vector Machine, Mallika and Selvamuthukumar (2021) have constructed an efficient approach for diagnosing diabetes (SVM). The proposed hybrid optimization approach combined the Crow Search algorithm (CSA) and the Binary Grey Wolf Optimizer (BGWO) in order to fully harness the potential of SVM in the diabetes detection system. The efficacy of the proposed hybrid optimization-based SVM (henceforth CS-BGWO-SVM) technique was thoroughly investigated using the real-world databases, such as the UCIPima Indian standard dataset and the diabetes type dataset from the Data World repository. The results of empirical study demonstrated that CS-BGWO-SVM is a more efficient classification method with exceptional accuracy.

Wang et al. (2015) have proposed using a learning model based on SVM for the diagnosis of breast cancer. The model comprised of six alternative kernel functions and SVM structures, including an a-SVM and a C-SVM. For model hybridization for different base classifiers, a Weighted Area Under the Receiver Operating Characteristic Curve Ensemble (WAUCE) technique was proposed. A number of datasets were evaluated, including the Wisconsin Breast Cancer (WBC), Wisconsin Diagnostic Breast Cancer (WDBC), and Surveillance, Epidemiology and End Results (SEER) databases. Comparing the created model to prior studies based on a single SVM reveals a considerable increase in diagnostic accuracy. The method's ostentatiousness and lack of training time are the drawbacks found.

3. Grey Wolf Optimization algorithm (GWO)

Mirjalili et al (2014) have proposed Grey Wolf Optimization algorithm (GWO) based on the concept of grey wolf hunting mechanisms and social hierarchy. In GWO algorithm, grey wolves are classified into four categories viz. alpha (α) which is the leader, beta (β) which assists the leader, delta (δ) which follows both prior wolves, and omega (ω) (Mirjalili et al 2014).

There is a social hierarchy in the operation of Grey Wolf Optimization algorithm. The alpha (a) is considered to be the best solution, followed by the beta (b), and the delta (d) is considered to be the third-best option. The candidate's solution that is still in left is omega (w). These wolves (w) are below the three wolves that came before them in the hierarchy. Figure 1 depicts the hierarchy of grey wolves.

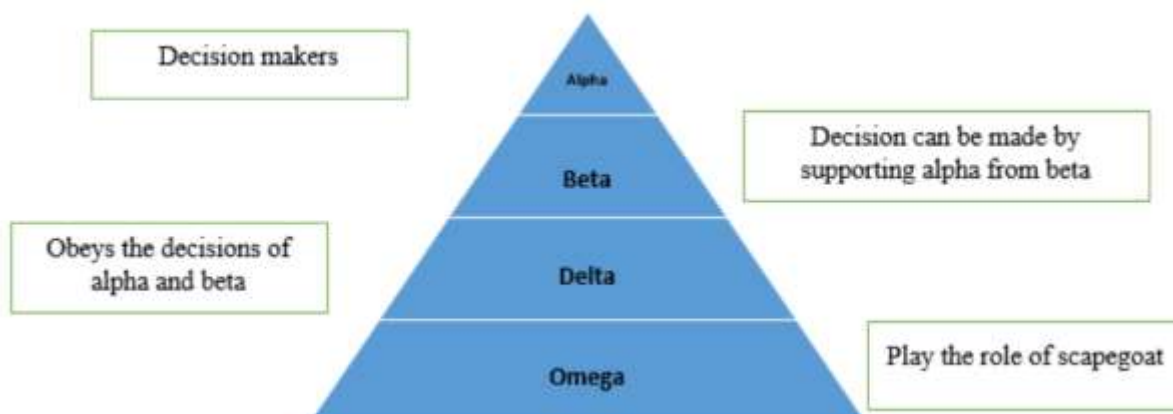


Figure 1 Hierarchy of Grey Wolves

In general, Grey wolves encircle their prey for hunting by using based on the equations (1) and (2).

$$D = |C \cdot X_p(t) - X(t)| \quad (1)$$

$$X(t + 1) = X_p(t) - A \cdot D \quad (2)$$

Where, D is the distance between the grey wolf and its prey. “t” represents the number of iteration. X_p represents the positional location of the prey and X represents positional location of the grey wolf.

$$A = 2ar_1 - a \quad (3)$$

$$C = 2r_2 \quad (4)$$

“a” represents a vector whose values decline linearly from 2 to 0 over the length of run, and r1 and r2 are random vectors within the interval [0, 1]. The values of A and C vectors determine how close the wolves will be to the prey.

3.1 Hunting Process

After the encirclement process, a grey wolf begins searching for the optimal option. Even if the optimal answer needs optimization, alpha wolf saves the optimal solution in each iteration and updates it if it is improved. Beta and delta are able to pinpoint the position of the prey. Consequently, each subspecies of grey wolf stores the optimal solution and uses the following equations to update the position of grey wolves. The position updation of grey wolves is represented in the equations (5) to (11).

$$D_a = |C_1 \cdot X_a(t) - X(t)| \quad (5)$$

$$D_b = |C_2 \cdot X_b(t) - X(t)| \quad (6)$$

$$D_d = |C_3 \cdot X_d(t) - X(t)| \quad (7)$$

$$X_1 = X_a(t) - A_1 \cdot D_a \quad (8)$$

$$X_2 = X_b(t) - A_2 \cdot D_b \quad (9)$$

$$X_3 = X_d(t) - A_3 \cdot D_d \quad (10)$$

$$X(t + 1) = X_1 + X_2 + X_3 \quad (11)$$

Here, $X_a(t)$, $X_b(t)$, $X_d(t)$ are the position vectors of three best solutions at a given iteration t . The coefficient vectors like A_1 , A_2 , A_3 , are essential to the operation of GWO. The exploitation portion is favoured by the coefficient values of A that lie between -1 and 1, which forces the search agents to converge (attack) in the direction of the prey. The search agents are forced to deviate from the prey in quest of a better or more effective solution when the coefficient values are greater than one or less than equal to one favouring the exploration phase. Coincident vector C is another GWO control parameter that encourages exploration. This parameter's value always falls between 0 and 2. It controls the role that the prey plays in determining the next location.

3.2 Exploitation Phase - Attacking Prey

In this part of the hunting process, a grey wolf attempts to halt the movement of its prey in order to attack it. This technique is implemented by lowering the value of a. The value of A is also decreased by the value a and lies between -1 and 1. The grey wolf can attack the victim if A is larger than -1 and less than 1. However, GWO suffers from stagnation in the local optimum, and researchers are attempting to identify several strategies to address this issue (Mirjalili et al., 2014).

3.3 Exploration Phase – Searching for Prey

Alpha, beta, and delta have an effect on the search method. These three groups are distinct from one another. Consequently, they require a mathematical equation in order to converge and assault prey. If the value of A is larger than 1 or less than -1, the search agents are driven to diverge from the prey. In addition, if A is larger than 1, the search agent attempts to locate

superior prey. C is an additional component element that affects the exploration phase in GWO (Mirjalili et al., 2014). When $C > 1$, the solution gravitates more toward the prey, and this contribution is hence considerable.

In summary, the GWO algorithm generates the random population. Alpha, beta, and delta assume the prey's position. The gap between possible solutions is then modified. After then, A is decreased from 2 to 0 to achieve equilibrium between the two phases. If A is greater than one, the search agents stop attacking the victim. If A is less than 1, then they pursue the prey. The GWO has achieved a successful conclusion and is now ended. Algorithm 2 covers the GWO algorithm in depth (Mirjalili et al., 2014).

Table 1 Pseudocode for GWO Algorithm

- 1: Initialize the grey wolfs (search agents)
- 2: Initialize a , A and C
- 3: Calculate the fitness value using eqn. 1
- 4: Compute the values for search agents X_α , X_β , X_δ
- 5: While ($iter < iter_{max}$)
- 6: **For each search agent**
- 7: Update the location of the present search agent by (11)
- 8: **End for**
- 9: Update a , A and C
- 10: Search agents evaluated by a fitness
- 11: Update X_α , X_β , X_δ
- 12: $iter = iter + 1$
- 13: **End While**
- 14: Display X_α and fitness value

4. Modified GWO

Traditional GWO tends to fall into the local optimum solution for complicated optimization issues. In this paper, conventional GWO method was modified by incorporating the population dynamic evolution operator and nonlinear convergence factor to successfully prevent slipping into the local optimum. By incorporating the population dynamic evolution operator, the search range of wolves in the GWO algorithm could be enlarged to the whole solution space with each iteration, hence increasing the likelihood of discovering the global optimal solution.

$$X_\alpha = X_1 \pm (B_u - B_l \cdot n + B_l) \quad (12)$$

$$X_\beta = X_2 \pm (B_u - B_l \cdot n + B_l) \quad (13)$$

$$X_\delta = X_3 \pm (B_u - B_l \cdot n + B_l) \quad (14)$$

where B_u is the upper band of search space and B_l is lower bound of the search space. and n is a random number with range between 0 and 1. The updated optimal solution vector is obtained by adding the three vectors X_α , X_β and X_δ .

5. SVM Algorithm

Boser, Guyon, and Vapnik (1992) have proposed the SVM classification method. The SVM method is commonly utilised in bioinformatics because to its high accuracy and capacity to manage data with vast dimensions (Cristianini & Shawe-Taylor 2000). SVM aims to

maximise the margin by discovering a hyper-plane between two distinct data categories. The hyper-plane linear model is described by the following equation:

$$f(x) = \text{sign}(W^T X + b) \quad (11)$$

Where, w = weight vector, b = bias, term x = input vector.

6. Experimental Design

6.1 Dataset Description

Wisconsin Diagnostic Breast Cancer (WDBC) database retrieved from the UCI Machine Learning Repository was used to validate the performance of the proposed Modified GWO - SVM technique. The WDBC dataset consists of 569 records, of these, 357 patients have been classified as having benign breast cancer, while the remaining patients have been classified as having malignant breast cancer. 32 features make up each record which comprises of a patient ID, a diagnosis, and 30 real-valued attributes. These parameters define the features of the cell nuclei that the digital picture of the FNA of the breast mass captured. The ten distinct characteristics of each cell nucleus are represented by the 30 real valued qualities, which are the radius, texture, perimeter, area, smoothness, compactness, concavity, concave point, symmetry, and fractional dimension. The mean value, standard error, and maximum value for each characteristic have all been recorded. The WDBC dataset's ten feature categories are summarised in Table 2.

Table 2 Summary of WDBC Dataset

Attributes	Measurement range		
	Mean	Standard error	Maximum
Radius	6.99–28.12	0.121–2.923	7.95–37.01
Texture	9.80–40.02	0.37–4.90	112.10–50.01
Perimeter	44.02–189.09	0.80–22.01	50.48–252.03
Area	144.04–2503.01	6.90–543.10	186.01–4255.00
Smoothness	0.054–0.164	0.003–0.035	0.072–1.102
Compactness	0.020–0.350	0.002–0.138	0.030–1.060
Concavity	0.001–0.501	0.000–0.400	0.000–1.255
Concave points	0.0001–0.202	0.000–0.055	0.000–1.296
Symmetry	0.108–0.305	0.009–0.080	0.158–0.668
Fractal dimension	0.051–0.098	0.001–0.031	0.057–0.210

7. Results and Discussion

The classification performance of the proposed model was determined by measuring various performance measures. The measures considered for assessing the performance of the proposed MGWO and SVM approach includes Accuracy, Sensitivity, Specificity, Precision, Recall, and F-Measure. The proposed algorithm was implemented using Matlab R2020a and LIBSVM (Version 3.3) (Chang 2011).

The primary purpose of this work was to improve classification performance and increase the accuracy of diagnosis of breast cancer by lowering the size of characteristics utilising the modified GWO. For a more accurate evaluation of the suggested technique, the following three cases were taken into account:

Case 1: Training: 60% of data; Testing: 40% of data

Case 2: Training: 70% of data; Testing: 30% of data

Case 3: Training: 80% of data; Testing: 20% of data

Initially, the proposed MGWO-SVM algorithm was evaluated using the Cancer dataset with all the ten features of the original dataset. The parameters of evaluation were computed and tabulated in Table 3. **Table 3 Classification Performance**

Case	Accuracy		Sensitivity		Specificity		Precision		F-measure	
	GWO-SVM	MGWO-SVM	GWO-SVM	MGWO-SVM	GWO-SVM	MGWO-SVM	GWO-SVM	MGWO-SVM	GWO-SVM	MGWO-SVM
Case 1	97.28	98.14	98.29	99.31	95.10	96.19	98.74	99.44	98.43	99.37
Case 2	96.24	97.19	95.50	96.49	97.51	98.40	97.20	98.19	96.33	97.33
Case 3	95.82	96.71	96.00	97.91	93.12	94.02	96.84	97.90	96.84	97.90

From the above table it is inferred that the proposed MGWO-SVM approach has recorded better results across the performance parameters like accuracy, Sensitivity, Specificity, Precision and F-measure when compared with GWO-SVM. Similarly, Case 1 scenario where the training data (60%) was higher than the testing data (40%) has recorded the best results.

8. Comparison with Previous Studies

To demonstrate the resilience of the proposed model, the results obtained from the proposed method was compared to the results of several recently developed models. Table ** discusses the results of the other recently developed models using the Wisconsin Breast Cancer Dataset (WDBC). Kamel et al. (2019) employed SVM as a classifier on the WDBC Dataset and determined its accuracy to be 85.89%. Adjusted BAT Algorithm for the WDBC Dataset (Tuba et al. 2016) has recorded an accuracy of 96.88%. Improved SVM parameters using Particle Swarm Optimization (PSO) proposed by Alhakbani, and al-Rifaie (2017) has attained an accuracy of 92.98%. Wang and Chen (2020) optimised SVM parameters using the Whale Optimizer Algorithm (WOA) in 2020 and attained an accuracy of 96.65%. Particle Search and Grid-SVM proposed by Ali and Abdullah (2020) recorded an accuracy of 97.45%. Support Vector Machine Parameters Optimization by Enhanced Fireworks Algorithm (EFA-SVM) proposed by Rube et al. (2016) recorded an accuracy of 96.31%. It is obvious from Table 4 that the proposed MGWO-SVM algorithm has produced the excellent accuracy and outperformed the other existing methods for breast cancer diagnosis.

Table 4 Performance Comparisons

Algorithm	Source	Accuracy (%)
SVM	Kamel et al. (2019)	85.89
PSO-SVM	Alhakbani & al-Rifaie (2017)	92.98
ABA-SVM	Tuba et al. (2016)	96.88
WOA-SVM	Wang & Chen (2020)	96.65
Grid-SVM	Ali & Abdullah (2020)	97.45
EFA-SVM	Tuba et al. (2016)	96.31
MGWO-SVM		97.28

9. Conclusions

An efficient model for inference in the field of medical diagnosis may be provided by the integration of multidimensional data with various categorization, feature selection, and dimensionality reduction algorithms. Data classification inside the clinical diagnosis system enables medical professionals to discover hidden data samples by training a vast quantity of actual datasets. Therefore, it is indispensable for timely prediction, analysis, and management

of breast cancer using an affective and reliable good breast cancer diagnosis model. Such a model will greatly benefit medical professionals in accurate prediction and treatment of cancer patients. Breast cancer is curable if it is found early, which may greatly improve the prognosis.

In this study, GWOs and SVM were employed to boost the accuracy of breast cancer diagnosis by selecting the appropriate features and data classification. The algorithm was developed and implemented using MATLAB and UCI dataset was used for experimental verification of the algorithm. This work has successfully demonstrated the application of grey wolf optimization algorithm (GWO) in enhancing the performance of support vector machine (SVM) for applications like breast cancer diagnosis.

Combining the SVM algorithm and the GWO to identify the subset of efficient features yielded the best results. Compared to previous algorithms, the accuracy, sensitivity, and specificity of the suggested technique were comfortably higher. This perspective can be utilised in the diagnosis of heart disease, diabetes, and other conditions in future research. Future study will focus particularly on comparing the proposed model to additional medical datasets. This work is a part of a larger study in which future enhancement focus on including feature selection techniques in the proposed algorithm and comparing the classification performance with other prominent models.

References

- Basir, M. A., & Hussin, M. S. (2021, September). Exploitation of Meta-Heuristic Search Methods with Bio-Inspired Algorithms for Optimal Feature Selection. In *2021 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAJET)* (pp. 1-5). IEEE.
- Tuba, E., Tuba, M. and Simian, D., 2016, July. Adjusted bat algorithm for tuning of support vector machine parameters. In *2016 IEEE congress on evolutionary computation (CEC)* (pp. 2225-2232). IEEE.
- Mirjalili, S., Mirjalili, S.M. and Lewis, A., 2014. Grey wolf optimizer. *Advances in engineering software*, 69, pp.46-61.
- Cristianini, N. and Shawe-Taylor, J., 2000. *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press.
- Wang, H., Zheng, B., Yoon, S.W. and Ko, H.S., 2018. A support vector machine-based ensemble algorithm for breast cancer diagnosis. *European Journal of Operational Research*, 267(2), pp.687-699.
- Rojas-Domínguez, A., Padierna, L.C., Valadez, J.M.C., Puga-Soberanes, H.J. and Fraire, H.J., 2017. Optimal hyper-parameter tuning of SVM classifiers with application to medical diagnosis. *Ieee Access*, 6, pp.7164-7176.
- Yin, X., Hou, Y., Yin, J. and Li, C., 2019, June. A novel SVM parameter tuning method based on advanced whale optimization algorithm. In *Journal of Physics: Conference Series* (Vol. 1237, No. 2, p. 022140). IOP Publishing.
- Mallika, C. and Selvamuthukumar, S., 2021. A hybrid crow search and grey wolf optimization technique for enhanced medical data classification in diabetes diagnosis system. *International Journal of Computational Intelligence Systems*, 14(1), pp.1-18.
- Singh, I., Bansal, R., Gupta, A. and Singh, A., 2020, November. A Hybrid Grey Wolf-Whale Optimization Algorithm for Optimizing SVM in Breast Cancer Diagnosis. In *2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC)* (pp. 286-290). IEEE.

- Khuriwal, N. and Mishra, N., 2018, March. Breast cancer diagnosis using adaptive voting ensemble machine learning algorithm. In *2018 IEEMA engineer infinite conference (eTechNxT)* (pp. 1-5). IEEE.
- Maglogiannis, I., Zafiroopoulos, E. and Anagnostopoulos, I., 2009. An intelligent system for automated breast cancer diagnosis and prognosis using SVM based classifiers. *Applied intelligence*, 30(1), pp.24-36.
- ling Chen, H., Yang, B., jing Wang, S., Wang, G., zhong Li, H. and bin Liu, W., 2014. Towards an optimal support vector machine classifier using a parallel particle swarm optimization strategy. *Applied Mathematics and Computation*, 239, pp.180-197.
- Ali, A.H. and Abdullah, M.Z., 2020. An efficient model for data classification based on SVM grid parameter optimization and PSO feature weight selection. *International Journal of Integrated Engineering*, 12(1), pp.1-12.
- Chang, C.C., 2011. " LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, 2: 27: 1--27: 27, 2011. <http://www.csie.ntu.edu.tw/~cjlin/libsvm, 2>.
- Alhakbani, H.A. and al-Rifaie, M.M., 2017, September. Optimising SVM to classify imbalanced data using dispersive flies optimisation. In *2017 Federated Conference on Computer Science and Information Systems (FedCSIS)* (pp. 399-402). IEEE.
- Kamel, S.R., YaghouzZadeh, R. and Kheirabadi, M., 2019. Improving the performance of support-vector machine by selecting the best features by Gray Wolf algorithm to increase the accuracy of diagnosis of breast cancer. *Journal of Big Data*, 6(1), pp.1-15.
- Wang, M. and Chen, H., 2020. Chaotic multi-swarm whale optimizer boosted support vector machine for medical diagnosis. *Applied Soft Computing*, 88, p.105946.
- Tuba, E., Tuba, M. and Beko, M., 2016, June. Support vector machine parameters optimization by enhanced fireworks algorithm. In *International Conference on Swarm Intelligence* (pp. 526-534). Springer, Cham.
- Wang, H.; Yoon, S.W. Breast cancer prediction using data mining method. In *Proceedings of the IIE Annual Conference Expo 2015, Nashville, TN, USA, 30 May–2 June 2015*; pp. 818–828.
- UCI. Breast Cancer Wisconsin Dataset. Available online: [https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+\(Diagnostic\)](https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic)) (accessed on 9 June 2019).
- Sotiriou, C., Neo, S.Y., McShane, L.M., Korn, E.L., Long, P.M., Jazaeri, A., Martiat, P., Fox, S.B., Harris, A.L. and Liu, E.T., 2003. Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proceedings of the National Academy of Sciences*, 100(18), pp.10393-10398.
- Jazaeri, A.A., Yee, C.J., Sotiriou, C., Brantley, K.R., Boyd, J. and Liu, E.T., 2002. Gene expression profiles of BRCA1-linked, BRCA2-linked, and sporadic ovarian cancers. *Journal of the National Cancer Institute*, 94(13), pp.990-1000.