

Swarm Intelligence Algorithms in Gene Selection Profile Based on Classification of Microarray Data: A Review

Abstract

Microarray data plays a major role in diagnosing and treating cancer. In several microarray data sets, many gene fragments are not associated with the target diseases. A solution to the gene selection problem might become important when analyzing large gene datasets. The key task is to better represent genes through optimum accuracy in classifying the samples. Different gene classification algorithms have been provided in past studies; after all, they suffered due to the selection of several genes mostly in high-dimensional microarray data. This paper aims to review classification and feature selection with different microarray datasets focused on swarm intelligence algorithms. We explain microarray data and its types in this paper briefly. Moreover, our paper presents an introduction to most common swarm intelligence algorithms. A review on swarm intelligence algorithms in gene selection profile based on classification of Microarray Data is presented in this paper.

Keywords: Microarray data, Gene Selection, Classification, Feature Selection, Swarm Intelligence Optimization.

I. INTRODUCTION

Cancer research continues to be one of the leading medical research fields which is one of the oldest sciences [1]. Research in cancer causes includes a wide range of areas. For the purpose of analyzing the causes and potential treatment purposes of diseases, several biological microarray experiments have already been performed. Threaten tumor diagnosis is critical as it is commonly quite difficult to treat patients who have been diagnosed in final phases. They are known to have important keys for identifying the fundamental problems of cure and disease prevention. Scientists are known for elucidating important key knowledge that is essential to the cure and prevention of diseases [3].

Deoxyribonucleic Acid (DNA) microarray data holds a major role in areas such as Molecular Biology and Medicine [4,5,6,7,8,9,10]. Profiles on gene expression may provide more details on accurate classification from cancer samples. This can be used not only for prediction but also for diagnosis, understanding and prognosis of the disease [11,12]. Gene expression microarray allows getting information about the expression levels of thousands of genes all at once. Microarrays are useful tools to investigate different aspects of gene

expression [13,14]. Gene classifications attempt to choose the genes are expressed or regulated. This technology can be beneficial for disease prognosis and the development of models to detect and predict diseases from genetic profiles [4,15]. The technology also tackles group similar data problems as regards the small number of usable features. Throughout this area, it is best to identify appropriate features. Decisive stages to reduce data dimensionality are necessary [16]. It is not very easy to identify particular types of cancerous cells because not enough DNA is required. If the standard data dimensionality is high, the classification would be difficult and less efficient [17,18].

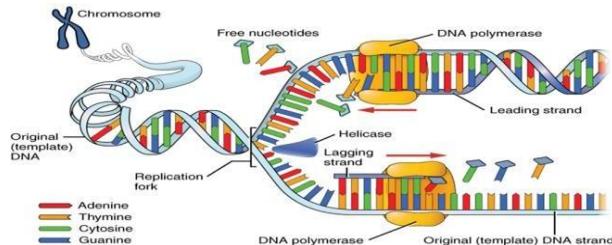


Fig. 1. DNA molecule and gene segment [19].

Cancer classifications are included in characteristics of the tumors and not on their areas. Tumors may grow wherever they like in any cell type, and this contributes primarily to the fact that the classification of brain cancer is so difficult. Decisive stages to reduce the data dimensionality are necessary [20,21]. Several other genetic markers of glioblastoma ability to survive have been identified in the recent study [22,23,24,25].

Feature selection (FS) is a pre-processing step that aims to improve performance and to facilitate classification and clustering processes. FS is a kind of multi-objective optimization problem. Its goal is to minimize feature size and maximize classification or clustering performance. FS methods are generally classified into three categories: filters, wrappers, and hybrid methods. In filter approaches, features are ranked according to a specific criterion like Chi-squared test, information gain etc. and those with high rank are selected. Wrapper methods evaluate the performance of feature subsets using a learning algorithm. The features in subsets can be selected sequentially [26,27].

Swarm intelligence (SI) is cooperation between information-bearing particles, each of which is ruled by a different kind of algorithm. Typically, so called SI structures are made up of multiple agents, the way of which is that of a wide population of basic agents that are in contact with one another and with the world. The inspiration tends to come from nature, particularly from biological systems. The agents have very basic laws that are rigidly enforced for any specific one, and while there is a structured command system that governs what the individual agents do throughout their lives, each agent is an independent agent that determines when to communicate with another agent and when not to, contributing to a group action that is unfamiliar to the individual agents. (In the wild) some examples of social intelligence involve ant colonies, bird flocking, animal herding, bacterial development, and fish schooling [28].

In terms of classification accuracy, Particle Swarm Optimization (PSO) is effectively selected compared with other existing selection methods. From the review [28,29,30,31,32] PSO's potential benefits for the feature selection are as shown in:

- PSO seems to have a strong ability to discover until the ideal cure is found. Different parts of the solution are being explored.
- PSO is super beneficial when selecting functionality because the particle swarm has memory, and all particles understand the solution while flying in an area of the problem.
- The reason the cost of implementing PSO is low is that it can be implemented very easily.
- PSO performs to connect ideas from different fields into a solution.
- PSO could also represent discrete and binary data.
- PSO is faster and more efficient than other feature selection technique and therefore does not involve complex vector math.
- PSO is a simple algorithm that is practical and effective, leading to promising results.

- PSO's efficiency is nearly unchanged by the problem dimension.

II. MICROARRAY DATA

According to Global Burden of Cancer, colon cancer is the second most common type of cancer deaths in the world. It is believed that in three American men and one in five American women will develop colon cancer in their lifetime [33,34]. Microarray technology has created a large microarray data collection reflecting gene expression developed from tissue and cell samples obtained. Gene expression data normally receives thousands of genes (sample). Therefore, these data are well known for their high, detailed, and broad range of detail [35,36]. Microarray evidence was instrumental in cancer detection and classification. Most microarray data sets contain thousands of genes, but there are a significant number of genes that do not make any effect on diseases. Intelligent algorithms to select genes are necessary because of microarray technology [37,38]. Microarray technology is an approach to explain the gene-by-gene interactions of genes. In addition to it, the microarray technique can quantify genes activity from the entire genome into one experiment [37,39].

Dementia is a broad group of brain illnesses leading to a decrease in cognition and memory capabilities on a long-term basis and often gradually. Common dementia is this type of form Alzheimer's Disease (AD), that represents 50% of cases [40]. A combination of clinical criteria such as neurological analysis, mental state evaluations and brain imaging is used for preliminary diagnosis for AD [41]. The tests of biomarkers such as amyloid-beta protein, tau protein expression of the Cerebrospinal Fluid (CSF) [42]. However, the procedure is both invasive and the CSF is obtained by lumbar punctures that are painful. Furthermore, the diagnosis of early stages of Alzheimer's or those with mild cognitive impairment may be stressful Mild Cognitive Impairment (MCI). Consequently, non-invasive biomarkers are important to develop, so that the early start diagnosis of AD or MCI may be done to preserve normal brain function [43].

The most common types of cancer in the human body are Leukemia [44]. Leukemia is a cancer of the blood-forming white cells in the marrow. WBC will be present in the blood of patients with cancer and is fatal. Leukemia can be categorized into two (2) types: chronic or acute leukemia. These types were classified based on when the disease starts, and the damage gets worst. Usually, chronic leukemia strikes individuals progressively and as it gets worse it compromises the adult or the elderly. For acute leukemia rapidly becomes critical condition and usually occurs in children. For a person affected by chronic leukemia, it will not appear at early stages as the stage of malignant cancer and thus the person will not have early signs and symptoms for the disease [45]. Chronic Lymphocytic Leukemia (CLL) and Chronic Myelogenous Leukemia (CML) are two kinds of leukemia that are a cause of blood cancer [46].

In early stage of acute leukemia, the irregular cells cannot damage the white blood cells. In the first stage, the leukemia cell increases rapidly and unregulated. Acute Lymphocytic Leukemia (ALL) and Acute Myelogenous Leukemia (AML)

are two main types of leukemic (AML), "Fig. 2", "Fig. 3" indicates the sample of blood microscopic images with different types of leukemia [45]. "Fig. 4" shown visualization of the process in microarray analysis.

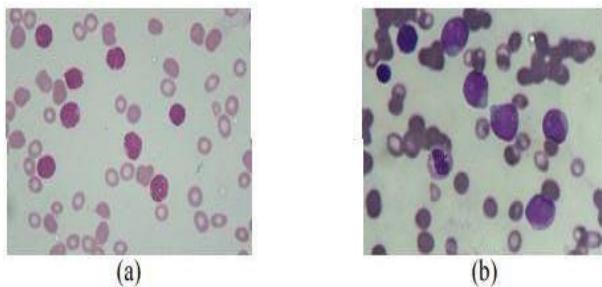


Fig. 2. (a) ALL. (b) AML [45]

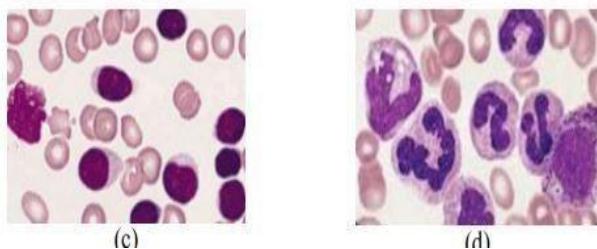


Fig. 3. (c) CLL. (d) CML [45].

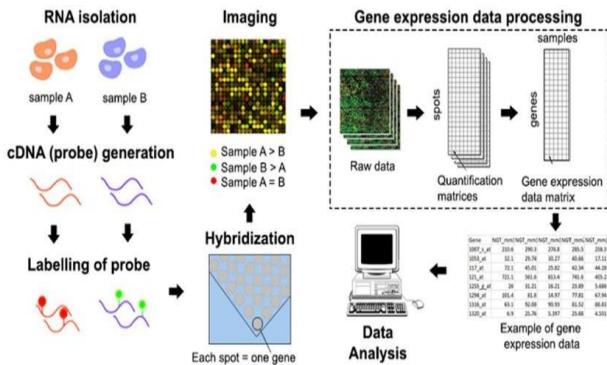


Fig. 4. Visualization of the process in microarray analysis [47]

Optimization algorithms are widely used in several different fields of study, including bioinformatics, biotechnology and biomedicine, which has been used to fix many different problems. Optimization as stated by Banga [48].

III. SWARM OPTIMIZATION

Swarm intelligence is derived from the swarming behavior of contributes. Group dwelling allows organisms to solve the problems which entities cannot or cannot address. Therefore, swarm intelligence is an intelligent way for individuals to solve it to further shortcomings [49,50,51,52].

A. Particle Swarm Optimization (PSO)

One of the latest metaheuristic techniques has been developed by Eberhart and Kennedy in 1995, is the PSO which is a stochastic global bird flocking technique inspired by social behavior. The algorithm models the study of a problem area by an individual or particle population. In PSO, there are a series of possible solutions such as particles [41,53]. The most popular algorithm on the PSO [54]. This was inspired by collective intelligence birds [55]. The PSO's key feature is the easy way to exchange information, based on a few equations, between agents. Here, a particle is called an agent and a group of them form a swarm. The particles of swarm change their positions based on the best location obtained (collective experience). The efficiency of the agents is assessed using a fitness function as normal. This pattern allows the emergence of global behavior from distinct individuals [56]. In an optimization problem, the motions of the particles indicate various solutions to that problem. However, in clustering, it depends on whether the class is more accurate [57,58,59].

Optimization of Particle Swarm is a heuristic algorithm focused on social behavior experiment, as suggested [60]. PSO is originally applied to mimic the behavior and data exchange of flying birds to solve the issues. Fig. 5 shown the flowchart of general PSO algorithm.

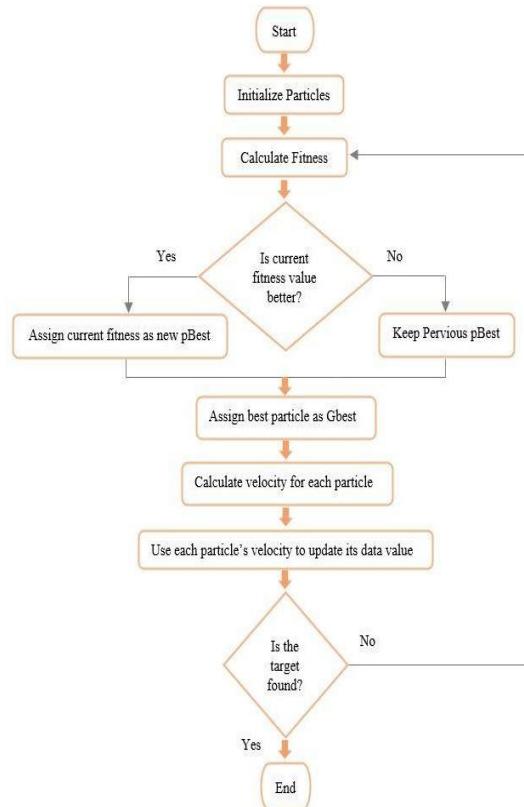


Fig. 5. Flow Chart of General PSO Algorithm [61].

PSO PSEUDO CODE

Stage 1: Initialization

The PSO is begun by a set of initial population. The primary population is particles by random positions and velocities.

Stage 2: Evaluation

For each particle, its fitness value is computed. Check each particle's fitness value one through one. Evaluate by the present best fitness value.

Stage 3: Compare for Obtain to Pbest

Evaluate and compare every particle fitness evaluation by the current particles to get Pbest.

Stage 4: Compare for Obtain to Gbest

Evaluate and compare fitness evaluation by the population's overall previous best to get Gbest.

Stage 5: Calculate of Function

Calculate of flowing functions for all particles.

$$Vid = w*Vid + \text{rand}() * (Pid - Xid) + C2 * \text{rand}() * (Pid - Xid) Xid(K+1) = Xid(K) + Vid(K+1)$$

Stage 6: Stopping criteria

In this step, the stage 2 and 5 are iterated till the predefined number of generations has been achieved. The best answer can be generated after termination [61].

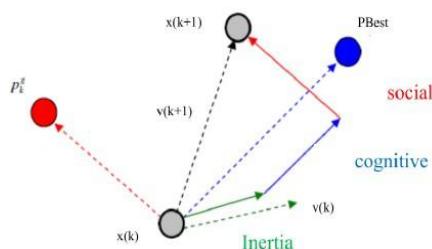


Fig. 6. The mechanism to handle particles movement [62].

One can observe that there have been three routes through which the particles move: inefficiency, cognitive and social. (see Fig. 6). To determine how far the particle moved, length it travelled in three different directions is considered [62].

B. Artificial Bee Colony (ABC)

The Artificial Bee Colony Algorithm was introduced in 2009, which is a swarm-based heuristic method. ABC is a multi-dimensional optimization algorithm that imitates the foraging behavior of bees. According to this algorithm, the purpose of the bees is to maximize the number of nectar sources and minimize the distance of the sources. For optimization problems, sources are represented by vectors. The vector dimension is the parameter number of the problem. Each source is a possible solution to the problem, and the source quality is represented by the amount of nectar; this is called the fitness value. Each source has a trial value and is set to 0 in the initialization phase. When a source is improved, this value remains the same; otherwise, it is incremented by 1. A colony has three types of bees: employed bees, onlooker bees, and scout bees. There is an equal number of employed and onlooker bees, whereas there is always a single scout bee. The algorithm

includes four steps [63,64]. See "Fig. 8" the typical behavior of honeybee foraging.

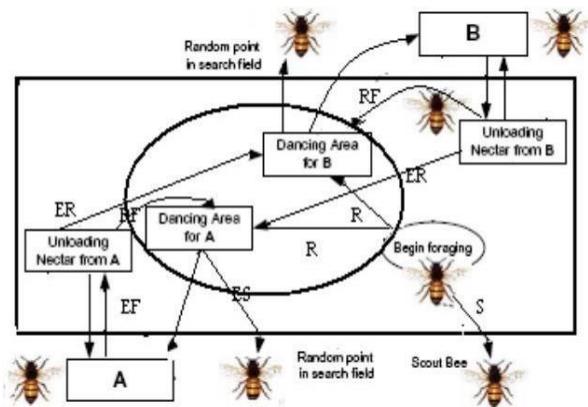


Fig. 7. Typical behavior of honeybee foraging [65,66]

C. Ant Colony Optimization (ACO)

Ant colony algorithm has been successfully implemented for solving complex optimization problems. An ant colony algorithm, which was inspired by ants who collect ant corpses and sort the ant larvae. Ant colony clustering has been successfully applied to numerous fields including image segmentation, text classification, and stock price prediction. While the classical ant colony algorithm has proved its usefulness in a variety of situations, it has drawbacks that need to be addressed. Some modifications are made to address these problems [67]. The flowchart of ACO algorithm shown in "Fig 8".

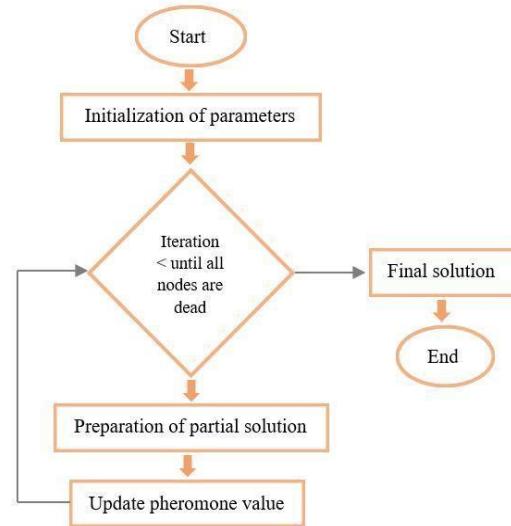


Fig. 8. Flowchart of ACO [67].

IV. RELATED WORK

Several researchers utilized evolutionary computation to develop solutions for selection problems. Lately, metaheuristic algorithms are used to carry out genetic selection, and their implementation has been studied. Despite the several methods suggested for gene selection, however, many other of them are, however, affected by local optimal stagnation problems and high calculation cost, which therefore cannot guarantee the optimal and significant use of metaheuristic algorithms in a huge search area of identified genes [68,69,70]. The researcher in [71] introduces the method for feature selection of Qualitative Mutual Information (QMI). In QMI, Random Forest Calculation is done, which separates correlated characteristics and reduces the gene redundant data. To label off the genes which are irrelevant, mutual information was utilized. MI tries to uncover each class variation on the variable. Random forest is an option for calculating the preference score in a time-intensive way, and there is a high chance of overfitting. The developed by [71] Considerable noise factor exists in the data; certain functions will be biased. This study discusses an enhanced version of cuckoo search and how it compares with cuckoo search and with harmony search.

However, in cuckoo search, more than one limit needs to be improved to make the algorithm work better. But the proposed [72] Cuckoo searches showed credibility only as a feature with

minimum features and evaluated. on the KNN classifier with an average accuracy of 99.13 percent of the data. Among the most often used methods for feature selection in particle swarm optimization, and previously, it has been advanced., it has been introduced by [73,74,75] for microarray data. Reference [54] using PSO for best fit functionality. PSO investigates the space for features and the number of features in the breast cancer data to select significant features. While [57] Pearson's correlation coefficient with PSO compared with Genetic Algorithm (GA) and a comparison between PSO and BPSO. Proposed by [55] proved Binary Particle Swarm Optimization (BPSO) has a better performance compared to GA for FS. Here, to handle different variables, BPSO is introduced for this research. While the authors in [56]. The authors used Multi-population Particle Swarm Optimization (MPSO) for feature selection to classify significant genes. MPSO develops searches better when compared to PSO. This is due to the PSO weakness that continues to fall into the local optima pit as well as the inability to try more varied solutions. Other than PSO, [57] He implemented ABC technique to choose the best genes for each binary optimal solution.

ABC becomes imitative of the bees' behavior, and the bees are to maximize the number of sources of nectar while reducing the path length of them. Introduced by [76], feature selection is an optimization involving multiple objectives, which enhances

TABLE .I THE REVIEW OF SWARM OPTIMIZATION ALGORITHMS FOR MICROARRAY DATA CLASSIFICATION

Ref.	Year	Methods	Problems	Datasets	Results	Feature Selection
[82]	2020	Their hybrid multi-objective cuckoo search algorithm was developed using evolutionary operators.	Investigate and select relevant genes with maximum accuracy.	<ul style="list-style-type: none"> Ovary Lung SRBCT CNS DLBCL Prostate Leukemia 	<ul style="list-style-type: none"> Ovary 97.5% Lung 93.7% SRBCT 100% CNS 73.7% DLBCL 94.5% Prostate 92.2% Leukemia 98.8% 	Evolutionary operators.
[83]	2020	<ul style="list-style-type: none"> PSO-SVM. PSO-ENSV. 	To choose the best subset of genes and reduce the dimensionality of the results.	<ul style="list-style-type: none"> Prostate Lung Breast Colon Breast 	PSO-ENSV is better than PSO-SVM and SVM with RBF kernel.	PSO-ENSV.
[84]	2020	<ul style="list-style-type: none"> Relief-F. Fishers Score. Information Gain. SNR 	It causes high computational complexity and low accuracy.	<ul style="list-style-type: none"> Prostate 	SNR with WOA 99.48% Relief-F test with MBOA 99.22 %	Wavelet.
[85]	2020	<ul style="list-style-type: none"> GA RF 	high dimensionality and low sparsity issues by feature selection.	<ul style="list-style-type: none"> Colon Leukemia DLBCL Ovarian CNS Prostate Breast 	<ul style="list-style-type: none"> Colon 96.77% Leukemia 100% DLBCL 100% Ovarian 100% CNS 93.33% Prostate 98.04% Breast 86.60% 	A multiple filter and GA wrapper-based hybrid approach (MF-GARF) that incorporates Random forest as fitness evaluator

[86]	2020	DSSA Combined with KNN	Population diversity and local adaptation to feature selection.	DSSA was tested on 23 datasets compared with the original SSA.	Several algorithms had been compared by a single way (PSO, GA, ALO and GOA). DSSA is to constantly improve and compare process for selecting the most important features.	Dynamic Salp swarm algorithm
[87]	2020	• ANN • ANN-GA	The microarray data are usually high-dimensional. the small number of sample sizes and the coverage of irrelevant variables. gene noise. Thus, getting information about the data sets and seeking a correlation among attributes can be difficult.	• Colon Tumor • Prostate Tumor • Lung Cancer	• Colon Tumor 86% • Prostate Tumor 82% • Lung Cancer 100%	Principal component analysis
[88]	2018	• GA-RBF	To detect cancer on different magnified images, automation must be used.	• Aster • Medcity	• 85.4 %.	Dual-tree and double-density 2-D wavelet transform and its coefficients.
[89]	2018	• ReliefF • PSO and SVM	Select a good feature combination in high dimensional small DNA microarray data.	• Colon. • SRBCT • Leukemia • Lung.	• Colon 98.1% • SRBCT 95.2% • Leukemia 90.2% • Lung 86.5%	ReliefF and PSO (RefPSO).
[90]	2018	• A Novel Complex Network and Hybrid GP+PSO	Find the optimal structure with the specified parameters.	• Leukemia. • Colon. • Lung. • Ovarian.	• Leukemia 99.6% • Colon 95.52% • Lung 100% • Ovarian 99.4%	• Pearson's Correlation • Spearman's Correlation • Euclidean Distance • Cosine Coefficient Fisher-ratio
[91]	2018	• ABC-GA	classify tumor type by microarray gene selection records	• Leukemia. • Colon.	• Leukemia 95.6% • Colon. 84.9%	Support Vector Machine recursive feature elimination

the predictive accuracy rates through the optimization. Through [77] Comparing ABC and CS-ABC, CS-ABC performed better than ABC with more accurate results. The premature convergence is handicapped by ABC. This needs CS help to replace its bad solutions. Other than that, [78] has used two archives guided multi-objective artificial bee colony algorithm for cost-sensitive feature selection. Two search mechanisms are utilized to improve the search capabilities of the algorithm. Nevertheless, compared to CS with a bee colony, CS has greater search efficiency due to its nature as a better entity than a bee colony. Thus, by having two archives, it will be a highly intricate and complex challenge. The idea that EO should be used as gene selection with a metaheuristic algorithm was also developed based on research [79,80]. These investigations have shown the benefits of classical operators who can help in methods of classification

EO such as crossover and mutation operators are the basis of the genetic algorithm's evolution. Crossover probabilistic selects two chromosomes based on fitness values from the current population and combines them to generate offspring [81]. Since an operator ensures a permanent fastening of the

population by reversing the value of bits of selected chromosomes into random positioning [82]. The application of various operators in a search space to provide the best search results and reduce the time it takes to search [83].

V. DISCUSSION

The developments in DNA microarray technology have contributed to faster research work of multiple genes at the same time. But big data in gene expression analysis do not seem to be able to come up with a reasonable number of recognized gene expression. All by ourselves in this survey, we analyzed several different papers to find out which profile genes help to classify microarray signals. The current methods require that there are only a few of the algorithms for handling the data, but a large percentage are processed. The statistic shows that nearly every "PSO" algorithm is used in almost every situation in the world. Based on what we know about swarming and swarm intelligence, we believe that pursuing the quest of analyzing the broad spectra of existing swarm algorithms for gene selection profile is, at best, a probable waste of time. However, the fact is that some of the latest algorithms, e.g., PSO and GA are also

used in combination with some other techniques and present very promising results in gene selection profile. Some of the novel algorithms that were developed still provided enough improvement in diagnosis/treatment to be considered useful and create an impact. Thus, as this paper intends, it is to support a researcher in picking out the most suitable algorithm and its setting for data mining tasks and endeavors. For this purpose, alongside the characteristics of the abundance of swarm algorithms for gene selection profile, it may well help to be aware of various practical implications of these algorithms and their use. Some researchers claim that different modifications, adaptations, and hybrids have worked very well to develop a very useful and practical artificial intelligence.

VI. CONCLUSION

Microarray devices can build databases of cancerous cells. Generally, training datasets used to classify cancer cells are rather small when compared to the huge number of genes involved. This paper provides a systematic survey of gene selection profiles for the classification of microarray results. Along with the review of machine learning algorithms, this research area has been primarily based on a limited number of gene fragments. We first studied several papers and tagged them as per table. The gene selection profile is described in each research paper of intelligent classification. We also showed a comparison between different datasets to show the comparison of the usage of data mining tools.

ACKNOWLEDGMENT

The authors would like to acknowledge Duhok Polytechnic University for providing all financial support and support for this study.

REFERENCES

- [1] Zebari, D. A., Zeebaree, D. Q., Abdulazeez, A. M., Haron, H., & Hamed, H. N. A. (2020). Improved Threshold Based and Trainable Fully Automated Segmentation for Breast Cancer Boundary and Pectoral Muscle in Mammogram Images. *IEEE Access*, 8, 203097-203116.
- [2] Hameed, I., Masoodi, S. R., Malik, P. A., Mir, S. A., Ghazanfar, K., & Ganai, B. A. (2018). Genetic variations in key inflammatory cytokines exacerbates the risk of diabetic nephropathy by influencing the gene expression. *Gene*, 661, 51-59.
- [3] Tang, J., Aleyani, S., & Liu, H. (2014). Feature selection for classification: A review. *Data classification: Algorithms and applications*, 37.
- [4] Petre, I. O., & Buiu, C. (2017, June). A colon cancer microarray analysis technique. In 2017 E-Health and Bioengineering Conference (EHB) (pp. 265-268). IEEE.
- [5] Wu, P., & Wang, D. (2018). Classification of a DNA microarray for diagnosing cancer using a complex network-based method. *IEEE/ACM transactions on computational biology and bioinformatics*, 16(3), 801-808.
- [6] Deng, S. P., Zhu, L., & Huang, D. S. (2015, December). Mining the bladder cancer-associated genes by an integrated strategy for the construction and analysis of differential co-expression networks. In *BMC genomics* (Vol. 16, No. 3, pp. 1-10). BioMed Central.
- [7] Deng, S. P., & Huang, D. S. (2014). SFAPS: An R package for structure/function analysis of protein sequences based on informational spectrum method. *Methods*, 69(3), 207-212.
- [8] Zeebaree, D. Q., Haron, H., & Abdulazeez, A. M. (2018, October). Gene selection and classification of microarray data using a convolutional neural network. In 2018 International Conference on Advanced Science and Engineering (ICOASE) (pp. 145-150). IEEE.
- [9] Ahmed, O., & Brifcani, A. (2019, April). Gene Expression Classification Based on Deep Learning. In 2019 4th Scientific International Conference Naja (SICN) (pp. 145-149). IEEE.
- [10] Zebari, D. A., Haron, H., Zeebaree, S. R., & Zeebaree, D. Q. (2019, April). Enhance the Mammogram Images for Both Segmentation and Feature Extraction Using Wavelet Transform. In 2019 International Conference on Advanced Science and Engineering (ICOASE) (pp. 100-105). IEEE.
- [11] Leung, Y. Y., Chang, C. Q., Hung, Y. S., & Fung, P. C. W. (2006, August). Gene selection for brain cancer classification. In 2006 International Conference of the IEEE Engineering in Medicine and Biology Society (pp. 5846-5849). IEEE.
- [12] Zeebaree, D. Q., Haron, H., Abdulazeez, A. M., & Zebari, D. A. (2019, April). Machine learning and Region Growing for Breast Cancer Segmentation. In 2019 International Conference on Advanced Science and Engineering (ICOASE) (pp. 88-93). IEEE.
- [13] Tell-Marti, G., Sarda, S. P., & Puig-Butille, J. A. (2019). Gene Expression Microarray: Technical Fundamentals and Data Analysis.
- [14] Zeebaree, D. Q., Haron, H., Abdulazeez, A. M., & Zebari, D. A. (2019, April). Trainable Model Based on New Uniform LBP Feature to Identify the Risk of the Breast Cancer. In 2019 International Conference on Advanced Science and Engineering (ICOASE) (pp. 106-111). IEEE.
- [15] Zhang, R., Gao, H., Liu, Y., Lu, Y., & Cui, Y. (2018, November). Biclustering of Gene Expression Data Based on Binary Artificial Fish Swarm Algorithm. In 2018 5th IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS) (pp. 247-251). IEEE.
- [16] Bolón-Canedo, V., Sánchez-Marono, N., Alonso-Betanzos, A., Benítez, J. M., & Herrera, F. (2014). A review of microarray datasets and applied feature selection methods. *Information Sciences*, 282, 111-135.
- [17] Tomašev, N., Radovanović, M., Mladenović, D., & Ivanović, M. (2013). The role of hubness in clustering high-dimensional data. *IEEE Transactions on Knowledge & Data Engineering*, (1), 1.
- [18] Zebari, N. A., Zebari, D. A., Zeebaree, D. Q., & Saeed, J. N. Significant features for steganography techniques using deoxyribonucleic acid: a review.
- [19] Ahmad, F. K., Norwawi, N. M., Deris, S., & Othman, N. H. (2008, August). A review of feature selection techniques via gene expression profiles. In 2008 International Symposium on Information Technology (Vol. 2, pp. 1-7). IEEE.
- [20] Mischel, P. S., Cloughesy, T. F., & Nelson, S. F. (2004). DNA-microarray analysis of brain cancer: molecular classification for therapy. *Nature Reviews Neuroscience*, 5(10), 782.
- [21] Zeebaree, D. Q., Haron, H., & Abdulazeez, A. M. (2018, October). Gene selection and classification of microarray data using convolutional neural network. In 2018 International Conference on Advanced Science and Engineering (ICOASE) (pp. 145-150). IEEE.
- [22] Rich, J. N., Hans, C., Jones, B., Iversen, E. S., McLendon, R. E., Rasheed, B. A., ... & West, M. (2005). Gene expression profiling and genetic markers in glioblastoma survival. *Cancer research*, 65(10), 4051-4058.
- [23] Jahwar, A. F., & Duhok, A. M. A. (2020). META-HEURISTIC ALGORITHMS FOR K-MEANS CLUSTERING: A REVIEW. *PalArch's Journal of Archaeology of Egypt/Egyptology*, 17(7), 12002-1220.
- [24] D. Q. Zeebaree, A. M. Abdulazeez, D. A. Zebari, H. Haron and H. Nuzly, "Multi-level fusion in ultrasound for cancer detection based on uniform lbp features," *Computers, Materials & Continua*, vol. 66, no.3, pp. 3363-3382, 2021.
- [25] Zebari, D. A., Haron, H., Zeebaree, D. Q., & Zain, A. M. (2019, August). A Simultaneous Approach for Compression and Encryption Techniques Using Deoxyribonucleic Acid. In 2019 13th International Conference on Software, Knowledge, Information Management and Applications (SKIMA) (pp. 1-6). IEEE.
- [26] Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of machine learning research*, 3(Mar), 1157-1182.
- [27] Othman, G., & Zeebaree, D. Q. (2020). The Applications of Discrete Wavelet Transform in Image Processing: A Review. *Journal of Soft Computing and Data Mining*, 1(2), 31-43.

[28] Dadaneh, B. Z., Markid, H. Y., & Zakerolhosseini, A. (2016). Unsupervised probabilistic feature selection using ant colony optimization. *Expert Systems with Applications*, 53, 27-42.

[29] Abdel-Basset, M., Fakhry, A. E., El-Henawy, I., Qiu, T., & Sangaiah, A. K. (2017). Feature and intensity based medical image registration using particle swarm optimization. *Journal of medical systems*, 41(12), 197.

[30] Su, H. (2019). Siting and sizing of distributed generators based on improved simulated annealing particle swarm optimization. *Environmental Science and Pollution Research*, 26(18), 17927-17938.

[31] Mandal, S. (2017). A modified particle swarm optimization algorithm based on self-adaptive acceleration constants. *International Journal of Modern Education and Computer Science*, 9(8), 49.

[32] Zebari, D. A., Zeebaree, D. Q., Saeed, J. N., Zebari, N. A., & Adel, A. Z. (2020). Image steganography based on swarm intelligence algorithms: A survey. *people*, 7(8), 9.

[33] Petre, I. O., & Buiu, C. (2017, June). A colon cancer microarray analysis technique. In 2017 E-Health and Bioengineering Conference (EHB) (pp. 265-268). IEEE.

[34] Zhang, X., He, T., Ouyang, L., Xu, X., & Chen, S. (2018, December). A Survey of Gene Selection and Classification Techniques Based on Cancer Microarray Data Analysis. In 2018 IEEE 4th International Conference on Computer and Communications (ICCC) (pp. 1809-1813). IEEE.

[35] Brazma, Alvis, and Jaak Vilo. "Gene expression data analysis." *FEBS Letters* 480.1 (2000): 17-24.

[36] Sherlock, Gavin. "Analysis of large-scale gene expression data." *Current opinion in immunology* 12.2 (2000): 201-205.

[37] Othman, M. S., Kumaran, S. R., & Yusuf, L. M. (2020). Gene Selection Using Hybrid Multi-Objective Cuckoo Search Algorithm with Evolutionary Operators for Cancer Microarray Data. *IEEE Access*, 8, 186348-186361.

[38] Lai, C. M., & Huang, H. P. (2020). A gene selection algorithm using simplified swarm optimization with multi-filter ensemble technique. *Applied Soft Computing*, 106994.

[39] Pino Angulo, A. (2018). Gene selection for microarray cancer data classification by a novel rule-based algorithm. *Information*, 9(1), 6.

[40] Zhai, Y., Yin, S., & Zhang, D. (2016). Association between antipsychotic drugs and mortality in older persons with Alzheimer's disease: A systematic review and meta-analysis. *Journal of Alzheimer's Disease*, 52(2), 631-639.

[41] Lin, S.W., Ying, K.C., Chen, S.C., Lee, Z.J.: Particle swarm optimization for parameter determination and feature selection of support vector machines. *Expert Syst. Appl.* 35(4), 1817–1824 (2008).

[42] Pérez-Grijalba, V., Pesini, P., Monleon, I., Boada, M., Tárraga, L., Ruiz-Laza, A., ... & Sarasa, M. (2013). Several direct and calculated biomarkers from the amyloid- β pool in blood are associated with an increased likelihood of suffering from mild cognitive impairment. *Journal of Alzheimer's Disease*, 36(1), 211-219.

[43] Chan, S. C., Wu, H. C., Lin, J. Q., & Zhang, Z. G. (2018, November). A Partial least squares-based regression approach for analysis of frontotemporal dementia gene markers in human brain gene microarray data. In 2018 IEEE 23rd International Conference on Digital Signal Processing (DSP) (pp. 1-5). IEEE.

[44] Hegde, R. B., Prasad, K., Hebbar, H., Singh, B. M. K., & Sandhya, I. (2019). Automated Decision Support System for Detection of Leukemia from Peripheral Blood Smear Images. *Journal of Digital Imaging*, 1-14.

[45] Harun, N. H., Bakar, J. A., Abd Wahab, Z., Osman, M. K., & Harun, H. (2020, April). Color Image Enhancement of Acute Leukemia Cells in Blood Microscopic Image for Leukemia Detection Sample. In 2020 IEEE 10th Symposium on Computer Applications & Industrial Electronics (SCAIE) (pp. 24-29). IEEE.

[46] Asl, A. A. S., & Zarandi, M. H. F. (2017, October). A type-2 fuzzy expert system for diagnosis of leukemia. In North American Fuzzy Information Processing Society Annual Conference (pp. 52-60). Springer, Cham.

[47] Hasri, N.N.M., et al., Improved Support Vector Machine Using Multiple SVM-RFE for Cancer Classification. *International Journal on Advanced Science, Engineering and Information Technology*, 2017. 7(4-2): p. 1589-1594.

[48] Banga, J. R. (2008). Optimization in computational systems biology. *BMC systems biology*, 2(1), 1-7.

[49] Krause, J., Ruxton, G. D., & Krause, S. (2010). Swarm intelligence in animals and humans. *Trends in ecology & evolution*, 25(1), 28-34.

[50] Murray, J. D. (2007). Mathematical biology: I. An introduction (Vol. 17). Springer Science & Business Media.

[51] Bonabeau, E., Dorigo, M., Marco, D. D. R. D. F., Theraulaz, G., & Théraulaz, G. (1999). Swarm intelligence: from natural to artificial systems (No. 1). Oxford university press.

[52] Yan-fei Zhu and Xiong-min Tang, "Overview of swarm intelligence," 2010 International Conference on Computer Application and System Modeling (ICCASM 2010), Taiyuan, 2010, pp. V9-400-V9-403, doi: 10.1109/ICCASM.2010.5623005.

[53] Prabha, K. A., & Visalakshi, N. K. (2014, March). Improved particle swarm optimization-based k-means clustering. In 2014 International Conference on Intelligent Computing Applications (pp. 59-63). IEEE.

[54] Coelho, R.A., d. R. N. Guimaraes, F., Esmin, A.A.A., 2014. Applying swarm ensemble clustering technique for fault prediction using software metrics. In: 2014 13th International Conference on Machine Learning and Applications. pp. 356-361.

[55] Hossain, M.S., Moniruzzaman, M., Muhammad, G., Ghoneim, A., Alamri, A., 2016. Big data-driven service composition using parallel clustered particle swarm optimization in mobile environment. *IEEE Trans. Serv. Comput.* 9 (5), 806-817.

[56] Esmin, A.A.A., Coelho, R.A., 2013. Consensus clustering based on particle swarm optimization algorithm. In: Proceedings of the 2013 IEEE International Conference on Systems, Man, and Cybernetics. IEEE Computer Society, pp. 2280-2285.

[57] Nanda, S.J., Panda, G., 2014. A survey on nature inspired metaheuristic algorithms for partitional clustering. *Swarm Evol. Comput.* 16, 1-18.

[58] Figueiredo, E., Macedo, M., Siqueira, H. V., Santana Jr, C. J., Gokhale, A., & Bastos-Filho, C. J. (2019). Swarm intelligence for clustering—A systematic review with new perspectives on data mining. *Engineering Applications of Artificial Intelligence*, 82, 313-329.

[59] Han, S. Y., Zhang, C. H., & Tang, G. Y. (2017). Approximation optimal vibration for networked nonlinear vehicle active suspension with actuator time delay. *Asian Journal of Control*, 19(3), 983-995.

[60] Hu, Z., Su, Q., & Xia, X. (2016). Multiobjective image color quantization algorithm based on self-adaptive hybrid differential evolution. *Computational intelligence and neuroscience*, 2016.

[61] Ng, W. S., Neoh, S. C., Htike, K. K., & Wang, S. L. (2017). Particle Swarm Feature Selection for Microarray Leukemia Classification. *Progress in Energy and Environment*, 2, 1-8.

[62] Premalatha, K., & Natarajan, A. M. (2010). Hybrid PSO and GA models for document clustering. *Int. J. Advance. Soft Comput. Appl.*, 2(3), 302-320.

[63] Rajasekhar, A., Lynn, N., Das, S., Suganthan, P., 2017. Computing with the collective intelligence of honey bees - a survey. *Swarm Evol. Comput.* 32, 25-48.

[64] Karaboga, D., & Akay, B. (2009). A survey: algorithms simulating bee swarm intelligence. *Artificial intelligence review*, 31(1-4), 61.

[65] Ghanbarzadeh, T. (2007). Multi-objective optimization using the bees algorithm. In Proceedings of International Virtual Conference on Intelligent Production Machines and Systems (IPROMS), Cardiff, UK.

[66] Karaboga, D., & Basturk, B. (2008). On the performance of artificial bee colony (ABC) algorithm. *Applied soft computing*, 8(1), 687-697.

[67] Liyan, D., Sainan, Z., Geng, T., Yongli, L., & Guanyan, C. (2013, November). Ant colony clustering algorithm based on swarm intelligence. In 2013 6th International Conference on Intelligent Networks and Intelligent Systems (ICINIS) (pp. 123-126). IEEE.

[68] Shreem, S. S., Abdullah, S., & Nazri, M. Z. A. (2016). Hybrid feature selection algorithm using symmetrical uncertainty and a harmony search algorithm. *International Journal of Systems Science*, 47(6), 1312-1329.

[69] Chen, K. H., Wang, K. J., Wang, K. M., & Angelia, M. A. (2014). Applying particle swarm optimization-based decision tree classifier for cancer classification on gene expression data. *Applied Soft Computing*, 24, 773-780.

[70] Nagpal, A., & Singh, V. (2018). A feature selection algorithm based on qualitative mutual information for cancer microarray data. *Procedia computer science*, 132, 244-252.

[71] Sudha, M. N., & Selvarajan, S. (2016). Feature selection based on enhanced cuckoo search for breast cancer classification in mammogram image. *Circuits and Systems*, 7(4), 327-338.

[72] Sakri, S. B., Rashid, N. B. A., & Zain, Z. M. (2018). Particle swarm optimization feature selection for breast cancer recurrence prediction. *IEEE Access*, 6, 29637-29647.

[73] Hameed, S. S., Petinrin, O. O., Osman, A., & Hashi, F. S. (2018). Filter-wrapper combination and embedded feature selection for gene expression data. *Int. J. Advance Soft Compu. Appl.*, 10(1).

[74] Ng, W. S., Neoh, S. C., Htike, K. K., & Wang, S. L. (2017). Particle Swarm Feature Selection for Microarray Leukemia Classification. *Progress in Energy and Environment*, 2, 1-8.

[75] Ozger, Z. B., Bolat, B., & Diri, B. (2019). A Probabilistic Multi-Objective Artificial Bee Colony Algorithm for Gene Selection. *J. UCS*, 25(4), 418-443.

[76] Jagadeesh, S., & Sugumar, R. (2017). A Comparative study on Artificial Bee Colony with modified ABC algorithm. *European Journal of Applied Sciences*, 9(5), 243-248.

[77] Kaveh, A., & Bakhshpoori, T. (2016). An efficient multi-objective cuckoo search algorithm for design optimization. *Advances in Computational Design*, 1(1), 87-103.

[78] Zhang, Y., Cheng, S., Shi, Y., Gong, D. W., & Zhao, X. (2019). Cost-sensitive feature selection using two-archive multi-objective artificial bee colony algorithm. *Expert Systems with Applications*, 137, 46-58.

[79] Sasikala, S., alias Balamurugan, S. A., & Geetha, S. (2015). A novel feature selection technique for improved survivability diagnosis of breast cancer. *Procedia Computer Science*, 50, 16-23.

[80] Yahya, A. A., Osman, A., & Abd Rahman Ramli, A. B. (2011). Feature selection for high dimensional data: an evolutionary filter approach.

[81] M. Fatimaezzahra, S. Mohamed, and E. Abdelaziz, "A combined cuckoo search algorithm and genetic algorithm for parameter optimization in computer vision," *Int. J. Appl. Eng. Res.*, vol. 51, pp. 12940-12954, Dec. 2017.

[82] Othman, M. S., Kumaran, S. R., & Yusuf, L. M. (2020). Gene Selection Using Hybrid Multi-Objective Cuckoo Search Algorithm with Evolutionary Operators for Cancer Microarray Data. *IEEE Access*, 8, 186348-186361.

[83] Qaraad, M., Amjad, S., El-Kafrawy, P., Fathi, H., & Manhrawy, I. I. (2020, June). Parameters Optimization of Elastic NET for High Dimensional Data using PSO Algorithm. In *2020 International Conference on Intelligent Systems and Computer Vision (ISCV)* (pp. 1-7). IEEE.

[84] Prabhakar, S. K., & Lee, S. W. (2020). Transformation based tri-level feature selection approach using wavelets and swarm computing for prostate cancer classification. *IEEE Access*, 8, 127462-127476.

[85] Saqib, P., Qamar, U., Khan, R. A., & Aslam, A. (2020, February). MF-GARF: Hybridizing Multiple Filters and GA Wrapper for Feature Selection of Microarray Cancer Datasets. In *2020 22nd International Conference on Advanced Communication Technology (ICACT)* (pp. 517-524). IEEE.

[86] Tubishat, M., Ja'afar, S., Alswaitti, M., Mirjalili, S., Idris, N., Ismail, M. A., & Omar, M. S. (2020). Dynamic salp swarm algorithm for feature selection. *Expert Systems with Applications*, 164, 113873.

[87] Cahyaningrum, K., & Astuti, W. (2020, August). Microarray Gene Expression Classification for Cancer Detection using Artificial Neural Networks and Genetic Algorithm Hybrid Intelligence. In *2020 International Conference on Data Science and Its Applications (ICoDSA)* (pp. 1-7). IEEE.

[88] Babu, T., Gupta, D., Singh, T., & Hameed, S. (2018, December). Colon Cancer Prediction on Different Magnified Colon Biopsy Images. In *2018 Tenth International Conference on Advanced Computing (ICoAC)* (pp. 277-280). IEEE.

[89] Liu, M., Xu, L., Yi, J., & Huang, J. (2018, February). A feature gene selection method based on ReliefF and PSO. In *2018 10th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA)* (pp. 298-301). IEEE.

[90] Wu, P., & Wang, D. (2018). Classification of a DNA microarray for diagnosing cancer using a complex network-based method. *IEEE/ACM transactions on computational biology and bioinformatics*, 16(3), 801-808.

[91] Deng, X., & Xu, Y. (2019, November). Cancer Classification Using Microarray Data By DPCAForest. In *2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI)* (pp. 1081-1087). IEEE.