Turkish Online Journal of Qualitative Inquiry (TOJQI)
Volume 12, Issue 8, July 2021: 8045 - 8054

Feature Selection and Classification to Identify Cancer in Microarray Gene Expression Profile

[1]Vishwas Victor, [2]Dr. Ragini Shukla

[1]Research Scholar, [2] Asso. Professor
[1] Department of CS & IT,
[1]Dr. C. V. Raman University, Bilaspur, India
[2] Department of CS & IT,
[2]Dr. C. V. Raman University, Bilaspur, India

Abstract

Cancer is one in all the dreadful diseases, which causes a substantial death rate in humans. Cancer is featured by associate irregular, unmanageable growth which will demolish and attack neighbouring healthy body tissues or somewhere else in the body. Microarray based mostly gene expression identification has been emerged as an economical technique for cancer classification, as well as for identification, prognosis, and treatment functions. In recent years, Deoxyribonucleic Acid microarray technique has gained a lot of attraction in both scientific and in industrial fields. It showed great importance in deciding the informative genes that can cause the cancer. This led to enhancements in early cancer diagnosis and in giving effective chemotherapy treatment. Studying cancer microarray gene expression data could be a difficult task because microarray is high dimensional-low sample dataset with loads of noisy or irrelevant genes and missing data. In this paper, we have a tendency to conduct a comprehensive study that focuses on exploring the main objectives and approaches that are applied using cancer microarray gene expression profile. We proceed by creating a classification for all approaches, and then conclude by investigating the foremost economical approaches that may be employed in this field.

Keywords: Cancer classification, Clustering approaches, Gene expression, Gene selection, Microarray.

I. Introduction:

The organic phenomenon profiles that are obtained from particular microarray experiments are widely used for cancer classification to make an efficient model. This model will differentiate traditional or different cancerous states by using chosen informative genes. However, studying microarray dataset in keeping with their gene expression profiles represents a difficult task. The complexness of the matter rises from the massive variety of options that contribute to a profile as compared to the terribly low number of samples unremarkably available in microarray analysis. Another challenge is the presence of noise (biological or technical) in the dataset that further affects the accuracy of the experimental results.

Microarrays, referred to as deoxyribonucleic acid chips or your time known as gene chips, are chips that are hybridized to a labelled unknown molecular extracted from a selected tissue of interest. This makes it attainable to measure at the same time the expression level in a cell or tissue sample for every gene represented on the chip. Deoxyribonucleic Acid microarrays are often used to determine which genes are being expressed during a given cell type at a particular time and beneath particular conditions. This allows us to check the gene expression in 2 completely different cell types or tissue samples, where we can find the more informative genes that are liable for inflicting a selected disease or cancer.

Recently, microarray technologies have unfolded several windows of chance to analyse cancer diseases using gene expressions. The first task of a microarray data analysis is to find out a computational model from the given microarray data which will predict the class of the given unknown samples. The accuracy, quality, and robustness are important components of microarray analysis. The accuracy of microarray dataset analysis depends on both the standard of the provided microarray data and the utilized analysis approach or objective. However, the curse of spatiality, the small number of samples, and the level of unsuitable and noise genes create the classification task of a test sample more challenging. Those inapplicable genes not solely introduce some uncalled-for noise to organic phenomenon knowledge analysis, but also increase the spatiality of the gene expression matrix. This ends up in the rise of the computational complexity in numerous subsequent research objectives such as classification and clustering.

Therefore, in our study, we have a tendency to focus on the main objectives and approaches that are applied on cancer microarray gene expression profile. We proceed by investigating the most economical approaches in this field. The rest of this paper organized as follow: Section 2 provides the reader some background material concerning microarray gene expression profile. Then, Section 3 illustrates and classifies the most approaches that are used recently for cancer microarray gene expression profile. Section 4, presents discussion and analysis concerning the foremost economical approaches that are presented throughout the paper. Finally, Section five concludes the paper.

II. Microarray Gene Expression Profile:

All living organisms include cells. As an example, Humans have trillions of cells and every cell contains an entire copy of the genome (the program for creating the organism) which is encoded in DNA. A gene may be a section of DNA that specifies the way to create a protein. For instance, Human DNA has about 30-35,000 genes. Gene expression is the method by which the information encoded in a gene is regenerate into an observable phenotype (most normally production of a protein). Therefore, gene expression is the degree to which a gene is active in a bound tissue of the body, measured by the amount of mRNA within the tissue. Individual genes may be switched on (exert their effects) or shifted according to the needs and circumstances of the cell at a selected time. It is worth mentioning that in cellular organisms, expression of the proper genes within the right order at the proper time is particularly crucial throughout embryonic development and cell differentiation. Thus, abnormalities of organic phenomenon could lead to the death of cells, or their uncontrolled growth, such as in cancer.

A Microarray consists of a solid surface onto that notable DNA molecules are with chemicals bonded at special locations in array. Moreover, every array location is usually known as a probe and contains several replicates of the same molecule. every probe represents the measurement for a single gene, and an array represents measurements for several genes (the molecules). this suggests that every array location is rigorously chosen thus on interbreed only with the messenger RNA molecules that corresponds to one sequence.

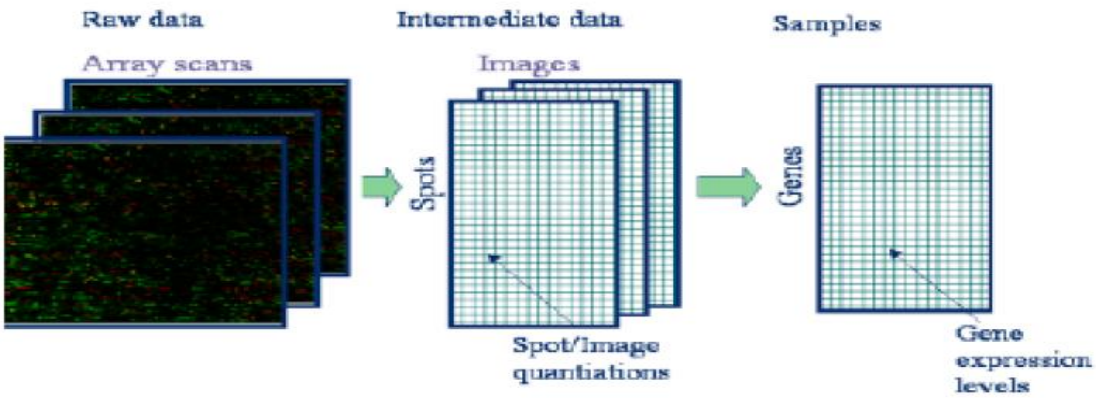

Figure 1: Generating Matrix from Microarray Experiments

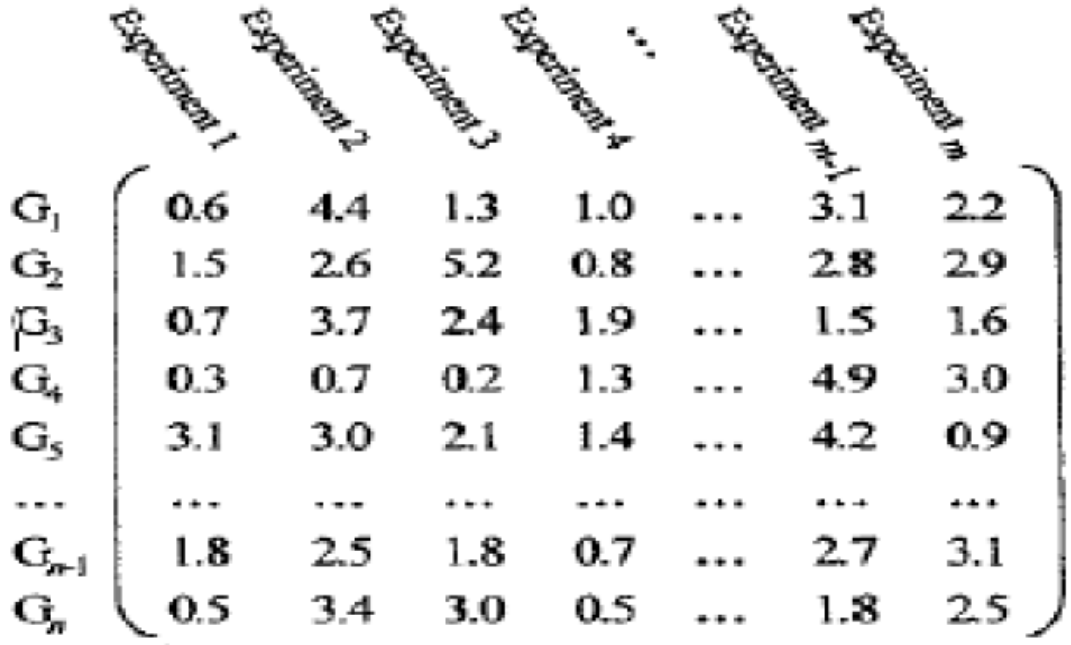

Figure 2: Example of microarray gene expression matrix

Fig. 1 shows how a gene expression matrix is generated. In the gene expression matrix, rows represent genes (as opposed to features/spots within the array) and columns represent measurements from totally different experimental conditions measured on individual arrays. just in case of cancer diagnoses, columns represent totally different sample tissue (cancerous tissue, or normal tissue) taken from totally different patients.

Generally, when multiple experiments are conducted, gene expression matrix are often viewed as a 2-dimensional array, indexed by an integer i identifying a notable gene Gi and an integer j identifying a specific experiment trial Ej. Then Aij is the relative quantity of hybridization (Gene expression level) for every gene Gi in experiment Ej. In Fig. 2, An example of the hybridization as we explained before gene expression matrix (A) for n genes assayed by m microarray experiments. Every entry represents the relative quantity of hybridization for every gene in each experiment. Typically, microarray matrixes contain thousands of rows (Genes) and dozens of columns (Experiments).

III. Objectives And Approaches:

Effective microarray experiments need careful planning that is based on clear objectives. The objectives of many studies using Desoxyribonucleic Acid microarrays can be divided into three main groups: gene selection, category discovery, and classification. Gene finding or gene selection is the method of selecting the littlest subset of informative genes that are most prophetical to its connected class. This helps in increasing the classifiers ability to classify samples more accurately. Class discovery issues with representing a new cancer or disease as a new class. Class prediction (classification) predicts the category of a brand-new specimen supported its expression profile. In this section, we are going to define these 3 objectives and illustrates the foremost efficient approaches that are employed in order to attain them in more details.

A. Gene Finding (Gene Selection):

The sequence finding studies are vital in microarray study since it's aimed to reduce the spatial property of microarray dataset by selecting the foremost informative genes. Moreover, gene-finding [1] ways usually perform class comparison to determine the genes whose expression is correlate to a quantitative measurement or a survival time. Sequence selection may be a method of choosing the smallest subset

of informative genes that area unit most predictive to its connected class for classification which maximizes the classifiers ability to classify samples accurately [10]. The best feature selection issue has been shown to be NP hard.

Notably, there are many benefits for gene selection method [6]. For diagnoses, it's less expensive to specialise in the expression of solely a few genes instead of on thousands of genes. This results in a reduction within the cost of clinical diagnosis. Also, the feature selection reduces the dimensionality issue, and this results in a reduction in computational cost. On the other hand, feature set selection typically gives rise to a far smaller and a lot of compact gene set.

Some gene selection ways don't assume any specific distribution model on the gene expression data and that they are referred to as model-free gene selection ways or sometimes called Filter technique [11]. Whereas alternative gene selection methods assuming bound models are remarked as model-based gene selection ways or could be known as Wrapper technique. In other hand, some scientist applied Filter technique and Wrapper method, this technique is known as Hybrid technique [7]. Moreover, hybrid gene selection strategies look for an optimal subset of options is constructed into the classifier construction, and can be seen as research within the combined area of feature subsets and hypotheses.

B. Class Discovery (Clustering):

Class discovery is completely different type of gene finding or class prediction since it doesn't involve any predefined classes. Class discovery involves analysing a given set of gene expression profiles with the goal of discovering subgroups that share common options. It involves grouping together specimens that are based on the similarity of their expression profiles with relevance the genes are represented on the array. Cluster analysis or clump is usually used for class discovery.

The objective of cluster expression profiles of tumours is to see new disease (cancer) classifications. Clustering [4] aims at dividing the info points (genes or samples) into groups (clusters) using measures of similarity, such as correlation or Euclidean distance. Discovery of a brand-new class is usually achieved by associate degree unsupervised machine learning method with the assistance of a clustering technique like hierarchical clustering, k-means clustering and self-organizing maps (SOM). it's referred to as unsupervised as a result of the grouping isn't driven by any composition external to the expression profiles, like tissue type, stage, grade or response to treatment.

C. Class prediction (Classification):

Class prediction or classification (including the assignment of labels to samples based on their expression patterns) is typically statistical or supervised machine learning methods. It often needs finding which genes are informative for distinctive the pre-defined classes, estimating the parameters of the mathematical relation that is used, and estimating the accuracy of the predictor. Class prediction may be a very helpful and useful data processing technique for medical issues of diagnostic classification, prognostic prediction and treatment choice. Also, most cancer studies in microarray expression identification have category comparison or class prediction objectives.

In supervised method we are supposed to coach the classifier before we tend to begin in classifying method, whereas unsupervised methodology we begin in classifying method with none coaching. Moreover, supervised methods are a unit typically more practical in cancer classification researches, and that they are used for cancer prediction as follows: A classifier is trained with a part of the samples in the cancer microarray dataset. Then, the trained classifier is used to predict the samples within the remainder of the dataset to evaluate the effectiveness of the classifier.

Microarray based mostly gene expression profiling [3] has become an important and promising dataset for cancer classification that are used for diagnosis and prognosis functions. The most important motivation for victimization microarray datasets is to classify unknown tissue samples according to their expression profiles. for instance, it is employed in classifying cancerous or traditional samples, or to discriminate different kinds or subtypes of cancer. Classification tasks are wide used in real-world applications, few of them involves only binary classifies and majority of them involve over 2 categories,

the so-called multi-class classification drawback. Moreover, since completely different subtypes of a cancer respond otherwise to the same therapy, it's necessary to diagnose the cancer kind of a patient properly, then customise the treatment for that patient. it's worth mentioning, that DNA microarrays [16] have been recently receiving huge attention in bi- and multi-cancer classification. Within the past decade, a number of feature selection and classification strategies are proposed for bi-class and multiclass cancer classification. In order to demonstrate the variations between the binary class classification approaches and multi category cancer classification approaches, within the following subsections we tend to summarize these approaches.

1) Binary Cancer Classification:

Classification tasks area unit widely employed in real-world applications, a number of them involves solely binary classifies and lots of them involve more than 2 classes, referred to as multi class classification drawback. Their application domain is diverse; for example, within the field of bioinformatics, and, within the cancer classification of microarrays. Within the literature, binary cancer classification issues have been such as extensively studied like for cancer, and carcinoma. It's worth mentioning, that there are several benchmarks for two-class cancer microarray that are accessible online. In Table 1, we summarize the foremost useful two-class cancer microarray datasets.

Table 1: BINARY CANCER MICROARRAY DATASETS

Cancer Microarray	No. Of Classes	No. Of Samples	No. Of Genes
Leukaemia	2	75	7129
Lung	2	175	12533
Colon	2	70	2000
Prostate	2	130	12600
Ovarian	2	260	15154
Breast	2	35	7129
Lymphoma	2	100	4026

Most of the projected binary class classification strategies in literature achieved correct result. There are many techniques that are applied for classifying two-class cancer microarray dataset together with statistical techniques, data processing methods, SVM (Support Vector Machine), k-NN (k- Nearest Neighbour), ANN (Artificial Neural Network), GA (Genetic Algorithms) [15], follow swarm optimisation (PSO) [22], Naïve Bayes (NB), call Trees (DTs). The Support Vector Machine (SVM) algorithm has verified to be one in every of the foremost powerful supervised learning algorithms in biological knowledge analysis together with microarray-based expression analysis. Also, SVM methodology used as binary categorical classifiers and it's been shown to systematically outperform other classification approaches together with weighted choice and k-nearest neighbours.

2) Multi Class Cancer Classification:

Recently, microarray technology has been considered as a important approach to classify multi classes (types) for cancer for early diagnosis and chemotherapy treatment functions [4]. As we noted, a number of systematic techniques are developed and studied to classify cancer types using gene expression data. However, most of those studies were confined towards binary factor selection issues and solely a really few considered multi class gene selection and classification. This is often since multi class factor selection and classification is considerably tougher than the binary problems. In Table 2, more beneficial benchmark multiclass cancer-related human organic phenomenon datasets that are gleaned from the literature are delineated. We have got chosen from all multi-class cancer microarray dataset that are used, five datasets, carcinoma, brain tumour, CNS, NCI60, and GCM. These datasets have less classification accuracy result, in comparison with different dataset like leukaemia and SRBCT dataset.

So, the analysis in these multi category cancer microarray datasets are challenge and open. In 1990, the National Cancer Institute sixty (NCI60) platforms included sixty human growth cell lines that represented nine cancer types. Table 3 presents description concerning NCI60 datasets. GCM is a more difficult microarray dataset that includes fourteen types of cancers. This information set contains expression data of 16,306 genes with the overall of 198 samples has been already divided into 2 elements, i.e., one hundred forty-four for training and therefore the different 54 for testing. Table 4, provides the general info of the GCM dataset.

Notably, multi-class cancer classifiers that are based on support vector machines are the foremost helpful and effective classifiers in activity correct cancer identification from microarray gene expression data. the primary generation of SVMs could solely be applied to binary classification tasks. However, most world diagnostic tasks, particularly cancer diagnostic are not binary. Therefore, many algorithms have emerged during the previous few years that permit multi-class classification with SVMs, like DAGSVM, a technique by Weston and Watkins (WW), and methodology by Crammer and Singer (CS). Furthermore, there are some novel methods in literature that aim to enhance the performance of SVM by combining with evolutionary algorithms, such as ESVM, and GASVM, or with Fuzzy algorithms just like FSVM.

However, there are several different multi class classifier that are planned within the literature like, applied mathematics approaches, Evolutionary algorithm, K-nearest neighbours (KNN), naïve Bayes (NB), neural networks (NN), and decision tree (DT). Furthermore, Artificial neural network (ANN) strategies give a lovely various to the higher than approach for an immediate multi-class classification drawback. Neural networks can map the computer file into totally different categories directly with one network. However, conventional neural networks sometimes manufacture lower classification accuracy than SVM. There are several multi-class cancer classification algorithms that primarily based on neural network, like FNN, ELM, WNN, PNN, and SANN. Also, economical multi-class cancer classification strategies that are supported applied mathematics techniques, such as, most chance classification (MLHD) methodology.

IV. Analysis And Discussion

As mentioned before, the objectives of different studies using DNA microarrays is classified into 3 major groups: gene finding, class discovery, and class prediction. In Table 5, we illustrate every objective with approaches and aim that has been used for analysing cancer microarray organic phenomenon profile.

Based on our study, we tend to conclude that cancer classification is a vital field of analysis for cancer microarray gene expression profile [5]. Also, most cancer studies in microarray expression identification really have class comparison or class prediction objectives. Moreover, microarray is considered an economical technique for cancer classification, as well as for diagnosis, prognosis, and treatment functions. In recent years, Deoxyribonucleic Acid microarray technique has gained more attraction in each scientific and in industrial fields, and it is important to note that the informative genes that cause the cancer to enhance early cancer diagnosing and to provide effective chemotherapy treatment. Classifying cancer microarray gene expression data could be a difficult task just because microarray is high dimensional-low sample dataset with lots of noisy or irrelevant genes and missing data. The inherent presence of a large number of irrelevant genes will increase the problem of the classification task influencing the discrimination power of relevant options [9]. Those irrelevant genes don't solely introduce some redundant noise to gene expression data analysis, however conjointly increase the spatiality of the gene expression matrix. This ends up in the rise of the computational complexity in numerous researches like classification and cluster. Therefore, finding an accurate gene selection methodology that could be efficient enough to reduce the spatiality and selecting informative genes area unit terribly difficult problems in cancer classification.

Table 2: MULTI CLASS CANCER MICROARRAY DATASETS

Cancer Microarray	No. Of Classes	No. Of Samples	No. Of Genes	Description
Lung	5	200	12600	Four lung cancer types and normal tissues
Brain	4	55	10367	Four malignant glioma types
CNS	5	85	7129	Central Nervous System Embryonal Tumour CNS consists of 5 subclasses: Medulloblastoma (MED), Malignant Glioma (MG), Atypical Teratoid / rhabdoid tumours (AT / RT), Normal Cerebellum (NC) and Primitive Neuroectodermal (PNET)
NCI60	9	60	57725	Nine various human tumours types
GCM	14	200	16306	Fourteen various human tumours types
SRBCT	4	65	2304	Small round blue cell tumours (SRBCT) of childhood are hard to classify by current clinical techniques.
Leukaemia	3	75	12582	AML, ALL and Mixed Lineage Leukaemia (MLL)

Table 3: NCI60 DATASET DESCRIPTION

Type Number	Cancer	Number of Cell Lines
1	Leukaemia	6 lines
2	Melanoma	8 lines
3	Lung	9 lines
4	Colon	7 lines
5	Brain	6 lines
6	Ovarian	7 lines
7	Breast	6 lines
8	Prostate	2 lines
9	Kidney	8 lines

Its value mentioning, that experimental studies indicate that direct multi class classifications are way more tough than binary classifications and that the classification accuracy might drop dramatically when the number of classes will increase. Therefore, rather than directly managing multi-class issues, several classification strategies for multi class issues use some combination of binary classifiers on a One-Versus All (OVA) or a One-Versus-One (OVO) comparison Basis. However, this manner of implementation leads to combining several binary classifiers and thus will increase system complexities. It conjointly causes a bigger computational burden and longer coaching time. for instance, the support vector machine (SVM) as a binary classifier tries to map the data from a lower-dimensional input space to a higher dimensional feature area in order that to create the information linearly separable into 2 classes.

Table 4: GCM DATASET DESCRIPTION

Type Number	Cancer	Number Of Sample
1	Breast	11
2	Prostate	10
3	Lung	11
4	Colorectal	11
5	Lymphoma	22
6	Bladder	11
7	Melanoma	10
8	Uterus	10

9	Leukaemia	30
10	Rental	11
11	Pancreas	11
12	Ovarian	11
13	Mesothelioma	11
14	Brain	20

Table 5: MAIN OBJECTIVES OF STUDYING MICROARRAY GENE EXPRESSION PROFILE

Objective	Approach	Aim
Gene Finding	Feature Selection	To reduce the dimensionality of microarray dataset by selecting the most informative genes
Class Discovery	Clustering	To determine new disease or cancer
Class Prediction	Classification	To classify samples (cancerous or normal) or to discriminate different types or sub-types of cancer

In literature there are many approaches for gene selection, however we observed that a multi-class cancer classification [20] that is combined with a gene choice technique, has not been investigated intensively. Thus, we conclude that we need to use gene choice method as a compulsory step before we begin cancer classification on microarray dataset. Also, we have a tendency to notice that ensemble classifiers also are applied for multi-class cancer classification however it doesn't generally improve the classification performance like SVM based classifier, or non-SVM primarily based classification strategies.

V. CONCLUSION

Microarray based mostly gene expression identification has become an important and promising approach that may be used for cancer classification. This can be a very important step for diagnosing and prognosis purposes. The foremost vital objective of microarray dataset is to classify unknown tissue samples according to their expression profiles. Microarray data suffers from the curse of spatial property, the little variety of samples, and the level of irrelevant and noise genes. These make the classification task of a test sample a really difficult problem. As a consequence, it's vital to eliminate those irrelevant genes and identify the informative genes that are why a feature selection drawback is crucial in gene expression data analysis. Therefore, the primary step in processing the expression data is to spot a small subset of genes that are primarily to blame for the cancer. it's needed to use a gene selection method as a compulsory step before we begin any cancer classification approach on a microarray dataset. Thus, we will conclude that the main objectives of many studies using DNA microarrays are often classified into three major groups: gene finding, class discovery, and class prediction.

REFERENCES

[1] Alanni, R., Hou, J., Azzawi, H. and Xiang, Y. (2019). Deep gene selection method to select genes from microarray datasets for cancer classification, BMC Bioinform. 2019, 20, 608.

[2] Almugren, N. and Alshamlan, H. (2019). A survey on hybrid feature selection methods in microarray gene expression data for cancer classification, IEEE Access, vol. 7, pp. 78533-78548, 2019.

[3] Alshamlan, H. M., Ghada H., Badr, and Alohali, Y. (2013). A Study of Cancer Microarray Gene Expression Profile: Objectives and Approaches, Proceedings of the World Congress on Engineering 2013 Vol II, WCE 2013, July 3 - 5, 2013, London, U.K.

[4] Ayadi, W., and Elloumi, M. (2011). Biclustering of microarray data, in Algorithms in Computational Molecular Biology: Techniques, Approaches and Applications; John Wiley & Sons: Hoboken, NJ, USA, 2011; pp. 651–663.

[5] Berrar, D., Downes, S. and Dubitzky, W. (2003). Multiclass cancer classification using gene expression profiling and probabilistic neural networks, in Pacific Symposium on Biocomputing, vol. 8, 2003, pp. 5–16.

[6] Bolón-Canedo, V., Sánchez-Marono, N., Alonso-Betanzos, A., Benítez, J.M. and Herrera, F. (2014). A review of microarray datasets and applied feature selection methods, Inf. Sci. 2014, 282, 111–135.

[7] Bonilla-Huerta, E., Hernandez-Montiel, A., Morales-Caporal, R. and Arjona-Liopez, M. (2016). Hybrid framework using multiple-filters and an embedded approach for an efficient selection and classification of microarray data, IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB) 13, vol. 1, pp. 12-26, 2016.

[8] Deb, S., Fong S. and Tian, Z. (2015). Elephant Search Algorithm for Optimization Problems, The Tenth International Conference on Digital Information Management (ICDIM 2015), 249-255, IEEE.

[9] El Akadi, A., Amine, A., El Ouardighi, A. and Aboutajdine, D. (2009). A new gene selection approach based on minimum redundancy-maximum relevance (mrmr) and genetic algorithm (ga), in Computer Systems and Applications, 2009. AICCSA 2009, IEEE/ACS International Conference on, 2009, pp. 69–75.

[10] Ghosh, M., Begum, S., Sarkar, R., Chakraborty, D. and Maulik, U. (2019). Recursive memetic algorithm for gene selection in microarray data, Expert Systems with Applications, vol. 116, pp. 172-185, 2019.

[11] Ke, W., Wu, C., Wu, Y. and Xiong, N. N. (2018). A new filter feature selection based on criteria fusion for gene microarray data, IEEE Access, vol. 6, pp. 61065-61076, 2018.

[12] Lv, J., Peng, Q., Chen, X. and Sun, Z. (2016). A multi-objective heuristic algorithm for gene expression microarray data classification, Expert Systems with Applications, vol. 59, pp. 13-19, 2016.

[13] Mohamed, N. S., Zainudin, S. and Othman, Z. A. (2017). Metaheuristic approach for an enhanced mrmr filter method for classification using drug response microarray data, Expert Systems with Applications, vol. 90, pp. 224-231, 2017.

[14] Pochet, N., De Smet, F., Suykens, J. A. and De Moor, B. L. (2004). Systematic benchmarking of microarray data classification: assessing the role of non-linearity and dimensionality reduction, Bioinformatics 20, vol. 17, pp. 3185-3195, 2004.

[15] Ram, P. K. and Kuila, P. (2019). Feature selection from microarray data: Genetic algorithm-based approach, Journal of Information and Optimization Sciences 40, vol. 8, pp. 1599-1610, 2019.

[16] Simon, R. (2009). Analysis of dna microarray expression data, Best practice and research Clinical haematology, vol. 22, no. 2, pp. 271–282, 2009.

[17] Tjaden, B. and Cohen, J. (2006). A survey of computational methods used in microarray data interpretation, in Applied Mycology and Biotechnology, Bioinformatics, 2006; 6: 7–18.

[18] Vural, H. and Subasi, A. (2015). Data Mining techniques to classify Microarray gene expression data using selection by SVD and information gain, Modeling of Artificial inteligence, vol. 6, no-2 pp. 171-182, 2015.

[19] Wang, L., Chu, F. and Xie, W. (2007). Accurate cancer classification using expressions of very few genes, IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 4, no. 1, pp. 40–53, 2007.

[20] Wang, Y., Tetko, I. V., Hall, M. A., Frank, E., Facius, A., Mayer, K. F. and Mewes, H. W. (2005). Gene selection from microarray data for cancer classification - a machine learning approach, Computational biology and chemistry 29, vol. 1, pp. 37-46, 2005.

[21] Zeebaree, D. Q., Haron, H. and Abdulazeez, A. M. (2018). Gene selection and classification of microarray data using convolutional neural network, in: 2018 International Conference on Advanced Science and Engineering (ICOASE), IEEE, pp. 145-150, 2018.

[22] Zhang, Y., Gong, D.W. and Cheng, J. (2015). Multi-objective particle swarm optimization approach for cost-based feature selection in classification. IEEE/ACM Trans. Comput. Biol. Bioinform. 2015, 14, 64–75.