Algorithms Implemented for Cancer Gene Searching and Classifications

Murad M. Al-Rajab and Joan Lu

School of Computing and Engineering, University of Huddersfield
Huddersfield, UK

U1174101@hud.ac.uk , j.lu@hud.ac.uk

Abstract. Understanding the gene expression is an important factor to cancer diagnosis. One target of this understanding is implementing cancer gene search and classification methods. However, cancer gene search and classification is a challenge in that there is no an obvious exact algorithm that can be implemented individually for various cancer cells. In this paper a research is conducted through the most common top ranked algorithms implemented for cancer gene search and classification, and how they are implemented to reach a better performance. The paper will distinguish algorithms implemented for Bio image analysis for cancer cells and algorithms implemented based on DNA array data. The main purpose of this paper is to explore a road map towards presenting the most current algorithms implemented for cancer gene search and classification.
Keywords: cancer; genes; searching algorithms, classification algorithms

1 Introduction

Cancer is one of the world's most serious diseases in modern society and a major cause of death worldwide. Traditional diagnostics methods are based mainly on the morphological and clinical appearance of cancer, but have limited contributions as cancer usually results from other environmental factors. There are several causes of cancer (carcinogens) such as smoke, radiation, synthetic chemicals, polluted water, and others that may accelerate the mutations and many undiscovered causes. On the other hand, a need to select the most informative genes from wide data sets, removal of uninformative genes and decreases noise, confusion and complexity and increase the chances for identification of diseases and prediction of various outcomes like cancer types is mandatory [1]. One of the challenging tasks in cancer diagnosis is how to identify salient expression genes from thousands of genes in microarray data that can directly contribute to the phenotype or symptom of disease [3]. The development of array technologies indicates the possibility of early detection and accurate prediction of cancer. Through these technologies, it is possible to get thousands of gene expression levels simultaneously through arrays, and also the ability to make use to know and find out whether it is cancer or not, and classify cancer [5]. Thus, there is a need to identify the informative genes that contribute to a cancerous state. An informative gene is a gene that is useful and relevant for cancer classification [6]. Cancer classification, which can help to improve health care of patients and the quality of life of individuals, is essential for cancer diagnosis and drug discovery [3]. Cancer classifica-

tion or prediction refers to the process of constructing a model on the microarray dataset and then distinguishing one type of samples from other types within the induced model [7]. Microarray is a device or a technology used to measure expression levels of thousands of genes simultaneously in a cell mixture, and finally produces a microarray data, which is also known as gene expression data. The task of cancer classification using microarray data is to classify tissue samples into related classes of phenotypes such as cancer versus normal [8]. A major problem in these microarray data is the high redundancy and the noisy nature of many genes or irrelevant information for accurate classification of cancer. Only a small number of genes may be important [9]. Early and accurate detection and classification of cancer is critical to the wellbeing of patients. The need for a method or algorithms for cancer identification is important and has a great value in providing better treatment and this can be done through analysis of genetic data. For practical use an algorithm has to be fast and accurate as well as easy to implement, test, and maintain. The optimal algorithm for a given task would have adequate performance with minimal implementation complexity [10]. To study the algorithms implemented for cancer gene search and classification, a long path of solid literature review must be constructed from Bioinformatics understanding passing through Bio-image processing and algorithms analysis toward cancer gene searching and selection algorithms implemented in the field and how these algorithms can be applied to classify cancer cells and how efficient they are. Due to the emergence of new technologies such as the micro array data, these new technologies produce large datasets characterized by a large number of features (genes); this is why feature selection (gene selection) has become very important in several fields such as Bioinformatics. Authors in [6, 11] introduced a new hybrid feature selection method that combines the advantages of filter strategy based on the Laplacian Score joint with a simple wrapper strategy. The suggested algorithm resulted in a fast hybrid feature selectors that can solve feature selection problems in high dimensional datasets and select a small subset full of informative genes that is most relative to cancer classification. Another research developed an automated system for robust and reliable cancer diagnoses based on gene microarray data as stated by the authors in [9]. They investigated that support vector machine classifier algorithms outperforms other algorithms such as K nearest neighbors, naive Bayes, neural networks and decision tree; and thus they could adopt the important genes for cancer tumor classifications. On the other hand the authors in [12], found the smallest set of genes that can ensure highly accurate cancer classifications from microarray data by using supervised machine learning algorithms. Moreover, the authors in [13], survived different feature selection techniques and their application for gene array data, they found two optimal search methods for cancer classification which are Genetic Algorithms (GA) and Tabu search (TS) to generate candidate genes for classifications. They argued that GA is an optimal search method that behaves like evolution processes in nature, while TS is a heuristic method that guides the search for optimal solution making use of flexible memory.

The main purpose of this paper is to explore a road map towards presenting the most current algorithms implemented for cancer gene search and classification.

The remainder of this paper will be structured as follows; Section 2 will discuss the common algorithms implemented in the research topic, on the other hand, section 3will give an overview of the algorithms, while, results and discussion will be presented in section 4. Finally, section 5 will conclude the paper.

2 Common Algorithms for Cancer Gene Search and Classification

The study of the algorithms is classified into two categories; first the algorithms that focus on gene expression analysis for cancer gene selection, and second, the algorithms that focus on Bio-Image analysis and performs cancer classification. These categories are discussed below:

2.1 Analysis of Cancer gene selection and classification Algorithms

Microarray data is being an influence to cancer diagnostics. Its accurate prediction to the type or size of tumors based on reliable and efficient classification algorithms, so that patient can be provided with better treatment or therapy response. The main issue behind microarray data is its high dimensionality which may lead to low efficiency in cancer gene classification and also makes it difficult to classify the related genes. Among thousands of genes whose expression levels are measured, not all are needed for classification [5]. Thus, one challenging task in cancer diagnosis is how to identify silent expression genes from thousands of genes in microarray data and how to select informative genes for classification that can assist to the symptom of disease [7]. Below is a summary of the most well implemented classification algorithms applied in the field and argued to be efficient for diverse cancer type's diagnosis and treatment.

Integrated Gene-Search Algorithm

The integrated algorithm is based on Genetic Algorithm (GA) and Correlation-based heuristics [1]. (Correlation-based feature selection) (CFS) for data preprocessing and data mining (decision tree and support vector machine algorithms) for making predictions. Thereafter, bagging and stacking algorithms were applied for further enhancement classification accuracy and the analysis of data was performed by WEKA data mining software. This work was proposed and successfully applied to the training and testing genetic expression data sets of ovarian, prostate, and lung cancers but also can be successfully applied to any other cancer like colon, breasted, bladder, leukemia, and so on. The Algorithm consists of two phases as shown in Figure 1, the iterative phase I, where data partitioning, execution of Decision Tree (DT) algorithm or any other data mining algorithms applied to the data set, then GA and CFS for gene reduction take place. After that, in phase II, data-mining algorithms are applied to the training and testing data sets generated from phase I and their results will be evaluated to determine the most significant gene set.

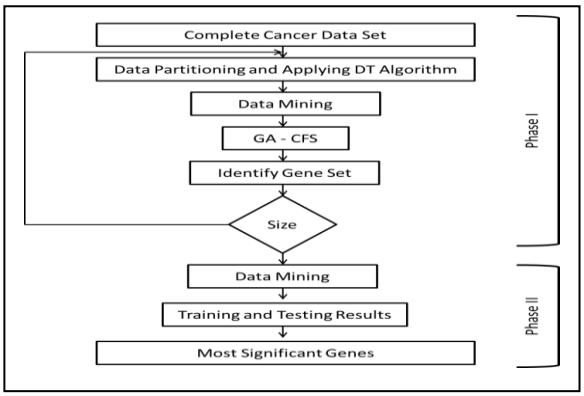

Fig. 1. Integrated Gene Search Algorithm

An integrated algorithm for gene selection and classification applied to microarray data for ovarian cancer

By applying a hybrid of algorithms (Genetic Algorithm "GA", Particle Swarm Optimization "PSO", Support Vector Machine "SVM", and Analysis of Variance "ANOVA") to select gene markers from target genes, finally fuzzy model is applied to classify cancer tissues [2]. Due to the huge amount of data types generated from gene expression and lack of systematic procedure to analyze the information instantaneously, in addition to avoid higher computational complexity, the need to select the most likely differential gene markers to explain the effects on ovarian cancer. It is concluded that the proposed algorithm has superior performance over ovarian cancer and can be applied and performed on other cancer diagnosis studies, and that is noticed from table 1.

Table 1. The Proposed Algorithm Accuracy of classification for various approaches

	The hybrid process of SVM and GA (%)	**The hybrid process of SVM and PSO (%)**	**The proposed algorithm (%)**
Colon	95.65%	97.13%	99.13
Breast	96.23%	97.95%	98.55

Source: Zne-Jung Lee, An integrated algorithm for gene selection and classification applied to microarray data of ovarian cancer, International Journal Artificial Intelligence in Medicine 42 (2008) 91.

A Bootstrapped Genetic Algorithm and Support Vector Machine to select genes for cancer classification.

The algorithm states that gene expression data obtained from microarrays have shown to be useful in cancer classification. A novel system is suggested for selecting a set of genes for cancer classification. The system is based on linear support vector machine and a genetic algorithm The proposed system considers two databases for the solution, one for the colon cancer and the other for the leukemia. It is argued that this proposed system of hybridization of genetic algorithm, support vector machine and bootstrapped methods is very efficient for classification problems. [4].

A Novel Embedded Approach composed of two main phases to the problem of cancer classification using gene expression data.

Phase one includes the use of gene selection to select the important predictive genes which make it later easier to be correctly classified. The second phase is to build powerful classifier models. For gene selection, a proposed of three filter approaches are analyzed, Information Gain (IG), Relief Algorithm (RA), and t-statistics (TA) to obtain a predictive reduced feature (gene) space containing the most informative genes. Later five well known classifier algorithms are utilized (Support Vector Machine (SVM), K Nearest Neighbor (KNN), Naïve Bayes (NB), Neural Network (NN), and Decision Tree (DT)) to classify nine famous available gene expression datasets. After the experiments, it was resulted that in 8 out of 9 datasets, SVMs classifier outperforms KNN, NB, NN and DT obviously in all cases [9].

Genetic Algorithm (GA) with an initial solution provided by t-statistics (t-GA) for selecting a group of informative genes from cancer microarray data.

The Decision Tree classifier (DT) is then built on the top of these selected genes. The performance of the proposed approach among other selection methods and indicated that t-GA has the highest accurate rate among different methods [14].

2.2 Cancer Classification through Bio image Analysis Algorithms

CAIMAN system (CAncer IMage ANalysis) [15] is an online algorithm repository that analyze the image produced by experiments relevant to cancer research (www.caiman.org.uk), three algorithms have been implemented to this project, an algorithm for measuring cellular migration, other one for vasculature analysis and an algorithm for image shading correction. The following table was a result of the estimation performance of the CAIMAN system (CAncer IMage ANalysis) , the three proposed algorithms were tested with two groups of five images each one of approximately 10kb in size and the other more than 1 Mb.The times are recorded from the moment the user opens the web page to the time the email with the results are received, as in Table 2 below:

Table 2. Proposed Algorithm Performance Estimation

Algorithm	Dimension (pixels)	Size (kb)	Time ± (s)
Migration	285 x 203	1001	62.6 ± 9.6
	127 x 900	1700	81.4 ± 16.7
Tracing	220 x 164	108	66.2 ± 20.3
	768 x 576	1300	207.4 ± 14.6
Shading	285 x 203	100	59.5 ± 14.1
	1270 x 900	1700	65.0 ± 15.6

Source: Constantino Carlos Reyes-Aldasoro, Michael K. Griffiths, Deniz Savas, Gillian M. Tozer, CAIMAN: An online algorithm repository for Cancer Image Analysis, Computer Methods and Programs in Biomedicine, Volume 103, Issue 2, August 2011, Page 103, ISSN 0169-2607, 10.1016/j.cmpb.2010.07.007.

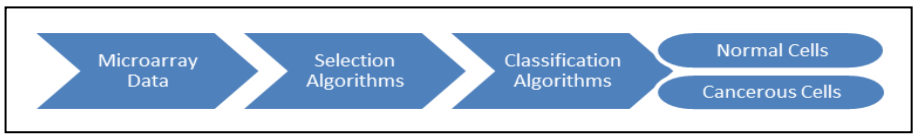

Fig. 2. Integrated Cancer Selection and Classification criteria

3 Algorithms Overview

It is noticed that to classify cancer cells into normal cells or cancerous cells, Selection and Searching Algorithms must be implemented first as shown in figure 2.

3.1 Searching and Selection Algorithms

Genetic Algorithm (GA) is a search algorithm. A GA is initiated with a set of solutions (chromosomes) called the population [1, 16]. Solutions from one population are taken and used to form a new population. This is motivated by a hope that the new population will be better than the old one. Solutions which are selected to form new solution are selected according to their fitness – the more suitable they are, the more chances they have to reproduce [14, 16], the chart of GA is presented in Figure 3.*Correlation-based feature selection (CFS) it* is a process of choosing or selecting a subset of original features so that the feature space is optimally reduced according to a certain evaluation criterion [17]. It reduces the number of features, removes irrelevant, redundant, or noisy data, and brings the immediate effects for applications [18]. *Particle Swarm Optimization (PSO)* is a population based search algorithm based on the simulation of the social behavior [19]. PSO is similar to GA in that the system is initialized with a population of random solutions. It is unlike GA, however, in that each potential solution is also assigned a randomized velocity, and the potential solutions "particles", are then "flown" through the problem space [20]. *Analysis of variance (ANOVA)* is an extremely important method in exploratory and confirmatory data analysis [21]. *Information Gain (IG)* is a method that attempts to quantify the best possible class predictability that can be obtained by dividing the full range of a given gene expression into two disjoint intervals corresponding to the down-regulation of the gene. It predicts samples in one interval to normal and samples in another interval to cancer [14].

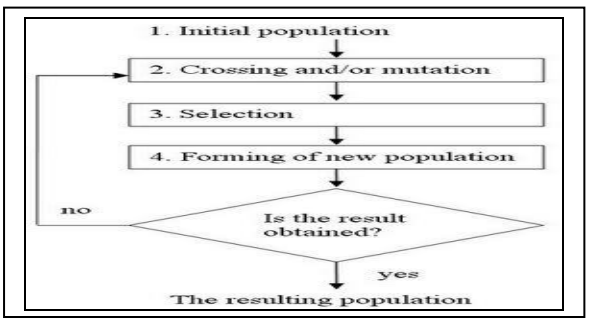

Fig. 3. Block Diagram of Genetic Algorithm

3.2 Classification Algorithms

Support Vector Machine (SVM) is considered popular classifier for microarray data [22]. It has an advantage applied in cancer diagnostic in that its performance appears not to be affected by using the set of full genes [9]. k- Nearest Neighbor (KNN) is one of the simplest learning algorithms, and applied to a variety of problem. It is used as a classifier among a given set of data and uses class labels of the most similar neighbor to predict the new class [9]. Naïve Bayes (NB) is a classifier that can achieve relatively good performance on classification tasks, based on the elementary Bayes' theory [9]. Decision Tree (DT) different methods exit to build a DT, in which a given data in a tree structure, with each branch representing an association between attribute values and a class label [9]. The most famous DT methods is the C4.5 algorithm, which partition the training data set according to tests on the potential of attribute values in separating the classes.

Table 3. Feature Selection Algorithms Specifications

Methods/ Technology involved	**Importance**	**Area/s**	**Advantages**	**Disadvantages**	**Problems**
Filter Selection Techniques	Compute the importance of each feature (gene) and then select the top ranked	Gene Selection	Simple Fast Easy scales to very high dimensional data	Univarate that means each feature is considered and treated separately, ignoring any correlation between features	Low classification performance
Wrapper Selection Technique	Selects subset of features that is useful to build a good classifier or predictor	Gene Selection	The ability to take into account the correlation between features and the interaction with the classifier	Prone to high risk of over fitting It require very intensive computation	Unfeasible for feature selection in high-dimensional data More complex

4 Results and discussion

In this paper, various algorithms were analyzed that perform the task of cancer gene search and classification by first selecting the informative genes and reducing the size and then distinguish the type of the cell tumor or not. Cancer gene selection is a preprocessing step used to find a reduced-sample size of microarray data. This can be achieved by two feature (gene) selection approaches as stated in Table 3. From the table it is found that both filter and wrapper models play a role in feature (gene) selection, but each has its pros and cons. Filter model is noticed to be fast but may give a low classification performance result, while the wrapper model takes time and more complex, but may give somehow a high performance result. Furthermore,it is noticed

from Table 4 (see appendix 1), that multiple algorithms implemented in integration and hybridization to analyze multiple kinds of cancer type. In addition, the efficiency of the algorithms was based on the cancer type and the algorithm implemented. The need for a scientific methodology to determine the efficient algorithm or integration of algorithms for cancer types was missed. We mean by algorithm efficiency how fast the algorithm to be implemented in terms of time and speed in order to analyze the cancer cells. Furthermore, Table 5 (see appendix 1) gives a summary for each individual algorithm and to which cancer type it was implemented. It is concluded from table 5 (see appendix 1), that Genetic Algorithm as a selection algorithm was implemented to almost all cancer types for a high performance, except the brain cancer, while Decision Tree and Support Vector Machine Algorithms were implemented to almost all types of cancer for high performance results. In addition figure 4 shows that the Integrated Algorithm for gene selection and classification has the highest accuracy 99% for colon and breast cancers, while the Bootstrapped Genetic Algorithm and Support Vector Machine give good performance accuracy without indicating the percentage. Also the Integrated Gene Search Algorithm has the second high performance up to 98% in accuracy results.

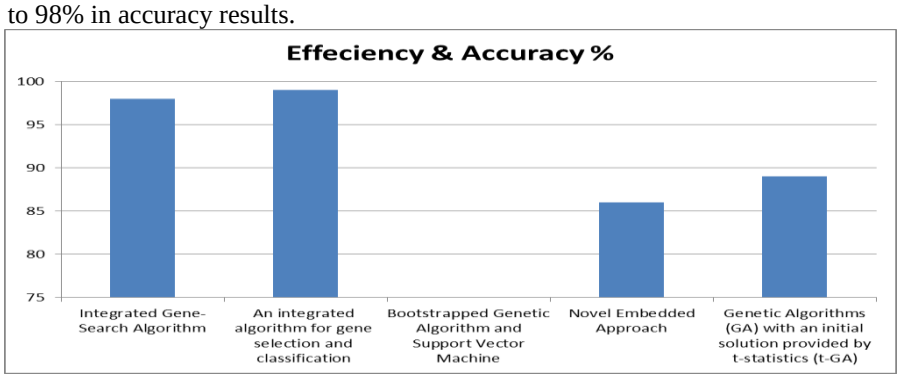

Fig. 4. Algorithm Efficiency and Accuracy

On the other hand, from the detailed review to many researchers' contributions, Table 6 (see appendix 1) summarizes out the most common Algorithms used for cancer gene search and classifications, most of these algorithms where implemented in an integrated model or hybridization methods as discussed, in order to give out an optimum desired result.

The main issue with the previous algorithms is the efficiency in performance, due that most of the suggested algorithms and technologies followed the hybridization methodology in order to achieve better in terms of efficiency and accuracy. When we talk about efficiency we mean less time and less memory, but the main concern will be saving time.

5 Conclusion and Future Work

It is concluded that there are multiple computational algorithms applied for cancer gene selection that are either filter or wrapper methods, each has its own advantages

or disadvantages and trying to reach a well performance result. On the other hand and in order to classify cancer cells, selection algorithms must be implemented first to reduce the microarray sample size and reach informative genes, then it would be easier to implement classifier algorithms to distinguish out tumor from normal cells. Moreover, the paper showed that most algorithms are implemented in an integration methodology and in a harmony in order to achieve a better performance result. Nevertheless, it was clear that the dominant algorithm applied in integration with other algorithms for gene selection was the Genetic Algorithm, while for classification was the Support Vector Machine; as both reached better results. The future work will be to analyze the processing time of each of the algorithms implemented in order to decide the best performance algorithm.

6 References

1. S. Shah and A. Kusiak, *Cancer gene search with data mining and genetic algorithms*, Computers in Biology and Medicine, vol.37, no.2, pp.251-261, 2007.
2. Zne-Jung Lee, An integrated algorithm for gene selection and classification applied to microarray data of ovarian cancer, International Journal Artificial Intelligence in Medicine 42 (2008) 81-93.
3. Huawen Liu, Lei Liu, Huijie Zhang, *Ensemble gene selection for cancer classification*, Pattern Recognition, Volume 43, Issue 8, August 2010, Pages 2763-2772, ISSN 0031-3203, 10.1016/j.patcog.2010.02.008.
4. Chen, X.-W., "Gene selection for cancer classification using bootstrapped genetic algorithms and support vector machines," Bioinformatics Conference, 2003. CSB 2003. Proceedings of the 2003 IEEE , vol., no., pp.504,505, 11-14 Aug. 2003
5. Chanho Park; Sung-Bae Cho, "Evolutionary ensemble classifier for lymphoma and colon cancer classification," Evolutionary Computation, 2003. CEC '03. The 2003 Congress on , vol.4, no., pp.2378,2385 Vol.4, 8-12 Dec. 2003
6. Mohamad, M.S.; Omatu, S.; Yoshioka, M.; Deris, S., "An Approach Using Hybrid Methods to Select Informative Genes from Microarray Data for Cancer Classification," Modeling & Simulation, 2008. AICMS 08. Second Asia International Conference on , vol., no., pp.603,608, 13-15 May 2008
7. Huawen Liu, Lei Liu, Huijie Zhang, Ensemble gene selection for cancer classification, Pattern Recognition, Volume 43, Issue 8, August 2010, Pages 2763-2772, ISSN 0031-3203
8. M.S. Mohamad, S. Omatu, S. Deris, and S.Z.M. Hashim, "A Model for Gene Selection and Classification of Gene Expression Data", International Journal of Artificial Life & Robotics, Springer, 11(2), 2007, pp. 219–222.
9. Osareh, A.; Shadgar, B., "*Microarray data analysis for cancer classification*," Health Informatics and Bioinformatics (HIBIT), 2010 5th International Symposium on , vol., no., pp.125,132, 20-22 April 2010
10. Jukka K. Nurminen, "*Using software complexity measures to analyze algorithms—an experiment with the shortest-paths algorithms*", Computers & Opera-

tions Research, Volume 30, Issue 8, July 2003, Pages 1121-1134, ISSN 0305-0548, 10.1016/S0305-0548(02)00060-6.
11. Solorio-Fernandez, S.; Martinez-Trinidad, J.F.; Carrasco-Ochoa, J.A.; Yan-Qing Zhang, "*Hybrid feature selection method for biomedical datasets*," Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), 2012 IEEE Symposium on , vol., no., pp.150,155, 9-12 May 2012
12. Lipo Wang; Feng Chu; Wei Xie, "Accurate Cancer Classification Using Expressions of Very Few Genes," Computational Biology and Bioinformatics, IEEE/ACM Transactions on , vol.4, no.1, pp.40,53, Jan.-March 2007
13. Jiexun Li; Hua Su; Hsinchun Chen; Futscher, B.W., "*Optimal Search-Based Gene Subset Selection for Gene Array Cancer Classification*," Information Technology in Biomedicine, IEEE Transactions on , vol.11, no.4, pp.398,405, July 2007
14. Jinn-Yi Yeh; Tai-Shi Wu; Min-Che Wu; Der-Ming Chang, "Applying Data Mining Techniques for Cancer Classification from Gene Expression Data," Convergence Information Technology, 2007. International Conference on , vol., no., pp.703,708, 21-23 Nov. 2007
15. Constantino Carlos Reyes-Aldasoro, Michael K. Griffiths, Deniz Savas, Gillian M. Tozer, "*CAIMAN: An online algorithm repository for Cancer Image Analysis"*, Computer Methods and Programs in Biomedicine, Volume 103, Issue 2, August 2011, Pages 97-103, ISSN 0169-2607, 10.1016/j.cmpb.2010.07.007.
16. Goldberg, D.E., "Genetic algorithms in search, optimization and machine learning". Addison Wesley, MA, 1989.
17. Lei Yu, Huan Liu, " Feature selection for high-dimensional data: A fast correlation-based filter solution", In in ICML (2003), pp. 856-863
18. R. Tiwari and M. P. Singh, "Correlation-based Attribute Selection using Genetic Algorithm," International Journal of Computer Applications (0975 – 8887), vol. 4, no. 8, pp. 28-34, August 2010.
19. Khanesar, M.A.; Teshnehlab, M.; Shoorehdeli, M.A., "A novel binary particle swarm optimization," Control & Automation, 2007. MED '07. Mediterranean Conference on , vol., no., pp.1,6, 27-29 June 2007
20. Eberhart, R.C.; Yuhui Shi, "Particle swarm optimization: developments, applications and resources," Evolutionary Computation, 2001. Proceedings of the 2001 Congress on , vol.1, no., pp.81,86 vol. 1, 2001
21. A. Gelman, "Analysis of Variance - Why it is More Important Than Ever", The Annals of Statistics, Vol. 33, No. 1. pp. 1-53, 2005
22. V. Vapnik, "*Statistical learning theory"*, Wiley, 1998

Appendix

Table 4. Efficient Algorithms for various cancer types

Algorithm	Embedded Algorithms	Cancer Type	Comments
Integrated Gene-Search Algorithm[1]	Genetic Algorithm Correlation-based heuristics Decision tree Support vector machine	Ovarian Prostate Lung Can be successfully applied to any other cancer like colon, breasted, bladder, leukemia, and so on.	High classification accuracy (94 – 98%)
An integrated algorithm for gene selection and classification [2]	Genetic Algorithm Particle Swarm Optimization Support Vector Machine Analysis of Variance Fuzzy Model	Ovarian Colon Breast	Superior performance for gene selection and classification (colon and breast 99% accuracy)
Bootstrapped Genetic Algorithm and Support Vector Machine [4]	Genetic Algorithm Support vector machine	Colon Leukemia	Well suited for feature (gene) selection problems
Novel Embedded Approach [9]	Information Gain Relief Algorithm t-statistics Support Vector Machine K Nearest Neighbour Naïve Bayes Neural Network Decision Tree	Lung Prostate Breast Leukemia Brain Colon Ovarian	Suport Vector Machines peroforms accuracies > 85% with the combination of Information Gain Decision Tree are the worst model in accuracy
Genetic Algorithms (GA) with an initial solution provided by t-statistics (t-GA) [14]	Genetic Algorithm T-statistics Decision Tree	Colon Leukemia Lymphoma Lung Central Nervous System (CNS)	Colon accuracy 89% Leukemia accuracy 94% Lymphona accuracy 92% Lung accuracy 98% CNS accuracy 77%
CAIMAN system (CAncer IMage ANalysis) [15]	Migration measurement Vasculature tracing Shading correction	Cancer related images	More algorithms can be implemented

Table 5. Cancer Types Algorithms

Cancer / Algorithm	Ovarian	Prostate	Lung	Colon	Breast	Bladder	Leukemia	Brain	Lymphoma	CNS
Genetic Algorithm	✓	✓	✓	✓	✓	✓	✓		✓	✓
Correlation based heuristics	✓	✓	✓	✓	✓	✓	✓			
Decision tree	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Support Vector Machine	✓	✓	✓	✓	✓	✓	✓	✓		
Particle Swarm Optimization	✓			✓	✓					
Analysis of variance	✓			✓	✓					
Fuzzy Model	✓			✓	✓					
Information Gain	✓	✓	✓	✓	✓		✓	✓		
Relief Algorithm	✓	✓	✓	✓	✓		✓	✓		
t-statistics	✓	✓	✓	✓	✓		✓	✓	✓	✓
K nearest Neighbor	✓	✓	✓	✓	✓		✓	✓		
Naïve Bayes	✓	✓	✓	✓	✓		✓	✓		
Neural Network	✓	✓	✓	✓	✓		✓	✓		

Table 6. Common Feature Selection and Classifications Algorithms

Selection Algorithms	**Classification Algorithms**
Genetic Algorithm (GA)	Support Vector Machine (SVM)
Correlation-based heuristics (Correlation-based feature selection) (CFS)	Bootstrapped SVM
Particle Swarm Optimization (PSO)	K-Nearest Neighbors (KNN)
Analysis of Variance (ANOVA)	Naïve Bayes
Information Gain (IG)	Neural Networks (NN)
Relief Algorithm (RA)	Decision Tree (DT)
t-statistics (TA)	Bagging and Stacking Algorithms
	Fuzzy Model