*Research Article*

# Incremental Search for Informative Gene Selection in Cancer Classification

**Fathima Fajila\* and Yuhanis Yusof**

Universiti Utara Malaysia, Sintok, Kedah, Malaysia
fajilanisper@gmail.com; yuhanis@uum.edu.my
**\*Correspondence:** fajilanisper@gmail.com

**Abstract:** Although numerous methods of using microarray data analysis for classification have been reported, there is space in the field of cancer classification for new inventions in terms of informative gene selection. This study introduces a new incremental search-based gene selection approach for cancer classification. The strength of wrappers in determining relevant genes in a gene pool can be increased as they evaluate each possible gene's subset. Nevertheless, the searching algorithms play a major role in gene's subset selection. Hence, there is the possibility of finding more informative genes with incremental application. Thus, we introduce an approach which utilizes two searching algorithms in gene's subset selection. The approach was efficient enough to classify five out of six microarray datasets with 100% accuracy using only a few biomarkers while the rest classified with only one misclassification.

**Keywords:** *Cancer classification; Gene's subset; Informative gene; Microarray; Wrappers*

## 1. Introduction

Cancer is characterized as uncontrolled cell growth in parts of the human body. Early prognosis can help in providing treatments to restrict further spreading of the disease within the body. Moreover, the cancer diagnosis is very important as there are a variety of subtypes which can be ameliorated with discriminant treatments. Recently, microarray invaded into cancer research field intending to find biomarkers for cancer classification. Innumerable solutions are being suggested related to cancer classification using computational techniques [1].

The cancer microarray provides gene-based knowledge related to cancer. Typically, a microarray dataset is a matrix consisting of rows and columns. Each row relates to a particular sample whereas each column relates to a particular gene. The size of the microarray matrix is vast as it provides thousands of genes compared to a low number of samples [2, 3]. This is a major problem in cancer microarray analysis [3]. Therefore, selecting the most relevant genes from the microarray datasets is important. Typically, microarray datasets include redundant and irrelevant genes in addition to informative genes. Thus, analysis facilitating the elimination of redundant and irrelevant genes while preserving the informative genes can lead to more efficient cancer classification.

Recently, computational approaches have received more attention in cancer classification [4-8]. However, the performance of many existing cancer classification approaches does not seem to be sufficiently robust [4, 8]. Some of them suffer from poor performance in terms of classification [9, 10] whereas others result in huge gene's subsets [8, 11, 12]. Many hybrid approaches have been proposed for cancer classification in recent decades [9, 13, 14].

Yang et al. [15] presented an approach with filter and swarm algorithm. Information Gain (IG) filter and correlation-based filter were used with binary particle swarm optimization algorithm in the study [15]. Gunavathi and Premalatha [9] proposed a hybrid approach with Genetic Algorithm (GA).

Two classifiers including Support Vector Machine (SVM) were used to classify ten binary cancer microarray datasets. However, the study [9] did not reveal significant performance improvements. Further, Motieghader et al. [16] presented a hybrid approach with GA and Learning Automata named as GALA. The proposed approach worked well for some datasets, yet provided low accuracy and large gene's subset on few datasets.

Moreover, a hybrid approach with IG filter and SVM was suggested by Gao et al. [14]. Five binary cancer microarray datasets were classified using the LIBSVM library. The three most highly prioritized genes were selected to demonstrate the competence of the study [14]. A hybrid approach with adaptive GA was suggested by Lu et al. [11]. However, the approach [11] produced large gene's subsets. Alshamlan [5] proposed a hybrid approach with a filter namely correlation-based feature selection filter. Artificial Bee Colony (ABC) algorithm was applied at the wrapper step giving the name as Co-ABC for the proposed method. Six microarray datasets were classified with SVM classifier. The study [5] demonstrated acceptable classification accuracy with few biomarkers. Recently, Alanni et al. [17] suggested IG filter and Gene Selection Programming (GSP) for cancer classification. The method [17] was demonstrated on ten cancer microarray datasets.

Even though there are numerous suggestions for gene selection provided from the existing studies, still selecting the best combination of wrapper and filter with an appropriate searching algorithm is challenging. Therefore, we propose a hybrid feature selection approach with incremental search strategy for cancer classification. Two searching algorithms are utilized for informative gene's subset selection. The detailed methodology is provided in section 2. The experimental results and a discussion are provided in section 3, and the conclusion in section 4.

## 2. Methodology

Fig. 1 display the pseudo-code of the hybrid feature selection method, referred to as Incremental Search for Informative Gene (ISIG) selection. The method has two steps. Initially a data pre-processing step is applied to remove irrelevant and redundant genes from the standard microarray dataset. In this study, the IG filter is employed for pre-processing. Typically, filters evaluate all the features with a criterion. Thus, ranking only the relevant genes by using a threshold value to eliminate the less prioritized genes is an issue with filters. Hence, we use $n-1$ rule where $n$ denotes the sample count in the dataset to selects the highly prioritized $n-1$ genes through pre-processing [18]. IG filter uses the entropy concept to rank the features. Thus, the proposed ISIG approach selects the highly ranked $n-1$ genes through pre-processing. The size of the gene's subset resultant from the pre-processing step can be further processed for biomarker selection. Thus, the subset selection step which uses the wrapper algorithm is applied as the second step for this task.

Wrappers evaluate feature subsets rather than individual features. Therefore, wrappers are computationally more complex compared to filters. This complexity exponentially increases with the number of features resulting from the initial wrapper application and is thus impractical for very large microarray datasets. Nevertheless, wrappers focus on the interactions between genes through subset evaluation which may result in an informative gene's subset. Hence, justifies the use of wrapper after pre-processing the data.

Another aspect of wrapper is the searching technique. The feature subsets resulted from the wrapper approach vary with regards to the searching algorithms. In this study, two searching algorithms, namely Best First Search (BFS) [19, 20] and Evolutionary Search (ES) [21, 22] are utilized for biomarker selection. In ISIG approach, ES is deployed prior to BFS, hence resulting in a consecutive search as shown in Figure 1. This will then produce a better gene selection as compared to a single search.

Furthermore, even though ES algorithm produces acceptable gene's subset selection, nonetheless, the size of the gene's subset selected are slightly large. On the other hand, BFS provides small gene's subsets yet, not as efficient as the ES. In the proposed ISIG approach, initially the ES is applied for the pre-processed gene's set. Then, in order to produce a gene's subset with only the most informative genes, BFS is applied to the resultant subset. Such an incremental search not only reduces the gene's subset but also improves the classification performance. The two most common classifiers

used in microarray data analysis namely Naïve Bayes (NB) [23] and SVM [24] are used to assess the competency of the method.

---

**Input**: Cancer microarray dataset
**Output**: Informative gene's subset
**Algorithm**:
1. Pre-process the dataset with IG filter
2. Select the gene's subset using the pre-processed dataset
   a. Evaluate the gene's subsets with ES
   b. Evaluate the gene's subsets with BFS
3. Return the most informative gene's subset
4. Classify the dataset with NB / SVM classifier
5. Return the classification accuracy

---

**Figure 1.** Pseudo-code of the proposed ISIG approach

## 3. Results and Discussion

To realize the proposed ISIG, six cancer microarray datasets namely; Colon cancer [25], Ovarian cancer (Kent Ridge Biomedical Data Repository), Leukemia3 [26], Mixed-Lineage Leukemia (MLL) [27], Small Round Blue Cell Tumor (SRBCT) [28], and Leukemia4 [27] were used for the evaluation. Two datasets are of binary class problem and the remainder represent multi-class problem. The proposed approach was implemented using the open-source Waikato Environment for Knowledge Analysis (WEKA version 3.8.1) software. Information about the datasets is depicted in Table 1.

The method's performance was evaluated using the two most popular evaluation metrics, namely classification accuracy and Receiver Operating Characteristic (ROC). The classification accuracy, number of Correctly Classified Instances (CCI) and ROC value are compared for each dataset. The performance of the classifiers without gene selection is given in Table 2, while Table 2-Table 5 provides the outcome of deploying ISIG. The number of biomarkers selected by each search strategy, as well as by ISIG is given in the parenthesis with the datasets. Table 2 illustrates the classification results obtained for the standard datasets where no data pre-processing was performed. Experimental results of gene's subsets selected with BFS, ES and ISIG approaches for all the datasets are given in Table 3, Table 4, and Table 5 respectively. The most informative gene's subset selected using ISIG is provided in Table 6.

**Table 1.** Cancer microarray datasets information

| Dataset | No. of classes | No. of genes | No. of samples | Description |
|---------|----------------|--------------|----------------|-------------|
| Colon | 2 | 2000 | 62 | Tumor: 40 and Normal: 22 |
| Ovarian | 2 | 15154 | 253 | Cancer: 162 and Normal: 91 |
| Leukemia3 | 3 | 7129 | 72 | AML: 25, B-cell: 38, and T-cell: 9 |
| MLL | 3 | 12582 | 72 | ALL: 24, MLL: 20, and AML: 28 |
| SRBCT | 4 | 2308 | 83 | BL: 11, NB: 18, EWS: 29, and RMS: 25 |
| Leukemia4 | 4 | 7129 | 72 | BM: 21, PB: 4, B-cell: 38, and T-cell: 9 |

*Note - AML: Acute Myeloid Leukemia, ALL: Acute Lymphoblastic Leukemia, MLL: Mixed Lineage Leukemia, BL: Burkitt's Lymphoma, NB: Neuroblastoma, EWS: Ewing's Sarcoma, RMS: Rhabdomyosarcoma, BM: Bone Marrow, and PB: Peripheral Blood*

**Table 2.** Classifier performance of the six cancer datasets without gene selection

| Dataset | NB | | | SVM | | |
|---------|----|----|----|-----|----|----|
| | Accuracy (%) | CCI (Total) | ROC (%) | Accuracy (%) | CCI (Total) | ROC (%) |
| Colon(2000) | 52.38 | 11(21) | 56.7 | 80.95 | 17(21) | 88.2 |
| Ovarian(15154) | 84.88 | 73(86) | 86.9 | 100.00 | 86(86) | 100.0 |
| Leukemia3(7129) | 91.66 | 22(24) | 91.0 | 95.83 | 23(24) | 96.7 |
| MLL(12582) | 95.83 | 23(24) | 96.7 | 100.00 | 24(24) | 100.0 |
| SRBCT(2308) | 96.42 | 27(28) | 97.4 | 100.00 | 28(28) | 100.0 |
| Leukemia4(7129) | 87.5 | 21(24) | 88.5 | 95.83 | 23(24) | 97.0 |

**Table 3.** Performance of biomarkers selected using BFS

| Dataset | NB | | | SVM | | |
|---|---|---|---|---|---|---|
| | Accuracy (%) | CCI (Total) | ROC (%) | Accuracy (%) | CCI (Total) | ROC (%) |
| **Colon(8)** | 90.47 | 19(21) | 97.1 | 90.47 | 19(21) | 94.1 |
| **Ovarian(3)** | 100.00 | 86(86) | 100.0 | 100.00 | 86(86) | 100.0 |
| **Leukemia3(3)** | 95.83 | 23(24) | 96.2 | 91.66 | 22(24) | 95.4 |
| **MLL(5)** | 100.00 | 24(24) | 100.0 | 91.66 | 22(24) | 97.5 |
| **SRBCT(7)** | 92.85 | 26(28) | 98.2 | 89.28 | 25(28) | 96.0 |
| **Leukemia4(5)** | 100.00 | 24(24) | 100.0 | 87.50 | 21(24) | 87.3 |

**Table 4.** Performance of biomarkers selected using ES

| Dataset | NB | | | SVM | | |
|---|---|---|---|---|---|---|
| | Accuracy (%) | CCI (Total) | ROC (%) | Accuracy (%) | CCI (Total) | ROC (%) |
| **Colon(16)** | 90.47 | 19(21) | 97.1 | 90.47 | 19(21) | 94.1 |
| **Ovarian(62)** | 98.83 | 85(86) | 99.9 | 100.00 | 86(86) | 100.0 |
| **Leukemia3(13)** | 95.83 | 23(24) | 100.0 | 95.83 | 23(24) | 95.9 |
| **MLL(15)** | 100.00 | 24(24) | 100.0 | 100.00 | 24(24) | 100.0 |
| **SRBCT(21)** | 100.00 | 28(28) | 100.0 | 100.00 | 28(28) | 100.0 |
| **Leukemia4(18)** | 100.00 | 24(24) | 100.0 | 95.83 | 23(24) | 96.9 |

**Table 5.** Performance of biomarkers selected using ISIG

| Dataset | NB | | | SVM | | |
|---|---|---|---|---|---|---|
| | Accuracy (%) | CCI (Total) | ROC (%) | Accuracy (%) | CCI (Total) | ROC (%) |
| **Colon(4)** | 90.47 | 19(21) | 98.5 | **95.23** | **20(21)** | **97.1** |
| **Ovarian(3)** | **100.00** | **86(86)** | **100.0** | **100.00** | **86(86)** | **100.0** |
| **Leukemia3(4)** | 95.83 | 23(24) | 99.5 | **100.00** | **24(24)** | **100.0** |
| **MLL(4)** | **100.00** | **24(24)** | **100.0** | 79.16 | 19(24) | 89.2 |
| **SRBCT(8)** | **100.00** | **28(28)** | **100.0** | 78.57 | 22(28) | 93.2 |
| **Leukemia4(5)** | **100.00** | **24(24)** | **100.0** | 91.66 | 22(24) | 93.2 |

Even though some datasets have provided 100% classification accuracy without gene selection (refer to Table 2), it is known that standard microarray datasets consist of irrelevant and redundant genes. Based on data depicted in Table3, Table 4, and Table 5, it is noted that by pre-processing the data, a better classification is produced. This can be seen for datasets SRBCT, Leukemia3, MLL, and Colon cancer. Both BFS and ISIG have provided the same classification accuracy for Ovarian cancer dataset and Leukemia4 with the same number of genes.

The proposed ISIG has correctly classified five datasets out of six with 100% accuracy and the rest with only one misclassification. On the other hand, the ES and BFS have classified 4 and 3 datasets with 100% accuracy respectively. Though the classification accuracy for Colon cancer is less than 100% (using both SVM and NB), ISIG has obtained the highest accuracy than the other two methods. Further, it is obvious that ES has selected the subset with the highest number of genes for all the datasets. Though the gene's subsets selected with ISIG is having one more gene in count than that of BFS for SRBCT and Leukemia3 datasets, ISIG has gained the highest performance for both SRBCT and Leukemia3 dataset. On the other hand, for Ovarian cancer and Leukemia4 datasets, both BFS and ISIG approaches have given a similar performance related to classification accuracy and biomarker count. Overall, none of the individual search methods obtained a higher accuracy than ISIG for any dataset. This shows the effectiveness of employing incremental search in gene's subset selection.

Colon cancer classification produced by ISIG is better than the ones reported in the literature [9, 11, 14, 29, 30]. Even though Alshamlan [5] and Motieghader et al. [16] obtained slightly higher accuracy for Colon cancer classification, yet they are of larger gene's subset. Bouazza et al. [31] suggested a filter approach which produced 30 genes for Ovarian cancer classification whereas we could obtain the same accuracy while using only three genes. The same outcome can be seen in the work proposed by Gunavathi & Premalatha [9] who proposed GA for gene selection. They provided

ten genes for Ovarian cancer whereas the proposed ISIG works well with three genes. As for the Leukemia3 dataset, the accuracy obtained in this study is greater than the ones reported by Mazumder and Veilumuthu [32]. When compared to the outcome of Alshamlan [5], even though accuracy is the same, yet the researchers achieved it with six genes whereas it is four genes in this study.

The competency of the approach related to accuracy and informative gene count obtained from this study for MLL dataset is superior in contrast to the approaches mentioned in the literature [16, 32]. Compared to earlier studies [27, 30, 33], ISIG has shown better performance for SRBCT dataset. However, the accuracy for SRBCT dataset is the same as mentioned in [5, 32] regardless the gene's subset. Further, the classification accuracy and the amount of biomarkers for Leukemia4 dataset is 100% and five in the proposed study whereas it is 100% yet with seven genes in [32]. Moreover, the performance of Leukemia4 classification is better than that of mentioned in the past [4, 7, 12].

Thus, the results achieved with the ISIG approach shows that it is comparatively good for cancer classification. The informative genes selected with ISIG approach is tabulated in Table 6.

**Table 6.** The informative biomarkers selected using ISIG for the six datasets

| Dataset | Genes |
|---|---|
| **Colon(4)** | A249, A765, A1679, A377 |
| **Ovarian(3)** | MZ244.95245, MZ434.68588, MZ2.7921478 |
| **Leukemia3(4)** | M27891_at, U05259_rna1_at, L08895_at, D00749_s_at |
| **MLL(4)** | 35164_at, 37539_at, 35614_at, 37988_at |
| **SRBCT(8)** | gene1601, gene1708, gene417, gene2050, gene2022, gene1207, gene2144, gene129 |
| **Leukemia4(5)** | M23197_at, M27891_at, U05259_rna1_at, L47738_at, D87292_at |

## 4. Conclusions

The paper presents a hybrid gene selection approach, namely ISIG which is based on applying BFS followed by ES to produce gene's subset. The ES selects comparatively informative gene's subsets compared to BFS. Yet, ES provides relatively large subsets while the deployment of BFS helps to reduce the size of the informative gene's subsets. The ISIG approach is assessed on six standard cancer datasets. Effectiveness of ISIG was investigated on both binary and as well as multi-class problems. It is learned that the proposed incremental search is beneficial as higher accuracy is obtained while using fewer number of genes. Hence, various combinations of incremental strategies are aimed to be proposed for feature selection in the future.

## References

[1] Alshamlan Hala, Ghada Badr and Yousef Alohali, "A comparative study of cancer classification methods using microarray gene expression profile", In: Herawan T., Deris M., Abawajy J. (eds) Proceedings of the First International Conference on Advanced Data and Information Engineering (DaEng-2013), Lecture Notes in Electrical Engineering, Springer, Singapore, Print ISBN: 978-981-4585-17-0, Online ISBN: 978-981-4585-18-7, pp. 389-398, Vol. 285, 2014, DOI: https://doi.org/10.1007/978-981-4585-18-7_44, Available: https://link.springer.com/chapter/10.1007/978-981-4585-18-7_44.

[2] Ali El Akadi, Aouatif Amine, Abdeljalil El Ouardighi and Driss Aboutajdine, "A two-stage gene selection scheme utilizing MRMR filter and GA wrapper", *Knowledge and Information Systems*, pp. 487-500, Vol. 26, No. 3, March 2011, DOI: https://doi.org/10.1007/s10115-010-0288-x, Available: https://link.springer.com/article/10.1007/s10115-010-0288-x#citeas.

[3] Chyh-Ming Lai, Wei-Chang Yeh and Chung-Yi Chang, "Gene selection using information gain and improved simplified swarm optimization", *Neurocomputing*, ISSN: 0925-2312, pp. 331-338, Vol. 218, 2016, DOI: 10.1016/j.neucom.2016.08.089.

[4] Mohammad Al-Batah, Belal Zaqaibeh, Saleh Ali Alomari and Mowafaq Salem Alzboon, "Gene Microarray Cancer Classification using Correlation Based Feature Selection Algorithm and Rules Classifiers", *International Journal of Online & Biomedical Engineering*, pp. 62-73, Vol. 15, No. 8, 2019, DOI: 10.3991/ijoe.v15i08.10617, Available: https://doi.org/10.3991/ijoe.v15i08.10617.

[5] Hala Mohammed Alshamlan, "Co-ABC: Correlation artificial bee colony algorithm for biomarker gene discovery using gene expression profile", *Saudi Journal of Biological Sciences*, ISSN: 1319-562X, pp. 895-903, Vol. 25, No. 5, 2018, DOI: 10.1016/j.sjbs.2017.12.012.

[6] Das Kaberi and Debahuti Mishra, "Hybridized univariate and multivariate filter based approaches for gene selection", *International Journal of Pharma and Bio Sciences*, ISSN: 0975-6299, pp. 1215-1226, Vol. 7, No. 3, 2016.

[7] Mazumder Dilwar Hussain and Ramachandran Veilumuthu, "An enhanced feature selection filter for classification of microarray cancer data", *ETRI Journal*, pp. 358-370, Vol. 41, No. 3, 2019, DOI: 10.4218/etrij.2018-0522, Available: https://onlinelibrary.wiley.com/doi/full/10.4218/etrij.2018-0522.

[8] Panda Mrutyunjaya, "Elephant search optimization combined with deep neural network for microarray data analysis", *Journal of King Saud University-Computer and Information Sciences*, ISSN: 1319-1578, pp. 940-948, Vol. 32, No. 8, 2020, DOI: 10.1016/j.jksuci.2017.12.002.

[9] Gunavathi Chellamuthu and Kandasamy Premalatha, "Performance analysis of genetic algorithm with kNN and SVM for feature selection in tumor classification", *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, pp. 1490-1497, Vol. 8, No. 8, 2014, Available: https://www.researchgate.net/publication/278036571.

[10] Wang Yadi, Xin-Guang Yang and Yongjin Lu, "Informative gene selection for microarray classification via adaptive elastic net with conditional mutual information", *Applied Mathematical Modelling*, ISSN: 0307-904X, pp. 286-297, Vol. 71, 2019, DOI: 10.1016/j.apm.2019.01.044.

[11] Lu Huijuan, Junying Chen, Ke Yan, Qun Jin, Yu Xue and Zhigang Gao, "A hybrid feature selection algorithm for gene expression data classification", *Neurocomputing*, ISSN: 0925-2312, pp. 56-62, Vol. 256, 2017, DOI: 10.1016/j.neucom.2016.07.080.

[12] B.H. Shekar and Guesh Dagnew, "L1-Regulated Feature Selection and Classification of Microarray Cancer Data Using Deep Learning", In: *Chaudhuri B., Nakagawa M., Khanna P., Kumar S. (eds) Proceedings of 3rd International Conference on Computer Vision and Image Processing, Advances in Intelligent Systems and Computing*, Print ISBN: 978-981-32-9290-1, Online ISBN: 978-981-32-9291-8, Vol. 1024, 2020, Springer, Singapore, DOI: 10.1007/978-981-32-9291-8_19, Available: https://link.springer.com/chapter/10.1007/978-981-32-9291-8_19#citeas.

[13] Hala M. Alshamlan, Ghada H. Badr and Yousef A. Alohali, "Genetic Bee Colony (GBC) algorithm: A new gene selection method for microarray cancer classification", *Computational biology and chemistry*, ISSN: 1476-9271, pp. 49-60, Vol. 56, 2015, DOI: 10.1016/j.compbiolchem.2015.03.001.

[14] Lingyun Gao, Mingquan Ye, Xiaojie Lu and Daobin Huang, "Hybrid method based on information gain and support vector machine for gene selection in cancer classification", *Genomics, proteomics & bioinformatics*, ISSN: 1672-0229, pp. 389-395, Vol. 15, No. 6, 2017, DOI: 10.1016/j.gpb.2017.08.002, Available: https://www.sciencedirect.com/science/article/pii/S1672022917301675.

[15] Yang Cheng-San, Li-Yeh Chuang, Chao-Hsuan Ke and Cheng-Hong Yang, "A Hybrid Feature Selection Method for Microarray Classification", *IAENG International Journal of Computer Science*, pp. 285-290, Vol. 35, No. 3, 2008.

[16] Motieghader Habib, Ali Najafi, Balal Sadeghi and Ali Masoudi-Nejad, "A hybrid gene selection algorithm for microarray cancer classification using genetic algorithm and learning automata", *Informatics in Medicine Unlocked*, ISSN: 2352-9148, pp. 246-254, Vol. 9, 2017, DOI: 10.1016/j.imu.2017.10.004.

[17] Alanni Russul, Jingyu Hou, Hasseeb Azzawi and Yong Xiang, "A novel gene selection algorithm for cancer classification using microarray datasets", *BMC medical genomics*, pp. 1-12, Vol. 12, No. 1, 15th January 2019, DOI: 10.1186/s12920-018-0447-6, Available: https://link.springer.com/article/10.1186/s12920-018-0447-6.

[18] Wang Yu, Igor V. Tetko, Mark A. Hall, Eibe Frank, Axel Facius, Klaus FX Mayer and Hans W. Mewes, "Gene selection from microarray data for cancer classification—a machine learning approach", *Computational biology and chemistry*, ISSN: 1476-9271, pp. 37-46, Vol. 29, No. 1, 2005, DOI: 10.1016/j.compbiolchem.2004.11.001.

[19] Pearl Judea, *Heuristics: Intelligent search strategies for computer problem solving*, 1st ed. United States: Addision Wesley, 1st January 1984.

[20] Stuart Russell and Norvig Peter, *Artificial intelligence: a modern approach*, 2nd ed., New Jersey, USA: Prentice Hall, 2003.

[21] Rechenberg Ingo, "Cybernetic solution path of an experimental problem", *Royal Aircraft Establishment Library Translation*, 1122, 1965, Available: https://ci.nii.ac.jp/naid/10000137330/.

[22] Schwefel H-P, "Kybernetische Evolution als Strategie der experimentellen Forschung in der Stromungstechnik", Diploma thesis, Technical Univ. of Berlin, 1965, Available: https://ci.nii.ac.jp/naid/10011278419/#cit.

[23] Bayes Thomas, "LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFR S", *Philosophical transactions of the Royal Society of London*, Print ISSN: 0261-0523, Online ISSN: 2053-9223, pp. 370-418, No. 53, 1st January 1763, DOI: 10.1098/rstl.1763.0053, Available: https://royalsocietypublishing.org/doi/abs/10.1098/rstl.1763.0053.

[24] Vapnik Vladimir N, "An overview of statistical learning theory", *IEEE transactions on neural networks*, Print ISSN: 1045-9227, Online ISSN: 1941-0093, pp. 988-999, Vol. 10, No. 5, September 1999, DOI: 10.1109/72.788640, Available: https://ieeexplore.ieee.org/abstract/document/788640.

[25] Alon Uri, Naama Barkai, Daniel A. Notterman, Kurt Gish, Suzanne Ybarra, Daniel Mack and Arnold J. Levine, "Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays", In *Proceedings of the National Academy of Sciences,* pp. 6745-6750, Vol. 96, No. 12, 1999, DOI: 10.1073/pnas.96.12.6745.

[26] Scott A. Armstrong, Jane E. Staunton, Lewis B. Silverman, Rob Pieters, Monique L. den Boer, Mark D. Minden, Stephen E. Sallan, Eric S. Lander, Todd R. Golub and Stanley J. Korsmeyer, "MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia", *Nature genetics*, pp. 41-47, Vol. 30, No. 1, 2002 DOI: 10.1038/ng765, Available: https://www.nature.com/articles/ng765z.

[27] Zhu Zexuan, Yew-Soon Ong and Manoranjan Dash, "Markov blanket-embedded genetic algorithm for gene selection", *Pattern Recognition,* ISSN: 0031-3203, pp. 3236-3248, Vol. 40, No. 11, 2007, DOI: 10.1016/j.patcog.2007.02.007.

[28] Javed Khan, Jun S. Wei, Markus Ringner, Lao H. Saal, Marc Ladanyi, Frank Westermann, Frank Berthold et al., "Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks", *Nature medicine,* pp. 673-679, Vol. 7, No. 6, 2001, DOI: 10.1038/89044, Available: https://www.nature.com/articles/nm0601_673.

[29] Rabia Aziz, CKa Verma and Namita Srivastava, "A fuzzy based feature selection from independent component subspace for machine learning classification of microarray data", *Genomics data,* ISSN: 2213-5960, pp. 4-15, Vol. 8, 2016, DOI: 10.1016/j.gdata.2016.02.012.

[30] Vural Halit and Abdülhamit Subaşı, "Data-mining techniques to classify microarray gene expression data using gene selection by SVD and information gain", *Modeling of Artificial Intelligence,* ISSN: 2312-0355, pp. 171-182, Vol. 2, 2015, DOI: 10.13187/mai.2015.6.171.

[31] Bouazza Sara Haddou, Khalid Auhmani, Abdelouhab Zeroual and Nezha Hamdi, "Selecting significant marker genes from microarray data by filter approach for cancer diagnosis", *Procedia Computer Science,* ISSN: 1877-0509, pp. 300-309, Vol. 127, 2018, DOI: 10.1016/j.procs.2018.01.126, Available: https://www.sciencedirect.com/science/article/pii/S1877050918301388.

[32] Mazumder Dilwar Hussain and Ramachandran Veilumuthu, "Cancer Classification with a Novel Hybrid Feature Selection Technique", *International Journal of Simulation--Systems, Science & Technology,* Print ISSN: 1473-8031, Online ISSN: 1473-804x, Vol. 19, No. 2, 2018, DOI: 10.5013/IJSSST.a.19.02.07, Available: https://ijssst.info/Vol-19/No-2/paper7.pdf.

[33] Mramor Minca, Gregor Leban, Janez Demšar and Blaž Zupan, "Visualization-based cancer microarray data classification analysis", *Bioinformatics*, pp. 2147-2154, Vol. 23, No. 16, 15 August 2007, DOI: 10.1093/bioinformatics/btm312.