

A SURVEY ON CANCER CLASSIFICATION USING DATA MINING TECHNIQUES

A.Daisy¹, R.Porkodi²

¹M.Phil Research Scholar, Department of Computer Science, Bharathiar University, Coimbatore-46, India

²Assistant Professor, Department of Computer Science, Bharathiar University, Coimbatore-46, India

Abstract: *Classification of data has been effectively utilized in the several areas such as scientific experiments, medical industry, credit approval, weather prediction, customer segmentation, target, marketing, fraud detection and diagnosis and prediction of different disease in biomedical domain. The accurate prediction of the various cancer types helps to providing the better treatment and minimization of the toxicity on the patients. Many data mining techniques have been proposed to predicting the cancer. This paper presents the various data mining techniques used for classification of cancer disease and its merits and demerits etc. The main objective of this analysis is to study the various data mining techniques used in classifying the cancer disease and to improve the accuracy of predicting the cancer in early stage and reduces the death rate.*

Keywords: Data Mining, Classification, Cancer Classification, Prediction.

1. INTRODUCTION

Data mining (Reena, G. 2011) is the iterative process of discovering the interesting knowledge from the large amounts of data stored in the data base. It is relatively young and interdisciplinary field of computer science, and it is the process of extracting the patterns from the large data sets by combining the various data mining techniques. The recent technical advances in the processing power, storage capacity, and inter connectivity of computer technology the data mining is the important tool.

The data mining algorithms (Ismaeel, A. G., et.al.2016) are extensively utilized to classify the cancer disease in the early stage. Recent days the several recent techniques are used to classify the cancer disease such as supervised one unsupervised classification methods. The supervised methods used are Naïve Bayes classifier, J48 Decision Trees and Support Vector Machines, whereas the unsupervised method is an adaptation of the K-means clustering method.

1.1 Classification of cancer

Cancers are classified in two ways: by the type of tissue in which the cancer originates (histological type) and by primary site, or the location in the body where the cancer first developed. The international standard for the classification and nomenclature of histologies is the International Classification of Diseases for Oncology, Third Edition (ICD-O-3).

From a histological standpoint there are hundreds of different cancers, which are grouped into six major categories:

- Carcinoma
- Sarcoma
- Myeloma

- Leukemia
- Lymphoma
- Mixed Types

1.1.1 Carcinoma

Carcinoma refers to a malignant neoplasm of epithelial origin or cancer of the internal or external lining of the body. Carcinomas, malignancies of epithelial tissue, account for 80 to 90 percent of all cancer cases. Epithelial tissue is found throughout the body. It is present in the skin, as well as the covering and lining of organs and internal passageways, such as the gastrointestinal tract. Carcinomas are divided into two major subtypes: adenocarcinoma, which develops in an organ or gland, and squamous cell carcinoma, which originates in the squamous epithelium.

Adenocarcinomas generally occur in mucus membranes and are first seen as a thickened plaque-like white mucosa. They often spread easily through the soft tissue where they occur. Squamous cell carcinomas occur in many areas of the body. Most carcinomas affect organs or glands capable of secretion, such as the breasts, which produce milk, or the lungs, which secrete mucus, or colon or prostate or bladder.

1.1.2 Sarcoma

Sarcoma refers to cancer that originates in supportive and connective tissues such as bones, tendons, cartilage, muscle, and fat. Generally occurring in young adults, the most common sarcoma often develops as a painful mass on the bone. Sarcoma tumors usually resemble the tissue in which they grow.

Examples of sarcomas are:

- Osteosarcoma or osteogenic sarcoma (bone)
- Chondrosarcoma (cartilage)
- Leiomyosarcoma (smooth muscle)
- Rhabdomyosarcoma (skeletal muscle)
- Mesothelial sarcoma or mesothelioma (membranous lining of body cavities)
- Fibrosarcoma (fibrous tissue)
- Angiosarcoma or hemangioendothelioma (blood vessels)
- Liposarcoma (adipose tissue)
- Glioma or astrocytoma (neurogenic connective tissue found in the brain)
- Myxosarcoma (primitive embryonic connective tissue)
- Mesenchymous or mixed mesodermal tumor (mixed connective tissue types)

1.1.3 Myeloma

Myeloma is cancer that originates in the plasma cells of bone marrow. The plasma cells produce some of the proteins found in blood.

1.1.4 Leukemia

Leukemias ("liquid cancers" or "blood cancers") are cancers of the bone marrow (the site of blood cell production). The word leukemia means "white blood" in Greek. The disease is often associated with the overproduction of immature white blood cells. These immature white blood cells do not perform as well as they should, therefore the patient is often prone to infection. Leukemia also affects red blood cells and can cause poor blood clotting and fatigue due to anemia. Examples of leukemia include:

- Myelogenous or granulocytic leukemia (malignancy of the myeloid and granulocytic white blood cell series)
- Lymphatic, lymphocytic, or lymphoblastic leukemia (malignancy of the lymphoid and lymphocytic blood cell series)
- Polycythemia vera or erythremia (malignancy of various blood cell products, but with red cells predominating)

1.1.5 Lymphoma

Lymphomas develop in the glands or nodes of the lymphatic system, a network of vessels, nodes, and organs (specifically the spleen, tonsils, and thymus) that purify bodily fluids and produce infection-fighting white blood cells, or lymphocytes. Unlike the leukemias which are sometimes called "liquid cancers," lymphomas are "solid cancers". Lymphomas may also occur in specific organs such as the stomach, breast or brain. These lymphomas are referred to as extranodal lymphomas. The lymphomas are subclassified into two categories: Hodgkin lymphoma and Non-Hodgkin lymphoma. The presence of Reed-Sternberg cells in Hodgkin lymphoma diagnostically distinguishes Hodgkin lymphoma from Non-Hodgkin lymphoma.

1.1.6 Mixed Types

The type components may be within one category or from different categories. Some examples are:

- adenosquamous carcinoma
- mixed mesodermal tumor
- carcinosarcoma
- teratocarcinoma

2. LITERATURE SURVEY

Nilashi, M., et.al. [3] presented the knowledge based system to classification of breast cancer by using the clustering, noise removal and classification techniques. In present techniques the expectation maximization (EM) used as the clustering method for clustering the data in the related groups. Then the classification and Regression trees are utilized to generate the fuzzy rules for classification of the breast cancer disease in the present knowledge based system of fuzzy rule techniques. In order to overcome the multi collinearity issue we include Principal component Analysts (PCA) in the present technique. The present

technique tested with Wisconsin Diagnostic Breast cancer and Mammographic mass datasets and the result shows that improve the prediction accuracy of breast cancer.

Mahapatra, R., et.al. [4] presents the single layer Artificial Neural network for classification of cancer patient. In order to achieve the various feature reduction schemes the set of simple classifier utilized in this paper. Initially the Principal component analysis (PCA), Factor Analysis (FA) and Discrete Fourier Transform (DFT) utilized to reducing the dimension, after that these reduction dimensions are used to build the intelligent classifiers by using the various functional link Artificial Neural Network (FLANN).

Elyasigomari, V., et.al. [5] proposed hybrid approach MRMR-COA-HS (Minimum Redundancy and Maximum Relevance-Cuckoo Optimization Algorithm-Harmony Search) to overcome the gene selection issue to classification of cancer. Initially the MRMR utilized in the pre processing stage to select the top 100 genes, then the selected genes are fed in to the wrapper set up which contains the COA-HS algorithm and the SVM classification technique used in present technique it provides the high accuracy then finally classification performance of the selected genes are measured in terms of accuracy.

Liu, H., et.al. [6] proposed the Ensemble Gene selection method (EGS) to chosen the multiple gene subsets for classification purpose. In this technique the genes are chosen based on the conditional mutual information. The result shows that the present gene subset has good discriminative capability for data classification. Additionally the number of selected genes of the present techniques also finds out self adaptively. In order to increase the diversity of the present technique the initial points are allocate to various genes with highest information. If the multiple gene subsets have been obtained the present technique provides the train base classifiers and then the result is integrated by the majority voting strategy. The result shows that the present technique is outperform than the traditional techniques.

Jeleń, Ł., et.al. [7] presents the application of pattern recognition and image processing techniques to examine the significance of the feature extraction from the fine needle aspiration biopsy images and the cause of the reducing large number features by utilizing the various feature selection methods with small loss of the classification accuracy. The present technique used to reduce the size of the feature vector size then perform the classification with less information and deliver the satisfied results. The reduction of the feature vector leads to reduce the complexity of classification. The best classification feed forward to the neural network when the correlation measures are utilized. The present technique with correlation feature vector reach good performance and able classify the breast cancer data with high accuracy while the feature vector size reduced iteratively.

Nguyen, T., et.al. [8] proposed the novel approach for classification for predicting the cancer through gene expression profiles which is build with supervised learning hidden markov models (HMMs). The gene expression of

each tumour is designed by HMM and the prominent discriminant genes are chosen by the present techniques depends on the modification of the analytic hierarchy process (AHP). The modified AHP allows quantitative factors which are used to rank the outcomes of the individual gene selection methods such as t-test, entropy, receiver operating characteristic curve, Wilcoxon test and signal to noise ratio. The result shows that the HMM is the powerful tool for cancer classification better than the classical classification techniques. The combinations of AHP-HMM provide the better stability and robustness to selection of gene and improve early detection, ease of use to the treatment of cancers in effective and efficient manner.

Xie, H., et.al. [9] presents random projection (RP) technique utilized to reduce the high dimensional features in to low dimensional space with the short duration to predicting the classification of cancer disease. In order to improve the accuracy of the random projection technique it's combining with other techniques such as Principle Component Analysis (PCA), Linear Discriminant Analysis (LDA), and Feature Selection (FS). The different combination of the methods tested with the microarray dataset. The result shows that the feature selection with random projection improves the classification accuracy better than the PCA and LDA.

Piao, Y., et.al. [10] presents the feature subset based ensemble method to classifying the multiple cancers by utilizing the miRNA expression data in order to generate the multiple subsets the feature relevance and redundancy considers. The present techniques utilize the C4.5 decision tree algorithm and SVM algorithm for classification. The present techniques tested with the sequence based miRNA expression datasets and validated with the 10 fold and leave one out cross validations. The result shows that the present technique reaches higher prediction accuracy than the traditional ensemble techniques.

Bharathi, A., & Natarajan, A. M. et al. [11] proposed a simple yet very effective method which is used to cancer classification using the very few gene expression. The aim of the present technique is the finding the smallest gene subsets for accurate cancer classification from micro array data by using supervised machine learning algorithms (SVM). The present techniques involves in two phase such as chosen some important genes by using the 2 way Analysis of Variance (ANOVA) ranking scheme, then test with the SVM classifier it provides the good accuracy.

Thein, H. T. T., & Tun, K. M. M. et al [12] presents the analysis of feed forward neural network and the island differential evolution propagation algorithm utilized to train this network. The aim of the present technique is a creating the effective tool for construct the neural models which helps to proper classification of different classes of breast cancer. The present techniques proposed two different migration topologies such as random topology and torus topology. The performances are tested with Wisconsin Breast Cancer Diagnosis problem and the result shows that the random topology

provides good classification accuracy compare to torus topology.

Dora, L., et.al. [13] proposed the novel Gauss Newton Representation based Algorithm (GNRBA) for classification of breast cancer. It uses the sparse representation with feature selection and evaluates the sparsity in a computationally efficient way. Then the present technique proposed new gauss Newton based classifier to find optimal weights for training samples for classification. The present techniques are tested with Wisconsin breast cancer database and Wisconsin Diagnosis breast cancer database from the UCI machine learning repository. The result shows that the present technique provides better accuracy, sensitivity, specificity, confusion matrices compare to traditional approaches.

Reis, S., et.al. [14] presents the analysis of categorize and automated classification of breast cancer by using the multi scale basic image features (BIF) and local Binary Patterns (LBP) combined with the random decision trees classifier used for the classification of breast cancer. The present techniques demonstrate the text based classification of Hematoxylin and Eosin (H&E) images from IBC. The result shows that the multi scale approach provides the good accuracy.

Kourou, K., et.al. [15] presents the recent Machine Learning (ML) approaches to predicting the cancer. The various predictive models are discussed based on ML techniques as well as various input features and data samples. The ML is the branch of artificial intelligence which is used to relate the problem of learning from the data samples in the concept of inference. The each learning process contains two phases. (i) Estimation of unknown dependencies in a system from the given dataset. (ii) Then the usage of the estimated dependencies to prediction the new outputs of the system. In this work the two main methods used such as supervised learning and unsupervised learning.

Krishnaiah, V., et.al. [16] presents the various data mining techniques in the several types of lung cancer datasets to enhance the lung cancer diagnosis. In this technique the most effective model to predict patients with lung cancer appears to be the Naive bayes which is used to follow the IF-THEN rules, decision trees and neural networks. The decision tree result is easier to read and interpret. The present techniques of predicting lung cancer can be further enhanced and expanded.

Kharya, S., et.al. [17] presents the several data mining techniques to diagnosis and prediction of breast cancer. The prediction of outcome of the disease is the one of the complex task to enhance the data mining applications. The usage of the computers with automated tools, the large volumes of the medical data are gathered and available within the medical research groups. The data mining techniques are popular research tool for medical researcher to predictions of the exploit patterns and related with large number of variables which is used to improve the prediction of disease using the historical datasets. The several data mining techniques are such as Decision trees, Digital Mammography classification using association rule mining and ANN, Association rule based classifier, neural

network based classifier system, Naive bayes classifier, support vector machine, logistic regression and Bayesian network. The result shows that the Bayesian network is perform well to predict out Breast Cancer and diagnosis. However the Bayesian networks requires large amount of probability data.

Chaurasia, V., & Pal, S., et.al.[20]presents to analysis the performances of several data mining techniques. The classification of Breast cancer data can be utilized to find out the result of some disease or discover the common nature of cancer disease. The several data mining techniques are used to analyse the cancer disease, the present novel approach used to find out the compare performances of decision tree classifier such as Sequential Minimal Optimization (SMO), K-Nearest Neighbor Classifier, andBest First Tree. The result shows that the performance of SMO provides good result compare to other classifier in terms of accuracy, low error rate and performance.

Agrawal, A., et.al. [23]presents to improve the prediction models for lung cancer using data mining techniques. In present techniques utilizes the ensemble voting of five decision tree based classifiers and Meta classifiers used to find out the lung cancer prediction in terms of accuracy and according to the ROC curve. Additionally the lung cancer outcome calculator was developed by using this present technique. The prediction quality of the technique is calculated by this calculator is very efficient to find out the lung cancer prediction.

3. DISCUSSION

The above survey provides the detailed description of classification of cancer using various data mining techniques as depicted in Table 1.

S. No	Author Name	Methods Used	Dataset Used	Merits	Demerits	Performance
1	Lavanya, D., & Rami, D. K. U. [18]	Decision tree classifier (CART)	Breast Cancer Datasets	Easy to generate rules	Need large amounts of memory to store the entire tree for deriving the rules	Accuracy=94.72%
2	Ramani, R. G., & Jacob, S. G. [19]	Hybrid feature selection	Gene Set Enrichment Analysis database (GSEA db)	Easy to use and improve accuracy	Complexity issue occur	Accuracy=87%

3	Jacob, S. G., &ramani, R. G. [21]	Random tree and Quinlan's C4.5 algorithm	Wisconsin Prognostic Breast Cancer (WPBC)	Classification accuracy improve	Limited utility for future enhancement	Accuracy=100%
4	Mishra, D., &Sahu, B. [22]	Multiple filter multiple wrapper approach (MFMW)	Leukemia Dataset	Easy to implement	Slow Execution and Lack of generality	Accuracy=100%
5	Shajahaan, S. S., Shanthi, S., &ManoChitra, V. [24]	Decision Tree	Breast cancer dataset	Classification accuracy improved and reduce problem complexity	Training time is relatively expensive	Accuracy=100%
6	Zheng, B., Yoon, S. W., & Lam, S. S. [25]	Hybrid of K-means and support vector machine algorithm	Breast Cancer Wisconsin (Original) Dataset	Reduce computational time	Not easy to interpret	Accuracy=97.38%
7	Salama, G. I., Abdelhalim, M., &Zeid, M. A. E. [26]	Multi Classifiers	(Wisconsin Breast Cancer (WBC), Wisconsin Diagnosis Breast Cancer (WDIBC))	Work well on both numeric and textual data	Computation time complexity occur	Accuracy =97.28%
8	Glaab, E., Bacardit, J., Garribaldi, J. M., &Krasnogor, N. [27]	Evolutionary machine learning technique	Microarray cancer datasets	Efficiently work with large scale dataset	Time consuming for training	Accuracy=96.6%

9	Yu, H., Ni, J., Dan, Y., & Xu, S. [28]	Skewed gene selection algorithm	Gene expression datasets	Handle both linear and non linear data	Decision rules is quite time consuming.	Accuracy=100% G-Mean=100%
10	Mandal, S. K. [29]	Logistic Regression classifier	Wisconsin Diagnosis Breast Cancer (WDBC) dataset	Reduce time complexity	Performance is based on number of outlier in data	Accuracy=97.90%
11	Salem, H., Atriya, G., & El-Fishawy, N. [30]	New Novel Approach based on gene expression profiles	Microarray gene expressions datasets	Improve the classification accuracy	Time complexity	Accuracy=100% Specificity =97.3% Sensitivity =99.78%

4. CONCLUSION

In this survey the several data mining techniques have been discussed in classification of cancer disease prediction. The several data mining techniques such as Artificial Neural Network (ANN), Ensemble gene selection methods, pattern recognition, Learning Hidden Markov Models, random projection (RP), Ensemble Method, SVM classifier, Random Topology, Novel Gauss Newton Representation, Machine Learning, Decision tree, Sequential Minimal Optimization, Multiple filter multiple wrapper approach and Skewed gene selection algorithm etc used in the literatures and these methods have both merits and demerits. According to the biomedical domain the information gain and genetic algorithm methodology have efficiently used for classification of cancer by using gene expression data. Initially the information gain is used for selects the significant features from the input patterns. Then the selected features are reduced by using the genetic algorithm (GA). The genetic algorithm has several major advantages such as does not need any mathematical requirements, the ergodicity of evolution operators makes GA very effective at performing the global search and the GA provides the great flexibility to hybridize with domain depend heuristics to make the efficient implementation for the specific problems. Then the information gain also has several advantages like it is used to reduce a bias towards multi valued attributes by taking the number of attributes with a large number of distinct values. Finally the gene expression profiles are utilizing to classify the human cancer disease chosen to improve the prediction of cancer classification. Further the research work can be extended to implement the hybrid or new classification algorithm to

classify gene expression dataset for better accuracy and prediction.

References

[1]Reena, G. (2011),”A survey of human cancer classification using micro array data”, International Journal of Computer Technology and Applications, 2(5).

[2]Ismaeel, A. G., & Mikhail, D. Y. (2016),”Effective data mining technique for classification cancers via mutations in gene using neural network”, arXiv preprint arXiv:1608.02888.

[3]Nilashi, M., Ibrahim, O., Ahmadi, H., &Shahmoradi, L. (2017), “A knowledge-based system for breast cancer classification using fuzzy logic method”, Telematics and Informatics, 34(4), 133-144.

[4]Mahapatra, R., Majhi, B., & Rout, M. (2012),”Reduced feature based efficient cancer classification using single layer neural network”, Procedia Technology, 6, 180-187.

[5]Elyasigomari, V., Lee, D. A., Screen, H. R., & Shaheed, M. H. (2017),” Development of a two-stage gene selection method that incorporates a novel hybrid approach using the cuckoo optimization algorithm and harmony search for cancer classification”, Journal of Biomedical Informatics, 67, 11-20.

[6]Liu, H., Liu, L., & Zhang, H. (2010), “Ensemble gene selection for cancer classification”, Pattern Recognition, 43(8), 2763-2772.

[7]Jeleń, Ł., Krzyżak, A., Fevens, T., &Jeleń, M. (2016), ”Influence of feature set reduction on breast cancer malignancy classification of fine needle aspiration biopsies”, Computers in biology and medicine, 79, 80-91.

[8]Nguyen, T., Khosravi, A., Creighton, D., &Nahavandi, S. (2015), “Hidden Markov models for cancer classification using gene expression profiles”, Information Sciences, 316, 293-307.

[9]Xie, H., Li, J., Zhang, Q., & Wang, Y. (2016), “Comparison among dimensionality reduction techniques based on Random Projection for cancer classification”, Computational biology and chemistry, 65, 165-172.

[10]Piao, Y.,Piao, M., &Ryu, K. H. (2017), “Multiclass cancer classification using a feature subset-based ensemble from microRNA expression profiles”, Computers in biology and medicine, 80, 39-44.

[11]Bharathi, A., & Natarajan, A. M. (2010), “Cancer Classification of Bioinformatics data using ANOVA”, International journal of computer theory and engineering, 2(3), 369.

[12]Thein, H. T. T., &Tun, K. M. M. (2015), “An approach for breast cancer diagnosis classification using neural network”, Advanced Computing, 6(1), 1.

[13]Dora, L., Agrawal, S., Panda, R., & Abraham, A. (2017), “Optimal breast cancer classification using Gauss–Newton representation based algorithm”, Expert Systems with Applications, 85(1), 134-145.

[14]Reis, S., Gazinska, P., Hipwell, J., Mertzaniidou, T., Naidoo, K., Williams, N., ...& Hawkes, D. J. (2017),

- “Automated Classification of Breast Cancer Stroma Maturity from Histological Images”, IEEE Transactions on Biomedical Engineering.
- [15] Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., & Fotiadis, D. I. (2015), “Machine learning applications in cancer prognosis and prediction”, Computational and structural biotechnology journal, 13, 8-17.
- [16] Krishnaiah, V., Narsimha, D. G., & Chandra, D. N. S. (2013), “Diagnosis of lung cancer prediction system using data mining classification techniques”, International Journal of Computer Science and Information Technologies, 4(1), 39-45.
- [17] Kharya, S. (2012), “Using data mining techniques for diagnosis and prognosis of cancer disease”, arXiv preprint arXiv:1205.1923.
- [18] Lavanya, D., & Rani, D. K. U. (2011), “Analysis of feature selection with classification: Breast cancer datasets”, Indian Journal of Computer Science and Engineering (IJCSE), 2(5), 756-763.
- [19] Ramani, R. G., & Jacob, S. G. (2013), “Improved classification of lung cancer tumors based on structural and physicochemical properties of proteins using data mining models”, PloS one, 8(3), e58772.
- [20] Chaurasia, V., & Pal, S. (2014), “A novel approach for breast cancer detection using data mining techniques”, International Journal of Innovative Research in Computer and Communication Engineering, 2(1), 2456-65.
- [21] Jacob, S. G., & Ramani, R. G. (2012, October), “Efficient classifier for classification of prognostic breast cancer data through data mining techniques”, In Proceedings of the World Congress on Engineering and Computer Science (Vol. 1, pp. 24-26).
- [22] Mishra, D., & Sahu, B. (2011), “Feature selection for cancer classification: a signal-to-noise ratio approach”, International Journal of Scientific & Engineering Research, 2(4), 1-7.
- [23] Agrawal, A., Misra, S., Narayanan, R., Polepeddi, L., & Choudhary, A. (2011, August), “A lung cancer outcome calculator using ensemble data mining on SEER data”, In Proceedings of the Tenth International Workshop on Data Mining in Bioinformatics (p. 5). ACM.
- [24] Shajahaan, S. S., Shanthi, S., & ManoChitra, V. (2013), “Application of Data Mining techniques to model breast cancer data”, International Journal of Emerging Technology and Advanced Engineering, 3(11), 362-369.
- [25] Zheng, B., Yoon, S. W., & Lam, S. S. (2014), “Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms”, Expert Systems with Applications, 41(4), 1476-1482.
- [26] Salama, G. I., Abdelhalim, M., & Zeid, M. A. E. (2012), “Breast cancer diagnosis on three different datasets using multi-classifiers”, Breast Cancer (WDBC), 32(569), 2.
- [27] Glaab, E., Bacardit, J., Garibaldi, J. M., & Krasnogor, N. (2012), “Using rule-based machine learning for candidate disease gene prioritization and sample classification of cancer gene expression data”, PloS one, 7(7), e39932.
- [28] Yu, H., Ni, J., Dan, Y., & Xu, S. (2012), “Mining and integrating reliable decision rules for imbalanced cancer gene expression data sets”, Tsinghua Science and Technology, 17(6), 666-673.
- [29] Mandal, S. K. (2017), “Performance Analysis of Data Mining Algorithms For Breast Cancer Cell Detection Using Naïve Bayes, Logistic Regression and Decision Tree”, International Journal Of Engineering And Computer Science, 6(2), 20388-20391.
- [30] Salem, H., Attiya, G., & El-Fishawy, N. (2017), “Classification of human cancer diseases by gene expression profiles”, Applied Soft Computing, 50, 124-134.