ISSN 1746-7659, England, UK
Journal of Information and Computing Science
Vol. 13, No. 3, 2018, *pp*.168-178

Fuzzy Discretization and Rough Set based Feature Selection for High-Dimensional Classification

Prema Ramasamy [1], Premalatha Kandhasamy [2]

[1] *Prema Ramasamy, Assistant Professor, New Horizon College of Engineering, Bangalore*
E-mail:premabit@gmail.com
[2] *Professor, Department of Computer Science and Engineering, Bannari Amman Institute of Techlology, Sathyamangalam.*

(Received May 11 2018, accepted July 16 2018)

Abstract. Contemporary biological technologies like gene expression microarrays produce extremely high-dimensional datasets with limited samples. Analysis of gene expression data is essential in microarray gene expression studies in order to retrieve the required information. Gene expression data generally contain a large number of genes but a small number of samples. The complicated relations among the different genes make analysis more difficult, and removing irrelevant genes improves the quality of results. In this regard, a new feature selection algorithm called 2-level MRMS is presented based on rough set theory. It selects a set of genes from microarray data by maximizing the relevance and significance of the selected genes. The paper also presents a novel discretization method, Gaussian Fuzzy Discretization based on fuzzy logic to discretize the continuous gene expression values. The performance of the proposed algorithm, along with a comparison with other related feature selection methods, is studied using the classification accuracy of k-Nearest Neighbor (kNN) and Support Vector Machine (SVM) on four microarray data sets. The experimental results show that the genes selected using 2-level MRMS feature selection give high classification accuracy than other methods.

Keywords: classification, feature selection, fuzzy discretization, high- dimensional data, maximum relevance and maximum significance, microarray data

1. Introduction

A microarray dataset [1] is a repository containing microarray gene expression data. The raw microarray data are images that are transformed into gene expression data matrices where rows represent genes, columns represent various samples such as tissues or experimental conditions and the numbers in each cell characterize the expression level of a particular gene in a particular sample. Figure 1 shows an example of an $M \times N$ gene expression dataset where M is the number of genes, and N is the number of samples.

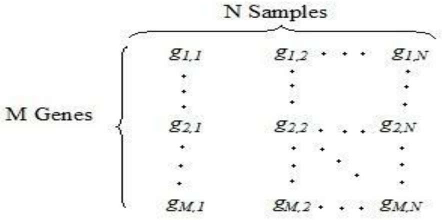


Fig. 1: Example of gene expression data

Data dimensionality reduction is one of the important machine learning tasks while dealing high-dimensional data with enormity on size, missing values and noise [2]. Gene expression dataset contains thousands of gene expression values, many of which may be irrelevant or redundant for classification [3]. Leaving out relevant attributes or keeping irrelevant attributes may affect the performance of the classification algorithm. Therefore statistical methods are required to identify a reduced search space are commonly used for classification [4]. There are many feature selection approaches to assist in classification of samples [5-9]. They are classified into four categories, namely filter approach, wrapper approach, embedded approach, and hybrid approach. A filter approach applies a statistical measure to assign a score to each feature without using a learning algorithm [10]. A wrapper approach uses learning techniques to evaluate the accuracy produced by the use of the selected features in the classification [11]. An embedded approach combines the feature selection step and classifier construction. A hybrid approach is a combination of both filter and wrapper-based methods [12]. In this paper, Rough Set Theory (RST) based feature selection method is used.

The rough set theory has been applied successfully to feature selection of discrete valued data [13,14,15]. It reduces the number of features of a dataset without considering any prior knowledge and using only the information contained within the dataset [16]. In this paper, 2-level MRMS feature selection method is proposed to select a set of genes from gene expression datasets by considering both relevance and significance of the selected genes. To compute the relevance and significance, the equivalence partitions of each gene is used. This can be automatically derived from the given datasets. So, RST needs no information other than the data set itself.

The RST feature selection process can only operate effectively with datasets containing discrete values. As gene expression datasets contain continuous value attributes, it is necessary to perform a discretization technique before gene selection. This paper presents a new discretization method, Gaussian Fuzzy Discretization (GFD) to discretize the continuous gene expression values. The discretization of numerical attributes can be performed before or after normalization [l7]. In this paer, the datasets are normalized using fuzzy logic. Then the normalized dataset can be discretized using mean and standard deviation.

The GF-discretized datasets are given as input to the feature selection methods. The rest of the paper is organized as follows. Section 2 discusses the related work in brief. Section 3 introduces the basic concepts of rough sets. The GFD process and the proposed feature selection method is explained in Section 4 for selecting relevant and significant genes.

2. Related Work

Hu et al. [18] proposed a feature subset selection technique based on a fuzzy-rough model. They used a symmetric function to compute fuzzy similarity relations between the objects with a numerical attribute and transform the similarity relation into a fuzzy equivalence one. So, this approach does not require discretizing the numerical data. Also they defined four attribute significance measures. Based on the measures, they constructed a forward hybrid attribute selection algorithm. Jenson and Shen [19] examined a novel approach based on fuzzy-rough sets, called fuzzy-rough feature selection. It overcomes the problems of noisy and real-valued data, as well as handling mixtures of nominal and continuous value attributes. FRFS achieves this by the use of fuzzy-rough sets, and a new measure of attribute significance, the fuzzy-rough degree of dependency. It also deals with real-valued decision attributes.

Yao and Zhao [20] discussed attribute reduction in decision-theoretic rough set models regarding different classification properties, such as: decision monotocity, confidence, coverage, generality and cost. Cornelis et al. [21] introduced a framework for fuzzy-rough set based feature selection. They provided a comprehensive typology of subset evaluation measures that can be used to define fuzzy decision reducts. Zhang et al. [22] studied the attribute reduction based on a discernibility matrix and used it to design correspondence attribute reduction algorithm. A simplified decision table was first introduced and then, a new measure of the significance of an attribute was defined for reducing the search space of the simplified decision table. Zahra and Reza [23] proposed a new fuzzy 2-level complementary system for classification of gene expression data. This approach exploits complementary learning and hierarchical organization, and complexity reduction and good interpretability are achieved.

3. Rough Sets

Let $I = (U, A \cup D)$ be an information system [24], where U is a non-empty set of finite objects (the universe) and A is a non-empty finite set of attributes and D is the set of decision attributes. This information system can be called as decision table. $\forall a \in A$, there exists a corresponding function $f_a: U \to V_a$, where V_a is the set of values that attribute a take. If $P \subseteq A$, there is an associated equivalence relation [24]:

$$IND(P) = \{(x, y) \in U^2 \,\big|\, \forall a \in P, f_a(x) = f_a(y)\} \tag{1}$$

The partition of U, generated by $IND(P)$ is denoted U/P. If $(x, y) \in IND(P)$ then x and y are indiscernible to P. The equivalence classes of the P-indiscernibility relation are denoted as $[x]_P$. Let $X \subseteq U$, the P-lower approximation $\underline{P}X$ and P-upper approximation $\overline{P}X$of set X can be defined as [24]:

$$\underline{P}X = \{x \in U | [x]_P \subseteq X\} \tag{2}$$

$$\overline{P}X = \{x \in U | [x]_P \cap X \neq \emptyset\} \tag{3}$$

The objects in $\underline{P}X$ can be with certainty classified as members of X on the basis of knowledge in P, while the objects in $\overline{P}X$ can be only classified as possible members of X on the basis of knowledge in P.

Let $P \subseteq A$ and D is the decision attribute (class label) then the positive region can be defined as [24]:

$$POS_P(D) = \bigcup_{X \in U/D} \underline{P}X \tag{4}$$

Positive region $POS_P(D)$ is the set of all objects of U that can be certainly classified to blocks of the partition U/D by means of P. Then the dependency degree $\gamma_P(D)$ can be calculated as [24]:

$$k = \gamma_P(D) = \frac{|POS_P(D)|}{|U|} \tag{5}$$

If $k = 1$, D depends totally on P if $0 < k < 1$, D depends partially on P, and if $k = 0$ then D does not depend on P.

4. The Proposed works

4.1. Gaussian fuzzy discretization

A fuzzy set A of a non-empty set X is defined as $< x, \mu_A(x) >$ where $x \in X$ and $(\mu_A(x))$ is the membership function of the fuzzy set A. A fuzzy set is a collection of objects with graded membership i.e. having degrees of membership [25].

In this paper, datasets are transformed by exploitation of a fuzzy membership function. A membership function is a curve that defines how each point in the input space is mapped to a membership value between 0 and 1. A fuzzy membership function that is used to represent vague, linguistic terms is the Gaussian which is given in Equation (6).

$$\mu_A(x) = \exp(-\frac{(x-m)^2}{2(k)^2}) \tag{6}$$

Where m is the centre and k is the width of the fuzzy set A.

Here, all the sample values for each feature are considered as a set. To find the membership function of this non-empty set, each feature values with respect to all the samples are fuzzified into three fuzzy qualifiers, small, medium and large. By applying the Gaussian membership function (Equation (7)), each feature values are normalized to a scale of 0 to 1, where 1 is the highest expression level and 0 is the lowest. Figure 2 shows the membership values of four random features.

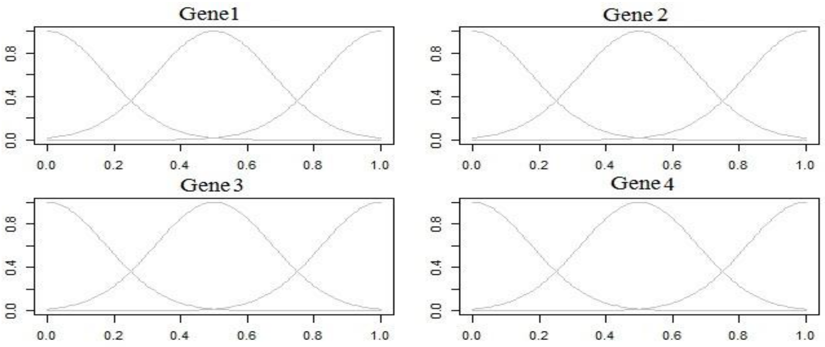


Fig. 2: Membership functions of four random features.

The normalized dataset can be discretized using mean μ and standard deviation σ computed over n values of that gene [1]. Any value larger than $(\mu + \sigma/2)$ is transformed to state 1; any value between $(\mu - \sigma/2)$ and $(\mu + \sigma/2)$ is transformed to state 0; any value smaller than $(\mu - \sigma/2)$ is transformed to -1.

4.2. 2-level maximum relevance-maximum significance

In high-dimensional data analysis such as microarray data, the dataset contains a number of insignificant features. The presence of such irrelevant and insignificant features may affect the performance of machine learning algorithms. So, the selected features should have high relevance with the classes and high significance in the feature set. The features with high relevance are expected to be able to predict the classes of the samples. However, if insignificant features are present in the feature subset, they may reduce the classification accuracy. A feature set with high relevance and high significance enhances the predictive capability. In this paper, the rough set theory is used to select the relevant and significant features or genes from high-dimensional gene expression data sets.

Let $C = \{A_1, A_2, \ldots A_i \ldots, A_j, \ldots, A_m\}$ denotes the set of m features of a given high-dimensional microarray dataset and K is the set of selected genes. The proposed algorithm maximizes dependency between a feature subset and a class label and also it maximizes the significance among the selected features. This paper uses the value of dependency degree as the relevance of the corresponding attributes. To what extent an attribute is contributing to calculate the dependency on decision attribute can be calculated by the significance of that attribute. The change in dependency when an attribute is removed from the set of condition attributes is a measure of the significance of the attribute. The higher the change in dependency, the more significant the attribute is. The significance of attribute A can be calculated as,

$$\sigma_C(D, A) = \gamma_C(D) - \gamma_{C-\{A\}}(D) \tag{7}$$

If the significance is 0, then the attribute is dispensable. In the first level of proposed algorithm, the top-100 features are selected which have high relevance with the classes. In the second level, the significance of those top-100 features is calculated. Then the top-25 features are selected which have high significance in the feature set. The 2-level MRMS algorithm is explained as follows.

Input: Set of conditional attributes C, Decision attribute (class label) D

Output: The top-25 features.

```
(i)    Initialize $C \leftarrow \{A_1, A_2, \ldots A_i \ldots, A_j, \ldots, A_m\}$, S← $\emptyset$
(ii)   Calculate the relevance $\gamma_C(D)$ of each feature $A_i \in C$
(iii)  Select the feature $A_i$ as the most relevant feature that has the highest relevance value with
       decision attribute.
(iv)   S← $A_i$, $C \leftarrow C - A_i$
(v)    Combine $S$ with each feature in $C$ and calculate the relevance of these attributes. Choose the
       highest. Add in $S$.
(vi)   Repeat this step to find the top-100 features.
(vii)  Find the significance of all the 100 features in $S$.
(viii) Then select the top-25 features with maximum significance.
```

5. Results

The proposed feature selection algorithm with GFD is examined in four different high-dimensional gene expression datasets. These datasets are publically available at Gene Expression Omnibus (GEO) website (https://www.ncbi.nlm.nih.gov/geo/) and details are given in Table 1.

In this paper, the experimental comparison of the feature selection method with several well-regarded feature selection methods in terms of classification accuracy is provided. The 2-level MRMS algorithm selects a subset of top-25 relevant features from all the datasets. To evaluate the performance of the proposed algorithm, the well known classifiers SVM and kNN are employed. The selected genes are utilized for training the classifiers. The performance is evaluated using 10-fold cross validation. The radial basis kernel function is used for SVM classifier. The number of instance considered for determination of similarity with classes as three for kNN. The performance of the proposed method is compared with feature selection methods, chi-square, information-gain, gain-ratio, reliefF. To demonstrate the performance of the algorithm, the top 5, 10, 15, 20 and 25 genes are selected as features for classification.

Table 1: Description of the datasets

Dataset	GEO accession id	No. of samples	No. of genes	No. of classes
Breast	GSE9574	29	22283	2
Ovarian	GSE12470	24	54675	2
Prostate	GSE32269	55	18288	2
Autism	GSE25507	146	54613	2

Table 2: The classification accuracy using SVM classifier

Classifiers		SVM				
Datasets	No. of features	chi-square	information-gain	gain-ratio	reliefF	Proposed algorithm
Breast	5	70.81	73.60	72.19	74.00	74.00
	10	79.00	77.60	79.19	80.80	82.57
	15	84.87	84.61	85.33	87.12	87.12
	20	88.34	90.00	86.63	89.19	92.70
	25	92.17	93.50	93.50	94.87	94.87
Ovarian	5	72.57	71.28	72.00	72.00	72.57
	10	78.83	78.41	79.17	79.38	79.38
	15	82.18	82.71	81.11	80.32	82.34
	20	85.67	85.00	85.40	86.69	87.77
	25	87.70	86.40	86.40	86.40	88.73
Prostate	5	76.20	77.00	77.10	74.17	78.83
	10	77.71	80.25	80.78	78.13	80.81
	15	85.34	86.63	85.71	85.71	86.63
	20	87.77	86.63	89.37	90.70	92.57
	25	89.80	88.50	89.11	92.20	93.00
Autism	5	80.33	79.83	78.65	77.14	80.72
	10	85.17	86.12	86.83	85.57	87.52
	15	92.35	91.67	89.80	91.80	93.15
	20	98.60	97.20	96.97	97.81	98.19
	25	98.60	97.20	97.20	98.80	100

Table 3: The classification accuracy using kNN classifier

Classifiers	kNN					
Datasets	No. of features	chi-square	information-gain	gain-ratio	reliefF	Proposed algorithm
Breast	5	71.51	74.52	72.25	74.00	75.14
	10	79.83	79.16	79.57	81.00	82.87
	15	85.72	86.68	87.30	88.70	88.70
	20	89.37	90.78	88.54	90.16	93.40
	25	93.00	93.50	94.00	94.15	95.85
Ovarian	5	72.20	71.16	70.78	70.00	72.00
	10	77.20	75.25	78.01	77.13	78.87
	15	80.17	80.00	80.77	79.06	80.66
	20	84.23	85.80	84.18	86.20	86.16
	25	87.23	85.62	86.15	86.20	87.92
Prostate	5	64.30	66.50	66.50	67.00	65.77
	10	70.13	72.82	70.82	72.75	72.89
	15	77.83	82.17	80.13	82.00	83.71
	20	84.27	86.92	85.73	88.25	90.23
	25	85.57	86.92	86.92	90.18	90.60
Autism	5	77.77	76.39	75.14	76.62	79.97
	10	84.58	84.58	85.14	84.38	85.18
	15	90.67	90.85	88.67	90.35	91.81
	20	95.20	95.85	95.85	97.81	98.80
	25	97.77	97.45	96.88	98.80	98.80

As a general conclusion, for all the datasets, the accuracy performance is improved by using 2-level MRMS feature selection mostly in all cases. Also, SVM performs better than kNN due to its suitability for high dimensional data.

Furthermore, the results in this study are validated in Tables 4 and 5 by conducting a statistical paired samples one-tailed test. This statistical test is used to verify whether there is any significant difference in the accuracy and the number of selected genes by using a significance interval of 95% (a = 0.05). The results of this study show that most of the p-values obtained are less than 0.05. This means that there is a significant difference in the accuracy and the number of selected genes in the method used in this study as compared to the other methods on all datasets, respectively.

Table 4: p-Values between proposed and other methods about accuracy with SVM

Methods	p-Values			
	Breast	Ovarian	Prostate	Autism
Proposed vs. chi-square	0.000	0.028	0.019	0.077
Proposed vs. information-gain	0.008	0.036	0.059	0.015
Proposed vs. gain-ratio	0.015	0.024	0.022	0.008
Proposed vs. reliefF	0.052	0.011	0.180	0.096

Table 5: p-Values between proposed and other methods about accuracy with kNN

Methods	p-Values			
	Breast	Ovarian	Prostate	Autism
Proposed vs. chi-square	0.001	0.013	0.003	0.003
Proposed vs. information-gain	0.010	0.058	0.044	0.009
Proposed vs. gain-ratio	0.013	0.051	0.025	0.033
Proposed vs. reliefF	0.012	0.027	0.018	0.021

In order to understand whether the proposed algorithm is able to extract interactions with a biological meaning, the differential gene subset selected by this method are analyzed by conducting the gene set enrichment analysis on the DAVID tool (Database for Annotation, Visualization, and Integrated Discovery). DAVID is able to provide a comprehensive set of functional annotation tools for investigators to understand the biological meaning behind a large list of genes [26]. The top-25 genes selected by using the proposed method are supplied into the DAVID website (https://david.ncifcrf.gov/home.jsp). The Functional Annotation Tool is utilized to achieve the Functional Annotation Clustering results (the Classification Stringency is set to High). The group Enrichment Score (ES) and the geometric mean of the member's p-values in a corresponding annotation cluster, is used to rank their biological significance. Thus, the top ranked annotation clusters most likely have consistently lower p-values for their annotation members. The larger the enrichment score, the more enriched is the gene subset. In this study, the first value from the top annotation cluster having the largest ES, and the concerned terms having similar biological meanings are presented in Table 6. It is seen that the genes selected by 2-level MRMS are related to genes that belong to the cancer-related Gene Ontology (GO) terms).

To further verify the proposed method's effectiveness, the biological annotations of top genes for each dataset are identified. Table 7 lists the significant shared GO terms or parent of GO terms used to describe the top-25 genes in each dataset for the process, function and component ontologies. For example, for the Breast dataset, the genes are mainly involved in cell differentiation, nucleoplasm and translation initiation factor activity.

The genes selected by the proposed feature selection algorithm using GF-discretized data are analyzed to identify the biological significance of them. Table 8 shows the genes related to the diseases. NR4A3 shows the early induction of the orphan nuclear receptor during cell death of the human breast cancer cell line [27]. The expression level of JUN modulates regulation of activator protein activity by estradiol in breast cancer cells [28]. Increased MUC1 immunoreactivity was observed in adencarcinomas of the breast, pancreas and ovary [29]. SPON1 was originally isolated from bovine ovarian follicular fluid as a stimulator of vascular smooth muscle cell proliferation [30]. KLK3 is a secreted protein that is widely used as a diagnostic marker for prostate cancer [31]. GFI1 plays a significant role in the down regulation of endogenous production of 1,25D in prostate cancer cells [32]. LXN prediction of personality disorder is susceptible to state effects of depression [33]. CD36 is over expressed in human brain that correlates with beta-amyloid deposition [34].

Table 6: The enrichment analysis results about annotation cluster by DAVID

Dataset	Annotation cluster	Enrichment score
Breast	GO:0000982~transcription factor activity, GO:0032870~cellular response to hormone stimulus, GO:0042493~response to drug, GO:0032570~response to progesterone	2.69
Ovarian	GO:0005615~extracellular space GO:0016021~integral component of membrane	1.23
Prostate	GO:0070062~extracellular exosome GO:0005887~integral component of plasma membrane GO:0005576~extracellular region	1.49
Autism	GO:0005886~plasma membrane GO:0005654~nucleoplasm GO:0016021~integral component of membrane GO:0003677~DNA binding	1.38

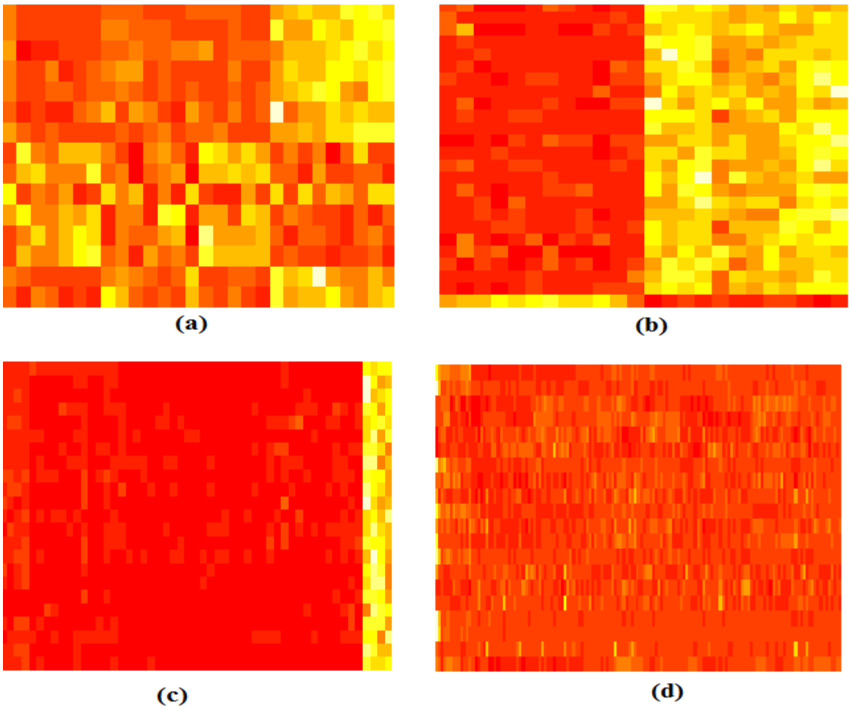
Table 7. Significant GO terms

Dataset	Biological Process	Cellular Component	Molecular Function
Breast	skeletal muscle cell differentiation	nucleoplasm	translation initiation factor
Ovarian	cell morphogenesis	extracellular space	flavin adenine dinucleotide binding
Prostate	immune response	extracellular region	carbohydrate binding
Autism	antibacterial humoral response	proteinaceous extracellular matrix	protein serine/threonine/tyrosine kinase activity

A heat-map is a two-dimensional representation of data in which values are represented by colors. Heat-maps originate in 2D displays of the values in a data matrix. Larger values are represented by small dark squares (pixels) and smaller values by lighter squares. Each row shows the expression levels of one selected feature, and each column is a sample. Figure 3 shows the heat-maps depicting the predictive performance of the genes selected by the proposed feature selection algorithm. From the Figure 3(a), it can be observed that there is a visible border between two classes of the Breast dataset. Figure 3(b) depicts a cut between two classes of the Ovarian dataset. The good performance of the selected genes for the Prostate dataset is also shown in Figures 3(c).

Table 8. Genes related to diseases

Dataset	Probe Id	Gene name
Breast	216979_at	NR4A3
	213281_at	JUN
	209909_s_at	TGFB2
	214056_at	MCL1
Ovarian	201010_s_at	TXNIP
	213693_s_at	MUC1
	206458_s_at	SPON1
	230943_at	SOX17
	202037_s_at	SFRP1
Prostate	204582_s_at	KLK3
	206589_at	GFI1
	207008_at	CXCR2
	211163_s_at	TNFRSF10C
Autism	218729_at	LXN
	221663_x_at	HRH3
	230100_x_at	PAK1
	209554_at	CD36



DOI https://doi.org/2024-JICS-22442 | Generated on 2025-04-09 07:43:31

Fig. 3: Heat-map (a) Breast; (b) Ovarian; (c) Prostate; (d) Autism.

6. Conclusion

This paper provides information on the performance of different feature selection techniques for microarray datasets. In this work, a new discretization method, GFD is introduced to discretize the continuous gene expression datasets. Also, 2-level MRMS feature selection is developed based on rough set theory. It identifies the subset of significant genes by maximizing the relevance and significance of the selected genes from four microarray datasets. The performance of the proposed algorithm is studied with respect to two important criteria. First, it compares the performance of the proposed method and some existing methods using the predictive accuracy of k-NN and SVM classifiers. The experimental results demonstrate that the classification accuracy is improved with the top-25 genes selected with the proposed algorithm. Secondly, the gene subset selected by the proposed method is analyzed by conducting the gene set enrichment analysis on the DAVID tool. The genes selected by 2-level MRMS are related to genes that belong to the similar Gene Ontology terms. The heat-maps are also visualized. The results illustrate that the proposed feature selection algorithm can be used to increase the quality of high-dimensional gene selection.

7. References

[1] Rinaldis ED, Lahm A. (2007). DNA Microarrays: Current Applications. Norfolk, UK: Horizon Bioscience.

[2] Pena, J.M., Lozano, J.A., Larranaga, P. and Inza, I. (2001). Dimensionality reduction in unsupervised learning of conditional Gaussian networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6), 590–603.

[3] Horng, J.T., Wu, L.C., Liu, B.J., Kuo, J.L., Kuo, W.H., and Zhang, J.J. (2009). An expert system to classify microarray gene expression data using gene selection by decision tree. *Expert Systems with Applications*, 36(5), 9072–9081.

[4] Arauzo Azofra, A., Aznarte, J. L., and Benítez, J. M. (2011). Empirical study of feature selection methods based on individual feature evaluation for classification problems. *Expert Systems with Applications*, 38(7), 8170–8177.

[5] Bhattacharyya, D.K, Kalita, J.K. (2013). Network Anomaly Detection: A Machine Learning Perspective. 1st ed. Boca Raton, FL, USA: CRC Press.

[6] Hoque, N, Bhattacharyya, D.K, Kalita J.K. (2014). MIFS-ND: A mutual information – based feature selection method. *Expert Systems with Applications*, 41(14), 6371–6385.

[7] Min N, Hu Q, Zhu W. (2014). Feature selection with test cost constraint. *International Journal of Approximate Reasoning*, 55(1): 167-179.

[8] Tabakhi, S, Moradi, P, Akhlaghian, F. (2014). An unsupervised feature selection algorithm based on ant colony optimization. *Engineering Applications of Artificial Intelligence*, 32: 112-123

[9] Jenson, R., Shen, Q. (2009). New Approaches to fuzzy-rough feature selection, *IEEE Transactions on Fuzzy Systems*, 17(4), 824-838.

[10] Guyon, I, Elisseeff, A. (2003). An introduction to variable and feature selection. Journal of Machine Learning Research, 3, 1157-1182.

[11] Blum, A.L, Langley, P. (1997). Selection of relevant features and examples in machine learning. Artificial Intelligence, 97(1), 245–271.

[12] Hsu, H.H, Hsieh, C.W, Lu, M.D. (2011). Hybrid feature selection by combining filters and wrappers. Expert Systems with Applications, 38(7), 8144–8150.

[13] Jensen, R, Shen, Q. (2004). Fuzzy-rough attribute reduction with application to web categorization. Fuzzy Sets and Systems, 141(3), 469–485.

[14] Wu, W, Zhang, W. (2004). Constructive and axiomatic approaches of fuzzy approximation operators. Information Sciences, 159(3), 233–254.

[15] Jensen, R. and Shen, Q. (2004). Semantics-preserving dimensionality reduction: rough and fuzzy-rough-based approach. IEEE Transactions on Knowledge and Data Engineering, 16 (12), 1457–1471.

[16] Yumin, C., Duoqian, M. and Ruizhi, W. (2010). A Rough Set approach to feature selection based on ant colony optimization. Pattern Recognition Letters, 31(3), 226-233.

[17] Han, J., Rodriguez J.C., Beheshti M. (2010). Discovering decision tree based diabetes prediction model. In Proceedings of the International Conference on ASEA: Advanced Software Engineering and Its Applications. (pp.99-109).

[18] Hu, Q., Xie, Z. and Yu, D. (2007). Hybrid attribute reduction based on a novel fuzzy-rough model and information granulation, Pattern recognition, 40(12), 3509- 3521.

[19] Jensen, R. and Shen, Q. (2007). Fuzzy-rough sets assisted attribute selection, Transactions on Fuzzy Systems, 15(1), 73-89.

[20] Yao, Y. and Zhao, Y. (2008). Attribute reduction in decision theoretic rough set models, Information Sciences, 178(17), 3356-3373.

[21] Cornelis, C., Jensen, R. and Hurtado, G. (2010). Attribute selection with fuzzy decision reducts. Information Sciences, 180(2), 209-224.

[22] Zhang, J., Dong, A., Niu, Y. and Nie, H. (2011). An Efficient Algorithm for Attribute Reduction Based on Discernibility Matrix. In Proceedings of the International Conference on Intelligent Computation and Bio-Medical Instrumentation. (pp. 175-178).

[23] Zahra, S. and Reza, G. (2012). Fuzzy-rough feature selection and a fuzzy 2-level complementary approach for classification of gene expression data. Scientific Research and Essays, 7(14), 1512-1520.

[24] Pawlak, Z. Rough Sets. (1991). Theoretical Aspects of Resoning About data. Dordrecht, The Netherlands: Kluwer.

[25] Zadeh, L.A. (1965). Fuzzy sets. Inform Control, 8: 338-353.

[26] Dennis, G. J. et al. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biology, 4.

[27] Ohkubo, T., Ohkura, N., Sasaki, K., Nagasaki, K., Hanzawa, H., Tsukada, T., Yamaguchi, K. (2000). Early induction of the orphan nuclear receptor NOR-1 during cell death of the human breast cancer cell line MCF-7. Molecular and. Cellular Endocrinology, 162(1), 151-156.

[28] Philips, A., Teyssier, C., Galtier, F., Rivier-Covas, C., Rey, J.M., Rochefort, H., Chalbos, D. (1998). FRA-1 expression level modulates regulation of activator protein-1 activity by estradiol in breast cancer cells. Molecular Endocrinology, 12(7), 973-985.

[29] Ho, SB., Niehans, G.A., Lyftogt, C., Yan, P.S., Cherwitz, D.L., Gum, E.T., Dahiya, R., Kim, Y.S. (1993). Heterogeneity of mucin gene expression in normal and neoplastic tissues. Cancer Research. 53(3), 641-651.

[30] Terai, Y., Abe, M., Miyamoto, K., Koike, M., Yamasaki, M., Ueda, M., Ueki, M., Sato, Y. (2001). Vascular smooth muscle cell growth-promoting factor/F-spondin inhibits angiogenesis via the blockade of integrin alphavbeta3 on vascular endothelial cells. Journal of Cellular Physiology, 188(3), 394-402.

[31] Xi, Z., Klokk, T.I., Korkmaz, K., Kurys, P., Elbi, C., Risberg, B., Danielsen, H., Loda, M., Saatcioglu, F. (2004). Kallikrein 4 is a predominantly nuclear protein and is overexpressed in prostate cancer. *Cancer Research*, 64(7), 2365-2370.

[32] Dwivedi, P.P., Anderson, P.H., Tilley, W.D., May, B.K., Morris, H.A. *J. (2007).* Role of oncoprotein Growth Factor Independent-1 (GFI1) in repression of 25-hydroxyvitamin D 1alpha-hydroxylase (CYP27B1): A comparative analysis in human prostate cancer and kidney cells. *Journal of Steroid Biochemistry and Molecular Biology, 103(3), 742-746.*

[33] Black, K.J., Sheline, Y.I. (1997). Personality disorder scores improve with effective pharmacotherapy of depression. 43(1), 11-18.

[34] Ricciarelli, R., D'Abramo, C., Zingg, J.M., Giliberto, L., Markesbery, W., Azzi, A., Marinari, U.M., Pronzato, M.A., Tabaton, M. (2004). CD36 overexpression in human brain correlates with beta-amyloid deposition but not with Alzheimer's disease. *Free Radical Biology and Medicine, 36(8), 1018-1024.*