Journal of Medical Systems (2019) 43: 208
https://doi.org/10.1007/s10916-019-1353-y

TRANSACTIONAL PROCESSING SYSTEMS

A Bio-inspired Algorithm based Multi-class Classification Scheme for Microarray Gene Data

LT. Thomas Scaria[1] · T. Christopher[2]

Received: 14 March 2019 / Accepted: 20 May 2019 / Published online: 29 May 2019

Abstract
Microarray gene data is widely known for its high dimensionality and volume. The utilization of microarray gene data is increasing now-a-days, owing to the advancement of medical science. Microarray gene data helps in diagnosing diseases quite accurately. However, processing microarray gene data is difficult and is usually not understandable. Taking this challenge into account, this work presents a user-friendly rule based classification model, which is easily understandable and does not demand users to have prior knowledge. The classification rules are formed with the help of cuckoo search optimization algorithm and the rules are pruned by the associative rule mining. Finally, the classification is performed with the help of the pruned rules. The performance of the proposed approach is satisfactory in terms of accuracy, sensitivity, specificity and time consumption.

Keywords Microarray gene · Disease analysis · Cancer classification · Association rules

Introduction

Microarray gene expression is widely utilized by the biological experts to track the gene expression level of a living being. Microarrays are utilized for assessing the gene expression by performing several different operations and one of the most common operations is the comparison between the genes of a cell, which are retained in different conditions. Hence, the microarray gene analysis has gained considerable research interest. However performing microarray gene analysis is not easier, as it involves voluminous data.

Usually, the process of microarray gene analysis intends to analyse the microarray data for gaining beneficial knowledge from the data. Data classification is one of the most common operations performed over data and can be attained by the classifiers. The classifier is trained with the sample data, such that it can differentiate between the entities with respect to the class. Though there are numerous classifiers to perform analysis by differentiating the entities, rule based classification is quite rare in the microarray data analysis field.

The rule based classification model requires a set of well formed rules, such that the healthcare professionals can diagnose the disease and the suitable treatment for the disease can be figured out. However, it is a difficult task to figure out the effective rules from a voluminous dataset with noise. In addition to this, the dimensionality of microarray data is quite larger and hence, there rule based classification technique may require greater number of rules to perform classification.

The bio-inspired algorithms prove better performance over the complex problems with greater search space. There are so many bio-inspired algorithms presented in the literature. Hence, it is better to utilize the bio-inspired algorithms to mine the rules being framed. Now-a-days, rule mining with the help of bio-inspired algorithms is becoming the new research trend and some of the samples are ABCMiner [1] and BeeMiner [2].

Recognizing the need of better classification technique and to present an aid to the healthcare professionals, this article intends to propose a rule based classification model based on Cuckoo Search Optimization Algorithm (CSOA). The main task of CSOA is to detect the perfect interval upon the rule based classification scheme to classify between the microarray gene data. The proposed work considers the microarray gene datasets with both binary and multi-classes of cancer. The

This article is part of the Topical Collection on *Transactional Processing Systems*

✉ LT. Thomas Scaria
lttscaria.19@gmail.com

[1] Department of Computer Science, St. Pius X College, Kasaragod, Kerala, India

[2] PG and Research Department of Information Technology, Government Arts College, Coimbatore, India

rules framed by the proposed approach are understandable and effective.

This article presents a web application to perform disease classification based on rules, which is beneficial for the healthcare professionals. This work aims to predict the disease by means of the association rules formed with respect to the association between the entities and the Cuckoo Search Optimization Algorithm (CSOA). The proposed approach classifies the microarray gene expression data.

This work considers different gene expression data with binary and multi-class datasets. The performance of this approach proves that this work shows better classification accuracy and hence, the proposed model can be utilized as an aid for biologists. Some of the work contributions are presented as follows.

- This work presents a rule based classification model, which is quite rare in the existing literature.
- The framed rules help in extracting better knowledge and train the classification model accordingly.
- The accuracy rate of the proposed approach is better, which is evident in the performance analysis.

The rest of the content is organized in the following way. Section 2 presents the review of literature with respect to microarray gene data analysis. The proposed rule based classification technique for microarray gene data is elaborated in section 3. The performance of the proposed approach is evaluated in section 4. Finally, section 5 concludes the article with the future possible enhancements.

Review of literature

This section reviews the related literature with respect to microarray gene data analysis.

A gene expression analysis for early lung cancer prediction technique is presented with the help of machine learning techniques in [3]. This work analyses the microarray gene expression data of the Kent Ridge bio-medical dataset to detect the lung cancer. The most optimal set of genes is selected from the microarray gene data to predict the cancer causing agent.

A microarray image gridding and segmentation technique is proposed in [4], which is executed on Graphics Processing Unit (GPU). This work intends to achieve better performance by utilizing the available resources in the Computed Unified Device Architecture (CUDA). It is claimed that the proposed approach shows better performance by consuming minimal time period.

A faster cDNA microarray gene expression data classifier is proposed for diagnosing the diseases in [5]. This work enhances the Gene Expression Graph based classifier for minimizing the computation time. This work filters the genes by means of the edge weight, in order to increase the classification accuracy and to reduce the false-positive rates.

In [6], a consensus gene selection criteria is proposed on the basis of distributed GPU with partial least-square based microarray data analysis is presented. This work measures the consistency and distinctiveness of gene expressions and the genes associated with the specific disease are figured out. The process of gene selection is accelerated as the work is implemented in distributed GPU. This work utilizes Diffused Large B Cell Lymphoma (DLBCL) and Prostrate cancer datasets.

A grouped gene selection of cancer based on adaptive sparse group lasso with respect to conditional mutual information is proposed in [7]. Initially, the genes are grouped by means of weighted gene co-expression network for the cancer microarray data. The conditional mutual information is utilized to compute the integrated and data driven weights. This paves way for enhancing the correlation between the genes in all the groups. By this way, the genes are classified and selected.

In [8], a microarray based cancer diagnostic system is proposed on the basis of repeated cross validation based ensemble feature selection. This work resamples the data by means of ensemble techniques and the features are selected by Repeated Cross Validation (RCV). The performance of this work is compared against Support Vector Machine (SVM) and recursive feature elimination techniques. The performance of this approach is evaluated in four different datasets.

A local nearest neighbour based feature weighting system is proposed for gene selection in [9]. This work considers the nearest neighbours based weighting approach by minimizing and maximizing the distances of within-class and between-class locally with respect to k nearest neighbour rule. This approach can be utilized in both binary and multi-class problems.

In [10], the feature selection and feature extraction techniques are combined together by utilizing deep learning in order to predict the outcome of the breast cancer. This work presents an unsupervised feature learning model by combining the Principal Component Analysis (PCA) and autoencoder neural network to detect the unique features of gene expressions. The AdaBoost algorithm based ensemble classifier is employed to predict the clinical outcomes of breast cancer.

The significance and functional similarity of gene identification with respect to the disease is proposed in [11]. This gene selection algorithm combines the information acquired from the protein interaction network and the gene expression profile. The significance of the gene is computed by comparing it with the other gene by utilizing mutual information. This work performs analysis on different cancer microarray datasets.

In [12], a system to predict the progression of cancer by employing gene interaction regularized elastic net is proposed. This work considers both the measurement and interaction information of the gene by establishing the elastic net. The discriminate features are chosen with the help of the grouping effect. This work is evaluated over ovarian and breast cancer datasets.

An auto-weighted least square method is presented for predicting the missing values in the microarray data in [13]. This work weights the neighbourhood genes with respect to a corresponding gene in terms of the gene significance. The convergence is enhanced with the help of an accelerating strategy. The missing values on the microarray data is predicted by this work as well.

In [14], a fast and scalable feature selection technique is proposed for gene expression data based on Hilbert-Schmidt independence criterion. The correlation between the gene expression data and the response variables are figured out by detecting the informative genes with the help of multivariate algorithm. This algorithm can be utilized on different response variables and is suitable of binary and multi-class classification.

A large-scale microarray data analysis meant for cloud-scale genomic signals is presented in [15]. This work is based on the cloud-scale distributed parallel based one dimensional wavelet based transformation for declaring a threshold. This idea retains the genes by means of denoising process to classify between the cancer types. However, this work processes the image based data.

In [16], an unsupervised gene selection technique is proposed on the basis of matrix factorization framework. This work introduces an unsupervised two-stage gene selection technique, in which the first stage clusters the genes by removing the redundant gene with the help of k-NN algorithm. The second stage selects the significant genes out of all the genes with the help of matrix factorization.

In [17], a gene retrieval system is presented on the basis of Local Fisher Discriminant Analysis (LFDA) and Support Vector Machine (SVM) is proposed. The LFDA is utilized for reducing the dimensionality of the microarray data and SVM is utilized for classification. A microarray data gene retrieval system based on Kernelized LFDA and Extreme Learning Machine (ELM) is proposed in [18]. In [19], an ensemble classification based technique is proposed on the basis of Information Gain Ratio (IGR) and classifiers k-NN, SVM and ELM.

Motivated by these existing works, this article intends to propose a gene classification system based on rule based classification model. The intention of this work is to make the entire process simpler with better understandability and is achieved by the meaningful rules.

Proposed approach

This section elaborates the proposed web application for disease prediction by forming rules. Initially, the overview of the work is presented as follows.

Work overview

Data classification is one of the widespread problems being faced by the data mining applications. It is a complex task to classify between the data, as the effectiveness of the classification system depends on the knowledge imparted to the classifier. Microarray data is quite voluminous and is pretty hard to impart knowledge to classifier. In addition to this, the microarray cancer data analysis is completely related to the human lives.

The objective of the microarray cancer data classification is to form a classification model, which enables the classifier to decide the best possible class of the entity. As this work handles the data related to human lives, the accuracy of the system is the main concern. Though there are numerous disease prediction and diagnostic systems available in the literature, the microarray data based classification techniques are scarce. Additionally, the classification techniques are hard to understand and in order to deal with this issue, this work attempts to propose a rule based classification scheme. The overall flow of the work is presented in Fig. 1.

The proposed approach deals with the voluminous microarray data for framing the rule set, such that the class associated with the entity can be figured out in the forthcoming process. There are several classification algorithms for microarray data, which are based on classifiers such as k-NN, SVM,

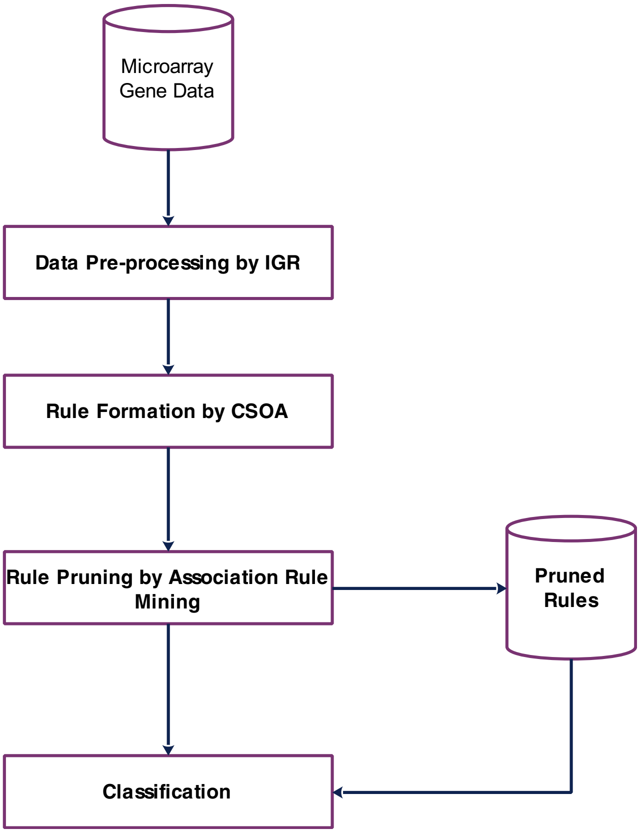

Fig. 1 Overview of the proposed approach

ELM, decision tree and so on. These classifiers perform better, however learning process of the classifiers depend on the statistical and mathematical ideas that can applied irrespective of the nature of domain. The classifiers are common for all sorts of classification problems and so the classifiers cannot derive problem specific knowledge with respect to the datasets.

Understanding the benefits of rule based classification techniques, this article presents a rule based classification scheme for microarray data. The proposed approach involves four significant phases and they are data pre-processing, rule formation, rule pruning and classification. The rules are framed by means of the bio-inspired algorithm namely CSOA. This work generates the significant rules for better classification accuracy. The proposed approach is elaborated in the following section.

Proposed Rule based Classification Approach

The rule based classification scheme requires more rules to ensure better classification but, it is quite tough to save more number of rules and to handle them. In addition to this, it is challenging to detect the appropriate rules for performing classification. The proposed work intends to address the above stated issues by forming dynamic rules with the help of CSOA.

The CSOA is completely based on the behaviour of cuckoo with the levy flight concept, as discussed in [20]. Basically, the CSO algorithm involves three significant phases and they are egg laying phase, quality of egg assessment and nest maintenance. The cuckoo bird lays eggs on the nest in a random fashion and the quality of eggs is assessed in the quality assessment phase. The eggs with better quality alone are suitable for the next generation. Suppose, if the nest contains eggs with poor quality, then the nests are discarded.

The proposed work achieves the goals in four different phases and they are data pre-processing, rule formation, rule pruning and classification phases. All these phases are explained as follows.

Microarray Data Pre-processing

The microarray dataset is high dimensional and hence, the dimensionality has to be reduced for better data processing. The high dimensionality of the microarray dataset slows down the process and degrades the performance of the system. Understanding the issue of high dimensionality, this work reduces the dimensionality by selecting the significant genes by employing Information Gain Ratio (IGR), as performed in our previous work [19].

The IGR chooses the significant genes by considering the information gain and the gene with maximal correlation level are chosen as significant genes. The IGR is computed by

$$IGR(F) = \left[\frac{IG_S - IG(F)}{IG_S + IG(F)}\right] \times 100 \tag{1}$$

Where IG_s and IG_F are given by

$$IG_S = -\left[\frac{r.o(c_1, S)}{|S|}\right] log_2 \left[\frac{r.o(c_1, S)}{|S|}\right] \tag{2}$$

$$IG(F) = \left[\frac{|F_i|}{|F|}\right] \times IG(F_i) \tag{3}$$

In the above equations, $\left[\frac{r.o(c_1,S)}{|S|}\right]$ denotes the probability of the repetitive occurrence of a gene, which is present in the class c_1. Let the feature f contains q unique values, which can be represented as $\{f_1, f_2, f_3, \ \ldots \ , f_q\}$. For a dataset with f features, the training dataset is formed as follows $\{c_1, c_2, c_3, \ldots, c_q\}$.

Hence, the IGR selects the genes with more significance by considering the correlation and information gain and these virulent genes are utilized for the forthcoming phases.

Rule Formation Phase

Rule formation is the heart of the proposed approach, which is meant for building the meaningful rule set and paves way easy classification. Hence, this work pays more attention to form better rules and the rules framed by the proposed work follows a specific structure. Consider that all the rules formed by the proposed work are represented in a standard format as given below.

$$Rl(rl_{i_1}, rl_{i_2}, \ldots CL, rl_{i_n}) \tag{4}$$

Here, rl_{i_n} is the different attributes and n is the total count of entities being present in each rule and is fixed on the basis of the count of genes. Suppose, if the user employs five genes to perform classification, then the classification rule is framed with the help of five genes. Each rule of the system involves two important parts and the standard form of the rule is as follows.

$$Rl : rl_i \rightarrow CL \tag{5}$$

Each and every rule is identified with the help of an identifier and is called as the index. Apart from this, each rule is treated with the operations such as max, min and mean. These operations are performed over the expression with respect to the class. The max and min are the two basic and significant operations, which can reveal the upper and the lower limit of the values.

Hence, the optimal upper and lower limits help in attaining better results and the limits are fixed by the CSO algorithm. The second important stage of rule formation is the rule detection, which intends to detect a particular kind of cancer out of the whole entities in the microarray data. The rules are detected in the following way. As mentioned earlier, each and every rule is computed with the max and mix values for every rule in the solution space. The lower and upper limit of a rule is computed by the following equations.

$$rl_i(lowl) = rl_i.m - q_1 \times (rl_i.max - rl_i.min) \tag{6}$$

$$rl_i(upl) = rl_i.m + q_2 \times (rl_i.max - rl_i.min) \tag{7}$$

In the above equations, rl_i. *max* and the rl_i. *min* are the maximum and minimum values of the rule item denoted as rl_i. With the upper and lower limits, the general range can be figured out. rl_i. *m* denotes the mean value of the microarray expression with specific identifier and is computed with respect to all the entities present in a specific class. q_1 and q_2 are the values ranging between 0 and 1. It is always ensured that the *lowl* and *upl* are unequal to each other and *lowl* is always lesser than the *upl*. As soon as the rules are detected, the fitness of the rules is figured out with the help of standard performance indicators such as True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN), as presented in eq. 8.

$$R_{fit} = \frac{TP}{TP + FN} \times \frac{TN}{TN + FP} \tag{8}$$

The gene data with respect to index is evaluated and when the value of the gene data exists between the *lowl* and *upl*, then it can be included in the rule. The class is predicted with the help of TP, TN, FP and FN. TP indicates that the total count of entities with predicted class and are included in the rule. TN denotes that the total count of entities with unpredicted class and are not included in the rule. FP indicates that the total count of entities with unpredicted class and are not included in the rule. FN indicates that the total count of entities with predicted class and are not included in the rule.

When the rules are formed by this way, the optimal solution to the problem is figured out by the CSO algorithm. This algorithm is based on the concept of egg laying by cuckoo on a randomly picked nest. The nests with high quality eggs alone are suitable for the next generation and the nests with poor quality eggs are discarded. In this case, the rules are eggs and the nests are randomly chosen.

For the rule item ri_1, the position of the cuckoo is randomly selected with the help of levy flights.

$$X_i^{(t+1)} = X_i^{(t)} + \alpha \oplus Levy(\lambda) \tag{9}$$

$$Levy(\lambda) = t^{-\lambda}; \lambda \in [0, 3] \tag{10}$$

Where λ is the step-length. In equation (9), $X_i^{(t+1)}$ is the microarray data identifier of a new rule item, which is on the nest *t*. $X_i^{(t)}$ is the microarray data identifier of a rule item on the nest *t*, which is the neighbour of the current solution. 훼 represents the step size and is greater than 0, mostly set as 1. $\oplus$ indicates the entry-wise multiplication, which is similar to the PSO algorithm. Employment of Levy flight boosts up the efficiency of navigating through the search space [20].

Suppose, when the position of egg has to be modified and it can be attained by changing the limit of the rule item as follows.

$$X_i^{(t+1)}.upl = X_i^{(t+1)}.max - \gamma \tag{11}$$

$$X_i^{(t+1)}.lowl = X_i^{(t+1)}.min + \gamma \tag{12}$$

$X_i^{(t+1)}.upl$ and $X_i^{(t+1)}.lowl$ denote the upper and the lower limit of the new rule and γ is the random number between 0 and 1. Hence, the rules are formed by considering the range of the data. Though, it is unnecessary to maintain all the rules and hence, the rules with optimal fitness rates are to be chosen and saved for future reference. This is done by the rule pruning step and is explained in the following section.

Rule Pruning Phase

The total count of rules computed by the rule formation phase is more and the count can be reduced with the help of this phase. The main task of this phase is to remove the duplicate and ineffective rules. Let the microarray training data is represented as MTD and CL is the class label. Each rule rl_i can be represented as in equation 5 and the support (*SUP*) and confidence (*CON*) rates of the rules are computed as follows.

$$SUP(Rl_i) = \frac{TP}{CL} \tag{13}$$

$$CON(Rl_i) = \frac{TP}{TP + FP} \tag{14}$$

The support of a rule is computed by the count of entities in the training dataset with matching items Rl_i for a given class label *CL*. The total count of matching rule items Rl_i with class label *CL* with the total count of matching rule items Rl_i is computed to determine the confidence of the rule. The overall algorithm of this work is presented as follows.

```
Proposed Algorithm
Input: Microarray gene dataset
Output: Disease classification
Begin
//Pre-processing
Pre-process the dataset by employing eqn.1;
//Rule formation
Rule set = NULL;
Choose a class from a set of classes;
Detect the rules and compute the fitness;
Add the rules with greatest fitness to the rule set;
//Rule pruning
Prune the rules by eqns (13) and (14);
//Classification
Compute the feasibility of the sample and forecast;
Return the result;
End;
```

Hence, the rules with maximal confidence are taken into account and stored for future classification. The data classification is presented as follows.

Table 1 Dataset details

Dataset	Classes	Total entities	Gene count
Colon	2	62	2000
Leukemia	2	72	7129
Lung	2	96	7129
SRBCT	4	83	2308
Lymphoma	3	62	4026
Leukemia	3	72	7129

Classification Phase

The entity is classified by computing the feasibility of the sample with respect to a particular class and is computed by the following

$$Feasibility = \frac{TP}{C_{cl}} \quad (15)$$

The feasibility of the sample is the ratio of the TP rate and the count of the entities with respect to a specific class being forecasted. The test entity is treated with the forecasting value of the rules as follows.

$$Forecast = \left(w1 \times R_{fit}\right) + (w2 \times feasibility) \quad (16)$$

In the above equation, w1 and w2 are the weighted parameters with respect to the fitness and the feasibility level. W1 is a random number between 0 and 1, w2 is computed by $(1-w1)$. The test sample is treated with these equations, so as to forecast the appropriate class for the sample. The class with the greatest feasibility is taken into account and suggested to the user. The classification accuracy of the microarray dataset is computed by the ratio of the perfectly classified test sample (CC_{Ts}) to the total count of samples in the dataset being classified correctly (N_c).

$$Classification\ accuracy = \frac{CC_{Ts}}{N_c} \quad (17)$$

By this way, the rule is provided with the classification accuracy. The performance of the proposed approach is evaluated in the forthcoming section.

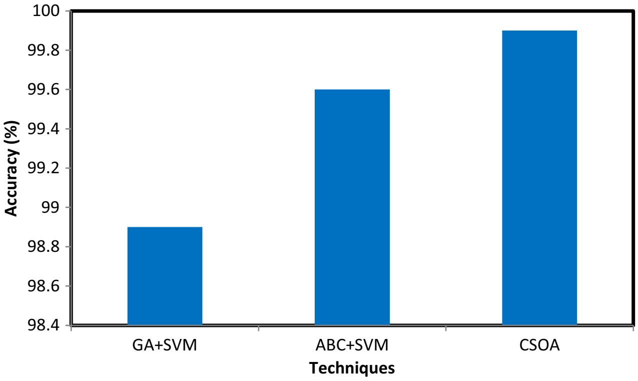

Fig. 2 Accuracy rate analysis

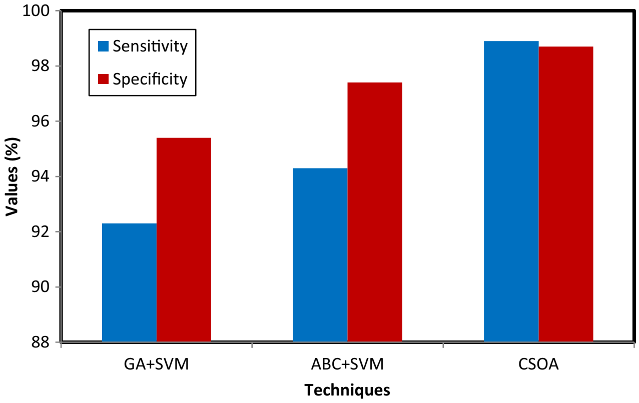

Fig. 3 Sensitivity and specificity rate analysis

Results and discussion

The performance of the proposed approach is evaluated over both binary and multiclass datasets. The binary class microarray datasets include colon, leukemia and lung cancer [21–23]. The considered multi-class microarray datasets are SRBCT, lymphoma and leukemia [24–26]. The initialized cuckoos are 200 and the termination condition is fixed by considering the count of iterations, which is 2500. The count of gene and classes available in the dataset are presented in Table 1.

The performance of the proposed work is tested and the experimental results are computed and compared with the existing approaches. The performance metrics considered are accuracy, sensitivity, specificity and time consumption. The experimental results attained by the proposed approach are as follows (Figs. 2, 3 and 4).

The performance of the proposed approach is compared against two recently proposed algorithms found in [27, 28] respectively. When the performance of the proposed approach is tested, the accuracy, sensitivity, specificity rates are proven to the maximum when compared to the existing approaches, while consuming minimal time period. The main reason for the better performance of the proposed approach is the utilization of CSOA and the rule pruning phase. This work can be applied to datasets with any classes without any prior

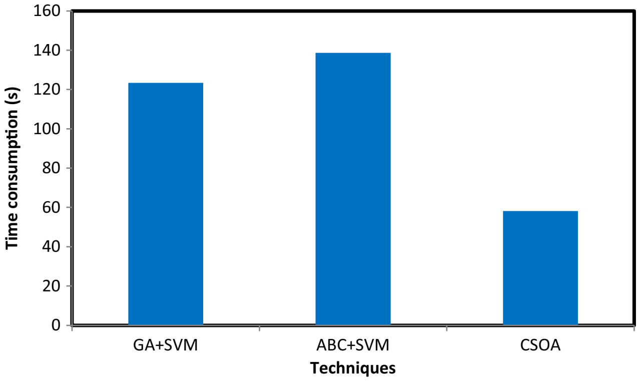

Fig. 4 Time consumption rate analysis

knowledge. As the rules are pruned with respect to the effectiveness, the performance of the proposed approach is enhanced. The following section concludes the paper.

Conclusions

This article proposes a rule based classification model that relies on CSOA, which is meant for microarray gene data. The microarray gene data is highly voluminous and is difficult to perform classification, which is required to distinguish between the type of cancer. This work trains the system with the sample train data obtained from the publicly available datasets and the proposed approach forms rules with the help of CSOA. The formed rules are then pruned with the help of associative rule mining, in order to select the efficient rules alone. Finally, the classification is performed by computing the feasibility and forecast values. The main advantages of this work are that the users need not to possess any prior knowledge and the classification can be done for any dataset. As the classification is based on rules, the proposed work is user-friendly and easily understandable. In future, this work is planned to be extended by considering the microarray image. In addition to this, the performance of different bio-inspired algorithms are planned to be incorporated into the system.

Compliance with ethical standards

Conflict of Interest The authors declare that they have no conflict of interest.

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Celik, M., Karaboga, D., Koylu, F., Artificial bee colony data miner (abc-miner). In: 2011 International Symposium on Innovations in Intelligent Systems and Applications (INISTA), June. pp. 96–100, 2011.
2. Talebi, M., Abadi, M., Beeminer: a novel artificial bee colony algorithm for classification rule discovery. In: 2014 Iranian Conference on Intelligent Systems (ICIS), February, pp. 1–5, 2014.
3. Pati, J., Gene Expression Analysis for Early Lung Cancer Prediction Using Machine Learning Techniques: An Eco-Genomics Approach. IEEE Access 7:4232–4238, 2018.
4. Katsigiannis, S., Zacharia, E., and Maroulis, D., MIGS-GPU: Microarray Image Gridding and Segmentation on the GPU. IEEE Journal of Biomedical and Health Informatics 21(3):867–874, 2017.
5. Hsieh, S.-Y., and Chou, Y.-C., A Faster cDNA Microarray Gene Expression Data Classifier for Diagnosing Diseases. IEEE/ACM Transactions on Computational Biology and Bioinformatics 13(1):43–54, 2016.
6. Wu, H.-C., Wei, X.-G., and Chan, S.-C., Novel Consensus Gene Selection Criteria for Distributed GPU Partial Least Squares-Based Gene Microarray Analysis in Diffused Large B Cell Lymphoma (DLBCL) and Related Findings. IEEE/ACM Transactions on Computational Biology and Bioinformatics 15(6):2039–2052, 2018.
7. Li, J., Dong, W., and Meng, D., Grouped Gene Selection of Cancer via Adaptive Sparse Group Lasso Based on Conditional Mutual Information. IEEE/ACM Transactions on Computational Biology and Bioinformatics 15(6):2028–2038, 2018.
8. Güney, H., and Öztoprak, H., Microarray-based cancer diagnosis: repeated cross-validation-based ensemble feature selection. Electronics Letters 54(5):272–274, 2018.
9. An, S., Wang, J., and Wei, J., Local-Nearest-Neighbors-Based Feature Weighting for Gene Selection. IEEE/ACM Transactions on Computational Biology and Bioinformatics 15(5):1538–1548, 2018.
10. Zhang, D., Zou, L., Zhou, X., and He, F., Integrating Feature Selection and Feature Extraction Methods With Deep Learning to Predict Clinical Outcome of Breast Cancer. IEEE Access 6:28936–28944, 2018.
11. Maji, P., and Shah, E., Significance and Functional Similarity for Identification of Disease Genes. IEEE/ACM Transactions on Computational Biology and Bioinformatics 14(6):1419–1433, 2017.
12. Zhang, L., Liu, H., Huang, Y., Wang, X., Chen, Y., and Meng, J., Cancer Progression Prediction Using Gene Interaction Regularized Elastic Net. IEEE/ACM Transactions on Computational Biology and Bioinformatics 14(1):145–154, 2017.
13. Yu, Z., Li, T., Horng, S.-J., Pan, Y., Wang, H., and Jing, Y., An Iterative Locally Auto-Weighted Least Squares Method for Microarray Missing Value Estimation. IEEE Transactions on NanoBioscience 16(1):21–33, 2017.
14. Gangeh, M. J., Zarkoob, H., and Ghodsi, A., Fast and Scalable Feature Selection for Gene Expression Data Using Hilbert-Schmidt Independence Criterion. IEEE/ACM Transactions on Computational Biology and Bioinformatics 14(1):167–181, 2017.
15. Harvey, B. S., and Ji, S.-Y., Cloud-Scale Genomic Signals Processing for Robust Large-Scale Cancer Genomic Microarray Data Analysis. IEEE Journal of Biomedical and Health Informatics 21(1):238–245, 2017.
16. Li, J., and Wang, F., Towards Unsupervised Gene Selection: A Matrix Factorization Framework. IEEE/ACM Transactions on Computational Biology and Bioinformatics 14(3):514–521, 2017.
17. Lt. Scaria, T., and Christopher, T., Microarray Gene Retrieval System based on LFDA and SVM. International Journal of Intelligent Systems and Applications (1):9–15, 2018.
18. Lt. Thomas Scaria, Christopher, T., Supervised Microarray Gene Retrieval System Based on KLFDA and ELM, International Journal of Advanced Intelligent Paradigms, Accepted, In Press, 2018.
19. Lt. Scaria, T., and Christopher, T., Ensemble Classification based Microarray Gene Retrieval System. ICTACT Journal on Soft Computing 9(1):1806–1812, 2018.
20. Yang, X. S., Nature-Inspired Metaheuristic Algorithms 128, (2008).
21. Alon, U., Barkai, N., Notterman, D., Gish, K., Ybarra, S., Mack, D., and Levine, A., Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. Proc. Nat. Acad. Sci. 96(12):6745–6750, 1999.
22. Golub, T., Slonim, D., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J., Coller, L., Downing, J., Caligiuri, M., Bloomfield, C., and Lander, E., Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 286(5439):531–537, 1999.
23. Beer, D. G., Kardia, S. L., Huang, C.-C., Giordano, T. J., Levin, A. M., Misek, D. E., Lin, L., Chen, G., Gharib, T. G., Thomas, D. G. et al., Gene-expression profiles predict survival of patients with lung adenocarcinoma. Nature Med. 8(8):816–824, 2002.
24. Khan, J., Wei, J. S., Ringner, M., Saal, L. H., Ladanyi, M., Westermann, F., Berthold, F., Schwab, M., Antonescu, C. R., Peterson, C. et al., Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. Nature Med. 7(6):673–679, 2001.

25. Alizadeh, A., Eisen, M., Davis, M., Rosenwald, A., Boldrick, J., Sabet, T., Powell, Y., Yang, L., Marti, G., Moore, T., Hudson, J., Lu, L., Lewis, D., Tibshirani, R., Sherlock, G., Chan, W., Greiner, T., Weisenburger, D., Armitage, J., Warnke, R., Levy, R., Wilson, W., Grever, M., Byrd, J., Botstein, D., Brown, P., and Staudt, L., Distinct types of diffuse large b-cell lymphoma identified by gene expression profiling. Nature 403(6769):503–511, 2000.

26. Armstrong, S. A., Staunton, J. E., Silverman, L. B., Pieters, R., den Boer, M. L., Minden, M. D., Sallan, S. E., Lander, E. S., Golub, T. R., and Korsmeyer, S. J., Mll translocations specify a distinct gene expression profile that distinguishes a unique leukemia. Nature Genet. 30(1):41–47, 2001.

27. Alshamlan, H. M., Badr, G. H., and Alohali, Y. A., Genetic bee colony (gbc) algorithm: a new gene selection method for microarray cancer classification. Comput. Biol. Chem. 56:49–60, 2015b.

28. Alshamlan, H. M., Badr, G. H., and Alohali, Y. A., Abc-svm: artificial bee colony and svm method for microarray gene selection and multi class cancer classification. Int. J. Mach. Learn. Comput. 6(3): 184, 2016.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.