

Comparison Analysis of the Pearson's Phi-Square Test and Correlation Metric Effectiveness to Form the Subset of Differently Expressed and Mutually Correlated Genes

Lyudmyla Yasinska-Damri^a, Sergii Babichev^b, and Igor Liakh^c

^a *Ukrainian Academy of Printing, Pid Goloskom street, 19, Lviv, 79000, Ukraine*

^b *Kherson State University, University street, 27, Kherson, 73000, Ukraine*

^c *Uzhhorod National University, University street, 14, Uzhhorod, 88000, Ukraine*

Abstract

The development of patients' health monitoring systems based on gene expression data is a very important direction of current bioinformatics. In this instance, the allocation of both differently expressed and mutually correlated gene expression profiles (GEP) which allow monitoring in real-time the patients' health with high accuracy is a very important step of this problem solution. There are various types of similarity metrics to identify the level of GEP proximity. In this research, we compare the Pearson chi-square test and correlation metric to evaluate the gene expression profiles proximity. The evaluation of appropriate metric effectiveness has been executed by applying the object's classification quality criteria such as accuracy, f-score and Matthews correlation coefficient (MCC). The simulation results have shown that the metric based on Pearson's phi-square coefficient is significantly effective in comparison with the correlation metric to allocate the mutually similar gene expression profiles and, this metric can be used when the differently expressed and mutually correlated GEP will be extracted using various clustering algorithms.

Keywords

Gene expression profiles, correlation metric, Pearson's chi-square test, gene expression profiles classification, classification quality criteria

1. Introduction and literature review

The extraction of a subset of differently expressed and mutually correlated gene expression profiles (GEP) to further create a decision support system regarding the various diseases diagnosis or the gene regulatory network (GRN) reconstruction involves assessing both the informativity and proximity of gene expression profiles by using both single methods or ensemble of appropriate methods to measure the degree of GEP proximity. Currently, the clustering and biclustering techniques are applied widely to solve this problem. The implementation of these methods allow identifying the differently expressed and mutually correlated GEP, however, their application is led to a high rate of subjectivity due to the imperfection of used quality measures. In addition, useful information may be lost due to removing the informative gene expression profiles that contain significant information about the condition of the investigated object. The application of hybrid models based on joint use of both machine learning and data mining techniques for creating models based on an ensemble of various methods in order to analyze and follow the formation of GEP subsets considering the type of the disease can be reasonable in this instance.

At a recent time, plenty of scientific papers have been devoted to the decision of the problems of measuring the degree of GEP informativeness in order to form the subsets of both differently

IntellITSIS'2022: 3rd International Workshop on Intelligent Information Technologies and Systems of Information Security, March 23–25, 2022, Khmelnytskyi, Ukraine

EMAIL: Lm.yasinska@gmail.com (L. Yasinska-Damri); sbabichev@ksu.ks.ua (S. Babichev); ihor.lyah@uzhnu.edu.ua (I. Liakh)

ORCID: 0000-0002-8629-8658 (L. Yasinska-Damri); 0000-0001-6797-1467 (S. Babichev); 0000-0001-5417-9403 (I. Liakh)



© 2022 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)

expressed and mutually proximate GEP in terms of the investigated objects recognizing accuracy. Thus, in [1] the authors considered the questions aimed at the detection of gene expression profiles of miRNA molecules. They have allocated from each library approximately 2,744,989 rows from 9,888,123 ones during the experiment carried out. As a result, 2,565 siRNAs molecules were discovered. The questions aimed at the comparative analysis of various types of classifiers application effectiveness to identify the differently expressed GEP using errors of both the first and second kind were considered in [2-4]. The principal shortcoming of the proposed method is that we a prior cannot have full information about the classes to which genes treat. For this reason, these techniques have a high rate of subjectivity.

A comparative analysis of different hybrid models aimed at the extraction of subsets of GEP to decide the problem of differently expressed and mutually similar gene expression profiles allocation to create the cancer disease classifier based on gene expression data was carried out in [5-7]. The authors considered various steps of gene expression data pre-processing: from filtering with the following statistical analysis of the experimental data in order to solve the feature selection task to the evaluation of various types of hybrid models based on the joint use of clustering algorithms and classifiers. To evaluate the effectiveness of the respective approach, the authors have used various criteria based on an estimation of the objects classification results. Additionally, the authors have considered various combinations of current machine learning and data mining techniques. In this review, the following filtration and feature selection methods have been considered: maximization of mutual information based on Shannon entropy criterion; chi-squared test; technique based on correlation analysis; Fisher and Laplacian measure; random forest ranking technique; method based on a probabilistic random function; logarithmic transformation; method based on maximum relevance minimum redundancy; information gain technique. The authors in the appropriate hybrid models have used various combinations of gene grouping methods and the investigated objects classification techniques.

The hereinbefore presented brief review allows inferring that the problem of objective extraction of differently expressed and mutually similar genes in terms of high-resolution ability when the disease diagnostic has not an unambiguous solution nowadays. In many instances, the acceptable classification accuracy was reached when a small quantity of gene expression profiles was applied. To reconstruct a qualitative gene regulatory network (in order to understand the particularities of genes interconnection) it is necessary to use a larger quantity of differentially expressed and mutually correlated gene expression profiles.

Papers [8,9] present the partial solution of this problem. Proposed by the authors hybrid model assumes joint application of Shannon entropy, various types of statistical quality criteria, the SOTA clustering algorithm where the correlation distance was applied as the proximity metric, and various types of binary classifiers. The authors have proposed the step-by-step procedure of the GEP division with the evaluation of each step carried out effectivity by applying both the clustering and classification quality criteria. The fuzzy inference technique was applied to do the final decision about the differently expressed and mutually similar GEP selection. To the authors' minds, the application of an ensemble of the quality criteria contributes to higher objectivity when subsets of mutual similar and differently expressed genes extraction. However, the proposed technique has some shortcomings. At first, it focused on the dataset containing only two classes of the investigated objects. The authors have not considered the multi-class datasets. Thus, it will be better to extend the types of classifiers using datasets that contain a higher number of classes. The second shortcoming regards a limited number of datasets used during the simulation process performing. Thus, it is necessary to validate this model using various types of other gene expression profiles datasets [10,11].

A presented brief review of the current research in this subject area indicates that the problem of the extraction of mutually correlated and differently expressed GEP considering the type of disease is actual and, at the present time, this problem has no unambiguous decision. Its effective solution can be obtained using current techniques of computer science (data mining and machine learning) which are applied successfully in different fields of both applied and scientific research nowadays [12-15]. The choice of the gene expression profiles proximity metric that allows objectively grouping the mutually correlated and differently expressed GEP is the principal stages of this problem solve. The current research aims to the comparative analysis of various proximity metrics such as Pearson's chi-square test and correlation-based metric to assess the GEP proximity.

The **goal of the paper** is the comparative analysis of correlation-based metric and Pearson's chi-square test to assess the GEP proximity using various types of the classification quality criteria as the main criteria to assess the respective metric efficiency.

2. Problem statement

Let, the experimental gene expression data be represented as follows:

$$G = \{e_{sp}\}, s = \overline{1, n}; p = \overline{1, m}, \quad (1)$$

where: n is the number of genes that determine the state of the investigated samples; m is the number of the samples.

The main measure for the formation of subsets of differently expressed and mutual similar GEP in this instance be a target function:

$$F(e_i, e_j) = \min f(e_i, e_j), \quad (2)$$

where: e_i and e_j are the gene expression profiles i and j respectively; $f(\cdot)$ is the proximity function used to assess the proximity level of i -th and j -th GEP.

In our research, we investigate as the GEP proximity function Pearson's chi-square (χ^2) coefficient and correlation metric. The results of the objects classification were used to assess the appropriate similarity function effectiveness.

3. Materials and methods

3.1. Metrics and criteria to evaluate the GEP proximity

Pearson's statistical chi-square (χ^2) measure tests the hypothesis that the values of GEP are distributed according to the same law [16]. Let the k -th gene expression profile be presented as a numeric vector of expression values: $e_k = (e_{kp}), p = \overline{1, m}$, where m means the number of the study samples or conditions of the experiment carried out to form the gene expression data. If the range of the appropriate gene expression values $[e_k^{\min}, e_k^{\max}]$ divides into d non-intersection intervals $[e_{ks}^r, e_{kp}^r], r = \overline{1, d}$, then, the number of expression values allocated within the respective interval r can

be determined in the following way: $m_r = \sum_{j=1}^m [e_{ks}^r < e_{kj} \leq e_{kp}^r]$.

In the case of the Pearson's chi-square test traditional application when using categorized data, initially it is necessary to assess the number of the investigated vector values belonging to the appropriate interval. At the second stage, it is necessary to assess the expected amount of samples in the respective range taking into account the probability of the appropriate sample allocation in the corresponding range: $m'_r = p_r \cdot m$. Chi-square coefficient in this instance can be evaluated in the following way:

$$\chi^2 = \sum_{r=1}^d \frac{(m_r - m'_r)^2}{m'_r} \quad (3)$$

The hypothesis that the studied data values are distributed accordingly to a certain distribution is accepted or rejected is done on the basis of comparative analysis of the criterion (3) with boundary value taking into consideration both the amount of the freedom grades and the likelihood of the result receiving. The null hypothesis is rejected (the data distribution does not correspond to the appropriate distribution) if the chi-square value is greater than the boundary value. Otherwise, this hypothesis is accepted.

When we process the gene expression profiles, the value of the expressions in profiles is proportional to the quantity of appropriate specific gene. If we compare two GEP e_i and e_j , we

assume, that the expression values in the first and second profiles are expected and evaluated respectively. Then, equation (3) for the chi-square criterion calculation takes the form:

$$\chi^2 = \sum_{s=1}^m \frac{(e_{is} - e_{js})^2}{e_{is}} \quad (4)$$

Higher gene expression profiles proximity level suits to a smaller value of the criterion (4). This fact is a basis to form the subsets of mutual similar GEP based on Pearson's chi-squared test.

The second measure which is used in our research assumes the calculation of pairwise correlation distance between appropriate GEP in order to assess the degree of their consistency. As was noticed hereinbefore, the main goal of GEP pre-processing (features selection) is the allocation of differently expressed and mutually correlated GEP, which can allow us to identify the investigated samples contained allocated GEP as the attributes with the highest accuracy. We assume that the allocated gene expression profiles which correspond to hereinbefore listed requests should have a high level of mutual correlation and we can use the Pearson's correlation coefficient (since the gene expression profiles values are the numeric ones) to form the subset of differently expressed and mutually correlated gene expression profiles:

$$d_{cor}(e_i, e_j) = 1 - \frac{\sum_{s=1}^m (e_{is} - \bar{e}_i)(e_{js} - \bar{e}_j)}{\sqrt{\sum_{s=1}^m (e_{is} - \bar{e}_i)^2} \sqrt{\sum_{s=1}^m (e_{js} - \bar{e}_j)^2}}, \quad (5)$$

where \bar{e}_i and \bar{e}_j are average value of the e_i and e_j gene expression profiles respectively.

Similar to the previous case (the use of Pearson's chi-squared test), the minimal value of the criterion (5) suits to higher proximity level of the investigated GEP.

The evaluation of the used distance functions (Pearson's chi-square coefficient and correlation distance) effectiveness was done based on the analysis of the investigated samples classification by calculation of respective quality criteria based on errors of both the first and the second kinds. Within the framework of the simulation process implementation, we used the Random Forest (RF) binary classifier [17,18] to assess the appropriate metric effectiveness. An effectiveness of this classifier implementation to identify the gene expression data was proven in [8]. As experimental data, we used the samples of patients examined on lung cancer. In accordance with the data description, the investigated samples can be divided into two groups: health patients and patients with lung cancer tumors. The quality of data classification was evaluated using criteria that contain as components the first and the second types errors. Table 1 presents the confusion matrix used to calculate the classification quality criteria.

Table 1

Confusion matrix

The real state of the patient according to the diagnosis results	Result of the object classification	
	Patients with tumor	Healthy patients
Patients with tumor	True positive values (TPV)	False negative values (FNV)
Healthy patients	False positive values (FPV)	True negative values (TNV)

To assess the efficiency of the hereinbefore listed metrics we applied the traditional classification quality criteria such as:

- Accuracy (ACC), defined as the total probability of correct results prediction by the classifier use:

$$ACC = \frac{TPV + TNV}{TPV + FPV + TNV + FNV} \quad (6)$$

- F-score (FS), is a measure of the accuracy of a current model operation and it can be used to assess the binary classifier effectiveness that classifies the samples into negative and positive ones. F-score combines the Precision (PRC) and Recall (RCL) in the following way:

$$FS = \frac{2 \cdot PRC \cdot RCL}{PRC + RCL}, \quad (7)$$

where:

$$PRC = \frac{TPV}{TPV + FPV}; \quad RCC = \frac{TPV}{TPV + FNV}$$

- Matthews correlation coefficient (*MCC*) is a measure to assess the binary classifier effectiveness [19]:

$$MCC = \frac{(TPV \cdot TNV) - (FPV \cdot FNV)}{\sqrt{(TPV + FPV) \cdot (TPV + FNV) \cdot (TNV + FNV) \cdot (TNV + FNV)}} \quad (8)$$

Higher values of measures (6) - (8) correspond to the higher efficiency of the data classification procedure implementation.

3.2. The stepwise procedure of the simulation process implementation

The algorithm with a stepwise procedure implemented within the framework of the simulation process to assess the efficiency of the used measures is shown in Figure 1. Its practical implementation assumes the following:

Stage I. Forming the gene expression data as a matrix and vector of the method to evaluate the distance function between the gene expression profiles.

1.1. An analysis of the gene expression data and the forming these data as a matrix with rows and columns which are represented the investigated samples and genes that characterize the corresponding samples.

1.2. Formation of a vector of distance functions calculation methods for further estimation of mutual proximity of gene expression profiles.

Stage II. Formation of a triangular distance matrix contained distance values between the gene expression profiles by the application of appropriate distance function.

2.1. Selection of the first method from the formed vector of distance function calculation methods ($n = 1$).

2.2. Calculation of the distance function value for all pairs of GEP that make up the matrix of gene expression data. Formation of a triangular distances matrix between the GEP.

2.3. Selection of k_{max} number of the mutually similar and differently expressed GEP considering the current distance function.

Stage III. Classification of the investigated objects and form the vector of the classification quality criteria.

3.1. Selection of two the nearest gene expression profiles according to the current distance function ($k = 2$).

3.2. Initialization of the classification stage of objects containing as attributes of the selected genes expression values ($p = 1$). To increase the classification objectivity, this procedure was repeated 10 times ($p_{max} = 10$) with the redistribution of objects in the training and test data subsets.

3.3. Implementation of the data classification procedure at the appropriate stage p .

3.4. Calculation of the quality criteria by the formulas (6) – (8).

3.5. Increasing the number of nearby GEP per unit and going to step 3.2 of this procedure. If the number of expression profiles of genes used as attributes during the object classification procedure implementation reaches the maximum value, then the formation of a matrix of the classification quality criteria.

Stage IV. The received results analysis.

4.1. Creation of the diagrams of the used classification quality criteria for various increasing quantities of the mutually correlated and differentially expressed GEP using appropriate distance metric.

4.2. The simulation results analysis. Doing the conclusion about the efficiency of the tested proximity metrics.

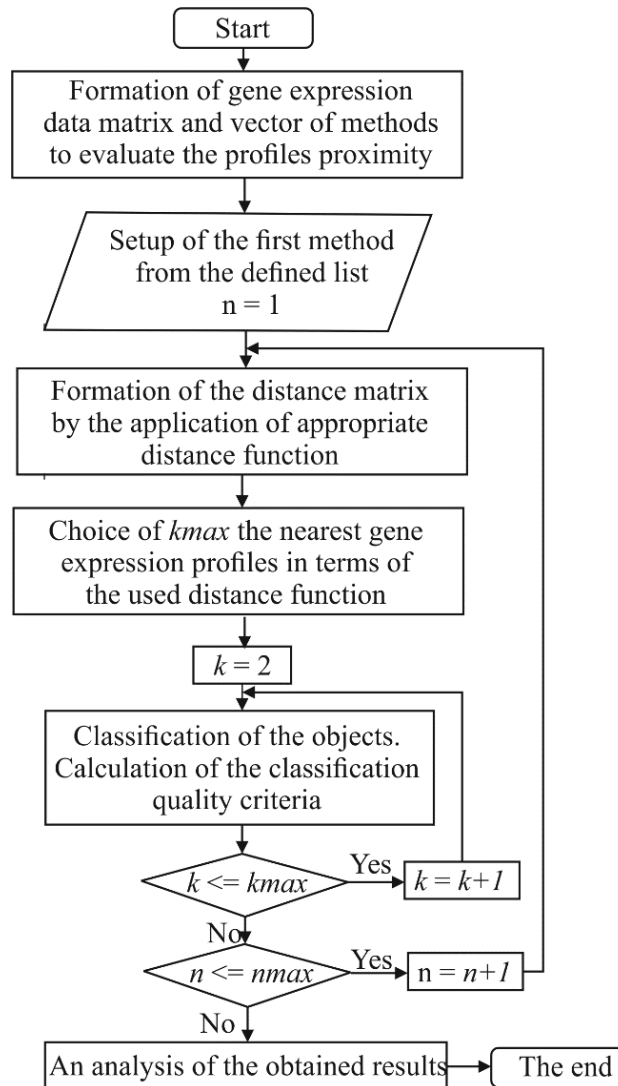


Figure 1: Structure block chart of algorithm to evaluate the effectiveness of the Pearson's chi-square test and correlation proximity metric

4. Experiment, results and discussion

The experimental basis for the hereinbefore presented algorithm implementation was the dataset GSE19188 [20]. This dataset contains the experimental results of testing the various patients on lung cancer (156 patients). As a result, the expression values of 54675 genes were assessed. The simulation process was carried out using the R programming language. 156 samples were studied for the experimental data formation. Taking into consideration the data annotation, the tested samples were shared into two groups: healthy patients (65 samples) and patients with tumor of cancer (91 patients). As we have noted early, each of the samples contained in total 54675 genes, half of which was non-expressed for all samples (genes expression was zero). These genes were removed from the dataset at the first step. We have used the results of the research presented in [4], where the authors have applied a hierarchical clustering procedure with joint use of SOTA clustering algorithm with correlation proximity metric and binary classifiers to fixation an optimal hierarchical level of the gene expression data partition in terms of the samples classification accuracy. The authors allocated 401 differently expressed and mutually correlated gene expression data. The use of these data as the attributes has

allowed the authors to get approximately 94% classification accuracy. This genes were used as the experimental dataset during the simulation process implementation.

Figure 2 shows the charts of both the Pearson's chi-square coefficient and correlation distance values distribution using both a box-and-whiskers diagram and a kernel-density plot. An analysis of the obtained charts allows us to conclude that in both cases the data have outliers that correspond to significantly higher values of the distance between the respective GEP. These profiles should not be used at a further stage of data classification.

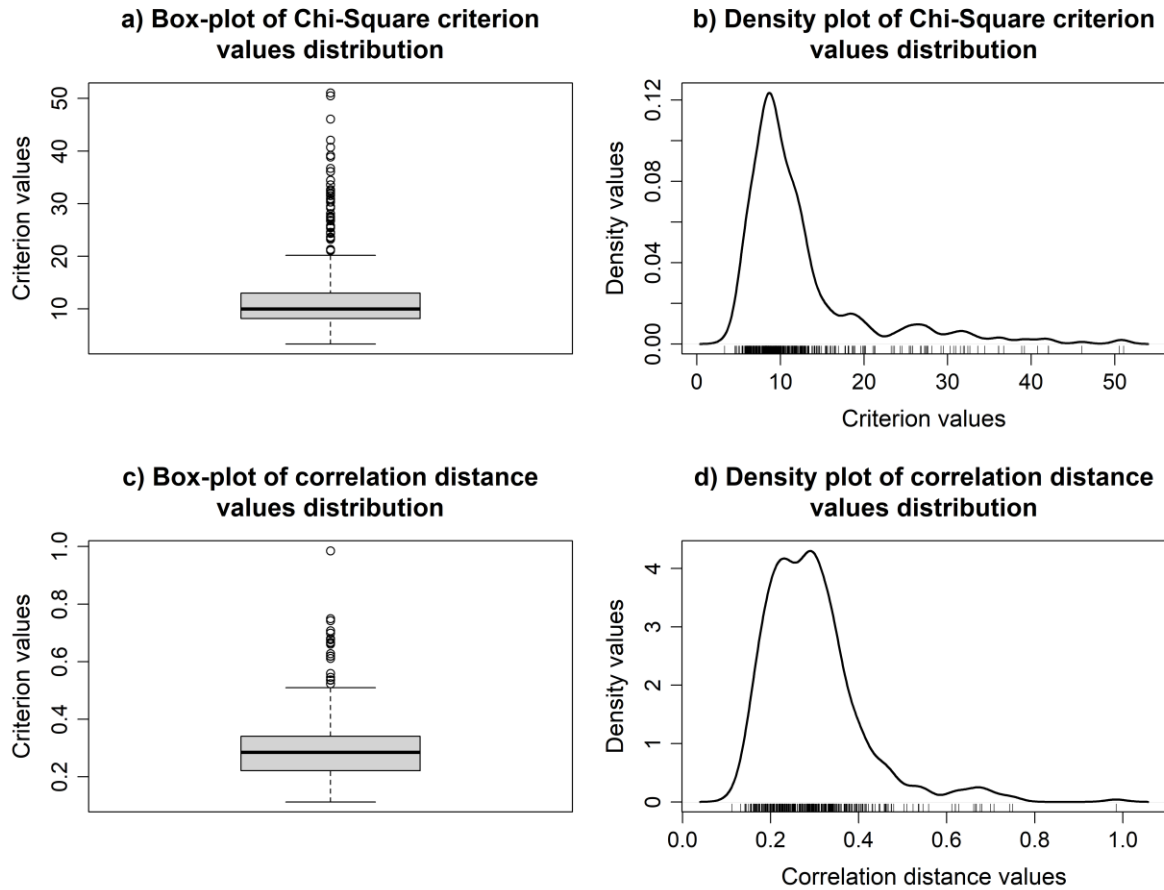


Figure 2: Charts of distribution of the Pearson phi-square coefficient (a,b) and correlation distance (c,d) values

Figures 3-5 show the results of the examined samples classification. The nearest and differently expressed gene expression profiles in terms of the applied distance metrics were used as the attributes of the examined samples. The number of GEP was increasing from 2 to 100 during the simulation procedure implementation. The results are presented as the charts of classification quality criteria calculated by the formulas (6) – (8) versus the number of GEP.

The received charts analysis allows concluding the use of the chi-square test is more reasonable in comparison with the correlation measure in terms of various criteria that were used when the simulation procedure was executing. As it can be seen in the charts, when we applied the correlation measure to form the subset of the nearest and differently expressed GEP, the samples classification results are significantly worse in comparison with the results got with the use of Pearson's chi-square coefficient as the distance function.

The obtained results create the conditions for increasing the objectivity of the most informative GEP extraction due to the careful selection of the distance functions which can be used as the component in complex distance metric calculated based on the use of an ensemble of the most effective distance functions.

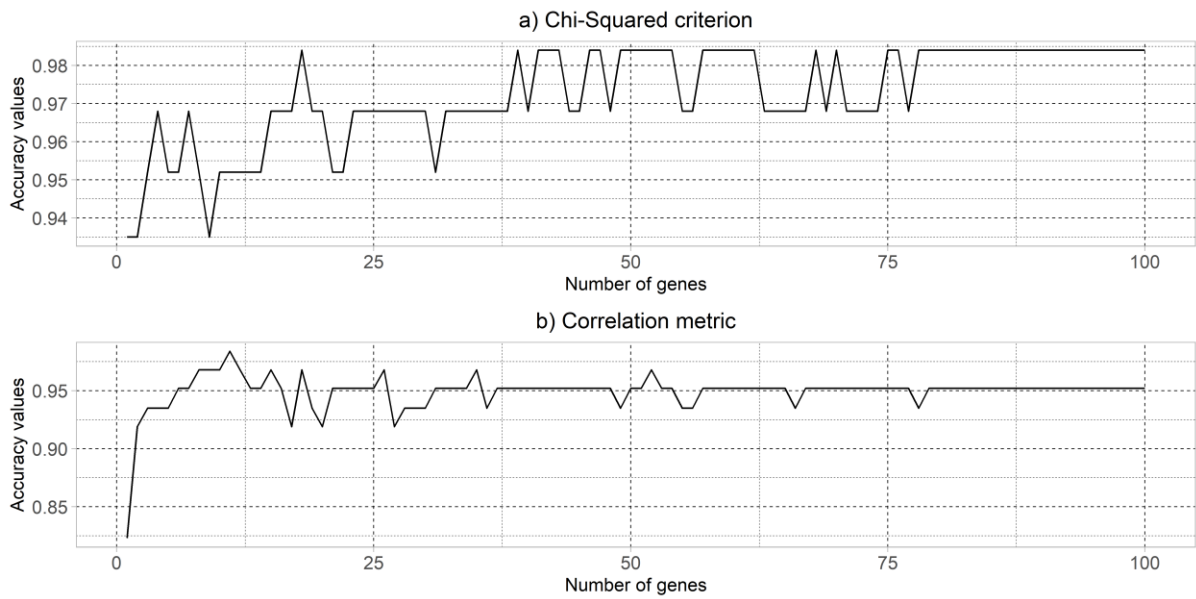


Figure 3: Charts of classification accuracy values when the increasing number of the nearest and differently expressed GEP

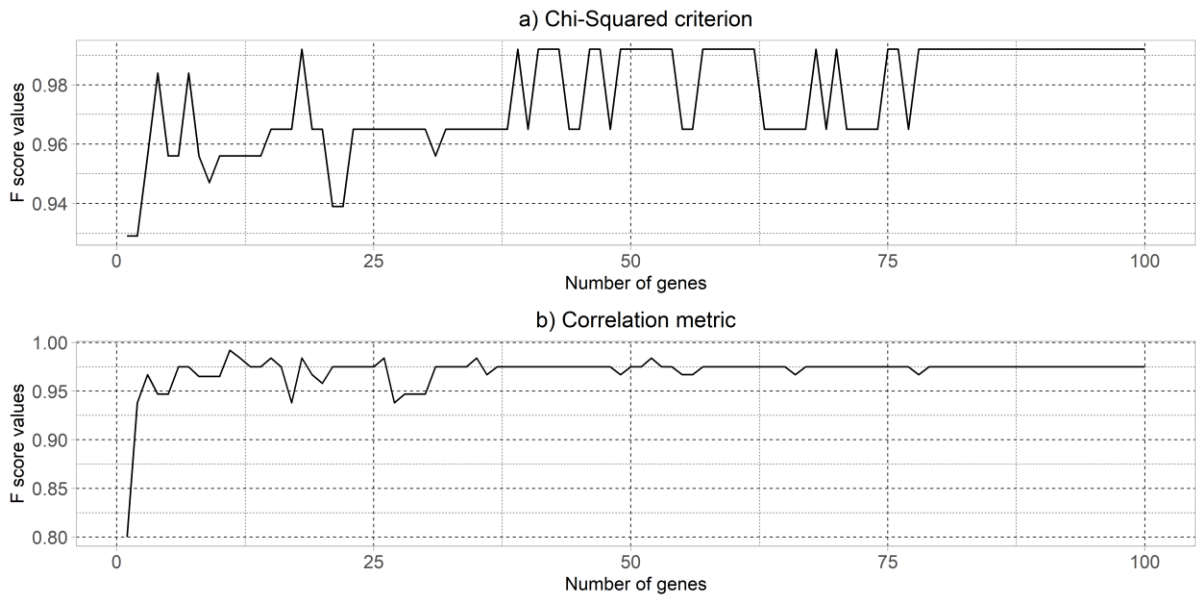


Figure 4: Charts of F-score measure values when the increasing number of the nearest and differently expressed GEP

5. Conclusions

In this research, we have carried out the comparative analysis of two distance functions: Pearson chi-square coefficient and correlation distance to assess the GEP proximity. The results of the investigated objects classification have been used to evaluation of the appropriate distance function effectiveness. The classification accuracy, F-score and Matthews correlation coefficient have been used as the classification quality criteria within the framework of our research. The dataset GSE19188 gene expression profiles of patients studied for early-stage lung cancer has been used as the experimental data. Taking into consideration the data annotation, the tested samples were shared into two groups: healthy patients (65 samples) and patients with tumor of cancer (91 patients). We have applied 401 differently expressed and mutually correlated gene expression data as the experimental dataset during the simulation process implementation.

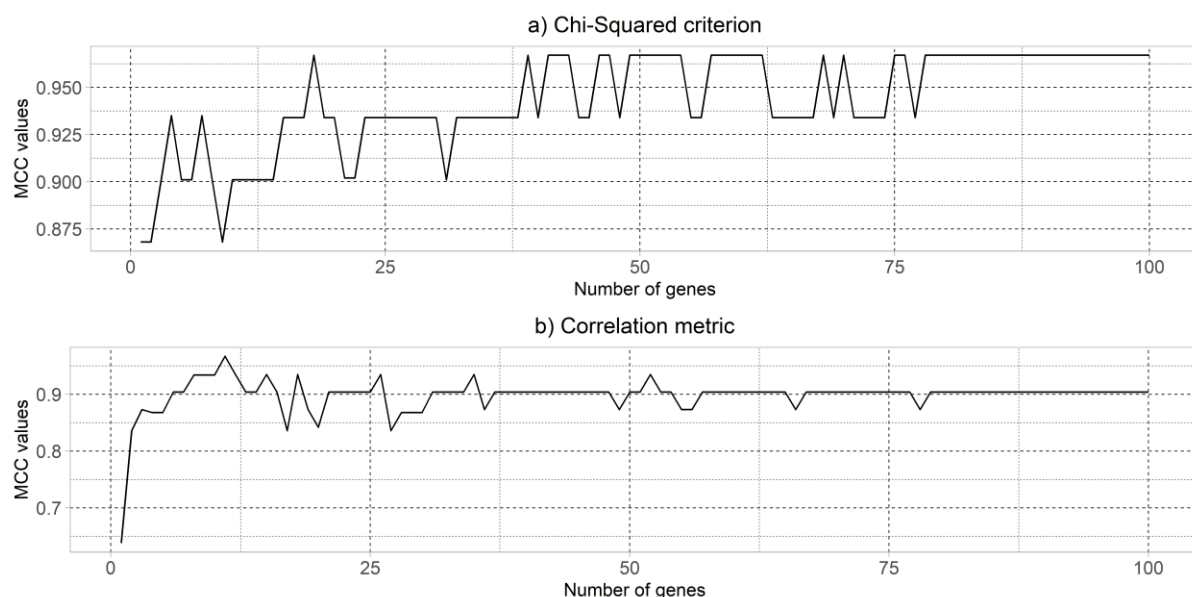


Figure 5: Charts of Matthews correlation coefficient values when the increasing number of the nearest and differently expressed GEP

The stepwise procedure of increasing the nearest gene expression profiles from 2 to 100 with the implementation of data classification and calculation of the classification quality criteria has been implemented during the simulation process. The charts of classification quality criteria versus the number of gene expression profiles for each of the used distance functions have been obtained as the simulation results.

An analysis of the obtained charts has allowed us to conclude about the lower efficiency of the correlation distance metric in comparison with the Pearson's phi-square coefficient both in absolute value and sensitivity. When using correlation distance metric for the subset of the nearest gene expression profiles formation, the results of the objects' classification that make up a subset of the testing data are significantly worse than the results obtained with the use of Pearson's phi-square coefficient as the distance function. The obtained results also create the conditions for increasing the objectivity of the most informative gene expression profiles extraction due to the careful selection of the distance functions which can be used as the component in complex distance metric calculated based on the use of an ensemble of the most effective distance functions. This is a further perspectives of the authors' research.

6. References

- [1] L. Wang, F. Song, H. Yin, et al. Comparative microRNAs expression profiles analysis during embryonic development of common carp, *Cyprinus carpio*. *Comparative Biochemistry and Physiology - Part D: Genomics and Proteomics*, 37 100754 (2021). doi: 10.1016/j.cbd.2020.100754
- [2] M. A. Marchetti, D.G. Coit, S.W. Dusza, et al. Performance of Gene Expression Profile Tests for Prognosis in Patients with Localized Cutaneous Melanoma: A Systematic Review and Meta-Analysis. *JAMA Dermatology*, 156(9) 2020 953-962. doi: 10.1001/jamadermatol.2020.1731
- [3] K.C. Howlader, M.S. Satu, M.A. Awal, et al. Machine learning models for classification and identification of significant attributes to detect type 2 diabetes. *Health Information Science and Systems*, 10 (1) 2 (2022). doi: 10.1007/s13755-021-00168-2
- [4] L. Zhou, Y. Zhu, T. Zong, Y. Xiang. A feature selection-based method for DDoS attack flow classification. *Future Generation Computer Systems*, 132 (2022) 67-79. doi: 10.1016/j.future.2022.02.006
- [5] N. Almgren, H. Alshamlan. A survey on hybrid feature selection methods in microarray gene expression data for cancer classification. *IEEE Access*, 7 8736725 (2019) 78533-78548. doi: 10.1109/ACCESS.2019.2922987

- [6] S. Park, G. Yi. Development of Gene Expression-Based Random Forest Model for Predicting Neoadjuvant Chemotherapy Response in Triple-Negative Breast Cancer. *Cancers*, 14 (4) 881 (2022). doi: 10.3390/cancers14040881
- [7] S.M. Snow, K.A. Matkowskyj, M. Maresh, et al. Validation of genetic classifiers derived from mouse and human tumors to identify molecular subtypes of colorectal cancer. *Human Pathology*, 119 (2022) 1-14. doi: 10.1016/j.humpath.2021.10.002
- [8] S. Babichev, J. Škvor. Technique of Gene Expression Profiles Extraction Based on the Complex Use of Clustering and Classification Methods. *Diagnostics*, 10 (8) 584 (2020). doi: 10.3390/diagnostics10080584
- [9] S. Babichev, J. Krejci, J. Bicanek, V. Lytvynenko, V. Gene expression sequences clustering based on the internal and external clustering quality criteria, 2017, Proceedings of the 12th International Scientific and Technical Conference on Computer Sciences and Information Technologies, CSIT2017, 1 (2017) 91-94. doi: 10.1109/STC-CSIT.2017.8098744
- [10] L.-H. Lee, C.-H. Chen, W.-C. Chang, et al. Evaluating the performance of machine learning models for automatic diagnosis of patients with schizophrenia based on a single site dataset of 440 participants. *European Psychiatry*, 65 (1) e1 (2022). doi: 10.1192/j.eurpsy.2021.2248
- [11] K.-N. Heo, J.-Y. Lee, Y.-M. Ah. Development and validation of a risk-score model for opioid overdose using a national claims database. *Scientific Reports*, 12 (1) 4974 (2022). doi: 10.1038/s41598-022-09095-y
- [12] P. Vitynskyi, R. Tkachenko, I. Izonin, H. Kutucu. Hybridization of the SGTm Neural-Like Structure Through Inputs Polynomial Extension, 2018, Proceedings of the 2018 IEEE 2nd International Conference on Data Stream Mining and Processing, DSMP, 8478456 (2018) 386-391. doi: 10.1109/DSMP.2018.8478456
- [13] M. Haghghat, L. Browning, K. Sirinukunwattana, et al. Automated quality assessment of large digitised histology cohorts by artificial intelligence. *Scientific Reports*, 12(1) 5002 (2022). doi: 10.1038/s41598-022-08351-5
- [14] M.R. Sabour, M. Besharati, G.A. Dezvareh, M. Hajbabaie, M. Akbari. Application of artificial neural network with the back-propagation algorithm for estimating the amount of polycyclic aromatic hydrocarbons in Tehran Oil Refinery, Iran. *Environmental Nanotechnology, Monitoring and Management*, 18 100677 (2022). doi: 10.1016/j.enmm.2022.100677
- [15] N. Shakhovska, V. Vysotska, L. Chyrun. Intelligent systems design of distance learning realization for modern youth promotion and involvement in independent scientific researches, 2017, *Advances in Intelligent Systems and Computing*, 512 (2017) 175-198. doi: 10.1007/978-3-319-45991-2_12
- [16] B. Liu, W. Gou, H. Feng. Pathological investigations and correlation research of microfibrillar-associated protein 4 and tropoelastin in oral submucous fibrosis. *BMC Oral Health*, 21 (1) 588 (2021). doi: 10.1186/s12903-021-01962-w
- [17] L. Breiman. Random forests. *Machine Learning*, 45 (2001) 5-32.
- [18] S. van Gaal, A. Alimohammadi, A., A.Y.X. Yu, et al. Accurate classification of carotid endarterectomy indication using physician claims and hospital discharge data. *BMC Health Services Research*, 22 (1) 379 (2022). doi: 10.1186/s12913-022-07614-1
- [19] G. Canbek, T. Taskaya Temizel, S. Sagiroglu. BenchMetrics: a systematic benchmarking method for binary classification performance metrics. *Neural Computing and Applications*, 33 (21) (2021) 14623-14650. doi: 10.1007/s00521-021-06103-6.
- [20] J. Hou, J. Aerts, B. den Hamer, et al. Gene expression-based classification of non-small cell lung carcinomas and survival prediction. *PLoS ONE*, 5 e10312 (2010). doi: 10.1371/journal.pone.0010312