



Javanbakht T., Chakravorty S. (2022). Optimization of machine learning algorithms for proteomic analysis using topsis. *Journal of Engineering Sciences*, Vol. 9(2), pp. E7-E11, doi: 10.21272/jes.2022.9(2).e2

Optimization of Machine Learning Algorithms for Proteomic Analysis Using TOPSIS

Javanbakht T.^{1*}, Chakravorty S.²

¹ Department of Computer Science, University of Quebec in Montreal,
201, President Kennedy St., Montreal, Quebec H2X 3Y7, Canada;

² Biju Patnaik University of Technology, Chhend Colony, Rourkela, Odisha 769004, India

Article info:

Submitted: August 21, 2022
Accepted for publication: November 24, 2022
Available online: November 28, 2022

*Corresponding email:

javanbakht.taraneh@courrier.uqam.ca

Abstract. The present study focuses on a new application of the TOPSIS method for the optimization of machine learning algorithms, supervised neural networks (SNN), the quick classifier (QC), and genetic algorithm (GA) for proteomic analysis. The main hypotheses are that the change in the weights of alternatives could affect the ranking of algorithms. The obtained data confirmed this hypothesis for their ranking. Moreover, adding labor as a cost criterion to the list of criteria did not affect this ranking. This was because candidate 3 had better fuzzy membership degrees than the two other candidates concerning their criteria. This work showed the importance of the value of the fuzzy membership degrees of the cost criterion of the algorithms in their ranks. The values of the fuzzy membership degrees of the algorithms used for proteomic analysis could determine their priority according to their score differences. One of the advantages of this study was that the studied methods could be compared according to their characteristics. Another advantage was that the obtained results could be related to the new ones after improving these methods. The results of this work could be applied in engineering, where the analysis of proteins would be performed with these methods.

Keywords: multi-criteria decision making, TOPSIS, prediction, proteomic analysis.

1 Introduction

The multi-criteria decision-making (MCDM) techniques are essential tools that have found their applications in diverse fields of science and engineering. Data optimization based on predicting the best alternative and ranking candidates can be carried out using these algorithms. In recent years, some MCDM algorithms have been used to determine the possible optimal alternatives. In these methods, linguistic values are used for the assessment of weights of criteria in order to obtain the ranks of alternatives [1-3].

The technique for order of preference by similarity to ideal solution (TOPSIS) is an MCDM method which is more appropriate than other such techniques in the first technique. Profit and cost criteria are analyzed according to their difference for the candidates [4-6]. The advantages of TOPSIS are important in science and engineering, which are the use of application, universality, consideration of distances to the ideal solutions and simplicity on computation and presentation [7, 8].

Proteomic analysis, which investigates the structure and function of proteins, is an important issue in engi-

neering and its related fields. In the proteomic analysis, the dysregulated proteins in patients are selected, and the protein biomarkers that can influence the metabolisms of organs, such as bone, cartilage, liver, etc., are identified [9-12]. The proteomic pattern identification of diseases, such as different varieties of cancers, for which there are no early specific symptoms, is considered a necessary investigation as no long survival rate is obtained when these diseases are diagnosed in advanced stages [13].

TOPSIS has been used previously to analyze proteomic data and predict disease stages or cell expression profiles [14, 15]. These investigations use this feature ranking approach to select the most discriminative proteins from proteomics data for cancer staging. Moreover, introduced proteins are also investigated, which can potentially be applied in medical practice. Lower computational complexity and more production of general results without overfitting are the advantages of the approach used in these studies.

Although different machine learning algorithms such as neural network (NN), the quick classifier (QC), and genetic algorithm (GA) have been applied for protein analysis [16-21], their optimization in a comparative

study remains a novel topic to be investigated. To our knowledge, algorithms such as supervised neural network (SNN), QC and GA have not yet been optimized with TOPSIS for proteomic analysis. This novel work will provide insight into this method for optimizing these algorithms.

In this work, we investigated the optimization of three algorithms, SNN, QC, and GA, for the proteomic analysis with TOPSIS. The novelty of the current work relies on the feature ranking of the algorithms and their optimization with two series of modifications carried out in our analysis. We modified the weight values and the number of criteria in separate analyses and used TOPSIS as the appropriate optimization method to distinguish the profit and cost criteria. To our knowledge, the optimization of the algorithms used for proteomic analysis has not been investigated with this method.

The rest of the paper includes the following sections. In Section 2, we present the methodology of our paper, including information on the classification algorithms applied for the proteomic analysis and the TOPSIS method. In Section 3, we present our results in three series of analyses. We discuss the results and conclude the paper in Sections 4 and 5, respectively.

2 Research Methodology

This research work focused on optimizing classification algorithms for proteomic analysis with the TOPSIS method. The quantitative approach presented here was based on this method's characteristics, which allowed analyzing these algorithms according to their characteristic differences.

Three machine learning algorithms: SNN (candidate 1), QC (candidate 2), and GA (candidate 3), were analyzed in this work as their characteristics were previously investigated by Swiatly and her colleagues [13].

We used the TOPSIS code in python. The steps of this method were presented in our previous work [22].

We analyzed three algorithms: SNN (candidate 1), QC (candidate 2), and GA (candidate 3). These candidates are indicated as C-1, C-2, and C-3, respectively. The terms considered for the level of the candidates' criteria, such as recognition capability, cross-validation, sensitivity, and specificity, which were all the profit criteria, were as follows: very high, medium, high, and high for C-1, high, high, high and medium for C-2, and very high, medium, very high and medium for C-3, respectively. The terms used in the table were chosen according to the values of the characteristics of these algorithms, as indicated in table 3 of the paper of Swiatly and her colleagues [13].

3 Results

Table 1 shows the triangular fuzzy membership degrees of the candidates' characteristics and their mean values.

The alternative weight value of 0.5 was assigned for all the profit criteria: recognition capability, cross-validation, sensitivity, and specificity. In the criteria matrix, the term "true" was used for the profit criteria.

Table 1 – Triangular fuzzy membership degrees of candidates' characteristics

Candidates/ Criteria	Recognition capability	Cross validation	Sensitivity	Specificity
Triangular fuzzy degrees				
C-1	0.8, 0.9, 1.0	0.4, 0.5, 0.6	0.6, 0.7, 0.8	0.6, 0.7, 0.8
C-2	0.6, 0.7, 0.8	0.6, 0.7, 0.8	0.6, 0.7, 0.8	0.4, 0.5, 0.6
C-3	0.8, 0.9, 1.0	0.4, 0.5, 0.6	0.8, 0.9, 1.0	0.4, 0.5, 0.6
Mean values				
C-1	0.9	0.5	0.7	0.7
C-2	0.7	0.7	0.7	0.5
C-3	0.9	0.5	0.9	0.5

Table 2 shows the values of the candidates' distances from the best and worst alternative of the candidates, their similarity coefficients (CC_i), and rankings.

Table 2 – The distances from the best alternative and the worst alternative, the similarity coefficients, and the rankings

Candidates	d_i^*	d_i^-	CC_i	Ranking
C-1	0.063	0.061	0.493	1
C-2	0.072	0.050	0.413	2
C-3	0.071	0.051	0.417	3

Figure 1 shows the distances from the best and worst alternatives and the similarity coefficients of candidates with the black, red, and green lines, respectively.

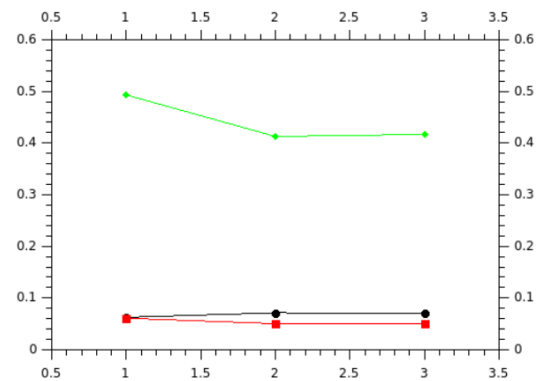


Figure 1 – The distances from the best alternative and the worst alternative and the similarity coefficients of candidates

In the second series of analyses with TOPSIS, we determined how the change in the criteria weights could affect the algorithm's output with the same entry data in the evaluation matrix. The alternative weights for each criterion are shown in Table 2. The recognition capability, cross-validation, sensitivity, and specificity weights for all the candidates were 0.9, 0.9, 0.9, and 0.2, respectively. These characteristics were considered as the profit criteria.

Table 3 shows the values of the candidates' distances from the best and worst alternatives of the candidates, their similarity coefficients (CC_i), and rankings.

Figure 2 shows the distances from the best and worst alternatives and the similarity coefficients of candidates with the black, red, and green lines, respectively.

Table 3 – The distances from the best alternative and the worst alternative, the similarity coefficients, and the rankings

Candidates	d_i^*	d_i^-	CC_i	Ranking
C-1	0.078	0.045	0.366	3
C-2	0.065	0.062	0.491	2
C-3	0.064	0.063	0.497	1

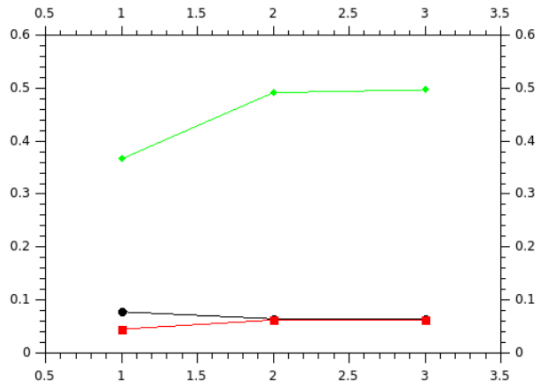


Figure 2 – The distances from the best alternative and the worst alternative and the similarity coefficients of candidates

Table 4 – Triangular fuzzy membership degrees of candidates' characteristics

Candidates/ Criteria	Recognition capability	Cross validation	Sensitivity	Specificity	Labor
Triangular fuzzy degrees					
C-1	0.8, 0.9, 1.0	0.4, 0.5, 0.6	0.6, 0.7, 0.8	0.6, 0.7, 0.8	0.6, 0.7, 0.8
C-2	0.6, 0.7, 0.8	0.6, 0.7, 0.8	0.6, 0.7, 0.8	0.4, 0.5, 0.6	0.8, 0.9, 1.0
C-3	0.8, 0.9, 1.0	0.4, 0.5, 0.6	0.8, 0.9, 1.0	0.4, 0.5, 0.6	0.1, 0.2, 0.3
Mean values					
C-1	0.9	0.5	0.7	0.7	0.7
C-2	0.7	0.7	0.7	0.5	0.9
C-3	0.9	0.5	0.9	0.5	0.2

Table 5 shows the values of the candidates' distances from the best and worst alternatives of the candidates, their similarity coefficients (CC_i), and rankings.

Table 5 – The distances from the best alternative and the worst alternative, the similarity coefficients, and the rankings

Candidates	d_i^*	d_i^-	CC_i	Ranking
C-1	0.092	0.046	0.334	3
C-2	0.104	0.053	0.337	2
C-3	0.055	0.104	0.656	1

Figure 3 shows the distances from the best and worst alternatives and the similarity coefficients of candidates with the black, red, and green lines, respectively.

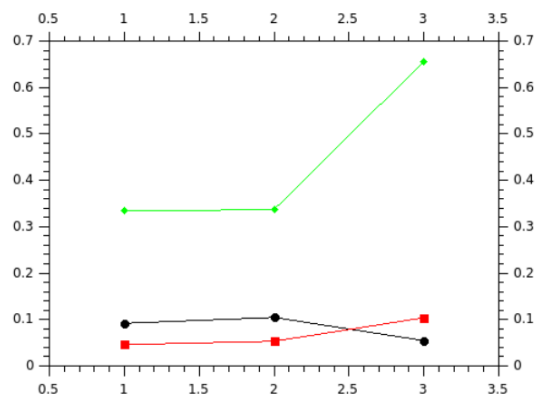


Figure 3 – The distances from the best alternative and the worst alternative and the similarity coefficients of candidates

In the third series of analyses, we added a cost criterion to the entry matrix. Recognition capability, cross-validation, sensitivity, and specificity were the profit criteria.

We considered labor as a cost criterion for the candidates.

The terms considered for the level of the candidates' criteria, such as recognition capability, cross-validation, sensitivity, specificity, and labor, were as follows: very high, medium, high, medium and height for C-1, high, medium, high, medium and very high for C-2, and very high, medium, very high, medium and low for C-3, respectively.

Table 4 shows the triangular fuzzy membership degrees of the candidates' characteristics and their mean values.

The alternative weight values of 0.9, 0.9, 0.9, 0.2, and 0.5 were assigned for recognition capability, cross-validation, sensitivity, specificity, and labor, respectively. In the criteria matrix, the terms "true" and "false" were used for the profit criteria and the cost criterion, respectively.

4 Discussion

The comparison of the obtained data in the tables and figures presented in this paper revealed that the change in the weights of alternatives could affect the ranking of algorithms.

Moreover, the ranking did not change after adding labor as a cost criterion to the list of criteria. This was because the GA, as candidate 3, had better fuzzy membership degrees than the two other candidates concerning their criteria, so the GA could be ranked in first place. In contrast, this algorithm was ranked in the second position before adding labor to the list of criteria. This showed the importance of the values of the fuzzy membership degrees of the cost criterion of algorithms in their ranks.

The obtained results show how the values of the fuzzy membership degrees of the algorithms used for the proteomic analysis could determine their priority according to their score differences. It is worth noting that the improvement of the algorithms SNN (candidate 1) and QC (candidate 2) could change the ranking when labor was considered as a cost criterion. For this purpose, it would be required to increase the recognition capability, cross-validation, sensitivity and specificity of the first and second candidates and reduce their labor. The change in the ranking of these algorithms

could add further information for the proteomic analysis. This issue will be addressed in a further investigation.

The future of proteomic analysis depends on technological developments to resolve current instrument limitations, the integration of biological and protein research approaches to investigate the biological sciences and molecular medicine, and the development of quantitative, tissue-specific proteomes for emphasizing low-abundance metabolic proteins, a direct look at an evolving disease process and extracting the whole factors that are involved in the analysis [23-25].

These developments will help recognize a specific process error before their potentially incorrect diagnosis [26]. The TOPSIS method can be used to predict disease stages in each investigation. The presented approach can also be used to analyze other materials [27-30]. As demonstrated in a series of studies, this approach is applicable to the selection of the analysis methods of biological materials such as DNA, RNA, etc. [31-33]. The same method can be applied to the investigation of the effects of their properties on their ranks [34-37].

Further investigations are required to improve these algorithms for the analysis of proteins, as their sensitivity and specificity could be improved for their applications in science and engineering.

References

1. Balioti, V., Tzimopoulos, C., Evangelides, C. (2018). Multi-criteria decision making using TOPSIS method under fuzzy environment. Application in spillway selection. *Proceedings*, 2, 637. <https://doi.org/10.3390/proceedings2110637>.
2. Tlas, M., Ghani, B. A. (2020). Interactive software for classification and ranking procedures based on multi-criteria decision-making algorithms. *Computational Ecology and Software*, 10(3), 133–150.
3. Rohanah, S. (2018). An evaluation of students performance using TOPSIS and Entropy approaches. *IOSR Journal of Research, Method in Education*, 8(6), 1–6. <https://doi.org/10.9790/7388-0806010106>.
4. Sałabun, W., Wątróbski, J., Shekhovtsov, A. (2020). Are MCDA methods benchmarkable? A comparative study of TOPSIS, VIKOR, COPRAS, and PROMETHEE II methods. *Symmetry*, 12(9), 1549. <https://doi.org/10.3390/sym12091549>.
5. Chang, S. - H., Tseng, H.-E. (2008). Fuzzy Topsis decision method for configuration management. *International Journal of Industrial Engineering*, 15(3), 304-313. <https://doi.org/10.23055/ijietap.2008.15.3.147>.
6. Bulgurcu, B. (2012). Application of TOPSIS technique for financial performance evaluation of technology firms in Istanbul stock exchange market. *Procedia*, 62, 1033-1040. <https://doi.org/10.1016/j.sbspro.2012.09.176>.
7. Kochkina, M. V., Karamyshev, A. N., Isavnin, A. G. (2017). Modified multi-criteria decision making method development based on “AHP” and “TOPSIS” methods using probabilistic interval estimates. *The Turkish Online Journal of Design, Art and Communication TOJDAC*, 1663-1674. <https://doi.org/10.7456/1070DSE/144>.
8. Abidin, M. Z., Rusli, R., Shariff, A. M. (2016). Technique for order performance by similarity to ideal solution (TOPSIS)- entropy methodology for inherent safety design decision making tool, *Procedia Engineering*, 148, 1043-1050. <https://doi.org/10.1016/j.proeng.2016.06.587>.
9. Álvarez, J. V., Bravo, S. B., Chantada-Vázquez, M. P., Barbosa-Gouveia, S., Colón, C., López-Suarez, O., Tomatsu, S., Otero-Espinar, F. J., Couce, M. L. (2021). Plasma proteomic analysis in Morquio A disease. *International Journal of Molecular sciences*, 22, 6165. <https://doi.org/10.3390/ijms22116165>.
10. Balbao, E., Marín, T., Oyarzún, J. E., Contreras, P. S., Hardt, R., van den Bosch, T., Alvarez, A. R., Rebolledo-Jaramillo, B., Klein, A. D., Winter, D., Zanlungo, S. (2021). Proteomic analysis of Niemann-pick type C hepatocytes reveals potential therapeutic targets for liver damage, *Cells*, 10, 2159. <https://doi.org/10.3390/cells10082159>.
11. Chen, Y., Yao, H., Zhang, N., Wu, J., Gao, S., Guo, J., Lu, X., Cheng, L., Luo, R., Liang, X., Wong, C. C. L., Zheng, M. (2021). Proteomic analysis identifies prolonged disturbances in pathways related to cholesterol metabolism and myocardium function in the COVID-19 recovery stage. *Journal of Proteome Research*, 20, 7, 3463-3474. <https://doi.org/10.1021/acs.jproteome.1c00054>.
12. Filbin, M. R., Mehta, A., Schneider, A. M., Kays, K. R., Guess, J. R., Gentili, M., Fenyves, B. G., Charland, N. C., Gonye, A. L. K., Gushterova, I., Khanna, H. K., LaSalle, T. J., Lavin-Parsons, K. M., Lilley, B. M., Lodenstein, C. L., et al. (2021). Longitudinal proteomic analysis of severe COVID-19 reveals survival-associated signatures, tissue-specific cell death, and cell-cell interactions. *Cell Reports Medicine*, 2, 100287. <https://doi.org/10.1016/j.xcrn.2021.100287>.

5 Conclusions

This paper investigated the prediction of machine learning algorithms for analyzing proteomic patterns with TOPSIS. The applied method consists of the calculations with the evaluation matrices, including the entry data for which the matrices of weight values and criteria were considered for different analyses. In the presented work, the algorithms previously used for the analysis of proteins were analyzed, and their rankings were compared. We used vector normalization in the TOPSIS method for the data normalization. The results of the two first series of the analysis showed that the third candidate could replace the first one in the rankings when the criteria weights were modified.

In addition, the third candidate could still maintain the first position in the ranking when labor was added as the cost criterion to the list of criteria. The obtained results showed that the optimization of the ranks could be affected by modifying the candidates' weights.

Optimizing the algorithms can lead to a better understanding of proteomic analysis with applications in science and engineering.

13. Swiatly, A., Horala, A., Hajduk, J., Matysiak, J., Nowak-Markwitz, E., Kokot, Z. J. (2017). MALDI-TOF-MS analysis in discovery and identification of serum proteomic patterns of ovarian cancer. *BMC Cancer*, 17, 472, 1–9.
14. Saghapour, E., Kermani, S., Sehhati, M. (2017). A novel feature ranking method for prediction of cancer stages using proteomics data. *Plos One*, 12(9), e0184203. <https://doi.org/10.1371/journal.pone.0184203>.
15. Mallik, S., Zhao, Z. (2019). Multi-objective optimized fuzzy clustering for detecting cell clusters from single-cell expression profiles. *Genes*, 10(8), 611. <https://doi.org/10.3390/genes10080611>.
16. Torng, W., Altman, R. B. (2017). 3D deep convolutional neural networks for amino acid environment similarity analysis. *BMC Bioinformatics*, 18, 302. <https://doi.org/10.1186/s12859-017-1702-0>.
17. Muzio, G., O'Bray, L., Borgwardt, K., Notes, A. (2021). Biological network analysis with deep learning. *Briefings in Bioinformatics*, 22(2), 1515-1530. <https://doi.org/10.1093/bib/bbaa257>.
18. Chiu, C. -Y., Hsieh, S. -Y., Wong, K. -S., Lai, S. -H., Chen, J. -K., Huang, J. -L. (2015). The value of total protein in guiding management of infectious parapneumonic effusion by using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Journal of Microbiology, Immunology and Infection*, 48(5), 483-489. <https://doi.org/10.1016/j.jmii.2013.11.013>.
19. Fan, N. -J., Gao, C. -F., Wang, X. -L., Zhao, G., Liu, Q. -Y., Zhang, Y. -Y., Cheng, B. -G. (2012). Serum peptidome patterns of colorectal cancer based on magnetic bead separation and MALDI-TOF mass spectrometry analysis. *J. Biomed. Biotechnol.*, 2012, 985020. <https://doi.org/10.1155/2012/985020>.
20. Unger, R. (2004). The Genetic Algorithm Approach to Protein Structure Prediction. *Structure and Bonding*, 110, 2697-2699. <https://doi.org/10.1007/b13936>.
21. Lv, Y. et al. (2006). Improved Genetic Algorithm for Multiple Sequence Alignment Using Segment Profiles (GASP). International Conference on Advanced Data Mining and Applications, In Lecture notes in computer science, 4093, 388-395. https://doi.org/10.1007/11811305_43.
22. Javanbakht, T., Chakravorty, S. (2022). Prediction of human behavior with TOPSIS. *Fuzzy Extension and Applications*, 3(2), 109-125. <https://doi.org/10.22105/jfea.2022.326185.1197>.
23. Bark, S. J., Hook, V. (2007). The future of proteomic analysis in biological systems and molecular medicine. *Mol. Biosyst.*, 3(1), 14-17. <https://doi.org/10.1039/b611446>.
24. Dastmalchi, M., Dhaubhadel, S. (2015). Proteomic insights into synthesis of isoflavonoids in soybean seeds. *Proteomics*, 15, 10, 1646-1657. <https://doi.org/10.1002/pmic.201400444>.
25. Spitzer, A. R., Chace, D. (2008). Proteomics- and metabolomics-based neonatal diagnostics in assessing and managing the critically ill neonate. *Clinic in Perinatology*, 35(4), 695-716. <https://doi.org/10.1016/j.clp.2008.07.019>.
26. Conrads, T. P., Fusaro, V. A., Ross, S., Johann, D., Rajapakse, V., Hitt, B. A., Steinberg, S. M., Kohn, E. C., Fishman, D. A., Whitely, G., Barrett, J. C., Liotta, L. A., Petricoin, E. F., Veenstra, T. D. (2004). High-resolution serum proteomic features for ovarian cancer detection, *Endocrine-Related Cancer*, 11, 163-178. <https://doi.org/10.1677/erc.0.0110163>.
27. Javanbakht, T., David, E. (2020). Rheological and physical properties of a nanocomposite of graphene oxide nanoribbons with polyvinyl alcohol. *Journal of Thermoplastic Composite Materials*, 0892705720912767. <https://doi.org/10.1177/0892705720912767>.
28. Javanbakht, T., Laurent, S., Stanicki, D., Frenette, M. (2020). Correlation between physicochemical properties of superparamagnetic iron oxide nanoparticles and their reactivity with hydrogen peroxide. *Canadian Journal of Chemistry*, 98, 601-608. <https://doi.org/10.1139/cjc-2020-0087>.
29. Javanbakht, T., Sokolowski, W. (2015). Thiol-ene/acrylate systems for biomedical shape-memory polymers. *Shape Memory Polymers for Biomedical Applications*, 157-166. <https://doi.org/10.1016/B978-0-85709-698-2.00008-8>.
30. Javanbakht, T., Ghane-Motlagh, B., Sawan, M. (2020). Comparative study of antibiofilm activity and physicochemical properties of microelectrode arrays. *Microelectronic Engineering*, 229, 111305. <https://doi.org/10.1016/j.mee.2020.111305>.
31. Pakpour, S., Olishvska, S., Prasher, S., Milani, A. S., Chénier, M. R. (2013). DNA extraction method selection for agricultural soil using TOPSIS multiple criteria decision-making model. *American Journal of Molecular Biology*, 3, 215-228. <https://doi.org/10.4236/ajmb.2013.34028>.
32. Tripathy, J., Dash, R., Pattanayak, B. K., Mishra, S. K., Mishra, T. K., Puthal, D. (2022). Combination of reduction detection using TOPSIS for gene expression data analysis, *Big Data and Cognitive Computing*, 6(1), 24. <https://doi.org/10.3390/bdcc6010024>.
33. Singh, S., Li, H. (2021). Comparative study of bioinformatic tools for the identification of chimeric RNAs from RNA Sequencing, *RNA Biology*, 18, S1, 254-267. <https://doi.org/10.1080/15476286.2021.1940047>.
34. Djavanbakht, T., Carrier, V., André, J. M., Barchewitz, R., Troussel, P. (2000). Effets d'un chauffage thermique sur les performances de miroirs multicouches Mo/Si, Mo/C et Ni/C pour le rayonnement X mou. *Journal de Physique IV, France*, 10, 281-287. <https://doi.org/10.1051/jp4:20001031>.
35. Krishnamoorthy, K., Mahalingam, M. (2015). Selection of a suitable method for the preparation of polymeric nanoparticles: Multi-criteria decision making approach. *Advanced Pharmaceutical Bulletin*, 5(1), 57-67. <https://doi.org/10.5681/apb.2015.008>.
36. Javanbakht, T., Hadian, H., Wilkinson, K. J. (2020). Comparative study of physicochemical properties and antibiofilm activity of graphene oxide nanoribbons. *Journal of Engineering Sciences*, 7(1), C1-C8. [https://doi.org/10.21272/jes.2020.7\(1\).c1](https://doi.org/10.21272/jes.2020.7(1).c1).
37. Negi, R. S., Bisht, R. S., Singh, R. K., Prasad, L. (2019). Physico-mechanical and antibacterial properties of pine gum/epoxy composites with/without silver nanoparticles. *Macromolecular Materials and Engineering*, 304, 7, 1800744. <https://doi.org/10.1002/mame.201800744>.