TELKOMNIKA Telecommunication, Computing, Electronics and Control
Vol. 18, No. 3, June 2020, pp. 1643~1649
ISSN: 1693-6930, accredited First Grade by Kemenristekdikti, Decree No: 21/E/KPT/2018
DOI: 10.12928/TELKOMNIKA.v18i3.14847

The comparison study of kernel KC-means and support vector machines for classifying schizophrenia

Sri Hartini, Zuherman Rustam
Department of Mathematics, University of Indonesia, Indonesia

Article Info

Article history:

Received Aug 15, 2019
Revised Jan 4, 2020
Accepted Feb 23, 2020

Keywords:

Fast fuzzy clustering
KC-means
Kernel function
Schizophrenia classification
Support vector machines

ABSTRACT

Schizophrenia is one of mental disorder that affects the mind, feeling, and behavior. Its treatment is usually permanent and quite complicated; therefore, early detection is important. Kernel KC-means and support vector machines are the methods known as a good classifier. This research, therefore, aims to compare kernel KC-means and support vector machines, using data obtained from Northwestern University, which consists of 171 schizophrenia and 221 non-schizophrenia samples. The performance accuracy, F1-score, and running time were examined using the 10-fold cross-validation method. From the experiments, kernel KC-means with the sixth-order polynomial kernel gives 87.18 percent accuracy and 93.15 percent F1-score at the faster running time than support vector machines. However, with the same kernel, it was further deduced from the results that support vector machines provides better performance with an accuracy of 88.78 percent and F1-score of 94.05 percent.

Corresponding Author:

Sri Hartini,
Department of Mathematics,
University of Indonesia, Depok 16424, Indonesia,
Email: sri.hartini@sci.ui.ac.id

1. INTRODUCTION

Schizophrenia is a lifelong mental disorder that induces the mind, feeling, and behavior [1]. It is characterized by unusual thoughts or experiences, disorganized speech patterns, and decreased participation in social life [2]. The World Health Organization stated that schizophrenia is not a common mental disorder; however, it affects more than 23 million people worldwide [3]. Approximately 220,000 people in England and Wales are suffering from this ailment [4]. Moreover, from the study of one million six hundred thousand person includes 568 subjects with clinically relevant psychotic syndromes, this mental disorder is more common among males than females [5].

Treatment of this mental disorder is usually permanent and often involves a combination of medications, psychotherapy, and coordinated particular care services. Therefore, it could be better assuming someone suffering this ailment is detected early to provide adequate family time to deliberate on a bunch of treatments that could be offered by this person. Various studies have already been developed in many ways. For instance, the linear support vector machines was implemented to detect the first episode of schizophrenia-spectrum disorder (FES) in patients and to differentiate those that are healthy with an accuracy of 62.34 percent [6]. In another study, the multilayer perceptron neural network (MLP NN) algorithm obtained 82 percent accuracy using the public gene expression omnibus (GEO) genome-wide expression dataset [7].

There were also k-NN with 85.7 percent accuracy, decision tree with 62.50 percent accuracy, and Naive Bayes with an accuracy of 87.71 for schizophrenia dataset which including in the latter dataset [8]. Moreover, with a total of 212 participants which consist of 141 schizophrenia patients and 71 healthy controls, the regularized SVM model gives accuracy of 86.6% in the training set of 127 individuals, meanwhile validation accuracy of 83.5 percent in an independent set of 85 people [9]. Meanwhile, SVM was able to classify the schizophrenia dataset with an accuracy of 90.1 percent and 95.0 percent in at least one simulation using linear and Gaussian kernel [10]. Mean accuracy of more than 90 percent was also obtained using Twin SVM with linear and Gaussian kernel [11]. According to these previous research, it was deduced that support vector machines is one of the well classifiers. As the other method, there is also kernel KC-means that known is better than the former KC-means in the classification task using the cluster analysis. This research, therefore, aims to expand the previous research by comparing the performance of kernel KC-means and support vector machines, both with the polynomial kernel function, using the same dataset as the last two methods.

2. RESEARCH METHOD

2.1. Material

In this research, the Northwestern University Schizophrenia dataset [12], consisting of 171 schizophrenia and 221 non-schizophrenia samples was used. The instances are described by 64 features, which were numerically and categorically characterized. Detailed information about these features are shown in Table 1.

Table 1. Northwestern University chizophrenia dataset features

Feature number	Feature name	Description
1–34	Scale for the Assessment of Positive Symptoms (SAPS) 1–34	The answer for questionnaire SAPS which scale from 0 to 5
35–59	Scale for the Assessment of Negative Symptoms (SANS) 1–25	The answer for questionnaire SANS which scale from 0 to 5
60	Gender	The gender of the patient (male or female)
61	Dominant Hand	Left or right hand
62	Race	Caucasian, African, American or other
63	Ethnic	Caucasian, African-American, or other
64	Age	Age of the patient

2.2. Methods

This research makes use of two methods, namely kernel KC-means and support vector machines. The performances of these two methods are evaluated using 10-fold cross-validation. The dataset was divided randomly into ten mutually exclusive folds with an approximately equal number of samples [13]. Each fold then was taken as the validation data for testing the classification model, while the rest was taken for building model and has a role as the training data. This process was repeated until each fold have already been validation data, which was beneficial to the repeated random sub-sampling [14]. The use of the cross-validation approach in evaluating classifiers is considered due to its ability to avoid pitfalls in comparing these two methods [15].

2.2.1. Kernel function

The real-world applications often need more than linear functions. Therefore, as an alternative solution, kernel offers new representation by projecting the data into feature space that has higher dimension for the computational time of linear learning machines to be increased [16]. The kernel function for every $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ is defined in (1) [17].

$$\mathrm{K}(x, y) = (\phi(x), \phi(y)) \tag{1}$$

where ϕ is the mapping function from input data to the feature space. Besides, the distance between two mapped points in kernel representation is defined in (2) [17].

$$\mathrm{d(x, y)} = \sqrt{\mathrm{K}(x, x) - 2\mathrm{K}(x, y) + \mathrm{K}(y, y)} \tag{2}$$

This formula is substituted to replace the Euclidean distance commonly used.

There are several types of kernel functions, including linear, polynomial, Gaussian radial basis function (RBF), and sigmoid kernel function, with formulas are shown in Table 2. However, this research uses the polynomial kernel function with the polynomial order h from 1 to 10. Based on Liu et al. [18], this kind of kernel function is appropriate for conditions where all the training data normalized.

Table 2. Kernel functions

	Name	Formula
1.	Linear kernel function	$\mathrm{K}(x,y) = x \,.\, y$
2.	Polynomial kernel function	$\mathrm{K}(x,y) = (x \,.\, y + 1)^h$
3.	Gaussian Radial Basis Function (RBF) kernel function	$\mathrm{K}(x,y) = \exp\left(-\frac{\lVert x - y \rVert^2}{2\sigma^2}\right)$
4.	Sigmoid kernel function	$\mathrm{K}(x,y) = \tanh(\kappa x \,.\, y - \delta)$

2.2.2. Kernel KC-means

This method is a combination of K-Means, Fuzzy C-Means algorithm, and kernel function [19] with the purpose to makes the running time of fuzzy c-means faster with the same performance when only use its method. The kernel KC-means is also known as fast fuzzy clustering based on kernel. Kernel KC-means is a modified version of KC-means [20] proposed by Atiyah et al. Moreover, we have to minimize k-means and fuzzy c-means objective function, with the modification on the distance definition. Assume x_i is the i-th sample, and $\boldsymbol{v_j}$ is the j-th centroid. When applying k-means using kernel function, then our goal is to optimize the objective function shown in (3) [17].

$$\mathrm{J} = \sum_{i=1}^{n} \sum_{j=1}^{c} \mathrm{r}_{ij} \left(\mathrm{K}(x_i, x_i) - 2\mathrm{K}(x_i, v_j) + \mathrm{K}(v_j, v_j) \right) \tag{3}$$

where

$$\mathrm{r}_{ik} = \begin{cases} 1 & , \text{if } \mathrm{k} = \arg\min \mathrm{d}^2(\mathrm{x}_i, \mathrm{v}_j) \\ 0 & , \text{otherwise} \end{cases} \tag{4}$$

Meanwhile, when applying fuzzy kernel c-means, our goal is to optimize the objective function (5) [21]:

$$\mathrm{J}_m = \sum_{i=1}^{n} \sum_{j=1}^{c} (\mathrm{u}_{ij})^m \left[\mathrm{K}(x_i, x_i) - 2\mathrm{K}(x_i, v_j) + \mathrm{K}(v_j, v_j) \right] \tag{5}$$

where u_{ij} is the membership value of the i-th sample in the j-th cluster that satisfies these two conditions:

$$\sum_{j=1}^{c} \mathrm{u}_{ij} = 1, \; i = 1, 2, \dots, n \tag{6}$$

$$0 < \sum_{i=1}^{n} \mathrm{u}_{ij} < n, j = 1, 2, \dots, c \tag{7}$$

with $c < n$ being a positive integer. The algorithm of this method is given in Figure 1.

2.2.3. Support vector machines

This method is appropriate to uses when there are precisely two classes. According to Vapnik et al. [22], data points are considered as support vectors, and the goal, therefore, is to find the best hyperplane that able t separate them into classes. The scheme of this method is seen in Figure 2 [23], where H is the hyperplane, W is the normal vector to the hyperplane, while m is the minimum distance between positive and negative hyperplanes.

Moreover, the kernel approach is used to resolve a simple hyperplane that is useful in a classification problem with only two classes. In other words, this method brings the form of mapping input into a space that has higher dimension to support nonlinear classification problems where the hyperplane causes the maximum separation between each class [24]. Therefore, the optimization problem observed is defined to minimize the following function [25] while determining the value of weight $\mathrm{w} \in \mathbb{R}^{\mathrm{n}}$ and the bias $\mathrm{b} \in \mathbb{R}^{\mathrm{n}}$.

$$\frac{1}{2} \lVert w \rVert^2 \tag{8}$$

$$\text{s.t} \quad y_i(w^T \cdot x_i + b) \geq 1 \quad , \forall i = 1, 2, \dots N \tag{9}$$

Then the decision function to optimize the margin is defined as (10) [25]:

$$f(x) = \text{sign}(w \cdot x + b) \tag{10}$$

where the value of w and b is obtained using the formula in (11) and (12), respectively.

$$w = \sum_{i=1}^{N} a_i y_i x_i \tag{11}$$

$$b = \frac{1}{N_s} \sum_{i \in S} (y_i - \sum_{m \in S} a_m y_m x_m) \tag{12}$$

with S is the vector set where $a_i \neq 0$ for every $i \in S$ and $N_s = |S|$. This method is still considered due to its excellent performance that does not only maximizes margins but also minimizes existing errors [26].

Input: dataset $X = \{x_1, x_2, \dots, x_n\}$, the number of clusters c, the maximum number of iterations allowed T, and ε $(0 < \varepsilon < 1)$
Output: $V = \{v_1, v_2, \dots v_c\}$, $U = [u_{ij}]$, $1 \leq i \leq n$, $1 \leq j \leq c$

1. Initialization: $V^0 = \{v_1, v_2, \dots v_c\}$;
2. Compute the value of $d(x_i, v_j)$ using equation (2) for $1 \leq i \leq n$ and $1 \leq j \leq c$;
3. Assign the elements x_i to the clusters according to
$$r_{ik} = \begin{cases} 1 & , \text{if } k = \arg\min d^2(x_i, v_j) \\ 0 & , \text{otherwise} \end{cases}$$
4. Update the cluster centers by taking $t = t + 1$ and
$$v_j^{(t)} = \frac{\sum_{i=1}^{n} r_{ik} x_i}{\sum_{i=1}^{n} r_{ik}}, \quad j = 1, 2, \dots, c$$
5. If $\|V^{(t-1)} - V^{(t)}\| < \varepsilon$ or $T = t$, then go to the next step. Otherwise, go back to step (2);
6. Let V be the set of final centers of clusters obtained from step 1 to 5;
7. Set V as the new dataset;
8. Compute the membership and assign the elements x_i to the clusters according to
$$u_{ij}^{(t)} = \frac{1}{\sum_{k=1}^{c} \left(\frac{d(x_i, v_j)}{d(x_i, c_k)} \right)^{\frac{2}{m-1}}} \quad \text{where } m \neq 1, 1 \leq i \leq n \text{ and } 1 \leq j \leq c;$$
9. Update the cluster centers according to
$$v_j^{(t)} = \frac{\sum_{i=1}^{n} (u_{ij})^m x_i}{\sum_{i=1}^{n} (u_{ij})^m}, \quad j = 1, 2, \dots, c$$
10. Recover corresponding data set clusters based on the result of the step before;

End

Figure 1. Kernel KC-means algorithm

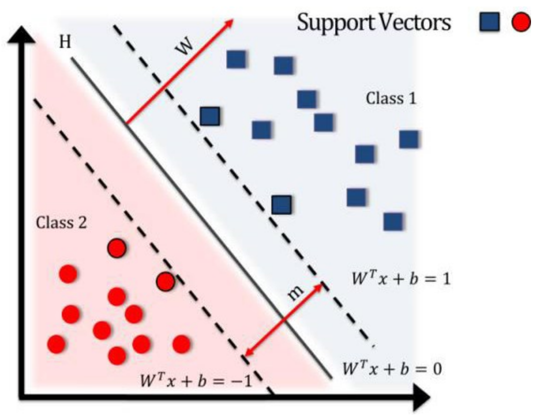

Figure 2. Support vector machines classification scheme

2.2.4. Performance measure

This research uses accuracy and F1-score to measure performance based on the result of the confusion matrix. However, the proper use of metrics from the confusion matrix is necessary because it affects all statistical comparison metrics [27]. Therefore, accuracy is a subjective measure used to evaluate a classifier on a set of test data with the classifier's prediction correctly divided by the total number of instances [28]. Meanwhile, F1-score, which considers both the precision and sensitivity, is used to distinguish the correct prediction of labels within different classes [29]. These two formulas are shown in (13) and (14):

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \tag{13}$$

$$\text{F1} - \text{Score} = \frac{2 * \text{sensitivity} * \text{precision}}{\text{sensitivity} + \text{precision}} \tag{14}$$

with sensitivity and precision is given in (15) and (16)

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP+FN}} \tag{15}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{16}$$

TP is the count of schizophrenia samples correctly diagnosed, FP is the number of non-schizophrenia samples incorrectly diagnosed, TN is the number of non-schizophrenia samples correctly diagnosed, and FN is the total of schizophrenia samples incorrectly diagnosed [30].

3. RESULTS AND ANALYSIS

The performance of both kernel KC-means and support vector machines using polynomial kernel were examined with k-fold cross-validation where $k = 10$. In 10-fold cross-validation, ten folds with approximately equal size were formed from the dataset after randomly divided [13]. Furthermore, one fold is the validation data, while the rest is training data. This process repeats for ten times, which gives advantages of over repeated random sub-sampling [14]. The performance of kernel KC-means using polynomial kernel is shown in Table 3. In this table, the running time varied and was conducted under or in 0.1 seconds. Therefore, it is concluded that the highest value of accuracy and F1-score is obtained when the sixth-order polynomial kernel is utilized. Meanwhile, the performance of support vector machines using polynomial kernel is shown in Table 4. We can see in this table that the accuracy and F1-score is not as various as kernel KC-means.

Table 3. The performance of kernel KC-means with polynomial kernel

Polynomial Degree	Accuracy (%)	F1-Score (%)	Running Time (s)
1	74.36	85.29	0.10
2	72.50	84.06	0.05
3	72.50	84.06	0.02
4	76.92	86.96	0.03
5	82.05	90.14	0.04
6	87.18	93.15	0.02
7	71.79	83.58	0.03
8	76.92	86.96	0.04
9	76.92	86.96	0.02
10	66.67	80.00	0.03

Table 4. The performance of support vector machines with polynomial kernel

Polynomial Degree	Accuracy (%)	F1-Score (%)	Running Time (s)
1	88.78	94.05	0.55
2	88.78	94.05	0.15
3	88.78	94.05	0.15
4	88.78	94.05	0.15
5	88.78	94.05	0.41
6	88.78	94.05	0.09
7	88.27	93.77	0.08
8	88.52	93.91	0.08
9	88.52	93.91	0.10
10	88.27	93.77	0.08

The accuracy and F1-score of this method are higher with 88.78 and 94.05 percent, respectively, in the sixth-order polynomial kernel. Besides, the fastest running time and the best performance is yielded when the sixth-order polynomial kernel. Therefore, the best performance of kernel KC-means and support vector machines which were obtained when the sixth-order polynomial kernel is compared in Table 5. In this table, we can conclude that support vector machines give better performance because of the accuracy and F1-score are 1.6% and 0.9% higher, respectively, than the performance of kernel KC-means. However, KC-means delivers the performance in faster running time than support vector machines. Therefore, both methods deliver excellent results with the KC-means producing better running time.

Table 5. The comparison between kernel KC-means and support vector machines with the sixth-order polynomial kernel function

Method	Accuracy (%)	F1-Score (%)	Running Time (s)
Kernel KC-means	87.18	93.15	0.02
Support Vector Machines	88.78	94.05	0.09

4. CONCLUSION

Treatment associated with a mental disorder, especially Schizophrenia, should be administered early. The medication is necessary because this mental disorder affects behavior, mind, and feeling. This research uses data which consists of 171 schizophrenia and 221 non-schizophrenia samples from Northwestern University to compare the performance of kernel KC-means and support vector machines in the schizophrenia classification task. From the experiment, kernel KC-means was found to give better running time with an accuracy of 87.18 percent and F1-score of 93.15 percent, which is lower than the support vector machines. The latter method still gives better accuracy and F1-score with 88.78 and 94.05 percent, respectively.

In this study, kernel KC-means, as well as support vector machines, provided high performance. However, there is space for improvement. For future researches, it was then encouraged to explore the possibility of constructing new models to obtain better performance, especially considering that polynomial kernel in both methods did not deliver above 90 percent of accuracy. It is also possible to implement these methods in other datasets; therefore, the model is examined in different ways with the result that it would be considered as the appropriate method which gives an accurate diagnosis.

ACKNOWLEDGEMENTS

This research was full supported financially by University of Indonesia, with a DRPM PUTI Q2 2020 research grant scheme. The authors would like to appreciate all the reviewers involved in the improvement of this article.

REFERENCES

[1] Wang W. C., Lu M. L., Chen V. C. H., Ng M. H., Huang K. Y., Hsieh M. H., Hsieh M. J., McIntyre R. S., Lee Y., Lee C. T. C, "Asthma, corticosteroid use, and schizophrenia: A nationwide population-based study in Taiwan," *PLoS One*, vol. 12, no. 3, pp. 1-12, 2017.
[2] The National Institute of Mental Health. [Online]. Available: https://www.nimh.nih.gov/. Accessed: 27 June 2019.
[3] World Health Organization. [Online]. Available: https://www.who.int/. Accessed: 27 June 2019.
[4] Royal College of Psychiatrists, "Report of the National Audit of Schizophrenia (NAS) 2012,"London: *Healthcare Quality Improvement Partnership*, pp. 23, 2012.
[5] Kirkbride J. B., Fearon P., Morgan C., et al., "Heterogeneity in incidence rates of schizophrenia and other psychotic syndromes: Findings from the 3-center AeSOP study," *Arch Gen Psychiatry*, vol. 63, no. 3, pp. 250-258, 2006.
[6] Mikolas P., Hlinka J., Skoch A., Pitra Z., Frodl T., Spaniel F., Hajek T., "Machine learning classification of first-episode schizophrenia spectrum disorders and controls using whole brain white matter fractional anisotropy," *BMC Psychiatry*, vol. 18, no. 97, pp. 1-7, 2018.
[7] Yilacioglu K., Konuk M., "Classification of schizophrenia patients by using genomic data: A data mining approach," *The Journal of Neurobehavioral Sciences*, vol. 2, no. 3, pp. 1-3, 2015.
[8] Saylan C. C., Yilancioglu K., "Classification of schizophrenia and bipolar disorder by using machine learning algorithms," *The Journal of Neurobehavioral Sciences*, vol. 3, no. 3, pp. 92-95, 2016.
[9] Chin R., You A. X., Meng F., Zhou J., Sim K., "Recognition of schizophrenia with regularized support vector machine and sequential region of interest selection using structural magnetic resonance imaging," *Scientific Reports*, vol. 8, pp. 1-10, 2018.
[10] Rampisela T. V., Rustam Z., "Classification of schizophrenia data using support vector machine (SVM)," *IOP Conf. Series: Journal of Physics: Conf. Series,* vol. 1108, pp. 1-7. 2018.
[11] Rustam Z., Rampisela T. V., "Support vector machines and twin support vector machines for classification of schizophrenia data," *International Journal of Engineering and Technology*, vol. 7, no. 4, pp. 6873-6877, 2018.
[12] Wang L., Kogan A., Cobia D., Alpert K., Kolasny A., Miller M. I., Marcus D., "Northwestern University Schizophrenia Data and Software Tool (NUSDAST)," *Frontiers in Neuroinformatics*, vol. 7, no. 25, 2013.

[13] Kohavi R., "A study of cross-validation and bootstrap for accuracy estimation and model selection," *Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI). Montreal*, vol. 2, pp. 1137-1143, 1995.
[14] Raju S. K., Murty M. R., Rao M. V., Satapathy S. C., "Support vector machine with k-fold cross validation model for software fault prediction," *International Journal of Pure and Applied Mathematics*, vol. 118, no. 20, pp. 321-334, 2018.
[15] Salzberg S. L., "On comparing classifiers: Pitfalls to avoid and a recommended approach," *Data Mining and Knowledge Discovery*, vol. 1, pp. 317-328, 1997.
[16] Cristianini N., Shawe Taylor J., "An introduction to support vector machines and other kernel-based learning methods," Cambridge: *Cambridge University Press*, 26-51, 2014.
[17] Bishop C. M., "Pattern recognition and machine learning," New York: *Springer*, 291-299, 2006.
[18] Liu L., Shen B., Wang X., "Research on kernel function of support vector machine. In: Huang YM, Chao HC, Deng DJ, Park J. *Editors*. Advanced Technologies, Embedded, and Multimedia for Human-centric Computing," *Lecture Notes in Electrical Engineering, Springer*, vol. 260, 2014.
[19] Rustam Z., Hartini S., "Classification of breast cancer using fast fuzzy clustering based on kernel," *IOP Conf. Series: Materials Science and Engineering*, vol. 546, pp. 1-10, 2019.
[20] Atiyah I. A., Mohammadpour A., Taheri S. M., "KC-means: A fast fuzzy clustering," *Hindawi: Advance in Fuzzy Systems*, pp. 1-8, 2018.
[21] Bezdek J. C., "Pattern recognition with fuzzy objective function algorithms," New York: *Plenum Press*, pp. 174-191. 1981.
[22] Stitson M. O., Weston J. A. E., Gammerman A., Vovk V., Vapnik V., "Theory of support vector machines," *Technical Report CSD-TR-96-17*. Royal Holloway, University of London. 1996.
[23] Leal N., Leal E., Sanchez G., "Marine vessel recognition by acoustic signature,"*ARPN Journal of Engineering and Applied Sciences*, vol. 10, no. 20, pp. 9636, 2015.
[24] Arfiani, Rustam Z., Pandelaki J., Siahaan A., "Kernel spherical k-means and support vector machine for acute sinusitis classification*," IOP Conf. Series: Materials Science and Engineering*, no. 546, pp. 1-10, 2019.
[25] Rustam Z., Zahras D., "Comparison between support vector machines and fuzzy c-means as classifier for intrusion detection system," *IOP Conf. Series: Journal of Physics: Conf. Series*," no. 1028 pp. 1-7, 2018.
[26] Rustam Z., Syarifah M. A., Siswantining T., "Recursive particle swarm optimization (RPSO) schemed support vector machine (SVM) implementation for microarray data analysis on chronic kidney disease (CKD)," *IOP Conf. Series: Materials Science and Engineering*, no. 546, pp. 1-7, 2019.
[27] Hossin M., Sulaiman M. N., "A review on evaluation metrics for data classification evaluations," *International Journal of Data Mining & Knowledge Management Process (IJDKP)*, vol. 5, no. 2, pp. 1-11, 2015.
[28] Parker C., "An analysis of performance measures for binary classifiers," *IEEE International Conference on Data Mining,* pp. 517-526, 2011.
[29] Sokolova M., Japkowicz N., and Szpakowicz S., "Beyond accuracy, F-score and ROC: A family of discriminant measures for performance evaluation. In: Sattar A, Kang B. *Editors*. AI 2006: Advances in Artificial Intelligence. AI 2006. *Lecture Notes in Computer Science*. vol. 4304. Berlin, Heidelberg: Springer, 2006.
[30] Rahideh A., Shaheed M. H., "Cancer classification using clustering based gene selection and artificial neural networks," P*roceeding of the 2nd International Conference on Control, Instrumentation and Automation (ICCIA)*, pp. 1175-1180, 2011.