ASM Sc. J., 12, Special Issue 5, 2019 for ICoAIMS2019, 190-198

Generalized Mean Distance-based k Nearest Centroid Neighbor Classifier

Nordiana Mukahar[1,2*] and Bakhtiar Affendi Rosdi[1]

[1]*Intelligent Biometric Group, School of Electrical and Electronic Engineering, Universiti Sains Malaysia*

[2]*Faculty of Electrical Engineering, Universiti Teknologi Mara*

The classic k nearest neighbor (kNN) is a well-known non-parametric classif8er used in the pattern recognition task. Nonetheless, the sensitivity towards neighborhood size, k and outliers' sample will seriously deteriorate the classification performance of the kNN classifier. The selection of nearest neighbor in the kNN is based on the closeness towards the test sample and it lacks information in terms of geometrical distribution. In this paper, a generalized mean distance-based k nearest centroid neighbor classifier (GMDkNCN) is proposed by adopting the concept of generalized mean using centroid distance. In order to capture more class information, the k local mean vector of each class is used, and k generalized mean centroid distances are calculated using k local mean vectors per class. The categorical nested generalized mean centroid distance of each class is then developed for a class decision of a new coming test sample. The advantage of adopting k generalized mean centroid distances and categorical nested generalized mean centroid distance of each class allowing to exploit more class information from local mean centroid vectors with different weighted contributions. In this way, the test sample is assigned to the class with the minimum nested generalized mean centroid distance. Experiments are performed on the eighteen UCI data sets and compared the proposed method to the three state-of-art kNN and two k nearest centroid neighbor-based classifiers. The experimental results demonstrate that the proposed GMDkNCN performs better than five competing classifiers with robust classification performance.

Keywords: k nearest neighbor; k nearest centroid neighbor; generalized mean centroid distance; pattern recognition

I. INTRODUCTION

Non-parametric classifiers provide simple and effective strategies in pattern classification. Especially, the conventional k nearest neighbor (kNN) (Stevens *et al.*, 1967) has been widely used in the supervised learning that involves dataset to predict the class of new coming input sample with unknown class. It has been well known that the kNN classifier offers several benefits in classification such as the simplicity in implementation, effectiveness, and intuitiveness. Even though the kNN hold several significant advantages, yet it still presents some issues to solve.

The first issue is related to the implementation of simple majority vote in the kNN in which misclassification might occur particularly on minority classes. It is because they often sacrifice minority classes and embrace majority classes. All nearest neighbors in the kNN are treated with equal importance regardless its distance to the test sample to be classified and this can be a problem when a closer neighbor is more reliable in the class prediction of the test sample.

Next issue is related to the distance metric to measure the similarity between two samples. In the kNN rule, the closeness of two samples is commonly measured using the Euclidean distance metric and the value returned by similarity metrics may be affected by such noisy samples. It treats all samples equally, and therefore, the selection of neighbors in the kNN is sensitive to outlier or noisy samples (Fukunaga, 1990; Ma *et al.*, 2016). For a test sample that lies in the areas with high concentrated

*Corresponding author's e-mail: nordi741@tganu.uitm.edu.my

distribution, its neighborhood is dominated by the class with higher density than the other classes kNN may lead to misclassification and it shows that the kNN is sensitive to the variance of the distribution.

Several weighting strategies (Gou *et. al.*, 2012; Keller *et. al.*, 1985; Derrac *et al.*, 2016) have been proposed by researchers in attempt to solve the drawback in the kNN classifier that assume nearest neighbors have equal importance with identical weight. In order to overcome the influence of outliers, a local mean-based strategy classifier has been introduced (Zeng *et. al.*, 2009; Mitani *et. al.*, 2006; Gou *et al.*, 2014). In local mean-based approach, the distance is measured between test sample and k local mean vector of testing samples of each class. Experimental results of the local mean-based kNN show significant performance in response to existing outliers and less sensitive to the choice of k. The local mean-based concept is further extended with different distance metrics such as harmonic mean and generalized mean (Pan *et. al.*, 2017; Gou *et al.*, 2019) to give appropriate weight on more reliable local mean vector. Generalized mean distance with local mean concept (Gou *et al.*, 2019) introduces controllable variable, p to enhance the positive influence from the neighbors and it gives significant effect to the classification performance. The first step in local mean-based kNN is to obtain k nearest neighbors of each class using Euclidean distance. Nevertheless, the value returned by similarity metric of Euclidean distance may be affected by such noisy samples. It treats all samples equally, and therefore, the selection of k nearest neighbors in the local mean-based kNN is sensitive to outlier or noisy (Fukunaga, 1990; Ma *et al.*, 2016).

The nearest centroid neighborhood concept (Chaudhuri, 1996) defines the neighborhood by taking into account not only the proximity of prototypes to a given input sample but also their symmetrical distribution around it. The k Nearest Centroid Neighbor (kNCN) classifier was proposed by Sanchez et al. (Sanchez, 1997) to overcome the deficiency in the classic kNN classifier. Instead of finding the nearest distance of nearest neighbors, the kNCN finds nearest centroid neighbors. Similar to kNN, the kNCN apply majority votes to select nearest centroid neighbor for classification decision. Several works using the kNCN-based classifiers (Gou *et al.*, 2012; Jaafar *et al.*, 2016; Sanchez *et al.*, 1998) have shown that the kNCN achieves good accuracy than the kNN-based classifiers in practice. Similar to kNN, kNCN makes the inappropriate assumption that k centroid neighbors have an identical weight. It overlooks the importance of some neighbors which are not close to the test sample which beneficial for more accurate classification.

This paper proposes generalized mean distance k nearest centroid neighbor (GMDkNCN) classifier that efficiently combine the strategy of local mean based on centroid distance (Basche & Lichman, 2013) with the generalized distance (Gou *et. al.*, 2019), taking the benefit of the generalized distance and local mean- concept to improve the classification performance of the kNCN.

The rest of this paper is structured as follows. Section 2 presents related classifiers. Section 3 introduces the proposed GMDkNCN classifier. In Section 4, experimental data set and the comparative results between the proposed method and the competing classifiers will be described. Finally, Section 5 presents concluding remarks.

II. RELATED WORKS

This section elaborates the related classifiers which motivate to the development of the GMDkNCN classifier in this work. These classifiers are the kNCN and generalized mean distance k nearest neighbor (GMDkNN).

A. k Nearest Centroid Neighbor Classifier

The kNCN classifier uses centroid distance as a distance metric to select nearest neighbors. According to the kNCN classifier, the nearest neighbors must be closed enough to the test sample and symmetrically distributed around it (Sanchez *et al.*, 1997). Given the testing sample, x, and a set of training samples $T = \{y_i | y_i \in R^p\}_{i=1}^{N}$ with N training samples in a p -dimensional feature space. $C = \{c_i | c_i \in \{c_1, c_2, \dots, c_m\}\}_{i=1}^{N}$ denotes the corresponding class label set of T, where y_i has the label c_i and m represents the number of classes. The nearest centroid neighbors of a test sample, x is searched iteratively in four steps.

(i) Calculate the distances of training samples, y_i to x using equation 1. The first nearest centroid neighbor, $y_{ncn,1}$ is the closest sample to the x and it is included in the set of k nearest centroid neighbors, $NCN_k(x) = \{y_{ncn,r}\}_{r=1}^{k}$.

$$d(x, y_i) = \sqrt{(x - y_i)^T (x - y_i)},\ 1 \le i \le N \qquad (1)$$

(ii) Find the centroid point, y^c_{rj} which is the mean of selected nearest centroid neighbor(s) and the remaining samples using equation 2.

$$y^c_{rj} = \frac{\sum_{r=1}^{r-1} y_{ncn,r} + y_i \in T - \{y_{ncn,r}\}}{r} \quad (2)$$

(iii) Calculate the centroid-distances, $d(x, y^c_{rj})$between y^c_{rj} to x using equation 3. The nearest centroid distance to x is selected as r -th nearest centroid neighbor, $y_{ncn,r}$.

$$d(x, y^c_{rj}) = \sqrt{(x - y^c_{rj})^T (x - y^c_{rj})} \quad (3)$$

(iv) Assign the class, c with majority vote to x.

$$c_x = \arg\max_{c_j} \left\{ \sum_{(y_{ncn,i}, \omega_i^{ncn}) \in NCN_k(x)} \delta(c_j = \omega_i^{ncn}) \right\} \quad (4)$$

$$1 \le r \le k \text{ and } 1 \le j \le m$$

Where $\delta(c_j = \omega_i^{ncn})$ is the Kronecker delta function with the value of one if $c_j = \omega_i^{ncn}$ or zero otherwise

B. Generalized Mean Distance k Nearest Neighbor Classifier

Generalized mean distance k nearest neighbor is introduced to give more appropriate weighted classification contributions from different neighbors (Gou *et al.*, 2019). In addition, the negative influence of outliers is suppressed by adjusting the controllable variable p to elevate the positive influence from the nearest neighbor. In this way, more weight are given to the closer neighbor and the strength of weight classification can be dynamically adjusted through different values of p.

Consider the set of k nearest neighbors is indicated as $y_j^{NN} = \{y_{i,j}^{NN} \in R^p\}_{j=1}^k$ from class c_i. The generalized mean distances of the first r distances from the first r k nearest neighbors to the test sample, x from class c_i is obtained through equation 5.

$$g(x, U_{i,r}^{NN}) = (1/r \sum_{j=1}^{r} (d((x, y_{i,j}^{NN}))^p))^{1/p} \quad (5)$$

The set of k generalized mean distance from class c_i is then represented as $g^i = g(x, U_{i,1}^{NN}, x, U_{i,2}^{NN}, \dots, x, U_{i,kj}^{NN})$. Subsequently, the nested generalized mean distance based on the k generalized mean distance for class c_i is calculated using equation 6 where $i = 1,2, \dots, m$.

$$G(x, g^i) = (1/k \sum_{r=1}^{k} (g((x, y_{i,r}^{NN}))^p))^{1/p} \quad (6)$$

In this way, more weights are given to the closer neighbor and the strength of weight classification can be dynamically adjusted through different values of p.

III. THE PROPOSED GENERALIZED MEAN DISTANCE k NEAREST CENTROID NEIGHBOR CLASSIFIER

In this paper, a new variant of the kNCN classifier based on local mean centroid vector and generalized mean distance is introduced, termed generalized mean distance k nearest centroid neighbor classifier. Let $T = \{y_i | y_i \in R^p\}_{i=1}^N$ be a set of training samples with N training samples in a p -dimensional feature space. $C = \{c_i | c_i \in \{c_1, c_2, \dots, c_m\}\}_{i=1}^N$denotes the corresponding class label set of T, where y_i has the label c_i and m represents the number of classes. Subset of training set, T_{c_i} is $T_{c_i} = \{y_j^i | y_j^i \in R^p\}_{i=1}^{N_i}$ with the total number of training samples from class c_i is N_i. The class label of a test sample, x is determined by the proposed GMDkNCN as follows

(i) Obtain k nearest centroid neighbors of test sample, x from the set T_{c_i} and it is denoted as $y_j^{NCN} = \{y_{i,j}^{NCN} \in R^p\}_{i=1}^k$.

(ii) Calculate k local centroid mean vector, $U_i^{NCN} = \{u_{i,j}^{NCN} \in R^p\}_{i=1}^N$using k nearest centroid neighbors from class c_i as follows

$$U_{i,j}^{NCN} = 1/k \sum_{j=1}^{k} y_{i,j}^{NCN} \quad (7)$$

(iii) Compute the Euclidean distance between test sample, x and local centroid mean vector, $u_{i,k}^{NCN}$ from class c_i and the corresponding distances are denoted as $D_i^{NCN} = d(x, u_{i,1}^{NCN}), d(x, u_{i,2}^{NCN}), \dots, d(x, u_{i,kj}^{NCN})$.

iv) The generalized mean distance of the first r distances from the first r local centroid mean vectors to x for class c_i is obtained as

$$gmd\left(x, U_{i,r}^{NCN}\right) = (1/r \sum_{j=1}^{r}(d\left(\left(x, y_{i,j}^{NCN}\right)\right)^{p}))^{1/p} \quad (8)$$

where $i = 1,2,\ldots,m$. Let k generalized mean distances of class c_i is represented as in equation 9.

$$g^{i} = gmd\left(x, U_{i,1}^{NCN}\right), gmd\left(x, U_{i,2}^{NCN}\right), \ldots, gmd\left(x, U_{i,kj}^{NCN}\right) \quad (9)$$

v) Find a new nested generalized mean distance based on the k generalized mean distance for class c_i using equation 10 where $i = 1,2,\ldots,m$.

$$G\left(x, g^{i}\right) = (1/k \sum_{r=1}^{k}(gmd((x, U_{i,r}^{NCN}))^{p}))^{1/p} \quad (10)$$

vi) Finally, the test sample, x is assigned to the class which has minimum nested generalized mean distance.

$$c_x = \underset{c_i}{\operatorname{argmin}}\, G(x, g^{i}) \quad (11)$$

IV. EXPERIMENTS

This section presents the description of data sets involved in this work and the experimental results. Extensive experiments are conducted to evaluate and compare the proposed GMDkNCN classifier with two kNCN-based classifiers (kNCN (Sanchez *et al.*, 1997) and LMkNCN (Gou *et al.*, 2012)) and other three competitive kNN-based classifiers: kNN (Stevens *et al.*, 1967), GMDkNN (Gou *et al.*, 2019), MLMkHNN (Pan *et al.*, 2017). Eighteen real-world data sets from UCI machine learning data sets repository (Bache & Lichman, 2013) is used in the experiments and the classification performance is measured using accuracy rate.

A. Datasets

The selected real-world dataset used in the experiments are summarized in Table 1. The eighteen real-world datasets are downloaded from the UCI machine learning repository (Bache & Lichman, 2013). Table 1contains the information of the real-world data sets such as the data name, number of samples, attributes and classes. As shown in Table 1, there are ten two-class classification problems, while the others are the multi class classification problems. In these data sets, the sizes of the samples are between 19 and 990, and the size of dimensionality is between 3 and 90.

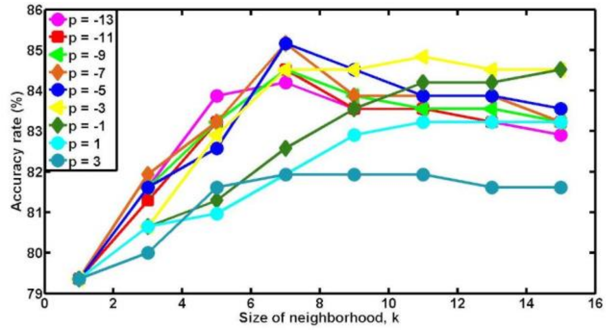
Table 1. The eighteen real world data sets used in the experiments

Name	Samples	Attributes	Classes
Vertebral	310	6	2
Balance	625	4	3
Breast Cancer	699	9	2
Bupa	345	6	2
Haberman	306	3	2
Hayes Roth	132	5	3
Hepatitis	155	19	2
Ionosphere	351	34	2
Iris	150	4	3
Knowledge	403	5	4
Libras	360	90	15
Mammographic	961	5	2
Seeds	210	7	3
Segmentation	19	7	2
Sonar	208	60	2
Vehicle	846	18	4
Vowel	990	10	11
Wpbc	198	32	2

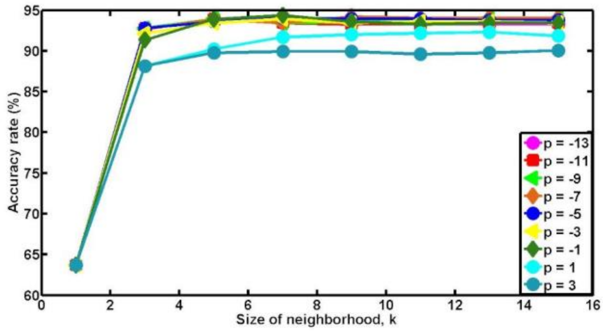
B. Experimental Results

The classification performance comparison is made between the proposed classifier, GMDkNCN with the competing classifiers, kNCN (Sanchez *et al.*, 1997), LMkNCN (Gou *et al.*, 2012), kNN (Stevens *et al.*, 1967), GMDkNN (Gou *et al.*, 2019) and MLMkHNN (Pan *et al.*, 2017) by conducting experimental comparisons on eighteen real-world data sets. Two experiments are conducted to evaluate the classification performance of the proposed classifier. In GMDkNCN, the value of parameter p is determined in the first experiment by varying the values from -13 to 3 in step 2. The choice of value p is recommended by Gou et al. (Gou *et al.*, 2019) and similar protocol is adopted to choose parameter p of the GMDkNN. The value of parameter p which resulting the best classification accuracy rate will be selected as the optimum value for the second experiment. Second experiment is conducted to evaluate the classification performance of the GMDkNCN classifier as a function of the neighborhood size, under different values of the neighborhood size, k with the range from 1 to 15 in step 2. All classifiers are evaluated based on 10-fold cross validation at which the experiments are done 10 times for each data set with 10 different training and test sets.

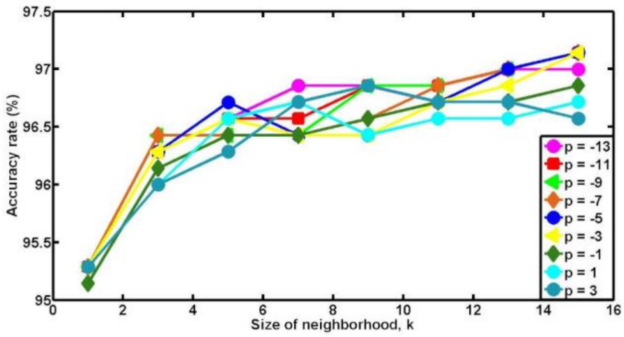
Figure 1 depicts the classification accuracy rates of the proposed GMDkNCN classifier through different values of p and k on each real-world data set. Based on Figure 1 it can be on observed that different values of p give significant impact to the classification accuracy. The value of p that resulting the best classification accuracy is considered as the optimum value to be used for the second experiment and it is stated in each data set as can be seen in Figure 1. Negative value of p produce better classification accuracy on 11 out 18 data sets used in this work as can be seen in Figure 1. These findings indicate that the classification accuracy of the proposed GMDkNCN classifier is directly influenced by different values of p of each data set. The improvement of classification accuracy is clearly seen when k becomes large on most of the data sets except for three data sets (Hayes Roth, Sonar and Vowel) that show decline trend. It shows that the proposed GMDkNCN is capable of selecting more accurate neighbors to obtain satisfactory result using negative values of p and has robust performance on different values of k.



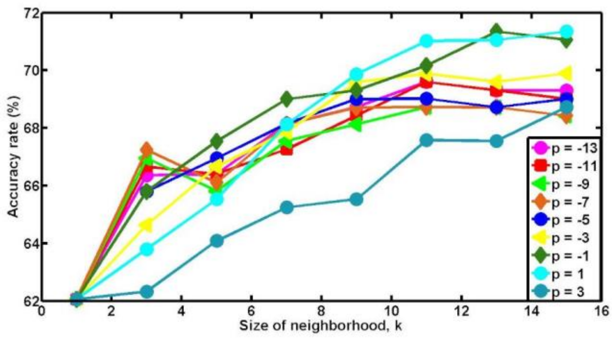
(a) Vertebral, (p = -7)



(b) Balance, (p = -1)

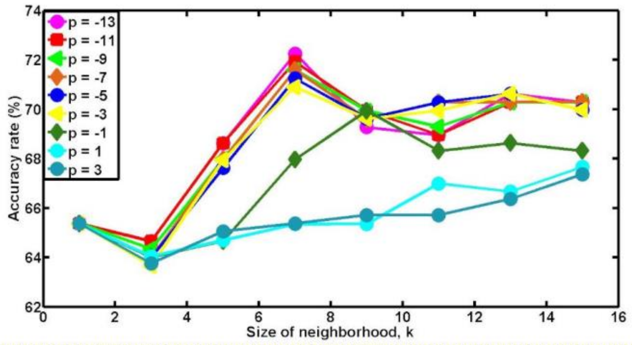


(c) Breast cancer, (p = -5)

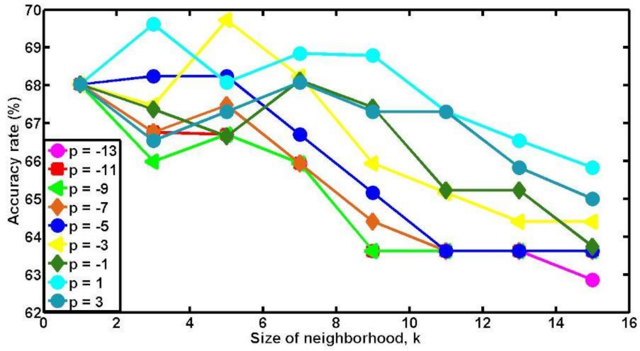


(d) Bupa, (p = 1)

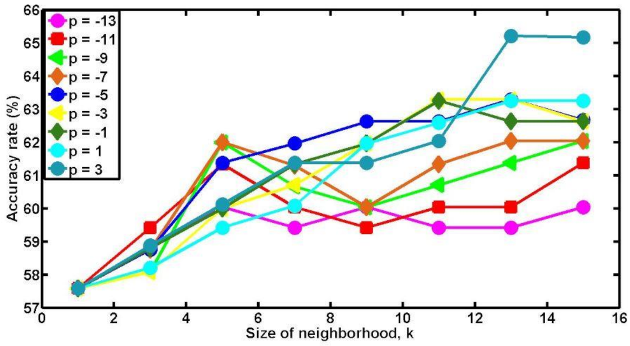
Figure 1. Classification accuracy of the GMDkNCN through different values of p and k on each real-world UCI data sets



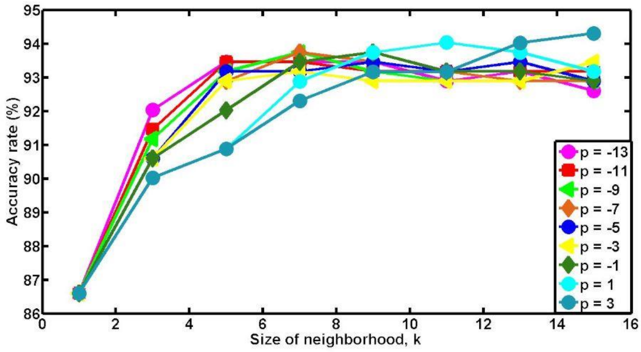
(e) Haberman, (p = -13)



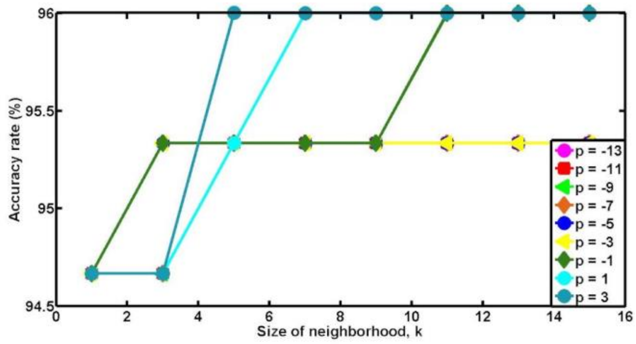
(f) Hayes Roth, (p = -3)



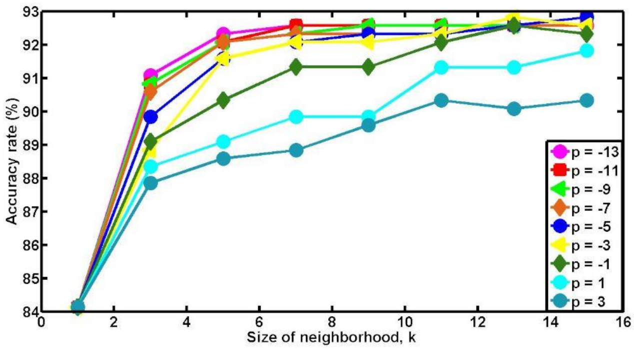
(g) Hepatitis, (p =3)



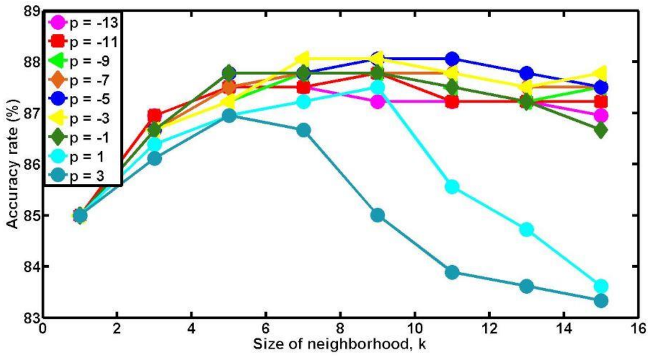
(h) Ionosphere, (p = 3)



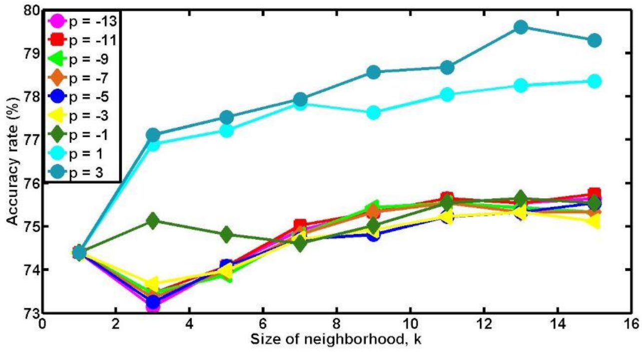
(i) Iris, (p = 3)



(j) Knowledge, (p = -11)

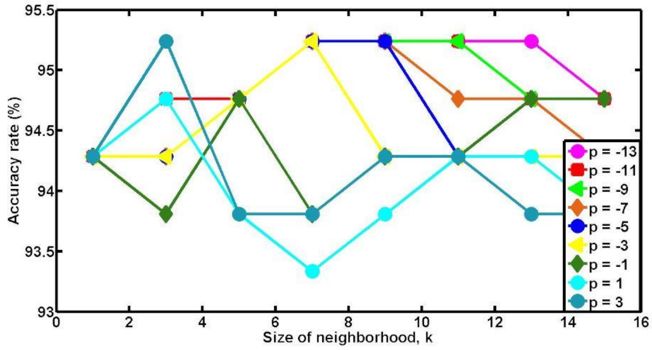


(k) Libras, (p = -3)

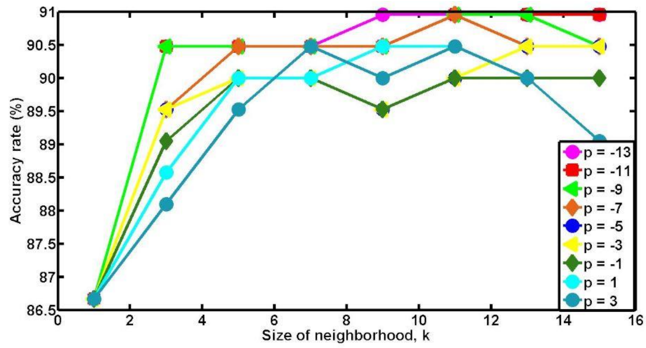


(l) Mammographic ,(p = 3)

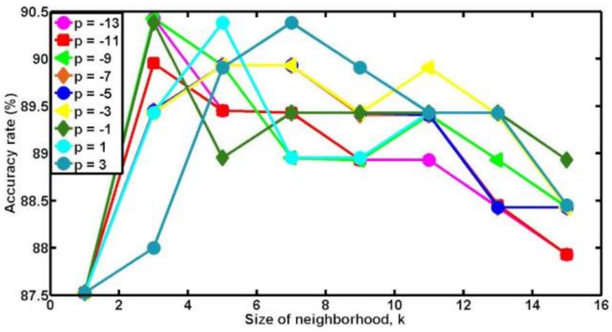
Figure 1. continued



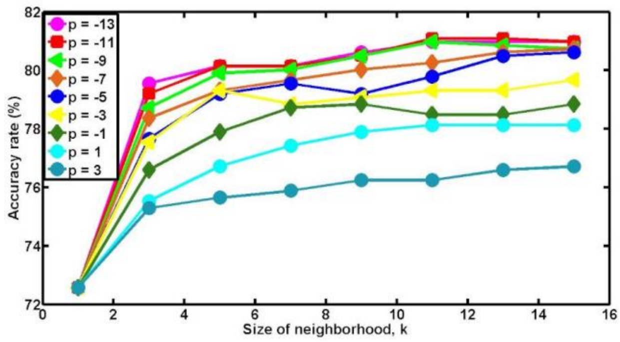
(m) Seeds, (p = 3)



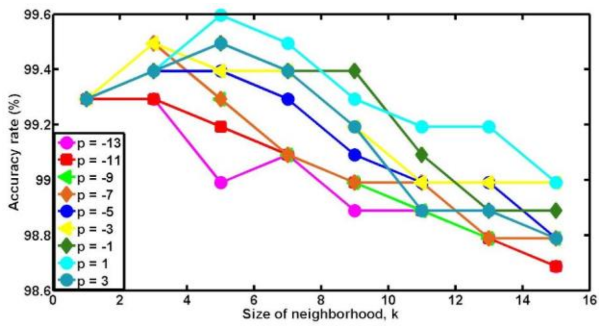
(n) Segmentation, (p = -13)



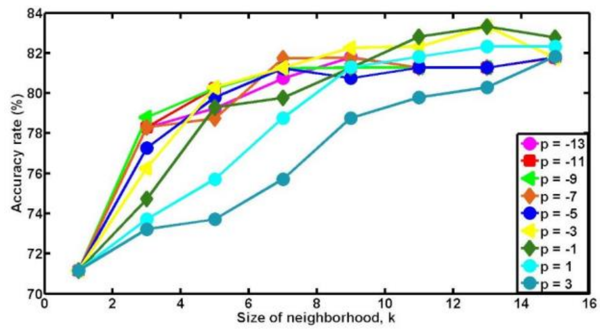
(o) Sonar, (p = -13)



(p) Vehicle, (p = -11)



(q) Vowel, (p = 1)



(r) Wpbc, (p = -3)

Figure 1. continued

The result in Table 2 is reported in terms of the average accuracy and the best value of the neighborhood size, k. Classification results of each dataset is determined by averaging 10 classification accuracy rates. The results show that that the proposed GMDkNCN classifier has the highest average of classification accuracy rates on real-world data sets in comparison to the other competing classifiers. It can be observed that the proposed GMDkNCN classifier achieves the highest rank on 8 out of 18 real-world data sets as highlighted in Table 2. These findings indicate that the strategy of integrating local centroid mean vector and generalized mean distance used in this work make it capable to capture more reliable local mean centroid vector and consequently elevate the classification performance. It is apparent from Table 2 that the classification performance of the GMDkNCN classifier is more superior to its closest competing technique, GMDkNN classifier. The possible explanation of this positive result could be due to the consideration of spatial distribution. In addition, the degree of positive influence of local mean centroid vector is further increased by the introduction of p value.

V. CONCLUSIONS

New classification technique termed generalized mean distance k nearest centroid neighbor (GMDkNCN) is introduced in this paper, which integrating the concept of local mean centroid vector and generalized mean distance.

Table 2. Classification comparison of the GMDkNCN classifier and five competing classifiers on the eighteen real-world data sets in terms of the average accuracy rate (%) with the values of k in the parentheses

Dataset	**kNN** (Stevens *et al.*, 1967)	**kNCN** (Sanchez *et al.*, 1997)	**LMkNCN** (Gou *et al.*, 2012)	**GMDkNN** (Gou *et al.*, 2019)	**MLMkHNN** (Pan *et al.*, 2017)	**GMDkNCN**
Vertebral	79.67(9)	83.54 (9)	83.87(11)	80.96(3)	81.62(15)	**85 16**(7)
Balance	89.42(15)	91.19(11)	91.34(15)	91.66(15)	92.94(15)	**94.38**(7)
Breast Cancer	96.99(5)	97.14(15)	96.85(7)	96.85(7)	**97.28**(15)	97.14(15)
Bupa	65.79(15)	69.93(13)	69.31(15)	66.69(13)	68.15(15)	**71.34**(15)
Haberman	**75.17**(15)	73.60(15)	67.98(15)	67.03(15)	68.35(15)	72.24(7)
Hayes Roth	68.02(1)	68.02(1)	69.56(7)	**72.80**(7)	68.18(5)	69.72(5)
Hepatitis	66.41(11)	**67.04**(11)	63.29(15)	64.58(15)	63.83(15)	65.20(13)
Ionosphere	86.60(1)	**94.59**(7)	94.02(13)	89.74(11)	91.16(15)	94.30(15)
Iris	96.00(5)	**96.66**(13)	96.00(7)	**96.66**(7)	**96.66**(5)	96.00(5)
Knowledge	85.84(3)	88.85(3)	91.32(15)	91.57(15)	92.31(15)	**92.82**(15)
Libras	85.00(1)	85.00(1)	87.50(5)	**88.88**(13)	88.33(5)	88.05(7)
Mammographic	**80.33**(7)	79.28(15)	78.87(13)	78.77(15)	75.44(9)	79.60(13)
Seeds	94.28(1)	94.28(1)	94.28(1)	**95.71**(5)	**95.71**(3)	95.23(3)
Segmentation	87.14(3)	88.09(9)	90.47(9)	89.52(11)	89.52(7)	**90.95**(9)
Sonar	87.52(1)	87.52(1)	**90.85**(3)	89.90(7)	89.90(7)	90.42(3)
Vehicle	72.57(1)	77.78(13)	78.24(13)	77.78(15)	78.01(13)	**81.08**(11)
Vowel	99.29(1)	99.29(1)	99.59(5)	99.49(3)	99.39(3)	**99.59**(5)
Wpbc	80.28(11)	81.26(11)	82.31(13)	80.28(15)	79.28(9)	**83.31**(13)
Average accuracy rate	83.13	84.62	84.76	84.38	84.23	**85.92**

The proposed strategy is focused to capture more reliable local centroid vector based on centroid distance. In addition, the proposed GMDkNCN introduces adjustable variable p to enhance the positive influence of the nearest centroid neighbors. In this way, different weights contribution can be properly assigned to local mean centroid vector by dynamically adjusting the values of p. The experimental results on the eighteen real-world data sets demonstrate the improvement on the classification performance of the GMDkNCN classifier and outperform five benchmark classifiers (kNN (Stevens *et al.*, 1967), kNCN (Sanchez *et al.*, 1997), LMkNCN (Gou *et al.*, 2012), GMDkNN (Gou *et al.*, 2019) and MLMkHNN (Pan *et al.*, 2017)).

VI. ACKNOWLEDGEMENT

This work is supported by Post Graduate Conference Fund Universiti Sains Malaysia 2019 and USM Bridging Fund 304/PELECT/6316321.

VII. REFERENCES

Bache, K & Lichman, M 2013, UCI machine learning repository. viewed 1 January 2017, <https://archive.ics.uci.edu/ml/index.php>.

Chaudhuri, BB 1996, '*A new definition of neighborhood of a point in multi-dimensional space*,' Pattern Recognition Letter, vol. 17, pp. 11-17.

Derrac, J, Chiclana, F, Garca, S & Herrera, F 2016, '*Evolutionary fuzzy k-nearest neighbors algorithm using interval-valued fuzzy sets'*, Information Sciences, vol. 329, pp. 144-63.

Fukunaga K 1990, '*Introduction to statistical pattern recognition'*, Academic Press, New York, USA.

Gou, J 2012, '*Weighted k -nearest centroid neighbor classification the knn classifier'*, Journal Computational and Information Systems, vol. 2, pp. 851-860.

Gou, J, Du, L, Zhang, Y & Xiong, T 2012, ' *A new distance-weighted k-nearest neighbor classifier'*, Journal of Information & Computational Science, vol. 9, pp. 1429-1436.

Gou, J, Ma, H, Ou, W, Zeng, S, Rao, Y & Yang, H 2019, '*A generalized mean distance-based k-nearest neighbor classifier'*, Expert Systems with Applications, vol. 115, pp. 356-372.

Gou, J, Yi, Z, Du, L & Xiong, T 2012, '*A local mean-based k-nearest centroid neighbor classifier*', The Computer Journal, vol. 55, pp. 1058-1057

Gou, J, Zhan, Y, Rao, Y, Shen, X, Wang, X & He, W 2014, '*Improved pseudo nearest neighbor classification*', Knowledge-Based Systems, vol. 70, pp. 361-375.

Indyk, P & Motwani, R 1998, 'Approximate nearest neighbors: towards removing the curse of dimensionality', in *Proceedings of the 13th annual ACM Symposium on Theory of computing*, pp 604-613.

Jaafar, H, Mukahar, N & D. Athiar D 2016, '*A Methodology of Nearest Neighbor: Design and Comparison of Biometric Image Database'*, in *Proceedings of the IEEE Student Conference on Research and Development (SCOReD)*, Kuala Lumpur, Malaysia.

Keller, JM, Gray, MR & Givens, JA 1985, '*A fuzzy k-nearest neighbor algorithm'*, IEEE transactions on systems, man, and cybernetics, vol. 4, pp. 580-585.

Ma, H, Wang, X & Gou, J 2016, '*Pseudo nearest centroid neighbor classification'*, in *Proceedings of the Frontier Computing*, pp. 103-15.

Mitani, Y & Hamamoto, Y 2006, '*A local mean-based nonparametric classifier'*, Pattern Recognition Letters, vol. 27, pp. 1151-1159.

Pan, Z, Wang, Y & Ku, W 2017, '*A new k-harmonic nearest neighbor classifier based on the multi-local means'*, Expert Systems with Applications, vol. 67, pp. 115-125.

Sanchez, JS, Pla, F & Ferri FJ 1998, '*Improving the k-NCN classification rule through heuristic modifications*', Pattern Recognition Letter, vol. 19, pp. 1165-1170.

Sanchez, JS, Pla, F & Ferri, FJ 1997, '*On the use of neighbourhood-based non-parametric classifiers*', Pattern Recognition Letter, vol. 18 pp. 1179-1186.

Stevens, KN, Cover, TM & Hart, PE 1967, 'Nearest neighbor', *IEEE Transactions of Information Theory*, vol. 13, no. 1, pp. 21-27.

Zeng, Y, Yang, Y & Zhao, L 2009, '*Pseudo nearest neighbor rule for pattern classification*', Expert Systems with Applications, vol. 36, pp. 3587-3595.