



Journal of Applied Sciences

ISSN 1812-5654

science
alert

ANSI*net*
an open access publisher
<http://ansinet.com>

How the Parameters of K-nearest Neighbor Algorithm Impact on the Best Classification Accuracy: In Case of Parkinson Dataset

¹Chih-Min Ma, ²Wei-Shui Yang and ¹Bor-Wen Cheng

¹Department of Industrial Engineering and Management, National Yunlin University of Science and Technology, Taiwan, Republic of China

²Department of Business Administration, Nan Kai University of Technology, Taiwan, Republic of China

Abstract: Parkinson's disease is a chronic and progressive neurological disease, which affects nervous system in a division of the brain that controls muscle movements. This disease is complicated and can be difficult to diagnose accurately in their early stages that prompts the researchers to try various classification methods to separate the healthy from the PD subjects. The K-Nearest Neighbor (KNN) classifier is one of the most heavily usage and benchmark in classification. In this study about KNN approach, there are two specific issues to be explored. The first one is to determine and obtain the optimal value of k ; another issue is to identify the effects of distance metric and normalization in KNN classifier with Parkinson dataset. This study utilizes a series of data split which the percentage of training data gradually increase from 5% to Leave One Out cross-validation. As can be seen from the results, the classification accuracy in the classification of raw PD dataset, either Euclidean distance or Manhattan distance is used to calculate the accuracy that are always lower than normalized dataset. The best classification accuracy corresponds to the k neighboring points is also different under various ratios of training and testing data, its means that the best answer k is not arbitrarily chosen, it should be obtained by calculating carefully. While the percentage of the training set accounts for 95%, using normalization by min-max with Euclidean distance has achieved promising performance, $96.73 \pm 5.97\%$ (when $k = 1$). As to the accuracy of the LOOCV, 96.41 ($k = 1$), is very close to the optimum value and there is no variation. Although, the results of LOOCV would not the best, it has been more popular in the case of stringent forecast.

Key words: KNN algorithm, classification accuracy, Parkinson's disease

INTRODUCTION

Parkinson's Disease (PD) is an important nervous system disorder, which affects nervous system in a division of the brain that controls muscle movements. Symptoms of Parkinson's disease, slowness and poverty of movement, resting tremor, rigidity, bradykinesia, etc., are commonly diagnosed in patients (Singh *et al.*, 2007), which generally affects people over the age of 60 and it is more common in men than in women. PD is an irreversible brain lesion. If this kind of disease can be diagnosed and treated earlier, drug treatments would mitigate or relieve the effects of the symptoms.

Up to now, the causes of Parkinson disease are not so clear. There is a lack of clear laboratory tests that can help to confirm the diagnosis sporadic PD. Particularly in early stages, the disease can be difficult to diagnose accurately. Scientists and researchers are doing a lot of researches to investigate the causes of this disease. They are trying to study many possible causes, including aging

and environmental poisons. One of the researchers, was performed and developed a method to identify different voice patterns through the use of voice disorders to diagnose Parkinson's disease (Little *et al.*, 2007). Related studies indicate that sound distortion occurs on 90% of patients with Parkinson's disease (Ho *et al.*, 1998). Subsequently, Little *et al.* (2009) used a different approach, Support Vector Machine (SVM), got overall classification accuracy reaches 91.4%. The researcher concludes that combined specific algorithms with harmonic noise ratio would be the best way to classify the healthy from PD subjects (Little *et al.*, 2009).

Later on some researchers used the dataset from Little *et al.* (2007). The work done by Resul (2010) emphasize on comparing the accuracy of different types of classification methods for effective evaluate and diagnosis of Parkinson's. Various classifiers, Regression, DMNeural, Neural Network and Decision Tree, have been used to identify the Parkinson's disease. As a result, the neural network obtains the optimal classification accuracy

of 92.9%. On the basis of accuracy, KNN classifier shows the best to distinguish between Parkinson's disease and those who do not have it. The K-Nearest Neighbor (KNN) classifier is one of the most heavily usage and benchmark in classification. Even comparing the most advanced machine learning approaches, this simple and easy-to-use algorithm still can produce superior results. However, there are several critical issues affecting the performance of KNN, mainly including the choice of k, the new class label of the data prediction, the selection the distance metric and its data pre-processing method.

Many scholars in the past mostly used arbitrarily chosen k values to be calculated and tried, such as pre-selected k = 3 (Yang *et al.*, 2012), k = 1, 2, ..., 10 (Sierra *et al.*, 2011), k = 1, 2, ..., 50 (Wang *et al.*, 2007) or k = square (number of samples) (Mitra *et al.*, 2002; Bhattacharya *et al.*, 2012). But there isn't clear enough to provide the necessary guideline for determining the k-value. Generally, Larger values of k presents less sensitive to noise and makes smoother between the boundaries of classes. As a result, choosing the same (optimal) k becomes almost impossible for different applications. Historically, the optimal k for most of the datasets is always located between 3 and 10. However, this is not absolutely certain criteria. Determining the value of k is actually supposed to be discussed further.

In addition, using different distance metrics to calculate the accuracy of classification may produce different results. Many scholars have been promoted to develop a variety of distance metrics for enhancing the accuracy of KNN algorithm (Wang *et al.*, 2007; Bhattacharya *et al.*, 2012; Yang *et al.*, 2012). Furthermore, database preprocessing and model's validation methods also have an impact on the accuracy of classification. In statistical external validation, the whole dataset splits into training and testing sets randomly, using this skill to check the quality of the classifier. Here, how to determine the split ratio of the dataset and compare the possible similarities and differences in each execution, this will be another issue to be explored. Throughout the above description of k neighboring points, distance function and data pattern, PD has been classified using KNN classifier, the parameters impact on the best classification accuracy has been discussed either. The aim of this article is twofold.

The first one is to determine the value of k, that is, what is the optimal value for the neighboring constant 'k', and how should it be obtained? Exhaustive search is used to calculate the accuracy under different k values.

The second one is to cross calculate the classification accuracy of each distance metric by transforming the dataset into different types of

normalization. This will help to identify the effects of distance metric and normalization in KNN classifier on the classification of PD dataset.

MATERIALS AND METHODS

The Parkinson dataset used in this study, obtained from UCI machine learning repository. The dataset is composed of 22 attributes sustained biomedical voice measurements including 32 subjects from both sexes, of which 23 were diagnosed with PD. There are 195 instances in the data set, including 48 normal and 147 PD patients. The main purpose is to distinguish between healthy people with PD from the data, according to "status" variable which is assign to 0 for healthy people and 1 for PD. The attributes of dataset are given in Table 1 (Little *et al.*, 2007).

K-Nearest neighbor classifier: KNN is a supervised learning algorithm and perceived as a simple but powerful classification, even for complex applications, capable of yielding high-performance results (Dzuida, 2010). In the beginning of 1970's, KNN has been already used in statistical estimation and pattern recognition as a non-parametric technique. The k samples which are closest to the query sample are found based on a similarity measure among the training data. The class label of the query sample is determined by using majority voting (Shakhnarovich *et al.*, 2005). In the study, distance measurements like Euclidean and Manhattan are used to calculate the distances of the samples. The various

Table 1: Describing the attribute information presented by some kind of vocal signals with Parkinson's disease

Attribute name	Description
MDVP: Fo (Hz)	Average vocal fundamental frequency
MDVP: Fhi (Hz)	Maximum vocal fundamental frequency
MDVP: Flo (Hz)	Minimum vocal fundamental frequency
MDVP: Jitter (%)	Measures of variation in fundamental frequency
MDVP: Jitter (Abs)	
MDVP: RAP	
MDVP: PPQ	
Jitter: DDP	
MDVP: Shimmer	Measures of variation in amplitude
MDVP: Shimmer (dB)	
Shimmer: APQ3	
Shimmer: APQ5	
MDVP: APQ	
Shimmer: DDA	
NHR	Two measures of ratio of noise to tonal components in the voice
HNR	
Status	Health status of the subject (1)-Parkinson's disease, (0)-Healthy
RPDE D2	Two nonlinear dynamical complexity measures
DFA	Signal fractal scaling exponent
Spread 1	Three non-linear measure of fundamental frequency variation
Spread 2	
PPE	

k-values in KNN classifier are used and compared with each other, which effects have been investigated and pointed out the best one.

Euclidean distance: The euclidean distance between any two instances is the length of the line segment connecting them. In this study, the dataset is composed of 22 attributes is represented in 22-dimensional space. If $x = (x_1, x_2, \dots, x_{22})$ and $y = (y_1, y_2, \dots, y_{22})$ are two points in Euclidean 22-space, then the distance from x to y is given by:

$$d(x,y) = \sqrt{\sum_{i=1}^{22} (x_i - y_i)^2}$$

Manhattan distance: The manhattan distance is the distance that would be traveled to get from one data point to the other, measured along axes at right angles. The name implies the Manhattan street in grid layout which causes the shortest path between two points. The equation for this distance between a point $x = (x_1, x_2, \dots, x_{22})$ and a point $y = (y_1, y_2, \dots, y_{22})$ is given by:

$$d(x,y) = \sum_{i=1}^{22} |x_i - y_i|$$

Normalization: Distance measure calculated directly from the training set has one major shortcoming, that is, the variable has a different measurement scales or a mixture of numerical and categorical variables. For example, in this study, attribute “MDVP: Fo (Hz)” is based on Hertz, and MDVP: Jitter (%) is based on percent then Hertz will have a much higher influence on the distance calculated. One of the solutions is to standardize the training set, such as normalization. It is a data preprocessing tool used in data mining system. By normalizing the attributes of a dataset, their values are scaled to a small-specified range, such as 0.0-1.0. Normalization is particularly useful for classification algorithms. There are many methods for data normalization. In this study, KNN classifier predicts the results of classification adopting raw data and data after four normalization approaches, includes min-max normalization, z-score normalization, max normalization and normalization by decimal scaling. Generally, the data subjected to preprocessing in the classification process will improve the performance of classification (Kumar and Zhang, 2007; Polat *et al.*, 2008). Twenty two continuous-valued attributes in Parkinson's disease dataset are normalized by following methods. Details about these four normalization methods are described as below:

- **Min-max normalization:** Linear transformation of the raw data was carried out. Suppose that v_{min} and v_{max} are the minimum and maximum of attribute variable. Every value v from the raw dataset will be mapped into value v' using the following equation:

$$v' = \frac{v_{max} - v_{min}}{v - v_{min}}$$

- **Z-score normalization:** The values of attribute are normalized using the mean and standard. A new value v' is obtained using the following expression:

$$v' = \frac{v - v_{ave}}{v_{std}}$$

where, v_{ave} and v_{std} are the mean and standard deviation of attribute X, respectively

- **Max normalization:** The values of attribute are normalized using every value v from the original interval divided by the maximum of attribute variable v_{max} . New value v' is obtained using the following expression:

$$v' = \frac{v}{v_{max}}$$

where v_{max} is the maximum of attribute variable

- **Normalization by decimal scaling:** Normalizes by moving the decimal point of values of attribute. The number of decimal points depends on the maximum absolute value. A new value v' corresponding to v is obtained using the following expression:

$$v' = \frac{v}{10^i}$$

where i is the smallest integer such that $\max(|v'|) < 1$.

- **Cross-validation:** Choosing the optimal value for k-neighborhood is best done by first inspecting the data. In general, a large k-neighborhood value is more precise as it reduces the overall noise but there is no guarantee. Cross-validation is another way to determine a good k-neighborhood value by using an independent data set to validate the value of k-neighborhood.

In K-fold cross-validation, the original sample is randomly divided into K sub-samples of approximately equal size. One of the K sub-samples is keep as the testing data, and the rest of the K-1 sub-samples are used as training data. The process is repeated K times, with each of the K sub-samples used for validation exactly

once. The average of K results from the folds is yielded to a single estimation.

In this study, started with 3-fold cross-validation, 67% training and 33% testing data split, has been used and randomly chosen from dataset. The best accuracy of classification is yielded with such a matrix out of a given set of data and a given distance function. This phase mainly discusses the value of k and these different data patterns, what the impacts will be on the best classification accuracy. Then, this study will use a series of data split which the percentage of training data gradually increase from 5% to LOOCV (Leave one out cross-validation). The data set is tested further for finding the causes of the true optimum value of k. That is, this study will try to find the best value of k under a variety of split ratio of training dataset and with the different distance function.

RESULTS AND DISCUSSION

Range of k: This study was conducted by the value of $k = 1$, progressively increasing k neighboring points from 1-133 (67% training dataset) to test the range of k. Euclidean distance is used here to examine the accuracy by different raw dataset and normalized datasets. The results indicate that if k is even number, the accuracy is less than the condition of odd, $k+1$ and $k-1$. In addition, as k increases to 65 or more, the accuracy of all data patterns is almost close to 75.38%. This means that when the k neighboring points is large enough, the accuracy will naturally tends to the expected value of the sample, in this data set including 48 normal and 147 PD patients.

Excluding all even numbers of k, this study re-defines the range of k value, which is odd and less than 65. In training and testing of KNN classifier, still 67-33% train-test data split has been randomly chosen from the dataset. The effects of k-value in KNN classifier on the classification accuracy including Euclidean and Manhattan distance have been investigated. Both of the obtained classification results are plotted in Fig. 1 and 2. As can be seen from these figures, the classification accuracy in the classification of raw PD dataset, either Euclidean distance or Manhattan distance is used to calculate the accuracy that are almost the lowest. On the other hand, it is noteworthy that the accuracy from normalized datasets is relatively better, especially normalized by the min-max normalization or decimal scaling has a higher accuracy and make no difference around $k = 1$ among the two figures.

Best accuracy of classification: The overall data set splits into training and testing sets as following. The percentage

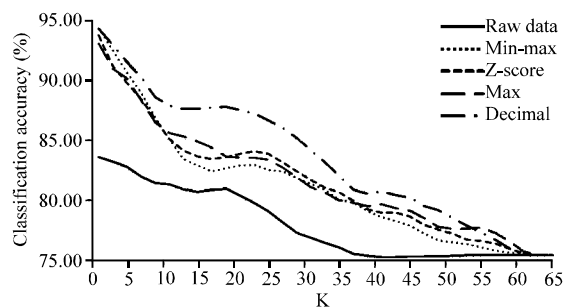


Fig. 1: KNN accuracy for different k-value based on euclidean distance, training data (67%)/testing data (33%), k is odd and less than 65

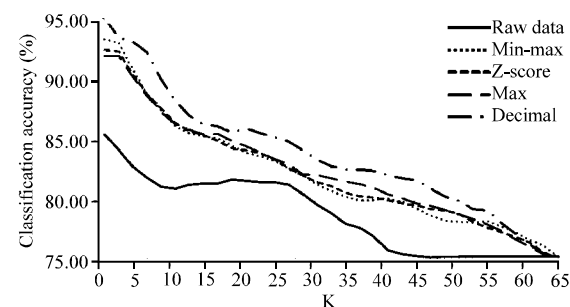


Fig. 2: KNN accuracy for different k-value based on Manhattan distance, training data (67%)/testing data (33%), k is odd and less than 65

of the training data set at least 5%, gradually up to LOOCV, is randomly selected. Using the chosen raw and normalized datasets estimate the classification accuracy of the KNN classifier. The above steps performed repeatedly 100 times under each percentage shown in Table 2. The results from the calculations then can be averaged to yield a single estimation. As can be seen from the table, found that the less the percentage of training set, the lower the accuracy, if there lower enough, the accuracy will tends to the expected value of the sample also. As the percentage is greater than 75%, the accuracy of normalized datasets reaches 95% or more. The results imply that the percentage of training sets shall split the larger the better, the classification accuracy will be relatively close to the optimum value, accordingly, and then it can obtain an acceptable optimum value, not exactly one. This is why in most studies; researchers always use a 10-folds cross-validation, 90% training and 10% testing data split, to find their answers. In fact, the results from 10-folds cross-validation may not hit the target, but close.

In addition, each best k value is acquired from different training and testing data split respectively, only a small portion of k is likely to belong to 3, 5 or 7, the others are almost equal to 1. This represents under a

Table 2: Effects of k-value in KNN classifier on the classification accuracy used different percentage of the training data set on raw and normalized datasets, (based on euclidean distance and manhattan distance)

Training (%)	Metric	Raw data	Min-Max	Z-score	Max	Decimal
5.0	Euclidean	75.43±0.27 (7)	77.78±6.69 (1)	76.96±5.52 (1)	77.92±5.58 (1)	79.84±6.66 (1)*
	Manhattan	75.62±3.67 (5)	78.61±5.40 (1)	78.68±5.52 (1)	77.30±6.78 (1)	78.75±6.63 (1)
10.0	Euclidean	76.06±4.44 (5)	80.91±4.83 (1)	81.71±4.03 (1)	81.30±4.72 (1)	83.14±4.43 (1)
	Manhattan	78.03±5.23 (3)	83.37±3.24 (1)	82.50±3.71 (1)	82.64±4.13 (1)	83.45±3.97 (1)*
15.0	Euclidean	78.26±3.91 (3)	84.34±3.56 (1)	84.08±3.88 (1)	84.58±3.04 (1)	86.23±3.45 (1)
	Manhattan	79.26±4.27 (5)	85.15±3.35 (1)	84.64±3.97 (1)	85.19±3.77 (1)	86.83±3.35 (1)*
20.0	Euclidean	79.38±3.30 (1)	87.12±3.13 (1)	86.64±3.10 (1)	85.90±3.44 (1)	87.28±2.98 (1)
	Manhattan	80.85±3.39 (1)	87.17±3.41 (1)	87.21±2.86 (1)	86.53±3.17 (1)	87.94±2.90 (1)*
25.0	Euclidean	80.43±3.58 (1)	88.30±2.92 (1)	88.17±2.83 (1)	87.71±3.08 (1)	88.44±3.14 (1)
	Manhattan	81.73±2.88 (1)	88.86±3.04 (1)	88.43±2.93 (1)	87.82±2.89 (1)	90.23±3.15 (1)*
30.0	Euclidean	80.23±3.79 (1)	89.57±2.63 (1)	88.63±2.83 (1)	88.26±3.17 (1)	89.59±2.55 (1)
	Manhattan	82.57±2.91 (1)	89.80±2.88 (1)	89.40±2.96 (1)	89.29±2.61 (1)	91.13±2.93 (1)*
35.0	Euclidean	81.84±3.47 (1)	90.87±2.52 (1)	89.89±2.84 (1)	89.25±2.78 (1)	91.44±2.95 (1)
	Manhattan	83.21±3.17 (1)	91.01±2.50 (1)	90.59±2.78 (1)	89.83±2.71 (1)	92.34±2.68 (1)*
40.0	Euclidean	82.25±3.38 (1)	91.81±2.75 (1)	90.53±3.06 (1)	90.35±2.82 (1)	91.72±2.52 (1)
	Manhattan	83.60±3.02 (1)	91.64±2.81 (1)	90.97±2.82 (1)	90.34±2.96 (1)	92.96±2.56 (1)*
45.0	Euclidean	83.10±2.98 (1)	92.14±2.80 (1)	90.64±3.07 (1)	90.69±2.72 (1)	92.78±2.23 (1)
	Manhattan	84.35±3.12 (1)	91.84±2.73 (1)	91.49±2.97 (1)	90.62±2.79 (1)	93.41±2.38 (1)*
50.0	Euclidean	82.93±3.07 (1)	92.89±2.67 (1)	91.35±2.74 (1)	91.65±2.55 (1)	93.18±2.60 (1)
	Manhattan	84.74±3.36 (1)	92.85±2.57 (1)	91.84±2.74 (1)	90.95±2.43 (1)	94.08±2.39 (1)*
55.0	Euclidean	83.07±3.50 (1)	93.08±2.52 (1)	92.21±3.01 (1)	92.30±2.56 (1)	93.61±2.44 (1)
	Manhattan	85.15±3.45 (1)	92.52±2.67 (1)	91.67±2.80 (1)	91.54±2.24 (1)	94.09±2.28 (1)*
60.0	Euclidean	83.63±3.24 (3)	93.74±2.42 (1)	92.35±3.08 (1)	93.00±2.79 (1)	93.55±2.46 (1)
	Manhattan	84.49±3.16 (1)	93.05±3.03 (1)	92.50±2.83 (1)	91.56±2.69 (1)	94.65±2.19 (1)*
65.0	Euclidean	83.63±3.72 (3)	94.34±2.76 (1)	93.24±3.05 (1)	93.18±2.52 (1)	93.96±2.40 (1)
	Manhattan	85.25±3.34 (1)	93.79±2.96 (3)	92.63±2.89 (1)	92.04±2.96 (1)	94.90±2.30 (1)*
70.0	Euclidean	83.90±4.59 (1)	94.33±2.82 (1)	93.79±3.14 (1)	92.72±3.11 (1)	94.33±2.79 (1)
	Manhattan	85.36±3.94 (1)	93.62±3.07 (1)	93.00±2.77 (1)	92.22±2.58 (1)	94.74±2.56 (1)*
75.0	Euclidean	84.21±4.29 (3)	95.00±3.26 (1)	93.00±3.79 (1)	93.00±3.55 (1)	94.54±3.13 (1)
	Manhattan	85.88±4.24 (1)	94.75±3.13 (3)	93.71±3.18 (1)	92.02±3.26 (1)	95.10±2.78 (1)*
80.0	Euclidean	85.38±4.23 (5)	95.36±2.98 (1)	94.31±4.04 (1)	93.51±3.63 (1)	94.36±3.80 (1)
	Manhattan	85.21±4.59 (1)	94.13±3.70 (3)	93.05±3.88 (3)	92.72±3.69 (1)	95.87±3.05 (1)*
85.0	Euclidean	85.69±5.29 (5)	95.69±3.48 (1)*	94.17±4.72 (1)	93.59±3.89 (1)	95.38±3.90 (1)
	Manhattan	85.28±5.32 (7)	94.28±4.47 (3)	93.62±3.90 (1)	92.24±4.37 (3)	95.79±3.78 (1)
90.0	Euclidean	84.84±6.95 (5)	96.68±4.14 (1)*	94.68±4.06 (1)	94.58±5.04 (1)	95.95±4.09 (1)
	Manhattan	85.89±7.66 (3)	94.74±5.45 (3)	94.21±4.76 (3)	91.53±5.98 (3)	96.47±3.82 (1)
95.0	Euclidean	85.84±9.53 (5)	96.73±5.97 (1)*	95.75±6.82 (1)	94.55±6.95 (1)	96.54±6.81 (1)
	Manhattan	85.44±11.16 (1)	94.99±6.75 (1)	93.79±7.75 (1)	93.14±7.98 (3)	96.30±5.71 (1)
LOOCV	Euclidean	86.15 (5)	96.41 (1)*	95.38 (1)	94.36 (1)	95.9 (1)
	Manhattan	85.64 (3)	94.87 (5)	93.33 (1)	92.82 (1)	95.9 (1)

(k)*: Optimal accuracy in k value

different ratio of training and testing data, its best classification accuracy corresponds to the k neighboring points is different. However, there are many scholars mostly used arbitrarily chosen k values to be calculated, such as pre-selected k = 3 (Yang *et al.*, 2012), k = 1, 2, ..., 10 (Sierra *et al.*, 2011), k = 1, 2, ..., 50 (Wang *et al.*, 2007) or k = Square (number of samples) (Mitra *et al.*, 2002; Bhattacharya *et al.*, 2012). How should determine the value of k, there are no specific description. This case uses nearly global search methods to explore the range of k. It is feasible to calculate a smaller dataset. As the training set = 95.0%, using normalization by min-max normalization with Euclidean distance, have the highest average accuracy which is 96.73±5.97 (k = 1). As to the accuracy of the LOOCV, 96.41 (k = 1), is very close to the optimum value and there is no variation, the results would have been more popular in the case of stringent forecast.

CONCLUSION

Constantly improved algorithm techniques have been developed so far to secure the accuracy of classification. As mentioned above, there are many methods that are in use for recognizing PD today, whether in the field of medical research or public health. KNN is a simple but useful algorithm technology. It has the potential to become a good supportive for the experts to improve the accuracy and reliability of diagnosis, as well as making the diagnosis fewer possible errors and more time-efficient.

From the results of this study and by comparing other literature, such as, Ene (2008) using neural network-based approach has derived the highest classification accuracy 81.28%. Resul (2010) also uses four well-known classification methods, best classification accuracy achieved by the neural network classifier 92.9%.

As can be seen from this study, the classification accuracy in the classification of raw PD dataset, either Euclidean distance or Manhattan distance is used to calculate the accuracy that are almost the lowest and under a different ratio of training and testing data, its best classification accuracy corresponds to the k neighboring points is also different. That means that k can not be arbitrarily chosen, it should be calculated carefully. The training set = 95.0%, using normalization by min-max normalization with Euclidean distance, have the highest average accuracy 96.73 ± 5.97 (when $k = 1$). The results points out different distance functions with different normalized methods also have a decisive influence on the optimum value. As to the accuracy of the LOOCV, 96.41 ($k = 1$), is very close to the optimum value and there is no variation. Although, the results of LOOCV would not the best, it has been more popular in the case of stringent forecast.

REFERENCES

- Bhattacharya, G., K. Ghosh and A.S. Chowdhur, 2012. An affinity-based new local distance function and similarity measure for kNN algorithm. *Pattern Recognition Lett.*, 33: 356-363.
- Dzuida, D.M., 2010. *Data Mining for Genomics and Proteomics: Analysis of Gene and Protein Expression Data*. John Wiley and Sons, New York, ISBN: 9780470593400, Pages: 328.
- Ene, M., 2008. Neural network-based approach to discriminate healthy people from those with Parkinson's disease. *Ann. Univ. Craiova Math. Comp. Sci. Ser.*, 35: 112-116.
- Ho, A.K., R. Ianssek, C. Marigliani, J.L. Bradshaw and S. Gates, 1998. Speech impairment in a large sample of patients with Parkinson's disease. *Behav. Neurol.*, 11: 131-137.
- Kumar, A. and D. Zhang, 2007. Hand-geometry recognition using entropy-based discretization. *IEEE Trans. Inform. Forensics Security*, 2: 181-187.
- Little, M.A., P.E. McSharry, S.J. Roberts, D.A.E. Costello and I.M. Moroz, 2007. Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. *BioMed. Eng. Online*, Vol. 6. 10.1186/1475-925X-6-23
- Little, M.A., P.E. McSharry, E.J. Hunter, J. Spielman and L.O. Ramig, 2009. Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. *IEEE Trans. Biomedical Eng.*, 56: 1015-1022.
- Mitra, P., C.A. Murthy and S.K. Pal, 2002. Unsupervised feature selection using feature similarity. *IEEE Trans. Pattern. Anal. Mach. Intell.*, 24: 301-312.
- Polat, K., S. Kara, A. Guven and S. Gunes, 2008. Utilization of discretization method on the diagnosis of optic nerve disease. *Comput. Methods Prog. Biomedicine*, 91: 255-264.
- Resul, D., 2010. A comparison of multiple classification methods for diagnosis of Parkinson disease. *Expert Syst. Appl.*, 37: 1568-1572.
- Shakhnarovich, G., T. Darrell and P. Indyk, 2005. *Nearest-Neighbor Methods in Learning and Vision, Theory and Practice*. MIT Press, USA., ISBN: 0-262-19547-X.
- Sierra, B., E. Lazkano, I. Irigoien, E. Jauregi and I. Mendialdua, 2011. K nearest neighbor equality: Giving equal chance to all existing classes. *Inform. Sci.*, 181: 5158-5168.
- Singh, N., V. Pillay and Y.E. Choonara, 2007. Advances in the treatment of Parkinson's disease. *Prog. Neurobiol.*, 81: 29-44.
- Wang, J., P. Neskovic and L.N. Cooper, 2007. Improving nearest neighbor rule with a simple adaptive distance measure. *Pattern Recognition Lett.*, 28: 207-213.
- Yang, W., K. Wang and W. Zuo, 2012. Fast neighborhood component analysis. *Neurocomputing*, 83: 31-37.