ANO Detection with K-Nearest Neighbor Using Minkowski Distance

Hui-Lee Ooi, Siew-Cheok Ng, and Einly Lim
University of Malaya/ Department of Biomedical Engineering, Kuala Lumpur, Malaysia
Email: huilee_ooi@siswa.um.edu.my, siewcng@um.edu.my, einly_lim@um.edu.my

***Abstract*—Full unloading of the left ventricle in patients implanted with ventricular assist devices over a long period of time has been reported to cause aortic valve fusion and thrombosis. Therefore, it is vital to detect the occurrence of aortic valve non-opening (ANO) in these patients. In the present study, measurements were obtained from four greyhounds under a wide range of operating conditions, which covered varying levels of blood volume, vascular resistances, contractilities and pumping speeds. K nearest neighbor classifier was implemented on two features, i.e. root mean square and standard deviation of pump speed amplitude. Classification performance using different Minkowski distance metrics (Manhattan distance, Euclidean distance, Minkowski distance with a power of 3 and Chebyshev distance) and k parameter (odd numbers ranging from 1 to 29) were evaluated. Results showed that the compared metrics achieved similar performance (accuracy of 93%) and concurred unanimously with regards to the optimal numbers of k parameters (k=9) for the data tested.**

***Index Terms*—aortic valve non-opening, ANO detection, k-nearest neighbor, Minkowski distance**

I. Introduction

Heart failure refers to a condition where the heart is unable to perform its function effectively to meet the physiological needs of the body. It is prevalent among men and women across all ethnic groups and is ranked as the top underlying cause of death worldwide [1]. Despite the disease being highly common and potentially deadly, there is currently no permanent cure. Fortunately, there are several options to treat this global disease with emerging advances of technology. Ventricular assist device (VAD) is one such option that provides mechanical circulatory function designed to aid the failing heart in pumping blood throughout the body. The device improves the quality of life and prolongs the longevity of heart failure patients by means of bridge to transplantation, bridge to recovery as well as destination therapy.

However, the biomechanics of aortic valve undergoes substantial alteration [2] in patients implanted with VADs and this spurred the rising needs to detect detrimental states caused by over-pumping. Aortic valve non opening (ANO) occurs when the aortic valves remain closed throughout the cardiac cycle with zero flow due to increasing pumping speed [3]. This may eventually lead to more serious complications such as aortic stenosis, aortic regurgitation [2] or even aortic fusion [4] if left unattended. Hence, it is imperative that optimal speed is applied to satisfy the varying needs of physiological changes in the assisted patients.

Despite being one of the simplest and oldest classifier [5], K nearest neighbor (KNN) offers an appealingly straightforward classification solution based on the class of their nearest neighbor, achieving competitive results when combined with prior information in supervised learning. It performs non-parametric discrimination [5] and thus do not impose any assumption on the distribution of the tested data, a useful trait that is highly desirable in practical applications [2], [6]-[10]. Without involving any preprocessing of the training data, KNN is also known as instance-based or lazy learning algorithm [11] as the classification process can only be carried out with the presence of test data.

One drawback of KNN is that it is often limited by constraints [12] of computational complexity and biased by the k parameter, which determine the region size of nearest neighbor. It was known that the computation complexity of a large number of data in KNN often leads to long testing time. Euclidean distance [13] and Mahalanobis [14] distance are two of the most popular distance metric applied in the implementation of KNN. In an attempt to allow more efficient execution, different types of normalized Minkowski distance including Euclidean distance for ANO detection are studied in this paper.

II. Methodology

A. Data Acquisition

The data were acquired from four different anesthetized greyhounds implanted with VentrAssist VAD and underwent several stages of pumping states with different levels of preloads and afterloads. The inflow cannula of the device was connected to the apex of the left ventricle whereas the outflow cannula was attached to the ascending aorta. Initially set at 4 kHz, the sampling rate for the signals was then down-sampled to 200Hz [15]. As illustrated in Fig. 1, the experiment started out with speed ramp under healthy condition.

Manuscript received September 18, 2013; revised November 20, 2013.

doi: 10.12720/ijsps.1.2.208-211

Different levels of cardiac contractibility, afterload and preload were later induced for investigation.

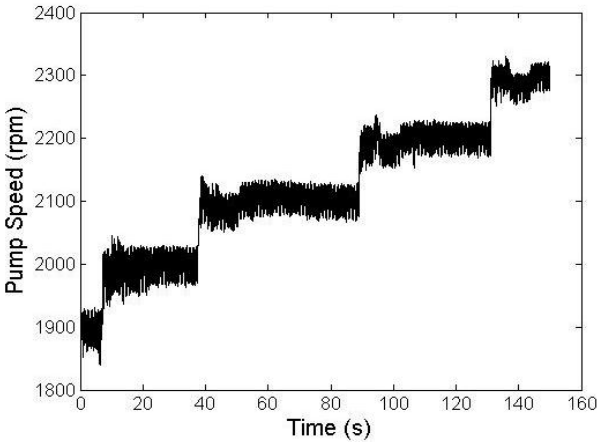


Figure 1. Speed ramp signal taken from greyhound.

B. Pump States Verification

Non-invasively obtained pump speed signal was used in the investigation of ANO occurrence. In order to independently identify the states contained in the signals, additional physiological signals such as left ventricular pressure (LVP) and aortic pressure (AoP) were used for verification, as shown in Table I.

TABLE I. STATE IDENTIFICATION CRITERIA

	ANO	Normal ejection
Maximum left ventricular pressure (LVP) [3]	Max LVP>AoP	Max LVP<AoP
Aortic flow (AoQ) [3]	~0	Positive
Presence of dicrotic notch [16]	No	Yes

C. Cycle Approximation

Fig. 2 showed how cycle was determined from the data. Noise was removed by having the pump speed signal going through a low pass filter with cutoff frequency of 10Hz. Features were extracted from each cardiac cycle to characterize the signal from its morphology. Combination of root mean square and standard deviation of signal amplitude was used as features in this study. By superimposing the filtered signal and averaged signal from the pump speed, estimation of cycle was indicated by the alternate intersection points.

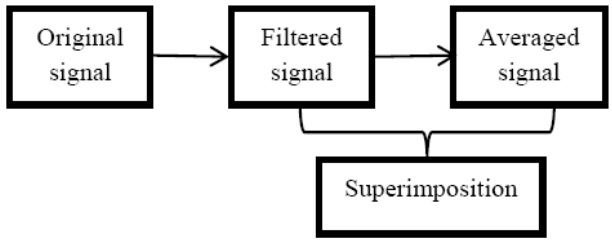


Figure 2. Flow of data cycle extraction.

D. Feature Normalization

To prepare the data for classification, a preprocessing process known as normalization was first executed to form appropriate generalization of data for higher accuracy and efficiency [17]. The distance function for completing the classification is closely related to the scale of the values of features presented. To avoid features with large range from outweighing features of relatively lower magnitude and thus causing unnecessary weightage bias [17], all the features were scaled to the same range. Two types of normalization methods were applied in this study, namely z-score normalization and min-max normalization.

The z-score normalization was implemented by the function:

$$X_{test\_norm} = \frac{X_{test} - mean(X_{train})}{std(X_{train})} \quad (1)$$

$$X_{train\_norm} = \frac{X_{train} - mean(X_{train})}{std(X_{train})} \quad (2)$$

The min-max normalization was implemented by the function:

$$X_{test\_norm} = \frac{X_{test} - min(X_{train})}{max(X_{train}) - min(X_{train})} \quad (3)$$

$$X_{train\_norm} = \frac{X_{train} - min(X_{train})}{max(X_{train}) - min(X_{train})} \quad (4)$$

where X_{test_norm} represents normalized test data, X_{train_norm} represents normalized train data, X_{train} represents test data without normalization and X_{train} represents train data without normalization.

E. Distance Metric

For each new arriving test data, certain similarity or closeness criteria was computed from the training data [5]. Decision was made on the test data by gauging the output from these functions, based on the class of the training data of the highest proximity. Distance metric plays a pivotal role in KNN classification as it will affect the performance in terms of accuracy as well as time and space complexity.

In this study, several variation of Minkowski distance function was employed for evaluating the similarity between training set and testing set for classification purposes.

Generally, Minkowski distance is given by:

$$D = \sqrt[p]{\sum \left| x_{train\_norm} - x_{test\_norm} \right|^p} \quad (5)$$

By setting p=1, Manhattan distance is obtained. Also known as L_1 distance, taxicab distance or city block distance, this metric is essentially the sum of differences across dimensions.

$$D = \sum \left| x_{train\_norm} - x_{test\_norm} \right| \quad (6)$$

This distance function was previously applied for KNN in various dataset [18]-[20].

By setting p=2, Euclidean distance is obtained:

$$D = \sqrt[2]{\sum \left| x_{train\_norm} - x_{test\_norm} \right|^2} \quad (7)$$

Euclidean distance is fairly popular for the implementation of KNN approach in previous literatures [18], [20] and [21]:

By setting p=∞, Chebyshev distance, also known as Chessboard distance is produced [9], [20].

$$D = \lim_{p\to\infty} \left(\sum \left| x_{train\_norm} - x_{test\_norm} \right|^p \right)^{1/p} = max\left(\left| x_{train\_norm} - x_{test\_norm} \right| \right) \quad (8)$$

F. K Parameter Optimization

For KNN, a small k parameter may cause the result to be susceptible to noise whereas a large k value may unnecessarily include too many points from the other class. Hence, it is imperative to fine tune the optimal value of k in order to achieve best performance.

In evaluating and comparing the different size of nearest neighbor, ten-fold cross validation was performed on the data consisting of 10321 cycles of ANO state and 9976 cycles of normal ejection states. The dataset was first randomly divided into ten mutually exclusive subsets of the same proportions. One subset was used as the testing set whereas the remaining nine sets act as training sets. This procedure was repeated for each fold, with one subset from the original training subset taking turn in being the testing set. The performance of classification was taken from averaging the results from each fold. With minimal effect of individual subset characteristic, a more reliable estimate of performance is reflected from the multiple validations [22].

Hence, different odd values of k parameters ranging from 1 to 29 was tested and compared under cross validated testing to seek optimization. Classification performance was evaluated by sensitivity (true positive), specificity (true negative) and accuracy.

III. Results and Discussion

Fig. 3 shows the distribution of the normal ejection data and ANO data for a ramp study on greyhounds. It can be seen from the distribution that ANO data lies in a more concentrated clusters while normal ejection states shows a more dispersive trend from the feature on pump speed. Classifier such as KNN plays a major role in differentiating these two classes so that ANO detection can be accomplished.

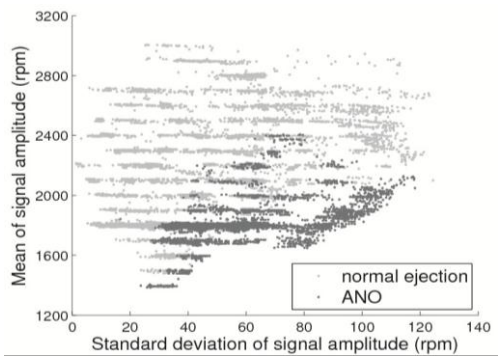


Figure 3. Data spread of normal ejection and ANO state distribution.

Although Minkowski distance generally is said to be more sensitive to the scales of the feature magnitude as compared to the Mahalanobis distance, the inherent weakness can be resolved by normalization of all features prior to the classification task. Two tested scaling methods, namely z-score normalization and min-max normalization were found to yield same sensitivity, specificity and accuracy from the cross validated results of this study.

Initially, the performance improves with the increase of k parameter, due to the expansion of nearest neighbor size to include more neighboring points that were taken into consideration. However, with further increase in the said parameter, the performance has reached a plateau state where the classification capability limit has already been reached for the data. Hence, from the plotted Fig. 4, it can be observed that k=9 is optimal for effective classification of ANO from normal ejection state for all tested distance metrics.

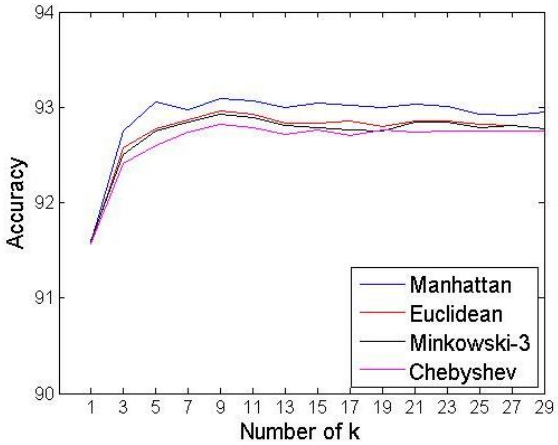


Figure 4. Comparison of performance of different distance metric and k parameter.

Generally all Minkowski distance has yielded comparable performance in terms of sensitivity, specificity and accuracy as shown in Table II.

TABLE II. Statistical Performance of the Different Distance Metric for k=9

Types of Minkowski distance	Percentage		
	Sensitivity	Specificity	Accuracy
Manhattan distance	90.8	95.4	93.1
Euclidean distance	90.7	95.3	93.0
Minkowski distance with p=3	90.7	95.3	92.9
Chebyshev distance	90.5	95.2	92.8

IV. Conclusion

All the tested distance metrics have unanimously shown that KNN is an appropriate approach to perform classification task to differentiate between ANO and normal ejection states for VAD, achieving accuracy at around 93%

REFERENCES

[1] A. S. Go, D. Mozaffarian, V. L. Roger, E. J. Benjamin, J. D. Berry, W. B. Borden, and M. B. Turner, "Heart disease and stroke statistics—2013 update: A report from the american heart association," *Circulation*, vol. 127, no. 1, pp. e6-e245, January 1, 2013.

[2] M. Granegger, F. Moscato, S. Mahr, G. Wieselthaler, and H. Schima, "Assessment of the aortic valve opening during rotary blood pump support," *ASAIO J*, vol. 57, p. 75, 2011.

[3] P. J. Ayre, N. H. Lovell, R. W. Morris, M. L. Wilson, and J. C. Woodard, "Identifying physiologically significant pumping state transitions in implantable rotary blood pumps used as left ventricular assist devices: An in-vivo study," pp. 445-448.

[4] A. G. Rose, S. J. Park, A. J. Bank, and L. W. Miller, "Partial aortic valve fusion induced by left ventricular assist device," *The Annals of Thoracic Surgery*, vol. 70, no. 4, pp. 1270-1274, 2000.

[5] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21-27, 1967.

[6] D. Hongxiang, B. Zhansheng, and Y. Xilin, "The application of improved K-Nearest neighbor classification in topic tracking," pp. V2-64-V2-68.

[7] Y. Liao and V. R. Vemuri, "Use of K-nearest neighbor classifier for intrusion detection," *Computers & Security*, vol. 21, no. 5, pp. 439-448, 2002.

[8] Q. P. He and W. Jin, "Fault detection using the k-nearest neighbor rule for semiconductor manufacturing processes," *IEEE Transactions on Semiconductor Manufacturing*, vol. 20, no. 4, pp. 345-354, 2007.

[9] C. Shahabi, M. R. Kolahdouzan, and M. Sharifzadeh, "A road network embedding technique for k-nearest neighbor search in moving object databases," *GeoInformatica*, vol. 7, no. 3, pp. 255-273, 2003.

[10] O. Golubitsky and S. M. Watt, "Distance-based classification of handwritten symbols," *International Journal on Document Analysis and Recognition*, vol. 13, no. 2, pp. 133-146, 2010.

[11] E. K. Garcia, S. Feldman, M. R. Gupta, and S. Srivastava, "Completely lazy learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 9, pp. 1274-1285, 2010.

[12] N. Bhatia, "Survey of nearest neighbor techniques," *IJCSIS*, vol. 8, no. 2, pp. 302-305, 2010.

[13] N. Roussopoulos, S. Kelley, and F. Vincent, "Nearest neighbor queries," *ACM Sigmod Record*, vol. 24, no. 2, pp. 71-79, 1995.

[14] J. Blitzer, K. Q. Weinberger, and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," in *Proc. Advances in Neural Information Processing Systems 18*, 2006, pp. 1473-1480.

[15] D. M. Karantonis, S. L. Cloherty, D. G. Mason, R. F. Salamonsen, P. J. Ayre, and N. H. Lovell, "Automated non-invasive detection of pumping states in an implantable rotary blood pump," in *Proc. Conf. IEEE Eng Med Biol Soc.*, 2006, pp. 5386-5389.

[16] E. Lim, S. Dokos, R. F. Salamonsen, F. L. Rosenfeldt, P. J. Ayre, and N. H. Lovell, "Numerical optimization studies of cardiovascular–rotary blood pump interaction," *Artificial Organs*, vol. 36, no. 5, pp. E110-E124, 2012.

[17] L. A. Shalabi, Z. Shaaban, and B. Kasasbeh, "Data mining: A preprocessing engine," *Journal of Computer Science*, vol. 2, no. 9, pp. 735, 2006.

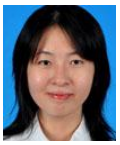
[18] J. Wang, P. Neskovic, and L. N. Cooper, "Improving nearest neighbor rule with a simple adaptive distance measure," *Pattern Recognition Letters*, vol. 28, no. 2, pp. 207-213, 2007.

[19] V. Garcia, E. Debreuve, and M. Barlaud, "Fast k nearest neighbor search using GPU," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2008, pp. 1-6.

[20] P. S. Lakshmi, V. Sushma, and T. Manasa, "Different similarity measures for text classification using Knn."

[21] D. O. Loftsgaarden and C. P. Quesenberry, "A nonparametric estimate of a multivariate density function," *The Annals of Mathematical Statistics*, pp. 1049-1051, 1965.

[22] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proc. International Joint Conference on Artificial Intelligence*, 1995, pp. 1137-1145.



Hui-Lee Ooi received her B.Eng. degree in electronic engineering from Universiti Sains Malaysia, in year 2012. She is currently pursuing MEngSc in University of Malaya. Her research interest is pumping state detection.

Siew-Cheok Ng received his B.Eng, MEngSc and PhD from Universiti of Malaya. He is now a lecturer in Department of Biomedical Engineering, University of Malaya in Malaysia. His research area is biomedical signal processing.

Einly Lim received her B.Eng and MEngSc from University of Malaya in Biomedical Engineering. She then received her PhD from University of New South Wales. She is now a senior lecturer in Department of Biomedical Engineering, University of Malaya in Malaysia. Her research areas are cardiac modeling, cardiac rehabilitation, physiological control and ventricular assist devices.