e-ISSN: 2948-4065
Volume 3 Number 1 July 2025, 9 ~ 14
DOI: https://doi.org/10.37698/1xcynk87

A Hybrid Bearing Monitoring Model Integrating Cosine-Difference and Weightiness in Nearest Neighbors Algorithm

J.S. Yap[1], M.H. Lim[1], M. Salman Leong[1] and W.K. Ngui[2]
[1]Institute of Noise and Vibration, Universiti Teknologi Malaysia, Kuala Lumpur 54100, Malaysia
[2]Faculty of Mechanical & Automotive Engineering Technology. Universiti Malaysia Pahang, 26600 Pekan, Pahang, Malaysia

ARTICLE INFO

Article history:

Received May 06, 2025
Revised Jun 08, 2025
Accepted Jul 15,2025

Keywords:

Bearing Faults,
Condition Monitoring,
Vibration Analysis,
Machine Learning

Conflict of Interest:

None

Funding:

ABSTRACT

This study presents a novel hybrid bearing condition monitoring model, CosWNN, which integrates cosine difference and weightiness within the k-nearest neighbors algorithm. The model addresses key challenges in vibration analysis, specifically the need for efficient computational resources and the scarcity of real-world faulty bearing data. By minimizing signal processing requirements and maintaining classification accuracy with limited data, CosWNN achieves an average accuracy of 77.1%, outperforming traditional nearest neighbors algorithms by 4.4% to 49.5%. Despite these advancements, the model's performance diminishes with fewer training samples, indicating the necessity for further optimization, including the adjustment of the quantity of nearest neighbors and the incorporation of data augmentation techniques. The study underscores the potential of CosWNN for robust bearing fault detection and its applicability in scenarios with constrained data and computational resources.

Corresponding Author: Yap Jee Siang, Institute of Noise and Vibration, Universiti Teknologi Malaysia, Kuala Lumpur 54100, Malaysia. Tel: +603-26154924. E-mail: yapjeesiang@gmail.com

© J.S. Yap, M.H. Lim, M. Salman Leong and W.K. Ngui
This is an open-access article under the CC BY-SA 4.0 international license.

1. Introduction

Monitoring of operating bearing's condition is important to reduce occurrence of machine failure as well with safety mishap [1]. Vibration analysis is one of the main approaches by acquiring the machine's vibration signal using accelerometer, these signals exhibit different characteristic depending on types of bearing fault compared with healthy conditions [2]. As the artificial intelligence (A.I) gaining research attentions, A.I. algorithms are widely applied in monitoring model to classify the bearing conditions [3]. In this study, a bearing monitoring model by integrating Cosine-Difference and Weightiness in Nearest Neighbors (CosWNN) is developed, with objectives to improves the real-world's shortcoming of the following:

- Requirement of computation resources: More powerful processors and memory are essential for signal processing and model training [4].
- Limited data for Model's training: Real-world faulty bearing data is hard to acquired due to operating machine is either broke down or forced stopped when fault is alerting [5].

2. Literature Survey

A brief survey was conducted in this study, the recent solutions of mentioned challenges were discussed, and the survey finding was summarized.

2.1 Challenges

2.1.1 Challenge 1: Efficiency in Computation Resources

Signal processing and A.I. model training is associated to costly data acquisition panels and complicated program which increasing cost of condition monitoring system [6], [7]. Erica et. al explored an affordable-cost and compact alternative by measuring acoustic noise and sampled by a microcontroller, however the vibration data transformation such as FFT (Fast-Fourier-Transform), and top-flat windowing were inevitable in the system [6]. Thani et. al proposed an A.I. model which analyses raw vibration data using Auto-Encoder as feature extractor, a classification of 90.3% (F1-score) was achieved however the model training time was longer [4].

2.1.2 Challenge 2: Training of Model with Limited Data

Most of the intelligent condition monitoring system are data-driven, the classification performance depends on the amount of training information [8]. Yang et. al adopted Conditional-Generative-Adversarial-Network (CGAN) for synthesize training data and processed it into 2-D image [8], [9]. Zhang et. al applied few-shot learning with pair of Siamese Convolutional Neural Networks (CNN), accuracy of 82.8% was found with only 60 training samples in raw vibration signals [9].

2.2 Survey Finding

As the survey finds, there is lack of studies which addressed the efficient computational resource and limited training data challenges. Therefore, CosWNN aims to perform acceptable classification results by taking data inputs without signal processing, while retaining performance with limited training data.

3. Methodology

3.1 Algorithms

Nearest Neighbors algorithm (also known as k-NN) is a classic A.I. method to determine the calculated distance between testing and training data [10]. The training data in k-NN is stored in the model rather than trained as compared to other A.I. methods. Hence the computation resources required for model training is excluded for k-NN [11].

3.2 Hyperparameters

In this study, multiple k-NN algorithms were evaluated using MATLAB's Classification Learner Toolbox [12], subsequently the CosWNN model was developed by integrating two of the k-NN algorithms. **Table 1** shows the algorithms studied and corresponding hyperparameters.

Table 1. The algorithms studied and corresponding hyperparameters.

k-NN Algorithm	Abbreviation	Quantity of Neighbors	Difference Metric	Weightiness
Fine-k-NN	Fk-NN	1	Euclidean	Identical
Medium-k-NN	Mk-NN	10	Euclidean	Identical
Coarse-k-NN	Ck-NN	100	Euclidean	Identical
Cosine-k-NN	Cosk-NN	10	Cosine	Identical
Cubic-k-NN	Cubk-NN	10	Minkowski	Identical
Weighted-k-NN	Wk-NN	10	Euclidean	1/Squared
Proposed Hybrid Model	CosWNN	10	Cosine	1/Squared

3.3 Cosine-Difference

The proposed hybrid model is established on the Cosine-Difference [2] as below

$$\begin{aligned} \mathrm{Dc}\,(\mathrm{J},\mathrm{K}) &= 1 - \mathrm{Sc}(\mathrm{J},\mathrm{K}) \\ &= 1 - \frac{\mathrm{J}\,.\,\mathrm{K}}{||\mathrm{J}||\ .\ ||\mathrm{K}||} \\ &= 1 - \frac{\sum_{x=1}^{n} \mathrm{J}_x \mathrm{K}_x}{\sqrt{\sum_{x=1}^{n} \mathrm{J}_x^2 \cdot \sum_{x=1}^{n} \mathrm{K}_x^2}} \end{aligned} \tag{1}$$

Note: Dc is the Cosine-Difference, Sc is the Angular Similarity, J is the Testing data in time-series, and K is the Stored data in time-series.

3.4 Weightiness

Encompassing Weightiness on the Cosine-Distance can potentially boost A.I. algorithms' result compared to identical weight [10], [13]. The Weightiness technique manipulated in the proposed algorithm is '1/ Squared' as shown

$$w_i(J, K) = \frac{1}{D_c(J,K)^2} \quad (2)$$

Note: w is the Weightiness.

3.5 Set of Rules

The algorithm of proposed hybrid model is described in Table 2 as following

Table 2. The Algorithms of Proposed Hybrid Model.

CosWNN Set of rules
Initial:
J, c, K // *J*: testing information; c: class; *K*: stored information
for n **to training information dimension do:**
Determine the cosine-difference *Dc*(J,*K*) as illustrated in *formula (1)*
end for
Choose the anticipated k (quantity of nearest neighbors)
Sort the Dc by rising sequence
Calculate the observation of each class of the highest k
Multiply the w for each observation
Yield:
Delegate *J* to the maximum gathered w of class, c

3.6 Data Source

2 sets of bearing data were adopted for algorithm are: (a) Case-Western-Reserve-University Bearing (CWRU); (b) SpectraQuest Machinery Fault Simulator by Universiti-Teknologi-Malaysia (UTM). The explanation of the datasets is elaborated in **Table 3**.

Table 3. The bearing description of used datasets.

Dataset	**CWRU**	**UTM**
Bearing Model	6205-2RS JEM SKF	SpectraQuest customised bearing
Motor Load (hp)	0	
Rotational Speed (rpm)	1772	1800
Sampling Rate (Hz)	12000	8000
Class of the condition of bearings	No-Defect-(N), Outer-Race-Defect-(OR), Inner-Race-Defect (IR), and Ball-Defect-(B)	

3.7 Model Training

Several nearest neighbors of model versions were trained with different training amount of training samples is listed in **Table 4**. All the models were tested with separate labelled samples.

Table 4. The amount of training samples used for model training.

Training Sample Amount	**T100**	**T80**	**T60**	**T40**	**T20**
No-Defect-(N) Outer-Race-Defect-(OR) Inner-Race-Defect (IR) Ball-Defect-(B)	150	120	90	60	30
Overall	600	480	360	240	120

3.8 Quantification Of A.I Results

Classification Accuracy is used to quantify the results of A.I. models as commonly adopted by other researchers [16]. The Classification Accuracy is explained as follows

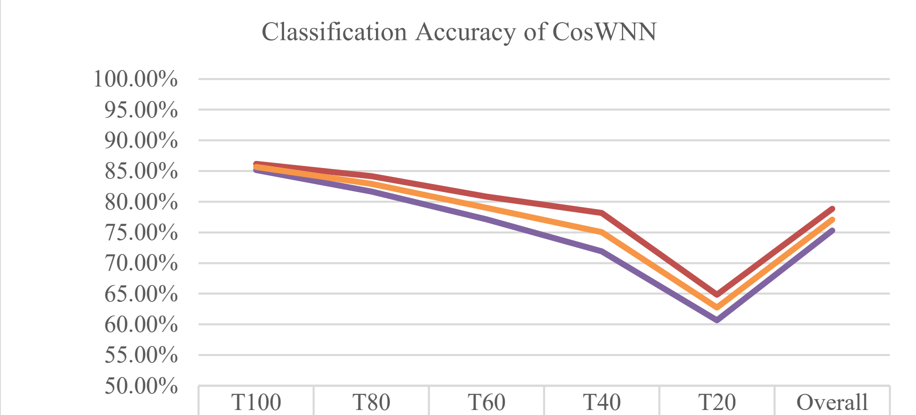
$$\text{Classification Accuracy (\%)} = \frac{\text{TP+TN}}{\text{TP+UP+FN+UN}} \times 100 \qquad (3)$$

Note: TP =observed true positives; TN = observed true negatives;
UP =observed untrue positives; and UN = observed untrue negatives.

4. OUTCOMES & discussions

4.1 Findings

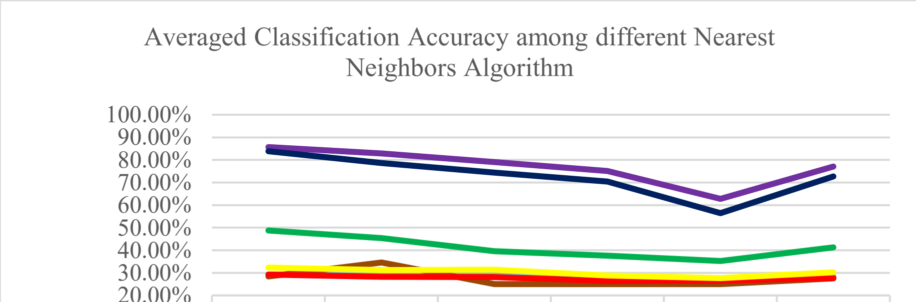
The results of the proposed hybrid model for CWRU and UTM datasets and an average of the two datasets are displayed in **Figure 1**. Both CWRU and UTM datasets produced similar results with overall accuracy of 78.8% and 75.3% respectively. the highest accuracies were achieved when the amount of training samples was 100% utilized, and degradations were observed when the training samples were reduced. An average classification accuracy of 77.1% was calculated for the proposed hybrid model.



	T100	T80	T60	T40	T20	Overall
CWRU	86.17%	84.17%	80.83%	78.17%	64.83%	78.83%
UTM	85.17%	81.67%	77.13%	71.91%	60.67%	75.31%
Average	85.67%	82.92%	78.98%	75.04%	62.75%	77.07%

Figure 1. The classification accurateness of proposed hybrid model.

The average results are compared with other nearest neighbors are illustrated in **Figure 2**. All algorithms except for CosWNN and Cosk-NN were found to perform poorly in bearing condition's classification below 50%. Cosk-NN obtained a relatively decent performance with an average of 72.7%. By integrating the weightiness into Cosk-NN, the proposed hybrid model showed an improvement of 4.4% over Cosk-NN. Nevertheless, CosWNN's classification degradation up to 22.9% in event of the reduction of training samples were observed.



	T100	T80	T60	T40	T20	Overall
CosWNN	85.67%	82.92%	78.98%	75.04%	62.75%	77.07%
Fk-NN	48.75%	45.42%	39.59%	37.59%	35.25%	41.32%
Mk-NN	29.25%	29.25%	29.92%	27.33%	26.25%	28.40%
Ck-NN	28.42%	34.59%	25.09%	25.00%	25.00%	27.62%
Cosk-NN	83.84%	78.59%	74.34%	70.42%	56.42%	72.72%
Cubk-NN	29.25%	28.33%	28.17%	26.42%	26.58%	27.75%
Wk-NN	32.34%	31.25%	31.33%	28.83%	27.67%	30.28%

Figure 2. Averaged Classification Accuracy among different Nearest Neighbors Algorithm.

4.2 Discussions

Most of the nearest neighbors' difference metric were found to be ineffective to classify vibration signals without signal processing except for Cosine-Difference. The preference of Cosine distance over Euclidean and Minkowski distances shall be validated with other datasets, for example unprocessed signals of Machinery Fault Database (MaFaulDa) and Paderborn University[17].

Integrating Weightiness into the proposed model showed an average 4.4% over the predecessor (Cosk-NN), hence hybridizing Cosine-Difference and Weightiness method in the nearest neighbors could potentially improves the overall classification results.

Although the proposed hybrid model showed improvement over other algorithms, an average of 77.1% classification accuracy is not considerable excellent, and an averaged degradation up to 22.9% when training samples were reduced. More improvements of the condition monitoring model are recommended to fulfill the objectives of this study.

One of the recommendations is to optimize the selection of k (number of nearest neighbors) rather than using the default k=10. Other than that, if limited training data is available, data augmentation shall be explored as suggested by other studies [18], [19].

Conclusion

In conclusion, this study addresses the challenges of efficient computation resources and limited training data in bearing fault detection by developing a hybrid monitoring model integrating cosine-difference and weightiness into the k-nearest neighbors algorithm. The proposed CosWNN model achieved a classification accuracy of 77.1%, surpassing other algorithms by 4.4% to 49.5%, especially under conditions of insufficient training data. Despite this improvement, the model's performance degrades significantly with reduced training samples, highlighting the need for further optimization, such as refining the selection of the number of nearest neighbors (k) and exploring data augmentation techniques. The hybrid model demonstrates potential for effective bearing condition monitoring with limited faulty data and minimal signal processing resources, warranting further validation with additional datasets and enhancements.

Acknowledgments

The authors extend their heartfelt thanks to the Institute of Noise and Vibration, UTM, for their financial support of this study. This research was funded under the Higher Institution Centre of Excellence (HICoE) Grant Scheme, specifically the Ad vanced Health Monitoring for Turbomachinery project (R.K130000.7843.4J227) and the Health Monitoring and Integrity Assessment of Ageing Assets project (R.K130000.7843.4J228).

References

J. Xu, P. Xu, Z. Wei, X. Ding, and L. Shi, 'DC-NNMN: Across components fault diagnosis based on deep few-shot learning', Shock and Vibration, vol. 2020, 2020, doi: 10.1155/2020/3152174.

X. Liu, H. Huang, and J. Xiang, 'A personalized diagnosis method to detect faults in a bearing based on acceleration sensors and an fem simulation driving support vector machine', Sensors (Switzerland), vol. 20, no. 2, Jan. 2020, doi: 10.3390/s20020420.

Q. Lu, X. Shen, X. Wang, M. Li, J. Li, and M. Zhang, 'Fault Diagnosis of Rolling Bearing Based on Improved VMD and KNN', Math Probl Eng, vol. 2021, 2021, doi: 10.1155/2021/2530315.

M. A. Tnani, P. Subarnaduti, and K. Diepold, 'Efficient Feature Learning Approach for Raw Industrial Vibration Data Using Two-Stage Learning Framework', Sensors, vol. 22, no. 13, Jul. 2022, doi: 10.3390/s22134813.

H. Zhu, Z. Huang, B. Lu, F. Cheng, and C. Zhou, 'Imbalance domain adaptation network with adversarial learning for fault diagnosis of rolling bearing', Signal Image Video Process, vol. 16, no. 8, pp. 2249–2257, Nov. 2022, doi: 10.1007/s11760-022-02190-7.

E. Raviola and F. Fiori, 'A low-cost, small-size, and bluetooth-connected module to detect faults in rolling bearings', Applied Sciences (Switzerland), vol. 10, no. 16, Aug. 2020, doi: 10.3390/app10165645.

Y. Hu et al., 'Few-shot transfer learning with attention for intelligent fault diagnosis of bearing', Journal of Mechanical Science and Technology, vol. 36, no. 12, pp. 6181–6192, Dec. 2022, doi: 10.1007/s12206-022-1132-4.

J. Yang, J. Liu, J. Xie, C. Wang, and T. Ding, 'Conditional GAN and 2-D CNN for Bearing Fault Diagnosis with Small Samples', IEEE Trans Instrum Meas, vol. 70, 2021, doi: 10.1109/TIM.2021.3119135.

A. Zhang, S. Li, Y. Cui, W. Yang, R. Dong, and J. Hu, 'Limited data rolling bearing fault diagnosis with few-shot learning', IEEE Access, vol. 7, pp. 110895–110904, 2019, doi: 10.1109/ACCESS.2019.2934233.

B. KARABULUT, G. ARSLAN, and H. M. ÜNVER, 'A Weighted Similarity Measure for k-Nearest Neighbors Algorithm', Celal Bayar Üniversitesi Fen Bilimleri Dergisi, vol. 15, no. 4, pp. 393–400, Dec. 2019, doi: 10.18466/cbayarfbe.618964.

D. Lopez-Bernal, D. Balderas, P. Ponce, and A. Molina, 'Education 4.0: Teaching the basics of knn, lda and simple perceptron algorithms for binary classification problems', Future Internet, vol. 13, no. 8, Aug. 2021, doi: 10.3390/fi13080193.

M. Senthamil Selvi and S. Jansi Rani, 'Classification of Admission Data Using Classification Learner Toolbox', in Journal of Physics: Conference Series, IOP Publishing Ltd, Aug. 2021. doi: 10.1088/1742-6596/1979/1/012043.

A. Sharma, R. Jigyasu, L. Mathew, and S. Chatterji, 'Bearing Fault Diagnosis Using Weighted K-Nearest Neighbor', in Proceedings of the 2nd International Conference on Trends in Electronics and Informatics, ICOEI 2018, Institute of Electrical and Electronics Engineers Inc., Nov. 2018, pp. 1132–1137. doi: 10.1109/ICOEI.2018.8553800.

Case School of Engineering, 'Bearing Data Center', Case Western Reserve University Bearing Data Center. Accessed: Sep. 10, 2023. [Online]. Available: https://engineering.case.edu/bearingdatacenter

S. R. Saufi, M. F. Isham, Z. A. Ahmad, and M. D. A. Hasan, 'Machinery fault diagnosis based on a modified hybrid deep sparse autoencoder using a raw vibration time-series signal', J Ambient Intell Humaniz Comput, vol. 14, no. 4, pp. 3827–3838, Apr. 2023, doi: 10.1007/s12652-022-04436-1.

N. Fathiah Waziralilah, A. Abu, M. H. Lim, L. K. Quen, and A. Elfakharany, 'Bearing fault diagnosis employing Gabor and augmented architecture of convolutional neural network', Journal of Mechanical Engineering and Sciences, vol. 13, no. 3, pp. 5689–5702, 2019, doi: 10.15282/jmes.13.3.2019.29.0455.

S. R. Saufi, Z. A. Bin Ahmad, M. S. Leong, and M. H. Lim, 'Challenges and opportunities of deep learning models for machinery fault detection and diagnosis: A review', IEEE Access, vol. 7. Institute of Electrical and Electronics Engineers Inc., pp. 122644–122662, 2019. doi: 10.1109/ACCESS.2019.2938227.

J. W. Oh and J. Jeong, 'Data augmentation for bearing fault detection with a light weight CNN', in Procedia Computer Science, Elsevier B.V., 2020, pp. 72–79. doi: 10.1016/j.procs.2020.07.013.

P. Lyu, H. Zhang, W. Yu, and C. Liu, 'A novel model-independent data augmentation method for fault diagnosis in smart manufacturing', in Procedia CIRP, Elsevier B.V., 2022, pp. 949–954. doi: 10.1016/j.procir.2022.05.090.

The Author(s):

Yap Jee Siang (https://orcid.org/0000-0002-7935-4455) is a Research Officer at Institute of Noise & Vibration, Universiti Teknologi Malaysia Kuala Lumpur, Jalan Sultan Yahya Petra (Jalan Semarak),54100 Kuala Lumpur, Malaysia. Tel: +603-26154924. Email: yapjeesiang@gmail.com

Lim Meng Hee (https://orcid.org/0000-0002-0030-0455) is an Associate Professor at Institute of Noise & Vibration, Universiti Teknologi Malaysia Kuala Lumpur, Jalan Sultan Yahya Petra (Jalan Semarak),54100 Kuala Lumpur, Malaysia. Tel: +603-26154924. Email: limmenghee@gmail.com

M Salman Leong (https://orcid.org/0000-0002-2411-6089) is the Director at Institute of Noise & Vibration, Universiti Teknologi Malaysia Kuala Lumpur, Jalan Sultan Yahya Petra (Jalan Semarak),54100 Kuala Lumpur, Malaysia. Tel: +603-26154924. Email: limmenghee@gmail.com

Ngui Wai Keng (https://orcid.org/0000-0003-4781-1321) is an Associate Professor at Faculty of Mechanical & Automotive Engineering Technology, Universiti Malaysia Pahang, 26600 Pekan, Pahang, Malaysia. Tel: +60 9-424 6200. Email: wkngui@umpsa.edu.my