Deep Generalized Canonical Correlation Analysis for Motor Fault Diagnosis

Guo, Peikun; Liu, Dehong; Wang, Bingnan; Wang, Yebin; Inoue, Hiroshi; Kanemaru, Makoto

TR2025-085 June 17, 2025

Abstract

While motor current signature analysis (MCSA) is widely used for motor fault detection, it has shown limitations as a single modality sensor-based method in diagnosing some types of faults such as bearing roughness. This paper introduces a novel framework for motor fault diagnosis using multi-modal sensor data. The framework addresses the challenges of multi-modal sensor fusion in motor fault diagnosis by first aligning features from various sensors in a shared latent space using deep generalized canonical analysis (DGCCA) and then incorporating attention-based fusion to weigh the contribution of feature channels. Validated on datasets collected on motors with different bearing friction levels, the approach demonstrates significant performance gains over baseline methods, achieving high accuracy and robustness across varying operating conditions. The proposed method is positioned as a scalable and effective solution for industrial motor fault diagnosis applications.

2025 IEEE Industry Applications Society Annual Meeting (IAS)

Deep Generalized Canonical Correlation Analysis for Motor Fault Diagnosis

Peikun Guo, Dehong Liu, Bingnan Wang, Yebin Wang,
Mitsubishi Electric Research Laboratories (MERL)
Cambridge, MA, USA

Hiroshi Inoue, Makoto Kanemaru
Mitsubishi Electric Corporation
Amagasaki, Japan

***Abstract*—While motor current signature analysis (MCSA) is widely used for motor fault detection, it has shown limitations as a single modality sensor-based method in diagnosing some types of faults such as bearing roughness. This paper introduces a novel framework for motor fault diagnosis using multi-modal sensor data. The framework addresses the challenges of multi-modal sensor fusion in motor fault diagnosis by first aligning features from various sensors in a shared latent space using deep generalized canonical analysis (DGCCA) and then incorporating attention-based fusion to weigh the contribution of feature channels. Validated on datasets collected on motors with different bearing friction levels, the approach demonstrates significant performance gains over baseline methods, achieving high accuracy and robustness across varying operating conditions. The proposed method is positioned as a scalable and effective solution for industrial motor fault diagnosis applications.**

***Index Terms*—Sensor fusion, Canonical correlation, Deep learning, Fault diagnosis**

I. INTRODUCTION

Motors are integral to numerous industrial applications and electrified transportation systems. However, these machines often operate under harsh conditions such as high ambient temperature, high moisture, and overload [1]. Such adverse environments can lead to various motor faults, such as bearing wear, insulation aging, and eccentricity, etc. Among these faults, bearing faults are the most prevalent one, accounting for 30% to 40% of all motor failures, according to the report by the IEEE Industry Application Society and the Japan Electrical Manufacturers' Association [2]. If the motor fault is not detected and resolved, it could result in significant maintenance costs, financial losses, and even safety hazards [3]. Therefore, it is very important to monitor the health condition of motors and to perform timely maintenance on the faulty motor.

To detect motor faults, many different methods have been studied, including physical-model-based methods, machine learning methods, and deep learning methods. Physical-model based methods aim to extract fault signals from sensing data using domain knowledge. For example, the motor current signature analysis (MCSA) method has been a prevailing method in detecting fault signature frequency component induced in the stator current by asymmetric magnetic field caused by faults such as eccentricity and broken-bar fault [1], [4], [5]. The MCSA method typically works well in detecting the asymmetric type of faults for motors operating at steady status. When the motor under test operates under varying conditions, the MCSA method remains effective with the aid of advanced signal processing technologies [6]. However, the MCSA-based method exhibits some limitations in diagnosing different types of faults or estimating the fault severity level such as bearing wear and insulation aging. To address these challenges, machine learning (ML) techniques have been increasingly applied to motor fault diagnosis and severity estimation. Traditional ML algorithms such as artificial neural networks (ANN) and support vector machines (SVM) have shown promise in extracting hidden patterns from data that are difficult for humans to discern [7]–[10]. These techniques have demonstrated high accuracy in estimating the severity level of some faults. In recent years, deep learning (DL) methods have gained popularity for their ability to achieve higher classification accuracy and better performance in fault diagnosis [11]. Among the most popular deep learning models, convolutional neural networks (CNN) [12], recurrent neural networks (RNN) [13], autoencoders (AE) [14], and deep belief networks (DBN) [15], [16] have been successfully applied to various fault diagnosis tasks.

Despite the development of various methods, single modality (for example, current) sensor shows its limitation in detecting motor faults. For instance, current signals are valuable in detecting eccentricity faults, but barely useful in detecting bearing surface wear. While vibration signals are helpful in detecting mechanical faults such as unbalance or bearing roughness, they are not informative for detecting electrical faults such as short-circuit fault. Furthermore, motors are typically operating at varying load and varying speed conditions, fault signatures also vary accordingly, making it challenging to online monitor motors' health conditions. Integrating data from multiple sensors offers a more complete understanding of the motor's condition by capturing various operational aspects. This multi-modal approach enhances detection performance beyond what any single sensor can achieve on its own.

Multi-sensor data fusion has attracted a lot of attention in the machine learning community and has been studied in multiview representation learning methods. Multiview representation is a powerful approach in scenarios where multiple "views" or modalities of data are available [17]. A representation that can explain multiple data views is more likely to capture the essential characteristics than one that only fits a

Part of this work was done when Peikun Guo was an intern at MERL.

single view. Traditional multiview learning techniques often rely on Canonical Correlation Analysis (CCA), a classical statistical method using linear transformations. Existing deep learning-based CCA methods learn nonlinear transformations for better performance but mainly focus on two-view data [18], [19]. Deep CCA (DCCA) is an extension of CCA that addresses the limit of linear transformation by learning a neural network such that non-linear transformations of two vectors are maximally correlated [18]. Deep generalized CCA (DGCCA) generalize DCCA from two sets of vectors to multiple sets of vectors [18], [20].

In the context of motor fault detection, various sensor signals—such as stator current, vibration, and velocity—can be regarded as different views of the motor's condition [21]. However, due to the nonlinear nature of both the motor drive system and the underlying physical principles of each sensing modality, these views are often nonlinearly correlated.

In this paper, we propose a DGCCA-based framework for motor fault detection by integrating data collected from multi-modal sensors, where each modality is represented by a distinct set of data vectors. Our main contributions are as follows:

1) **DGCCA for fault detection**: We use DGCCA to fuse multi-modal sensor data in motor fault detection. By learning a deep neural network, raw multi-modal data are transformed non-linearly into feature space such that multi-modal features are well aligned in a shared latent subspace. These features serve as input vectors for fault detection and severity estimation, enhancing the model's overall performance.
2) **Combination with attention-based fusion**: We incorporate DGCCA with an attention mechanism to effectively fuse features from different sensor modalities. The attention mechanism dynamically learns the contribution of each feature channel based on its relevance to different fault types, enabling reliable fault detection performance.
3) **Experimental validation on real-world data**: We validate our proposed framework using experimental multi-modal sensor data. The results demonstrate that our approach can effectively handle the complexity and nonlinearity of motor systems at varying operating conditions, outperforming methods that rely solely on physical modeling or purely data-driven techniques.

II. Related Works

A. Canonical Correlation Analysis (CCA)

Given two sets of sensor data vectors $X_1 \in \mathbb{R}^{d_1 \times N}$ and $X_2 \in \mathbb{R}^{d_2 \times N}$ representing two different modalities respectively, where N denotes the number of vectors in each set, d_1 and d_2 are the vector lengths, CCA learns two linear transformations $A_1 \in \mathbb{R}^{d_1 \times r}$ and $A_2 \in \mathbb{R}^{d_2 \times r}$ such that the correlation between projected vectors $A_1^T X_1$ and $A_2^T X_2$ is maximized [22]. Denote the covariance of X_1 as $S_{11} = X_1 X_1^T \in \mathbb{R}^{d_1 \times d_1}$, the covariance of X_2 as $S_{22} = X_2 X_2^T \in \mathbb{R}^{d_2 \times d_2}$, and their cross-covariance as $S_{12} = X_1 X_2^T \in \mathbb{R}^{d_1 \times d_2}$, respectively. The CCA objective is to solve an optimization problem

$$\begin{aligned} A_1^*, A_2^* &= \operatorname*{argmax}_{A_1, A_2} \operatorname{corr}(A_1^T X_1, A_2^T X_2) \\ &= \operatorname*{argmax}_{A_1, A_2} \frac{A_1^T S_{12} A_2}{\sqrt{A_1^T S_{11} A_1}\sqrt{A_2^T S_{22} A_2}}. \end{aligned} \tag{1}$$

Let $Z \triangleq S_{11}^{-\frac{1}{2}} S_{12} S_{22}^{-\frac{1}{2}}$ and the Singular Value Decomposition (SVD) of Z be $Z = USV^T$. Then A_1^*, A_2^*, and the total maximum canonical correlation can be achieved as [23]

$$\begin{cases} A_1^* &= S_{11}^{-\frac{1}{2}} U, \\ A_2^* &= S_{22}^{-\frac{1}{2}} V, \\ \operatorname{corr}(A_1^{*T} X_1, A_2^{*T} X_2) &= \operatorname{Tr}[(Z^T Z)^{\frac{1}{2}}], \end{cases} \tag{2}$$

where $\operatorname{Tr}[\cdot]$ represents the trace of the matrix.

One of the key limitations of Canonical Correlation Analysis (CCA) is that it is restricted to linear transformations, which may not capture complex nonlinear relationships between data modalities.

B. Deep Canonical Correlation Analysis (Deep CCA)

Deep CCA (DCCA) follows the idea of CCA and learns non-linear transformations using a pair of neural networks [18]. Let $f_1(\cdot; W_1)$ and $f_2(\cdot; W_2)$ denote two independent neural networks with parameters W_1 and W_2 respectively. The objective of deep CCA is to optimize W_1 and W_2 such that the canonical correlation between the outputs of f_1 and f_2 given two input vectors X_1 and X_2 respectively, denoted as $F_1 = f_1(X_1; W_1)$ and $F_2 = f_2(X_2; W_2)$, can be maximized by finding two linear transformations U_1 and U_2. The objective of deep CCA is to achieve

$$\begin{aligned} \{W_1^*, W_2^*\} &= \operatorname*{argmax}_{\{W_1, W_2\}} \operatorname{CCA}(F_1, F_2) \\ &= \operatorname*{argmax}_{\{W_1, W_2\}} \operatorname{corr}(U_1^T F_1, U_2^T F_2). \end{aligned} \tag{3}$$

In order to update W_1 and W_2, a loss function that measures the canonical correlation must be calculated and back-propagated through the network. Similar to CCA, let C_{11}, C_{22}, and C_{12} be covariances of F_1 and F_2, and their cross-covariance, respectively, and define matrix $E \triangleq C_{11}^{-\frac{1}{2}} C_{12} C_{22}^{-\frac{1}{2}}$. Similar to (2), the linear transformations U_1 and U_2 can be achieved and the canonical correlation can be calculated. To optimize neural network parameters W_1 and W_2 in (3), the loss function of DCCA is defined as the negative canonical correlation, i.e.,

$$\mathcal{L}_{DCCA} = -\operatorname{Tr}[(E^T E)^{\frac{1}{2}}]. \tag{4}$$

Network parameters W_1 and W_2 can be updated by minimizing the DCCA loss $\mathcal{L}_{DCCA}$ in (4) (or equivalently maximizing the total canonical correlation).

C. Deep Generalized CCA (DGCCA)

Deep generalized CCA (DGCCA) extends DCCA from two sets of vectors to M ($M > 2$) sets of vectors. The DGCCA network consists of M fully connected layers that map the input vectors $X_m \in R^{d_m \times N}$ to the network output $F_m = f_m(X_m; W_m) \in R^{o_m \times N}$, where subscript $m(m = 1, ..., M)$ represents the mth view (modality), W_m is the set of parameters of the mth neural network, N is the number of data points (each data point is a vector of length d_m), and o_m is the size of output (feature) from the final layer of the mth neural network. Instead of maximizing the canonical correlation between all output pairs, DGCCA seeks to learn a shared latent representation G across M neural network outputs, each corresponding to a different view (modality), by solving an optimization problem

$$\min_{U_m \in R^{o_m \times r}, G \in R^{r \times N}, W_m} \mathcal{L}_{DGCCA} = \sum_{m=1}^{M} \|G - U_m^T F_m\|_F^2$$
$$s.t. \quad GG^T = I_r, \tag{5}$$

where $G \in R^{r \times N}$ is the shared representation we are interested in learning, $U_m \in R^{o_m \times r}$ represents the mth linear transformation which maps the o_m-dimensional feature of the mth modality to a r-dimensional ($r < o_m$ and $r << N$) representation. $I_r \in R^{r \times r}$ is the identity matrix. This loss function ensures that the outputs of DGCCA are optimally aligned within the shared latent subspace.

III. Proposed Method

We propose a novel approach for sensor fusion that leverages DGCCA combined with a feature attention mechanism. The overall framework is illustrated in Fig. 1, which includes three main modules: feature extraction, DGCCA, and attention fusion, with details described in the following subsections.

This method is designed to extract, align, and fuse features from multiple sensory inputs to enhance the performance of downstream tasks. Signals collected from various sensors—such as current and vibration sensors—attached to motors are first processed to extract modality-specific features. These features are then passed through the DGCCA module, which projects them into a shared latent space by maximizing their correlation. An attention-based fusion mechanism is subsequently applied to combine the correlated features through weighted aggregation, enabling effective fault detection and severity estimation.

A. Feature Extraction

In our experimental framework, we directly feed the model with raw time series inputs collected in our experiments. This approach is justified by the non-periodic nature of some fault features, such as fault feature induced by bearing friction or insulation, which can be more effectively captured in the time domain.

Without loss of generality, let $X_m \in \mathbb{R}^{d_m \times N}$ denote a set of vectors collected by the mth ($m = 1, ..., M$) sensor, where N is the total number of measurements. Each X_m is passed

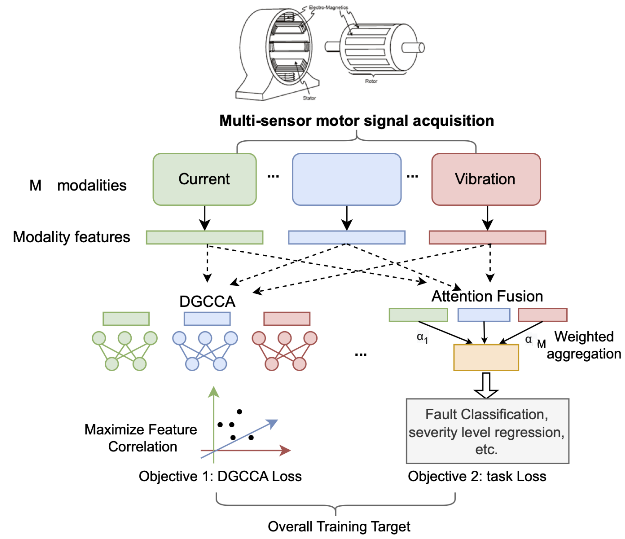


Fig. 1: Overview of the proposed multi-sensor motor fault detection framework.

through a corresponding feature extractor to obtain the feature representations $\{F_m = f_m(X_m, W_m) \in R^{o_m \times N}\}_{m=1}^M$. In particular, we use a 1D-CNN based ResNet-18 (without the last three layers) [24] as the backbone, which includes 4 convolutional layers, each includes 2 residual blocks with 32, 64, 128, and 256 channels respectively. The output from the final block is subsequently fed into the DGCCA module for additional analysis.

B. DGCCA on Feature data

The DGCCA network then processes these features $\{F_m = f_m(X_m, W_m) \in R^{o_m \times N}\}_{m=1}^M$ to align them in a shared latent space G, as formulated in (5).

To train the neural networks in DGCCA, it is necessary to compute the loss of the DGCCA objective $\mathcal{L}_{\text{DGCCA}}$. To this end, we compute the covariance matrices $C_m = F_m F_m^T \in R^{o_m \times o_m}$ for the output of each view. Subsequently, projection matrices $P_m = F_m^T C_m^{-1} F_m \in R^{N \times N}$ for $m = 1, ..., M$ that whitens the data are derived from these covariance matrices. Let $\bar{P} = \sum_{m=1}^M P_m$. It can be shown that the rows of G correspond to the top r (orthonormal) eigenvectors of $\bar{P}$ and the transformations are $U_m = C_m^{-1} F_m G^T$. We can express the objective loss function in (5) in an alternative form as follows

$$\mathcal{L}_{DGCCA} = \sum_{m=1}^{M} \|G - U_m^T F_m\|_F^2 = rM - \text{Tr}(G\bar{P}G^T). \tag{6}$$

C. Attention-Based Fusion

After obtaining the DGCCA-aligned features, we employ an attention-based mechanism to fuse the features effectively [25]. Specifically, a lightweight attention fusion module $g(\cdot; W_g)$ adaptively computes the attention weight vector $\alpha_m = g(F_m; W_g)$ for each feature representation, capturing the relative importance of different feature channels. The attention fusion module includes a two-layer fully connected network with the sigmoid activation function to ensure the attention weight values are in the range of [0,1]. The fused feature representation F is obtained by taking a weighted sum of the DGCCA-aligned features,

$$F = \sum_{m=1}^{M} \alpha_m \circ F_m, \tag{7}$$

where $\circ$ represents element-wise product. Note that the weight vector α_m is channel specific and dynamically derived from the attention mechanism for each data point. This step ensures that the most informative features from each sensor are emphasized, leading to a more robust and discriminative fused representation.

We then use the fused representation for fault detection and severity estimation using a simple feedforward neural network $h(\cdot; W_h)$ to map the fused features to the final prediction classes.

D. Training and testing Processes

We use cross entropy to measure the difference between the network prediction $\hat{y}$ and the ground truth y. Mathematically, this is expressed as

$$\mathcal{L}_{CE}(y, \hat{y}) = -\sum_{i} y_i \log(\hat{y}_i). \tag{8}$$

The total loss of the framework is a weighted sum of the prediction error and the DGCCA loss $\mathcal{L}_{DGCCA}$, i.e.

$$\mathcal{L}_{total} = \mathcal{L}_{CE} + \mu \mathcal{L}_{DGCCA}. \tag{9}$$

The parameters of the feature extractor neural networks $\{W_m\}_{m=1}^{M}$, the attention fusion module W_g, and the feedforward neural network W_h are then updated in the training process by minimizing the total loss. The Adam optimizer with a learning rate 0.0005 is adopted in the training process to implement the optimization process [26]. The detailed steps of the training process are summarized in Algorithm 1.

The testing process is conducted on separate datasets, which are used as inputs to the trained network. The resulting outputs are then compared with the ground truth labels to evaluate the performance of the proposed framework.

IV. Experiments

We evaluate the proposed framework on datasets collected on motors with different bearing roughness levels. We show in Fig.2 (a) a picture of our experiment setup and in Fig. 2 (b) an illustration of the system architecture. A 0.75kW three-phase squirrel-cage induction motor is used for the experimental study. A magnetic powder brake, whose torque can be tuned by changing its input operating current, is used as the load. Data are collected on the motor using multiple sensors of different modalities including three-phase motor current, vibration, velocity, and air gap, etc. The whole motor drive system is enclosed in a clear cage for safety purpose. To create varying levels of roughness on the bearing surface, micro metal particles of three different sizes are added. Smaller particles tend to create smoother surfaces, while larger particles result in rougher textures.

Algorithm 1 Sensor Fusion with Deep Generalized CCA and Feature Attention

Input: Data $\{x_m \in \mathbb{R}^{t_m \times N}\}_{m=1}^{M}$ from M sensors,
Initialize feature extractors and the DGCCA network,
for $i = 1$ to *epoch* **do**
  Preprocess data as input of DGCCA neural network $\{X_m = \Phi(x_m) \in \mathbb{R}^{d_m \times N}\}_{m=1}^{M}$;
  Compute the output feature of the DGCCA neural network $\{F_m = f_m(X_m, W_m) \in R^{o_m \times N}\}_{m=1}^{M}$;
  Compute the GCCA loss of output features $\{F_m\}$:
  - Compute covariance matrices $\{C_m = F_m F_m^T\}_{m=1}^{M}$,
  - Derive projections $\{P_m = F_m^T C_m^{-1} F_m\}_{m=1}^{M}$,
  - Form the sum projection matrix $\bar{P} = \sum_{m=1}^{M} P_m$,
  - Compute G, $\{U_m\}$, and DGCCA loss $\mathcal{L}_{DGCCA}$;

  Apply attention fusion module $g(\cdot; W_g)$ on $\{F_m\}_{m=1}^{M}$ to obtain fused features F;
  Perform classification for fault detection or severity estimation using fused features F as input of a feedforward neural network $h(\cdot; W_h)$;
  Update parameters $\{W_m\}$, W_g, and W_h by minimizing the total loss $\mathcal{L}_{total}$;
end for
Output: Network parameters and final feature representations $\{F_m\}_{m=1}^{M}$, and projection matrices $\{P_m\}_{m=1}^{M}$.

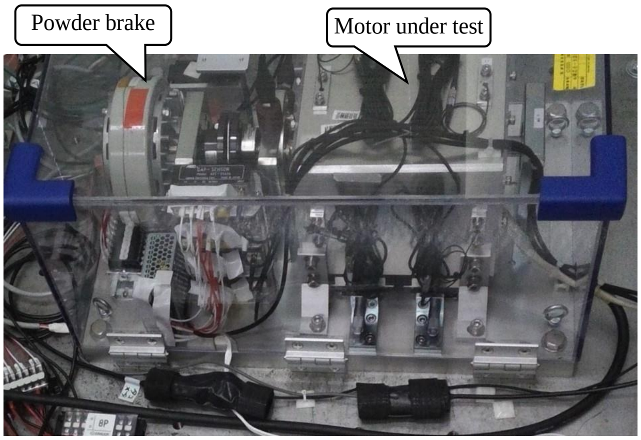
Time-domain sensor signals are sampled at 10kHz, and stored as 60-second recordings. To increase the variety of the dataset, the motor under test is controlled to operate at varying load and speed conditions. Details of the dataset are listed in Table I. We aim to monitor bearing health conditions based on multi-modal ($M = 3$) sensor data collected on motors of different friction levels (small/ medium/ large).

TABLE I: Summary of the dataset metadata for bearing wear classification.

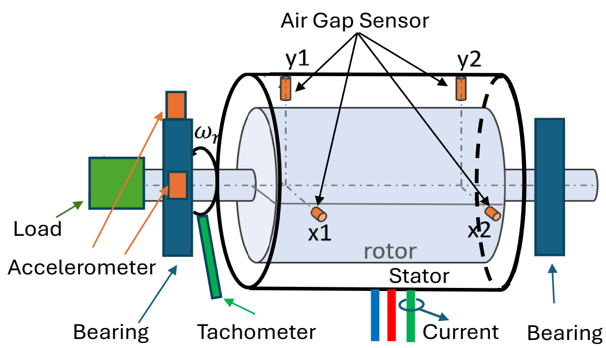
Task	Labels	# of Samples	Channels
Bearing Friction	Small	31	current,
	Medium	10	vibration
	Large	20	air gap

In Fig. 3 we show example snapshots of time-domain sensor data of stator current, acceleration, and air gap at small bearing friction and at large bearing friction respectively. We observe that three different modalities exhibit different frequency characteristics, magnitude ranges, noise levels. From the plot, it is visually challenging to identify the correspondence between the plot and the level of bearing roughness.

We train our model on a NVIDA RTX A2000 12GB GPU, and the weight of DGCCA loss $\mu = 0.05$ using collected datasets. During the training process, we add random noise to the original datasets, with an equivalent signal-to-noise ratio SNR = 20dB, to improve the robustness. The dropout technique [27] is also applied on the fused features with a

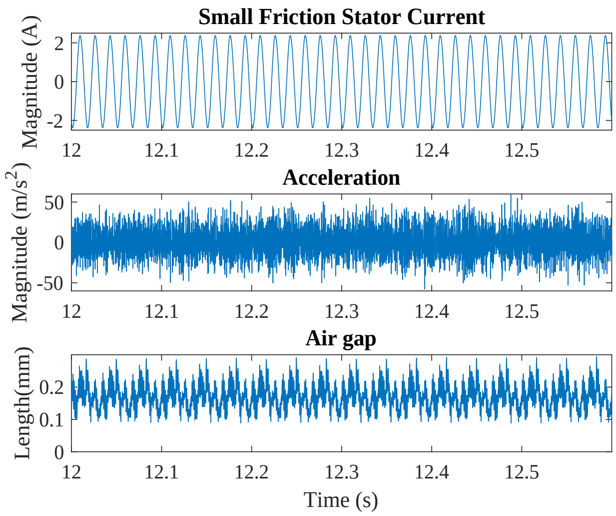


(a)

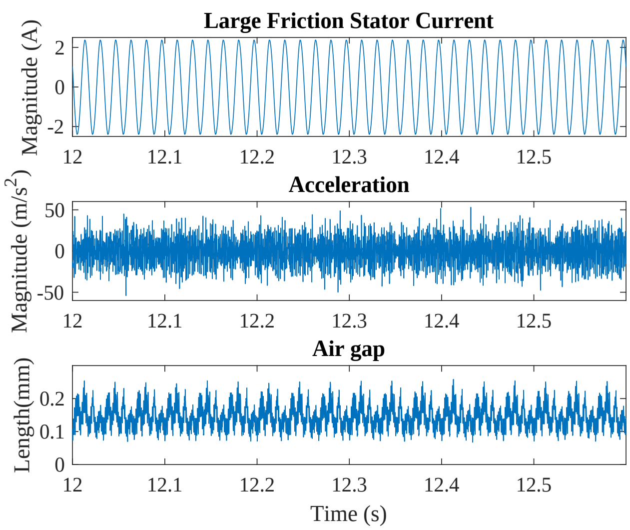


(b)

Fig. 2: (a) Experimental setup; (b) System architecture of the experimental setup.



(a)



(b)

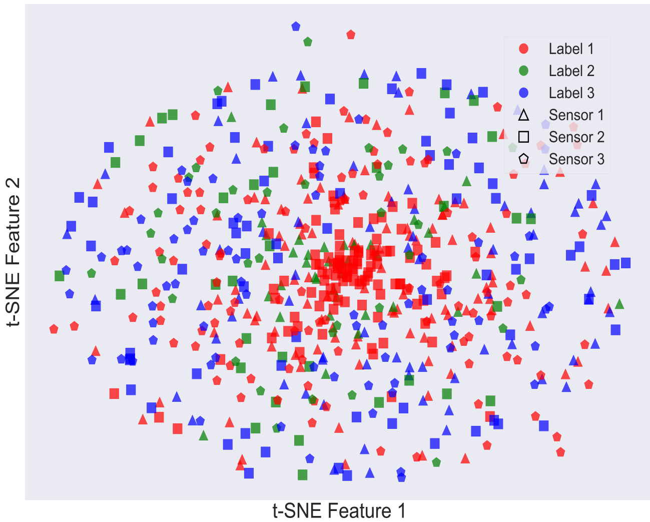
Fig. 3: Time-domain measurements of different modalities at (a) small bearing friction and (b) large bearing friction.

dropout rate $p_{drop} = 0.5$ to prevent overfitting in neural networks by randomly dropping units of feature channels.

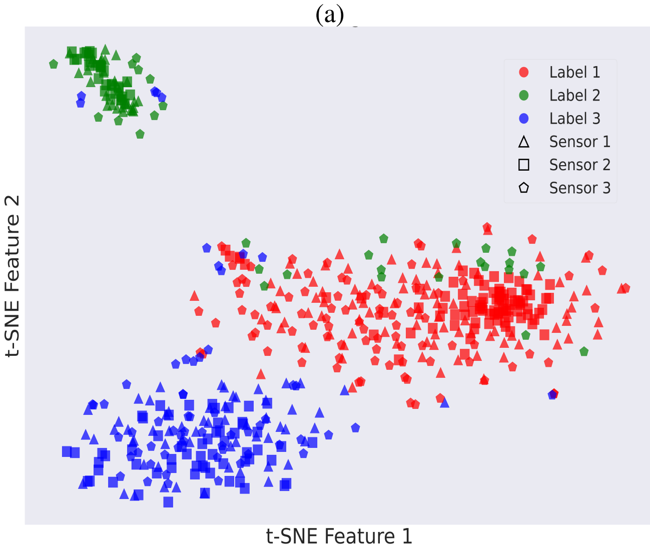
In Fig. 4 (a) and (b) we show example plots of t-distributed stochastic neighbor embedding (t-SNE) [28] visualization of features from the validation set, before and after DGCCA alignment process respectively. In the figure, different colors indicate different bearing wear levels (small, medium, and large) while different symbol shapes represent different sensor modalities (current, vibration, and air gap length). We observe that before DGCCA alignment, data points are scattered in the feature space, which are difficult to be classified. With our proposed DGCCA method, data points are well aligned in the shared latent space for different severity levels, which consequently results improved performance.

In Table II, we compare the average performance of estimating bearing-wear condition over 10 runs using a single channel, multi-channel using ResNet without feature alignment or attention fusion, and our proposed method, respectively. The accuracy and AUROC scores are calculated using the one-vs-one approach by breaking down multi-class classification problems into multiple binary classification problems [29]. It is clear that the proposed method significantly outperform those of single-channel setups. Specifically, the proposed method achieved an accuracy of 98.7% and an AUROC of 0.99, compared to the best single-channel performance of 94.5% accuracy and 0.94 AUROC with vibration data, while some channel (such as current) even failed to have good performance, which is reasonable because bearing wear has negligible impact on the stator current physically. Our method is also better than the ResNet multi-channel method, which is of 97.0% accuracy and 0.96 AUROC, due to the attention-based feature fusion. The experimental results demonstrate the efficacy of leveraging multiple sensor channels for motor fault detection and severity estimation.

From the experiment we observe that each sensor modality contributes unique and complementary information. For instance, current signals are valuable in detecting eccentricity faults, but meaningless in detecting bearing surface wear. Inte-



(a)



(b)

Fig. 4: (a) T-SNE of the features from validation set of bearing data before CCA alignment. (b)T-SNE dimensionality reduction visualization of the features from validation set of bearing data. Different colors indicate different wear levels, and shapes denote different sensory channels.

grating data from multiple sensors provides a comprehensive view of the motor's condition, capturing diverse aspects of its operation. By combining these modalities, the proposed method can detect and diagnose faults more accurately and reliably than using any single modality alone.

The use of DGCCA plays a crucial role in aligning features from different sensor channels into a shared latent space. This alignment ensures that the information from each sensor is effectively integrated, enhancing the overall feature representation. Additionally, the attention fusion mechanism adaptively weighs the importance of each modality and each feature channel, allowing the model to focus on the most relevant features for the fault detection and severity estimation task. This

TABLE II: Performance of bearing wear estimation

Setup	**Channel**	**Accuracy (%)**	**AUROC**
Single Channel	Vibration	94.5	0.94
	Current	54.4	0.37
	Air Gap	78.3	0.92
ResNet Multi-Channel	All	97.0	0.96
Proposed Method	All	98.7	0.99

AUROC: Area Under Receiver Operating Characteristic curve.

combination of DGCCA and attention fusion leads to superior performance, as evidenced by the significant improvements over the baseline multi-channel approach.

V. Conclusion

We proposed a Deep Generalized Canonical Correlation Analysis (DGCCA)-based sensor fusion method incorporating with attention fusion and applied it in the motor bearing wear estimation problem. The effectiveness is validated on experimental datasets with significantly improved performance. This consistent performance improvement highlights the generalizability of our approach across different sensor types and operational conditions, making it a versatile solution for various sensor fusion problems including motor fault diagnosis scenarios. Implementing our multi-sensor fusion approach can lead to more reliable and intelligent condition monitoring systems, ultimately improving operational safety and efficiency.

While our method demonstrates substantial improvements, it depends on the availability of multiple sensor channels, which may not always be accessible in all operational settings. Future work could explore the adaptability of the model to scenarios with limited sensor data.

References

[1] M. E. H. Benbouzid, "A review of induction motors signature analysis as a medium for faults detection," *IEEE Transactions on Industrial Electronics*, vol. 47, no. 5, pp. 984–993, 2000.

[2] P. Albrecht, J. Appiarius, R. McCoy, E. Owen, and D. Sharma, "Assessment of the reliability of motors in utility applications-updated," *IEEE Transactions on Energy Conversion*, no. 1, pp. 39–46, 1986.

[3] P. Zhang, Y. Du, T. G. Habetler, and B. Lu, "A survey of condition monitoring and protection methods for medium-voltage induction motors," *IEEE Transactions on Industry Applications*, vol. 47, no. 1, pp. 34–46, 2010.

[4] R. Kryter and H. Haynes, "Condition monitoring of machinery using motor current signature analysis," Oak Ridge National Lab., Tech. Rep., 1989.

[5] S. Nandi, H. A. Toliyat, and X. Li, "Condition monitoring and fault diagnosis of electrical motors—a review," *IEEE Transactions on Energy Conversion*, vol. 20, no. 4, pp. 719–729, 2005.

[6] V. A. Kelkar, D. Liu, H. Inoue, and M. Kanemaru, "Sparsity-driven joint blind deconvolution-demodulation with application to motor fault detection," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2023, pp. 1–5.

[7] F. Filippetti, G. Franceschini, C. Tassoni, and P. Vas, "Recent developments of induction motor drives fault diagnosis using AI techniques," *IEEE Transactions on Industrial Electronics*, vol. 47, no. 5, pp. 994–1004, 2000.

[8] L. Batista, B. Badri, R. Sabourin, and M. Thomas, "A classifier fusion system for bearing fault diagnosis," *Expert Systems with Applications*, vol. 40, no. 17, pp. 6788–6797, 2013.

[9] M. Cerrada, R.-V. Sánchez, C. Li, F. Pacheco, D. Cabrera, J. V. De Oliveira, and R. E. Vásquez, "A review on data-driven fault severity assessment in rolling bearings," *Mechanical Systems and Signal Processing*, vol. 99, pp. 169–196, 2018.

[10] M. A. Awadallah and M. M. Morcos, “Application of AI tools in fault diagnosis of electrical machines and drives-an overview,” *IEEE Transactions on Energy Conversion*, vol. 18, no. 2, pp. 245–251, 2003.
[11] S. Khan and T. Yairi, “A review on the application of deep learning in system health management,” *Mechanical Systems and Signal Processing*, vol. 107, pp. 241–265, 2018.
[12] W. Zhang, C. Li, G. Peng, Y. Chen, and Z. Zhang, “A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load,” *Mechanical systems and signal processing*, vol. 100, pp. 439–453, 2018.
[13] Z. An, S. Li, J. Wang, and X. Jiang, “A novel bearing intelligent fault diagnosis framework under time-varying working conditions using recurrent neural network,” *ISA Transactions*, vol. 100, pp. 155–170, 2020.
[14] J. Yang, G. Xie, and Y. Yang, “An improved ensemble fusion autoencoder model for fault diagnosis from imbalanced and incomplete data,” *Control Engineering Practice*, vol. 98, p. 104358, 2020.
[15] S. Zhang, S. Zhang, B. Wang, and T. G. Habetler, “Deep learning algorithms for bearing fault diagnostics—a comprehensive review,” *IEEE Access*, vol. 8, pp. 29 857–29 881, 2020.
[16] H. Shao, H. Jiang, H. Zhang, and T. Liang, “Electric locomotive bearing fault diagnosis using a novel convolutional deep belief network,” *IEEE Transactions on Industrial Electronics*, vol. 65, no. 3, pp. 2727–2736, 2017.
[17] C. Xu, D. Tao, and C. Xu, “A survey on multi-view learning,” *arXiv preprint arXiv:1304.5634*, 2013.
[18] G. Andrew, R. Arora, J. Bilmes, and K. Livescu, “Deep canonical correlation analysis,” in *International conference on machine learning*. PMLR, 2013, pp. 1247–1255.
[19] D. R. Hardoon, S. Szedmak, and J. Shawe-Taylor, “Canonical correlation analysis: An overview with application to learning methods,” *Neural computation*, vol. 16, no. 12, pp. 2639–2664, 2004.
[20] A. Benton, H. Khayrallah, B. Gujral, D. A. Reisinger, S. Zhang, and R. Arora, “Deep generalized canonical correlation analysis,” *arXiv preprint arXiv:1702.02519*, 2017.
[21] S. Shao, R. Yan, Y. Lu, P. Wang, and R. X. Gao, “DCNN-based multi-signal induction motor fault diagnosis,” *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 6, pp. 2658–2669, 2019.
[22] H. Hotelling, “Relations between two sets of variates,” in *Breakthroughs in statistics: methodology and distribution*. Springer, 1992, pp. 162–190.
[23] N. Martin and H. Maes, “Multivariate analysis,” *London, UK: Academic*, 1979.
[24] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
[25] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention is all you need,” *Advances in Neural Information Processing Systems*, 2017.
[26] K. Diederik, “Adam: A method for stochastic optimization,” *(No Title)*, 2014.
[27] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: a simple way to prevent neural networks from overfitting,” *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.
[28] L. Van der Maaten and G. Hinton, “Visualizing data using t-SNE.” *Journal of Machine Learning Research*, vol. 9, no. 11, 2008.
[29] M. Galar, A. Fernández, E. Barrenechea, H. Bustince, and F. Herrera, “An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes,” *Pattern Recognition*, vol. 44, no. 8, pp. 1761–1776, 2011.