Label-Consistent Convolutional Dictionary Learning for Machine Inspection

Saurabh Sahu[⋆], Kriti Kumar[⋆†], Angshul Majumdar[†], A Anil Kumar[⋆], M Girish Chandra[⋆]

[⋆] Embedded Devices and Intelligent Systems, TCS Research, Bangalore, India.
[†] Indraprastha Institute of Information Technology Delhi, New Delhi, India.
Email: {sahu.saurabh1, kriti.kumar, achannaanil.kumar, m.gchandra}@tcs.com, angshul@iiitd.ac.in

***Abstract*—Machine fault diagnosis is crucial for predicting and preventing unexpected failures. In most practical application scenarios of machine inspection, access to data is limited. Moreover, the labeled instances are few, which makes it challenging for the existing data-driven techniques to learn efficient representations. To address this issue, we propose a novel label-consistent convolutional dictionary learning for machine fault classification. This method employs a joint optimization formulation for learning the dictionary atoms, corresponding coefficients, and weights associated with a label-consistency term. The label-consistency term added in the joint formulation helps in learning class-discriminative dictionaries. The features generated from the class-discriminative dictionaries result in robust inferencing when used with an external classifier. The performance of the proposed method is evaluated using the publicly available CWRU dataset for bearing fault diagnosis. The results demonstrate the superior performance of the proposed method compared to other state-of-the-art deep learning and dictionary learning techniques for the data limited scenario. Furthermore, with only 10% of training data, the proposed method achieved an accuracy of around 85%, while the other methods provided less than 50%.**

***Index Terms*—Convolutional dictionary learning, Classification, Sparse representation, Supervised learning, Machine fault diagnosis**

I. Introduction

Maintaining the health of industrial equipment is crucial to avoid economic losses, the decline in production levels, and potential hazards to human life. Numerous industries are adopting more advanced digital technologies involving artificial intelligence for machine condition monitoring to enhance process efficiency and reduce downtime [1]–[3]. In the past, researchers have employed various feature-based (domain-crafted) [4] and data-driven approaches [5] to solve complex machine fault diagnosis problems. Although techniques involving domain-specific features are more intuitive and comprehensible, determining the relevant and discriminative features is a challenge and require domain expertise. Hence, data-driven techniques have recently gained significant attention due to the ability to learn meaningful representations from the data on their own [6].

Most data-driven techniques employ Deep Neural Networks (DNNs) [7] that learn useful patterns and relationships in data through a complex nonlinear mapping between the neural network layers. Thus, DNNs have been applied to a wide variety of problems in machine condition monitoring, including fault classification [8], [9], remaining useful life prediction [10], and anomaly detection [11]. Generally, DNNs require a lot of labeled data for training to perform well. However, data is limited in most practical application scenarios of machine inspection. Also, annotations are seldom available, making the classification problem more challenging. To address this, Dictionary Learning (DL) based techniques have been used that can learn efficient representations from limited data [12].

In DL, a sparse representation of the signal is obtained as a linear combination of a small number of basis vectors from the dictionary, and its associated sparse coefficients learned from the data. However, the computational cost of computing the sparse representation for the entire signal is high. Thus, patch-based processing is usually employed where the entire signal is divided into low-dimensional overlapping blocks (i.e., patches in image processing) [13]–[15]. However, the learned basis vectors of patch-based dictionary learning exhibit shift-variant behavior, i.e., basis vectors tend to contain shifted versions of each other. As a result, it fails to capture the underlying structure of the entire signal, since each block or patch is synthesized independently. Additionally, the learned dictionaries for an entire signal exhibit high redundancy due to the separate learning of neighboring and overlapping blocks. To address these limitations, Convolutional Dictionary Learning (CDL) has been proposed to learn shift-invariant dictionaries from the signals and has been successfully applied in various signal and image processing applications [16]–[18]. In our current context of machine health monitoring, the shift-invariant property of CDL is beneficial in extracting periodic impulses, that are typical signatures of a mechanical fault [19].

The existing methods based on CDL do not utilize label information while learning the dictionary; hence the representations learned are not class-discriminative. Including the label information can aid in learning class-discriminative dictionaries, thereby providing effective features for classification that are particularly useful for data limited scenario. To enable this, in this work, we propose a novel framework called Label-Consistent Convolutional Dictionary Learning (LC-CDL) that incorporates the label-consistency term into the CDL formulation for classification tasks. This approach employs a joint optimization formulation that learns the dictionary atoms, corresponding coefficients, and weights associated with a label-consistency term together, that facilitates the learning of class-discriminative convolutional dictionaries from the data. The learned dictionaries are utilized to generate class-discriminative features, which are then fed as input to an external classifier for robust inferencing. We have applied this method for bearing fault diagnosis as bearings are the most critical and vulnerable parts of industrial machines. Experimental results are presented using the publicly available Case Western Reserve University (CWRU) dataset for bearing fault diagnosis. The results demonstrate the enhancement in performance obtained by the inclusion of the label-consistency term in the CDL formulation. The proposed LC-CDL is shown to outperforms other state-of-the-art methods for limited data scenario.

To provide the necessary information of the proposed method, the rest of the paper is organised as follows. Section II provides a brief overview of dictionary learning techniques. Section III then discusses the proposed label-consistent CDL framework for classification. Section IV presents the results and discussion, and Section V concludes the paper.

II. Background on Dictionary Learning

A brief overview of Dictionary Learning (DL) and Convolution Dictionary Learning (CDL) is provided in this section.

A. Dictionary Learning

Given the data $\boldsymbol{s} \in \mathbb{R}^n$ of length n, a sparse representation of $\boldsymbol{s}$ can be obtained by a linear combination of few atoms selected from a learned Dictionary $\boldsymbol{D} \in \mathbb{R}^{n \times L}$, where L denotes the number of atoms. The basic DL formulation is expressed as:

$$\arg\min_{\boldsymbol{D},\boldsymbol{x}} \|\boldsymbol{D}\boldsymbol{x} - \boldsymbol{s}\|_2^2 + \lambda \|\boldsymbol{x}\|_1 \tag{1}$$

where $\boldsymbol{x} \in \mathbb{R}^L$ are the coefficients and λ associated with the l_1 norm controls the sparsity of the coefficients $\boldsymbol{x}$. Alternating minimization [20] is one of the popular method used for solving (1).

B. Convolutional Dictionary Learning

In the case of CDL, let $\{\boldsymbol{s}_k\}_{k=1}^{K}$ be the k^{th} input data sample of length n with K training samples. As shown in [17], a set of M distinct dictionary atoms $\{\boldsymbol{d}_m\}_{m=1}^{M}$ and the associated coefficient maps $\{\boldsymbol{x}_{m,k}\}_{m=1}^{M}$ of the same size as the data sample $\boldsymbol{s}_k$ are obtained by solving the following:

$$\arg\min_{\{\boldsymbol{d}_m\},\{\boldsymbol{x}_{m,k}\}} \frac{1}{2}\sum_{k=1}^{K}\left\|\sum_{m=1}^{M}\boldsymbol{d}_m * \boldsymbol{x}_{m,k} - \boldsymbol{s}_k\right\|_2^2 + \lambda \sum_{m,k} \|\boldsymbol{x}_{m,k}\|_1 \tag{2}$$
$$\text{s. t. } \|\boldsymbol{d}_m\|_2 = 1 \quad \forall\, m$$

where $*$ denotes convolution operation, and the additional constraint of $\|\boldsymbol{d}_m\|_2 = 1$ is applied to compensate the scaling ambiguity between the dictionary atoms $\boldsymbol{d}_m$ and the coefficient maps $\boldsymbol{x}_{m,k}$.

The convolution operation of $\boldsymbol{d}_m$ can be expressed as: $\boldsymbol{D}_m\boldsymbol{x}_{m,k} = \boldsymbol{d}_m * \boldsymbol{x}_{m,k}$, where $\boldsymbol{D}_m \in \mathbb{R}^{n \times n}$ denotes the convolution matrix. Taking $\boldsymbol{X}_m = [\boldsymbol{x}_{m,0}, ..., \boldsymbol{x}_{m,K}]$ and $\boldsymbol{S} = [\boldsymbol{s}_1, ..., \boldsymbol{s}_K]$, (2) can be re-written as:

$$\arg\min_{\{\boldsymbol{D}_m\},\{\boldsymbol{X}_m\}} \frac{1}{2}\left\|\sum_{m=1}\boldsymbol{D}_m\boldsymbol{X}_m - \boldsymbol{S}\right\|_F^2 + \lambda \sum_{m} \|\boldsymbol{X}_m\|_1 \tag{3}$$
$$\text{s. t. } \|\boldsymbol{D}_m\|_2 = 1 \quad \forall\, m$$

Since the problem in (3) is not jointly convex in both the variables $\{\boldsymbol{D}_m\}$ and $\{\boldsymbol{X}_m\}$, alternating minimization is employed to solve for each variable keeping the other variable fixed. The updates for $\{\boldsymbol{D}_m\}$ and $\{\boldsymbol{X}_m\}$ can be obtained using ADMM iterations as shown in [17]. Note as opposed to DL, that approximates the entire data sample by a linear combination of a few dictionary atoms, in CDL, the entire data sample is represented as a sum over a set of dictionary atom convolutions with associated coefficient maps.

III. Label-Consistent CDL

The proposed method utilizes CDL technique to capture the discriminative local patterns from the data samples for classification tasks. Fig. 1 presents the block diagram of the proposed LC-CDL method. Referring to Fig. 1, given the labelled data samples, class discriminative dictionaries are learnt for generating efficient features (block A and B). These features are later utilized to learn an external classifier (block C) for robust classification. The training and test phase of the proposed method are described in detail below.

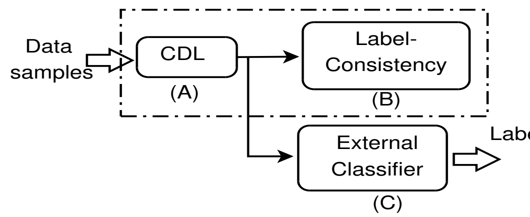


Fig. 1: Block Diagram of the Proposed Label-Consistent CDL

A. Training Phase

A joint optimization is carried out for learning the dictionary atoms $\{\boldsymbol{D}_m\}$, their corresponding coefficients $\{\boldsymbol{X}_m\}$ and weights $\{\boldsymbol{W}_m\}$ associated with the label-consistency term together for M filters. This unified framework enables discriminant dictionary atoms to be learnt from small amount of labelled training data. The joint optimization formulation is expressed as:

$$\begin{aligned}\arg\min_{\{\boldsymbol{D}_m\},\{\boldsymbol{X}_m\},\{\boldsymbol{W}_m\}} \frac{1}{2}\left\|\sum_{m=1}^{M}\boldsymbol{D}_m\boldsymbol{X}_m - \boldsymbol{S}\right\|_F^2 + \lambda\sum_{m=1}^{M}\|\boldsymbol{X}_m\|_1 \\ + \frac{\eta}{2}\left\|\boldsymbol{Q} - \sum_{m=1}^{M}\boldsymbol{W}_m\boldsymbol{X}_m\right\|_F^2\end{aligned} \tag{4}$$
$$\text{s. t. } \|\boldsymbol{D}_m\|_2 = 1 \quad \forall m$$

where, $\boldsymbol{Q} \in \mathbb{R}^{C \times K}$ represents one hot encoded labels for C-classes, such that $\boldsymbol{Q}_{ij} = 1$ if the data sample j belongs to class i, and 0 otherwise. The first two terms in (4) ensure that the dictionary atoms and associated sparse coefficients are learnt for data samples $\boldsymbol{S}$ such that the reconstruction error is low. The last term is the label-consistency term which is added here to assist in the learning of class-discriminative dictionaries such that the same class features map to the same class label. The hyperparameters λ and η control the sparsity of the learnt coefficients, and trade-off between the reconstruction and label-consistency terms, respectively.

We employ alternating minimization [20] to solve (4) to estimate $\{\boldsymbol{D}_m\}, \{\boldsymbol{X}_m\}$ and $\{\boldsymbol{W}_m\}$. The sub-problems to solve for the respective updates are given as:

$$\{\boldsymbol{D}_m\} \leftarrow \arg\min_{\{\boldsymbol{D}_m\}} \frac{1}{2}\left\|\sum_{m=1}^{M}\boldsymbol{D}_m\boldsymbol{X}_m - \boldsymbol{S}\right\|_F^2 \tag{5}$$
$$\text{s. t. } \|\boldsymbol{D}_m\|_2 = 1 \quad \forall m$$

$$\{\boldsymbol{X}_m\} \leftarrow \arg\min_{\{\boldsymbol{X}_m\}} \frac{1}{2}\left\|\sum_{m=1}^{M}\boldsymbol{D}_m\boldsymbol{X}_m - \boldsymbol{S}\right\|_F^2 + \lambda\sum_{m=1}^{M}\|\boldsymbol{X}_m\|_1 + \frac{\eta}{2}\left\|\boldsymbol{Q} - \sum_{m=1}^{M}\boldsymbol{W}_m\boldsymbol{X}_m\right\|_F^2 \quad (6)$$

$$\{\boldsymbol{W}_m\} \leftarrow \arg\min_{\{\boldsymbol{W}_m\}} \left\|\boldsymbol{Q} - \sum_{m=1}^{M}\boldsymbol{W}_m\boldsymbol{X}_m\right\|_F^2. \quad (7)$$

The update for the dictionary atoms $\{\boldsymbol{D}_m\}$ in (5) is similar to the work in [17]. The weights associated with the label-consistency term $\{\boldsymbol{W}_m\}$ are updated using simple least squares [21] and is obtained as:

$$\boldsymbol{W}_m = (\boldsymbol{Q} - \sum_{j=1, j\neq m}^{M}\boldsymbol{W}_j\boldsymbol{X}_j)\boldsymbol{X}_m^{\dagger} \quad (8)$$

where † denotes the pseudo-inverse. The update for the coefficients $\{\boldsymbol{X}_m\}$ are obtained by variable splitting, by introducing an auxiliary variable $\boldsymbol{Y}_m$. Using this, (6) is re-written as:

$$\arg\min_{\{\boldsymbol{X}_m\},\{\boldsymbol{Y}_m\}} \frac{1}{2}\left\|\sum_{m=1}^{M}\boldsymbol{D}_m\boldsymbol{X}_m - \boldsymbol{S}\right\|_F^2 + \lambda\sum_{m=1}^{M}\|\boldsymbol{Y}_m\|_1 + \frac{\eta}{2}\left\|\boldsymbol{Q} - \sum_{m=1}^{M}\boldsymbol{W}_m\boldsymbol{X}_m\right\|_F^2 \quad (9)$$

$$\text{s. t. } \boldsymbol{X}_m - \boldsymbol{Y}_m = 0.$$

We employ ADMM [22] to solve (9). The steps are given as:

$$\boldsymbol{X}_m = \arg\min_{\{\boldsymbol{X}_m\}} \frac{1}{2}\left\|\sum_{m=1}^{M}\boldsymbol{D}_m\boldsymbol{X}_m - \boldsymbol{S}\right\|_F^2 + \frac{\eta}{2}\left\|\boldsymbol{Q} - \sum_{m=1}^{M}\boldsymbol{W}_m\boldsymbol{X}_m\right\|_F^2 + \frac{\rho}{2}\sum_{m=1}^{M}\|\boldsymbol{X}_m - \boldsymbol{Y}_m + \boldsymbol{U}_m\|_F^2 \quad (10)$$

$$\boldsymbol{Y}_m = \arg\min_{\{\boldsymbol{Y}_m\}} \lambda\sum_{m=1}^{M}\|\boldsymbol{Y}_m\|_1 + \frac{\rho}{2}\sum_{m=1}^{M}\|\boldsymbol{X}_m - \boldsymbol{Y}_m + \boldsymbol{U}_m\|_F^2 \quad (11)$$

$$\boldsymbol{U}_m = \boldsymbol{U}_m + \boldsymbol{X}_m - \boldsymbol{Y}_m \quad (12)$$

where ρ controls the convergence rate of an algorithm. Ideally, one would like to start with a small value of ρ and increase it over iterations. However, as this is a Split Bregman [23] type approach, the equality constraint is imposed at convergence by $\boldsymbol{U}_m$. Therefore we can keep the value of ρ to be fixed. The closed form update for $\boldsymbol{X}_m$, is obtained by expanding the (10) in terms of trace and equating the derivative with respect to $\boldsymbol{X}_m$ to 0. This results in the following update for $\boldsymbol{X}_m$:

$$\boldsymbol{X}_m = (\boldsymbol{D}_m^T\boldsymbol{D}_m + \eta\boldsymbol{W}_m^T\boldsymbol{W}_m + \rho)^{-1}(\boldsymbol{D}_m^T\boldsymbol{S} + \eta\boldsymbol{W}_m^T\boldsymbol{Q} - \boldsymbol{D}_m^T\sum_{j=1,j\neq m}^{M}\boldsymbol{D}_j\boldsymbol{X}_j - \eta\boldsymbol{W}_m^T\sum_{j=1,j\neq m}^{M}\boldsymbol{W}_j\boldsymbol{X}_j + \rho(\boldsymbol{Y}_m - \boldsymbol{U}_m)) \quad (13)$$

Subsequently, the update for $\boldsymbol{Y}_m$ is obtained using Soft Thresholding Algorithm [24] as:

$$\boldsymbol{Y}_m = \mathcal{S}_{\lambda/\rho}(\boldsymbol{X}_m + \boldsymbol{U}_m). \quad (14)$$

where, $\mathcal{S}_\gamma(V) = \text{sign}(V) \odot \max(0, |V| - \gamma)$ and $\odot$ denotes element-wise multiplication. The dual variable $\boldsymbol{U}_m$ in (12) corresponds to the constraint $\boldsymbol{X}_m - \boldsymbol{Y}_m = 0$ in (9), and is updated by simple arithmetic operations.

Note that $\{\boldsymbol{D}_m\}, \{\boldsymbol{X}_m\}$ and $\{\boldsymbol{W}_m\}$ are updated iteratively until the objective function in (4) convergences. This marks the end of the training phase where class-discriminative dictionaries are learnt. These dictionaries are utilized to generate coefficients that form class-discriminative features for classification. Any suitable classifier can be learnt using these features for robust inferencing.

B. Test phase

Given the test data samples $\boldsymbol{S}^{test}$, the corresponding coefficient $\{\boldsymbol{X}_m^{test}\}$ are estimated using the dictionary atoms $\{\boldsymbol{D}_m\}$ learnt in the training phase. The solution for $\{\boldsymbol{X}_m^{test}\}$ can be obtained by solving [17]:

$$\{\boldsymbol{X}_m^{test}\} \leftarrow \arg\min_{\{\boldsymbol{X}_m^{test}\}} \frac{1}{2}\left\|\sum_{m=1}^{M}\boldsymbol{D}_m\boldsymbol{X}_m^{test} - \boldsymbol{S}^{test}\right\|_F^2 + \lambda\sum_{m=1}^{M}\left\|\boldsymbol{X}_m^{test}\right\|_1. \quad (15)$$

We employ ADMM technique for updating (15) which is similar to the approach described in [17]. The computed coefficient $\{\boldsymbol{X}_m^{test}\}$ act as features that are fed to the trained external classifier for estimating the class labels. In this work, we have employed Support Vector Machine (SVM) for classifying machine faults.

IV. Results and Discussion

This section briefly describes the dataset and the baseline methods used for the performance evaluation of the proposed LC-CDL method. Subsequently, the results are discussed in detail, and an ablation study is included to provide additional insights into the proposed method.

A. Dataset Description

We have considered the publicly available CWRU [25] dataset for bearing fault classification. It contains vibration data collected at 12kHz for both healthy and faulty bearings acquired at the drive end and fan end of the motor. The experimental setup comprises of a 2 hp motor, a torque encoder, a dynamometer, and control electronics for data collection. Three different faults, namely (i) Ball fault, (ii) Inner Race fault, and (iii) Outer Race fault of different diameters (7mils, 14mils, and 21mils), are investigated. Data is collected for four distinct operating conditions of 0hp, 1hp, 2hp, and 3hp, with the motor speed varying from 1797 to 1720 rpm. For this work, we have considered the faulty bearing data corresponding to 7 mils diameter, obtained from the drive end of the motor, and data from all operating conditions (0-3hp) is combined for performance evaluation.

B. Baseline Methods

The proposed method is compared against two deep learning methods namely, standard 1-dimensional CNN (1D-CNN) and the recent Semi-supervised Time series classification (SemiTime) method [26]. Kindly note that the standard 1D-CNN method is implemented with the same number of layers and filters as that of the proposed LC-CDL for a fair comparison. It employs a single 1D-convolution layer with 2 filters followed by a fully connected layer with a ReLU activation function. While SemiTime employs 4-layer 1D-CNN with batch normalization as discussed in their work. It can work both in a supervised and semi-supervised setting. Here, the supervised setting of SemiTime is used for comparison. Additionally, we provide results for Label Consistent Dictionary Learning (LC-DL) [15] to demonstrate the potential of CDL over standard DL methods. Similar to the proposed method, LC-DL employs the label-consistency term but with a standard DL formulation to learn the class-discriminative dictionary and the coefficients. The learned dictionary is used to compute the coefficients which are fed to the external classifier.

C. Experimental Results

The raw data is normalized and split into windows of $n = 1024$ samples with a 50% overlap, resulting in 3744 samples. The performance of all the methods is evaluated using the accuracy metric with training sets of 10%, 20%, 30%, and 50%. For each training set, 10% of the data is used for validation, and the remaining is used for testing. In both LC-DL and the proposed LC-CDL, we use the SVM with a degree 2 polynomial kernel as the external classifier. However, in general, any suitable classifier can be used. The hyperparameter tuning for the proposed LC-CDL method is carried out using grid search, and the optimal values obtained are $\eta = 1$, $\lambda = 0.2$, $\rho = 200$, and $M = 2$. The results obtained

TABLE I: Classification Accuracy with Different Methods

Methods	Training data			
	10%	20%	30%	50%
1D-CNN with ReLU activation	0.49	0.64	0.81	0.90
SemiTime [26]	0.40	0.73	0.82	0.95
LC-DL [15]	0.43	0.57	0.68	0.88
LC-CDL	**0.85**	**0.90**	**0.92**	**0.97**

with different methods, averaged over 10 independent runs is presented in Table I, with the best-performing technique highlighted in bold. It can be observed that the proposed LC-CDL achieves better results compared to other benchmarks across all training sets. Among the deep learning based methods, it can be observed that when the training data $\geq 10\%$, SemiTime performs better than 1D-CNN. It is worth noting that with 50% training data, SemiTime performs similarly to LC-CDL. However, when the training data reduces (10% or 20%), all the deep learning based methods fail to learn efficient representations, resulting in poor accuracy. The proposed LC-CDL demonstrates superior performance compared to all other methods, including the LC-DL method. This can be attributed to the fact that the shift-invariant basis of LC-CDL efficiently captures the transient events of vibration signals, resulting in learning good representations. The shift-invariant property of CDL combined with label-consistency helps in learning class-discriminative features and hence, a single layer configuration of CDL with significantly less training data outperforms state-of-the-art methods. The convergence plot of the objective function in (4) with 20% training data is shown in Fig. 2. It shows that the proposed LC-CDL method converges quickly, within a few iterations.

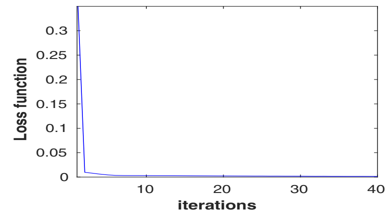
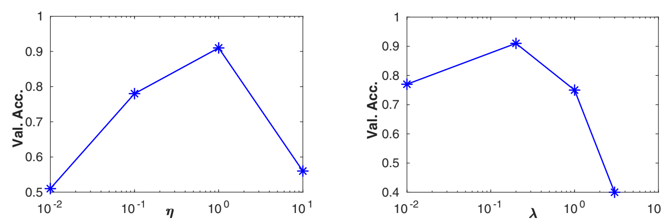


Fig. 2: Convergence of objective function

D. Ablation Study

Here we provide an ablation study to understand the impact of different hyperparameters in (4) on the performance of the method. Also, we present a comparative analysis of the proposed LC-CDL method with different configurations of CDL for additional insights.

1) Selection of hyperparameters: There are two hyperparameters associated with the LC-CDL formulation, namely λ that denotes the sparsity of the coefficients, and η that denotes the trade-off between the reconstruction and label-consistency term. To determine the optimal values of these hyperparameters, we varied them one at a time while keeping the other fixed and measured the resulting impact on the validation accuracy. Fig. 3a and 3b present the validation accuracy with 20% of the training data for different values of λ and η, respectively. For the case of η, it was observed



(a) Validation Accuracy vs η (b) Validation Accuracy vs λ

Fig. 3: Impact of η and λ on validation accuracy (20% training data)

that $\eta = 1.0$ gave the best accuracy on the validation set. This value provides equal importance to the reconstruction and label-consistency term. For the case of λ, the validation accuracy improved progressively up to $\lambda = 0.2$, beyond that a drop in performance was observed. These optimal value of η and λ, were employed to generate the results for the proposed LC-CDL method presented in Table I.

2) Configurations of CDL: The LC-CDL method proposed here employs an external non-linear classifier for classification. Two different configurations of CDL are considered for comparison with the proposed method: (i) LC-CDL without external classifier, and (ii) CDL without label consistency. The first configuration incorporates only blocks (A) and (B) of LC-CDL from Fig. 1, and do not use an external classifier (block C). Here, the weights of the label-consistency term acts as a linear classifier and used for classification. The second configuration employs blocks (A) and (C) of LC-CDL from Fig. 1, where the CDL learns the features without the label consistency term (blocks B). An external non-linear classifier is used on the learned features. Table II presents the experimental results obtained by the different configurations and the proposed LC-CDL method. Additionally, comparisons with 1D-CNN method with linear activation are also presented that is similar in nature to the LC-CDL without external classifier configuration. It can be observed that the proposed LC-CDL method (with external classifier) outperforms other methods. Although the first configuration yields marginally better results than the 1D-CNN, it is not able to capture the non-linearities in the data and hence does not perform well. The second configuration performs better than the former methods as it uses a non-linear classifier. However, the representations or features learnt by them are not class-discriminative as they do not utilize the label information. The joint learning of the CDL and the label consistency term in the proposed method facilitates the acquisition of class-discriminative features, that leads to superior performance with the help of an external non-linear classifier.

TABLE II: Different Configurations of the CDL Method

Methods	**Training data**	
	10%	**20%**
1D-CNN with linear activation	0.34	0.41
LC-CDL without external classifier	0.41	0.42
CDL without label consistency	0.65	0.76
LC-CDL with external classifier (Proposed Method)	**0.85**	**0.90**

V. Conclusion

This paper presents a novel label-consistent convolutional dictionary learning for classification tasks. The joint learning of the convolution dictionary with label consistency term allows class-discriminative dictionary to be learned from raw signals. The external classifier utilizes the features generated from class-discriminative dictionary to provide more reliable results. Experimental results on CWRU data demonstrate the superiority of the proposed method over state-of-the-art techniques for data limited scenarios; they also indicate the significance of the label-consistency term in the joint formulation. The results showed that the proposed LC-CDL method performed consistently well even with 10% training data, which was challenging for all other baseline methods.

It is important to note that the proposed method is applied for machine fault classification in this work, but the method is generic and can be used in other applications for classification tasks. In the future, a deep version of this method shall be explored for classification and regression tasks.

References

[1] Jay Lee, Edzel Lapira, Behrad Bagheri, and Hung-an Kao, "Recent advances and trends in predictive manufacturing systems in big data environment," *Manufacturing letters*, vol. 1, no. 1, pp. 38–41, 2013.

[2] Mou Ling Dennis Wong, M Zhang, and Asoke K Nandi, "Effects of compressed sensing on classification of bearing faults with entropic features," in *2015 23rd European Signal Processing Conference (EUSIPCO)*. IEEE, 2015, pp. 2256–2260.

[3] Mariela Cerrada, René-Vinicio Sánchez, Chuan Li, Fannia Pacheco, Diego Cabrera, José Valente de Oliveira, and Rafael E Vásquez, "A review on data-driven fault severity assessment in rolling bearings," *Mechanical Systems and Signal Processing*, vol. 99, pp. 169–196, 2018.

[4] René-Vinicio Sánchez, Pablo Lucero, Jean-Carlo Macancela, Mariela Cerrada, Rafael E Vásquez, and Fannia Pacheco, "Multi-fault diagnosis of rotating machinery by using feature ranking methods and svm-based classifiers," in *2017 International conference on sensing, diagnostics, prognostics, and control (SDPC)*. IEEE, 2017, pp. 105–110.

[5] Shen Yin, Xianwei Li, Huijun Gao, and Okyay Kaynak, "Data-based techniques focused on modern industry: An overview," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 1, pp. 657–667, 2014.

[6] Wei Zhang, Chuanhao Li, Gaoliang Peng, Yuanhang Chen, and Zhujun 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.

[7] Serkan Kiranyaz, Onur Avci, Osama Abdeljaber, Turker Ince, Moncef Gabbouj, and Daniel J Inman, "1d convolutional neural networks and applications: A survey," *Mechanical systems and signal processing*, vol. 151, pp. 107398, 2021.

[8] Wenjun Sun, Siyu Shao, Rui Zhao, Ruqiang Yan, Xingwu Zhang, and Xuefeng Chen, "A sparse auto-encoder-based deep neural network approach for induction motor faults classification," *Measurement*, vol. 89, pp. 171–178, 2016.

[9] Long Wen, Xinyu Li, Liang Gao, and Yuyan Zhang, "A new convolutional neural network-based data-driven fault diagnosis method," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 7, pp. 5990–5998, 2017.

[10] Min Xia, Teng Li, Tongxin Shu, Jiafu Wan, Clarence W De Silva, and Zhongren Wang, "A two-stage approach for the remaining useful life prediction of bearings using deep neural networks," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 6, pp. 3703–3711, 2018.

[11] Zhe Li, Jingyue Li, Yi Wang, and Kesheng Wang, "A deep learning approach for anomaly detection based on sae and lstm in mechanical equipment," *The International Journal of Advanced Manufacturing Technology*, vol. 103, pp. 499–510, 2019.

[12] Hao Tang, Hong Liu, Wei Xiao, and Nicu Sebe, "When dictionary learning meets deep learning: Deep dictionary learning and coding network for image recognition with limited data," *IEEE transactions on neural networks and learning systems*, vol. 32, no. 5, pp. 2129–2141, 2020.

[13] Alfred M Bruckstein, David L Donoho, and Michael Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM review*, vol. 51, no. 1, pp. 34–81, 2009.

[14] Yangyang Xu and Wotao Yin, "A fast patch-dictionary method for whole image recovery," *arXiv preprint arXiv:1408.3740*, 2014.

[15] Zhuolin Jiang, Zhe Lin, and Larry S Davis, "Label consistent k-svd: Learning a discriminative dictionary for recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 11, pp. 2651–2664, 2013.

[16] Vardan Papyan, Yaniv Romano, Jeremias Sulam, and Michael Elad, "Convolutional dictionary learning via local processing," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 5296–5304.

[17] Cristina Garcia-Cardona and Brendt Wohlberg, "Convolutional dictionary learning: A comparative review and new algorithms," *IEEE Transactions on Computational Imaging*, vol. 4, no. 3, pp. 366–381, 2018.

[18] Arthur Szlam, Koray Kavukcuoglu, and Yann LeCun, "Convolutional matching pursuit and dictionary training," *arXiv preprint arXiv:1010.0422*, 2010.

[19] Haitao Zhou, Jin Chen, Guangming Dong, and Ran Wang, "Detection and diagnosis of bearing faults using shift-invariant dictionary learning and hidden markov model," *Mechanical systems and signal processing*, vol. 72, pp. 65–79, 2016.

[20] David R Hunter, "Alternating minimization algorithms," *Wiley StatsRef: Statistics Reference Online*, pp. 1–10, 2014.

[21] Christopher M Bishop and Nasser M Nasrabadi, *Pattern recognition and machine learning*, vol. 4, Springer, 2006.

[22] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan Eckstein, et al., "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine learning*, vol. 3, no. 1, pp. 1–122, 2011.

[23] Tom Goldstein and Stanley Osher, "The split bregman method for l1-regularized problems," *SIAM journal on imaging sciences*, vol. 2, no. 2, pp. 323–343, 2009.

[24] Neal Parikh, Stephen Boyd, et al., "Proximal algorithms," *Foundations and trends® in Optimization*, vol. 1, no. 3, pp. 127–239, 2014.

[25] Zhenxiang Li, "Cwru bearing dataset and gearbox dataset of ieee phm challenge competition in 2009," 2019.

[26] Haoyi Fan, Fengbin Zhang, Ruidong Wang, Xunhua Huang, and Zuoyong Li, "Semi-supervised time series classification by temporal relation prediction," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 3545–3549.