ENFES: ENsemble FEw-Shot Learning For Intelligent Fault Diagnosis with Limited Data

Onat Gungor[1,3], Tajana Rosing[2], and Baris Aksanli[3]

[1]Department of Electrical and Computer Engineering, University of California San Diego
[2]Department of Computer Science and Engineering, University of California San Diego
[3]Department of Electrical and Computer Engineering, San Diego State University
ogungor@ucsd.edu, tajana@ucsd.edu, baksanli@sdsu.edu

***Abstract*—Fault diagnosis is a key component of predictive system maintenance. Big data collected from sensors helps create data-driven fault diagnosis methods. However, it may be extremely costly to label specific fault types in a collected dataset. Hence, prediction algorithms should perform well under limited supervision. Few-shot learning (FSL) can provide a great prediction performance using very limited labeled data by discovering similarity among input pairs. But selection of a single FSL method may be arduous due to changing working conditions. Ensemble FSL solves this problem by combining a variety of FSL methods systematically. We propose an ensemble FSL framework, *ENFES*, where we combine 5 different Siamese neural network architectures using an iterative majority voting classifier. Our transfer learning-oriented experiments show that *ENFES* can improve the best algorithm significantly while using very limited labeled data. We obtain up to 16.4% improvement over the best algorithm by only using 0.3% of the training data.**

***Index Terms*—fault diagnosis, predictive maintenance, few-shot learning, siamese neural network, ensemble learning**

I. Introduction and Related Work

Data-driven predictive maintenance uses sensor data to build machine learning models to find an optimal time to schedule maintenance [1]. Sensor data collection and processing is key to achieving good performance. Fault diagnosis helps establishing more efficient predictive maintenance systems by finding and classifying different fault types [2]. Due to advancements in machine learning, numerous intelligent fault diagnosis methods have been developed [3]–[5]. These methods require huge amounts of labeled training data to work well. However, it might not be feasible to obtain sufficient training samples for multiple fault types under all working conditions due to: 1) slowly occurring failure processes, 2) frequently changing working conditions, and 3) that some critical industrial systems are not allowed to run into faulty states [6]. Thus, intelligent fault diagnosis methods should work with limited data and changing operating conditions.

Few-shot learning (FSL) performs well with limited labeled data [7]–[9]. FSL-based fault diagnosis learns classifiers given only a few labeled examples of each fault type where the goal is to find similarity among input pairs. There are a variety of FSL approaches such as Siamese neural networks [10], matching networks [11], model-agnostic meta learning [12], and memory augmented neural network [13]; and FSL-based fault diagnosis works [6], [14]–[17]. These studies focus on using a single method for fault diagnosis. The performance of an algorithm can change with respect to dataset, working conditions, etc. [18]. Previous studies found that the best performing algorithm for a problem changes with respect to dataset [19], [20]. Thus, we need a systematic methodology to combine different methods. To achieve this, we utilize ensemble learning which is previously used for different predictive maintenance domains such as remaining useful life prediction [20], [21] and fault diagnosis [22], [23]. We also utilize transfer learning to adapt to changing working conditions. Transfer learning can enable transferring an already-learned model in an existing operating condition to a new condition that was not observed before [24].

This paper proposes an ensemble few-shot learning framework (*ENFES*) for intelligent fault diagnosis. In contrast to the state-of-the-art, we use multiple FSL methods and combine them via our majority voting framework to improve the prediction performance. We focus on Siamese neural network owing to its high accuracy in fault diagnosis [6]. Given pairs of input data, we first train five different Siamese neural networks. We combine the individual predictions using our iterative majority voting ensemble learner where the fault type voted the most by the classifiers is assigned as the final output. We evaluate our approach using Case Western Reserve University (CWRU) Bearing Datasets [25], a widely used benchmark for fault diagnosis. Since the working conditions can change frequently, we consider different transfer learning (TL) scenarios, where we train the model in one setting and test it in another. We observe that the best algorithm changes based on the underlying transfer learning scenario indicating that one algorithm would not work the best in all cases. *ENFES* can bring a significant performance improvement over the best algorithm with very limited training data, up to 16.4% and 10.7% improvement over the best method with only 60 and 90 samples (out of 19,800 samples), respectively.

II. Proposed Framework

Figure 1 depicts our proposed framework for intelligent fault diagnosis. Given pairs of input sensor data, we first train different Siamese neural networks. We collect the predictions of individual methods and then combine them using our

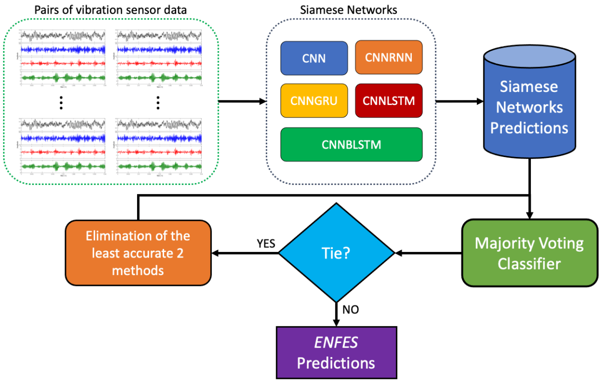


Fig. 1: Our Proposed Framework

majority voting classifier approach where the most voted fault type by the classifiers becomes the *ENFES* prediction.

A. Siamese Networks

Although there are several sensors collecting data, it may not be feasible to label every failure point. Few-shot learning (FSL) aims to solve this problem by using limited labeled data. The goal of training an FSL model is to discover similarity between inputs coming from different fault types. To train an FSL model, a support set is given containing a small set of labeled fault types. Given pairs of time series sensor data, training phase outputs the probability P that two input samples are the same. For testing, there are two different approaches: one-shot k-way testing and n-shot k-way testing. In one-shot k-way testing support set Ψ contains k samples $x_1, \ldots, x_k$ with distinct labels $y_1, \ldots, y_k$: $\Psi = \{(x_1, y_1), ..., (x_k, y_k)\}$. Given test sample $\tilde{x}$, the goal is to output the class $\tilde{c}$ that maximizes the similarity probability:

$$\tilde{c} = \underset{c}{\operatorname{argmax}}(P(\tilde{x}, x_c)), \ \ x_c \in \Psi \tag{1}$$

In n-shot k-way testing, support set has n samples from k classes $\Psi_1, \ldots, \Psi_n$. Given test sample $\tilde{x}$, we obtain class $\tilde{c}$ that maximizes the similarity probability over n samples:

$$\tilde{c} = \underset{c}{\operatorname{argmax}}(\sum_{j=1}^{n} P(\tilde{x}, x_{cj})), \ \ x_{cj} \in \Psi_j \tag{2}$$

We focus on Siamese Neural Network (SNN) due to its high accuracy in fault diagnosis [6]. SNNs contain two or more identical sub-networks (same network architecture and shared weights). Outputs of these networks are mapped to a high-dimensional feature vector to calculate the distance between the inputs. By connecting feature vector to a distance function and a fully connected neural network, we find the similarity probability between input pairs. Higher similarity probability indicates that two inputs are likely in the same category.

There are many alternatives for the sub-network selection, such as wide deep convolutional neural network (WDCNN) [26] and long short-term memory (LSTM) [27]. However, a single prediction method may not perform the best across different systems [18]. Ensemble learning solves this issue where predictions from multiple models are strategically combined. We create a CNN-based ensemble learning framework by considering 5 different architectures: convolutional neural network (*CNN*), CNN recurrent neural network (*CNNRNN*), CNN gated recurrent unit (*CNNGRU*), CNN long short-term memory (*CNNLSTM*), and CNN bi-directional long short-term memory (*CNNBLSTM*). We use the same CNN architecture across all methods as in [6], [26]. For the hybrid models (where 2 models are combined, e.g. *CNNRNN*), we connect CNN with the selected network structure consecutively, e.g., for CNNRNN, we connect CNN with 2 RNN layers with 32 and 16 nodes. We replace RNN layers with GRU, LSTM, and BLSTM for the *CNNGRU*, *CNNLSTM*, and *CNNBLSTM*, respectively. These SNNs provide single model fault type predictions for our test data, shown in Table I, where each model outputs its fault prediction for a given bearing. Then, we combine these predictions using our majority voting classifier.

TABLE I: Single Method Fault Type Predictions

Bearing ID	*CNN*	*CNNRNN*	*CNNGRU*	*CNNLSTM*	*CNNBLSTM*
1	Ball (0.014 in)	Ball (0.007 in)	Ball (0.014 in)	Ball (0.014 in)	Ball (0.021 in)
2	Outer Race (0.021 in)	Outer Race (0.014 in)	Outer Race (0.021 in)	Outer Race (0.014 in)	Outer Race (0.021 in)
...	...	...	...	...	...
250	Inner Race (0.007 in)	Inner Race (0.014 in)	Inner Race (0.007 in)	Inner Race (0.021 in)	Inner Race (0.007 in)

B. Majority Voting Classifier

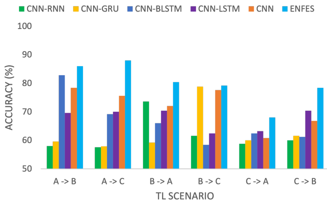
Majority voting classifier (MVC) is an ensemble learner that combines the class predictions of different methods. Assume that we have n different classifiers $\xi_1, \xi_2 \ldots \xi_n$ that map input data X to class $c_1, c_2 \ldots c_n$. MVC Ξ finds the class $\bar{c}$ that maximizes the weighted sum of correct class predictions [28]:

$$\Xi(X) = \bar{c} = \underset{c}{\operatorname{argmax}} \sum_{j=1}^{n} \omega_j I(\xi_j(X) = c) \tag{3}$$

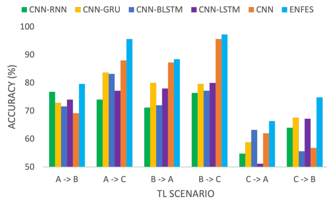
where $\omega_1, \omega_2 \ldots \omega_n$ are the classifier weights summing up to 1. I is the indicator function which is 1 if the classifier prediction j is class c, and 0 otherwise. If we set all weights equal to each other (i.e. $\omega_j = \frac{1}{n}$), this formulation becomes a *mode* function which outputs the class most voted by the classifiers. The main problem occurs when there is a tie among n classifiers. To break a tie, we propose a method that iteratively eliminates the least accurate $n-1$ methods based on the validation set accuracy. We continue classifier elimination until there is one method left to guarantee tie-breaking. We specifically start with an odd number of methods (5) to reduce possible tie scenarios among classifiers.

III. Experimental Analysis

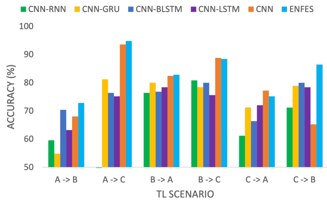
Dataset Description: We use Case Western Reserve University (CWRU) Bearing Datasets [25], a widely used benchmark for fault diagnosis. Rolling element bearing failure is one of the most frequent reasons for machine breakdown [29]. This data was collected at 12k samples/second and at 48k samples/second for drive end bearing experiments. We use the former data set that contains 19,800 training and



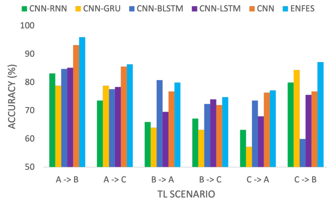
(a) **60 Training Samples**



(b) **90 Training Samples**



(c) **120 Training Samples**



(d) **150 Training Samples**

Fig. 2: Transfer Learning Methods Comparison

750 test samples. Bearing used in this experiment has three components: rolling element, inner race, and outer race. 9 different fault types are provided in the dataset based on the fault diameter (0.007, 0.014, and 0.021 inches) and the component (plus the normal bearing condition). Besides, three datasets (A, B, and C) are given representing different working conditions (based on motor speeds). We use transfer learning (TL) to transfer models across possible working conditions. For instance, transfer from dataset A to B trains model using dataset A, and tests this model on dataset B.

Experimental Setup: We construct the same setup in [6] as it led to great fault diagnosis performance: Sliding window of size 2048 points sliding with 80 points shift step, L1 distance to calculate distance between feature vectors, regularized cross-entropy loss function, Adam optimizer, batch size of 32, 15000 as number of epochs, and five-shot learning scenario. We repeat each experiment 10 times and report average values. All experiments are run on a PC with 16 GB RAM and an 8-core 2.3 GHz Intel Core i9 processor.

Results: We analyze the performance of our method under a transfer learning (TL) setting to represent changing working conditions. We have three different datasets corresponding to three working conditions: A, B, and C. We experiment with pairwise transfer learning scenarios: transfer model from A to B, A to C, B to A, B to C, C to A, and C to B. Since our focus is on fault type prediction using limited amount of data, we only use 60, 90, 120, and 150 samples from the training data (out of 19,800 samples) while using the entire test data. Fig. 2 presents the model performance comparison for different number of training samples and TL scenarios. In each sub-figure, x-axis represents the transfer learning scenario (e.g. $A \mapsto B$ denotes transfer from dataset A to B), and y-axis is the accuracy of the methods. Each color denotes a different method where *ENFES* is shown with light blue color. The performance of a single method changes with respect to TL scenario. For example, using 60 training samples, *CNNBLSTM* is the best method for $A \mapsto B$ while

TABLE II: *ENFES* Improvement Over Best Algorithm (%)

	Number of Training Samples			
TL Scenario	**60**	**90**	**120**	**150**
A->B	3.86	3.65	3.41	3
A->C	16.4	8.64	1.28	0.93
B->A	9.24	1.38	0.5	-0.99
B->C	0.51	1.67	-0.45	1.08
C->A	7.59	5.06	-2.59	1.05
C->B	11.38	10.65	8	3.33
Average	**8.17**	**5.18**	**1.69**	**1.4**

CNNRNN being best for the opposite transfer configuration (i.e. $B \mapsto A$). *ENFES* utilizes from all 5 methods to improve the prediction performance. Under 60 and 90 samples, *ENFES* outperforms other methods consistently. Table II presents our proposed approach improvement over the best method for all TL scenarios. At 60 and 90 training samples, *ENFES* improves the best method by up to 16.4% and 10.7% respectively (8.2% and 5.2% average). As we have more training samples, the improvement over the best algorithm decreases, yet we still obtain improvement up to 8% and 3.3% under 120 and 150 training samples (1.7% and 1.4% average).

The reason behind decreasing *ENFES* improvement with increasing data is due to the performance of the best algorithm. For instance, at 60 samples $C \mapsto B$ scenario, the best algorithm *CNNLSTM* provides 70% accuracy. As we reach 150 samples, the best algorithm *CNNGRU* can reach 85% accuracy. In this case, the room for improvement is limited by this method's accuracy. Hence, improvement for *ENFES* decreases from 11.4% to 3.3%. The performance of a single algorithm changes with respect to TL scenario, e.g. *CNNGRU* is the worst algorithm at $C \mapsto A$ scenario under 150 samples. Our method *ENFES* provides consistent accurate predictions at all instances (it is either the best or the second best approach). *ENFES* provides an average improvement over the best method if we use less than 1% of the entire training data. This result aligns with our claim where *ENFES* has highly efficient prediction using limited data.

IV. Conclusion

Predictive maintenance determines when maintenance actions are necessary based on continuous monitoring of machinery [30]. Fault diagnosis determines which fault occurred in a production system [2]. By using sensor data, intelligent fault diagnosis methods can be constructed, yet these methods require huge amount of labeled data to perform well. Few-shot learning (FSL) eliminates this restriction by discovering similarity among input pairs. Nevertheless, selection of a single FSL method may not provide optimal predictions. Different methods can be combined by using an ensemble learner. In this paper, we proposed ensemble few-shot learning approach for fault diagnosis. We specifically consider 5 different Siamese neural network structure and combine their fault type predictions via our majority voting classifier. We show that our approach can improve the classifier accuracy significantly under different transfer learning scenarios.

References

[1] R. K. Mobley, *An introduction to predictive maintenance*. Elsevier, 2002.

[2] M. Ardakani, M. Askarian, A. Shokry, G. Escudero, M. Graells, and A. Espuña, "Optimal features selection for designing a fault diagnosis system," in *Computer Aided Chemical Engineering*, vol. 38, pp. 1111–1116, Elsevier, 2016.

[3] Z. Chen, A. Mauricio, W. Li, and K. Gryllias, "A deep learning method for bearing fault diagnosis based on cyclic spectral coherence and convolutional neural networks," *Mechanical Systems and Signal Processing*, vol. 140, p. 106683, 2020.

[4] B. Zhang, Y. Miao, J. Lin, and Y. Yi, "Adaptive maximum second-order cyclostationarity blind deconvolution and its application for locomotive bearing fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 158, p. 107736, 2021.

[5] X. Wang, D. Mao, and X. Li, "Bearing fault diagnosis based on vibro-acoustic data fusion and 1d-cnn network," *Measurement*, vol. 173, p. 108518, 2021.

[6] 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.

[7] M. Fink, "Object classification from a single example utilizing class relevance metrics," *Advances in neural information processing systems*, vol. 17, pp. 449–456, 2005.

[8] L. Fei-Fei, R. Fergus, and P. Perona, "One-shot learning of object categories," *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, no. 4, pp. 594–611, 2006.

[9] D. Wu, F. Zhu, and L. Shao, "One shot learning gesture recognition from rgbd images," in *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pp. 7–12, IEEE, 2012.

[10] G. Koch, R. Zemel, and R. Salakhutdinov, "Siamese neural networks for one-shot image recognition," in *ICML deep learning workshop*, vol. 2, Lille, 2015.

[11] O. Vinyals, C. Blundell, T. Lillicrap, K. Kavukcuoglu, and D. Wierstra, "Matching networks for one shot learning," *arXiv preprint arXiv:1606.04080*, 2016.

[12] C. Finn and S. Levine, "Meta-learning and universality: Deep representations and gradient descent can approximate any learning algorithm," *arXiv preprint arXiv:1710.11622*, 2017.

[13] A. Santoro, S. Bartunov, M. Botvinick, D. Wierstra, and T. Lillicrap, "One-shot learning with memory-augmented neural networks," *arXiv preprint arXiv:1605.06065*, 2016.

[14] Y. Yang, H. Wang, Z. Liu, and Z. Yang, "Few-shot learning for rolling bearing fault diagnosis via siamese two-dimensional convolutional neural network," in *PHM-2020 Jinan*, pp. 373–378, IEEE, 2020.

[15] Q. Fang and D. Wu, "Ans-net: anti-noise siamese network for bearing fault diagnosis with a few data," *Nonlinear Dynamics*, pp. 1–18, 2021.

[16] S. Dixit and N. K. Verma, "Intelligent condition-based monitoring of rotary machines with few samples," *IEEE Sensors Journal*, vol. 20, no. 23, pp. 14337–14346, 2020.

[17] J. Wu, Z. Zhao, C. Sun, R. Yan, and X. Chen, "Few-shot transfer learning for intelligent fault diagnosis of machine," *Measurement*, vol. 166, p. 108202, 2020.

[18] O. Güngör, B. Akşanlı, and R. Aydoğan, "Algorithm selection and combining multiple learners for residential energy prediction," *Future Generation Computer Systems*, vol. 99, pp. 391–400, 2019.

[19] O. Gungor, J. Garnier, T. S. Rosing, and B. Aksanli, "Lenard: Lightweight ensemble learner for medium-term electricity consumption prediction," in *2020 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, pp. 1–6, IEEE, 2020.

[20] O. Gungor, T. S. Rosing, and B. Aksanli, "Opelrul: Optimally weighted ensemble learner for remaining useful life prediction,"

[21] Z. Li, K. Goebel, and D. Wu, "Degradation modeling and remaining useful life prediction of aircraft engines using ensemble learning," *Journal of Eng. for Gas Turbines and Power*, vol. 141, no. 4, 2019.

[22] C. Zhang, Y. He, B. Du, L. Yuan, B. Li, and S. Jiang, "Transformer fault diagnosis method using iot based monitoring system and ensemble machine learning," *Future Generation Computer Systems*, vol. 108, pp. 533–545, 2020.

[23] X. Li, H. Jiang, R. Wang, and M. Niu, "Rolling bearing fault diagnosis using optimal ensemble deep transfer network," *Knowledge-Based Systems*, vol. 213, p. 106695, 2021.

[24] K. Weiss, T. M. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *Journal of Big data*, vol. 3, no. 1, pp. 1–40, 2016.

[25] "Case western reserve university bearing data center website." https://csegroups.case.edu/bearingdatacenter/pages/project-history.

[26] W. Zhang, G. Peng, C. Li, Y. Chen, and Z. Zhang, "A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals," *Sensors*, vol. 17, no. 2, p. 425, 2017.

[27] R. R. Varior, B. Shuai, J. Lu, D. Xu, and G. Wang, "A siamese long short-term memory architecture for human re-identification," in *European conference on computer vision*, pp. 135–153, Springer, 2016.

[28] G. James, "Majority vote classifiers: theory and applications," 1998.

[29] W. A. Smith and R. B. Randall, "Rolling element bearing diagnostics using the case western reserve university data: A benchmark study," *Mechanical Systems and Signal Processing*, vol. 64, pp. 100–131, 2015.

[30] T. P. Carvalho, F. A. Soares, R. Vita, R. d. P. Francisco, J. P. Basto, and S. G. Alcalá, "A systematic literature review of machine learning methods applied to predictive maintenance," *Computers & Industrial Engineering*, vol. 137, p. 106024, 2019.