Sustainable Cities and Society 73 (2021) 103094

Contents lists available at ScienceDirect

Sustainable Cities and Society

journal homepage: www.elsevier.com/locate/scs

Fault classification in power distribution systems based on limited labeled data using multi-task latent structure learning

Mostafa Gilanifar [a], Hui Wang [*,a], Jose Cordova [a], Eren Erman Ozguven [a], Thomas I. Strasser [b], Reza Arghandeh [c]

[a] Florida A&M University-Florida State University College of Engineering, Tallahassee, FL, USA
[b] Electric Energy Systems at AIT Austrian Institute of Technology, Vienna, Austria
[c] Department of Computer science, Electrical Engineering and Mathematical sciences at Western Norway University of Applied Science, Bergen, Norway

ARTICLE INFO

Keywords:
Fault diagnosis
Power distribution systems
Distribution-level phasor measurement unit (D-PMU)
Machine learning
Multi-task latent structure learning

ABSTRACT

A significant issue for fault classification in power distribution systems is limited fault data for training classifiers to identify power failure types for remediation. Measurement data from power systems are mostly unlabeled without specified fault types, and labeled data with confirmed fault types are very limited, posing challenges to training classifiers with sufficient accuracy. Existing fault classification methods to deal with small labeled samples explore latent structures between labeled and unlabeled data. However, this line of methods has inaccurate assumptions on the relationship between unlabeled and labeled data and suffers from accuracy loss when dealing with very limited data that are labeled. This paper proposes a novel latent structure learning under a multi-task learning framework to supplement information and deal with the challenges in limited labeled data for fault classification. The proposed method not only takes advantage of the latent structure in unlabeled data that are not effectively utilized but also overcomes the limitations of latent structure learning by preventing classifiers from being overfitted to unlabeled data. The method was validated by an experimental study from distribution-level phasor devices in a hardware-in-the-loop testbed compared with state-of-the-art fault classification algorithms. The method is also demonstrated for the robustness against measurement noise.

1. Introduction

1.1. Background and related work

Fault classification or diagnosis in power distribution systems is essential to ensure faster fault isolation while reducing customer outage time. In recent years, new monitoring devices such as distribution-level phasor measurement units (D-PMU or micro-PMU), which provide phasor measurements for three-phase voltage and current, improve the situational awareness in power distribution systems. The faulty data available through these devices, along with state-of-the-art machine learning algorithms, provide the utility companies with more accurate fault classification or diagnosis for better remediation actions (Zhu et al., 2019).

Various fault diagnosis methods have been implemented for power distribution systems, which are mostly categorized under two major categories: impedance-based (e.g., El-Naily et al., 2020; Mora-Flarez et al., 2008) and traveling wave (e.g., Borghetti et al., 2010; Xiong et al., 2020) methods. The shortcoming of the impedance-based methods lies under their dependency on prior knowledge of network component characteristics such as physical specifications of conductors and transformer ratings, which may not be available or updated all the time (Gilanifar et al., 2020). On the other hand, traveling wave fault diagnostics methods are facing practical challenges due to mostly radial topology in power distribution systems with many short length branches. Moreover, traveling wave methods need high-frequency measurements that are costly and are not available all the time (Gilanifar, 2019). These methods are dependent on the topology of power distribution systems and the physical specifications of the system, such as transformer ratings that may not be available all time.

Aside from the aforementioned approaches, the development of advanced monitoring devices such as D-PMU and machine learning algorithms have brought more attention to data-driven fault diagnosis methods (Talaat et al., 2020). Machine learning utilizes the data from different types of faults that are collected by monitoring devices at a location to train a fault classifier. The advantage of this methodology is

* Corresponding author.
E-mail address: hwang10@fsu.edu (H. Wang).

https://doi.org/10.1016/j.scs.2021.103094
Received 8 February 2021; Received in revised form 25 May 2021; Accepted 11 June 2021
Available online 19 June 2021

that it is independent of the system topology and specifications and can provide more accurate and timely decisions that can be used for faster and better remediation actions. In the literature, Artificial Neural Networks (ANN) (Rahman et al., 2020; Sapountzoglou et al., 2020), and Support Vector Machines (SVM) (Borrás et al., 2016) are major machine learning-based anomaly detection technique in power systems. Researchers in Fan et al. (2018) investigated various autoencoder-based types and training schemes for anomaly detection. Moreover, Hu et al. (2019) developed a machine learning approach to automatically build a data-driven fault detection method by selecting an optimal set of sensors. A review on machine learning methods and their real-time applications in power systems can be found in Ahmad and Chen (2020). This line of methods usually requires a sufficient amount of data for the training.

1.2. Knowledge gap and proposed work

The **knowledge gap** in dealing with fault classification based on limited labeled data can be summarized as follows:

- ***There is a lack of effective fault classification methods to deal with very limited labeled data.*** Many machine learning-based fault diagnosis methods, including those research reviewed above, need data that are statistically sufficient for the training of classifiers (Gilanifar et al., 2019). However, electrical faults have a small sample size or are even rare events in the real-world power system, and the fault data types are insufficiently recorded and labeled (Gilanifar, 2019). The faulty events with limited records from a faulty location pose a significant challenge to most data-driven methods. The relevant research in the detection of faulty events with limited labeled data in power distribution systems is insufficient. Researchers in Gilanifar et al. (2020) proposed a multi-task learning (MTL) method to identify fault types in power distribution systems when all the measurement data are labeled in the supervised learning framework. However, when the labels in the training data are very limited, the learning accuracy can still suffer from a high classification error.
- ***Existing methods for classifying small labeled data impose inaccurate assumptions on the structure between the labeled and unlabeled data.*** Semi-supervised learning, a latent structure learning (LSL) algorithm, emerges recently to classify small-sample faulty data in power distribution systems (Zhou et al., 2018; 2019) using both labeled and unlabeled measurement data. This line of methods can generate the boundary of the classifiers passing through the data area with the least data density. However, such a classification strategy does not necessarily reflect true classes of data in reality (Chapelle et al., 2006). In addition, the LSL solely relies on the latent structure in the unlabeled data and labeled data from one single-source training data (e.g., one neighborhood or one node), potentially causing the trained classifier to be overfitted to the specific dataset that lacks generalization.

This paper proposes a novel fault classification method, named Multi-Task Latent Structure learning of Logistic Regression (MTLS-LR), to address the challenges in fault classification where very limited labeled data are available with confirmed types of electrical faults while a majority of data are unlabeled. The idea is to leverage multiple data measurements deployed at different parts of a power distribution system to supplement the information to the limited or even rare-event records of labeled data. The faulty events that occur at various locations captured by multiple measurement devices may exhibit different characteristics in the data. As such, we cannot implement traditional classification algorithms based on the merged data collected from multiple data sources. The information from different data sources may mislead the learning and jeopardize the fault classification accuracy.

The proposed MTLS-LR implements two main frameworks simultaneously: (1) a semi-supervised learning framework for the posterior class distributions when there are both labeled and unlabeled data, and (2) a multi-task learning framework that extracts common information and relatedness from multi-location data to improve learning. Under (2), the proposed method will find the similarity pattern or relatedness among similar-but-non-identical data sources as measured from different locations in a power distribution system. The relatedness will be utilized to guide the semi-supervised learning of the fault classifier under (1). The **contributions** of this research can be summarized below:

- *Application contribution*: One of the major bottlenecks for data-driven fault detection in power distribution systems is the lack of enough labeled fault event data to create training data sets for machine learning algorithms. This is due to the nature of faults which are rare events and also the historic inadequacy of monitoring devices at distribution sides in comparison to power transmission networks. This research develops an effective fault detection methodology for power distribution systems when dealing with limited labeled data or rare fault events.
- *Methodology contribution*: The method addresses the limitations in existing LSL methods that deal with small labeled data by leveraging multi-source historical information to prevent the classifier from being overfitted to unlabeled data.
- *Future significance*: The method and case studies demonstrate the values of unlabeled data scattered in the power distribution system that can be shared in improving fault classification. As such, the research motivates data sharing and exchange to support smart grid applications.

The remainder of this paper is organized as follows. Section 2 presents the schematic of the proposed MTLS-LR method and explains the two main parts of the proposed MTLS-LR, i.e., LSL and multi-task latent structure learning. A case study based on a hardware-in-the-loop (HIL) testbed consisting of multiple commercial distribution level phasor measurement units (D-PMUs) is explained in Section 3. The results are presented in Section 4, where Sections 4.1 and 4.2 discuss the potential of the algorithm in dealing with very limited labeled samples. The robustness of the proposed MTLS-LR against noisy fault measurements is studied in Section 4.3. Section 5 concludes the paper.

2. Multi-task latent structure learning for limited training data labeled

The schematic of the proposed MTLS-LR method is illustrated in

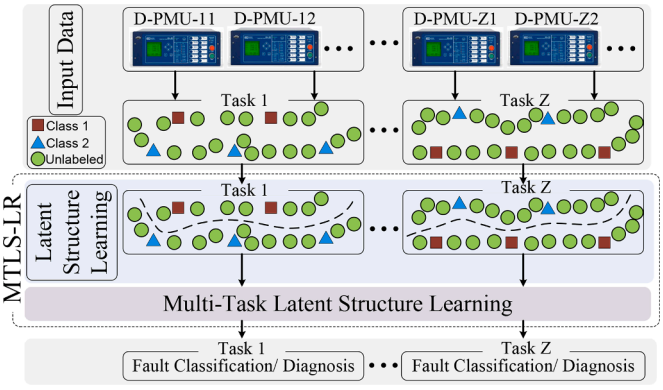

Fig. 1. Schematic of the proposed MTLS-LR. For each task, the majority of data are unlabeled (circles), whereas the labeled data (squares and triangles) are very limited. Each task focuses on the learning of fault classifiers by exploring latent structures between labeled and unlabeled data. The joint learning by multiple tasks can mitigate the limitation of LSL in each task.

Fig. 1. Historical data from multiple measurement devices, such as D-PMU, are collected from different nodes or neighborhood areas in a power distribution system. By integrating these data, multiple learning tasks are implemented simultaneously. Each task refers to the faulty event classification and diagnosis at a node or neighborhood area. The training datasets for each task include both the labeled and unlabeled samples, and labeled samples may include various types of short-circuit electrical faults measured. The majority of the training data are unlabeled, and labeled data can be very limited. For each learning task, a classification model is constructed to explore the latent structure between the labeled and unlabeled samples, aiming to encourage "closely-located" samples to be assigned with the same class label. The models from different tasks will be jointly learned by exploring the relatedness or similarity to mitigate the impact of limitations in LSL, as will be discussed in Section 1.2.

2.1. LSL for electrical fault classification

The LSL explores the structure between the labeled and unlabeled data to improve the classification of unlabeled data. The logic is that "similar" features, as reflected in the unlabeled data, belong to the same class. The training is achieved by modifying the boundaries of the classifier among the classes to pass through the data zones with the least density (Chapelle et al., 2006). As shown in Fig. 2, considering the features in unlabeled data can help modify the linear classification boundary (dash line) between classes 1 and 2 by forcing the boundary to pass through the data zones with a relatively lower density of unlabeled and labeled data.

Under the LSL framework, the estimation of a classifier is dependent on limited labeled data $D_L = \{(y_1, x_1), \cdots, (y_L, x_L)\}$ and unlabeled data $X_U = \{(x_{L+1}, \cdots, x_{L+U})\}$, where x_i is the data point, the y_i is the associated label, and L is the number of labeled data points. The LSL aims to train a classification function with the assistant of labeled data (D_L) and unlabeled data (X_U).

Different classifier models can be chosen for fault diagnosis. Without losing generality, this paper adopts a multinomial logistic regression (MLR) model to demonstrate the method since it is relatively convenient to obtain the posterior distribution of a fault class via a Bayesian framework (Li et al., 2010). The MLR can be represented as:

$$p(y_i = k|x_i, W) = \frac{exp\left(W^{(k)}x_i\right)}{\sum_{k=1}^{K} exp\left(W^{(k)}x_i\right)}, \tag{1}$$

where $x_i = [x_{i1}, \cdots, x_{il}]^T$ is a vector of l features. The term $W^{(k)}$ is the coefficients for class k and $W^{(k)}$ is the k^{th} column of the W for $k = 1, \cdots, K-1$ such that $W = [W^{(1)^T}, \cdots, W^{(K-1)^T}]^T$. The k in this paper represents the class and K shows the number of classes. Also, we can set $W^{(K)} = 0$ knowing that the MLR is not dependent on translation transformation on the $W^{(k)}$, which means that translation of $W^{(k)}$ will not affect the resulting probabilities. The posterior density can be:

$$p(W|Y_L, X_L, X_U) \propto p(Y_L|X_L, X_U, W)p(W|X_L, X_U) = p(Y_L|X_L, W)p(W|X_{L+U}),$$

where Y_L and X_L are the labels and feature vectors in D_L respectively, and X_{L+U} represents the $\{X_L, X_U\}$. Thus, the maximum a posteriori (MAP) of W can be estimated as:

$$\widehat{W} = \text{argmax}_W = \{l(W) + \log p(W|X_{L+U})\}, \tag{2}$$

where $l(W) = \log p(Y_L|X_L, W) = \log \prod_{i=1}^{L} p(y_i|x_i, W)$. In (2), the first term $l(W)$ denotes the log-likelihood function where all the labeled data D_L is considered. The second term $p(W|X_{L+U})$ considers the unlabeled data in the proposed LSL classifier. The term $p(W|X_{L+U})$ works as a prior on W that enforces the classifier to pass through the area with a smaller data density. Thus, this prior encourages "closely-located" data points for X to be assigned with the same class label. The math mechanism on how this prior works is briefly reviewed below.

This closeness between variables can be defined as a weighted graph $\mathscr{G} = (\mathscr{V}, \mathscr{E}, \mathscr{B})$. In this graph, $\mathscr{V}$ is the set of vertices of all the labeled and unlabeled data, $\mathscr{E}$ represents edges of the graph defined on $\mathscr{V} \times \mathscr{V}$, and $\mathscr{B}$ denotes weights of the graph defined on $\mathscr{E}$ ($\mathscr{B} \equiv \{\beta_{ij} \geq 0, (i,j) \in \mathscr{E}\}$). The Gaussian prior is adopted as:

$$p(W|\Gamma) \propto \exp\left\{-\frac{1}{2}W^T \Gamma W\right\}, \tag{3}$$

where Γ is the precision matrix that is defined as $\Gamma = \text{diag}(\lambda_1(X\Delta X^T + \tau I), \cdots, \lambda_{(k-1)}(X\Delta X^T + \tau I))$, where λ_k $k = 1, \cdots, (K-1)$ are some scale factors and $\tau > 0$ is a regularization parameter. Δ is a Laplacian matrix representing the graph $\mathscr{G}$. It can be shown that the $W^T\Gamma W$ can be expanded as:

$$W^T \Gamma W \propto \sum_{k=1}^{K-1} \lambda_k \left(\sum_{(i,j)\in\mathscr{E}} \beta_{i,j} \left[W^{(k)^T}(X_i - X_j) \right]^2 \right). \tag{4}$$

According to (3), the estimator of W can be obtained by the smallest value of $W^T\Gamma W$. Also, similar to Li et al. (2010), $\beta_{i,j}$ (weights of the graph) in this paper can be selected as $\exp(-\| X_i - X_j \|^2)$. Based on (4), the value of $\left[W^{(k)^T}(X_i - X_j) \right]^2$ should be smaller when the $\beta_{i,j}$ is large to reduce the $W^T\Gamma W$. Thus, by minimizing $\left[W^{(k)^T}(X_i - X_j) \right]^2$, the classifier can make those data that are connected with larger values of $\beta_{i,j}$ to have the same type of label that generates a smaller $\left[W^{(k)^T}(X_i - X_j) \right]^2$.

.

To obtain the MAP estimate of W in (2), an expectation-maximization (EM) algorithm is used. The EM algorithm is an iterative procedure with two steps in each iteration: E-step that computes the expectation value and M-step that maximizes the E-step. These steps at iteration t can be written as:

$$\textbf{E-Step: } P(W|W_t) \equiv E[\log p(W|D)|W_t], \tag{5}$$

$$\textbf{M-Step: } W_{t+1} \in \text{argmax}_W P(W|W_t), \tag{6}$$

where in Eq. (5), $D \equiv \{D_L, X_U\}$ denotes the set of labeled and unlabeled data. The main property of the above EM algorithm is that the $p(W_t|D)$ is non-decreasing for $t = 1, 2, \cdots$, and it converges to the local optima of the $p(W|D)$. For more details about the EM algorithm, please see Li et al. (2010).

Limitations with the LSL: As mentioned in the Introduction, the assumption that the classifiers' boundaries pass through data zones with low data density may not always be consistent with the reality (Goldberg et al., 2011). Fig. 3 illustrates an example where the latent structure learning (LSL) boundary (gray dash line) does not reflect the true

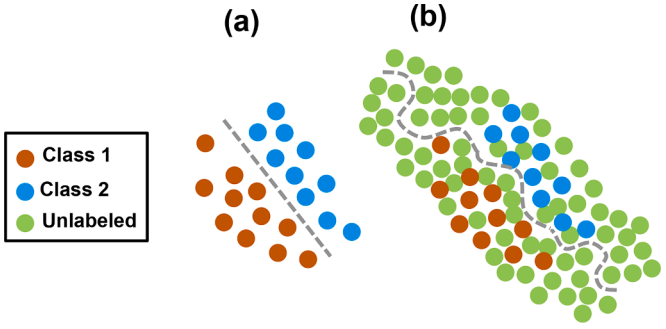

Fig. 2. The impact of unlabeled data on the classifier by LSL. (a) shows a linear classification boundary obtained from only labeled data. The existent of unlabeled data in (b) help construct a nonlinear classifier for classes 1 and 2 (red and blue dots)

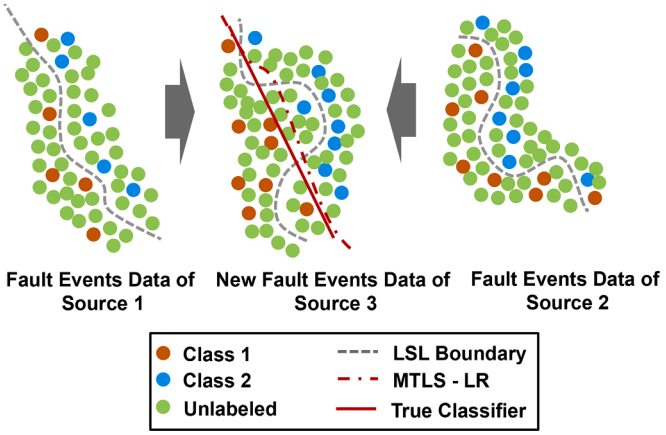

Fig. 3. Limitations with state-of-the-art latent structures learning (LSL), and multi-task learning can mitigate the problem that data are overfitted to unlabeled data by the traditional LSL methods.

boundary (solid red line) when the LSL boundary goes through the low-density data area. The LSL may overfit the trained classifier to the data. The next section will propose a multi-task LSL to address this limitation.

2.2. Multi-Task LSL

The training of LSL classifiers given limited labeled data can be improved under an MTL framework by leveraging the relatedness information among faulty events at different locations in a distribution system (Fig. 3). Such relatedness is reflected as the similar correlation patterns between the variables/features (x_i) and the response variable (y_i). Therefore, the logistic regressors (W) in the LSL associated with different parts of a system can be "similarly related".

Assume that faulty events occurred in Z different parts/locations within an electric grid. For each location, there are k possible types of faults (e.g., a line to ground, line to line, line to line to ground, etc.) where very limited data are labeled, whereas the majority are unlabeled in the training data. The objective is to identify the type of new faults in the testing data with the help of limited labeled data and an abundance of unlabeled data from not only the fault location of interest but also other locations in power distribution systems (i.e., multiple related-but-non-identical data sources). The input data is the voltage or current phasor measurements obtained from different locations of a network. In addition, for each location, a classifier represented by (1) will be learned.

The logistic regressors in the classifiers $\{W_1, \cdots, W_Z\}$ are assumed to be similarly related, where each of the $W_i, i = 1, \cdots, Z$ is for data source i. As summarized in (7), the proposed multi-task LSL aims to estimate the $\{W_1, \cdots, W_Z\}$ simultaneously given all data sources leveraging the relatedness between $\{W_1, \cdots, W_Z\}$. A hat symbol on top of the letters shows the estimated values.

$$\text{Objective: } \textit{To Estimate } \widehat{W}_1, \cdots, \widehat{W}_Z | (D_1, \cdots, D_Z), \tag{7}$$

where $D_i \equiv \{D_{Li}, X_{Ui}\}$, $i = 1, \cdots, Z$, contains both the labeled (D_{Li}) and unlabeled data (X_{Ui}). A regularization term is added to relate the learning of all the $\{W_1, \cdots, W_Z\}$ as follows:

$$\min_{\boldsymbol{W}} \ell(\boldsymbol{W}) + \lambda(norm(\boldsymbol{W})), \tag{8}$$

where $\ell(\boldsymbol{W})$ in Eqn 8 denotes a loss function of the latent structure learning classifier explained in Section 2.1 and $\boldsymbol{W} = [W_1, \cdots, W_Z]$, and λ is the coefficient.

By using the regularization term, one method to extract the relatedness among different tasks is the "*shared low-rank structures*". A norm is developed for regularization, which determines the rank of matrix $\boldsymbol{W}$. This norm extracts the similarity via a "shared low-rank structure" by letting the matrix $\boldsymbol{W}$'s rank be minimized. The minimization of matrix $\boldsymbol{W}$'s rank is an NP-Hard problem. Therefore, to estimate the rank function, a trace-norm (l_*) is implemented (Fazel et al., 2001), i.e., $\text{norm}(\boldsymbol{W}) = \| \boldsymbol{W} \|_*$. The trace-norm is defined as a function of singular values of matrix $\boldsymbol{W}$ as follows:

$$\| \boldsymbol{W} \|_* = \sum_{i=1}^{rank(\boldsymbol{W})} \sigma_i(\boldsymbol{W}), \tag{9}$$

where σ_i's are singular values calculated by a singular value decomposition (SVD) of the matrix $\boldsymbol{W}$. The proposed MTLS-LR includes multiple semi-supervised multinomial logistic regressions, i.e., semi-supervised softmax regressions for different tasks, which can obtain both the labeled and unlabeled input data. Moreover, a shared low-rank structure is implemented using a trace-norm (Eq. (9)) that shares the information amongst different tasks. The modeling structure the proposed MTLS-LR is shown in Fig. 4. The proposed MTLS-LR is a centralized method, where data from all tasks are available on a single machine and the parameters are computed using a standard single-thread algorithm. Furthermore, the proposed method processes all the signals from all the D-PMUs in the network.

2.3. Learning of the proposed MTLS-LR model

This section presents the learning procedures of the proposed MTLS-LR method. By adopting the "*shared low-rank structures*" method, the objective function of the proposed MTLS-LR can be derived as:

$$\min_{\boldsymbol{W}} \ell(\boldsymbol{W}) + \lambda \| \boldsymbol{W} \|_*, \tag{10}$$

where the coefficient $\boldsymbol{W}$ is postulated by a basis vector (B) multiplied with a coefficient matrix (C) as $\boldsymbol{W} = \boldsymbol{B}\boldsymbol{C}^T$ where $\boldsymbol{B} = \left[\vec{b_1}, \cdots, \vec{b_\nu}\right] \in \mathbb{R}^{p \times \nu}$ and $\boldsymbol{C} = \left[c_{ij}\right], i = 1, \cdots, L, j = 1, \cdots, \nu$, and ν is the rank of $\boldsymbol{W}$.

The matrix $\boldsymbol{B}$ is a subspace of matrix $\boldsymbol{W}$ that has smaller dimensions and plays a role in obtaining the relatedness among different faulty events recorded at various locations (i.e., different data sources). The matrix $\boldsymbol{C}$ may differ according to the data sources or fault locations. The aforementioned trace-norm extracts the relatedness from multiple faulty events at different locations of a system (Gilanifar and Parvania, 2021).

The problem defined in (10) is an unconstrained convex optimization problem with a non-smooth term, which is the l_* norm, presenting a significant challenge to solving the problem. One popular method (10) is the Accelerated Proximal Method (APM) (Gilanifar et al., 2019). Recently, the APM attracted more attention because of its capability of dealing with non-smooth optimization problems and its optimal convergence (Gilanifar et al., 2019; Nesterov, 1998). For more information about the APM procedures, please refer to Nesterov (1998).

The key procedures of the MTLS-LR method are summarized in Fig. 5. The parameters and hyperparameters are first initialized. In the training phase, the values of the hyperparameters should be determined by 10-fold cross-validation as outlined in Algorithm 1. Afterward, the MTLS-LR classifier is trained by using both limited labeled data and unlabeled data. The trained model will be applied to the testing dataset to estimate the likelihood of each fault type and determine the most probable type.

The performance is evaluated by calculating a confusion matrix based on the comparison between the predicted classes from the proposed MTLS-LR vs. the true classes. The classification error is used as an index to measure the percentage of faults detected mistakenly over all the faults available in the target fault location of interest.

Task Selection for the MTLS-LR: Great care should be exercised to include data sources (fault records from different locations) in the MTL. It is essential to identify the related tasks at multiple locations that can

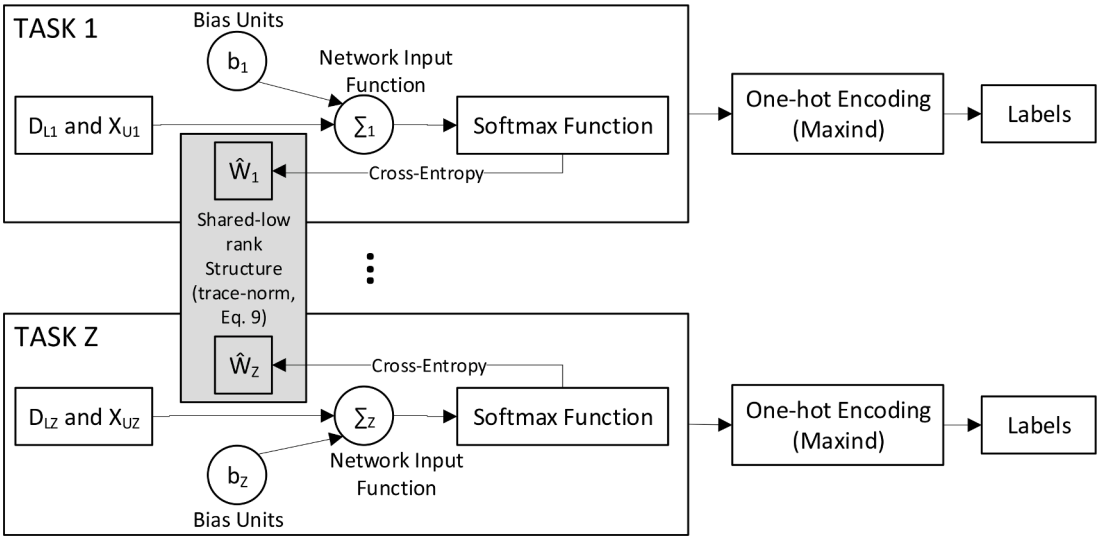

Fig. 4. Modeling structure of the proposed MTLS-LR.

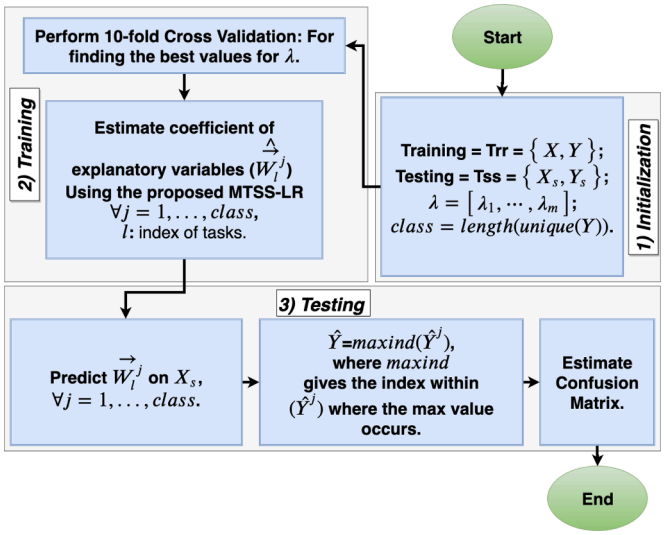

Fig. 5. A flowchart for the proposed MTLS-LR.

mostly contribute to the fault diagnosis in the target location while excluding those dissimilar tasks. A data-driven iterative procedure can be developed to explore the contributions from different combinations of learning tasks to the learning accuracy. Those tasks leading to the minimum classification errors will be considered as "related data" and selected for the training data (Gilanifar et al., 2019).

3. Case study

The proposed MTLS-LR method is validated by a case study on a D-PMU testbed. The results are compared with state-of-the-art machine learning algorithms on fault classification, given limited labeled data available for training.

3.1. D-PMU hardware-in-the-loop testbed

A realistic hardware-in-the-loop (HIL) testbed has been developed consisting of multiple commercial distribution level phasor measurement units (D-PMUs) from various vendors synchronized by a GPS clock. In this paper, the IEEE 123-node test feeder (as shown in Fig. 6) was used as an example since it is a common test feeder for fault diagnosis studies (Farajollahi et al., 2018; Gilanifar et al., 2020). It has overhead and underground lines, unbalanced loading, and multiple switching configurations (Farajollahi et al., 2018). The IEEE 123-node model was implemented in a multicore Opal-RT® real-time simulator with a rated voltage of 4.16 KV @ 60 Hz. Future research will also explore the potential of the proposed methodology generalizable for other applications in power systems.

This case study used actual data from commercial D-PMUs that are installed into the Opal-RT® target's Field-Programmable Gate Array (FPGA) output consoles via an amplifier. The D-PMUs computes and records the phasor values twice per cycle per nominal 60 Hz cycle that makes the output of 120 frames per second(von Meier et al., 2017). The D-PMU measurements are streamed to the open-source phasor data concentrator (OpenPDC) according to the IEEE C37.118 and IEC 61850 standards with their respective GPS-synchronized timestamp. Fig. 7 illustrates the physical testbed configuration. For more details regarding the HIL setup and the D-PMU specifications, please refer to Stifter et al. (2018).

The HIL generated different fault scenarios, including seven fault types in 7 locations over different power line segments. For each fault event, three-phase voltage and current magnitude and phase angle were recorded using multiple D-PMUs. These seven fault types include phase/line A to ground (AG), phase/line B to ground (BG), phase/line C to ground (CG), phases/lines A and B to the ground (ABG), phases/lines B and C to ground (BCG), phases/lines A and C to ground (ACG), and three-phase/line-to-ground (ABCG). Fig. 6 illustrates the location of faults by yellow arc symbols and the positions of the D-PMUs by blue stars.

This study created around 5000 fault events/samples with various fault impedance, type, location and then recorded the actual measurements from D-PMUs (Stifter et al., 2018). In our experimental setup, the fault impedance to ground is selected as follows: 0.01 Ω, 5 Ω, 10 Ω, 25 Ω, and 50 Ω. The minimum value of 0.01 Ohms was chosen to emulate the conditions of a bolted fault while the 50 Ohms was chosen as the high resistance scenario in our setup. Each fault case consists of pre-event and post-event conditions within a transient window (between no-fault and fault state) that is 0.2 s or 12 cycles. The transient window (the measurements between no-fault and fault state), as the input to the MTLS-LR method, is much more reduced and can also be employed to identify the fault occurring time. In this paper, 700 samples are generated for each fault location. The generated dataset of faulty events was used for training and testing in the proposed MTLS-LR method. It is worth mentioning that the proposed MTLS-LR is a data-driven machine

1. **Initialize:** $D_i = \{D_{Li}, X_{Ui}\}$; shows a training data which has both the labeled and unlabeled data for ith data source where $i = 1, \cdots, Z$, $\lambda = [\lambda_1, \cdots, \lambda_m]$.
2. Partition all the observations in D_i randomly into 10 equally parts such as $[p_{i1}, \cdots, p_{i10}]$, $\forall i$.
3. **For** $j = 1 : m$ (m is the number of suggested values for λ)
   (a) **For** $k \in \{1, \cdots, 10\}$
   - Obtain S_{ik} where S_{ik} is the training data for ith source that p_{ik} is excluded from it, ($S_{ik} = D_i \setminus p_{ik}$), $\forall i$.
   - $\hat{W}_i$ is obtained by fitting the MTLS-LR classifier on S_{ik} with λ_j, $\forall i$.
   - The predicted class ($\hat{C}_i^u$) for the observation u ($u = 1, \cdots, n$) of source i are obtained by predicting the trained classifier on p_{ik}, $\forall i$.
   - $Er_k = \frac{1}{n \times Z} \sum_i \sum_u I(C_i^u \neq \hat{C}_i^u) \times 100\%$, where I is the indicator function and C_i^u is the true class for observation u in ith source data.

   (b) **End For**
   (c) Calculate $\bar{Er} = \underset{k}{mean}(Er_k)$.
4. **End For**
5. Compute $\lambda^* = \lambda_{j^*} = \underset{j}{\text{argmin}} \{\bar{Er}\}$,

Algorithm 1. 10-fold Cross Validation steps used in the proposed MTLS-LR.

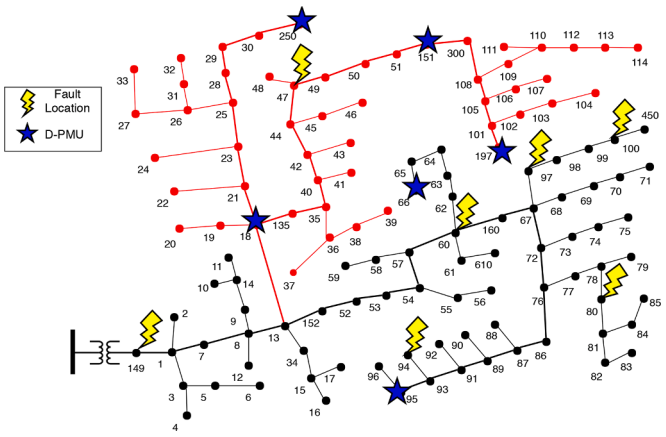

Fig. 6. An example IEEE 123-node test feeder that also shows locations of D-PMU and faults.

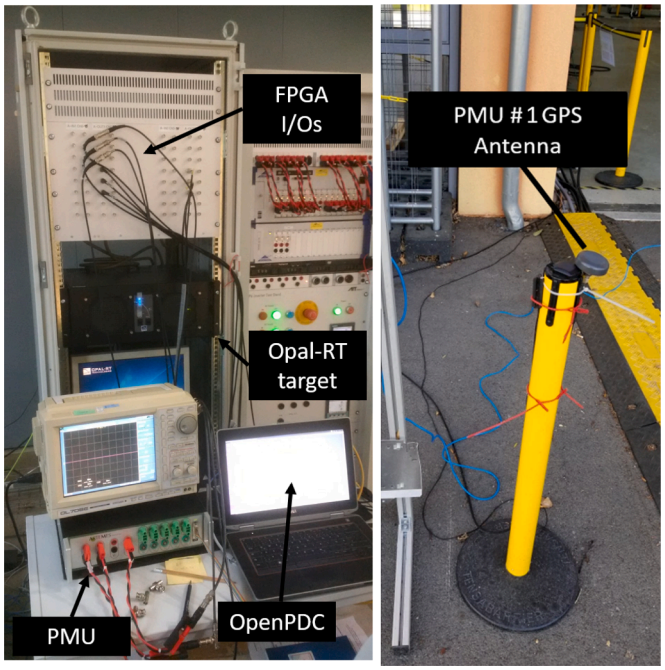

Fig. 7. Experimental Setup: (a) Opal-RT Target with D-PMU; (b) D-PMU GPS and its antenna.

learning method that works with voltage and current phasor measurement data from multiple D-PMU in different locations of a network. The major factor that would impact the fault classification accuracy is the similarity patterns in the signals measured at different locations and latent structures between labeled and unlabeled data no matter whether faults are on nodes or on lines.

For visualization, Fig. 8a shows an example of three-phase voltage magnitude measurements through a sequence of different fault types on Node 149 of the IEEE 123. Fig. 8b illustrates the magnitude for one of the voltage phases, as measured by different D-PMUs on Nodes 149, 95, and 197.

3.2. Experimental data description

The specifications of the training and testing datasets are presented in Table 1. The data is normalized by dividing each sample of a feature (such as the current or voltage column in the dataset) over vector-wise l_1 norm of that feature ($\text{column}_i = \text{column}_i ./ \textit{vecnorm}(\text{column}_i)$). The

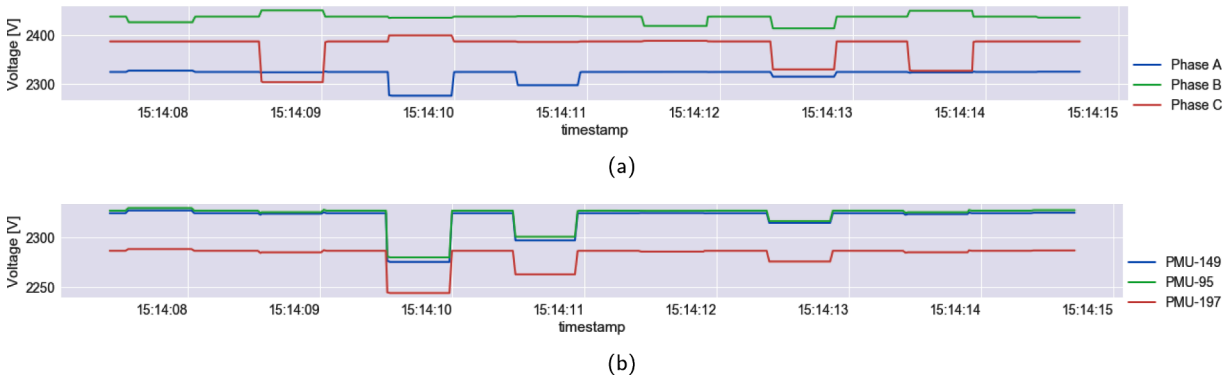

Fig. 8. Voltage measurement simulation based on the IEEE 123-node test feeder: (a) Three-phase measurements from node 149; (b) Phase A measurements from nodes 149, 95 & 197. Note that only degrading signals are included in the training and test dataset.

Table 1
Training and Testing Dataset that Used for Validation. Only 10 labeled data are available per location.

Test Feeders	IEEE 123-node
Tasks	2
Types of the Faults	7
Fault Impedance	0 to 50 ohm
Fault Locations Considered in Task 1	47
Fault Locations Considered in Task 2	100, 149, 94, 60, 97, 80
# of Training Samples (Percent)	3920 (80%)
# of Testing Samples (Percent)	980 (20%)

vecnorm is a MATLAB command for a vector-wise l_1 norm. It is worth noting that most data are unlabeled and only 10 data at each location are labeled with confirmed fault types for classification training.

Fig. 6 illustrates two sections of the network black and red colors that are corresponding to two learning tasks of fault classification. There are faults in node 47 in the red section. For the black section, there are faults in nodes 100, 149, 94, 60, 97, and 80. The fault types in task 1 are classified by using the data from both tasks 1 and 2 and the same with the classification for task 2. As such, the learning in each task utilized the measurements from all D-PMUs.

The training data consists of 80% of the total samples of faulty events, including both labeled and unlabeled data that account for the majority of the data. The remaining 20% were used for testing.

4. Results and discussions

This section presents the validation results of the proposed MTLS-LR based on the fault scenarios generated by the HIL testbed and the data in Table 1. The performance of MTLS-LR is compared with state-of-the-art machine learning methods for fault classification. This section further investigates the trade-off between labeled and unlabeled fault measurement data for different scenarios of lacking labeled data. In addition, the impact of noise in D-PMU measurement data on the performance of MTLS-LR is discussed by considering the deterioration of measurement data quality in real-world applications.

4.1. Classification error of MTLS-LR

The proposed MTLS-LR method was tested for limited labeled data, i.e., 10 labeled samples per location and type using voltage and current phasor signals. The confusion matrix of the proposed MTLS-LR is presented in Fig. 9. The overall classification errors of the MTLS-LR is 1.43%. It is noticeable that even with a very limited labeled data scenario, the proposed method could achieve a good accuracy of more than 98%.

The MTLS-LR method was also compared with classification

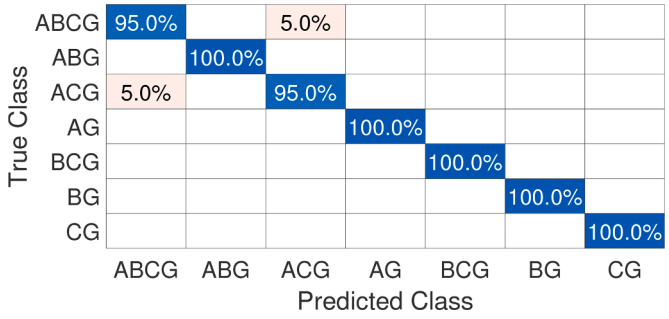

Fig. 9. Confusion Matrix of the proposed MTLS-LR method for the IEEE 123-nodes test feeder using both voltage and current phasor values. It should be noted that limited labeled data, e.g., 10 labeled data per location and type is used.

performance by using state-of-the-art supervised learning by using limited labeled data, i.e., 10 samples per location and type. These methods include support vector machine (SVM), and logistic regression (LR), and an LSL (semi-supervised learning) logistic regression method (Li et al., 2010). The comparison results are shown in Fig. 10. In this comparison, all the methods were tuned to improve the performance based on the same set of datasets. For instance, the radial basis function kernel for SVM was used, and the best parameters along with hyper-parameters were obtained after parameter tuning with 10-fold cross-validation. LR and SVM can only utilize the labeled data, while LSL and MTLS-LR use both the labeled and unlabeled data for the training step. Fig. 10 indicates that the MTLS-LR outperforms the LR and SVM by 66% and 50% in relative error reduction, respectively. The MTLS-LR also outperforms the LSL by 33%. It should be noted that other

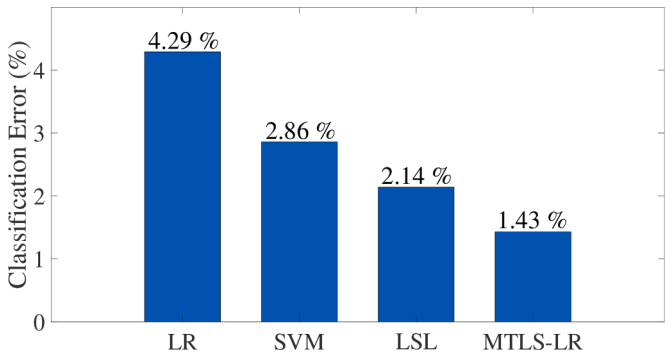

Fig. 10. Comparison of fault classification error between MTLS-LR and other machine learning algorithms considering both current and voltage as input data. The method demonstrated a clear advantage of MTLS-LR.

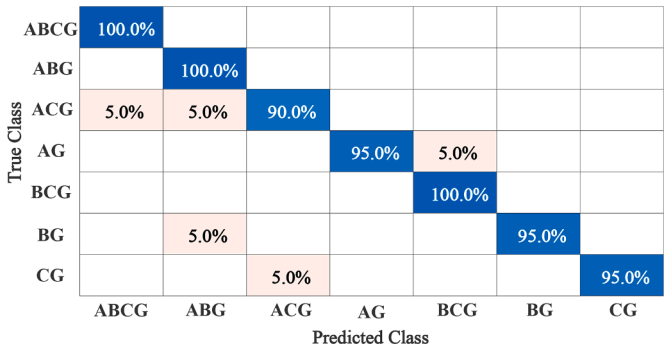

Fig. 11. Confusion Matrix of the feedforward deep neural network for the IEEE 123-nodes test feeder using both voltage and current phasor measurements.

learning philosophies such as reinforcement learning which is a sequential learning strategy that determines the decision at the next time given historical data, differs from our problem formulation that focuses on the small-sample learning of faults at one single time.

It is worth highlighting that (1) using unlabeled data (LSR method vs. LR and SVM methods) and shared information from different data resources (MTLS-LR vs. LSL) has advantages in improving fault classification/diagnosis trained by a small number of labeled data; and (2) the data measured from other related-but-non-identical faults at different locations can be leveraged to improve the learning accuracy under the MTL framework. As such, the results demonstrate the significant advantage of using the MTLS-LR when dealing with limited labeled data.

Comparison with a feedforward deep neural network (DNN): This section further compares the results obtained from the proposed MTLS-LR with a feedforward DNN. The same input data with the same features and the same function for normalization is implemented. The hyperparameters of DNN are selected through 5-fold cross-validation. For the cross-validation, a fully connected (FC) layer with ReLU activation function and dropout regularization in the hidden layers, along with the output layer fully connected with softmax activation is used. For selecting the best hyperparameters, 100 neurons were initially chosen for one FC layer. Then, different activation functions such as ReLU, leakyReLU, tanh, elu, as well as several optimizers such as Adam, RMSprop, SGD were tested for different learning rates to find the best activation function and optimizer. The results obtained from the cross-validation for the initial structure with the ReLU activation function and Adam optimizer, show a 98.97% validation accuracy. Moreover, various learning rates 1 and 2 were tested for finding the hyper-parameters of the Adam optimizer. The results showed 99.02% validation accuracy, i.e, the accuracy over the validation set, i.e., a part of the training dataset, when the learning rate 1 and 2 are respectively selected as $3e^{-3}$, 0.96, and 0.95.

The above selected hyper-parameters are used to find the other hyper-parameters of the DNN such as numbers of layers, neurons, and dropout probability for scenarios. According to the results, the best number of neurons for each hidden layer is selected as 200, the number of layers is 3, and the dropout rate is 0.1. After the DNN model is trained with the selected hyperparameters, it is tested on the same testing data that we used for our proposed MTLS-LR. The accuracy that is obtained from the DNN method is 96.43% (please see Figure 11) which is higher than LR and lower than SVM, LSL, and the proposed MTLS-LR method. The DNN is trained on only labeled data which is 10 samples per location and type. It even does not demonstrate superiority over traditional machine learning methods due to limited labeled data.

We record the computational time for the proposed method. It should be noted that all the calculations are conducted on a computer with an Intel Core i7-7500U and 2.70 GHz CPU. For the IEEE 123-nodes test feeder, the computational time for the offline training of the algorithm including the 10-fold cross-validation is around 365 seconds. The implementation of the obtained classification tool on testing data is much faster, i.e., less than 1 second and suitable for quasi-real-time implementation. To improve the algorithm, one can use better computing devices such as GPUs. Our future work is toward improving the performance of the algorithm for shorter computing times.

4.2. Ratio between labeled vs. unlabeled data

This subsection discusses the impact of changing the ratios between labeled and unlabeled samples on the classification performance. Table 2 shows the classification error in percentage when considering a different number of labeled and unlabeled data. There are 80 samples per location and type of fault in the training data. Thus, some entries in Table 2 do not have any results and are presented as dash (-) since the total number of labeled and unlabeled samples exceeds 80. The results imply that adding more labeled or unlabeled samples can help reduce the MTLS-LR classification error. For instance, adding 70 more labeled samples when there are only 6 labeled samples, would reduce the classification error by 86.9%. On the other hand, by introducing 70 unlabeled samples, the classification error is reduced by 23.5% when only six labeled samples are available.**Rare-event scenario with extremely limited labeled data:** The results in the first two columns in Table 2 reflect the rare-event scenarios with extremely limited labeled data, i.e., only 1 and up to 6 labeled data. It is observed that the proposed MTLS-LR method effectively reduced the classification error as more unlabeled data are included. For instance, by introducing 70 unlabeled samples, the classification error is reduced by 23.5% when only six labeled samples are available. The results showed the superiority of the MTLS-LR method in dealing with such extreme data-limited scenarios from the target fault location of interest.

The results in Table 2 also indicated that incorporating more labeled data in training would limit the improvement of adding more unlabeled data. In other words, the MTLS-LR becomes less advantageous compared to state-of-the-art classification methods when more labeled data from the same fault location and fault type are included. To show that the proposed MTLS-LR is more effective in dealing with the rare-event scenarios given with extremely limited labeled data, a new comparative study was conducted. In this study, the LSL and the proposed MTLS-LR were compared when the labeled data decreased to 1 sample per location and type.

As shown in Table 3, if the number of labeled samples is 1, the classification errors in both methods would increase (5.7% MLTS-LR vs. 21.43% LSL); however, the proposed MTLS-LR is affected much less than the LSL since it can leverage the information from other similar-but-non-identical data sources via multi-task learning.

4.3. Robustness of MTLS-LR against noise

This subsection investigates the robustness of the proposed MTLS-LR when the training data are affected by variation in fault events voltage or current phasor measurement data in real-world applications. Such variation (we can call it “noise” from the data science point of view) in fault measurement can be due to the load levels or fault impedance. The fault event measurement variation (aka noise) can potentially mislead the machine learning algorithm results.

In this analysis, the traditional LSL using logistic regression and the proposed MTLS-LR were implemented on the unlabeled data mixed with noises. First, the methods were applied to only four labeled data (L=4) and thirty unlabeled data (U=30) per location and type of fault. In addition, forty noisy unlabeled samples from the real-field simulation were added to the data, and the algorithm was tested for fault classification. The standard deviation of these forty noisy samples has a 2.81% difference with noiseless samples. It is observed that 26 out of 279 samples in the noisy data are out of the 1.5 interquartile ranges above the upper quartile or below the lower quartile. However, none of the normal samples is out of this range.

The classification errors of LSL and MTLS-LR in normal data and

Table 2
Classification error in percentage in the existence of various labeled (L) and unlabeled (U) samples (%) using only voltage phasor measurements. The first two columns show the rare-event scenarios when extremely limited data were labeled.

		L									
		1	6	10	20	30	40	50	60	70	80
U	**0**	36.428	10.9286	7.7142	5.7142	4.2857	3.5714	3.5714	3.5714	1.428	1.428
	5	34.285	10.2857	7.7142	5.7142	4.2857	3.5714	3.5714	3.5714	1.428	-
	10	34.285	10.2857	7.1428	5.7142	4.2857	3.5714	3.5714	3.5714	1.428	–
	20	33.571	10.2857	7.1428	4.2857	3.5714	2.8571	2.8571	2.1428	–	–
	30	33.571	9.6429	6.4285	3.5714	3.5714	3.5714	2.8571	–	–	–
	40	32.142	10.2857	6.4285	3.5714	3.5714	3.5714	–	–	–	–
	50	32.142	10.2857	6.4285	3.5714	2.8571	–	–	–	–	–
	60	32.142	8.3571	6.4285	2.8571	–	–	–	–	–	–
	70	32.142	8.3571	5.7142	–	–	–	–	–	–	–
	79	31.428	–	–	–	–	–	–	–	–	–

Note: There are 80 samples per location and fault type in the training data

Table 3
Impact of rare-event scenario on the proposed MTLS-LR and LSL methods using both voltage and current phasor measurements.

Methods	10 labeled samples/type/location	1 labeled samples/type/location
LSL	2.14%	21.43%
MTLS-LR	1.43%	5.7%

noisy data scenarios are summarized in Fig. 12. It can be seen that adding these noisy unlabeled data increased the classification errors for both LSL and MTLS-LR based on very limited labeled data. However, the relative increase error in the proposed MTLS-LR is smaller than the LSL method. Specifically, the classification errors of the proposed MTLS-LR and LSL are increased by 21.40% and 36.84%, respectively. The result indicates that the proposed MTLS-LR method can effectively overcome the limitations of the traditional LSL methods by guiding the learning of latent structure in the existent of misleading information in the noisy data. As such, the proposed MTLS-LR is less sensitive to the noise in the measurement data.

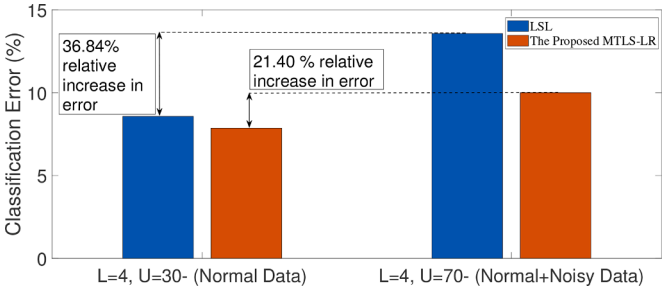

Fig. 12. Classification error of LSL using logistic regression and MTLS-LR based on the unlabeled data mixed with noisy patterns (%). The comparisons between the corresponding black bars or orange bars show the robustness of the proposed MTLS-LR over traditional latent stucture learning.

5. Conclusion

Fault classification and diagnosis can help develop fault detection, location, isolation, and service restoration (FLISR) solutions in power distribution systems. Nevertheless, electrical faults are much less recorded, whereas most data recorded by monitoring devices are unlabeled, presenting a grand challenge to train an accurate classifier. Traditional methods utilize latent structure between labeled and unlabeled data to improve the learning accuracy. However, this methodology has significant limitations in its inaccurate assumption on the relationship between labeled and unlabeled data. This paper develops a fault classification method, named MTLS-LR, based on very limited data that are labeled with fault types. The idea is to extract similar information from historical data in different sources/locations in the power distribution system to guide the exploration of the latent structure between labeled and unlabeled data while preventing the classifier from being overfitted to unlabeled data. As such, the **contribution** of this paper is to overcome the limitation in traditional LSL in dealing with fault diagnosis based on small records of labeled fault types while effectively utilizing an abundance of unlabeled data scattered in power distribution systems that have not been utilized in the prior research. The findings of this paper's case study can be highlighted as:

- The results show that the MTLS-LR method performs better than traditional fault classification methods, especially when labeled data are limited.
- The proposed MTLS-LR method is less vulnerable to noisy measurements in real-world applications.
- The proposed MTLS-LR method can improve traditional transfer learning methodology by (1) integrating it with latent structure learning and (2) modeling the between-data relatedness as the similar correlation patterns between measurement data and the likelihood of fault types.
- The study also discusses the ratio between labeled and unlabeled data to explore the applicable conditions of the proposed MTLS-LR. More unlabeled data can significantly reduce the classification error when the labels are very limited.
- This work can motivate data sharing in power distribution systems for smart grid applications.
- The proposed MTLS-LR method was validated using actual fault measurements obtained from multiple commercial D-PMUs in a hardware-in-the-loop testbed.

Future research direction can focus on cost-effectively finding electrical fault locations given limited labeled data.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Ahmad, T., & Chen, H. (2020). A review on machine learning forecasting growth trends and their real-time applications in different energy systems. *Sustainable Cities and Society, 54*, 102010. https://doi.org/10.1016/j.scs.2019.102010

Borghetti, A., Bosetti, M., Nucci, C. A., Paolone, M., & Abur, A. (2010). Integrated use of time-frequency wavelet decompositions for fault location in distribution networks: Theory and experimental validation. *IEEE Transactions on Power Delivery, 25*(4), 3139–3146. https://doi.org/10.1109/TPWRD.2010.2046655

Borrás, M. D., Bravo, J. C., & Montaño, J. C. (2016). Disturbance ratio for optimal multi-event classification in power distribution networks. *IEEE Transactions on Industrial Electronics, 63*(5), 3117–3124. https://doi.org/10.1109/TIE.2016.2521615

Chapelle, O., Chi, M., & Zien, A. (2006). A continuation method for semi-supervised SVMs. *Proceedings of the 23rd international conference on machine learning* (pp. 185–192). ACM.

El-Naily, N., Saad, S. M., & Mohamed, F. A. (2020). Novel approach for optimum coordination of overcurrent relays to enhance microgrid earth fault protection scheme. *Sustainable Cities and Society, 54*, 102006. https://doi.org/10.1016/j.scs.2019.102006

Fan, C., Xiao, F., Zhao, Y., & Wang, J. (2018). Analytical investigation of autoencoder-based methods for unsupervised anomaly detection in building energy data. *Applied Energy, 211*, 1123–1135. https://doi.org/10.1016/j.apenergy.2017.12.005

Farajollahi, M., Shahsavari, A., Stewart, E. M., & Mohsenian-Rad, H. (2018). Locating the source of events in power distribution systems using micro-PMU data. *IEEE Transactions on Power Systems, 33*(6), 6343–6354. https://doi.org/10.1109/TPWRS.2018.2832126

Fazel, M., Hindi, H., Boyd, S. P., et al. (2001). A rank minimization heuristic with application to minimum order system approximation*, 6. Proceedings of the american control conference* (pp. 4734–4739). Citeseer.

Gilanifar, M. (2019). *Heterogeneous Data fusion for performance improvement in electric power systems*. The Florida State University. Ph.D. thesis.

Gilanifar, M., Cordova, J., Wang, H., Stifter, M., Ozguven, E. E., Strasser, T. I., & Arghandeh, R. (2020). Multi-task logistic low-ranked dirty model for fault detection in power distribution system. *IEEE Transactions on Smart Grid, 11*(1), 786–796. https://doi.org/10.1109/TSG.2019.2938989

Gilanifar, M., & Parvania, M. (2021). Clustered multi-node learning of electric vehicle charging flexibility. *Applied Energy, 282*, 116125. https://doi.org/10.1016/j.apenergy.2020.116125

Gilanifar, M., Wang, H., Konila Sriram, L. M., Ozguven, E. E., & Arghandeh, R. (2019). Multi-task Bayesian spatiotemporal gaussian processes for short-term load forecasting. *IEEE Transactions on Industrial Electronics*. https://doi.org/10.1109/TIE.2019.29282751–1

Gilanifar, M., Wang, H., Ozguven, E. E., Zhou, Y., & Arghandeh, R. (2019). Bayesian spatiotemporal gaussian process for short-term load forecasting using combined transportation and electricity data. *ACM Transactions on Cyber-Physical Systems, 4*(1). https://doi.org/10.1145/3300185

Goldberg, A. B., Zhu, X., Furger, A., & Xu, J.-M. (2011). Oasis: Online active semi-supervised learning. *Twenty-fifth AAAI conference on artificial intelligence*.

Hu, R., Granderson, J., Auslander, D., & Agogino, A. (2019). Design of machine learning models with domain experts for automated sensor selection for energy fault detection. *Applied Energy, 235*, 117–128. https://doi.org/10.1016/j.apenergy.2018.10.107

Li, J., Bioucas-Dias, J. M., & Plaza, A. (2010). Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning. *IEEE Transactions on Geoscience and Remote Sensing, 48*(11), 4085–4098.

Mora-Flarez, J., Melandez, J., & Carrillo-Caicedo, G. (2008). Comparison of impedance based fault location methods for power distribution systems. *Electric Power Systems Research, 78*(4), 657–666. https://doi.org/10.1016/j.epsr.2007.05.010

Nesterov, Y. (1998). Introductory lectures on convex programming volume I: Basic course. *Lecture Notes*.

Rahman, M. A., Asyhari, A. T., Leong, L., Satrya, G., Hai Tao, M., & Zolkipli, M. (2020). Scalable machine learning-based intrusion detection system for IoT-enabled smart cities. *Sustainable Cities and Society, 61*, 102324. https://doi.org/10.1016/j.scs.2020.102324

Sapountzoglou, N., Lago, J., De Schutter, B., & Raison, B. (2020). A generalizable and sensor-independent deep learning method for fault detection and location in low-voltage distribution grids. *Applied Energy, 276*, 115299. https://doi.org/10.1016/j.apenergy.2020.115299

Stifter, M., Cordova, J., Kazmi, J., & Arghandeh, R. (2018). Real-time simulation and hardware-in-the-loop testbed for distribution synchrophasor applications. *Energies, 11*(4), 876. https://doi.org/10.3390/en11040876

Talaat, M., Alsayyari, A. S., Alblawi, A., & Hatata, A. (2020). Hybrid-cloud-based data processing for power system monitoring in smart grids. *Sustainable Cities and Society, 55*, 102049. https://doi.org/10.1016/j.scs.2020.102049

von Meier, A., Stewart, E., McEachern, A., Andersen, M., & Mehrmanesh, L. (2017). Precision micro-synchrophasors for distribution systems: A summary of applications. *IEEE Transactions on Smart Grid, 8*(6), 2926–2936. https://doi.org/10.1109/TSG.2017.2720543

Xiong, Q., Feng, X., Gattozzi, A. L., Liu, X., Zheng, L., Zhu, L., Ji, S., & Hebner, R. E. (2020). Series arc fault detection and localization in dc distribution system. *IEEE Transactions on Instrumentation and Measurement, 69*(1), 122–134.

Zhou, Y., Arghandeh, R., & Spanos, C. J. (2018). Partial knowledge data-driven event detection for power distribution networks. *IEEE Transactions on Smart Grid, 9*(5), 5152–5162. https://doi.org/10.1109/TSG.2017.2681962

Zhou, Y., Arghandeh, R., Zou, H., & Spanos, C. J. (2019). Nonparametric event detection in multiple time series for power distribution networks. *IEEE Transactions on Industrial Electronics, 66*(2), 1619–1628. https://doi.org/10.1109/TIE.2018.2840508

Zhu, J., Shen, Y., Song, Z., Zhou, D., Zhang, Z., & Kusiak, A. (2019). Data-driven building load profiling and energy management. *Sustainable Cities and Society, 49*, 101587. https://doi.org/10.1016/j.scs.2019.101587