This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2019.2938989, IEEE Transactions on Smart Grid

Multi-Task Logistic Low-Ranked Dirty Model for Fault Detection in Power Distribution System

Mostafa Gilanifar[1], Jose Cordova[1], Hui Wang[1], Matthias Stifter[2], Eren E. Ozguven[1], Thomas I. Strasser[3], and Reza Arghandeh[4]

***Abstract*—This paper proposes a Multi-task Logistic Low-Ranked Dirty Model (MT-LLRDM) for fault detection in power distribution networks by using the distribution Phasor Measurement Unit (PMU) data. The MT-LLRDM improves the fault detection accuracy by utilizing the similarities in the fault data streams among multiple locations across a power distribution network. The captured similarities supplement the information to the task of fault detection at a location of interest, creating a multi-task learning framework and thereby improving the learning accuracy. The algorithm is validated with real-time PMU streams from a hardware-in-the-loop testbed that emulates real field communication and monitoring conditions in distribution networks. The results showed that the MT-LLRDM outperforms other state-of-the-art classification methods using actual synchrophasor data achieved from a power hardware-in-the-loop testbed.**

***Index Terms*—Fault Detection, Power Distribution Networks, Synchrophasors, Hardware-in-the-loop, Multi-task Learning, low-rank structure.**

I. INTRODUCTION

Electrical faults cause undesired abrupt changes in voltage and current waveforms depending on the fault location, type, and grid conditions [1]. It is worthwhile mentioning that the smart grid revolution has created a paradigm shift in distribution networks. For example, the advent of two-way power flow due to distributed energy resources (DER) introduce new challenges for traditional protection system in power distribution networks which are designed initially for one-way power flow [2], [3]. Moreover, the controllable loads and DERs impose more variability at load side which makes it harder to distinguish type of faults especially the high impedance ones. These transformations impose different challenges on existing distribution networks monitoring systems and promote advanced tools such as distribution synchrophasors (D-PMU) to better observe, understand and manage power distribution systems[4]. The D-PMU data provide more opportunities for solving complex problems in power systems such as fault detection [5], [6], [7], topology detection [8], and load flow [9] which need higher quality data. Generally speaking, D-PMU devices are installed to support the operation of power distribution networks in different ways. Utilizing the data from D-PMU for fault type detection can be an added advantage for using such devices. Traditionally, the fault detection problem has been addressed by using current magnitude measurements. However, there have been previous studies remarking the use of voltage provided by synchrophasor units in the fault detection problem such as [10]. The [3] established the advantages of utilizing voltage measurements in conjunction with current measurements. For further information regarding the role of D-PMU data in advanced operation and control of the distribution networks readers can refer to NASPI [11] and U.S. Department of Energy reports [12].

This paper focuses on the fault type detection in distribution network with processing the D-MU data using a novel Multi-Task machine learning method. Using the D-PMU data for fault type detection is with this assumption that D-PMU will be available in distribution networks for the purpose of advanced situational awareness and support different functions in power distribution network operation. This paper provides an added advantage with utilizing D-PMU for fault type detection. Moreover, having a better knowledge of the type of fault provides a better foundation for the Fault location, isolation, and service restoration (FLISR) process. Additionally, having supplementary awareness of the fault type present in the system speeds up the fault mitigation process performed by the utilities technical crews. It is worthwhile mentioning that this algorithm can serve as an extra layer of intelligence in addition to the legacy operation and protection system.

Traveling-wave and impedance-based methods are among the most notable fault detection techniques. The disadvantage of the impedance methods is that they rely on the knowledge of the network components characteristics. Even with the introduction of Geographical Information Systems (GIS) tools into electric utilities, the components information are not updated regularly in practice, and the GIS data logs contain a variety of errors in spatial mapping of devices, wrong transformer ratings, and/or incorrect physical specifications of conductors [13]. Although Traveling-wave methods have proved to be accurate in transmission networks, they are facing more complexity in distribution networks due to mostly radial topology with many short length branches. They also require high-frequency measurements for reliable performance. Such high-resolution measurement data is expensive and may not be available all the time. Furthermore, with increasing complexities and uncertainties in a distribution networks, wide-area monitoring methods are proposed. These methods are usually used in distribution systems [14], [15], [16], [17].

With the development of information science and machine learning techniques, researchers enabled to develop advanced

*This research is partly supported by U.S. NSF 1640587 and 1744131 awards, and European Union's Horizon 2020 award 654113.

[1]M. Gilanifar and J. Cordova were Ph.D. candidates of IME and ECE Dept. at Florida State University when the work was completed; H. Wang and E. E. Ozguven are Assistant Professors with IME, and CEE Dept., Florida State University, Fl., USA. mg14m at my dot fsu dot edu

[2]M. Stifter is a Ph.D. student at TU Vienna, Institute of Energy Systems and Electrical Drives.

[3]T. I. Strasser is a Senior Scientist with AIT Austrian Institute of Technology, Center of Energy, Vienna, Austria.

[4]R. Arghandeh is a Professor in Western Norway University of Applied Science, Bergen, Norway.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2019.2938989, IEEE Transactions on Smart Grid

analytical approaches such as the neural network (NN) [18], [19], [20], support vector machines (SVM) [21], and principal component analysis (PCA) [22]. All of the machine learning methods mentioned above fall under the category of single-task learning (STL) methods whose classification performance is affected by the missing data or low measurement resolution of the training dataset.

This paper proposes a novel multi-task learning (MTL) framework named the *multi-task logistic low-ranked dirty model (MT-LLRDM)* to improve the accuracy of fault detection in power distribution networks. In the proposed method, the identification of the fault types at each location is based on integrating the recorded fault events obtained from phasor measurement units (PMUs) in multiple locations. The MTL framework will explore the similarity of the data patterns on voltage and current phasors induced by the faults among multiple locations to supplement more information about the nature of faults at another location. Although the correlation among the PMU data and fault types are different among multiple locations, the way how PMU data reflect the nature of the faults exhibit some similarities. For instance, voltage signals for a single-line-to-ground fault would indicate similar reduction patterns no matter where the fault occurs.

In the field of machine learning, finding the similarity between datasets from multiple tasks (e.g. sources, domains, locations) is still a significant challenge. The literature on multi-task learning present different regularization terms to extract shared knowledge by applying a weighted penalty term to the objective function [23], [24]. There are MTL methods dealing with the joint learning of Gaussian Process (GP) models for multiple tasks where multiple kernels characterize the between-task similarity. Therefore, the prediction at the testing dataset can be achieved by a weighted sum of the learned kernels [25]-[26]. However, the available studies on improving the performance of MTL when dealing with fault classification is still scant.

The contributions of this paper are summarized as follows:

1) From the methodology point of view, this paper proposes a novel MT-LLRDM algorithm to distinguish similarities among different tasks (datasets) by capturing both *common set of features* and *shared low-rank structures* in the data.
2) From the application point of view, this paper utilized the proposed MT-LLRDM approach for improving the fault detection performance in power distribution networks using D-PMU data. Furthermore, the validation of the MT-LLRDM algorithm is performed with a real-time hardware-in-the-loop testbed that emulates real field conditions of a distribution network.

The remainder of this paper is organized as follows. Section II explains the proposed MT-LLRDM. Section III describes the hardware-in-the-loop PMU testbed used for proof-of-concept validation, and then Section IV represents the results and conducts comparisons with state-of-the-art MTL, and machine learning methods to demonstrate the effectiveness of the proposed method. Section V concludes the paper.

II. Multi-task Logistic Low Ranked Dirty Model

This section presents the Multi-task Logistic Low Ranked Dirty Model (MT-LLRDM) for the fault detection problem. The overall framework of the proposed MT-LLRDM is shown in Fig. 1. Fault events measured by PMUs at different locations are the input to the MT-LLRDM algorithm. In this paper, the learning of fault classifier at each location is considered as a "task". The classifiers for all the tasks based on the logistic regression loss function are jointly learned by the proposed MT-LLRDM. Subsection II-A explains the building process for a multi-task multiclass logistic regression model starting from a binary logistic regression. Subsection II-B is a review of the existing regularization-based MTL methods. In subsection II-C, the formulation of the proposed MT-LLRDM is presented.

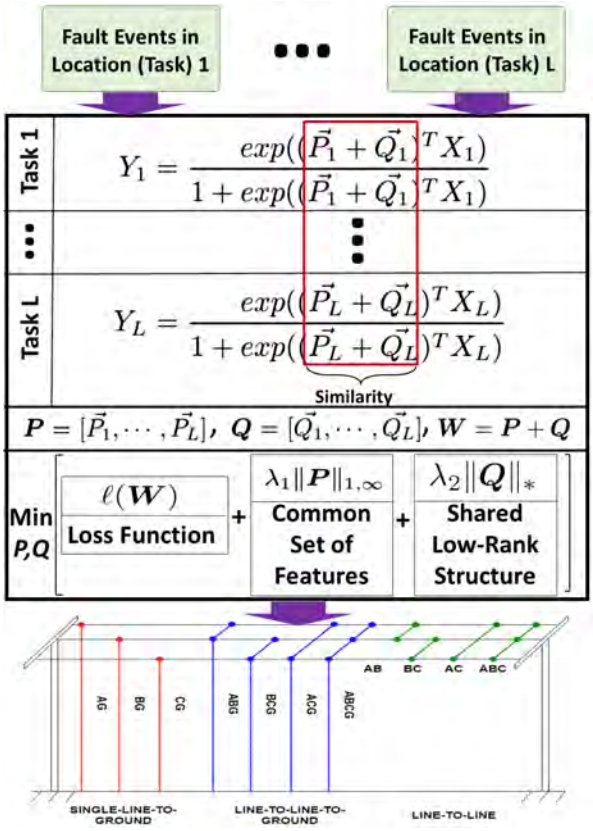


Fig. 1. Overall framework of the proposed MT-LLRDM for fault detection in power distribution. Fault types: Phase-A-to-ground (AG), Phase-B-to-ground (BG), Phase-C-to-ground (CG), Phases-A-B-to-ground (ABG), Phases-A-C-to-ground (ACG), Phases-B-C-to-ground (BCG), three-line-to-ground (ABCG), line-to-line-AB (AB), line-to-line-AC (AC), line-to-line-BC (BC), three-line-to-line (ABC).

A. Multi-Task Multiclass Logistic Regression Model

The multiclass logistic regression is a statistical learning method that is used to classify the types of faults by estimating the probabilities of each type. This multiclass logistic regression can be turned into k binary logistic regression problems that k is the number of classes. Suppose that we have k different faults happened in one location (single-task). For each fault type ($i = 1, 2, \cdots, k$), values of that type are turned into positive examples ($Y = 1$), and the values of rest of fault types are turned into zero ($Y = 0$). Then, for fault type i, the binary logistic regression is formulated as

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2019.2938989, IEEE Transactions on Smart Grid

$$h_i(X) = prob(Y_i = 1|X) = \frac{exp(\vec{W}^T X)}{1 + exp(\vec{W}^T X)},\ X \in \mathbb{R}^{n\times p}, \tag{1}$$

where $prob(Y_i = 1|X)$ takes a value between 0 and 1. Also, $X = (\vec{X}_1, \vec{X}_2, \cdots, \vec{X}_p) \in \mathbb{R}^{n\times p}$ is a set of explanatory variables including the voltage or current phasor measurements obtained from different PMUs. Moreover, $\vec{W} \in \mathbb{R}^p$ is coefficients of the explanatory variables (X). Once all the k binary logistic regressions are trained on the training data, they are used to predict the fault types in the testing data. First, the probability that the testing data belongs to each class (h_i(testing data)) is computed using the trained logistic regression classifiers. Then, we will pick the class for which the corresponding binary logistic regression classifier outputs the highest probability and return the class label $(1, 2, \cdots, k)$ as the prediction for the testing data. In short, two steps should be considered as

- Training: Train k binary logistic regression ($h_i(X)$) for each fault type i, $i = 1, 2, \cdots, k$,
- Testing: Select the fault type i that maximizes $h_{(i)}$(testing data) on a new testing data.

The fault classification obtained by a multiclass logistic regression at a certain location can be improved by using the recorded data from multiple locations under an MTL framework. This subsection explains the modeling structure of the multi-task logistic regression by exploring the similarities in PMU data patterns associated with electrical faults across different locations in a power network. This similarity is reflected by an alike relationship between the set of explanatory variables (X) and types of the faults (Y) across different locations in a power network. Therefore, the coefficients $\vec{W}$ in Eq. 1 can be *similarly related* under the MTL framework. Mathematical formulation of such similarity is discussed in the next subsections.

The proposed MT-LLRDM is estimated by learning the logistic regression classifiers from different locations ($l = 1, \cdots, L$) as shown in Fig. 1. The objective of MT-LLRDM is to estimate the $\boldsymbol{W} = \left[\vec{W}_1, \vec{W}_2, \cdots, \vec{W}_L\right]$ simultaneously given the data X for each fault location as well as the assumed similarity between $\vec{W}_1, \vec{W}_2, \cdots, \vec{W}_L$. It is noteworthy that the similarity comes from the $\boldsymbol{W}$ in different locations not between the X and Y. Mathematically, the learning objective is to jointly estimate $\hat{\boldsymbol{W}}|\left\{(X_1, Y_1), \cdots (X_L, Y_L)\right\}$. The methods to model the similarity are discussed next.

B. Review of Regularization Methods for Characterizing Similarity

This subsection reviews the formulation of the regularization-based MTL methods to capture between-task similarity. The two widely used perspectives to characterize the similarity between different tasks (here different fault locations) are summarized below:

1) Common Set of Features: The general formulation of "common set of features" perspective is to solve the following problem:

$$\min_{\boldsymbol{W}} \quad \ell(\boldsymbol{W}) + \lambda\|\boldsymbol{W}\|_{1,\infty}, \tag{2}$$

where $\boldsymbol{W} = [\vec{W}_1, \cdots, \vec{W}_L] \in \mathbb{R}^{p\times L}$ is a group of coefficients for all different tasks, and λ is a regularizer coefficient. Also, $\|\boldsymbol{W}\|_{1,\infty}$ is defined as

$$\|\boldsymbol{W}\|_{1,\infty} = \sum_j max(|W_{j1}|, \cdots, |W_{jL}|), \tag{3}$$

where W_{jk} is the the j^{th} row and k^{th} column of $\boldsymbol{W} \in \mathbb{R}^{p\times L}$. This norm penalizes the sum of the maximum absolute values of each row to encourage each row of $\boldsymbol{W}$ to have zero elements. Therefore, this regularizer can find a common set of input variables that have an effect on the fault classification among different locations. Also, $\ell(\boldsymbol{W})$ denotes a least square loss function for the different tasks. By assuming the logistic regression loss function, $\ell(\boldsymbol{W})$ forms as follows:

$$\ell(\boldsymbol{W}) = \sum_{l=1}^{L}\sum_{j=1}^{n_l} log(1 + exp(-Y_{l,j}(\boldsymbol{W}_l^T X_{l,j}))), \tag{4}$$

where n_l is the number of events for fault location l and $X_{l,j}$ is the j^{th} observation of X in location l. $\boldsymbol{W}_l$ is the coefficients for task l. A good example of using this perspective is Multi-Task Joint Feature Learning (MT-JFL) [27].

2) Shared Low-Rank Structure: The general formulation of the MT-LR is to solve the following problem, i.e.,

$$\min_{\boldsymbol{W}} \quad \ell(\boldsymbol{W}) + \lambda\|\boldsymbol{W}\|_*, \tag{5}$$

where the $\boldsymbol{W}$, $\ell(\boldsymbol{W})$ and λ are the same as Eq. 2. A trace norm regularizer ($\|\boldsymbol{W}\|_*$) is usually defined as a sum of singular values of $\boldsymbol{W}$, i.e.,

$$\|\boldsymbol{W}\|_* = \sum_{i=1}^{rank(\boldsymbol{W})} \sigma_i(\boldsymbol{W}), \tag{6}$$

where σ_i's are singular values of the matrix $\boldsymbol{W}$ obtained by a singular value decomposition. A good example of using this perspective is Multi-Task Logistic trace-norm Regression (MT-LR) [28]. The $\|\boldsymbol{W}\|_*$ provides a low-rank structure that includes common basis vectors shared across different faults happened in multiple locations. Suppose that $rank(\boldsymbol{W}) = \nu$. The component $\boldsymbol{W}$ can then be represented on a basis vector multiplied with a coefficient matrix 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 $\boldsymbol{C} = \left[c_{ij}\right], i = 1, \cdots, L$ and $j = 1, \cdots, \nu$. The basis vectors $\boldsymbol{B}$ span a low-dimensional subspace of matrix $\boldsymbol{W}$ and capture the similarities among different faults happened in different locations. The coefficient matrix $\boldsymbol{C}$ can be different for different locations. The two perspectives defined above capture the similarity from two different norms. Inspired by these two perspectives, this paper develops an improved method to capture the similarity in the next subsection.

C. The Proposed MT-LLRDM Formulation

As proposed by Jalali et al. [23], the variables coefficients ($\boldsymbol{W}$) can be decomposed into a group sparse component (named as $\boldsymbol{P}$) and a sparse component (named as $\boldsymbol{Q}$) in a *Dirty Model* as

$$\vec{W}_l = \vec{P}_l + \vec{Q}_l, \quad \vec{P}_l, \vec{Q}_l \in \mathbb{R}^p, \tag{7}$$

where $\vec{W}_l$, $\vec{P}_l$, and $\vec{Q}_l$ are l^{th} column of the $\boldsymbol{W}$, $\boldsymbol{P}$ and $\boldsymbol{Q}$ respectively. The group sparse component is assumed to capture the between-task similarity.

Based on the two norms explained above, this paper further develops a low-rank structure for the matrix $\vec{Q}_l$ to incorporate

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2019.2938989, IEEE Transactions on Smart Grid

more inter-location similarities that could not be captured by the group sparse component, leading to a logistic low-ranked version of *Dirty model* (MT-LLRDM). The objective of the MT-LLRDM is proposed as follows:

$$\min_{\boldsymbol{P},\boldsymbol{Q}} \quad \ell(\boldsymbol{W}) + \lambda_1 \|\boldsymbol{P}\|_{1,\infty} + \lambda_2 \|\boldsymbol{Q}\|_*, \tag{8}$$

where each column of $\boldsymbol{P}$ and $\boldsymbol{Q}$ corresponds to a fault location (task), i.e., $\boldsymbol{P} = [\vec{P}_1, \cdots, \vec{P}_L]$ and $\boldsymbol{Q} = [\vec{Q}_1, \cdots, \vec{Q}_L]$, and $\boldsymbol{W} = \boldsymbol{P} + \boldsymbol{Q}$. Moreover, λ_1 and λ_2 are the coefficients of the norms.

ALGORITHM 1: 10-fold Cross Validation for finding the best values of λ_1 and λ_2

1) **Initialize:** $Trr = \{X, Y\}$: Training data;
   $\lambda_1 = [\lambda_{11}, \cdots, \lambda_{1m}]$; $\lambda_2 = [\lambda_{21}, \cdots, \lambda_{2m}]$.
2) Find an equal random partition of $[p_1, \cdots, p_{10}]$ of all observations in Trr,
3) **For** $i = 1 : m$ (number of suggested values for λ_1 and λ_2)
   a) **10-fold cross validation:**
   For $k \in \{1, \cdots, 10\}$
   - Define $S_k = \{S_X, S_Y\}$ where $S_k = Trr \setminus p_k$,
   - n = length(p_k),
   - $\hat{\vec{\boldsymbol{W}}}_l^{(jik)} =$ MT-LLRDM.Fit($\lambda_{1i}^j, \lambda_{2i}^j; S_X, S_Y$),
   - $pred_l^{(jik)}$ = predict.MT-LLRDM ($\hat{\vec{\boldsymbol{W}}}_{\boldsymbol{l}}^{(jik)}, p_k$),
   - Er_k = Miss-classification error,

   b) **End For**
   c) Compute $\bar{Er} = \underset{k}{average}(Er_k)$,
4) **End For**
5) Obtain $\lambda_{1i^*}^j$ and $\lambda_{2i^*}^j$ by $i^{*j} = \underset{i}{\operatorname{argmin}} \{\bar{Er}\}$,

The norm regularization term $l_{1,\infty}$ penalizes the sum of the maximum absolute values of each row that encourages entire rows of the matrix to have zero elements. This regularizer introduces the group-sparse structure that encourages feature-selection by identifying the most appropriate features in any of the l tasks. Also, the trace-norm regularizer helps generate the low-rank structure in $\boldsymbol{Q}$. Moreover, λ_1 and λ_2 are coefficients of these two norms, and $\ell(\boldsymbol{W})$ denotes a least square loss function for task l as defined in Eq. 4.

The l_* norm of component $\boldsymbol{Q}$ provides a low-rank structure that has common basis vectors shared across multiple fault locations. Suppose that $rank(\boldsymbol{W}) = \nu$. The component $\boldsymbol{Q}$ can then be represented on a basis vector multiplied with a coefficient matrix as $\boldsymbol{Q} = \boldsymbol{B}\boldsymbol{C}^T$ where $\boldsymbol{B} = \left[\vec{b_1}, \cdots, \vec{b_\nu}\right] \in \mathbb{R}^{p \times \nu}$ and $\boldsymbol{C} = \left[c_{ij}\right], i = 1, \cdots, L$ and $j = 1, \cdots, \nu$. The basis vectors $\boldsymbol{B}$ span a low-dimensional subspace of matrix $\boldsymbol{Q}$ and capture the similarities between different fault locations. The coefficient matrix $\boldsymbol{C}$ can be different for different fault locations.

In summary, the proposed MT-LLRDM introduces a novel way to capture the similarities inspired by the *shared low-rank* and *common set of features* perspectives as described in Subsection II-B. The difference between the proposed MT-LLRDM and the state-of-the-art regularization-based MTL methods is shown in Table I. This method proposes the following hypothesis: If some similarly related information cannot be characterized by the common set of features, then it is likely to be captured by the shared low-rank structure, hence increasing the chance of capturing more similarities among multiple locations. Real-field data scenarios will be conducted in Section III to validate the effectiveness of this hypothesis.

Table I
DIFFERENCE BETWEEN THE PROPOSED MT-LLRDM AND OTHER MTL METHODS

Methods	Perspective	Norms	Application
MT-JFL [27], Dirty model [23]	Common Set of Features	$l_{1,\infty}$	Data dimensionality is large
MT-LR [28], SLR [24], and Robust [29]	Shared Low-Rank	l_*	Number of tasks is large
Proposed MT-LLRDM	Common Set of Features and Shared Low-Rank	$l_{1,\infty}$ and l_*	When both data dimensionality and number of tasks are large

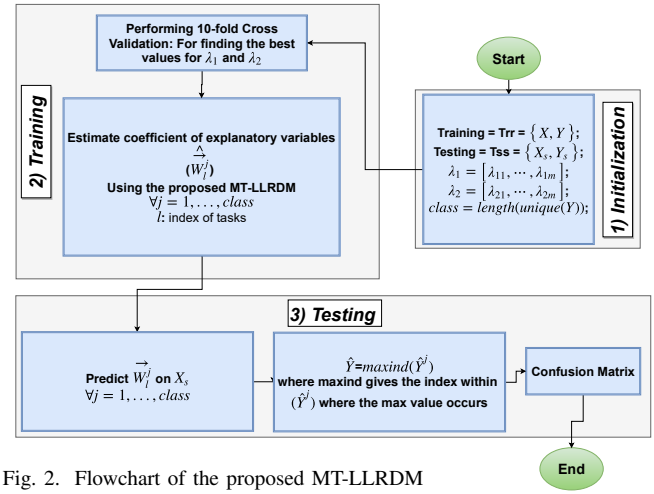


Fig. 2. Flowchart of the proposed MT-LLRDM

The proposed method in Eq. 8 is an unconstrained convex optimization problem, whose function is non-smooth. This non-smoothness exists in the l_* and $l_{1,\infty}$, posing a challenge to solve the problem and run the algorithm several times. For solving the non-convex Eq. 8, one way is to use the Accelerated Proximal Method (APM) [24]. Because of its optimal convergence rate and its capability in dealing with large-scale non-smooth optimization problems, it has gained more attention in recent years [30]-[31]. The APM consists of two main components including Proximal Operator and Proximal Operator Computation. Please refer to [31] for detailed procedures about the APM.

The proposed MT-LLRDM algorithm is summarized in Fig. 2. The first step is initializing the parameters including λ_1 and λ_2. After initializing, we train the MT-LLRDM on the training data for each class (fault type) and then, we test the trained MT-LLRDM on new test data. As explained in Section II-A, by using the one-vs-all method, we pick the class i that maximizes the classifier's output on the test data. Also, before training the MT-LLRDM in the training step, we run a 10-fold cross validation for finding the best values of λ_1 and λ_2. This 10-fold cross validation is explained in the Algorithm 1 in details. After finding the best values for λ_1 and λ_2, we train the MT-LLRDM by those best values and then, test the trained algorithm on the test data. After that, we choose a class that has the highest probability (h_i) for each fault event.

Finally, the confusion matrix is calculated by comparing the predicted and true classes. The confusion matrix is a square

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2019.2938989, IEEE Transactions on Smart Grid

matrix that consists of columns and rows listing the number of instances as "predicted class" vs. "actual class" respectively. After obtaining the confusion matrix, misdetection and false alarm rate error can be obtained. The misdetection error is equivalent to the False Negative (FN) error, which is the total number of entries above the main diagonal in the confusion matrix. On the other hand, false alarm rate error, which is equivalent to the False Positive (FP) error and is equal to the total number of entries below the main diagonal in the confusion matrix, can be estimated. Moreover, the misclassification rate, which is a summation of the misdetection and false alarm rate error, is calculated.

The Fig. 2 provides a way of integrating the fault event data at multiple locations to improve fault detection at a location of interest. Great care should be exercised for the selection of fault data locations to be included in the algorithm. It is not always better to include all the fault locations. Those locations where faults and PMU data exhibit very different correlations due to certain undiscovered reasons may negatively impact the learning accuracy. Therefore, it is necessary to conduct a location (or task) selection based on different combinations of locations. If there are L locations in a power network, there are $\sum_{i=1}^{L-1} \binom{L-1}{i}$ combinations of locations as inputs for the fault detection at a certain location of interest. The combination of the locations leading to the smallest classification errors will be reported as the optimal results.

III. Case Study and Data Description

In order to validate the proposed MT-LLRDM algorithm, we have developed two different test frameworks with multiple virtual and physical PMUs in a real-time power systems hardware-in-the-loop environment. We performed fault classification on the actual data set from the PMU streams obtained during fault conditions. This section describes the technical specifications and features of the mentioned testbeds. The comprehensive description regarding our developed PMU hardware-in-the-loop testbed is available in [32] by the authors. The first test feeder, IEEE 37-nodes, has largely unbalanced loads with uncommon connection configuration which is challenging for fault detection. The second test feeder, IEEE 123-node, is more common and it is widely used in previous studies [33].

A. IEEE 37-nodes Test Feeder

The first use case in this paper is the IEEE 37-nodes test feeder which is an actual feeder with a delta configuration, very unbalanced at 4.8 kV @ 60 Hz rated voltage and frequency with a substation transformer with 2% resistance and 8% reactance. The test feeder data sheet and specifications such as load parameters and line segment lengths can be found in [34]. The IEEE 37-nodes test feeder was selected as it is a challenging case for fault detection due to its unusual phase unbalance. The PMU streams were utilized to capture the fault events in a data repository utilized for machine learning algorithms.

The overall architecture of the PMU HIL testbed is illustrated in Fig. 3. The system consists of the IEEE 37-node test feeder model described above with virtual PMUs

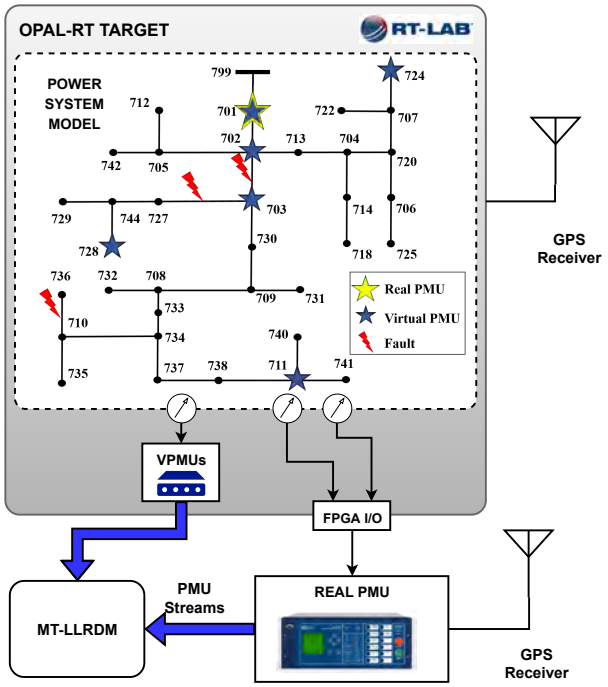


Fig. 3. PMU HIL setup for the IEEE 37-nodes test feeder that is used for fault detection application.

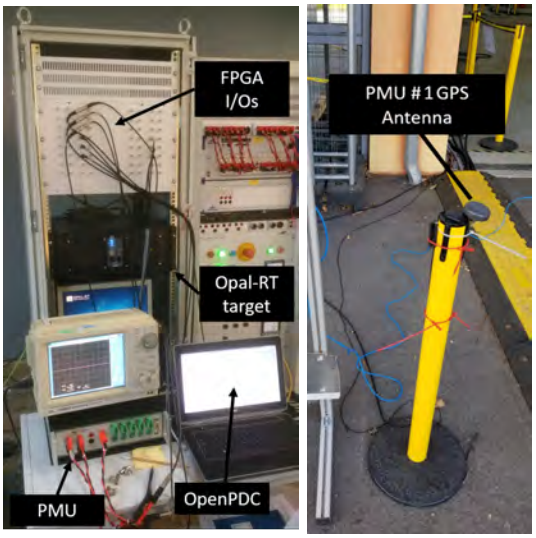


Fig. 4. Implemented HIL setup for fault detection: (a) PMU and Opal-RT Target; (b) PMU GPS Antenna.

measurements streams simulated inside the Opal-RT® Digital Real-Time Simulator (defined as *Target* from now on) and two actual commercial Class A (complying Standard IEC 61000-4-30) PMUs. To achieve measurements from the grid, the commercial PMUs are connected to the Opal-RT® target's Field-Programmable Gate Array (FPGA) output consoles. The phasor measurement method has a rating of 512 samples per nominal 60 Hz cycle with a streaming output of 120 frames per second. The PMU measurements are streamed to the open-source phasor data concentrator (OpenPDC) in compliance with the IEEE C37.118 standard with their respective GPS-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2019.2938989, IEEE Transactions on Smart Grid

synchronized timestamp. The advantage of this HIL setup is the capability to validate scenarios with multiple actual and/or virtual PMUs. Fig. 4 depicts the physical testbed configuration. For further details, please refer to the paper [32] by the authors.

The data collected from the PMU streams are utilized to perform fault classification. The PMUs provide three-phase magnitudes and angles from their respective node location in real-time. Different types of fault events were created on the IEEE 37 bus model at the locations shown in Fig. 3. The fault impedance in each fault events was changed from 0.01 to 50 Ohms and was chosen in accordance with [35]. In our experiments with HIL testbed, 99 fault events were created for each of the following fault types: *(i)* single-line-to-ground (AG), *(ii)* line-to-line (AB), and *(iii)* three-line-to-ground (ABCG). Moreover, we have simulated the faults in three different line segments to validate our algorithm. Table II shows a summary of the monitored nodes on the IEEE 37-nodes test feeder.

Fig. 5 illustrates the fault data produced in this study by showing a scatter plot for magnitude vs. angle measurements of one phase at node 728 of the IEEE 37-node test feeder. This figure displays all the combinations of fault types and fault locations for a total of 891 fault events.

B. IEEE 123-nodes Test Feeder

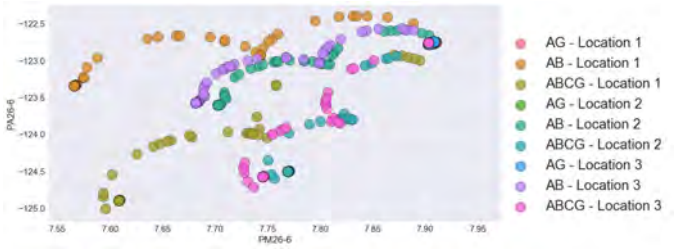


Fig. 5. Scatterplot for voltage magnitude vs. phase angle measurements from all 891 fault events simulated at the IEEE 37-nodes test feeder. Measurements as seen from node 701.

In order to further validate the proposed MT-LLDRM algorithm, a second testbed was developed in a similar setup as discussed in section III-A. The IEEE 123-nodes test feeder was modeled in the multicore Opal-RT® real-time simulator. The IEEE 123-nodes test feeder has a rated voltage of 4.16 kV @ 60 Hz. It has both overhead and underground lines, unbalanced loading and multiple switching configurations which is a more typical use case. This test feeder has been used in other PMU-based fault detection studies such as [33]. The test feeder specifications such as loads, line lengths, transformer ratings, etc. can be found in [34]. In this study, we have used the default switching configuration. As with the first test feeder, the IEEE 123-nodes test feeder was selected due to the unbalanced characteristics that are challenging for fault detection.

For the IEEE 123-nodes test feeder, the simulation setup is similar to the one used in the IEEE 37-nodes test feeder. It was simulated in an HIL framework in Opal-RT with communication and protocol IEEE C37.118 and IEC 61850 standards compliance. These PMU streams are collected by OpenPDC and serve as a data repository of recorded fault events. In essence, we have expanded the framework shown

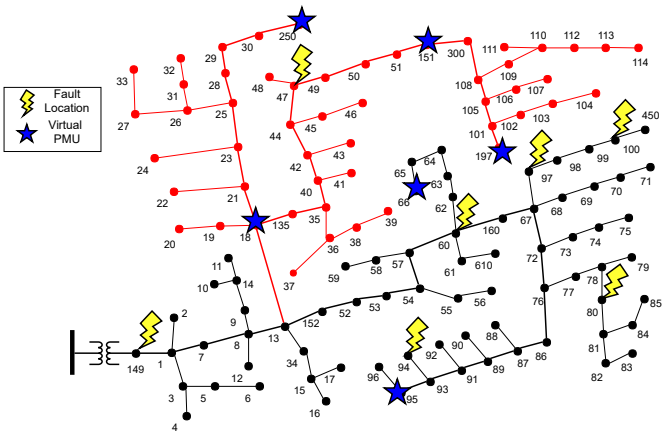


Fig. 6. PMU and fault locations in the IEEE 123-nodes test feeder.

in Fig. 3 with the IEEE 123-nodes test feeder shown in Fig. 6 as the simulated network.

The network of virtual PMUs monitors normal conditions prior to setting different fault types (balanced and unbalanced) to obtain random fault scenarios for detection and classification testing purposes. Table II and Fig. 6 show the locations of the physical and virtual PMUs inside the IEEE 123-nodes test feeder. Fault sequences of different types, fault locations, and random fault impedances have been simulated to generate a dataset which is used to train and validate the fault classification algorithm. Fourteen fault locations have been placed on different line segments with seven fault line-to-ground fault types: AG, BG, CG, ABG, BCG, ACG, and ABCG. Additionally, the fault impedances have been assigned from a uniform random distribution that ranges from 0.01 to 50 Ohms. Fourteen locations have been chosen (see Table II) according to the voltage drop and distance from the main feeder. Faults have been placed generating 100 events with changing fault impedances giving a total of 9,702 unique fault events for this testbed. As an example, a fault sequence is depicted in 7a, showing the measurements from node 95 for three phases. Additionally, Fig. 7b shows one of the phases for different PMU location (i.e., nodes 149, 95, and 197). As expected, it can be observed that faults vary in magnitude with different locations and faulted phases.

Table III provides a brief comparison for the two case studies used to validate the proposed MT-LLRDM methodology.

Table II
TEST FEEDERS' PMU AND FAULT LOCATIONS

IEEE 37-node test feeder			**IEEE 123-node test feeder**		
Nodes with PMU	Fault Types	Fault Location (Lines)	Nodes with PMU	Fault Types	Fault Location (Lines)
701 702 703 711 724 728	AG AB ABCG	702-703 703-727 710-736	18 66 95 197 151 250	AG, BG, CG AB, BC, AC ABCG	149 47 60 94 97 80 100

IV. RESULTS AND DISCUSSIONS

In this section, we validate our method using the fault events data produced by the hardware-in-the-loop framework described in section II. Two different test feeders have been

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2019.2938989, IEEE Transactions on Smart Grid

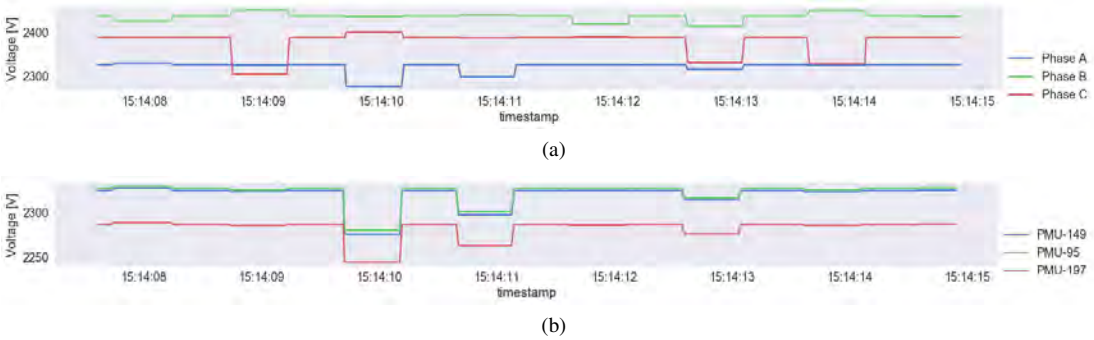


Fig. 7. Voltage measurements from the IEEE 123-nodes test feeder: (a) Three-phase measurements from node 149; (b) Phase A measurements from nodes 149, 95 & 197.

Table III
COMPARISON OF THE TEST FEEDERS MODELING

Feature	Case Study #1	Case Study #2
Test feeder	IEEE 37-nodes	IEEE 123-nodes
Simulation tool	Opal-RT/RT-Lab	Opal-RT/RT-Lab
Number of nodes	37	123
Fault locations	3	7
Types of faults	3	7
Real PMUs	1	-
Virtual PMUs	6	6
Synchronization	GPS / PTP	DRTS system clock
Measurements	V and I phasors	V and I phasors
Communication	C37.118	C37.118
Fault Events	**891**	**9,702**

used, the IEEE 37-nodes and the IEEE 123-nodes test feeders, where we analyze over 10,000 events. The main goal is to perform fault detection and accurately classify each observed fault into its true fault type.

Table IV
DATASET SPECIFICATIONS

Feature	IEEE 37-nodes	IEEE 123-nodes
Number of Tasks (Fault locations)	3	7
Training per Task (20%)	60	140
Testing per Task (80%)	240	560

As a first case study, we perform fault detection on the IEEE 37-nodes test feeder testbed and compare our results with state-of-the-art multitask learning (MTL) and single-task (STL) learning methods. The dataset specifications are shown in Table IV. For this usecase, the number of tasks is three which is the number of pre-recorded fault locations. In the MT-LLRDM, the fault detection at each location is considered as a different task where different fault events have been recorded. As shown previously in Fig. 1, the fault classification is performed with meaningful information from other fault locations. In other words, we use the fault data from three different locations as three different tasks from which all three locations are used as a training data set. Each task has all available information from the seven PMUs used in this case study.

We specify our training set as 20% of the total observed fault events. The remaining 80% is used for testing purposes. We divided our experiments by type of measurement (e.g., voltage and current) utilizing the magnitude and phase angle values from seven PMUs as mentioned in section III for training the proposed machine learning algorithm. We separate our experiments to reflect the fact that some feeders may not have both measurements. Therefore, we validate our approach with voltage and current measurements separately.

By knowing that the magnitude and phase angle measurements from the same fault type will have a similar signal profile, MT-LLRDM attempts to capture the similarity among different locations in a power system network. MT-LLRDM extracts the similarity of the fault data occurring in different locations by finding those variables that have the same effects on types of the faults (*common set of features*) in conjunction with a low-dimensional subspace (in $\boldsymbol{Q}$), which has the shared information among different locations (Figure 1). In total, we classify each testing event by fault type, i.e. single-line-to-ground (AG), line-to-line (AB) and three-line-to-ground (ABCG).

Table V
MEAN MISCLASSIFICATION RATE FOR ALL FAULT LOCATIONS OF 37-NODE TEST FEEDER AND DIFFERENT METHODS (%)

Phasor Measurements	STL		MTL		
	SVM	**NN**	**MT-JFL**	**MT-LR**	**MT-LLRDM**
Voltage	7.9	5.65	2.75	2.68	1.05
Current	7.28	5.29	3.60	3.32	1.76

In order to validate the results of MT-LLRDM in the fault detection process, we utilize the *misclassification rate* which measures the percentage-wise ratio of wrong detected faults over the total number of faults in that location. Let C_i be the class (i.e. fault type) of any given fault (single-line-to-ground, line-to-line or three-line-to-ground) and $\hat{C}_i$ the class prediction for the ith observation or event under fault conditions. Then, the misclassification rate in percentage (MR) is calculated as follows:

$$MR = \frac{1}{n}\sum_{i} I(C_i \neq \hat{C}_i) * 100\% \tag{9}$$

where n is total number of observations (or faults). As an example, if out of a total of 100 fault events, 10 were classified incorrectly (not accurately assigned a fault type), then the total misclassification rate is 10%. In Table V, we show the mean misclassification error using voltage and current phasor values for over three different locations. We compare the proposed method with state-of-the-art single-task and multi-task learning algorithms for fault detection in power systems. For the single-task learning algorithms, we choose SVM and NN, which have been most widely used for fault detection. In this paper, we have tried to obtain the best possible results

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2019.2938989, IEEE Transactions on Smart Grid

of SVM and NN methods for this data set by optimizing the hyperparameters and kernel automatically using "fitcecoc" and "patternnet" functions in the MATLAB® by minimizing five-fold cross-validation loss. The MT-LLRDM outperforms both SVM and NN by a relative error decrease of 86% and 81% respectively. Among multi-task learning algorithms, we chose MT-LR and MTL-JFL as two methods that are using just one of the "common set of features" or the "shared low-rank" strutures as explained in Section II-C.

Our MT-LLRDM using voltage phasor measurements outperforms the MT-LR and MT-JFL by a relative error decrease of 61% and 62%, respectively. This accuracy improvement shows that using the proposed MT-LLRDM can extract more similarities in fault events data from different locations. Fault detection based on current phasor measurements follows a similar trend showing a relative error decrease of 75% and 67% for MT-LLRDM in comparison to SVM and NN respectively. Moreover, the MT-LLRDM outperforms the MT-LR and MT-JFL by a relative error decrease of 47% and 51% respectively. One interesting observation is that using PMU voltage measurements with MT-LLRDM provides slightly better results for fault detection in comparison to using current phasor values.

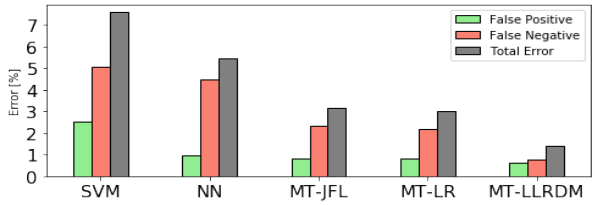


Fig. 8. Fault detection error using different classification methods for the IEEE 37-Nodes Test Feeder.

We also compare the false-positive and false-negative errors as shown in Fig. 8. The proposed MT-LLRDM shows an improvement of 75% and 36% compared with SVM and NN in false positive error and improvement of 85% and 83% in false negative error. The same trend of accuracy improvement can be observed when comparing the MT-LLRDM with the MTL methods. The proposed MT-LLRDM outperforms the MT-LR and MT-JFL by 22% and 25% in false positive error and 65% and 66% in false negative error.

We further validate our MT-LLRDM with a second dataset simulated in a similar hardware-in-the-loop setup as with the IEEE 37-nodes test feeder. For this testbed, we utilize the IEEE 123-nodes test feeder where we simulated a larger number of fault events at more different locations as described in section III-B. In this case study, we have seven different fault locations that represent the total number of tasks for our MT-LLRDM. Similarly to the previous case, we use 20% of the total number of recorded fault events as training data and 80% for testing purposes. The dataset is provided by 6 PMUs streams (see Fig. 6) where we have recorded 7 different fault types: phase-A-to-ground (AG), phase-B-to-ground (BG), phase-C-to-ground (CG), phases-A-B-to-ground (ABG), phases-A-C-to-ground (ACG), phase-B-C-to-ground (BCG) and three-phase-to-ground (ABCG).

In table VI, we compare the misclassification rates results of our proposed MT-LLRDM methodology with both multitask and single-task learning algorithms. For fault detection based on voltage measurements, the MT-LLRDM outperforms the multitask learning algorithms, MT-LR and MT-JFL, by a relative error decrease of 39% and 52%, respectively. Furthermore, the MT-LLRDM presents a relative error decrease of 73% and 76% for SVM and NN respectively, showing that MT-LLRDM outperforms the single-task learning methods. A similar trend can be observed for the current-based fault detection misclassification results.

Table VI
MEAN MISCLASSIFICATION RATE FOR ALL FAULT LOCATIONS OF 123-NODE TEST FEEDER AND DIFFERENT METHODS (%)

Phasor Measurements	STL		MTL		
	NN	SVM	MT-JFL	MT-LR	MT-LLRDM
Voltage	21.25	14.71	10.43	8.25	5.01
Current	17.78	13.90	10.41	8.07	4.98

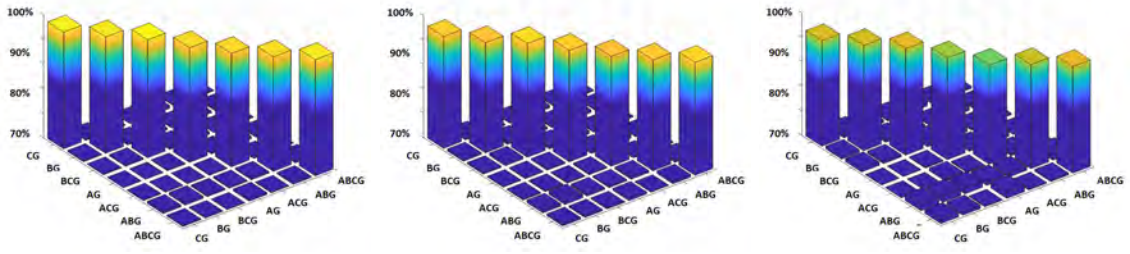
Fig. 9 depicts the comparison of fault detection accuracy for the IEEE 123-nodes usecase for the proposed MT-LLRDM (Fig. 9a), MT-LR (Fig. 9b), and SVM (Fig. 9c methods. The prediction accuracy is shown by the diagonal bars while the positive-negative and negative-positive errors are represented by the off-diagonal bars. It can be observed that the MT-LLRDM outperforms the other two algorithms in detecting each fault type.

It is worth remarking that although the misclassification rate for the IEEE 123-nodes test feeder outperforms the state-of-the-art methods, it is higher than the results presented for the IEEE 37-nodes test feeder. This is due to the fact that the IEEE 37-nodes test feeder is a smaller network with a higher number of PMUs which provides more information about the grid. In turn, the IEEE 123-nodes has a larger number of nodes, longer line segments and less number of PMUs, making the fault detection more difficult.

We further 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 37-nodes test feeder, the computational time for fault detection including all the training (with 10-fold cross validation) and testing part is around 324 seconds. Also, for the IEEE 123-nodes test feeder, the computational time is around 359 seconds.

To show the relationship between number of PMUs and the accuracy of fault detection, a sensitivity analysis has been done for IEEE 123 with using different number of PMUs with following scenarios: (i) two PMUs at nodes 18 and 95, (ii) three PMUs located at nodes 18, 95, and 197, (iii) four PMUs at nodes 18, 95, 197, and 66, (iv) five PMUs at nodes 18, 95, 197, 66, and 151, and (v) six PMUs at nodes 18, 95, 197, 66, 151, and 250. As it is shown in Fig. 10, it can be observed that having 5 PMUs may be the sufficient number of PMUs in the IEEE 123-nodes test feeder and adding more number of PMUs doesn't lead to a massive drop in the fault detection error. Going from 5 PMUs to 6 PMUs only reduces the error by 1.38% and 1.58% for the voltage and current respectively. This amount of reduction can be ignored, and 5 PMUs can be chosen for this test feeder to reduce the cost for creating a PMU monitoring system. On the other hand, by

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2019.2938989, IEEE Transactions on Smart Grid



(a) (b) (c)

Fig. 9. Confusion Matrix for different methods for the IEEE 123-nodes test feeder. Diagonal terms are correct identifications and off-diagonal ones are misclassifications. mACC for multi-class detection accuracy: (a) MT-LLRDM; (b) MT-LR; (c) SVM distance.

considering the recent trends in PMU market, the industry is observing a price reduction for synchrophasor devices due to the reduction in microprocessors cost, the economy of scale for PMU production, the advent of more PMU producers, and availability of new open-source PMU devices such as OpenPMU initiative [36]. The OpenPMU makes a complete PMU unit assembled for almost $1000. Moreover, a promising initiative is the Open-Box PMU presented by Pinte et al. [37] with a low-cost around $250.

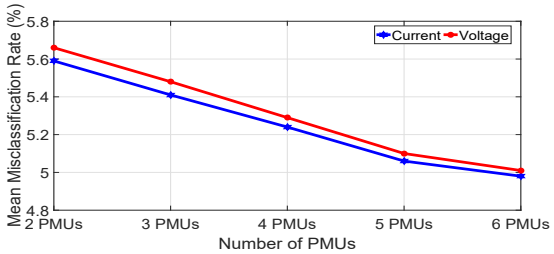


Fig. 10. Mean Misclassification rate (in %) obtained by MT-LLRDM for the IEEE 123-Nodes Test Feeder for different number of PMUs.

V. CONCLUSIONS

The highlights of this paper can be summarized as follows:

- The novel multi-task framework, MT-LLRDM, is proposed to capture similar information on fault events across different parts of the distribution networks.
- For improving the learning performance of the MT-LLRDM, a low-ranked structure of the *dirty model* is proposed that simultaneously employs the "common set of features" and "shared low-rank" structures.
- Based on actual PMU streams from a PMU HIL testbed, the proposed MT-LLRDM outperforms both state-of-the-art STL methods including SVM, NN and MTL methods including MT-LR and MT-JFL.

Future work will include the use of the proposed MT-LLRDM in the fault location application. Additionally, a Hardware-in-the-loop framework with an interface to power electronics inverters will be implemented to further validate the algorithm's performance with the presence of distributed energy resources.

REFERENCES

[1] M. Bollen and I. Gu, *Signal processing of power quality disturbances*, ser. IEEE Press series on microelectronic systems. IEEE Press, 2006.

[2] H. Mohsenian-Rad, E. Stewart, and E. Cortez, "Distribution synchrophasors: Pairing big data with analytics to create actionable information," *IEEE Power and Energy Magazine*, vol. 16, no. 3, pp. 26–34, May 2018.

[3] M. Pignati, L. Zanni, P. Romano, R. Cherkaoui, and M. Paolone, "Fault detection and faulted line identification in active distribution networks using synchrophasors-based real-time state estimation," *IEEE Transactions on Power Delivery*, vol. 32, no. 1, pp. 381–392, Feb 2017.

[4] A. von Meier and R. Arghandeh, "Chapter 32 - Every Moment Counts: Synchrophasors for Distribution Networks with Variable Resources," in *Renewable Energy Integration (Second Edition)*, L. E. Jones, Ed. Boston: Academic Press, Jan. 2017, pp. 435–444.

[5] Y. Zhou, R. Arghandeh, H. Zou, and C. J. Spanos, "Nonparametric event detection in multiple time series for power distribution networks," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 2, pp. 1619–1628, Feb 2019.

[6] Y. Zhou, R. Arghandeh, and C. J. Spanos, "Partial knowledge data-driven event detection for power distribution networks," *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 5152–5162, Sep. 2018.

[7] J. Cordova, C. Soto, M. Gilanifar, Y. Zhou, A. Srivastava, and R. Arghandeh, "Shape Preserving Incremental Learning for Power Systems Fault Detection," *IEEE Control Systems Letters*, vol. 3, no. 1, pp. 85–90, Jan. 2019.

[8] G. Cavraro and R. Arghandeh, "Power Distribution Network Topology Detection With Time-Series Signature Verification Method," *IEEE Transactions on Power Systems*, vol. 33, no. 4, pp. 3500–3509, Jul. 2018.

[9] A. Tbaileh, H. Jain, R. Broadwater, J. Cordova, R. Arghandeh, and M. Dilek, "Graph Trace Analysis: An object-oriented power flow, verifications and comparisons," *Electric Power Systems Research*, vol. 147, pp. 145–153, Jun. 2017.

[10] S. Lotfifard, M. Kezunovic, and M. J. Mousavi, "Voltage sag data utilization for distribution fault location," *IEEE Transactions on Power Delivery*, vol. 26, no. 2, pp. 1239–1246, April 2011.

[11] N. D. T. Team, "Synchrophasor Monitoring for Distribution Systems: Technical Foundations and Applications," North American Synchrophasor Initiative, Tech. Rep., 2018.

[12] A. von Meier and R. Arghandeh, "Diagnostic Applications for Micro-Synchrophasor Measurements," US Department of Energy, Tech. Rep., 2014.

[13] W. Luan, J. Peng, M. Maras, J. Lo, and B. Harapnuk, "Smart meter data analytics for distribution network connectivity verification," *IEEE Transactions on Smart Grid*, vol. 6, no. 4, pp. 1964–1971, July 2015.

[14] M. Izadi, M. Farajollahi, and A. Safdarian, *Switch Deployment in Distribution Networks*. Singapore: Springer Singapore, 2018, pp. 179–233. [Online]. Available: https://doi.org/10.1007/978-981-10-7001-3-8

[15] S. Kazemi, "Reliability evaluation of smart distribution grids," G4 Monografiaväitöskirja, Aalto University, School of Electrical Engineering, 2011. [Online]. Available: http://urn.fi/URN:ISBN:978-952-60-4241-1

[16] R. Arghandeh and Y. Zhou, *Big Data Application in Power Systems*. Elsevier Science, 2017. [Online]. Available: https://books.google.com/books?id=V-xGDgAAQBAJ

[17] M. Izadi and A. Safdarian, "A mip model for risk constrained switch placement in distribution networks," *IEEE Transactions on Smart Grid*, 2018.

[18] L. M. Konila Sriram, M. Gilanifar, Y. Zhou, E. Erman Ozguven, and R. Arghandeh, "Causal markov elman network for load forecasting in multinetwork systems," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 2, pp. 1434–1442, Feb 2019.

[19] J. J. Q. Yu, Y. Hou, A. Y. S. Lam, and V. O. K. Li, "Intelligent fault detection scheme for microgrids with wavelet-based deep neural networks," *IEEE Transactions on Smart Grid*, pp. 1–1, 2018.

[20] K. S. L. Madhavi, M. Gilanifar, Y. Zhou, E. E. Ozguven, and R. Arghandeh, "Multivariate deep causal network for time series forecasting in interdependent networks," in *2018 IEEE Conference on Decision and Control (CDC)*, Dec 2018, pp. 6476–6481.

[21] S. Ekici, "Classification of power system disturbances using support vector machines," *Expert Systems with Applications*, vol. 36, no. 6, pp. 9859–9868, Aug. 2009.

[22] Y. Chen, L. Xie, and P. R. Kumar, "Power system event classification via dimensionality reduction of synchrophasor data," in *2014 IEEE 8th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, Jun. 2014, pp. 57–60.

[23] A. Jalali, P. Ravikumar, and S. Sanghavi, "A Dirty Model for Multiple Sparse Regression," *IEEE Transactions on Information Theory*, vol. 59, no. 12, pp. 7947–7968, Dec. 2013.

[24] J. Chen, J. Zhou, and J. Ye, "Integrating Low-rank and Group-sparse Structures for Robust Multi-task Learning," in *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '11. New York, NY, USA: ACM, 2011, pp. 42–50.

[25] M. Gilanifar, H. Wang, L. M. Konila Sriram, E. E. Ozguven, and R. Arghandeh, "Multi-task bayesian spatiotemporal gaussian processes for short-term load forecasting," *IEEE Transactions on Industrial Electronics*, pp. 1–1, 2019.

[26] M. Gilanifar, "Heterogeneous data fusion for performance improvement in electric power systems," Ph.D. dissertation, The Florida State University, 2019.

[27] G. Obozinski, B. Taskar, and M. I. Jordan, "Joint covariate selection and joint subspace selection for multiple classification problems," *Statistics and Computing*, vol. 20, no. 2, pp. 231–252, Apr 2010.

[28] T. Jebara, "Multi-task Feature and Kernel Selection for SVMs," in *Proceedings of the Twenty-first International Conference on Machine Learning*, ser. ICML '04. New York, NY, USA: ACM, 2004, pp. 55–.

[29] J. Chen, J. Liu, and J. Ye, "Learning incoherent sparse and low-rank patterns from multiple tasks," *ACM Trans. Knowl. Discov. Data*, vol. 5, no. 4, pp. 22:1–22:31, Feb. 2012. [Online]. Available: http://doi.acm.org/10.1145/2086737.2086742

[30] A. Beck and M. Teboulle, "A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, Jan. 2009.

[31] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course*, ser. Applied Optimization. Springer US, 2004.

[32] M. Stifter, J. Cordova, J. Kazmi, and R. Arghandeh, "Real-Time Simulation and Hardware-in-the-Loop Testbed for Distribution Synchrophasor Applications," *Energies*, vol. 11, no. 4, p. 876, Apr. 2018.

[33] M. Farajollahi, A. Shahsavari, E. M. Stewart, and H. Mohsenian-Rad, "Locating the source of events in power distribution systems using micro-pmu data," *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 6343–6354, Nov 2018.

[34] W. H. Kersting, "Radial distribution test feeders," in *2001 IEEE Power Engineering Society Winter Meeting. Conference Proceedings (Cat. No.01CH37194)*, vol. 2, Jan 2001, pp. 908–912 vol.2.

[35] E. Personal, A. García, A. Parejo, D. F. Larios, F. Biscarri, and C. León, "A comparison of impedance-based fault location methods for power underground distribution systems," *Energies*, vol. 9, no. 12, 2016. [Online]. Available: http://www.mdpi.com/1996-1073/9/12/1022

[36] D. Laverty, R. J. Best, P. Brogan, I. Al Khatib, L. Vanfretti, and D. Morrow, "The openpmu platform for open-source phasor measurements," *Instrumentation and Measurement, IEEE Transactions on*, vol. 62, pp. 701–709, 04 2013.

[37] B. Pinte, M. Quinlan, A. Yoon, K. Reinhard, and P. W. Sauer, "A one-phase, distribution-level phasor measurement unit for post-event analysis," in *2014 Power and Energy Conference at Illinois (PECI)*. IEEE, 2014, pp. 1–7.

Mostafa Gilanifar Mostafa Gilanifar, Ph.D. (S'19) is a postdoctoral research associate in the ECE department at the University of Utah. He received his Ph.D. in Industrial & Manufacturing Engineering from Florida State University in May 2019. His research interests include data analytics, data fusion, and multi-task learning models for heterogeneous spatiotemporal data obtained from smart cities and smart grids.

Jose Cordova Jose Cordova, Ph.D. (M'07) works for the Electrical Power Research Institute (EPRI) as an Engineer/Scientist II. He received his Ph.D. in Electrical Engineering from the Electrical and Computer Engineering Department of FAMU-FSU College of Engineering. He received his master's degree in Electrical Engineering in 2013 from FSU under a Fulbright Scholarship from the U.S. Department of State. His research interests include power quality event detection and data mining for decision support in smart cities and smart grids as well as technical interconnection requirements of DERs in power systems.

Hui Wang Hui Wang received his Ph.D. degree in industrial engineering from the University of South Florida in 2007. He is currently an assistant professor of Industrial and Manufacturing Engineering at Florida State University. His recent research focuses on cloud data fusion to support artificial intelligence (AI) in Internet-of-Things (IoT) applications including inter-connected flexible cybermanufacturing systems, supply chain network for on-site/on-demand production, and smart power systems.

Matthias Stifter Matthias Stifter received his degree from Vienna University of Technology, Austria. From 2003 to 2007 he was with TU Vienna and TU Graz. From 2007 to 2018 he was with AIT, Energy Department - Electric Energy Systems. He is lecturer at the University for Applied Science Technikum Vienna and at TU Vienna. Since 2018 he is Senior Data Scientist for Grid Analytics at Omnetric. His research interests include integration of distributed generation, voltage control and smart meters and data analytics.

Eren E. Ozguven Eren E. Ozguven, Ph.D. (M'11) is an Associate Professor at the Dept. of Civil & Environmental Eng. at Florida A&M University-Florida State University. Dr. Ozguven holds a Ph.D. degree in Civil and Environmental Engineering from the Rutgers University (New Brunswick, NJ, USA) with concentration in emergency supply transportation operations. His research interests include smart cities, urban mobility, traffic safety and reliability, emergency transportation, and intelligent transportation systems.

Thomas Strasser Thomas Strasser (SM'13) received his Ph.D. in mechanical engineering from Vienna University of Technology, Austria, in 2003. He is currently a Senior Scientist with the Center for Energy, the AIT Austrian Institute of Technology. He is active as a Senior Lecturer (Privatdozent) with the Vienna University of Technology. Dr. Strasser was the recipient of the Venia Docendi (habiliation) from the same university in 2017. He is a member of international IEC and IEEE standardization working groups.

Reza Arghandeh Dr. Reza Arghandeh (SM'18) is a Full Professor in the Dept of Computing, Mathematics, and Physics at the Western Norway University of Applied Sciences, Norway. He is also a Senior Data Scientist at StormGeo Company in Bergen, Norway. He has been an Assistant Professor in the Electrical and Computer Engineering Dept at Florida State University, USA during 2015-2018. He completed his Ph.D. in Electrical Engineering at Virginia Tech (2013) and spent two years as a postdoc at the University of California-Berkeley. His research interests include data analysis and decision support for smart grids and smart cities.