Transfer Learning for Smart Grid Data Analytics

Dr Yan Xu
Cham Tao Soon Professor in Engineering
Director, Center for Power Engineering
Nanyang Technological University
Singapore

1 **Background:**

- Smart grid & data resources
- Advanced data-analytics for smart grid

2 **Introduction:**

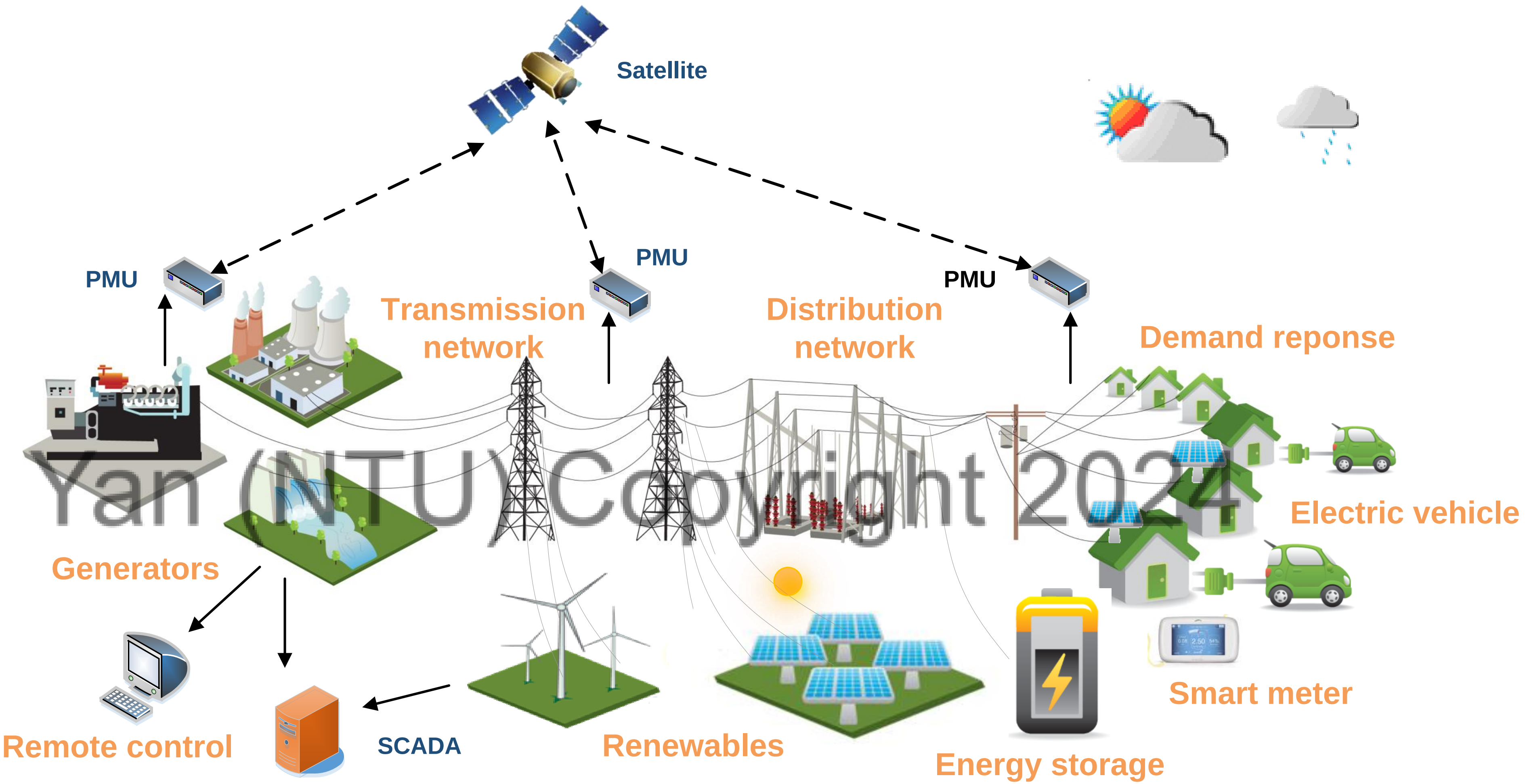
- Conventional machine learning
- Transfer learning

3 **Transfer learning:**

- Power system stability assessment → **from learned faults to unlearned faults**
- Power converter fault diagnosis → **from learned converters to unlearned converters**
- Masked-load forecasting → **from original load to masked-load by distributed energy resources (DERs)**

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

What is a "Smart Grid"?



Smart Grid: a modernized power grid with high-level renewables, more distributed energy resources, and wide-spread deployments of advanced ICT

NetZero & Carbon Neutrality

Data Resources in Smart Grid

Wide-spread deployments of advanced ICT can provide more data and information about the power system at different levels

Grid Monitoring System
(Phasor measurement unit (PMU), SCADA, etc.)

Customer Meters
(Residential smart meter, Industrial metersetc.)

Asset Sensors
(PQ sensor, battery management system, PD sensors, etc.)

Source of figures: website (searched in Google)

Illustration of Grid Data

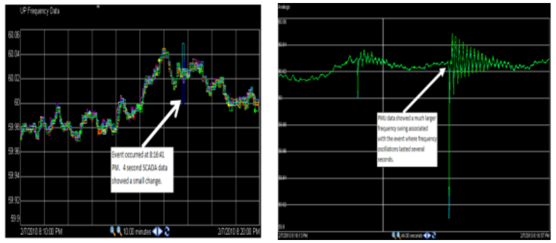


Illustration of Customer Data

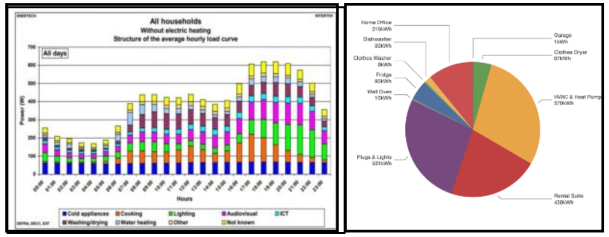
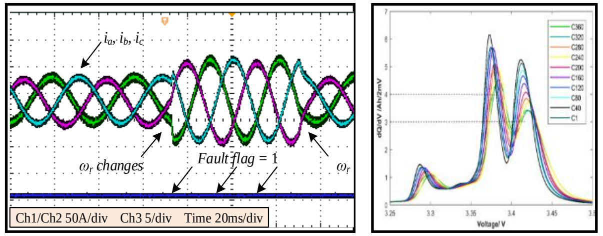


Illustration of Asset Data



How to make use of these data to support power system's monitoring, operation & control ?

Advanced AI and Data-Analytics for Smart Grid

NTU SODA Group's Research Works

Feature selection & extraction | Supervised learning | Unsupervised learning | Reinforcement learning

Research Problems

Grid Level
- Data-driven power system stability analysis
  - Assessment
  - Control
- Data-driven power system control & operation
  - LFC
  - OPF

Customer Level
- Non-instructive load monitoring (NILM)
- Data-driven home energy management (HEM)
- Residential Load Forecasting

Asset Level
- Power converter system fault diagnosis
- Li-ion battery health monitoring
- Asset Risk Assessment

Funding sources

NATIONAL RESEARCH FOUNDATION
PRIME MINISTER'S OFFICE
SINGAPORE
Research . Innovation . Enterprise

Ministry of Education
SINGAPORE

Singtel

LITEON

jtc
Breaking New Ground

ROLLS ROYCE

infineon

SPgroup
新加坡能源集团

Motivations for transfer learning

Conventional Machine Learning for Smart Grid Applications

Offline training

Training Database → Input / Output → **Machine learning algorithm**
- Classification/Regression
- SVM/DNN/ELM/DT…

Training

Online application

Real-time Measurement → Input → Machine learning model → output

Assumptions
- Training and application data are in the **same feature space**
- Training and application data follow the **same distribution**

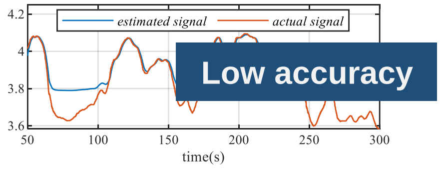
However, in practice, such assumptions may not always hold

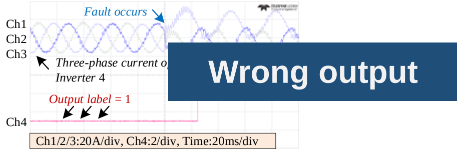
System A (Training)

System B (Apply)

Machine learning model

- The application **target may be different**
- The application target may have **a limited amount of training data**
- The target data may have a **different distribution**





Machine learning model may not work in practice

Transfer Learning – Notations and Definitions

- ***Domain*** : $D=\{X, P(X)\}$, where X is the feature space $X=\{x_1, x_2, \ldots, x_n\}$ and $P(X)$ is the marginal probability distribution.

- ***Task*** : $T=\{Y, f(\cdot)\}$), where Y is the label space and $f(\cdot)$ is the objective predictive function, which is not observed but can be learned from the training data, and consists of pairs (x_i, y_i), where $x_i \in X$ and $y_i \in Y$. The function $f(\cdot)$ can predict the corresponding label, $f(x)$, of a new instance x.

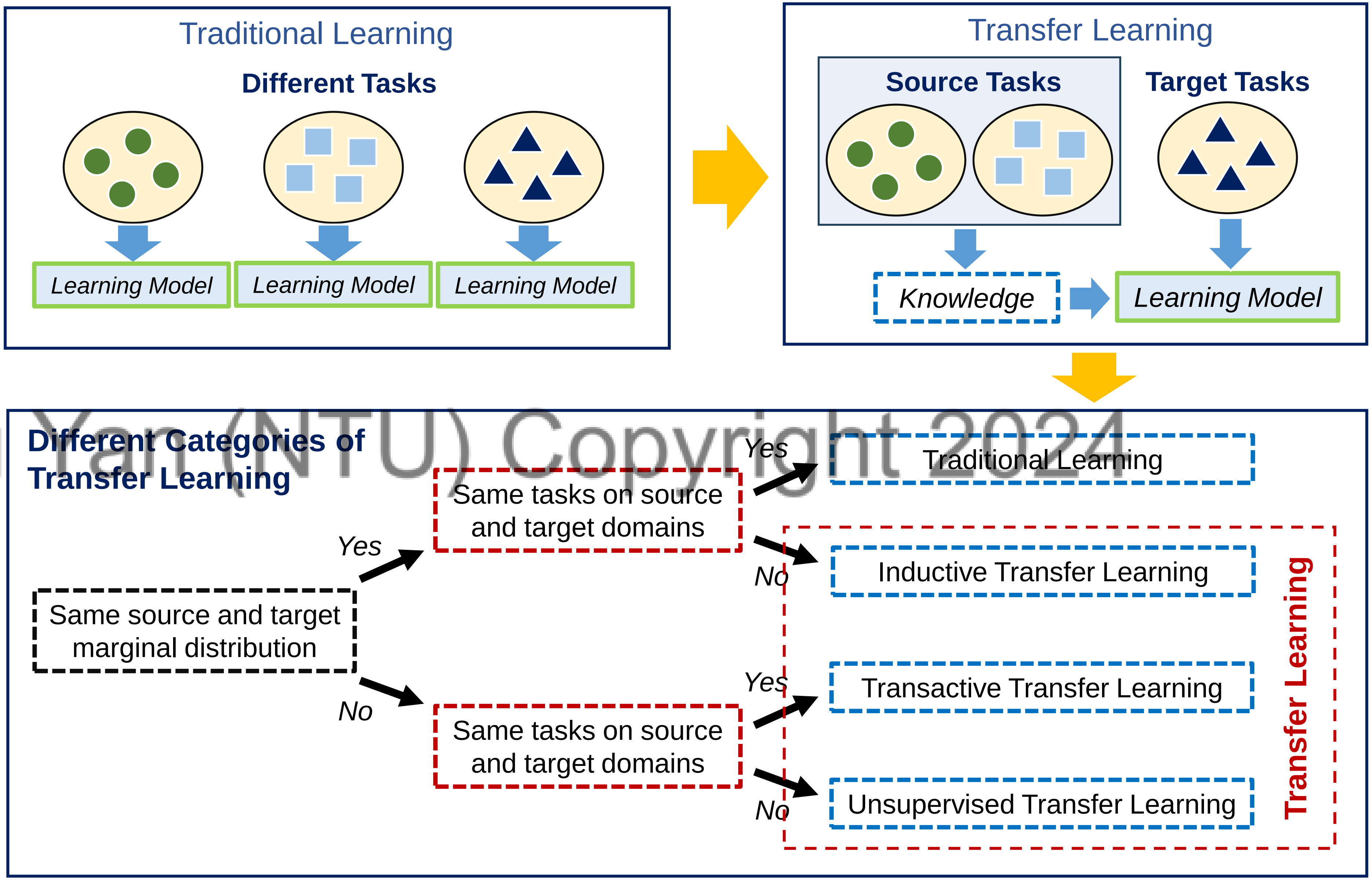
- ***Source domain*** : $D_s = \{(x_{si}, y_{si})\}$, $(i = 1, 2, \ldots, n)$, where $x_{si} \in X_S$ and $y_{si} \in Y_S$
- ***Target domain*** : $D_t = \{(x_{ti}, y_{ti})\}$, $(i = 1, 2, \ldots, m)$, where $x_{ti} \in X_T$ and $y_{ti} \in Y_T$ (in most cases, $m<<n$)

- ***Transfer Learning (TL)*** : Given a source domain D_s and its learning task T_s , a target domain D_t and its learning task T_T, TL aims to help improve the learning of the target predictive function $f_t(\cdot)$ in D_t using the knowledge in D_s and T_s , where $D_s \neq D_t$, or $T_s \neq T_T$.
- $D_s \neq D_t$ implies that either the features are different between the two datasets $X_S \neq X_T$, or their marginal distributions are different $P_S(X) \neq P_T(X)$
- $T_s \neq T_T$ implies that $Y_S \neq Y_T$ or $P(Y_S|X_S) \neq P(Y_T|X_T)$

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

S. J. Pan and Q. Yang, "A survey on transfer learning," IEEE Transactions on Knowledge and Data Engineering, 2009.

Transfer Learning – Illustration and Different Categories



S. J. Pan and Q. Yang, “A survey on transfer learning,” IEEE Transactions on Knowledge and Data Engineering, 2009.

Transfer Learning – Strategies

A. Instance-based Transfer Learning

- Principle: select or reweight instance that are close to the target (**instance weighting**)
- Advantage/Suitability: simple implementation, domains differ only in marginal distributions
- Limitations: weight is unknown and difficult to be obtained

B. Feature-based Transfer Learning

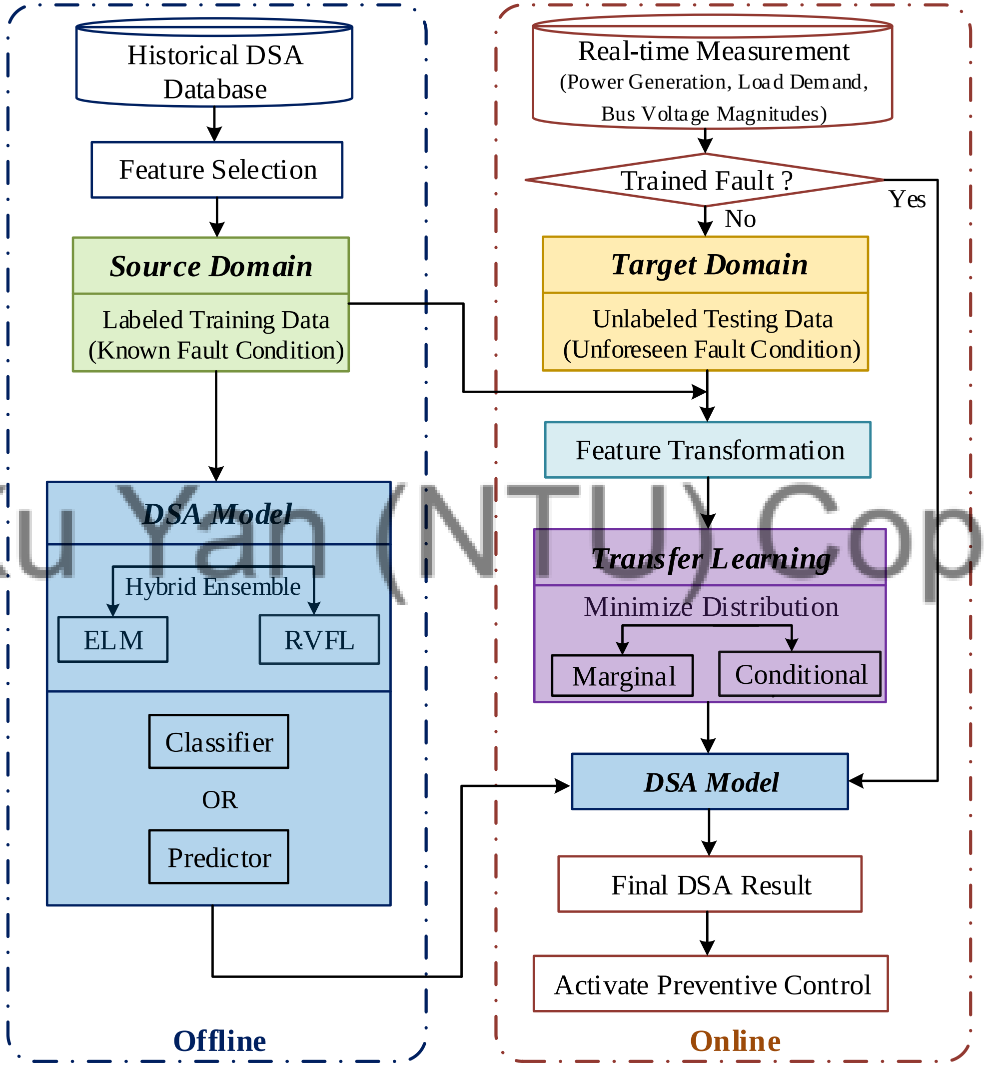
- Principle: find the common latent features and use them as a bridge to transfer knowledge (**feature transformation**), minimize marginal/conditional distribution difference preserve properties or potential structures of the data, find correspondence between features
- Advantage/Suitability : transfer between different feature spaces
- Limitations: negative transfer

C. Parameter-based Transfer Learning

- Principle: individual models for related tasks **share some parameters or prior distributions** of hyperparameters
- Advantage/Suitability: same task, same distribution of source and target domain
- Limitation: transferred parameter selection

F. Zhuang et al., “A comprehensive survey on transfer learning,” Proc. IEEE, vol. 109, no. 1, pp. 43–76, Jan. 2021.

[1] Transfer Learning for Power System Stability Assessment with Unlearned Faults



Problem descriptions:

- For pre-fault SA, one model is trained for one fault.
- Only a limit number of faults are considered.
- For practical application, unknown potential faults may happen.
- How to use one model to assess many different potential unknown faults?

Transfer learning:

- SA model is a classifier based on hybrid ensemble model.
- *RELIEF-F* algorithm is used to select the critical features.
- **Feature transformation with an adaptation matrix** via minimizing marginal and conditional distribution differences (**MMD**) between the unknown features and the known features.

$$\boldsymbol{MMD}^2(D_s, D_t) = \left\| \frac{1}{n}\sum_{i=1}^{n} f(x_{si}) - \frac{1}{m}\sum_{i=1}^{m} f(x_{ti}) \right\|^2 + \sum_{c=1}^{22} \left\| \frac{1}{n_c}\sum_{x_{si}\in D_s^c} f(x_{si}) - \frac{1}{m_c}\sum_{x_{ti}\in D_t^c} f(x_{ti}) \right\|^2$$

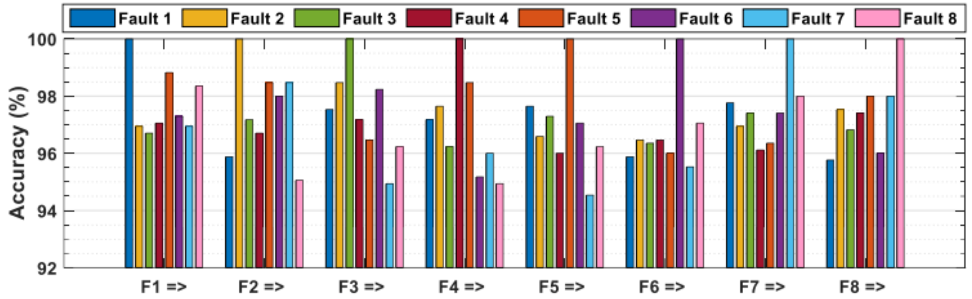
Byproduct:

- Using one model to assess many potential faults.
- The correlation between different faults can be revealed, thus different faults can be aggregated as one.

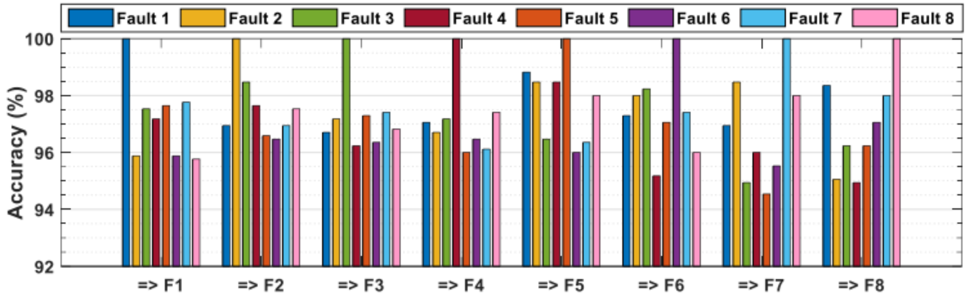
C. Ren and Y. Xu "Transfer Learning-Based Power System Online Dynamic Security Assessment: Using One Model to Assess Many Unlearned Faults," *IEEE Transactions on Power Systems*, 2019.

[1] Transfer Learning for Power System Stability Assessment with Unlearned Faults

Online Testing Results



(a) each of fault is transferred to the remaining 7 faults

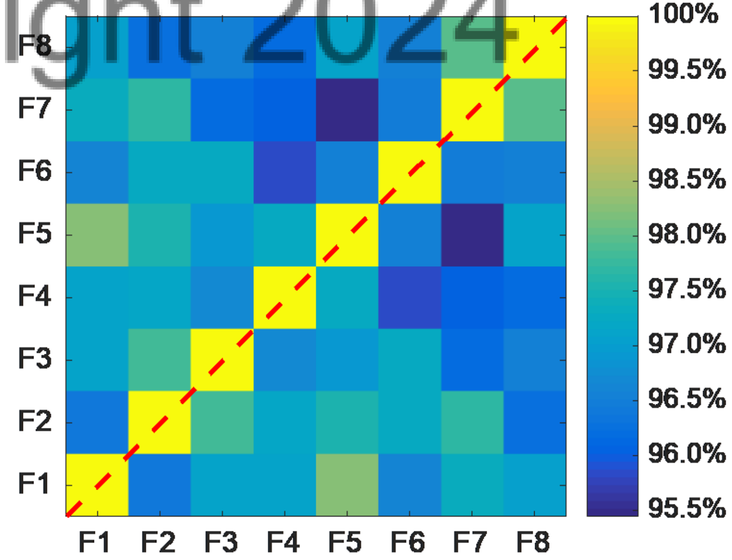


(b) each of 7 faults is transferred to the remaining 1 fault

Different structure of the incremental broad learning for (a) Increment of enhancement hidden nodes, (b) Increment of features, (c) Increment of enhancement hidden nodes, features, and new training instances

AVERAGE ACCURACY OF DIFFERENT METHODS

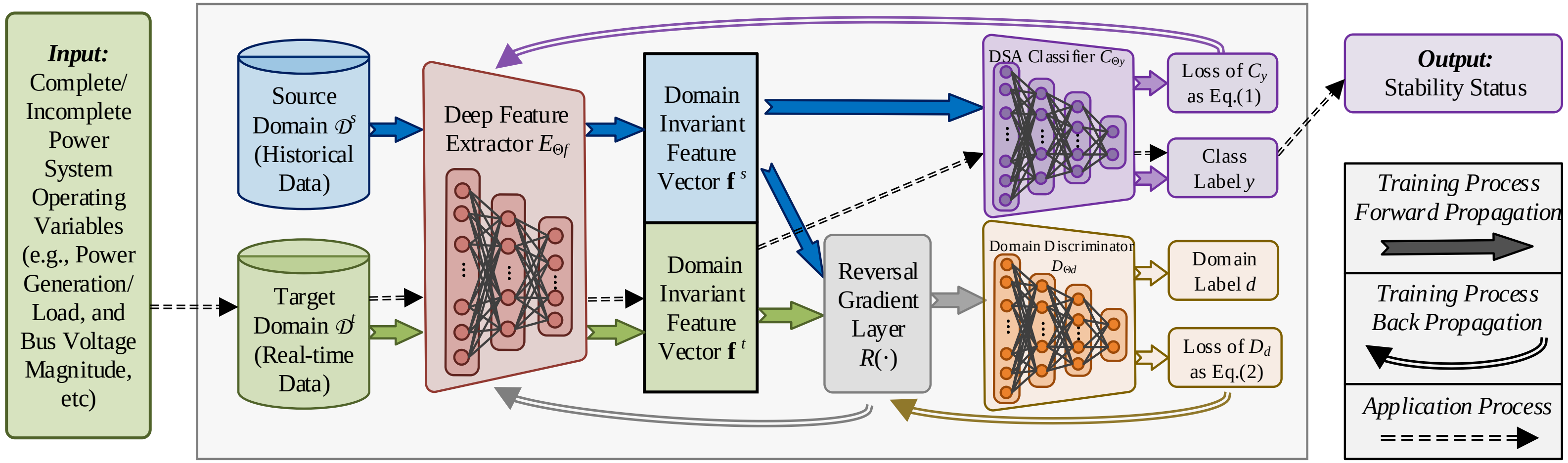
Method	Average Accuracy
Original DSA Model without Transfer Learning	82.25%
Proposed method	97.27%



Mutual Transfer Accuracy Matrix

C. Ren and Y. Xu "Transfer Learning-Based Power System Online Dynamic Security Assessment: Using One Model to Assess Many Unlearned Faults," *IEEE Transactions on Power Systems*, 2019.

[2] Transfer Learning for Power System Stability Assessment with Unlearned Faults and Missing Data



Proposed adversarial training framework for knowledge transfer

Problem description:

- In practice, the **unlearned fault with incomplete data** may occur at the same time. Under this scenario, the above **feature-based TL method and GAN-based method** will be ineffective, since the incomplete data inputs are from the unlearned faults.

Principle of adversarial training:

- **1) Feature learning** to extract the common impact features of two domains in one feature space from the input data for different faults, named **domain-invariant features;** 2) by **fooling the domain discriminator** with such features, the distribution of source domain and target domain becomes more similar; 3) the **SA classifier trained by source domain can be used** for unlabeled instances in target domain.
- The feature learning stage can also extract the domain-invariant features by **incomplete target domain** data, hence the proposed method can also accurately work with missing data.

C. Ren and Y. Xu, "An Integrated Transfer Learning Method for Power System Dynamic Security Assessment for Unlearned Faults with Missing Data," *IEEE Trans. Power Syst.*, 2021.

[2] Transfer Learning for Power System Stability Assessment with Unlearned Faults and Missing Data

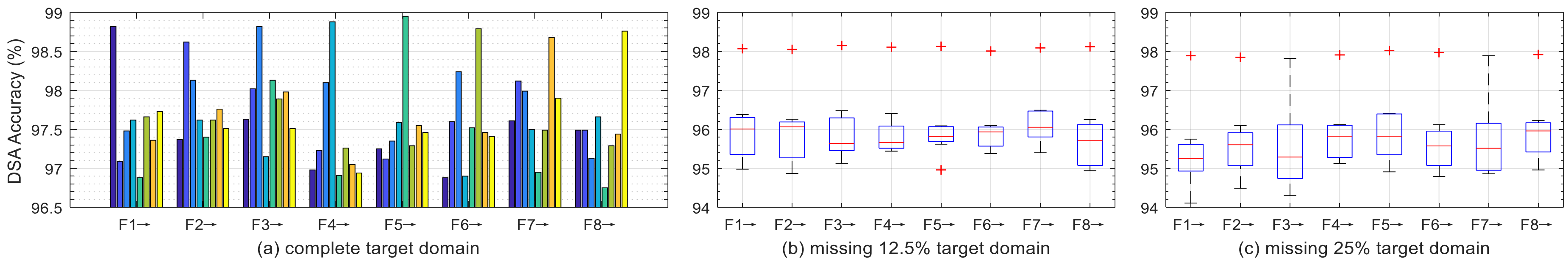


Figure. Testing results of the proposed TL method. (a) complete target domain data; (b) 12.5% missing target domain data; (c) 25% missing target domain data.

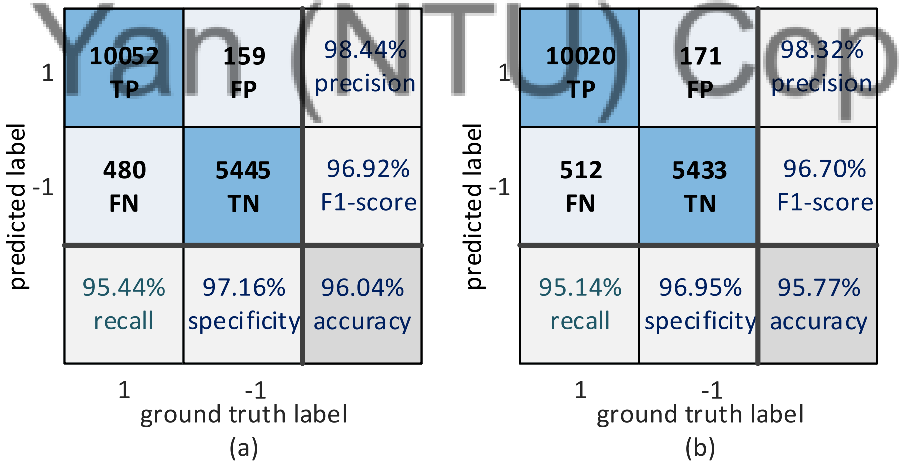


Figure. Confusion Matrix of (a) 12.5% missing target domain data; (b) 25% missing target domain data.

TABLE I
AVERAGE DSA PERFORMANCE OF DIFFERENT METHODS.

Method		Average DSA performance on target domain with the complete data			
		accuracy	precision	specificity	F1-score
Without TL	Ensemble learning	82.25%	90.59%	84.14%	85.66%
	DT	78.84%	87.37%	78.53%	82.98%
	LSTM	85.76%	92.03%	86.06%	88.70%
TL-based	Ref [2]	97.27%	98.81%	97.81%	97.89%
	Proposed method	97.68%	99.18%	98.48%	98.21%

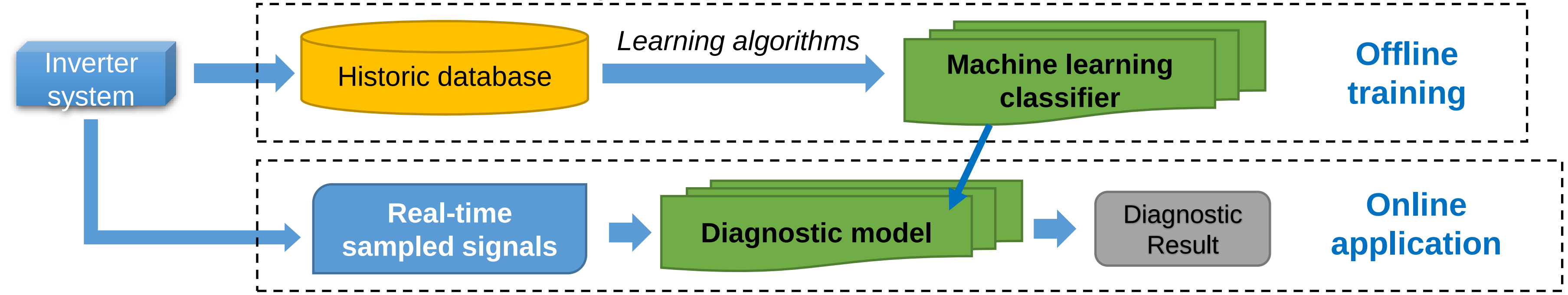
TABLE II
AVERAGE DSA ACCURACY OF DIFFERENT METHODS.

Method		Average DSA accuracy on target domain with the incomplete data	
		12.5% missing data	25% missing data
Without TL (e.g., DT, LSTM)		×	×
TL-based	Ref. [2]	×	×
	Proposed method	96.04%	95.77%

C. Ren and Y. Xu, "An Integrated Transfer Learning Method for Power System Dynamic Security Assessment for Unlearned Faults with Missing Data," *IEEE Trans. Power Syst.*, 2021.

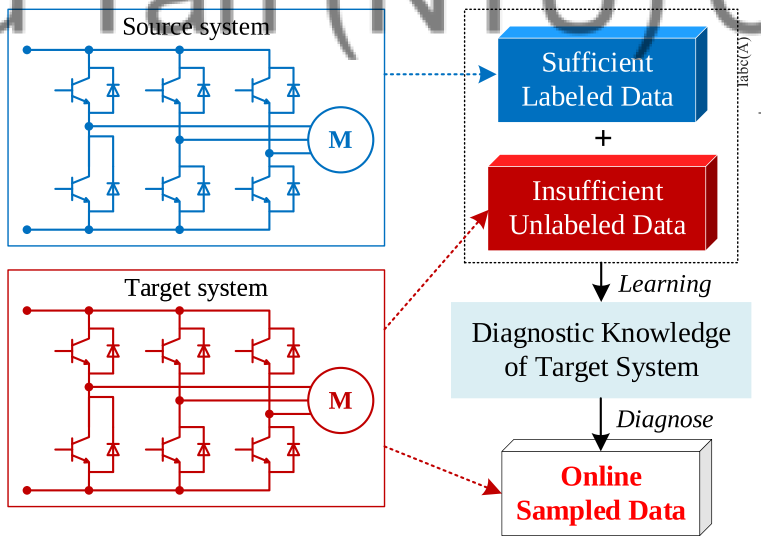
[3] Transfer Learning for Open-Circuit Fault Diagnosis of Different Inverter Systems

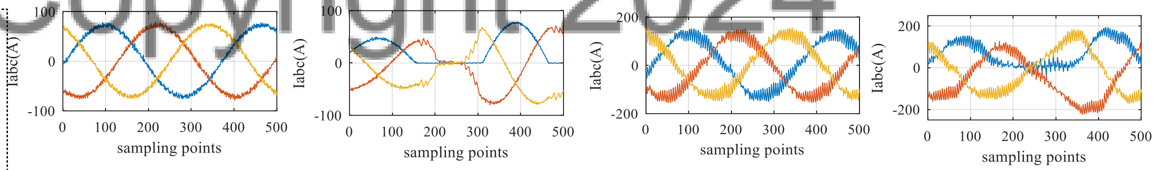
Conventional data-driven inverter fault diagnosis process



- ❑ The diagnostic model is trained by a fault database of **a specific inverter system**
- ❑ Only work for the corresponding system but not for **an unlearned inverter system**

Transferrable fault diagnosis (Adapt one trained diagnostic model for different inverter systems)





System A

labeled source dataset
$D_s = \{(x_{si}, y_{si})\}$

System B

unlabeled target dataset
$D_t = \{x_{tj}\}$

Extract diagnostic knowledge of the target system
- ❑ x_s and x_t, y_s and y_t belong to the same space
- ❑ The marginal, conditional distribution of source, target data are different, i.e. $P_s(x_s) \neq P_t(x_t)$, $P_s(x_s|y_s) \neq P_t(x_t|y_t)$

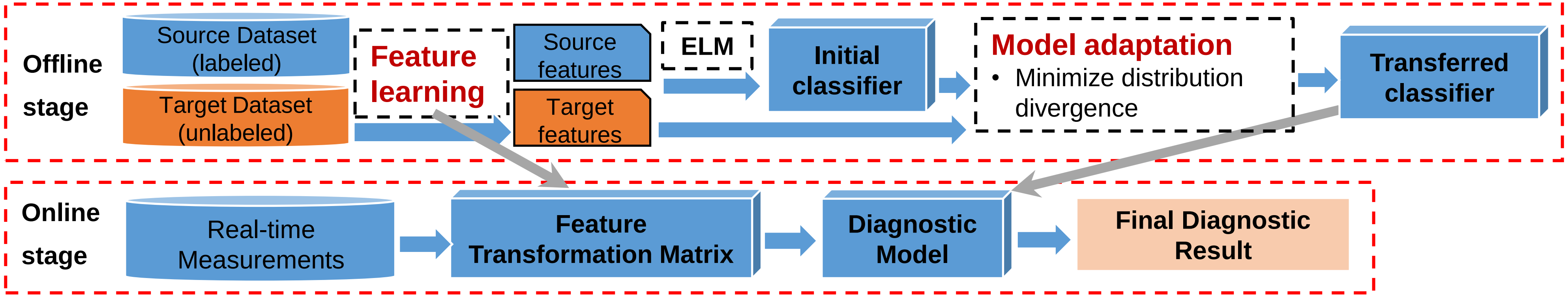
Y. Xia and Y. Xu, "A Transferrable Data-Driven Method for IGBT Open-Circuit Fault Diagnosis in Three-Phase Inverters," *IEEE Trans. Power Electron.*, 2021.

1. Background

2. Introduction

3. Transfer
- Power System Stability Assessment
- **Converter System Fault Diagnosis**
- Masked-Load Forecasting

[3] Transfer Learning for Open-Circuit Fault Diagnosis of Different Inverter Systems



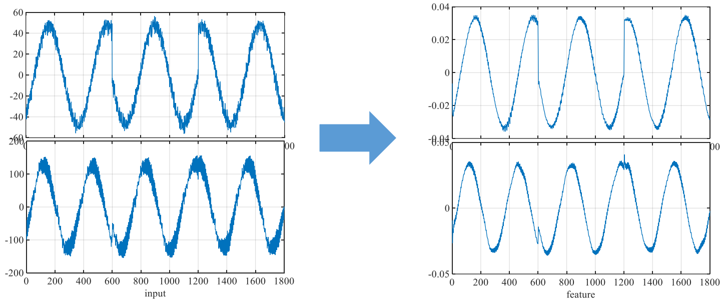
1. Feature Transformation

➢ Pre-process the training and testing data by integrating them into **a common feature space:**

1) Transform original data into **PCA** subspace

2) **Geodesic Flow Kernel** (GFK) is used to extract the path between the two subspaces

$$\mathbf{H}=\begin{bmatrix} X_{ps}U_1 & R_{ps}U_2 \end{bmatrix}\begin{bmatrix} Z_1 & Z_2 \\ Z_2 & Z_3 \end{bmatrix}\begin{bmatrix} U_1^T X_{ps}^T \\ U_2^T R_{ps}^T \end{bmatrix} \quad q=g(x)=\sqrt{\mathbf{H}}x$$



Original data **Extracted features**

2. Model Adaptation

➢ Minimize the distribution divergence between source and target data:

1) An initial diagnostic model is trained by **ELM**

$$f(q)=\beta\cdot\varphi(\alpha\cdot q+b)=\beta\cdot \boldsymbol{P} \quad \text{(update } \beta\text{)}$$

2) **MMD** (**evaluate the distribution divergence between two systems**)

$$\boldsymbol{D}^2(Q_s,Q_t)=\left\|\frac{1}{n}\sum_{i=1}^{n}f(q_{si})-\frac{1}{m}\sum_{i=1}^{m}f(q_{ti})\right\|^2+\sum_{c=1}^{22}\left\|\frac{1}{n_c}\sum_{q_{si}\in Q_s^c}f(q_{si})-\frac{1}{m_c}\sum_{q_{ti}\in Q_t^c}f(q_{ti})\right\|^2$$

3) Optimize β (output weight)

$$\theta=\arg\min_{\beta}(\boldsymbol{J}+\boldsymbol{D}^2(Q_s,Q_t)+\boldsymbol{R}(Q_s,Q_t))$$

optimization problem θ

$$=\arg\min_{\beta}\left(\left\|(\boldsymbol{Y}-\beta^T\boldsymbol{P})\boldsymbol{E}\right\|^2+tr(\beta^T\boldsymbol{PMP}\beta)+tr(\beta^T\boldsymbol{PLP}\beta)\right)$$

Training loss function | *MMD minimization* | *Laplacian regularization (improve manifold feature)*

∂θ/∂β=0 ➡ $\beta=\left((\boldsymbol{E}+\lambda\boldsymbol{M}+\eta\boldsymbol{L})\boldsymbol{P}\right)^{-1}\cdot\boldsymbol{EY}^T$

Y. Xia and Y. Xu, "A Transferrable Data-Driven Method for IGBT Open-Circuit Fault Diagnosis in Three-Phase Inverters," *IEEE Trans. Power Electron.*, 2021.

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

[3] Transfer Learning for Open-Circuit Fault Diagnosis of Different Inverter Systems

Offline Test Results

- In practical application, a comprehensive target dataset may not be available
- A minibatch as target dataset is selected covering one or several fault labels

Average Test Accuracy Performance under Different Target and Test Datasets

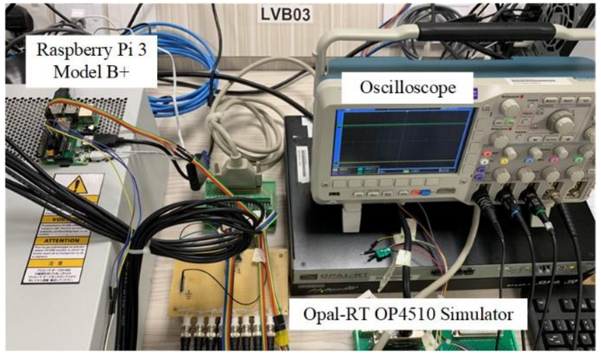
System A	System B		Average accuracy
Source Dataset	Target Dataset	Test Dataset	
Labeled 6600 samples with 22 labels	No transfer learning	{Label=1~22}	62.79 %
	60% of {Label=1}	40% of {Label=1}, {Label=2~22}	82.82 %
	60% of {Label=1, 2}	40% of {Label=1,2}, {Label=3~22}	83.36 %
	60% of {Label=1~7}	40% of {Label=1~7}, {Label=8~22}	87.05 %
	60% of {Label=1, 2, 5, 11, 19}	40% of {Label=1, 2, 5, 11, 19}, {Label=3~4, 6~10, 12~18, 20~22}	87.06 %
	60% of {Label=1~13, 16~20}	40% of {Label=1~13, 16~20}, {Label=14~15, 21~22}	91.24 %
	60% of {Label=1~22}	40% of {Label=1~22}	96.76 %

Comparison with Other Intelligent Algorithms

Methodology	Average accuracy	
	Without normalization	With normalization
KNN	46.73 %	55.58 %
ELM	55.70 %	62.79 %
BN	24.12 %	64.79 %
DT	43.97 %	52.07 %
FFT + PCA + BN	46.36 %	
FFT + ReliefF + ELM/RVFL ensemble	64.42 %	
FFT + ReliefF + RVFL ensemble	51.45 %	
Proposed method	**88.53 %**	

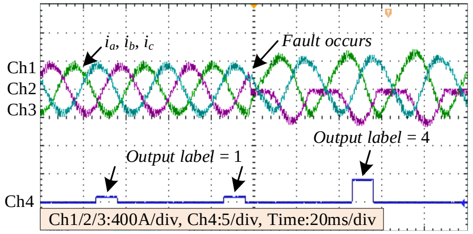
Average accuracy for randomly selected target and test datasets

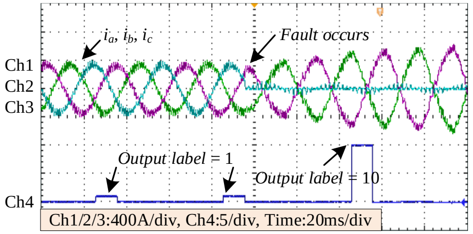
Real-Time Results



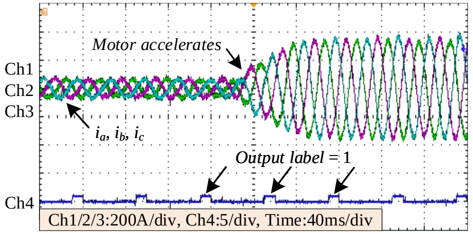
Experimental platform

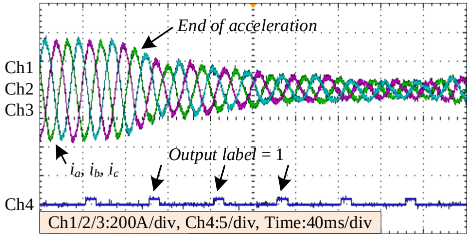
Diagnostic results





Under Different speeds





Y. Xia and Y. Xu, “A Transferrable Data-Driven Method for IGBT Open-Circuit Fault Diagnosis in Three-Phase Inverters,” *IEEE Trans. Power Electron.*, 2021.

[4] Transfer Learning for Fault Diagnosis of Multiple Inverters in a Noisy Microgrid

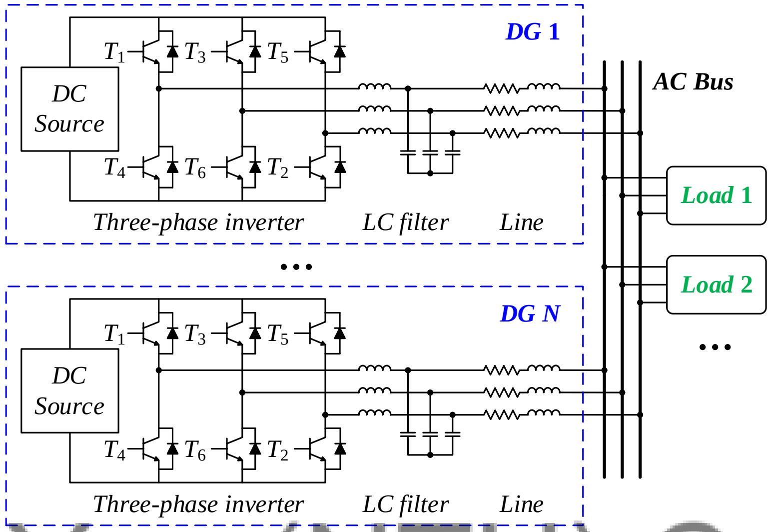


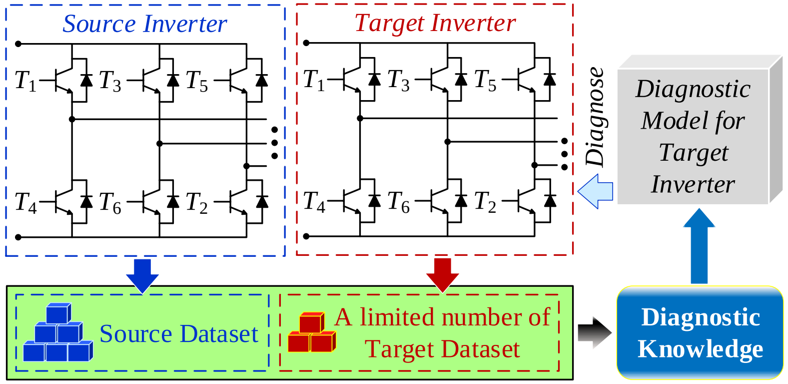
TABLE 1.1 Fault Labels of Faulty Inverter Localization

Faulty Inverter	Label	Faulty Inverter	Label
No fault	1	…	…
DG 1	2	DG $N-1$	N
DG 2	3	DG N	$N+1$

TABLE 1.2 Fault Labels of Switch Fault Classification

Faulty Switch	Label	Faulty Switch	Label
No fault	1	T_4	5
T_1	2	T_5	6
T_2	3	T_6	7
T_3	4		

- ➢ Two-level multi-classification problem: locating the faulty inverters and diagnose the faults
- ➢ Each fault label indicates a specific status of fault condition (faulty inverter, faulty switch)



One trained classifier may only work in a specific inverter

>>> transferability of data-driven models

labeled source dataset $D_s = \{(x_{Si}, y_{Si})\}$ $(i = 1, 2, \ldots, N_S)$

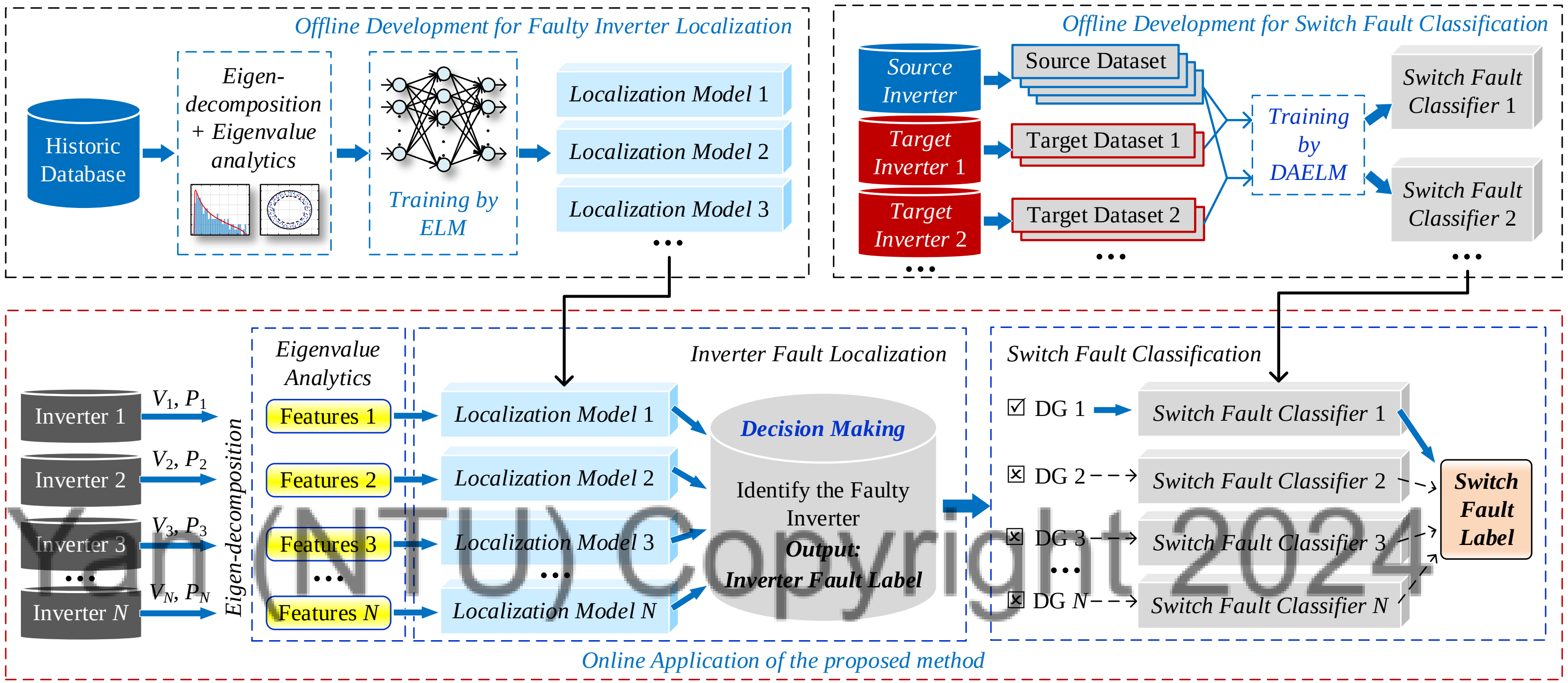
labeled target dataset $D_T = \{(x_{Tj}, y_{Tj})\}$ $(j = 1, 2, \ldots, N_T)$

$N_S >> N_T$ (limited number of **labeled target data**)

Diagnostic knowledge of the target inverter

Y. Xia, Y. Xu, and N. Zhou, "A Transferrable and Noise-Tolerant Data-Driven Method for Inverter Open-Circuit Fault Diagnosis in Microgrids," *IEEE Trans. Ind. Electron.*, 2023.

[4] Transfer Learning for Fault Diagnosis of Multiple Inverters in a Noisy Microgrid



an ELM-based initial classifier is trained ($D_s = \{(x_{Si}, y_{Si})\}$)

$f(\boldsymbol{\rho}) = \beta \cdot \varphi(\alpha\boldsymbol{\rho} + b) = \beta \mathrm{H}$ **>>> Optimize β by DAELM**

Model Adaptation Objective:

$$\min_{\beta} \frac{1}{2}\left(\|\beta\|^2 + r_S \cdot \sum_{i=1}^{N_S}\|\varepsilon_S^i\|^2 + r_T \cdot \sum_{j=1}^{N_T}\|\varepsilon_T^j\|^2\right) \qquad \begin{aligned} \varepsilon_S^i &= t_S^i - \beta \cdot \mathbf{H}_S^i,\ i = 1,2,\ldots,N_S \\ \varepsilon_T^j &= t_T^j - \beta \cdot \mathbf{H}_T^j,\ j = 1,2,\ldots,N_T \end{aligned}$$

To find β fit in both source and target domain.

$$L\left(\beta, \varepsilon_S^i, \varepsilon_T^j, \lambda_S, \lambda_T\right) = \frac{1}{2}\left(\|\beta\|^2 + r_S \cdot \sum_{i=1}^{N_S}\|\varepsilon_S^i\|^2 + r_T \cdot \sum_{j=1}^{N_T}\|\varepsilon_T^j\|^2\right) - \gamma_S\left(\beta \cdot \mathbf{H}_S^i - t_S^i + \varepsilon_S^i\right) - \gamma_T\left(\beta \cdot \mathbf{H}_T^j - t_T^j + \varepsilon_T^j\right)$$

Loss function

To solve, the partial derivatives $\partial L/\partial\beta$, $\partial L/\partial\varepsilon_S$, $\partial L/\partial\varepsilon_T$, $\partial L/\partial\gamma_S$, $\partial L/\partial\gamma_T$ are set as zero:

$$\begin{cases} \partial L / \partial\beta = 0 \rightarrow \beta = \gamma_S \mathbf{H}_S^{\mathrm{T}} + \gamma_S \mathbf{H}_T^{\mathrm{T}} \\ \partial L / \partial\varepsilon_S = 0 \rightarrow \gamma_S = r_S \cdot \varepsilon_S^{\mathrm{T}} \\ \partial L / \partial\varepsilon_T = 0 \rightarrow \gamma_T = r_T \cdot \varepsilon_T^{\mathrm{T}} \\ \partial L / \partial\gamma_S = 0 \rightarrow \beta \cdot \mathbf{H}_S - t_S + \varepsilon_S = 0 \\ \partial L / \partial\gamma_T = 0 \rightarrow \beta \cdot \mathbf{H}_T - t_T + \varepsilon_T = 0 \end{cases}$$

Solution: $\gamma_S = \left(\beta \cdot \mathbf{H}_S - \gamma_T \cdot \mathbf{H}_T^{\mathrm{T}} \cdot \mathbf{H}_S\right)\left(\mathbf{H}_S^{\mathrm{T}} \cdot \mathbf{H}_S\right)^{-1}$

$$\beta = \left(r_S \cdot t_S \cdot \mathbf{H}_S^T + r_T \cdot \mathbf{H}_T \cdot \mathbf{H}_T^T\right)\left(\mathbf{I} + r_S \cdot \mathbf{H}_S \cdot \mathbf{H}_S^T + r_T \cdot \mathbf{H}_T \cdot \mathbf{H}_T^T\right)^{-1}$$

Y. Xia, Y. Xu, and N. Zhou, "A Transferrable and Noise-Tolerant Data-Driven Method for Inverter Open-Circuit Fault Diagnosis in Microgrids," *IEEE Trans. Ind. Electron.*, 2023.

[4] Transfer Learning for Fault Diagnosis of Multiple Inverters in a Noisy Microgrid

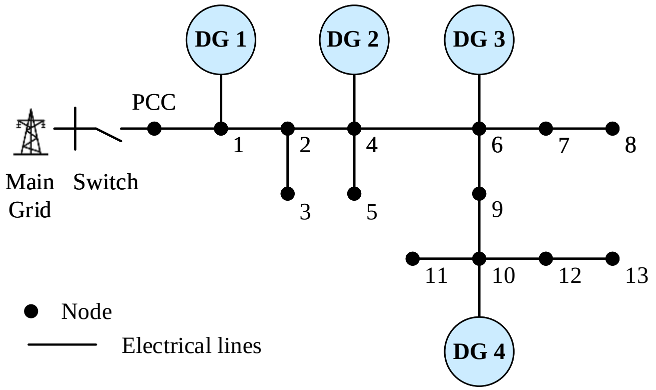


Figure. The topology of the 13-bus microgrid system.

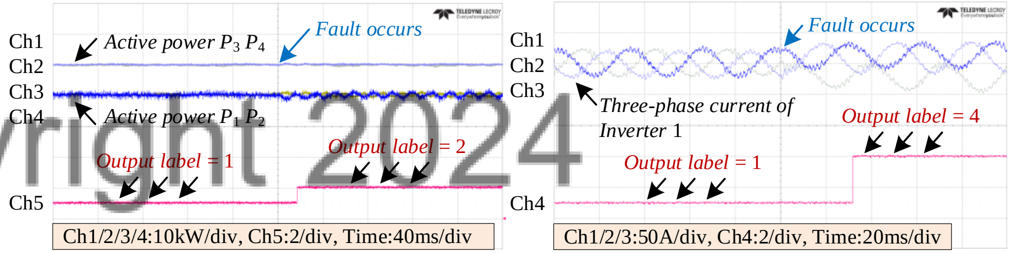
Figure. Real-time experimental test based on Opal-RT.

TABLE Test Accuracy Performance Under Different Target Datasets

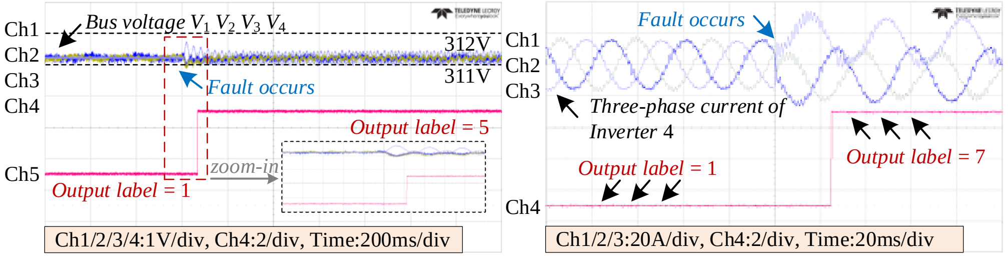
Inverter 1 Source Dataset	Inverter 3 Target Dataset	Inverter 3 Test Dataset	Testing accuracy
2100 instances with 7 labels	NA (without Transfer Learning)	1400 instances with 7 labels	60.07 %
	80 instances of {Label=1}		76.25 %
	30 instances of {Label=1}, 30 instances of {Label=2}.		79.81 %
	30 instances of {Label=1}, 20 instances of {Label=3}, 20 instances of {Label=5}.		92.04 %
	10 instances of {Label=1} ~ {Label=7}		99.64 %

TABLE Comparison With Other Intelligent Algorithms

Methodology	Average accuracy
ELM	70.41 %
DT	87.29 %
ELM ensemble	86.21 %
ReliefF + ELM/RVFL ensemble	89.71 %
PCA + BN	88.07 %
Proposed transferrable method	**94.43 %**



Real-time experimental results when inverter 1 is under T_3 open -circuit fault (a) faulty inverter localization (b) switch fault classification.



Real-time experimental results when inverter 4 is under T_6 open -circuit fault (a) faulty inverter localization (b) switch fault classification.

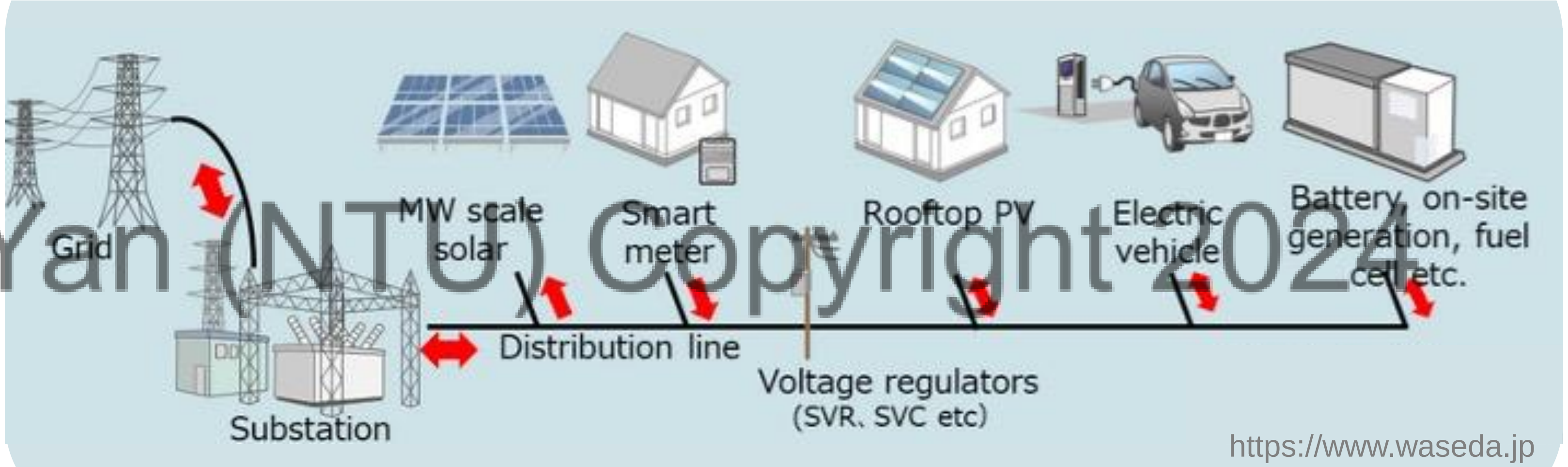
Y. Xia, Y. Xu, and N. Zhou, “A Transferrable and Noise-Tolerant Data-Driven Method for Inverter Open-Circuit Fault Diagnosis in Microgrids,” *IEEE Trans. Ind. Electron.*, 2023.

[5] Transfer Learning for Forecasting Masked-Load due to Behind-the-Meter DERs

Load demand is being masked by distributed energy resources (DERs)

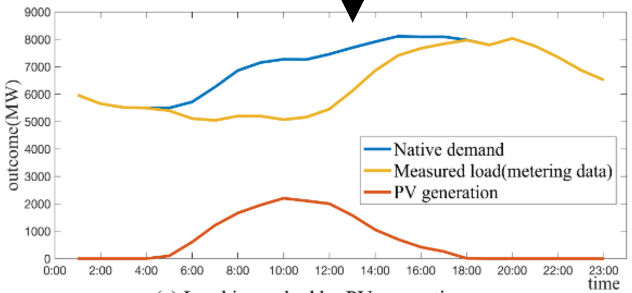
- **Distributed generation (DG):** rooftop PV, small wind turbine...
- **Energy storage system (ESS):** residential batteries, UPS...
- **Flexible loads:** Electric vehicles (EV), smart appliances...

- Behind-the-meter (BTW) installation
- Continuous growing with less visibility

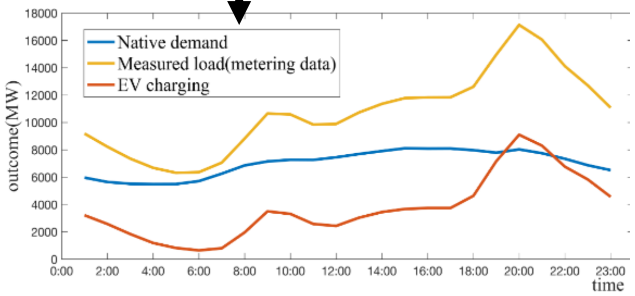


https://www.waseda.jp

Measured load is no longer merely native demand, as it contains DERs



(a) Load is masked by PV generation:



(b) Load is masked by EV charging:

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

[5] Transfer Learning for Forecasting Masked-Load due to Behind-the-Meter DERs

Problem descriptions:
- Residential load is masked by a **mix of different distributed energy resources (DERs)**;
- DERs are installed behind the meter, thus **information about them are not available**;
- Only available datasets for supervised learning is **historical unmasked load** and **present masked load**.

- **Available datasets:**

Historical unmasked load → Transfer → Present masked load

$x_T = x_S$ + DERs

$$D_S = \{(x_S^i, y_S^i)\}_{i=1}^{n_S}$$

- Large dataset
- sufficient to train a model

$$D_T = \{(x_T^i)\}_{i=1}^{n_T}$$

- Small dataset, unsupervised
- Since DERs in frequent developments

Note that D_S and D_T has different but related relationship: x_S is latent in x_T, since $x_T = x_S$ + DERs

- **Objective**

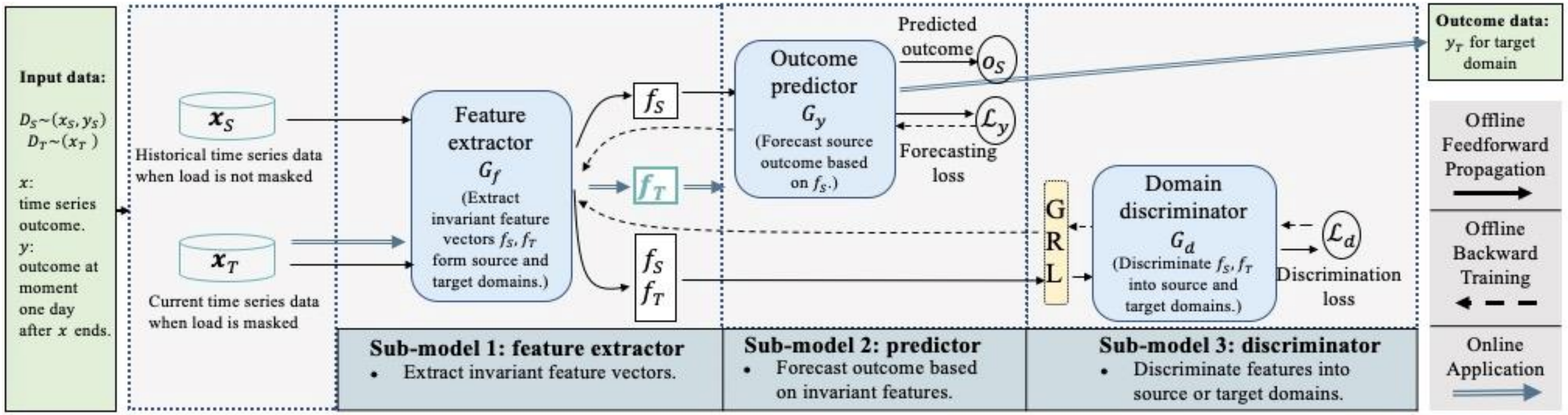
Find an input-output pattern g(·) for masked load forecasting:

$$y_T = g(x_T)$$

with knowledge in D_S and D_T.

(Transfer learning)

[5] Transfer Learning for Forecasting Masked-Load due to Behind-the-Meter DERs



Proposed Framework: Domain Adaptation Neural Network (DANN)

Firstly, feature vectors f_S and f_T are extracted from D_S and D_T. Then, a domain discriminator will discriminate f_S and f_T into D_S or D_T. Intuitively, when a discriminator fails to distinguish f_S and f_T from each other, that means f_S and f_T are in the similar distribution. Based on this idea, a feature extractor is trained against the discriminator, aiming to fool the discriminator. As a result, f_S and f_T has similar data distribution, and f_T could be compatible to a forecasting model which is trained with f_S.

Offline training: Backward training
- Update parameters as:
- $G_y' = G_y - \lambda_y \frac{\partial \mathcal{L}_y}{\partial G_y}$
- $G_d' = G_d - \lambda_d \frac{\partial \mathcal{L}_d}{\partial G_d}$;
- $G_f' = G_f - \lambda_{f1} \frac{\partial \mathcal{L}_y}{\partial G_f} + \lambda_{f2} \frac{\partial \mathcal{L}_d}{\partial G_f}$

- D_S is unmasked load
- D_T is masked load

Online application
- Input data from target domain, x_T
- Input x_T to feature extractor then outcome predictor, calculate y_T.

Z. Zhou, Y. Xu, and C. Ren, "A Transfer Learning Method for Forecasting Masked-Load With Behind-the-Meter Distributed Energy Resources," *IEEE Trans. Smart Grid*, 2022.

[5] Transfer Learning for Forecasting Masked-Load due to Behind-the-Meter DERs

Test Settings

- Assume load is masked by a mix of DERs (PV, EV)
- To mimic dynamic development of DERs, different penetration of target datasets are tested.

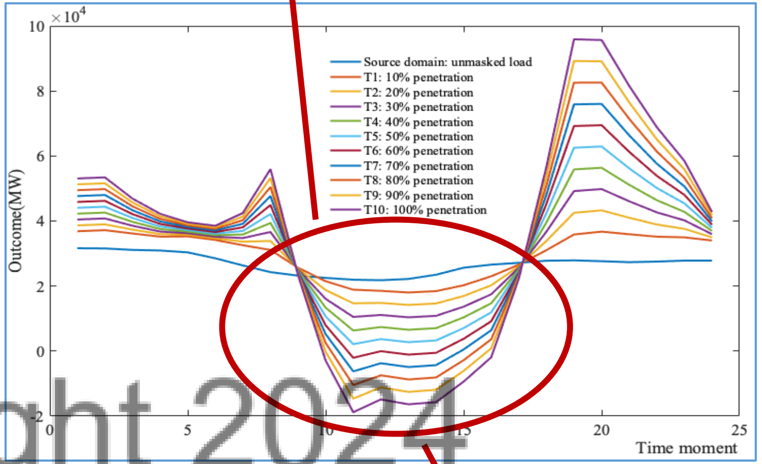
$$Dataset_T = Dataset_S - p_1 * PV + p_2 * EV$$

$$p_1, p_2\ adjust\ \frac{ave(p_1 * PV)}{ave(Dataset_S)}\%, \frac{ave(p_2 * EV)}{ave(Data_{setS})}\% = [10\%, 20\%, \ldots, 100\%]$$

- **Benchmark models:**

1. **Unmasked-load to forecast masked-load: $src \rightarrow tgt$**
2. **Masked-load to forecast masked-load: $tgt \rightarrow tgt$**

When DER penetration level increases:
"duck curve" of masked load becomes more apparent;
difference between source data and target data are larger;



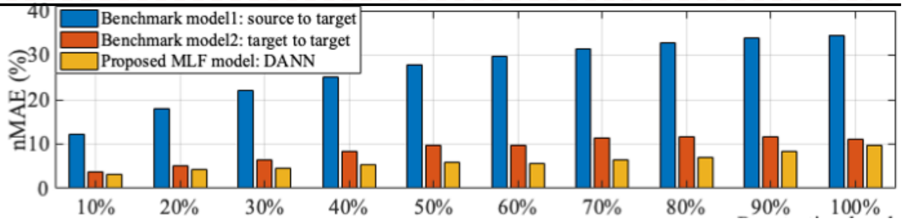
Test Results

Models	nMAE (%) in Different DERs Penetrations Levels									
	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
1	12.28	17.97	21.96	25.16	27.72	29.85	31.43	32.67	33.73	34.48
2	3.69	5.06	6.32	8.26	9.75	9.60	11.43	11.70	11.67	11.05
3	3.21	4.13	4.59	5.30	5.81	5.60	6.40	6.99	8.36	9.61

*1: Benchmark model 1: $src \rightarrow tgt$

*2: Benchmark model 2: $tgt \rightarrow tgt$

*3: Proposed model: DANN



At the lower DER penetration levels, accuracy improvement by TL is not significant since the load is just slighted masked (data distribution is not changed much). With growing DER levels, effectiveness of TF is more and more evident.

Z. Zhou, Y. Xu, and C. Ren, "A Transfer Learning Method for Forecasting Masked-Load With Behind-the-Meter Distributed Energy Resources," *IEEE Trans. Smart Grid*, 2022.

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Publications in Transfer Learning and Funding Acknowledgments

[1] C. Ren and Y. Xu, "Transfer Learning-Based Power System Online Dynamic Security Assessment: Using One Model to Assess Many Unlearned Faults," *IEEE Trans. Power Systems*, 2019.

[2] C. Ren and Y. Xu, "An Integrated Transfer Learning Method for Power System Dynamic Security Assessment for Unlearned Faults with Missing Data," *IEEE Trans. Power System*, 2021.

[3] Y. Xia and Y. Xu, "A Transferrable Data-Driven Method for IGBT Open-Circuit Fault Diagnosis in Three-Phase Inverters," *IEEE Trans. Power Electronics*, 2021.

[4] Y. Xia, Y. Xu, and N. Zhou, "A Transferrable and Noise-Tolerant Data-Driven Method for Inverter Open-Circuit Fault Diagnosis in Microgrids," *IEEE Trans. Industrial Electronics*, 2023.

[5] Z. Zhou, Y. Xu and C. Ren, "A Transfer Learning Method for Forecasting Masked-Load With Behind-the-Meter Distributed Energy Resources," *IEEE Trans. Smart Grid*, 2022.

[6] Y. Xia, Y. Xu, S. Mondal, and A. K. Gupta, "A Transfer Learning-Based Method for Cyber-Attack Tolerance in Distributed Control of Microgrids," *IEEE Trans. Smart Grid*, 2023.

Ministry of Education SINGAPORE

Rolls-Royce

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE


NANYANG
TECHNOLOGICAL
UNIVERSITY
SINGAPORE
Thank You!