SAND2024-05006C

Cyber-Physical Data Fusion & Threat Detection with LSTM-Based Autoencoders in the Grid

Authors: George Fragkos, Logan Blakely, Shamina Hossain-McKenzie, Adam Summers, Christopher Goes

Project: ***griDNA Sandia LDRD***

George Fragkos

Principal Member of Technical Staff @ Sandia National Laboratories

IEEE Kansas & Power Energy Conference 2024

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Agenda

- Introductory Notes
- Electric Grid Testbed, Dataset & Threat Model
- Cyber-Physical Data Fusion
- LSTM-Based Autoencoders for Cyber-Physical Data Fusion & Threat Detection
- Future work

Cybersecurity Challenges in the Grid

Cyber-Physical Power Grid

- The power grid has undergone a significant transformation in recent years
- High penetration of Distributed Energy Resources (DER)
- Integration with cyber communication networks and modern digital components
- Vast amounts of cyber and physical data are generated by the DER communication types and interfaces within the grid

Source: https://www.vifindia.org/article/2022/may/02/war-in-ukraine

Challenges

- There is a need:
  - For a robust mechanism to ingest and process this data
  - To fortify the grid against potential threats that can compromise its integrity and disrupt its functionality
- Examples of cyber-attacks: 2015 Ukraine grid cyber-attack, 2013 Metacalf snipper attack
- How can we use high-fidelity cyber-physical data to protect the cyber-physical power grid?

Artificial Intelligence (AI) for Cyber-Physical Security

Current Research Work Summary

- A big part of the research work focuses on utilizing Deep Neural Networks (DNNs) for identifying physical disturbances in the grid's measurements
- Another big portion performs threat detection using only cyber data originating from the IT environment
- There is a significant research gap for blending the physical with the cyber data
- ***"How can we effectively fuse the cyber-physical data generated in the grid to ensure situational awareness and detect not only cyber threats but also physical disturbances?"***

What do we propose?

- A cyber-physical threat detection methodology through data fusion
- Using a Long-Short Term Memory (LSTM)-based Autoencoder (AE)
- To integrate the temporal and structural patterns of cyber-physical data
- Generated by a Sandia's testbed that simulates a part of the electric grid

Image created with the Llama 3 Generative model from Meta AI

Electric Grid Testbed, Dataset & Threat Model

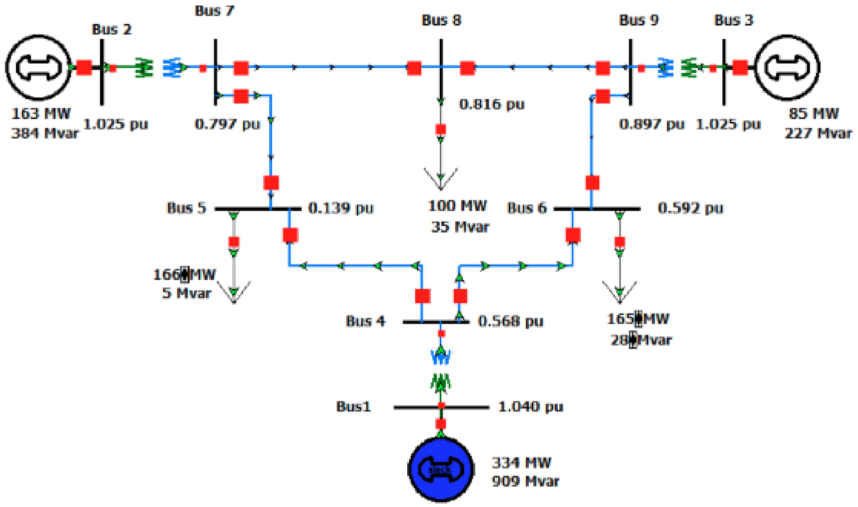
Electric Grid Testbed & Dataset

- Emulation environment: real-time digital simulator (RTDS) that enables streaming C37.118 data from PMUs in the RTDS WSCC 9-bus models and SCEPTRE
- SCEPTRE: Sandia Emulation Tool - It provides a comprehensive ICS/SCADA modeling and simulation capability that captures the cyber-physical impacts of targeted cyber events on critical infrastructure
- Cyber Features: Collected from 3 different relays in each of the three substations – packet RTTs and packet retransmissions
- Physical Features: Collected from 8 different PMUs – frequency, per-phase voltage, per-phase current

SCEPTRE Logo

Threat Model

- Physical events: a generator and line outage event (mitigation: load shedding)
- Cyber event: a Denial of Service (DoS) attack,
- Cyber-Physical event: generator and line outage events + DoS that impedes the load-shedding signal issued by the control center
- Result: Unstable system as defined by frequency instability



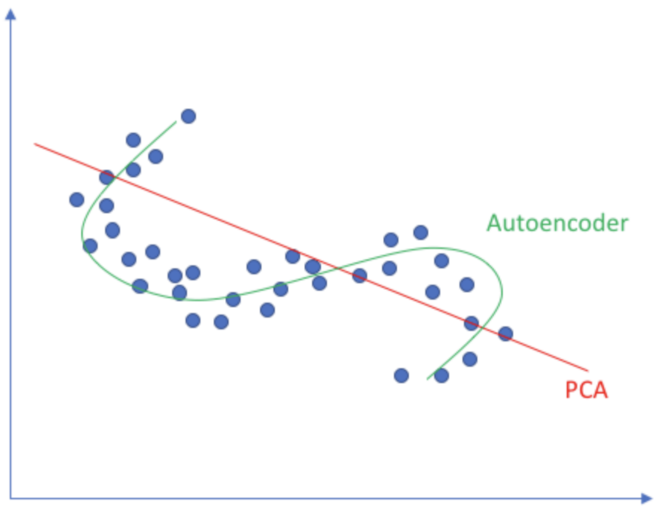
WSCC 9-Bus System

High Dimensionality & Multimodal Cyber-Physical Data

Problem

- Analyzing and extracting meaningful insights from cyber-physical data require specialized techniques and handling
- As the dimensionality of the input space increases, the complexity of the classification task also increases
- There is a growing interest in reducing the dimensionality of the input space to enhance the predictive performance of classification models

Linear vs nonlinear dimensionality reduction



Approaches for dimensionality reduction

PCA
[Principal Component Analysis]

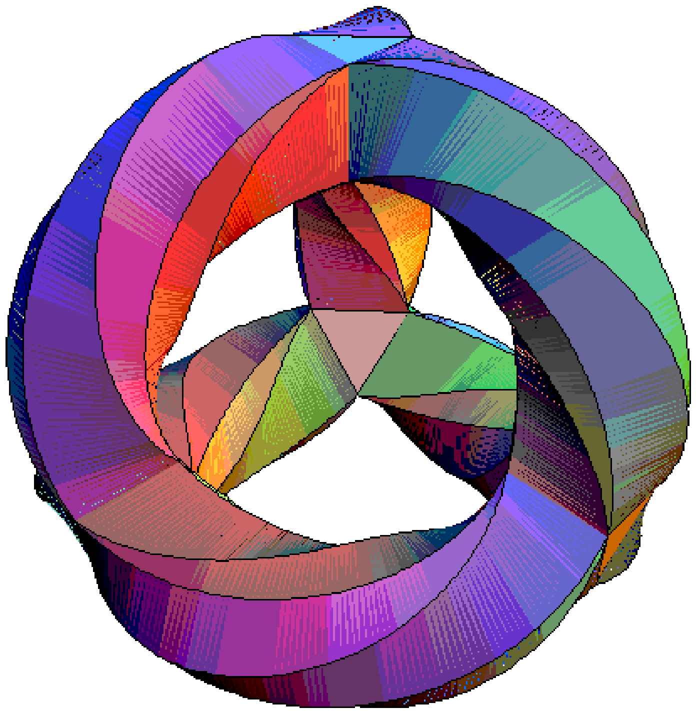
LDA
[Linear Discriminant Analysis]

Assumption: The data lies on a linear subspace

Not always true in real-world scenarios!

Cyber-Physical Data Fusion

- It is a dimensionality reduction technique used to understand the underlying structure of complex high-dimensional cyber-physical data
- It aims to uncover the intrinsic low-dimensional manifold on which the data points lie
- By preserving the local and global relationships between the cyber-physical data points, manifold learning provides a more meaningful representation for further analysis, i.e., threat detection in the electric grid



Autoencoders for Cyber-Physical Data Fusion

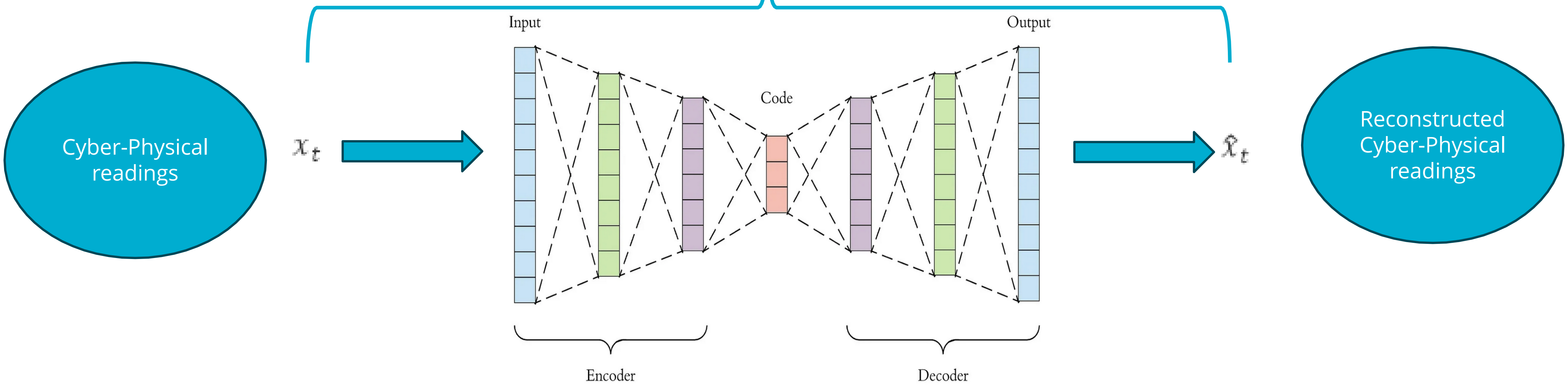
Description

- Autoencoders (AE) are a type of Artificial Neural Networks (ANN) used for:
  - Unsupervised Learning
  - Dimensionality reduction/Compression
  - Data Fusion

How do they work? / Architecture

- **Encoder**: Compresses the input cyber-physical data into a lower-dimensional space using an encoder network – bottleneck layer
- **Decoder**: Then it reconstructs the input data back into the original space using a decoder network
- It learns an **internal representation/code** to perform useful transformations on the input data (middle layer)
- Finds a codification of the input cyber-physical data by learning non-linear combinations of their features

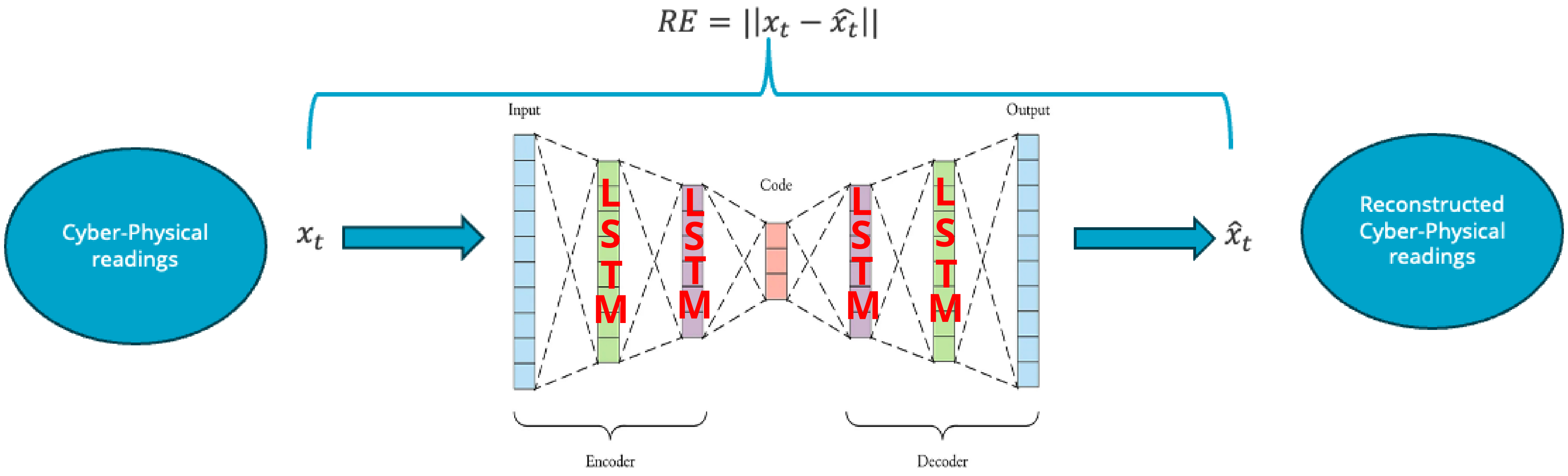
$$RE = ||x_t - \hat{x}_t||$$



LSTM-Based Autoencoders for Cyber-Physical Data Fusion

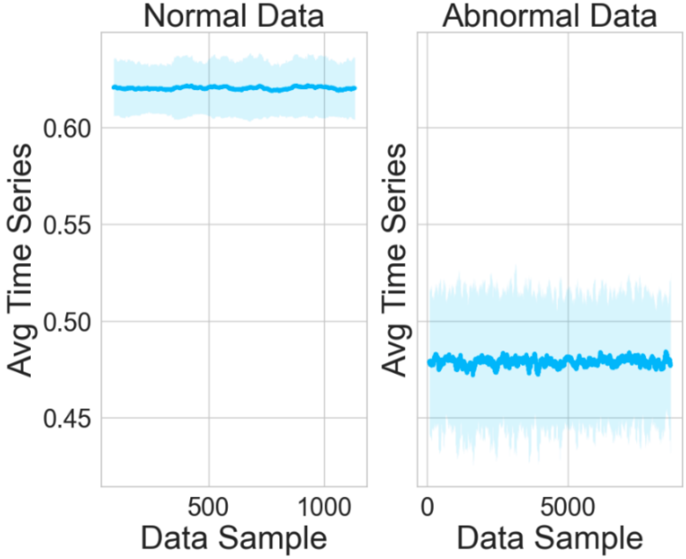
Why LSTMs?

- Plain AE might not capture the temporal ordering of the cyber-physical data
- LSTM are a type of Recurrent Neural Networks (RNN) that can capture the temporal patterns in time-series data
- LSTM-based AE: the only difference with the plain AEs is that the encoder and decoder are built using LSTM units instead of simple linear neural network layers
- The LSTM architecture within the AE enables memorizing past units and utilizing this memory to make predictions about future cyber-physical inputs



LSTM-Based Autoencoders for Cyber-Physical Threat Detection

Evaluation Analysis

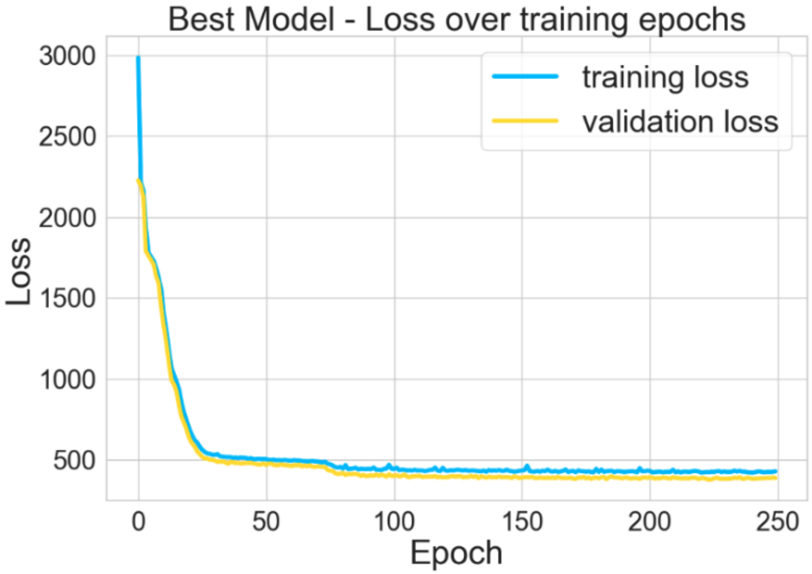


Avg Normalized cyber-physical time series

Distinct pattern between normal and abnormal data

Hyperparameters	*Grid-search values*
Number of layers	[2, 3, 4]
Weight Decay	[0.01, 0.001, 0.0001, 0.00001]
Learning Rate	[0.1, 0.001, 0.0001]
Dropout Rate	[10%, 20%, 30%, 40%]
Batch Size	[32, 64, 128, 256, 512]
Optimizer	[Adam, Adadelta, Adagrad, SGD, RMSprop]
Latent Space Reduction	[35%, 55%, 75%]

Grid-Search Hyperparameters Optimization

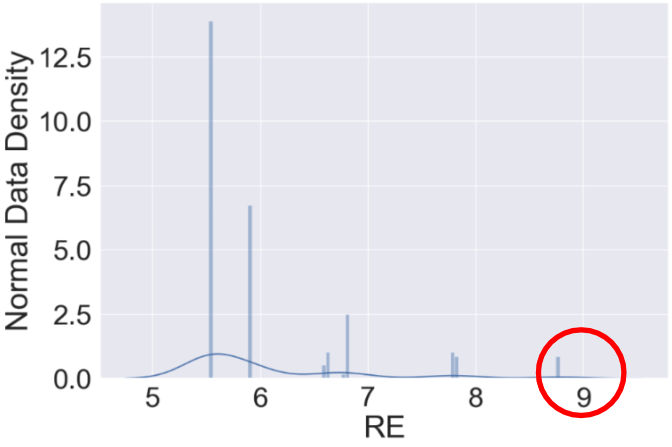


Training & validation loss of the optimized LSTM-based AE model

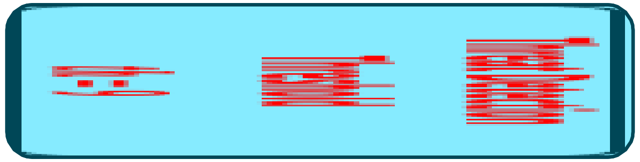
Optimized LSTM-based AE model is trained **ONLY on normal data**

LSTM-Based Autoencoders for Cyber-Physical Threat Detection

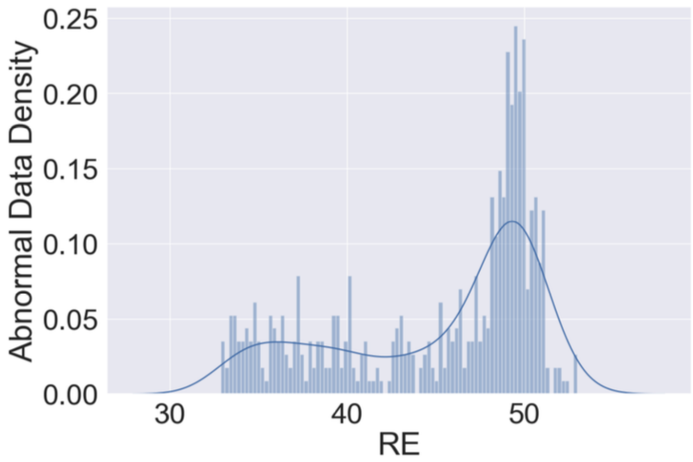
Evaluation Analysis



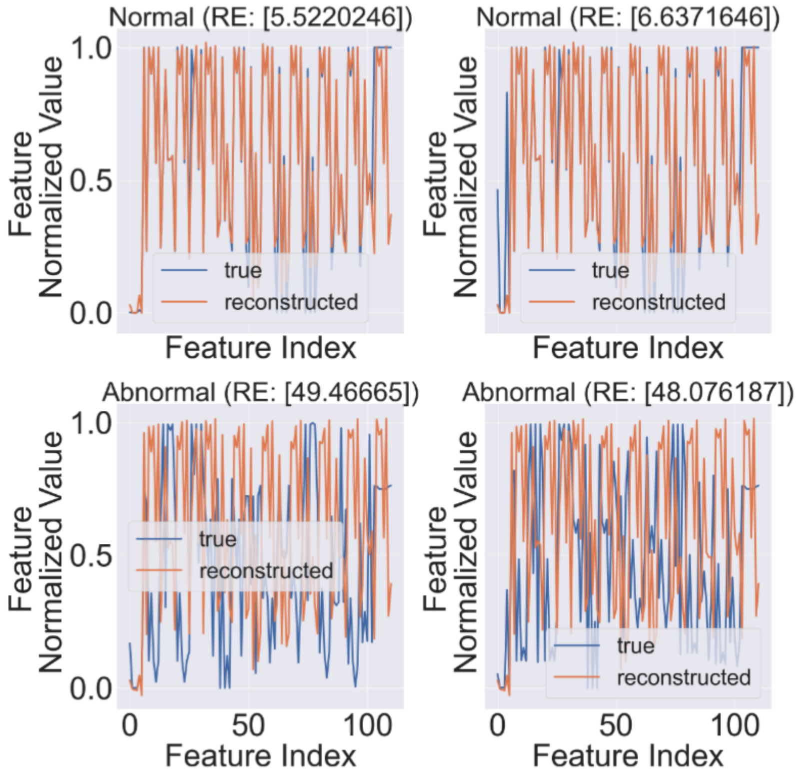
RE Histogram on the normal data of the trained LSTM-based AE model



$$Abormality(x_i') = \begin{cases} 1 & \text{if } RE(x_i') \geq th \\ 0 & \text{if } RE(x_i') < th \end{cases}$$



RE Histogram on the abnormal data of the trained LSTM-based AE model



Model's feature input values, reconstructed feature values and errors

Model	Accuracy	Avg Training Time (~10,000 data)	Avg Testing Time (1 data)
Cyber-Physical LSTM-based AE	100%	~22 minutes	<0.8 sec
Cyber-only LSTM-based AE	96.19%	~6 minutes	<0.4 sec
Physical-only LSTM-based AE	98%	~17 minutes	<0.8 sec

Future Work

- Team will investigate how the cyber-physical attack can be located within the grid
- Evaluate the combination of the LSTM-based AE model's learned latent space with other ML models
- Stealthier cyber-attacks will be examined

Contact: gfragko@sandia.gov

SAND2023-05402C

THANK YOU