PNNL-30935

Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

Bridging Power System Protection Gaps with Data-driven Approaches

January 2021

T Yin
J Lian
J Buckheit
R Fan

Prepared for the U.S. Department of Energy
under Contract DE-AC05-76RL01830

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from the
Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062;
ph: (865) 576-8401
fax: (865) 576-5728
email: reports@adonis.osti.gov

Available to the public from the National Technical Information Service
5301 Shawnee Rd., Alexandria, VA 22312
ph: (800) 553-NTIS (6847)
email: orders@ntis.gov <http://www.ntis.gov/about/form.aspx>
Online ordering: http://www.ntis.gov


This document was printed on recycled paper.
(8/2010)

PNNL-30935

Bridging Power System Protection Gaps with Data-driven Approaches

T Yin
J Lian
J Buckheit
R Fan

January 2021

Prepared for
the U.S. Department of Energy
under Contract DE-AC05-76RL01830

Pacific Northwest National Laboratory
Richland, Washington 99352

Summary

Protection is a critical function in power systems to avoid equipment damage, maintain personnel safety, and support system reliability. However, current protective relay technology cannot adequately protect equipment and personnel from effects of some events; these deficiencies are termed protection gaps. In this research, a data-driven approach is proposed to complement traditional protection technology and distinguish fault conditions from transients caused by normal operations. A convolutional neural network (CNN) based fault detection approach is implemented to achieve data translation invariance of the time-series input data. As a result, the data-driven method can accurately detect system faults despite variation and noise in the input data. In addition, using the CNN–based method avoids the complicated manual feature extraction procedure required by many traditional data-driven methods. The effectiveness of the proposed approach is tested on four kinds of protection gaps: high impedance faults, transformer/generator inter-turn faults, distribution system PV circuit faults, and the mis-operation situations of Zone 3 line protection relays operating under system stress. Finally, a transfer learning method is also proposed to address the common issue of data-driven methods for which real-world training data are scarce. Extensive study results demonstrate that the proposed approach can accurately bridge power system protection gaps.

Acronyms and Abbreviations

ANN	Artificial Neural Network
CNN	Convolutional Neural Network
MLP	Multilayer Perceptron
SVM	Support Vector Machine

Contents

Figures

Tables

1.0 Introduction

Electric energy grids are subject to faults and failures that result in unsafe conditions for humans and can damage power equipment or even cause an entire system to break down. The objective of power system protection is to isolate the faults or the affected components as fast as practicable without interrupting the capability of the system to serve the electric loads. Therefore, the protection schemes must accurately identify faults from other system transients caused by normal operations, such as capacitor bank switching, load changes, and motors starting, etc.

The key component for power system protection is the protective relay that monitors the system and makes decisions to interrupt the circuits if faults are detected. Traditional protective relays are designed to issue tripping commands when certain preset thresholds are exceeded. However, it is sometimes very difficult to determine accurate thresholds, because they usually depend on many factors such as operating conditions, equipment parameters, system transients, and fault types. The threshold settings represent a trade-off between protection sensitivity and security. Therefore, in practice, traditional relays cannot provide reliable or secure protection against faults or transients under certain circumstances. “Protection gap” is a term that describes the inadequacies in the protection schemes where existing technology would either be incapable of detecting the existence of a fault, or would misjudge normal operations and inadvertently perform a tripping action when such action is not desired. For example, when some faults happen the monitored measurements may resemble normal conditions. In this case, relays cannot detect this type of faults due to insufficient sensitivity. However, these faults, including transmission line high impedance faults, transformer/generator inter-turn faults, and distributed energy resource (mainly PV) minor circuit faults etc., are detrimental to the system. Another example of protection gaps is the misoperation of relays that trips healthy components during normal operations. The most common misoperation of line protection is that of so-called “Zone 3” relays falsely tripping transmission lines under overloaded conditions, which could lead to major disturbances or blackouts.

To solve the problem, we propose a new data-driven approach to bridge power system protection gaps. Instead of replacing existing relaying technologies, the proposed data-driven approaches will complement traditional relays and specifically target the protection gaps. The objective is to increase the overall resilience of the system with additional technology provided by these proposed data-driven approaches, working alongside traditional relay technologies. We have built and trained a convolutional neural network (CNN) to find mappings between raw instrumental measurements and corresponding faults or normal conditions. Compared with traditional relays that are based on preset thresholds, data-driven approaches can capture differences in both the magnitude and pattern of the measurements under normal operations and fault conditions. In addition, the proposed data-driven approach overcomes three major limitations of traditional ANNs. First, a traditional ANN does not take into account the temporal correlation of data. In contrast, the convolutional layers and pooling layers in a CNN model help learn the shape of the data and preserve the translation invarianceand of the time-series data; thus, it can accurately detect a fault even if it varies. Second, many traditional ANN-based methods require a complicated, time-consuming signal-feature extraction procedure that uses either a discrete Fourier transform or a discrete wavelet transform to preprocess the raw input data. However, the CNN-based approach can learn the signal features automatically during the training process; the researchers do not need to manually perform the complicated feature extraction, which saves much effort and time. Third, training an ANN system requires a large amount of data; it is very difficult to use ANN-based schemes in a real-world system with little training data. The proposed data-driven approach applies a transfer learning method to address this issue by leveraging a previously trained CNN model that has captured certain fault characteristics and patterns.

2.0 Modelling of Protection Cases

To test and proof the feasibility of the proposed data-driven approach, we have chosen four protection cases that represents typical protection gaps: high impedance faults, transformer inter-turn faults, PV circuit faults, and the mioperation situations of Zone 3 line protection relays operating under system stress. The first step is to model these protection cases and generate enough data for training and testing the proposed data-driven method.

2.1 High Impedance Fault

A HIF usually happens when a quasi-conductive object (such as a tree branch) contacts the line, or the transmission line breaks and falls to the ground. Because the quasi-conductive objects or ground have high resistance, the induced fault current is usually very low. Therefore, it is very difficult for traditional relaying technologies to detect the existence of a HIF. However, a HIF represents a great threat to personnel and public safety. The voltage around the downed conductor is very high; thus, people who walk near the downed conductor could be shocked if they were unaware of the hazard. In addition, the constant arcing caused by the HIF would easily ignite wild fires in the forests during dry and hot seasons. In the real world, around 5–20% of all distribution faults are HIFs, so they represent a serious public safety hazard.

A generic HIF model containing antiparallel DC sources, diodes, and variable resistors is used in this paper, as shown in Figure 1. The subscripts p and n in the model stand for positive and negative, respectively. When a HIF occurs, the model will be connected to the instantaneous phase voltage Vph(t). When the instantaneous phase voltage Vph(t) is larger than the positive DC voltage Vp, the positive cycle of fault current flows toward ground through the left-hand path in Figure 1. When the instantaneous phase voltage Vph(t)is smaller than the negative DC voltage Vn, the negative cycle of fault current flows from the ground through the right-hand path in Figure 1. Otherwise, neither the left nor the right path conducts. When simulating with the generic HIF model, the values of Vp,Vn,Rp, and Rn are randomly set to reflect the realistic characteristics of HIF: nonlinear impedance, time-varying parameters, and random nature.

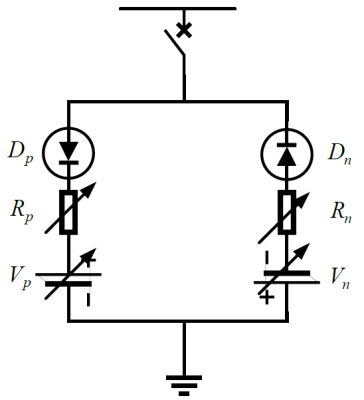
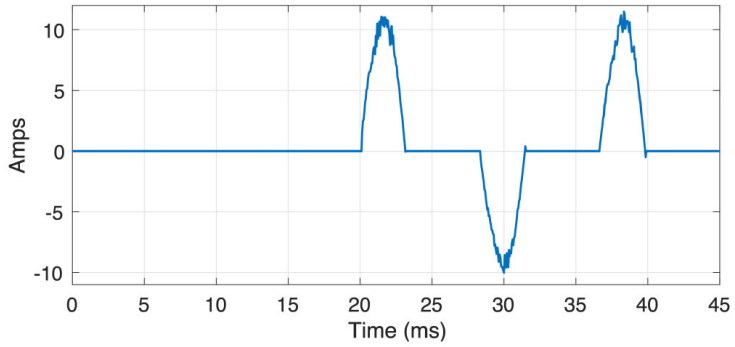


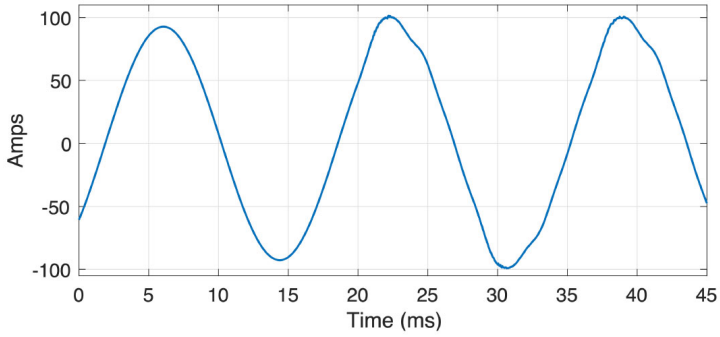
Figure 1. Generic HIF model

An example of using the generic HIF model is shown in Figure 2, where the fault happens at 18 ms. Figure 2(a) shows the modeled fault current at the HIF. The nonlinear fault current only conducts during the peak of the phase voltage (when Vph(t) > Vp or Vph(t) < Vn). Figure 2(b) shows the respective branch current as measured by transmission line terminal relays. The magnitude of the line current changes little during the HIF; thus, a traditional overcurrent relay cannot detect the HIF. On the other

hand, the HIF causes a distortion of the sinusoidal current waveform. It is possible to use a data-driven method to capture the fault pattern and differentiate the fault from normal transients.



(a) HIF current



(b) Current measured by the relay at line terminal

Figure 2. HIF and line current using generic HIF model

The above system is built in ATP-EMTP software, and the test system is IEEE 34-bus system. A snip-shot of the ATP-EMTP system is shown in Figure 3.

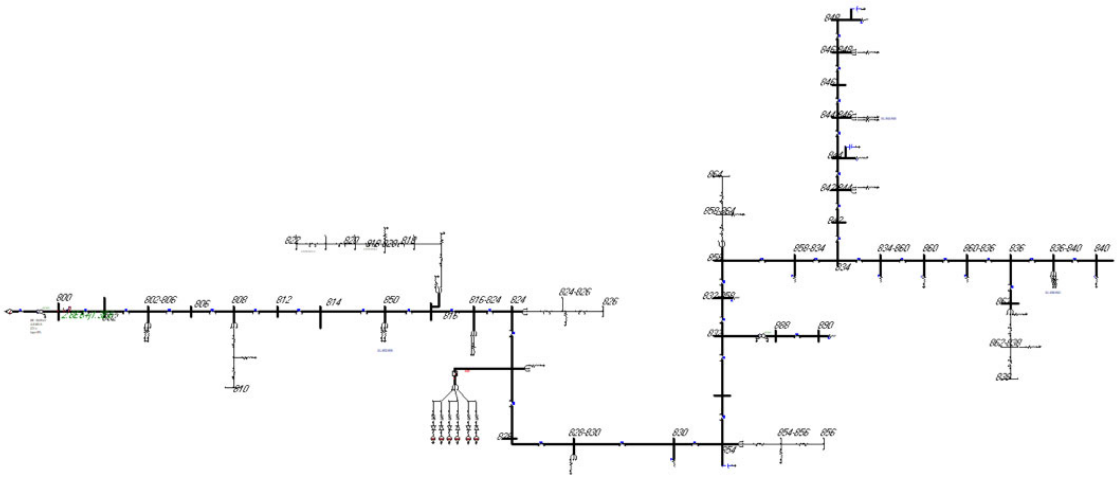


Figure 3. HIF test system in ATP-EMTP

2.2 Transformer Inter-turn Fault

A transformer inter-turn fault in another kind of protection gaps that has been existing for decades. During normal operation of power transformers, the windings and iron cores generate heat because of losses in conduction, hysteresis, and eddy currents. This heat would slowly damage the insulation of the windings over years of operation, unnoticed by the protection relays. If the winding insulation is worn out, one or more sequential pairs of turns will be shorted, which causes the "inter-turn" (or turn-turn) faults. The challenge to protective relaying technology is that the low fault currents make these inter-turn faults extremely difficult to detect. When an inter-turn fault begins in the transformer, usually only a very small portion of the windings is shorted. Therefore, the primary-secondary turn ratio and the magnetic flux will remain almost unchanged, resulting in negligible changes in terminal voltage and current measurements [10]. However, because the short resistance is so low, the fault current through the shorted circuit could reach thousands of amperes. The extremely high fault currents generate copious heat and cause local thermal overloading, which ultimately evolves to catastrophic failure if the inter-turn fault is not detected and isolated in its earliest stage.

The inter-turn fault model is based on a single-phase, saturable-core transformer model, as shown in Figure 4, with two additional branches added to the transformer windings (either primary or secondary side). The inter-turn fault location is determined by parameters α and β, and the severity of the inter-turn fault is determined by the value of $|\alpha - \beta|$ and the fault impedance in the short circuit.

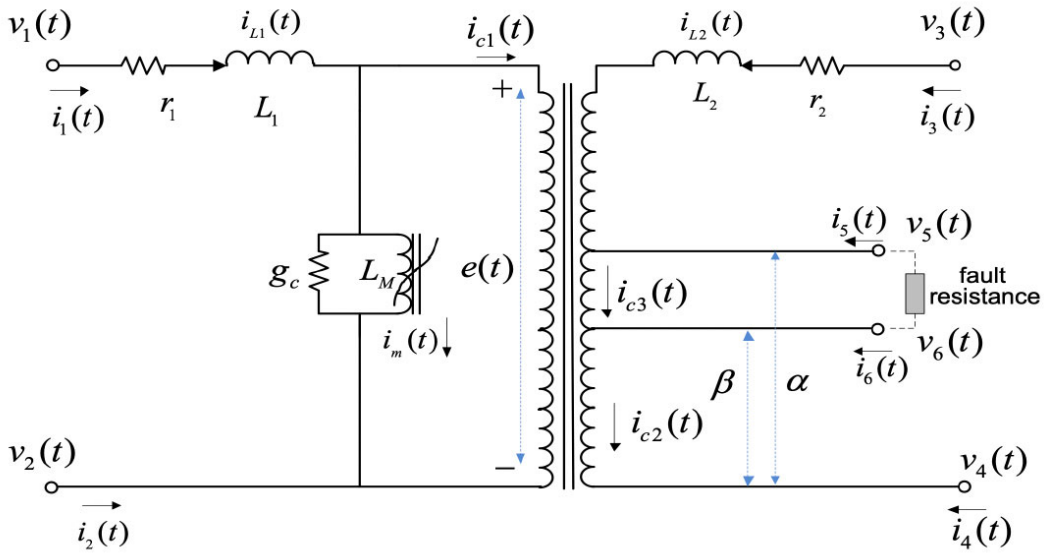


Figure 4. Transformer inter-turn fault model

It is easy to construct a delta-wye or wye-delta three-phase transformer by connecting three of the single-phase transformers shown in Figure 5. An example of terminal measurement of a 40 MVA 115/35 kV three-phase transformer with a secondary-side 1% inter-turn fault is shown in Figure 5. Hardly any change is noticeable in voltage or current measurements when the inter-turn fault happens, because the overall primary-secondary ratio stays almost the same. Traditional protective relays using percentage differential protection or negative-sequence differential protection methods cannot detect such a small inter-turn fault either.

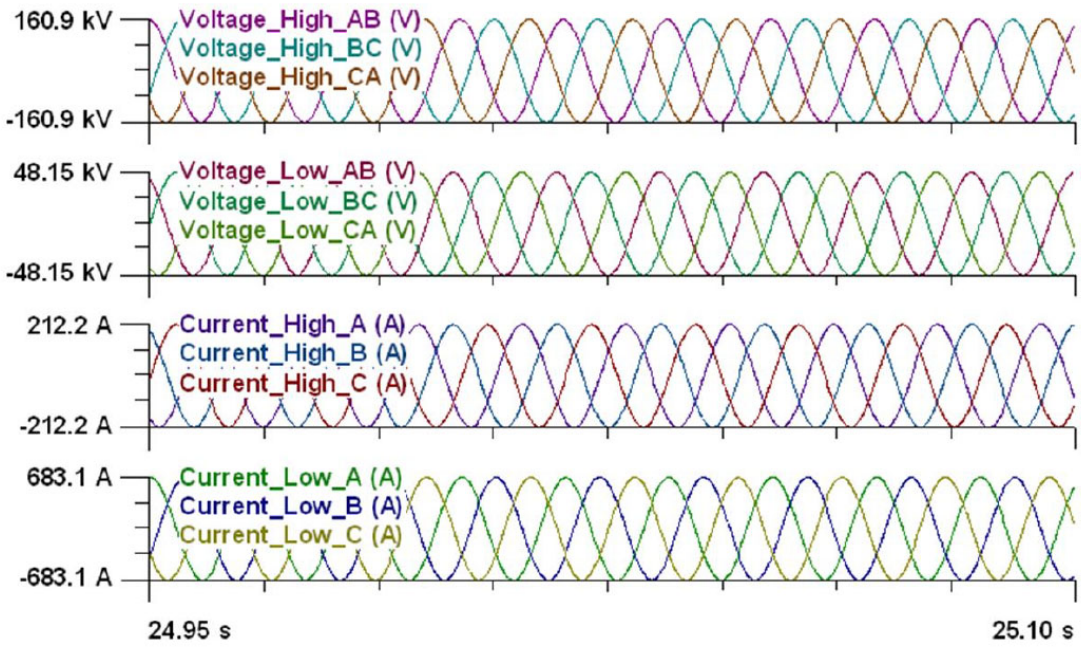


Figure 5. Transformer inter-turn fault terminal measurements

The above system is built in WinIGS software. A snip-shot of the transformer inter-turn fault test system is shown in Figure 6.

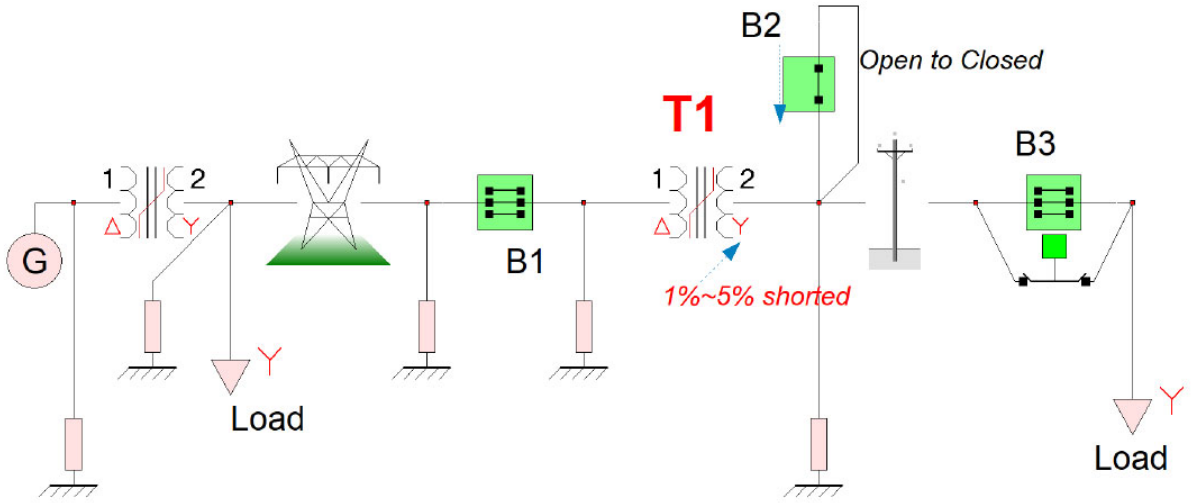


Figure 6. Transformer inter-turn fault test system

2.3 PV-Connected Distribution System Fault

Inverter-based DER such as solar photovoltaic (PV) acts like a voltage-controlled positive-sequence current source, with little or no zero sequence and negative sequence content. Rotating machines in DER

act like voltage sources, much like the grid itself. The behavior of rotating machines on the grid is well understood; simplified fault current models are available, with dynamic and transient models also available if needed. The rotating models, software tools, and machine type tests, which are the basis for model parameters, have evolved together over more than 100 years of operating experience. Inverters are much newer and much different than rotating machines; they do not provide much fault current, they can follow the terminal voltage angle very quickly, and there are no standard type tests for simplified fault models. This makes it harder to perform protection analysis and increases the chance for errors.

A rotating machine is represented with a voltage source behind impedance, or Thevenin equivalent (Figure 7 left), and it provides 5-6 times rated current to a fault on its terminals. During a fault, the phase relationship between terminal voltage and current can change suddenly because the Thevenin source angle does not change very much, due to inertia and the relatively slow machine excitation controls. On a radial distribution feeder, such DER behaves similarly to the substation source, but is not as strong.

Inverter-based PVs are represented with a voltage-controlled current source in parallel with an impedance, or Norton equivalent (Figure 7 right). Fast-acting inverter controls limit the fault contribution to no more than 2 times rated current, and usually no more than 1.1 times rated current. The inverter controls may also act quickly to hold a constant phase angle between current and voltage, so the source angle can change quickly. On a radial distribution feeder, such PV provides little fault current on their own, although certain types of interconnection transformer may contribute significant ground fault current.

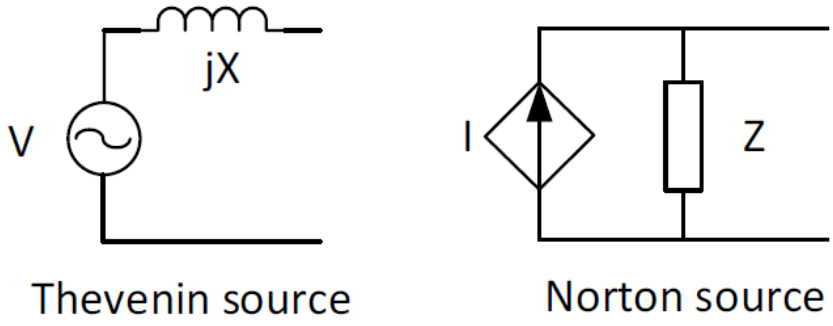


Figure 7. Controlled current source (Norton equivalent) for inverter-based PV, and Thevenin equivalent voltage source for rotating machine

The inverter model was developed to mimic the behavior of a real single-phase inverter, in simplified form. First, block diagram logic was implemented to maintain real power output at the steady-state value, subject to a limit on the RMS value. Because of this, under low-voltage conditions the inverter current increased to a limit of around 1.1 per-unit. Second, a phase-locked loop (PLL) was implemented using a quarter-cycle transport delay for single-phase inverters. After any disturbance, the PLL acted to bring the output voltage and current in back in phase. This also had the side effect of appearing to control reactive power, but that was not the PLL's main purpose. We only wanted to obtain realistic results for the dynamic angle behaviors during fault conditions.

A real PLL would provide the same function, but perform differently. The logic for both magnitude control and PLL were used to drive a controlled current source component in ATP. The ATP test system is based on IEEE 13-bus test feeders, and it is shown in Figure 8.

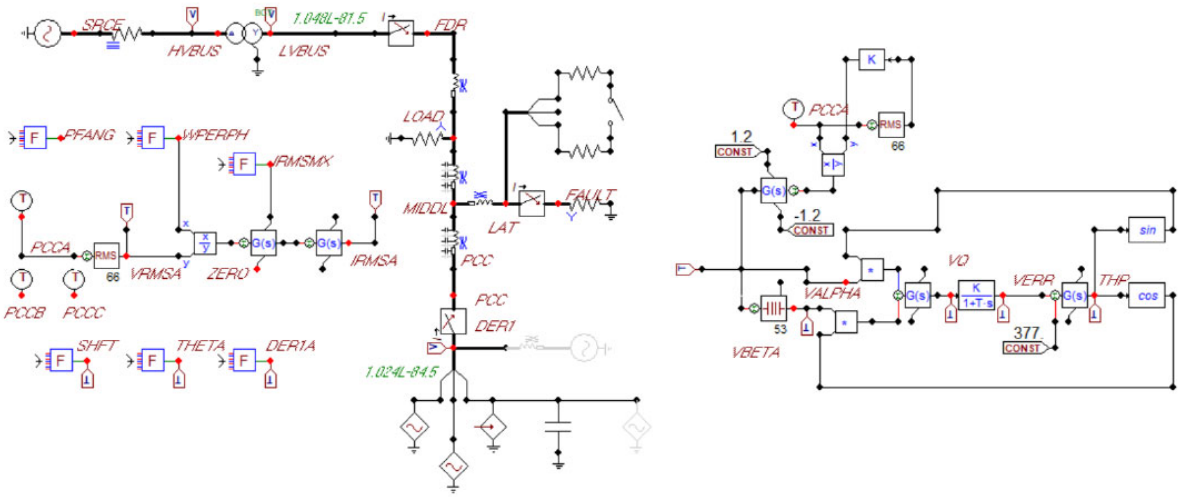
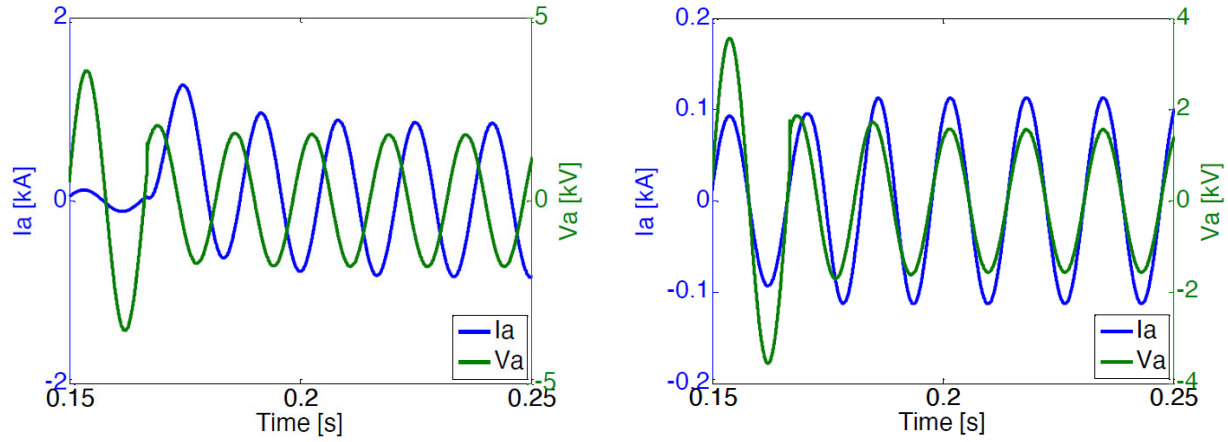


Figure 8. PV Distribution Faults on IEEE 13-Bus Feeder

It shows in Figure 9 the phase A voltage (green) and current (blue) for PV, during a line-line fault (LLF), as represented by Thevenin (left) and controlled Norton (right) sources. In both cases, the pre-fault current and voltage are in phase, the fault occurs at about 0.167 seconds, and the post-fault voltage is about 0.46 per-unit. The Thevenin source current, representative of a rotating machine, increases to about 6 per-unit, and the current lags the voltage by nearly 90 degrees. The controlled Norton source current, representative of an inverter, increases to about 1.1 per-unit. The current lags the voltage for only one and a half cycles and by less than 90 degrees. After that, the PLL brings the voltage and current back in phase.



(a) Thevenin source

(b) Norton source

Figure 9. PV system LLF response

2.4 Zone 3 False Trip during System Stress

The most popular protection scheme for transmission line protection is distance protection. Distance relaying for transmission circuits provide a more secure protection scheme as compared with overcurrent based protection schemes. Distance relays simply trip when the impedance they "see" falls within the

characteristic of the relay. When applied to transmission lines, depending on the fault type, the equivalent per unit length impedance may vary. For example, for a three phase fault the per unit length impedance of the line equals the positive sequence impedance of the line. For a single line to ground fault the equivalent per unit impedance is approximately equal to the average of the positive, negative and zero sequence impedance of the line. For the purpose of standardizing the distance relay design for three phase circuits, the relays should be so designed as to "see" an equivalent impedance that is approximately equal to the positive sequence impedance of the circuit per unit length times the distance to the fault. This is easily achieved with numerical relays by providing appropriate algorithms.

The typical distance relay includes three zones for protecting a transmission line, as shown in Figure 10. A typical practice is to set the zone 1 of the distance relay to about 80% of the line impedance (i.e. to reach 80% of the length of the line). This operation is fast with just a small delay (two to three cycles) to avoid tripping on transients. This practice allows line protection for the majority of the faults along the line. The 80% figure is selected to make sure that zone 1 operation (which normally does not have any appreciable time delay) does not operate on faults past the line. In other words we have a safety margin of 20%. Zone 2 is typically set to reach 125% of the line length. Time delays are moderate in the order of 10 to 20 cycles to coordinate with other fast tripping schemes. And zone 3 is typically set to reach 100% of the line plus 150% of the next line. The time delays for zone 3 are typically 30 or more cycles.

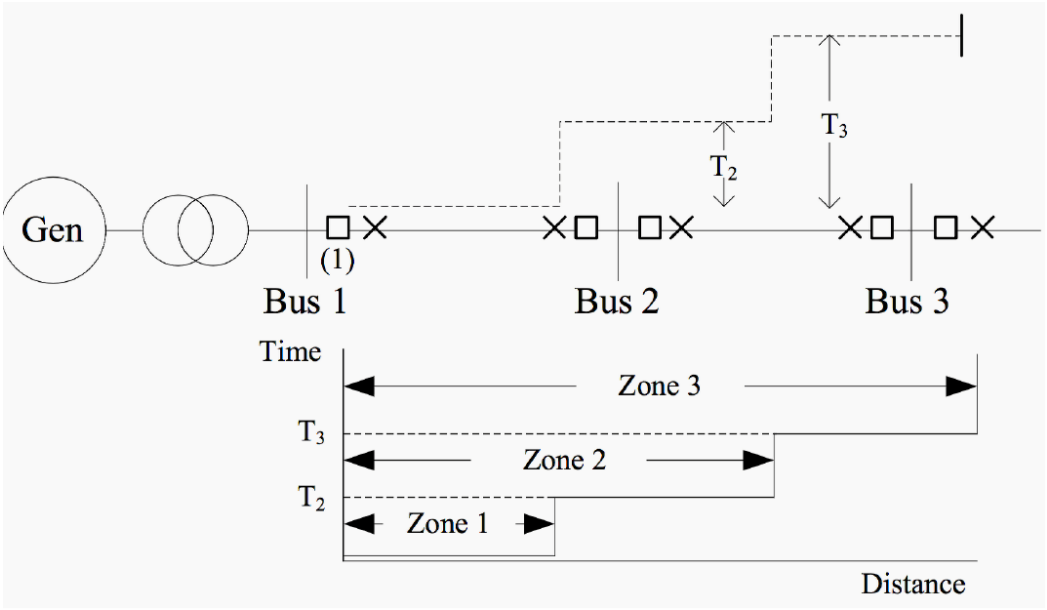


Figure 10. Distance relays with three protection zones

The typical relay characteristic of the three zones are shown in Figure 11. The problem of distance relay is that the safety of Zone 3 protection may not be guaranteed during system stress. During normal conditions, the distance relay will "see" an impedance outside of the relay charateristics, ie. around the blue dot in Figure 11. Therefore, the relay will not take any action and that ensures the safety operation of the power system. However, when the system is significantly overloaded, or if power osillaiton occurred in the system, there is a chance that the impedance "seen" by the relay moves into the Zone 3 region (the red dot). In that case, the relay will make a trip decision to cut off the line. As a consequence, the operation of the system is disturbed. Because there is no fault in the system, the power flow will be changed. During high-load periods, this would significantly increase the power flow on other lines, which could cause cascaded Zone 3 failures.

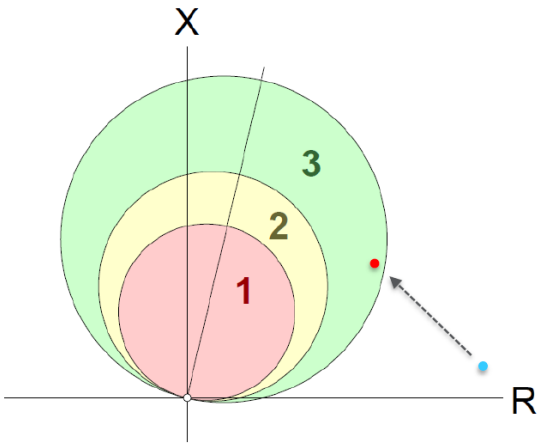


Figure 11. Relay characteristics of distance protection

The test system for Zone 3 protection is shown in Figure 2.13. The test system is also built in ATP-EMTP. In the test system we can select the fault location on different transmission lines to mimic the Zone 1, Zone 2, Zone 3 and out of Zone faults. The location, inception time, type, and severety of the faults can be randomly selected to generate different scenarios. In addition, the system loads can be also changed to produce the overloaded scenarios.

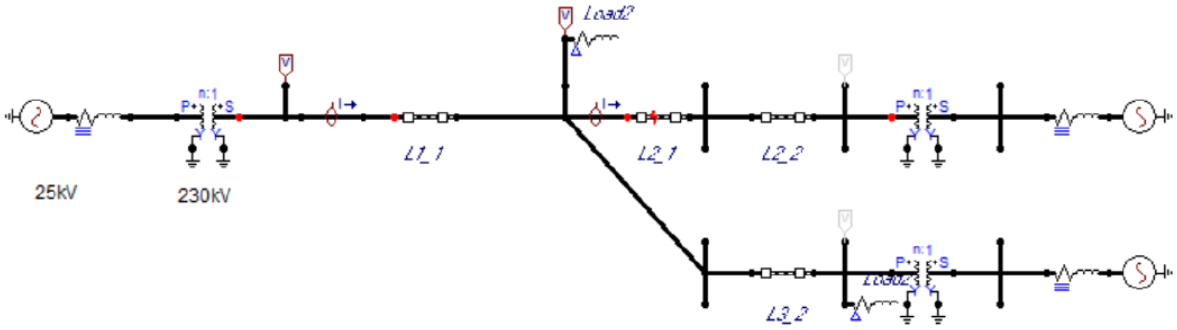


Figure 12. Zone 3 protection test sytem

2.5 Data Generation and Automation

One of the major challenge for data-driven approaches is the requirement of excessive data. In our project, we need to generate thousands or even tens of thousands of data for each protection gap case. Here is a simple math, if each piece of data generation takes 10 minutes (build the model, set the parameter and transient events, run the simulation and store the data properly), and we plan to generate 6,000 piece of data for each protection gap case, the total time required to generate those data will be $10 * 6000 * 4 = 240{,}000\ min = 4{,}000\ hours = 500\ business\ days$ ($8\ hour/day$). This calculation has not even counted the possible delay because of human mistakes.

Obviously, it will be impossible to achieve that mission by manually generating the data. To solve the problem, we have designed an automatic data generation procedure, as shown in Figure 13. We used a python controller to control the overall data generation. Before start running the python code, we manually built a well-designed base case for each of the four protection gap cases. In the base case, we specified the desired variables we would like to change randomly so that we could change the scenerios

of the simulation cases. For example, the fault impedance and fault inception time can be set as variable x and y. Thus, randomly changing the value of x and y would make the simulation case a new one with new data to be generated. When we started running the python codes, we would read in the base case and randomly change the value of the indentified variables (for example x and y). Then we saved the new simulation case to a new case so that we could run the simulation and generated the corresponding data. We could continue the creation of new cases and generation of new data until we reached the desired amount of data. These steps were all achieved by using the python controller we designed in this project.

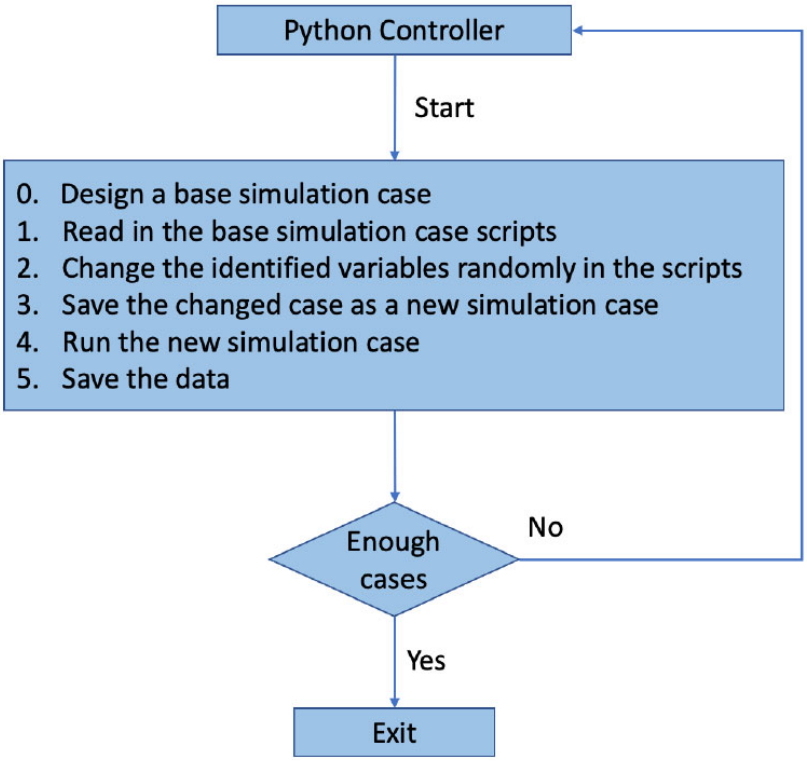


Figure 13. Data generation automation

3.0 Data and Models

3.1 Proposed Data-Driven Approaches

The novel data-driven approach we used to bridge the protection gaps is call deep convolutional neural network (CNN), and it is shown in Figure 14. The proposed model has an input layer, four CNN layers, two fully-connected (dense) layers and a sigmoid output layer.

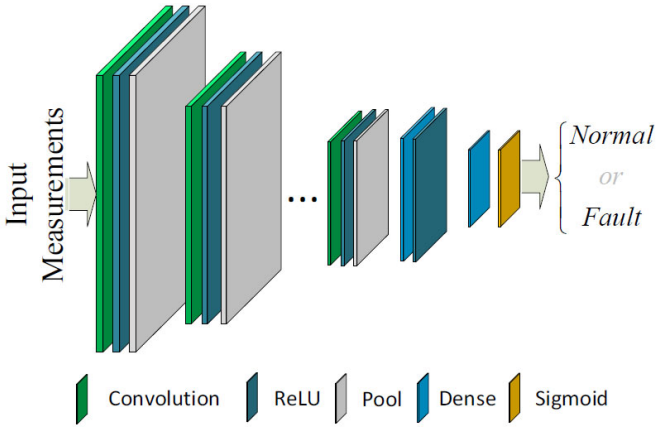


Figure 14. Proposed CNN model

The model takes the current waveforms as inputs and classifies them into either normal or fault conditions. Each CNN layer consists of convolution, rectified linear unit (ReLU) activation, and max-pooling functions. Figure 15 shows the operation of the convolution operation. The convolution function can help the model capture detail patterns of the input, and in a CNN, it is

$$(f * g)[n] = \sum_{i}^{n} f[i] \times g[i]$$

where f is the filter feature, g is the input that corresponds to the filter, and n is the size of the filter.

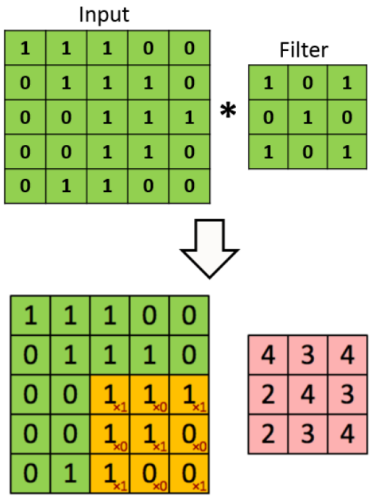


Figure 15. Illustration of the convolutional function in a CNN model

The output of the convolution operation x is passed to a ReLU activation, which is defined as

$$ReLU(x) = x^{+} = max(0, x)$$

The operation of ReLU activation function is shown in Figure 16. If the input x is larger than zero, the output will be the x itself; otherwise, the output will be zero. A nonlinear activation is required if we want the CNN be able to do some interesting computation. However, we do not want the nonlinear function to be too complicated to cause any gradient-diminishing problem in the deeper layers of a model. This simple ReLU activation function adds nonlinearity to the neural network, without complexing the derivative during the backpropagation process. ReLU handles the vanishing-gradient problem quite well, and it is less computationally expensive than traditional tanh and sigmoid activation operations.

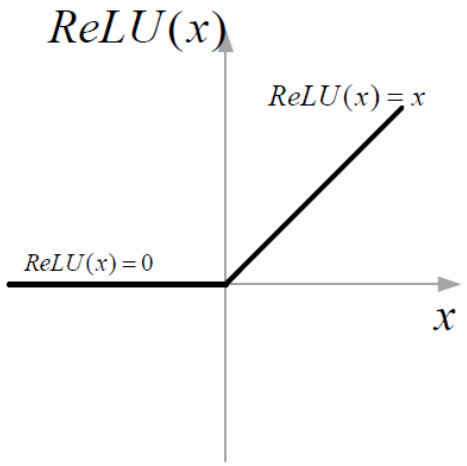


Figure 16. ReLU activation function

A max-pooling operation follows the ReLU activation is used to reduce the dimensionality and allow filters in deeper layers to learn a general overview of the input patterns. Its operation is very simple and straightforward, as shown in Figure 17. The output of 2×2 max-pooling operation is the corresponding maximum value of each 2×2 square of the 4×4 input.

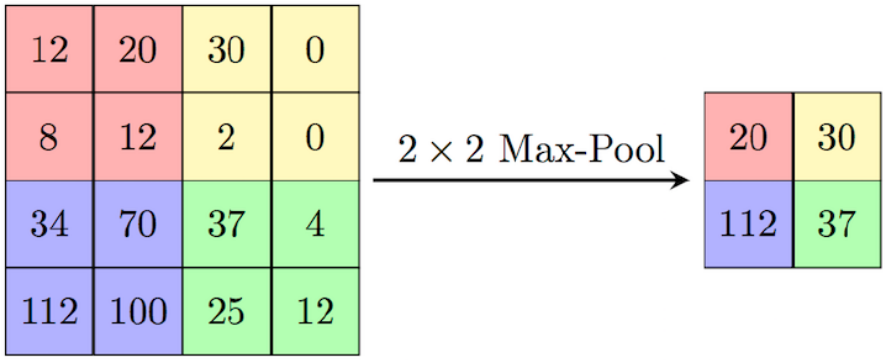


Figure 17. Max-Pooling function

At the end layer of the CNN, a fully-connected layer and an output layer using a sigmoid activation function is used to determine whether the input measurements are faults or normal transients. The sigmoid function is

$$\hat{y} = \frac{e^x}{e^x + 1}$$

The output of the sigmoid S(x) will be rounded to either 0 or 1, which stand for fault or normal scenario in our cases, as shown in Figure 18. To train the neural network and update all the parameters, a cross-entropy is used as the loss function, which is defined as

$$J = -\frac{1}{m}\sum_{j}^{m}\left[y^{(j)} log\left(\hat{y}^{(j)}\right) + \left(1 - y^{(j)}\right) log\left(1 - \hat{y}^{(j)}\right)\right]$$

where y is the true label of the input, m is the number of input data, and the superscript j stands for the jth observation. The CNN updates its parameters (or weights) of each layer with the objective of decreasing the loss J through a back-propagation process.

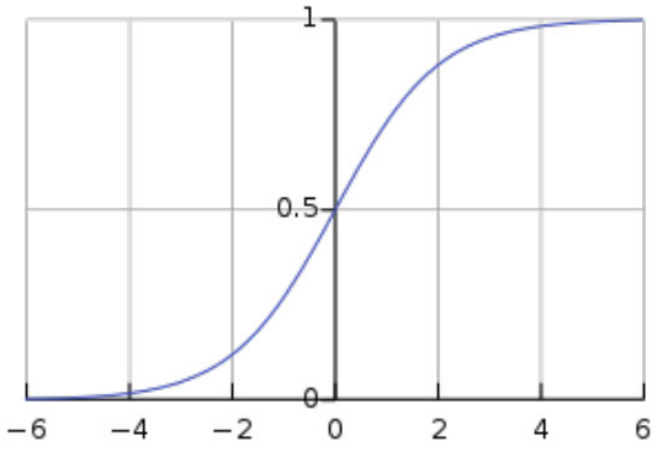


Figure 18. Sigmoid activation function

3.2 Conventional Data-Driven Approaches

We also designed two conventional data-driven methods to test their capability of bridging protection gaps. These two methods are multi-layer perceptron (MLP) method and supported vector machine (SVM) method. MLP is the most common deep learning network that has been used because of its simple structure and fast computation speed. A typical MLP network consists of the input, hidden and output layers, as shown in Figure 19. The single-ended voltage and current measurements are fed to the input layer and the fault distance is predicted from the output layer. The hidden layers contain many regression perceptrons to learn how to map the input with the output. The number of the hidden layers can be increased to a large number to build a “deep neural network” for learning complicated features. The MLP uses a four-layer neural network to discriminate fault and normal transients. The SVM is also a conventional data-driven method used for classification. The SVM is to find out the optimal boundaries with maximum margins to separate the faults and normal transients. The SVM method first uses a wavelet transform to capture the time and frequency information; then it uses a linear kernel to build the classifier.

The performance of proposed CNN-based approach is compared with the two conventional MLP and SVM methods for all the four protection gaps cases.

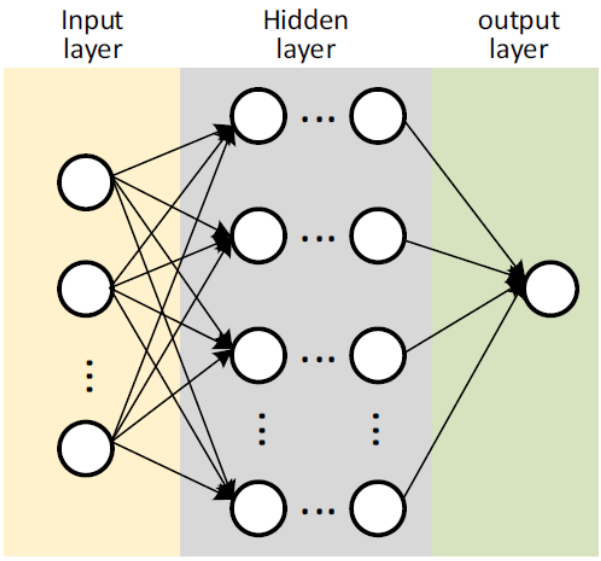


Figure 19. Multi-layer Perceptron

3.3 Evaluation Metrics

In this project, using a data-driven approach to bridge protection gaps is basically a binary classification method, and its output (prediction result) is either 1 (fault) or 0 (normal). The most popular evaluation metric is "Accuracy," which stands for the ratio of the number of correct predictions to the total number of predictions. However, this metric does not provide enough information about false predictions made by the neural network, which could be a problem when the number of positive labels is much more (or less) than the number of negative labels. To reasonably evaluate the performance of data-driven approaches, here we also use another metric: the "F1 score."

In a binary classification problem, the label is either P ositive or N egative; and the actual prediction is either True or False. Therefore, the prediction results of a binary classification can be divided into four categories:

1) True Positive (TP): the label y is positive and the prediction $\hat{y}$ is positive;

2) True Negative (TN): the label y is negative and the prediction $\hat{y}$ is negative;

3) False Positive (FP): the label y is negative and the prediction $\hat{y}$ is positive;

4) False Negative (FN): the label y is positive and the prediction $\hat{y}$ is negative.

The metric "Accuracy" is defined as follows:

$$Accuracy = \frac{T_P + T_N}{T_P + T_N + F_P + F_N}$$

To calculate the “F1 score,” we need to first calculate the “Precision” and “Recall.” “Precision” stands for the ratio of correctly predicted positive observations to the total predicted positive observations, which is defined as

$$Precision = \frac{T_P}{T_P + F_P}$$

“Recall” is the ratio of correctly predicted positive observations to all observations with actual positive labels, which is defined as

$$Recall = \frac{T_P}{T_P + F_N}$$

Finally, the metric “F1 score” is calculated as

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall}$$

With above equations, we can see that “Accuracy” focuses on TP and TN while “F1 score” also considers FP and FN . In this paper, we will use both “Accuracy” and “F1 score” as evaluation metrics.

4.0 Simulation Results

In this section, we will show the results of using the proposed data-driven approach (CNN) and the conventional data-driven approaches (MLP and SVM) to bridge the four protection gaps. Because the CNN and MLP both have the training/validation process, their performance will be compared for each epoch of training. The SVM has a different procedure of finding the optimal classification than the CNN and MLP; thus, only the final results of SVM will be presented.

4.1 High Impedence Fault

The HIF model in Section II is implemented in the IEEE 34-bus feeder, 24.9 kV test system. We generated around 4,500 transient scenarios (both HIF and normal) using this test system. The various normal events include load changes, sectionalizer switching, capacitor bank changes, DER generation, induction motor operation, and transformer tap changes. We selected the fault and normal transient parameters randomly. The HIF fault resistor varied between 150 and 700 ohms, and the conducting voltage varied between 65% and 85%. During simulation of the transient scenarios, we used various system conditions, including different loading (10%–90%) and voltage levels (0.95 p.u.–1.05 p.u.), different event locations and transient inception angles, and different line phases. The input measurements for the three methods were the time-domain sampled current measurements from the feeder terminal. The sampling rate was 15 kHz, and the input data length was two cycles. In addition, 2% white noise was added to the data to mimic real-world situations. From the generated transients, we randomly selected around 3,600 cases for training and 900 cases for testing.

During the training of the CNN and MLP model, we used 10% of the data for validation (to prevent overfitting). The cross-entropy losses of MLP and CNN methods during the training process are shown in Figure 20. The SVM method was not included in Figure 20 because it does not use training or validation algorithms as neural networks do. Figure 20 shows that the cross-entropy losses of all the schemes decreased as the training continued, until they finally reached steady states. The losses with MLP-based schemes are clearly greater than those with a CNN-based scheme, indicating that those schemes cannot compete with the CNN-based scheme in differentiating HIFs from normal transients. The test results shown in Table 1 validate this assumption: the accuracy values of the SVM-based and MLP-based schemes are 94.5% and 94.8%, respectively, and the respective F1 scores are 93.6% and 94.0%. In contrast, the CNN-based scheme has a much higher accuracy (98.9%) and F1 score (98.7%). Therefore, the proposed CNN model can better detect the HIFs than traditional SVM or MLP methods.

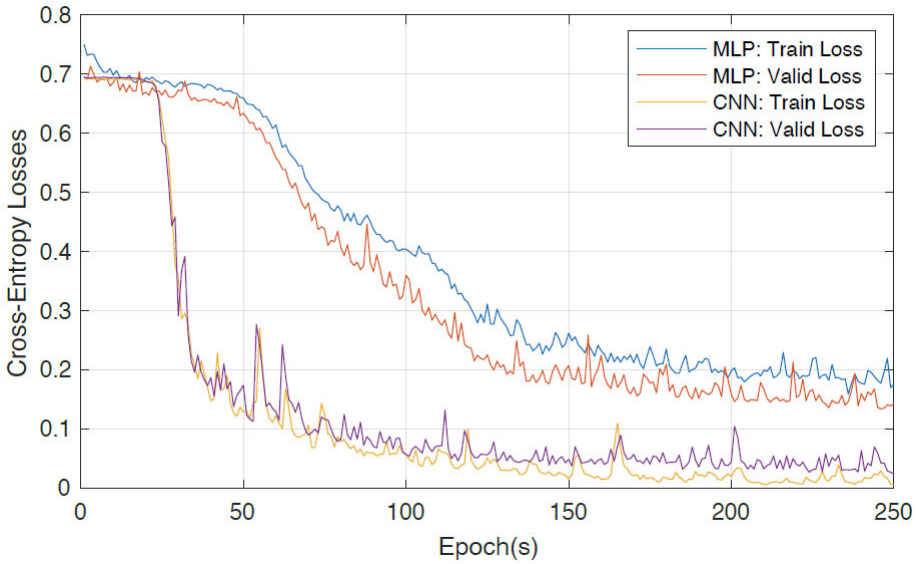


Figure 20. Cross-entropy losses of different methods for HIF

Table 1. Test results for HIFs

	SVM	**MLP**	**CNN**
True Positive T_P	366	370	393
False Positive F_P	16	17	3
False Negative F_N	34	30	7
True Negative T_N	487	486	500
Accuracy	94.5%	94.8%	98.9%
F1 score	93.6%	94.0%	98.7%

4.2 Transformer Inter-Turn Fault

To simulate the transformer inter-turn faults, we used a 20 MVA 115/25 kV three-phase saturable-core transformer with an inter-turn fault model. We generated around 3,000 transient scenarios using this test system. The normal events include load changes, capacitor bank changes, DER generation, and induction motor operation. We have also simulated several external faults outside the protection zone (i.e., the transformer). Because these transients should not trip the transformer breaker, they were also included among the normal transients. As with the first protection gap, we also used various system conditions and randomly selected the fault and normal transient parameters during the simulation. The transformer fault location varied on both primary and secondary windings, with a fault level between 1% and 8%. The input measurements for the three methods were the time-domain sampled phase voltage and current

measurements of the transformer. The sampling rate was 20 kHz, and the input data length was two cycles. In addition, 2% white noise was added to the data to mimic real-world situations. From the generated transients, we randomly selected around 2,400 cases for training and 600 cases for testing.

During the training of the CNN and MLP model, we used 10% of the data for validation (to prevent overfitting). The cross-entropy losses of MLP and CNN methods during the training process are shown in Figure 21. Again, the SVM method was not included in Figure 21 because it does not use training/validation algorithms as neural networks do. The CNN model has much smaller losses than the MLP, indicating the CNN-based scheme is more accurate in differentiating inter-turn faults from normal transients. The test results shown in Table 2 validate this assumption: the accuracy values of the SVM-based and MLP-based schemes are 94.8%, and 95.6%, and the respective F1 scores are 93.9% and 94.7%. In contrast, the CNN-based scheme has a much higher accuracy (99.3%) and F1 score (99.2%). Therefore, the proposed CNN model can better detect transformer inter-turn faults than traditional SVM or MLP methods.

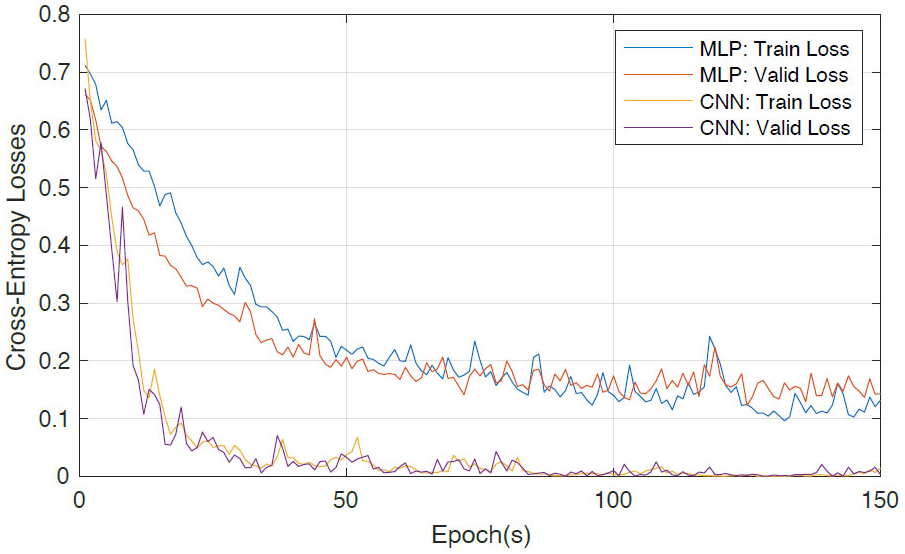


Figure 21. Cross-entropy losses of different methods for transformer inter-turn fault

Table 2. Test results for transformer inter-turn faults

	SVM	**MLP**	**CNN**
True Positive T_P	248	239	255
False Positive F_P	25	11	4
False Negative F_N	7	16	0
True Negative T_N	330	344	351
Accuracy	94.8%	95.6%	99.3%
F1 score	93.9%	94.7%	99.2%

4.3 PV Distribution System Faults

To simulate the PV distribution faults, we used the IEEE 13-Bus system. We generated around 10,000 transient scenarios using this test system. The normal events include load changes, capacitor bank changes, transformer tap changes, and induction motor operation. We also used various system conditions and randomly selected the fault and normal transient parameters during the simulation. The distribution system fault location varied all over the distribution feeder. The input measurements for the three methods (CNN, MLP and SVM) were the time-domain sampled current measurements at the relays. The sampling rate was 15 kHz, and the input data length was two cycles. In addition, 2% white noise was added to the data to mimic real-world situations. From the generated transients, we randomly selected around 8,500 cases for training and 1,500 cases for testing.

During the training of the CNN and MLP model, we used 10% of the data for validation (to prevent overfitting). The cross-entropy losses of MLP and CNN methods during the training process are shown in Figure 22. Cross-entropy losses **of** different methods for PV distribution system fault. Again, the SVM method was not included in Figure 22 because it does not use training/validation algorithms as neural networks do. The CNN model has much smaller losses than the MLP, indicating the CNN-based scheme is more accurate in differentiating PV distribution faults from normal transients. The test results shown in Table 3. Test results for PV distritbutuin system faults validate this assumption: the accuracy values of the SVM-based and MLP-based schemes are 97.6%, and 97.9%, and the respective F1 scores are 97.2% and 97.6%. In contrast, the CNN-based scheme has a much higher accuracy (99.5%) and F1 score (99.4%). Therefore, the proposed CNN model can better detect PV distribution system faults than traditional SVM or MLP methods.

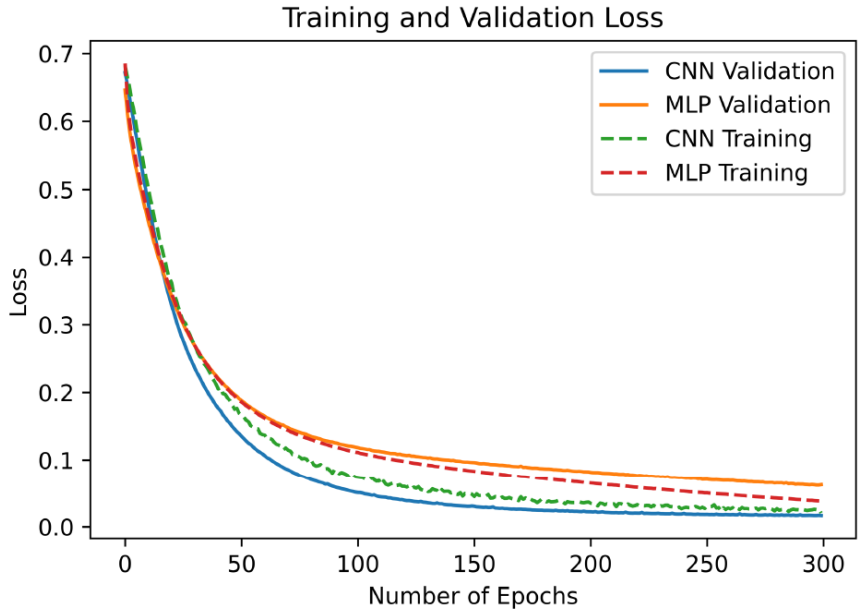


Figure 22. Cross-entropy losses of different methods for PV distribution system fault

Table 3. Test results for PV distritbutuin system faults

	SVM	**MLP**	**CNN**
True Positive T_P	591	587	604

False Positive F_P	15	6	1
False Negative F_N	19	23	6
True Negative T_N	815	824	829
Accuracy	97.6%	97.9%	99.5%
F1 score	97.2%	97.6%	99.4%

4.4 Zone 3 Protection Misoperation

To simulate the Zone 3 Protection situations, we generated around 6,000 transient scenarios using the test system. The normal events include severe loading situations, power swings, and out-zone external faults. We also used various system conditions and randomly selected the fault and normal transient parameters during the simulation. The input measurements for the three methods (CNN, MLP and SVM) were the time-domain sampled voltage and current measurements at the relays. The sampling rate was 15 kHz, and the input data length was two cycles. In addition, 2% white noise was added to the data to mimic real-world situations. From the generated transients, we randomly selected around 5,000 cases for training and 1,000 cases for testing.

During the training of the CNN and MLP model, we used 10% of the data for validation (to prevent overfitting). The cross-entropy losses of MLP and CNN methods during the training process are shown in Figure 23. Again, the SVM method was not included in Figure 23 because it does not use training/validation algorithms as neural networks do. The CNN model has much smaller losses than the MLP, indicating the CNN-based scheme is more accurate in determining if a Zone 3 operaiton is required. The test results shown in Table 4. Test results for Zone 3 Protection Misoperationvalidate this assumption: the accuracy values of the SVM-based and MLP-based schemes are 94.7%, and 95.9%, and the respective F1 scores are 94.6% and 95.9%. In contrast, the CNN-based scheme has a much higher accuracy (99.9%) and F1 score (99.9%). Therefore, the proposed CNN model can better determine if the Zone 3 operation is required than traditional SVM or MLP methods.

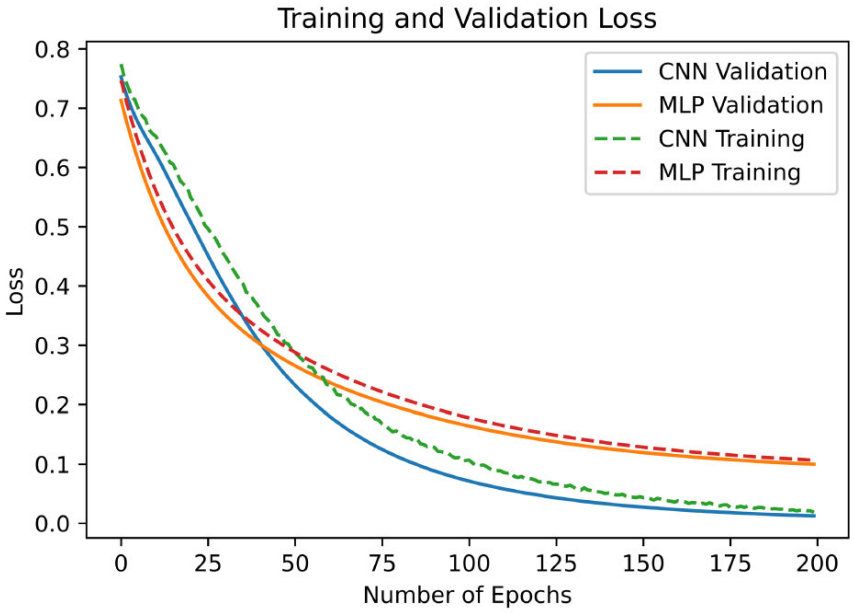


Figure 23. Cross-entropy losses of different methods for Zone 3 Protection Misoperation

Table 4. Test results for Zone 3 Protection Misoperation

	SVM	**MLP**	**CNN**
True Positive T_P	492	499	526
False Positive F_P	21	15	0
False Negative F_N	35	28	1
True Negative T_N	503	509	524
Accuracy	94.7%	95.9%	99.9%
F1 score	94.6%	95.9%	99.9%

5.0 Transfer Learning

The ultimate goal of this research is to use the proposed approach to handle real-world protection gaps. Training a neural network usually requires large amounts of data. However, available real-world protection gap data is usually limited, which presents a common barrier for ANN-based schemes in practical applications.

To address this problem, we propose to use the transfer learning technique in our data-driven approach. The basic concept of transfer learning is illustrated in Figure 24. When we have two different learning tasks (I and II), instead of training a neural network from the beginning (random initialization), we can leverage the learning experience from Task I and transfer it to the training process in Task II. Denote the Task I domain as DI and it consists of a feature space XI and a marginal probability P(XI)(DI ={XI,P(XI)}); the Task I TI consists of a label space yI and an objective predictive function f(·). Transfer learning aims to leverage the information in a new domain DII and improve the objective function f(·) for new task TII by using the knowledge in DI and TI, where DI ≠ DII or TI ≠ TII.

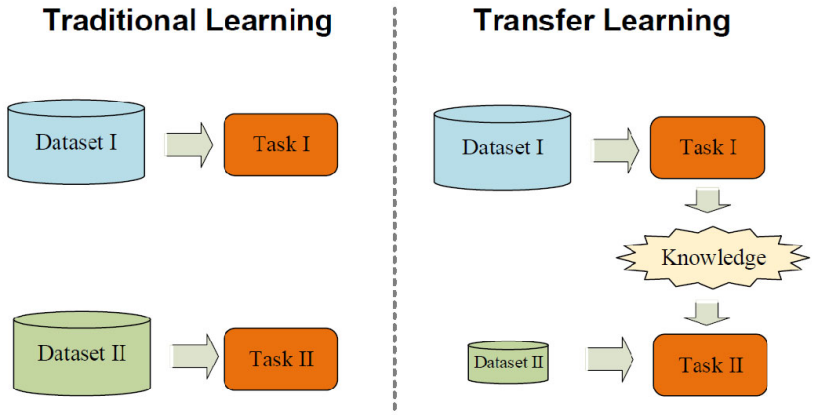


Figure 24. Traditional learning versus transfer learning

Transfer learning can be compared to a toddler having learned to tell chicken eggs from chickens (as Task I), he or she can easily learn to tell duck eggs from ducks (Task II) without needing much specific information about ducks and duck eggs. Though Task I (chicken eggs and chickens) and Task II (duck eggs and ducks) are different, the two tasks share some common features and patterns. The toddler can easily use the knowledge learned in Task I to complete learn Task II.

We first trained the CNN model in a system with abundant simulated data, which enabled it to capture the features and patterns of protection gaps. Then, when encountering similar or identical practical protection gaps with limited data, we can use the trained CNN model with partially frozen parameters to solve the problem via transfer learning. Because the training process is like fine-tuning with minimal changes, it requires much fewer data and the training speed is usually much faster.

We tested the transfer learning method in a new system, the IEEE 13-bus feeder 4.2 kV test system. The protection gap is still a HIF, but the HIF model is different from the one used in Section II. Instead of the antiparallel sources, diodes, and variable resistors shown in Figure 1, we used transient analysis of control systems for the HIF model. We also applied various system and HIF parameters and conditions to generate the data; the fault impedance varied from 10 to 200 ohms, and the conducting voltage varied from 60% to 80%. Because the test system and the HIF model were different from those used for Protection Gap I, the new generated data were different from the previous data.

However, this time we had far fewer data (≈240 sets). Half the data was used to train the CNN model, and the other data was used for testing. We compared the results of transfer learning, which leveraged knowledge from the previous Protection Gap I against the results from conventional machine learning, where the model was trained from scratch (random initialization).

The cross-entropy losses for training the CNN model are shown in Figure 25. We still used 10% of the training data for validation and terminated the training right before overfitting occurred for the best performance of both methods. Using transfer learning clearly resulted in much smaller cross-entropy losses. Test results for random initialization and transfer learning are shown in Table 5. The accuracy and F1 score for random initialization are as low as 54.6% and 55.7%. This means it could not learn the model from scratch with so few data, and its performance was nothing but random guessing. In contrast, transfer learning provides a much higher accuracy (97.5%) and F1 score (97.7%). Therefore, transfer learning is very effective to overcome the practical data limitation challenge.

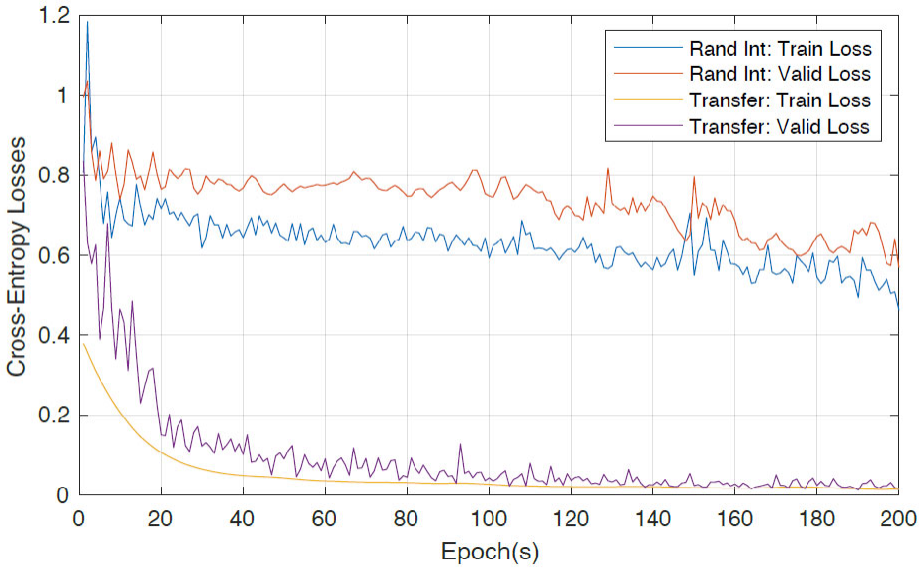


Figure 25. Cross-entropy losses from transfer learning and from random initialization

Table 5. Test results for transfer learning and for random initialization

	Random initialization	**Transfer learning**
True Positive T_P	34	63
False Positive F_P	23	1
False Negative F_N	31	2
True Negative T_N	31	53
Accuracy	54.6%	97.5%
F1 score	55.7%	97.7%

6.0 Conclusion

In this project we have proposed a new data-driven approach to bridge existing power system protection gaps. The new approach uses a CNN model to accurately differentiate fault events from normal transients. We have chosen four protection cases that represent typical protection gaps: high impedance faults, transformer inter-turn faults, PV circuit faults, and the mioperation situations of Zone 3 line protection relays operating under system stress. Results have proven the proposed method performs better than traditional data-driven approaches. The proposed approach also applies transfer learning to overcome the practical data limitation challenge, and its effectiveness has been demonstrated through comparison to traditional training methods.

7.0 References

[1] A. Y. Abdelaziz, S. F. Mekhamer, M. Ezzat, and E. F. El-Saadany, “Line outage detection using support vector machine (SVM) based on the phasor measurement units (PMUs) technology,” in Proc. IEEE Power Energy Soc. Gen. Meeting, San Diego, CA, USA, Jul. 2012, pp. 1–8.

[2] Y. Mohammadnian and et al., “Fault detection in distribution networks in presence of distributed generations using a data miningdriven wavelet transform,” IET Smart Grid, vol. 2, no. 2, pp. 163–171, 2019.

[3] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” nature, vol. 521, no. 7553, p. 436, 2015.

[4] V. Nair, G. E. Hinton, Rectified linear units improve restricted boltzmann machines, in: Pro-ceedings of the 27th International Conference on International Conference on Machine Learning,ICML’10, Omnipress, Madison, WI, USA, 2010, p. 807–8

Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by **Battelle** *Since 1965*

U.S. DEPARTMENT OF
ENERGY

902 Battelle Boulevard
P.O. Box 999
Richland, WA 99352
1-888-375-PNNL (7665)

www.pnnl.gov