EE394V Data Analytics in Power Systems
SVM Based Distribution Relays

Soham Roy and Dhruv Kumar Gupta

May 10, 2019

Abstract

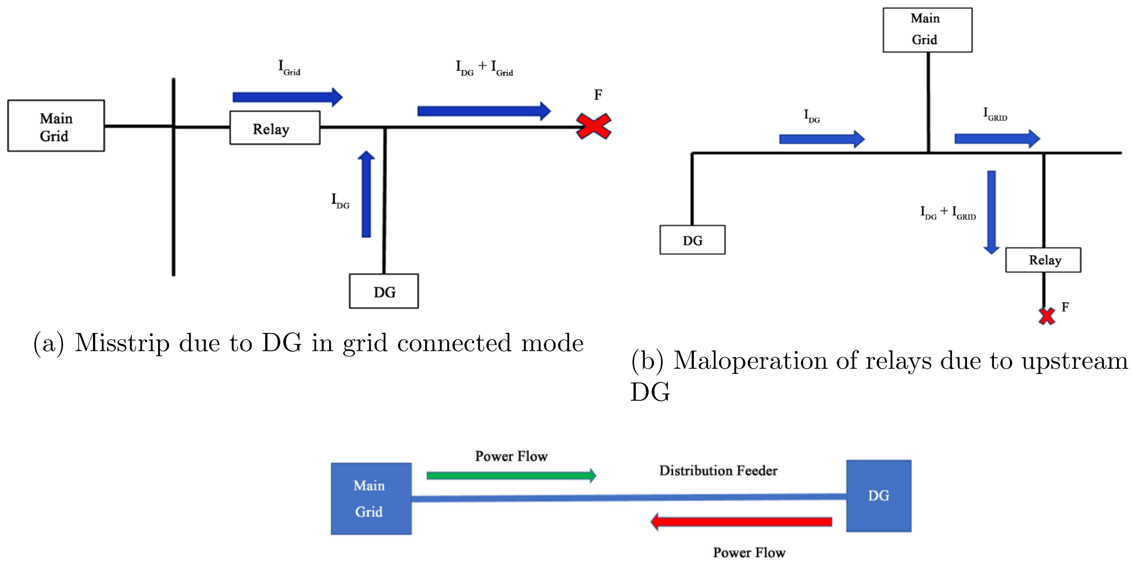
An increase in the penetration of Distributed Energy Resources (DERs) and Distributed Generation (DG) like solar and wind power systems is leading to diversification of power flow patterns. This calls for employing smart, data-driven, real-time settings in protection relays. Intelligent relay design in the transmission grid was proposed in [2], and an SVM-based approach for relay settings in the distribution system has been explored in [1]. Our project was targeted at implementing the SVM algorithm for training and testing relay settings in the simulation of some distribution system networks.

1 Introduction

Relays in conventional power systems involving centralized generation are based on the over-current (OC) principle and depend on prior knowledge of generation/load levels and system topology, which become more and more uncertain due to DERs. Most conventional relays work on thresholds set by the power system operators in conventional power systems before DERs came into play. With an increase in DG, the generation can become intermittent due to variation in solar and wind conditions. Moreover, the interfacing of such energy sources to the main grid using power electronics causes problems such as frequency fluctuation and harmonics. This leads to changing fault current characteristics [4].

Conventional OC relay thresholds are usually set based on two assumptions: (1) short circuit current (SCC) is much greater than current in healthy conditions, and (2) faults closer to the relay lead to higher measured SCC. These assumptions do not always hold true in modern power systems.

It is known that DG sources have much lower short circuit levels than the conventional grid, that is, they are much weaker sources. As shown in Figure 1(a), the current in a fault scenario comprises the sum of contributions from the grid, I_{grid} and from the DG installed in a downstream lateral, I_{DG}. But the upstream relay only reads I_{grid} and might not consider it as a fault. On the other hand, a DG installed further upstream (e.g. a solar farm) can lead to *maloperation* of relays downstream, as shown in Figure 1(b). Here, the relay falsely reads the current as a sum of I_{grid} and I_{DG}, considers it as a fault and sends a trip signal when it should not. One more issue that may arise with DG is that it makes the conventionally unidirectional power flow in distribution feeders bidirectional, and conventional relays are not trained to handle this [8]. This is depicted in Figure 1(c).



(a) Misstrip due to DG in grid connected mode

(b) Maloperation of relays due to upstream DG

(c) Bidirectional power flow in distribution feeders due to DG

Figure 1: Shortcomings of conventional relays in presence of DG

Artificial intelligence and machine learning have been gaining increasing popularity in power systems applications over the past years [10]. With the advent of data analytics, there is now a possibility of using classification algorithms like Support Vector Machine (SVM) for implementing smart decision-making by relays. The SVM algorithm has shown its effectiveness in other power systems studies like load forecasting [6] and electric market data analysis [7]. This data-driven approach is expected to be more accurate with lower false alarm rates. Changes in topology and load levels could also be incorporated in this kind of operation by using real-time SVM models through data streaming [1].

1.1 Support Vector Machine

In machine learning, a classification problem involves identifying the group or class to which a new observation belongs, based upon a training data set consisting of observations with known group associations. Such an approach, where we have a set of correctly identified observations as the training set, is called supervised learning [3].
The SVM algorithm classifier constructs a set of hyperplanes with the objective to maximize the separation (margin) between the groups. Contrary to perceptron learning, SVM aims at improving the confidence by minimizing the generalization error of the classifier.

1.1.1 Linearly Separable Case: Hard Margin

If a set of p-dimensional vectors can be separated by a $(p-1)$-dimensional hyperplane, the dataset is said to be linearly separable.

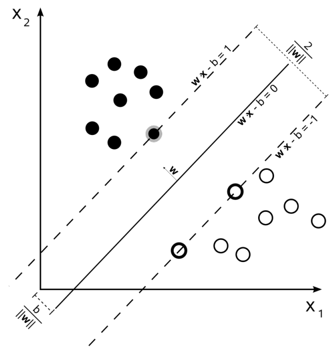


Figure 2: SVM operating principle

Suppose we are given a training data set of n points, $(x_1, y_1), ..., (x_n, y_n)$ where $y_i \in \{-1, 1\}$.
Let $f(x) = w^T x + b = 0$ represent the separating hyperplane as shown in Figure 2.
For feature vectors x_i with label $y_i = 1$:

$$w^T x_i + b \geqslant 1 \tag{1.1}$$

For feature vectors x_i with label $y_i = -1$:

$$w^T x_i + b \leqslant 1 \tag{1.2}$$

Thus for all $y \in \{-1, 1\}$:

$$y(w^T x + b) \geqslant 1 \tag{1.3}$$

The width of the separating hyperplane (i.e. $\perp$ distance between $w^T x + b = 1$ and $w^T x + b = -1$) is given as $\frac{2}{\|w\|}$.

SVM Optimization Problem:

$$\min_{w,b} \frac{1}{2} w^T w \tag{1.4}$$

$$s.t.\ y_i(w^T x_i + b) \geqslant 1, \qquad \text{for } i = 1, 2, ..., n \tag{1.5}$$

1.1.2 Non-linearly Separable Case

For cases where observations are not linearly separable, the algorithm can be modified either to account for the classification error or by using a non-linear boundary.

1.1.2.1 Soft Margin

This concept introduces a slack variable ($\xi_i \geqslant 0$) to each observation(x_i, y_i) such that:

$$y_i(w^T x_i + b) \geqslant 1 - \xi_i, \qquad \text{for } i = 1, 2, ..., n \tag{1.6}$$

$$\text{where, } \xi > 0 \rightarrow \text{Misclassification} \tag{1.7}$$

$$\xi = 0 \rightarrow \text{Correct classification} \tag{1.8}$$

Updated SVM Optimization Problem:

$$\min_{w,b} \frac{1}{2}w^T w + C\sum_{i-1}^{n} \xi_i \tag{1.9}$$

$$s.t.\ y_i(w^T x_i + b) \geqslant 1 - \xi_i, \qquad \text{for } i = 1, 2, ..., n \tag{1.10}$$

$$\text{and } \xi_i \geqslant 0, \qquad \text{for } i = 1, 2, ..., n \tag{1.11}$$

Here, C is a scalar value and a smaller C implies a softer margin.

1.1.2.2 Kernel Method

This algorithm makes use of a non-linear kernel function instead of taking the dot product of the observations. This allows the SVM to create a separating hyperplane in a transformed high-dimensional feature space which is able to classify observations which may not be linearly separable in the original input space.
Taking the Lagrangian Dual of the SVM problem:

$$\min_a L = \frac{1}{2}w^T w + \sum_{i=1}^{n} a_i[y_i(w^T x_i + b) - 1] \tag{1.12}$$

$$s.t.\ a_i \geqslant 0, \qquad \text{for } i = 1, 2, ..., n \tag{1.13}$$

Using Karush-Kuhn-Tucker (KKT) Condition:

$$\sum_{i=1}^{n} a_i y_i x_i = w,\ \sum_{i=1}^{n} a_i y_i = 0 \tag{1.14}$$

Combining the 2 sets of equations:

$$\max_a \sum_{i=1}^{n} a_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j y_i y_j \Phi(x_i)^T \Phi(x_j) \tag{1.15}$$

$$\text{s.t.} \sum_{i=1}^{n} a_i y_i = 0 \tag{1.16}$$

$$0 \leq a_i \leq C, \text{for } i = 1, 2, ..., n \tag{1.17}$$

To apply the kernel trick, the inner product$(x_i^T x_j)$ in eq.(2.15) is generalized to be $\Phi(x_i, x_j)$ Some common Kernel functions include:

- Polynomial (homogeneous): $\Phi(x_i, x_j) = (x_i, x_j)^d$
- Polynomial (inhomogeneous): $\Phi(x_i, x_j) = (x_i, x_j + 1)^d$
- Gaussian/Radial basis function (RBF): $\Phi(x_i, x_j) = exp(-\gamma\|x_i - x_j\|^2)$ for $\gamma > 0$
- Sigmoid: $\Phi(x_i, x_j) = \tanh(kx_i^T x_j - \delta)$ where $k > 0$ and $c < 0$

2 Problem Formulation

We use simulation data to train the Support Vector Machine (SVM) model for the main relays of each line such that they would give minimum rates of *maloperation* (relay wrongly predicts a

Algorithm 1 Fault classification using Gaussian Kernel SVM

Require: $(x_i, y_i) = ((P_i,\, Q_i), y_i)$, for each training data point and $x_j = (P_j,\, Q_j)$ for each testing data point

Find the optimal solution to (1.15) s.t. (1.16) and (1.17), where $\Phi(x_i)^T\Phi(x_j) = k(x_i, x_j) = \exp(-\gamma||x_i - x_j||^2)$.

Find the set of support vectors given by $a_i > 0 : w = \sum_{i:a_i>0} a_i y_i \Phi(x_i)$

Find b using $y_i(w^T\Phi(x_i) + b) = 1$

for testing data points $j = 1, 2, \ldots,$ **do**

  Find the prediction $f(x_j) = w^T\Phi(x_j) + b = \sum_i a_i k(x_j, x_i) + b$.

end for

healthy condition as a fault) and *misstrip* (relay predicts a fault as a healthy condition). In an actual distribution system, historic data may also be used. We stick to the popular choice of a Gaussian (RBF) Kernel which is known to have a high probability of finding the optimal hyperplane with few parameters and less numerical issues than other kernels [5]. The feature vector x_i for each line is of 12 dimensions consisting real power (P) and reactive power (Q) data, or voltage (V) and current (I) data. The true label y_i for each training/testing data point is 1 if it is an emergency fault condition and -1 otherwise. The predictions $f(x_i)$ for testing data points try to match y_i to the extent possible. *Maloperations* are false positives, i.e. the testing data points with y_i = -1 and $f(x_i)$ = 1 while *misstrips* are false negatives, i.e. those with y_i = 1 and $f(x_i)$ = -1. The fault classification algorithm is summarized in Algorithm 1.

3 Simulation Methodology

Our work used the data from various simulated scenarios of standard IEEE distribution systems (4Bus-YY-Bal, 13Bus, 34Bus) on OpenDSS. The generation of data and training/testing of SVM was done on MATLAB. The GridPV toolbox was used to interact with OpenDSS.

3.1 Data generation

For each test case, a particular line was chosen to simulate faults. Monitors, representative of data collection devices like PMUs, were placed on both ends of the chosen line. In order to generate multiple cases, the loads in the entire system were varied between 80% to 200% of their respective nominal load levels, with the assumptions that loads are independent and identically distributed. Further, the most common type of fault: an almost bolted (fault resistance $R_f = 0.1m\Omega$) Single Line to Ground (SLG) fault on each phase was simulated at the mid point of the particular line in order to generate fault data. The studies performed include:

- Training the relays with SVM using real power (P) and reactive power (Q) measurements as input parameters: denoted as SVM_{PQ}
- Training the relays with SVM using voltage (V) and current (I) measurements as input parameters: denoted as SVM_{VI}
- Studying the effect of increasing R_f on the SVM classification
- Addition of DG at some buses in the standard circuits and studying their effect. DG is

modeled as a *Generator* of *Mode=7* on OpenDSS, such that it supplies rated current whenever the voltage at its bus falls below 0.95 pu [9].

- Comparing the accuracy of the trained SVM relays with the conventional OC relays

Each case contains 3429 healthy cases and 21 fault cases. Both healthy and fault data comprise a set of 12-dimensional feature vectors and a set of labels.

3.2 Training and testing of SVM based relays

The total generated data for each case was randomly divided into training and test datasets in the ratio 80:20. The training data was first used to train the SVM using the MATLAB function *fitcsvm* with *KernelFunction* as *Gaussian* and *KernelScale* as *auto*. The testing data was used to verify the accuracy of the SVM and calculate rates of *maloperation* and *misstrip*.

3.3 Thresholding of conventional relays (OC Relays)

The minimum current threshold for conventional relays to classify a condition as a fault was chosen for a fault occurring at the end of the line with $R_f = 10\Omega$ and a load level of 120%. This setting is kept constant throughout the experiment as per the assumption that the relay has been programmed to operate before the inclusion of DG into the network.

4 Numerical Experiments

4.1 IEEE 4-Bus YY-Balanced System

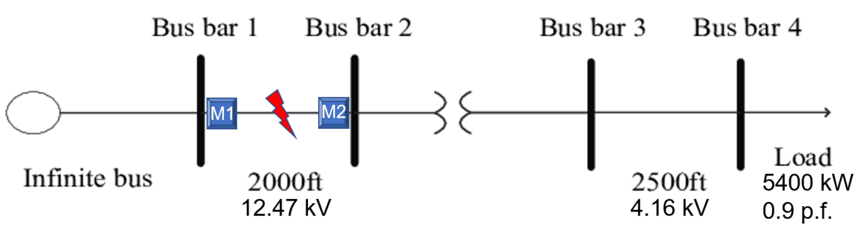
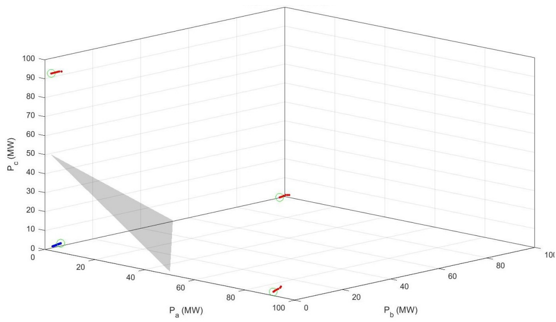


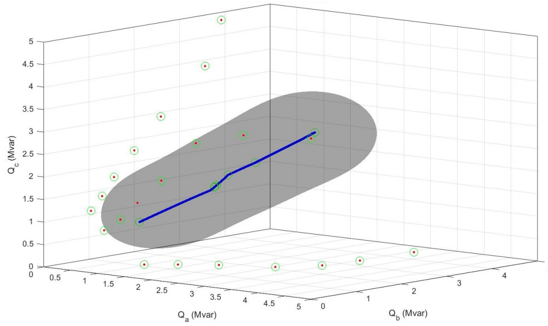
Figure 3: IEEE 4-bus network

The network is shown in Figure 3. The midpoint of line 1 was chosen as the fault location, and data from monitors M1 and M2 (at the two ends of the line) was used. For better visualization, only the real power from M1 and reactive power from M2 for each phase have been plotted in Figure 4, because it not possible to display a hyperplane for 12-dimensional vectors. As can be seen from Figure 4(a), the real power data from M1 is linearly separable and therefore a linear kernel is shown to separate fault data points from the healthy ones. The reactive power data from M2 shown in Figure 4(b) is however, not linearly separable and a Gaussian Kernel is shown for the purpose of classification.

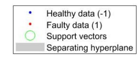
On comparison with the Conventional (OC) Relay for this network, the SVM based relays are seen to outperform in terms of prediction accuracy and *maloperations* as seen in Table 1.



(a) Monitor M1 - Real Power



(b) Monitor M2 - Reactive Power



(c) Legend

Figure 4: Visualization of healthy and fault data (without noise) for IEEE 4-Bus YY-Bal System

R_f	Metric (%)	SVM_{PQ}	SVM_{VI}	OC relay
0.1 mΩ	Maloperations	0	0	0.87
	Misstrips	0	0	0
	Accuracy	100	100	99.13
10 Ω	Maloperations	0	0	0.44
	Misstrips	0	0	0
	Accuracy	100	100	99.56

Table 1: IEEE 4-Bus YY-Balanced System

4.2 IEEE 13-Bus System

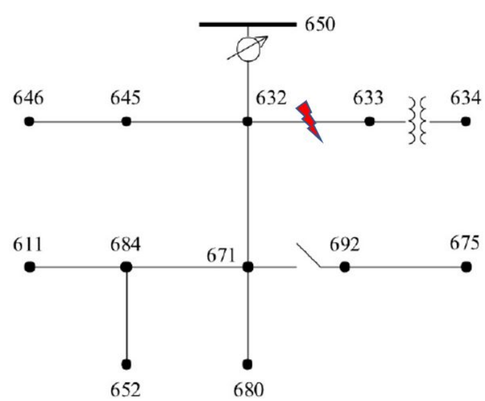
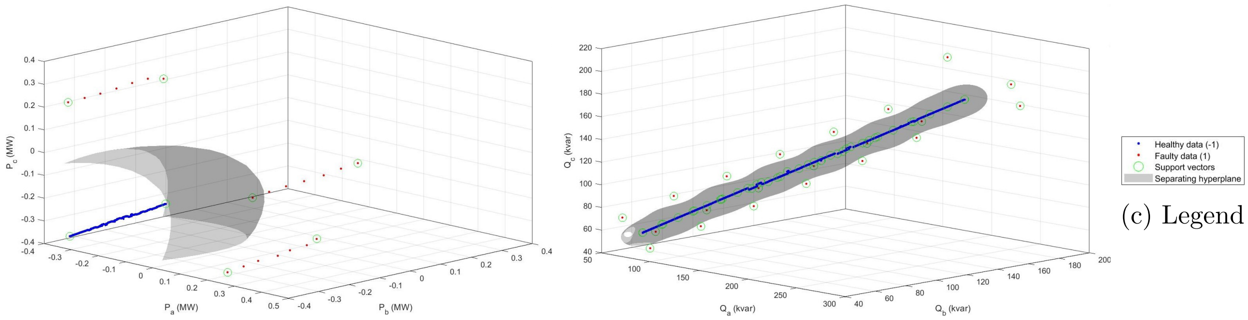


Figure 5: IEEE 13-bus network

The line between buses 632 and 633 was chosen for placing the fault, as shown in Figure 5. Gaussian Kernels used with 3-dimensional data are shown in Figures 6(a) and 6(b) for easy visualization. Next, DG units at unity power factor were connected to each load bus with real power rating equal to 40% of the load at their respective buses. The original system and the one with DG were both tested with varying R_f and the results are documented in Table 2. It can be seen that SVM_{VI} has an alarming 100% misstrip rate at high R_f. This shows that P and Q are better features for SVM classification than V and I since they are more easily separable. Moreover, P and Q contain information about both V and I. Therefore, only SVM_{PQ} is used for further study.



(a) Monitor M1 - Real Power

(b) Monitor M2 - Reactive Power

(c) Legend

Figure 6: Visualization of healthy and fault data (without noise) for IEEE 13-Bus System

R_f	Metrics (%)	SVM_{PQ}	SVM_{VI}	SVM_{PQ} with DG	OC relay with DG
0.1 mΩ	Maloperations	0	0	0	0.44
	Misstrips	0	0	0	0
	Accuracy	100	100	100	99.56
10 Ω	Maloperations	0	0	0	0.73
	Misstrips	0	100	0	0
	Accuracy	100	99.27	100	99.27

Table 2: IEEE 13-Bus System

4.3 IEEE 34-Bus System

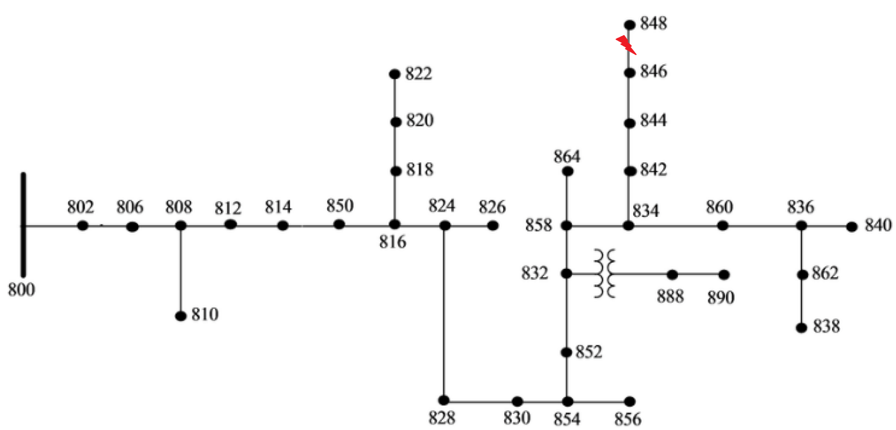
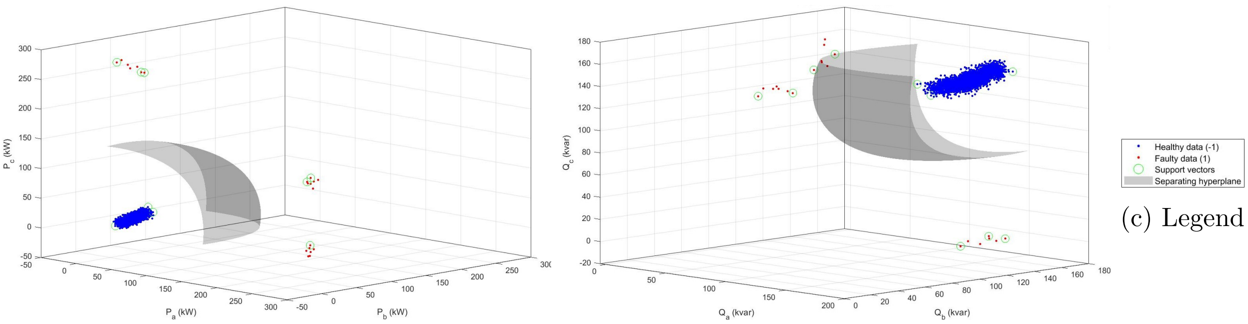


Figure 7: IEEE 34-bus network

The highest impedance (electrically weakest) three-phase bus was found to be bus 848 in Figure 7, using the GridPV function *highestImpedanceBus*. The line L23b connected to this bus was chosen to simulate faults. DG was later added to each load bus with 40% penetration, similar to the 13-bus case. Gaussian measurement noise of 10dB was also included in all cases to check the effects on the proposed relay accuracy. Figures 8(a) and 8(b) employ Gaussian Kernels with 3-dimensional data for easy visualization. From Table 3, it can be seen that SVM-based relays again perform better than conventional relays.



(a) Monitor M1 - Real Power

(b) Monitor M2 - Reactive Power

(c) Legend

Figure 8: Visualization of healthy and fault data (with 10dB noise) for IEEE 34-Bus System

R_f	Metrics (%)	SVM_{PQ}	SVM_{PQ} with DG	OC relay with DG
0.1 mΩ	Maloperations	0	0	0
	Misstrips	0	0	0
	Accuracy	100	100	100
10 Ω	Maloperations	0	0	0.29
	Misstrips	0	0	0
	Accuracy	100	100	99.71

Table 3: IEEE 34-Bus System (with 10dB noise)

5 Concluding Remarks

Conventional power system protection schemes were designed for unidirectional flow of power. However, with the onset of DERs, they face issues like *maloperation* and back feeding, and thus need to be upgraded to cater to this constantly changing network. In this project we explored the use of SVM, a machine learning technique to train the classifier for the main relay on a particular line, which can better identify faults as compared to conventional relays with DERs present in the system. The algorithm was tested on multiple test cases (IEEE 4-Bus-YY-Bal, 13-Bus and 34-Bus networks) and successfully proved its effectiveness even with varying load levels, fault resistances, measurement noise and DER penetration.

SVM-based distribution relays can have multiple advantages: (1) adaptability to changing topology and load levels, (2) better fault prediction - minimizing *misstrips* and unnecessary outages and (3) ability to be updated in real time using streaming data from PMUs.

This novel approach however has some downsides also: (1) it can require an expensive central computer and synchronous communications network for updating the classifier models in real time, and (2) it does not include a scheme for backup relays.

While in most of these simple systems 3-dimensional data of real power (P) or reactive power (Q) from one monitor alone was sufficient for classification, it is expected that using 12-dimensional data in more complex systems would prove to be substantially more beneficial. In each case, the accuracy of the SVM classifier based on P and Q was seen to be 100%, but it is expected to be slightly lower in actual work distribution systems which are much more complex. The SVM approach can be extended to training the classifier for the main relay of every line in the system with the help of more extensive computational resources.

Acknowledgements

We would like to thank Dr. Hao Zhu for her constant guidance and support throughout this project, and Sundaravaradan Navalpakkam Ananthan for his help with fault studies using OpenDSS and GridPV toolbox.

References

[1] X. Zheng, X. Geng, L. Xie, D. Duan, L. Yang and S. Cui, "A SVM-based Setting of Protection Relays in Distribution Systems," in *IEEE Texas Power and Energy Conference (TPEC)*. IEEE, 2018, pp. 1-6.

[2] B. Ravikumar, D. Thukaram, and H. Khincha, "Comparison of multiclass SVM classification methods to use in a supportive system for distance relay coordination," *IEEE Transactions on Power Delivery*, vol. 25, no. 3, pp. 1296-1305, 2010.

[3] "Statistical classification," *Wikipedia* [Online] Available: https://en.wikipedia.org/wiki/Statistical_classification

[4] S P George, S Ashok, M N Bandyopadhyay, "Impact of distributed generation on protective relays," in *2013 International Conference on Renewable Energy and Sustainable Energy (ICRESE)*. IEEE, 2013, pp. 1-5.

[5] Y. Zhang, M. D. Ilic, and O. K. Tonguz, "Mitigating blackouts via smart relays: a machine learning approach," *Proceedings of the IEEE*, vol. 99, no. 1, pp. 94-118, 2011.

[6] H. Jiang, Y. Zhang, E. Muljadi, J. Zhang, and W. Gao, "A short-term and high-resolution distribution system load forecasting approach using support vector regression with hybrid parameters optimization," in *IEEE Transactions on Smart Grid*, 2017.

[7] X. Geng and L. Xie, "A data-driven approach to identifying system pattern regions in market operations, in *IEEE Power & Energy Society General Meeting*, IEEE, 2015, pp. 15.

[8] U. Shahzad, S. Kahrobaee and S. Asgarpoor, "Protection of Distributed Generation: Challenges band Solutions", in *Energy and Power Engineering*, vol. 9, 2017, pp 614-653.

[9] J. Keller and B. Kroposki, "Understanding Fault Characteristics of Inverter-Based Distributed Energy Resources", *NREL Technical Report* 550-46698, January 2010.

[10] B. F. Wollenberg and T. Sakaguchi, "Artificial intelligence in power system operations," *Proceedings of the IEEE*, vol. 75, no. 12, pp. 1678-1685, 1987.