Proceedings of the 2022 Winter Simulation Conference
B. Feng, G. Pedrielli, Y. Peng, S. Shashaani, E. Song, C.G. Corlu, L.H. Lee, E.P. Chew, T. Roeder, and P. Lendermann, eds.

MACHINE LEARNING BASED SIMULATION FOR FAULT DETECTION IN MICROGRIDS

Joshua Darville
Temitope Runsewe
Abdurrahman Yavuz
Nurcin Celik

University of Miami
Department of Industrial and Systems Engineering
1251 Memorial Dr
Coral Gables, FL 33146, USA

ABSTRACT

Fault detection (FD) is crucial for a functioning microgrid (MG) but is particularly challenging since faults can stay undetected indefinitely. Hence, there is a need for real-time, accurate FD in the early phase of MG operations to mitigate small initial deviations from nominal conditions. To address this need, we propose an FD framework for MG operational planning. Our proposed framework is synthesized from i) a dataset generated by introducing faults into an MG with PV cells, ii) processing the dataset to train various machine learning (ML) models for FD, iii) benchmarking the resulting FD models using classification metrics, and iv) applying an appropriate fault mitigation strategy. Although noisy measurements were present during the experiment due to variations in ambient temperature and solar irradiance, our proposed FD model is shown to be both computationally efficient with an average training time of 1.76 seconds and accurate with a weighted F-score of 0.96.

ABBREVIATIONS

Acronym	Description
RES	renewable energy sources
PV	photovoltaics
MG	microgrid
ML	machine learning
FD	fault detection
MPPT	maximum power point tracking
IPPT	intermediate power point tracking
TSO	transmission system operator
US	United States

1 INTRODUCTION

Recently, renewable energy sources (RES) such as wind and solar have become ever more attractive as the rate of fossil fuel depletion and the frequency of extreme climate events increase. According to an annual electric power report in the United States (US), 3% of electricity generation in 2020 stemmed from solar power (EIA 2022).

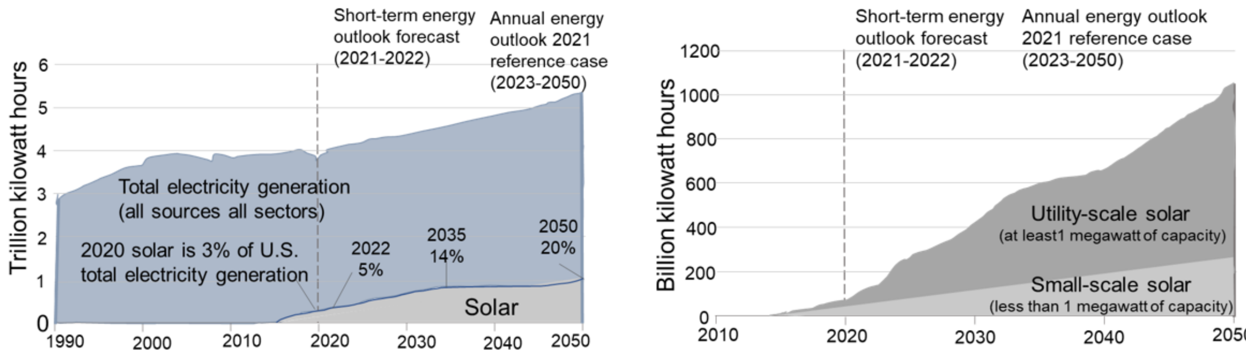

Figure 1: Annual U.S. electricity net generation between 1990-2050 for all sectors (left) and solar electricity net generation between 2010-2050 for all sectors (right).

The US Energy Information Administration also forecasted that solar power would be responsible for 4% of US electricity generation in 2021 and 5% in 2022 as shown in Figure 1 on the left. It is also projected that solar generation will produce as much as 14% of the US total energy supply in 2035 and increase to 20% in 2050. In 2011, solar at the distribution scale accounted for 68% of total U.S. solar electricity net generation. However, solar generation at the utility scale increased substantially in the US during the past decade as average construction cost for solar power plants fell. Now utilities produce the most solar generation with 68% of total solar generation in 2020 (Runsewe et al. 2020; Philippe et al. 2021). Subsequently, 13 GW of solar capacity was added in 2021 and this trend is expected to continue with an increase to 22 GW of solar capacity in 2022. Large additions to solar capacity at the utility scale are likely to continue due to declining production costs as depicted in Figure 1 on the right (EIA 2021). PV systems are susceptible to several anomalies due to their high sensitivity to external climate factors such as ambient temperature and solar irradiance. Thus, it is important to diagnose them as early as possible before large deviations from nominal operating conditions occur. PV systems can suffer from malfunctions caused by power electronics and possess very fast power dynamics which narrows the window for FD. Consequently, measurements from multiple sensors must be taken at a high enough frequency to capture the fault dynamics within an MG system. However, this high sampling rate generates ever-increasing amounts of data which may create a bottleneck for computationally intensive FD models (Sáenz et al. 2012). Since the reliability and service life of PV cells play an integral role in reducing the cost of PV systems, the aim of this study is to develop a fast and accurate machine learning (ML) model for real-time FD in the early phase of microgrid (MG) operations with PV systems. Various studies have investigated FD across a variety of power system configurations including PV generation (Darville and Celik 2020; Yavuz et al. 2020; Thanos et al. 2017; Mesham et al. 2020). Authors (Madeti et al. 2017) investigated FD in grid-connected PV systems and proposed an optimal location for electric current and voltage sensors to limit the cost increase due to the redundant nature of these devices. Likewise, a sensor-based FD analysis was also developed by (Saha et al. 2020) to detect partial shading of PV cells. PV systems containing maximum power point tracking (MPPT) and intermediate power point tracking (IPPT) controllers generally require advanced FD methods to detect faults. MPPT algorithms for partial shading such as dynamic leader based collective intelligence and memetic salp swarm algorithms are reported to be both fast and effective at reducing power losses during transmission. However, these algorithms mask signs of a fault leading to lower accuracy especially for low voltage and current values.

Statistical, ML and deep learning models are well positioned to be used for FD in PV systems since these models make inferences about the relationships between variables to make accurate predictions (Goodwin et al. 2022; Damgacioglu and Celik 2022; Wright et al. 2022). For example, (Hong and Pula 2022) proposed a 3D CNN model considering stacked 2D images (volumetric image) generated from the Gramian Angular Field transform as inputs to perform FD in PV systems. This model reflected over 90% accuracy score compared to 85% accuracy from (Benkercha et al. 2018) using a decision tree, and 85.82%

accuracy from (Chen et al. 2018) using random forest. A summary of previous FD models throughout the literature is presented in Table 1.

Table 1: Summary of previous FD models throughout the literature.

Study	Fault Type	Classification Method	Single- or Multi-classification
Hong et al. 2022	• Line-to-line • Shorted/open string in modules • Shorted/open strings in array	• Convolution neural network	• Multi classification
Benkercha et al. 2018	• Free fault • String fault • Short circuit fault • Line-line fault	• Decision tree	• Binary & Multi classification
Chen et al. 2018	• Line-line faults • Degradation • Open • Circuit • Partial shading	• Random forest	• Multi classification
Proposed	• Open circuit & Partial shading • MPPT/IPPT controller faults • Inverter fault • Current feedback sensor fault • Voltage sags	• Logistic Regression • Random Forest • Naive Bayes	• Binary classification

This study proposes a new FD framework for MG operational planning. As opposed to considering multiple types of faults within a multi-classification problem, the fault variations are aggregated into a single undesired effect on the MG system in a binary classification problem. Binary classification is used to simply detect whether there is a fault as opposed to identifying the type of fault present. This steers the ML model to focus on predicting the presence of a problem with higher accuracy as opposed to focusing on the various types of problems present. Subsequently, a transmission system operator (TSO) will identify the type of fault and employ an appropriate mitigation strategy as the predictive performance may wane from one class to another within a multi-classification problem. This approach provides a semi-automated control for an MG system where the TSO's domain knowledge is essential to appropriately rectify the type of fault or fault combinations present for a given scenario.

This article is organized as follows: Initially, section 1 presents a short description of the previous literature on fault detection. Next, section 2 outlines the proposed FD framework for MG operational planning including i) a description of each variable in the dataset, ii) the various methods used to preprocess the data for training, and iii) the types of ML models used for FD. Subsequently, section 3 discusses the results of the proposed FD framework for MG operational planning and the application of ML models for FD in MG systems. Finally, section 4 concludes the findings of our study and presents future work.

2 METHODOLOGY

In this article, we propose an FD framework for MG operational planning. Initially, the proposed FD framework uses statistical models for FD in MG systems. Here, real-time data from an MG system is used to train these FD models (Darville et. al 2022). Next, validation metrics are used to compare the FD performance of data-driven methods against real faults under both MPPT and IPPT modes.

In Figure 2, multi-sensor measurements are used to collect real-time data from an MG system. This data is then stored and preprocessed to train various ML algorithms for developing an array of FD models. Preprocessing is required to i) appropriately handle missing points, ii) encode the binary and ordinal variables, iii) ensure the classes of the response variable are balanced prior to analysis, and iv) scale all variables to a similar range to limit bias during training toward variables with a larger range than their counterparts.

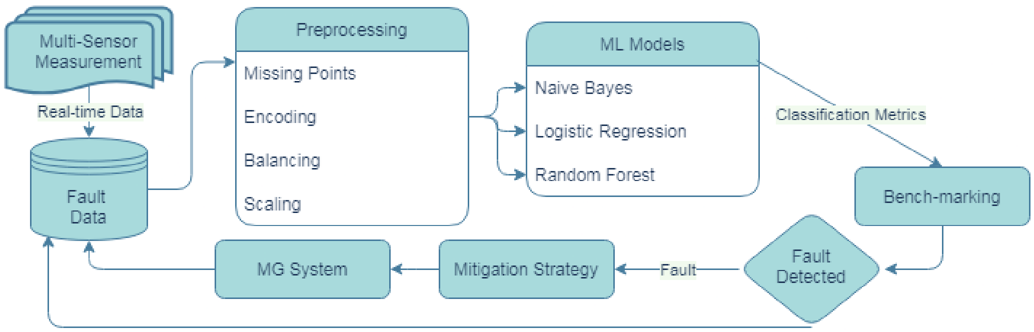

Figure 2: Proposed FD framework for MG operational planning.

The array of FD models obtained are benchmarked using classification metrics to determine the best ML model for this experimental data. After detecting a fault with high accuracy, an appropriate mitigation strategy can be implemented by the TSO to rectify the fault and restore the MG system to nominal operating conditions. Although ML models can be used for FD, their predictions must be implemented with domain knowledge from the TSO as the underlying ML algorithms learn from the data provided. Thus, understanding which variables should be used to describe a response and the various forms the response can take are pre-requisites for any data-driven method in a cyber-physical system such as an MG (Darville and Celik 2022).

2.1 Dataset Description

The Grid-connected PV System Faults (GPVS-Faults) data was collected from lab experiments by (Bakdi et al. 2021) where faults were manually introduced at various levels of severity halfway during these experiments in an MG system with PV generation. This dataset contains seven fault scenarios operating using MPPT and IPPT modes including: partial shading, open circuit, inverter, voltage sags, current feedback sensor, MPPT/IPPT controller in boost converter faults, and nominal operating conditions (no faults). Consequently, there are 16 distinct fault scenarios across both modes to train various ML models for FD in PV systems and reactive MG system maintenance. Based on the type of fault present, a TSO can employ a corresponding mitigation strategy outlined in Table 2.

Table 2: MG fault types and their respective mitigation strategies.

Fault	Type	Description	Mitigation Strategy	Strategy Source
F1	Inverter fault	Complete failure in one of the six IGBTs	Using specific inverter control (e.g., hysteresis controllers to account for the IGBT failures)	Tabbache et. al (2013)
F2	Feedback Sensor fault	One phase sensor fault 20%	Switching among different observers,	Wang et. al (2017)
F3	Grid anomaly	Intermittent voltage sags	Dynamic voltage restorer	Francis et. al (2014)
F4	PV array mismatch	10 to 20% nonhomogeneous partial shading	Bypass diodes Cell interconnection Shadow mode (string inverters) Module Level Power Electronics	Sinapis et. al (2016)

F5	PV array mismatch	15% open circuit in PV array	photovoltaic array reconfiguration methods	Yang et. al (2021)
F6	MPPT/IPPT controller fault	−20% gain parameter of PI controller in MPPT/ IPPT controller of the boost converter	Proportional Integral (PI) controller plus a phase-shift technique	Siriwat et al. (2016)
F7	Boost converter controller fault	+20% in time constant parameter of PI controller in MPPT/IPPT controller of the boost converter	Fuzzy controllers augmented using a feedforward compensation technique	Pena et. al (2002)

The faults in Table 2 are a function of the MG's state and subsequently the state variables (or predictors) used to denote the MG system are outlined in Table 3.

Table 3: Description of state variables used to denote the MG system.

Variable	Variable Type	Description
Time	Numeric	Time-real in seconds with an average sampling of 9.9989 microseconds
Ipv	Numeric	PV array current
Vpv	Numeric	PV array voltage
Vdc	Numeric	DC voltage
ia	Numeric	Phase_A current
ib	Numeric	Phase_B current
ic	Numeric	Phase_C current
va	Numeric	Phase_A voltage
vb	Numeric	Phase_B voltage
vc	Numeric	Phase_C voltage
Iabc	Numeric	Positive sequence estimated current magnitude
If	Numeric	Positive sequence estimated current frequency
Vabc	Numeric	Positive sequence estimated voltage magnitude
Vf	Numeric	Positive sequence estimated voltage frequency
Fault (Response)	Binary	"Yes or No" for fault present in {F1, F2, … , F7} in either IPPT or MPPT mode

In Table 3, domain knowledge about MG systems is required to select state variables to be considered in FD analysis. Therefore, appropriate state variables including i) time to track changes in an MG's behavior, ii) voltage and current values used to determine power output from PV arrays and at each phase within a 3-phase AC MG, iii) positive sequence components for voltage and current since they describe an MG operating under normal conditions (power flow from source to load), and iv) the resulting fault due to large variations in the previous variables (Bansal et al. 2018). After determining which state variables should be used to describe fault behavior within an MG system, these state variables are organized into a structured dataset for preprocessing and subsequently training ML models for FD.

2.2 Preprocessing Data

Preprocessing state variables is required to train FD models as previously illustrated in Figure 2. Here, we describe each of the four steps taken to prepare the data for training purposes. These four steps are applied concurrently to both the training and testing dataset which contain 80% and 20% of the original data. State variables with a significant number of missing values negatively impact the ML algorithm's training since there consequently many uninformative predictors to learn from. Thus, missing points are replaced in step 1 with the median value for each predictor since all predictors contain numeric values. Next, the binary response variable is encoded in step 2 such that the "Yes" class corresponding to the presence of a fault is labeled "1" and the "No" class corresponding to the absence of a fault is labeled "0". Subsequently, we check the class distribution of the response variable in step 3 since a severe skew in the class distribution can influence many ML algorithms to ignore the minority class entirely. Hence, ML algorithms are trained using a comparable amount of each possible outcome (or balanced classes) since the minority class is often the most important for predictions as shown in Figure 3.

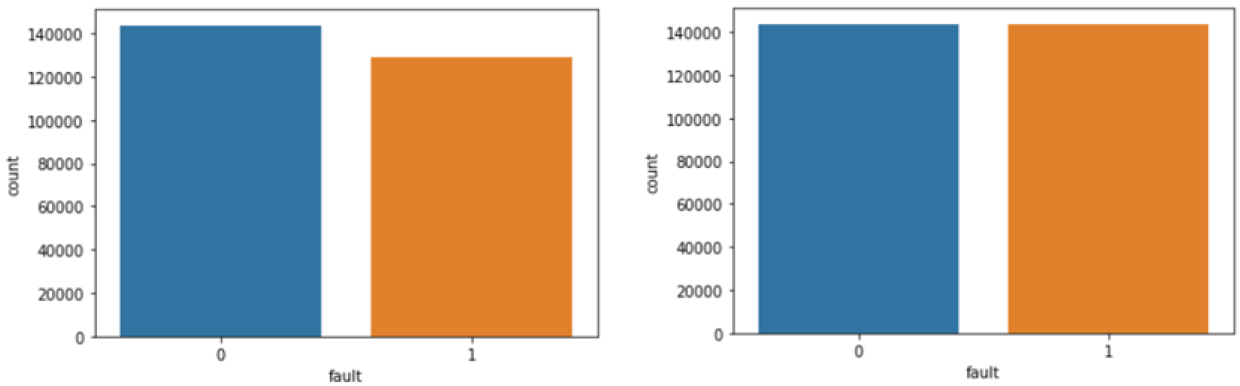

Figure 3: Imbalanced class distribution of response variable (left) and balanced class distribution of response variable after oversampling (right).

Figure 3 shows an imbalance class distribution for the response variable on the left, where class 0 is the majority class. The two most common approaches to rectifying an imbalance dataset are to i) delete data points from the majority class or under-sample the majority class and ii) to duplicate data points from the minority class or over-sample the minority class. Thus, class "1" is oversampled using synthetic minority oversampling technique (SMOTE) to match the number of observations in class "0" and balance the class distribution in Figure 3 on the right. Notably, oversampling was chosen due to the risk of losing valuable information by under-sampling the majority class. Finally, to mitigate bias toward predictors with a larger range of values, all predictors were scaled to a similar range in step 4 before training ML algorithms to develop FD models.

$$\frac{x_i - x_{min}}{x_{max} - x_{min}} \tag{1}$$

Therefore, the GPVS-Faults dataset is normalized using the min-max scaler function defined in (1). The level of impact scaling has on an FD model's predictive performance depends on the ML algorithm, where ML algorithms can be categorized based on their learning behavior (e.g., distance learners, gradient descent, trees) and decision boundary (e.g., linear and non-linear). Thus, the min-max scaler function was selected since it does not reduce the effect of potential outliers which may indicate a severe fault in an MG system.

2.3 Machine Learning Models

After preprocessing, ML algorithms suited for classification including Logistic Regression (LR), Random Forest (RF) and Naïve Bayes (NB) are trained to develop a robust FD model for an MG system. LR is a distance-based classification model that generates a linear decision boundary between classes within the response variable. LR assumes there is a linear relationship between each predictor and the log of odds of the response variable.

$$\hat{y} = \text{sigmoid}\left(\ln\left(\frac{P(k|x)}{1 - P(k|x)}\right) = b_0 + b_1x_1 + b_2x_2 + \cdots + b_nx_n\right) \tag{2}$$

LR is defined in (2). Hypothesis testing is then used to determine the significance of each predictor with respect to the response where the null hypothesis assumes the predictors have no effect on the response. Furthermore, LR assumes there are no leverage points, duplicate points, or multicollinearity (collinear predictors) as these factors tend to negatively impact the reliability of regression coefficients. Distance based models tend to scale poorly with larger datasets as the number of distance computations increase exponentially with each additional point. Therefore, statistical inferences drawn from the distributions of input variables can inform a current belief for faster predictions.

$$\mathrm{P(H|E)} = \frac{P(H) \times P(E|H)}{P(E)} \tag{3}$$

Bayes Theorem is the learning mechanism underlying the NB classifier and is defined in (3). Like LR, NB generates a linear decision boundary between classes within the response variable. Although the results from linear models are interpretable, a line or hyperplane that separates classes within the response variable may not exist. Thus, non-linear models which generate any decision boundary excluding a hyperplane are used to provide a more powerful separation tool. The tradeoff for this improved predictive performance is often a high model complexity which tends to make the resulting prediction intractable.

Non-linear models can generate a decision boundary that has a stair-case shape as with trees or a curvy shape as with neural networks (NNs). RF is a tree-based classification model that leverages the power of multiple decision trees called bagging (Altman and Krzywinski 2017) to generate a non-linear decision boundary between classes within the response variable. Since a random subset of available predictors are used to iteratively build single decision trees, RF combines the output of multiple decision trees or weak learners to generate the final output known as majority voting (Breiman 2001).

$$1 - \sum_{k=1}^{K} P_k^2 \tag{4}$$

$$\sum_{n=0}^{N} 2^n \tag{5}$$

Decision trees learn by dividing nodes into sub-nodes using a splitting criterion such as the Gini Impurity defined in (4). Furthermore, the total number of nodes within a decision tree or its size is defined in (5), where (n) is the depth of the tree. Parent nodes are continuously branched using the splitting criterion in (4) throughout the training process until only pure (or homogenous) terminal nodes remain. Thus, significant predictors can be identified based on how well they separate the GPVS-Faults dataset or reduce impurities during training. Moreover, tree-based regularization techniques can be employed by constraining (n) to mitigate any underfitting or overfitting observed during testing. Due to a computationally efficient decision rule, tree base ML models tend to scale well with larger datasets.

3 RESULTS AND DISCUSSION

Within the proposed FD framework, each ML model was tested for its ability to detect a fault during MG operations. The proposed ML models were coded in Python 3.6 and ran on a computer with an Intel i7 2.67 GHz processor and 16 GB RAM for replicability purposes. Training and validation data was collected from lab experiments conducted by (Bakdi et al. 2021) and is outlined in Table 3. The assumptions required to implement each ML model in (2) to (4) were both tested and addressed accordingly.

The assumption investigated in LR was the presence of multicollinearity. To check for collinear variables, a correlation matrix was used where coefficients closer to 1 reflect co-dependence between variables as shown in Figure 4.

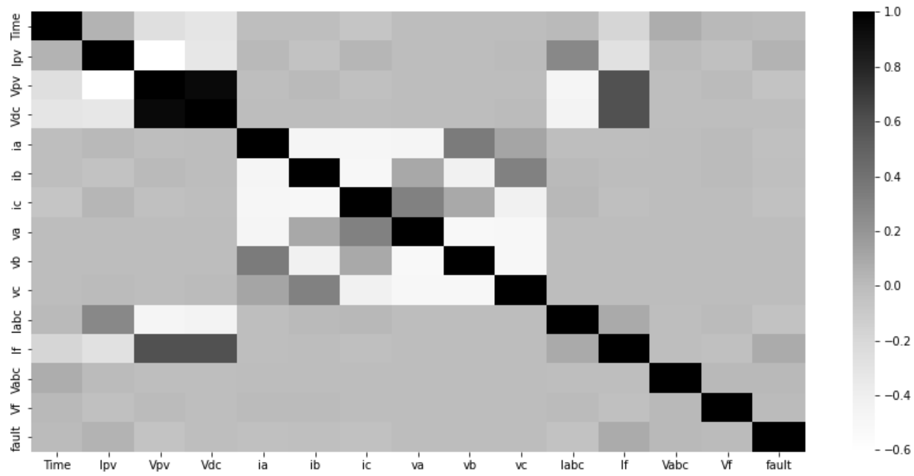

Figure 4: Correlation matrix to check for multicollinearity among variables.

Although Figure 4 shows the presence of multicollinearity among variables in the GPVS-Faults dataset, these correlations are observed and not introduced by creating new variables from existing ones. Hence, domain knowledge about the system and collection method must inform how multicollinearity should be addressed. In this study, data was collected from an MG system that produce three phase power where power is a function of voltage and current at each phase. Notably, the voltage and current at one phase (e.g., va-ia, vb-ib, vc-ic) are independent whereas there is a dependence between the voltage at one phase and the current at the other two phases. This observation is presumed to be the result of three symmetrical single-phase components operating simultaneously to satisfy a balanced three-phase loading within an MG (Weedy et al. 2012). Another correlation is observed between 'Vpv' and 'Fault' where the PV cells are manually exposed voltage imbalances generating faults such as F4 and F5 in Table 2. Therefore, correlations such as these are to be expected within an MG system and should not be manipulated or removed to improve the reliability of regression coefficients in (2).

After investigating assumptions associated with LR, assumptions associated with NB were investigated since RF does not require any preconditions to be implemented. NB assumes all the predictors are dependent, but this assumption is relaxed in NB classifier since each predictor requires it unique PDF for tractability (Taheri and Mammadov 2013). Since assumptions for LR, RF, and NB are satisfied, RF is applied to the GPVS-Faults dataset. The learning mechanism in (4) behind RF is used extract variables that influence the training process.

In Figure 5, the significance scores for RF are based on the decrease of Gini impurity when a variable is chosen to split a node. Since this importance is derived from the learning mechanism in (4), it does not directly reflect the relationship between variables and the response. Hence domain knowledge is required to better interpret the result for the MG system. Here, Figure 5 suggests PV cells voltage "Vpv" is the most significant variable for learning to classify "Faults" which was previously corroborated by Figure 4.

Moreover, applying the NB classifier results in the same significance scores across variables in the GPVS-Faults dataset. After addressing the assumptions required to implement each ML model and identifying variables that may influence the likelihood of a fault, the predictive performance of each ML model is reported in Table 4.

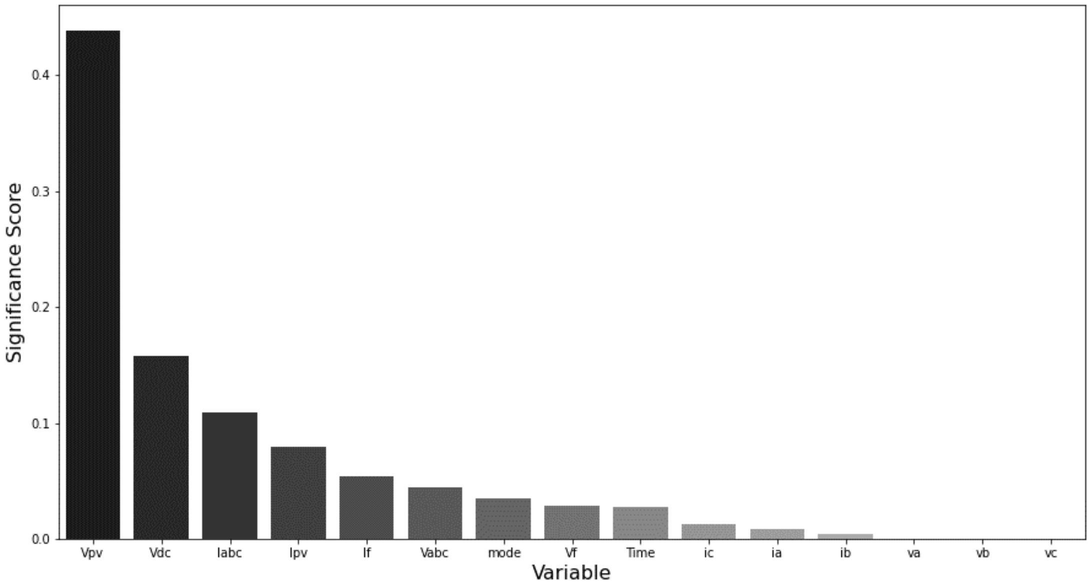

Figure 5: Variable significant for training RF.

In Table 4, the false positive (FP/n) and false negative (FN/n) ratios reflect type 1 and type 2 error respectively. With respect to MG operational planning, type 1 error measures the amount of misclassified negative values or rate at which the ML model is sending a false alarm to the TSO. Although these false alarms can incur an additional cost as the TSO time is consumed investigating them, it is the desired alternative compared to type 2 error. Type 2 error is of greater concern because it measures the amount of misclassified positive values or rate at which faults are being overlooked within an MG system by the ML model. Ideally both types of error should be minimized across each ML model; however, this is rarely observed in data science and domain knowledge should be used to define an operating region or an acceptable margin of error for the TSO. Furthermore, precision and recall describe the quality and quantity of data for each class of the binary response variable. However, neither precision nor recall can be used independently to benchmark the overall performance of an ML model. Therefore, an F-score combines both precision and recall into a metric to benchmark an ML model's overall performance.

Table 4: Results using ML models for FD.

Classifier	Classification Metrics							Training Time (sec)
	FP/n	FN/n	Class	Precision	Recall	F1 Score	Accuracy	
Logistic Regression	0.011	0.016	0 1	0.97 0.98	0.98 0.97	0.97 0.97	0.97 (0.05)	15.3 (+/- 0.35)
Random Forrest	0.046	0.029	0 1	0.97 0.96	0.96 0.97	0.96 0.96	0.96 (0.01)	1.67 (+/- 0.7)
Naive Bayes	0.181	0.187	0 1	0.63 0.63	0.64 0.63	0.63 0.63	0.63 (0.12)	0.07 (+/- 0.0)

Notably, the F-score and accuracy are equivalent in each class for each ML model since the class distributions of the response variable are balanced (see Section 2.2). Finally, K-fold cross validation is used to ensure the average accuracy and standard deviation that describe the predictive performance remain consistent across each ML model.

According to Table 4, RF shows a comparable training time of 1.67 seconds to learn compared to NB with the shortest training time at 0.07 seconds. However, NB shows poor predictive performance at 0.63 compared to LR at 0.97 and RF at 0.96. Although LR and RF have comparable predictive performance, LR has the longest training time at 15.3 seconds and will scale poorly data accrued during real-time MG operations compered to NB and RF. Therefore, amongst the ML models applied to the GPVS-Faults dataset in this study, RF should be chosen based on the F-score and time taken to train each model.

4 CONCLUSION

In this study we propose an FD framework for MG operational planning. Initially, the TSO uses domain knowledge to determine which state variables should be considered in FD analysis for MG systems. Preprocessing real-time data to train ML models for FD involves checking the assumption associated with a particular ML model using domain knowledge about MG systems before its implementation. Although a few correlations were observed between variables, they were not introduced by combining existing variables to synthesize new ones. Therefore, the variables producing these correlations should not be manipulated to improve the interpretability of a LR based FD model. Based on the predictor's significance during training across each ML model, insight on which state variables best describe an MG system's behavior were obtained. The ML algorithms were applied to the GPVS-Faults dataset including LR, RF, and NB. These ML models were chosen since that have unique learning mechanisms (e.g., distance, trees, probability) to develop a variety of FD models with either linear or non-linear decision boundaries. The classification performance metrics amongst the ML algorithms chosen showed LR can identify faults in MG systems best but takes the longest time to train and subsequently scale. Conversely, NB takes the least time to train but is also the least accurate when making predictions. Therefore, RF was shown to be the most appropriate ML model for FD in MG systems with PV generation with comparable accuracy to LR and comparable training time to NB.

REFERENCES

Altman, N., and M. Krzywinski, 2017. "Ensemble Methods: Bagging and Random Forests". *Nature Methods* 14(10):933-935.

Bakdi, A., W. Bounoua, A. Guichi, and S. Mekhilef. 2021. "Real-Time Fault Detection in PV Systems under MPPT using PMU and High-frequency Multi-sensor Data Through Online PCA-KDE-based Multivariate KL Divergence". *International Journal of Electrical Power & Energy Systems* 125:106457.

Bansal, Y., and R. Sodhi, 2018. "Microgrid Fault Detection Methods: Reviews, Issues and Future Trends". In *2018 IEEE Innovative Smart Grid Technologies-Asia (ISGT Asia)*. May 22nd-25th, Singapore, 401-406.

Benkercha, R., and S. Moulahoum. 2018. "Fault Detection and Diagnosis based on C4. 5 Decision Tree Algorithm for Grid Connected PV System". *Solar Energy* 173:610-634.

Breiman, L. 2001. "Random forests". *Machine learning* 45(1):5-32.

Chen, Z., F. Han, L. Wu, J. Yu, S. Cheng, P. Lin, and H. Chen. 2018. "Random Forest based Intelligent Fault Diagnosis for PV Arrays using Array Voltage and String Currents". *Energy Conversion and Management* 178:250-264.

Damgacioglu, H., and N. Celik. 2022. "A Two-stage Decomposition Method for Integrated Optimization of Islanded AC grid Operation Scheduling and Network Reconfiguration". *International Journal of Electrical Power & Energy Systems* 136:107647.

Darville, J., and N. Celik. 2020. "Microgrid Operational Planning using Deviation Clustering within a DDDAS Framework". In *Proceedings of the 2020 International Conference on Dynamic Data Driven Application Systems*. October 2nd-4th , Boston, MA, 77–84.

Darville, J., and N. Celik. 2020. "Simulation Optimization for Unit Commitment using a Region-based Sampling (RBS) Algorithm". In *IISE Annual Conference and Expo 2020*. November 1st-3rd, New Orleans, LA, 1424-1430.

Darville, J., J. Curia, and N. Celik. 2022. "Microgrid Operational Planning using a Hybrid Neural Network with Resource-aware Scenario Selection". *Simulation Modelling Practice and Theory* 119:102583.

U.S. Energy Information Administration (EIA). 2022. Annual Energy Outlook 2022. https://www.eia.gov/outlooks/aeo/pdf/AEO2022_ReleasePresentation.pdf, accessed 3rd March 2022.

U.S. Energy Information Administration (EIA). 2021. Solar Generation was 3% of U.S. Electricity in 2020, but we Project it will be 20% by 2050. https://www.eia.gov/todayinenergy/detail.php?id=50357#:~:text=Solar%20generation%20was%203%25%20of,will%20be%2020%25%20by%202050&text=According%20to%20our%20Electric%20Power,from%20all%20sources%20in%202020, accessed 3rd March 2022.

Francis, D., and T. Thomas. 2014. "Mitigation of voltage sag and swell using dynamic voltage restorer". In *2014 Annual International Conference on Emerging Research Areas: Magnetics, Machines and Drives (AICERA/iCMMD)*. July 1st-3rd Kottayam, Kerala, India, 1-6.

Goodwin, T., J. Xu, N. Celik, and C. H. Chen, 2022. "Real-time Digital Twin-based Optimization with Predictive Simulation Learning". *Journal of Simulation* 16(6):1-18.

Hong, Y., and R. A. Pula. 2022. "Detection and Classification of Faults in Photovoltaic Arrays using a 3D Convolutional Neural Network". *Energy* 246:123391.

Madeti, S. R., and S.N. Singh. 2018. "Modeling of PV System based on Experimental Data for Fault Detection using kNN Method". *Solar Energy* 173:139-151.

Madeti, S. R., and S. N. Singh. 2017. "Online Fault Detection and The Economic Analysis of Grid-connected Photovoltaic Systems". *Energy* 134:121-135.

Mesham, M., M. Fahmy, and N. Celik. 2020. "Simulation based Modeling for a Cybersecure Power Grid". In *2020 Spring Simulation Conference (SpringSim)*. May 19th -21st, Fairfax, VA, 1-12.

Pena, R. S., R. J. Cardenas, J. C. Clare, and G. M. Asher. 2001. "Control Strategies for Voltage Control of a Boost Type PWM Converter". In *2001 IEEE 32nd Annual Power Electronics Specialists Conference (IEEE Cat. No. 01CH37230)*. June 17th -21st, Vancouver, BC, Canada, 2:730-735.

Philippe, C. D., W. M. Hamilton, A. C. Penn, and J. M. Shultz. 2021. "Innovative Health Professional Leadership for a Climate-resilient Bahamas". *The Journal of Climate Change and Health* 4:100055.

Runsewe, T., O. Bafail, and N. Celik. 2020. "Performance Analysis of Waste Collection Programs in Material Recovery Facilities". In *IISE Annual Conference and Expo 2020*. November 1st-3rd, New Orleans, LA, 1401–1406.

Saha, S., M. E. Haque, C. P. Tan, M. A. Mahmud, M. T. Arif, S. Lyden, and N. Mendis. 2020. "Diagnosis and Mitigation of Voltage and Current Sensors Malfunctioning in a Grid connected PV system". *International Journal of Electrical Power & Energy Systems* 115:105381.

Sáenz, J. B., N. Celik, S. Asfour, and Y. J. Son. 2012. "Electric utility resource planning using Continuous-Discrete Modular Simulation and Optimization (CoDiMoSO)". *Computers & Industrial Engineering* 63(3):671-694.

Sakulchotruangdet, S., and S. Khwan-on. 2016. "Three-phase Interleaved Boost Converter with Fault Tolerant Control Strategy for Renewable Energy System Applications". *Procedia Computer Science* 86:353-356.

Sinapis, K., T. Rooijakkers, and C. Tzikas. 2016. Partner in Solar Energy Solutions: Shading Mitigation Strategies for PV Systems. https://sundaynl.nl/u/images/b2%20-%20kostas%20sinapis.pdf, accessed 3rd March 2022.

Tabbache, B., M. Benbouzid, A. Kheloui, J. M. Bourgeot, and A. Mamoune. 2013. "An Improved Fault-tolerant Control Scheme for PWM Inverter-fed Induction Motor-based EVs". *ISA transactions* 52(6):862-869.

Taheri, S., and M. Mammadov. 2013. "Learning the Naive Bayes Classifier with Optimization Models". *International Journal of Applied Mathematics and Computer Science* 23(4):787-795.

Thanos, A. E., M. Bastani, N. Celik, and C.H. Chen. 2017. "Dynamic Data Driven Adaptive Simulation Framework for Automated Control in Microgrids". *IEEE Transactions on Smart Grid* 8(1):209–218.

Wang, Z., D. M. Anand, J. Moyne, and D. M. Tilbury. 2017. "Improved Sensor Fault Detection, Isolation, and Mitigation using Multiple Observers Approach". *Systems Science & Control Engineering* 5(1):70-96.

Weedy, B. M., B. J. Cory, N. Jenkins, J. B. Ekanayake, and G. Strbac. 2012. *Electric Power Systems*. 5th ed. West Sussex, United Kingdom: John Wiley & Sons, Ltd.

Wright, W. J., J. Darville, N. Celik, H. Koerner, and E. Celik. 2022. "In-situ Optimization of Thermoset Composite Additive Manufacturing via Deep Learning and Computer Vision". *Additive Manufacturing* 58:102985.

Yang, B., H. Ye, J. Wang, J. Li, S. Wu, Y. Li, H. Shu, Y. Ren, and H. Ye. 2021. "PV Arrays Reconfiguration for Partial Shading Mitigation: Recent Advances, Challenges and Perspectives". *Energy Conversion and Management* 247:114738.

Yavuz, A., J. Darville, N. Celik, J. Xu, C.H. Chen, B. Langhals, and R. Engle. "Advancing Self-Healing Capabilities in Interconnected Microgrids via DDDAS with Relational Database Management". In *Proceedings of the 2020 Winter Simulation Conference*, edited by K.-H. Bae, B. Feng, S. Kim, S. Lazarova-Molnar, Z. Zheng, T. Roeder, and R. Thiesing, 2030 – 2041. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Yi, Z., and A. H. Etemadi. 2017. "Line-to-Line Fault Detection for Photovoltaic Arrays based on Multiresolution Signal Decomposition and Two-stage Support Vector Machine". *IEEE Transactions on Industrial Electronics* 64(11):8546-8556.

AUTHOR BIOGRAPHIES

JOSHUA DARVILLE is a Ph.D. candidate in industrial engineering at the University of Miami (UM) orginally from the Islands of The Bahamas. He acquired a dual Bachelor's degree in physics and mechanical engineering from Fisk University in 2017 and

Vanderbilt University in 2019. He was captain of the Fisk University rocket team which was the only HBCU to participate in the NASA student launch competition in 2017 to 2018. He then graduated from Vanderbilt to pursue his passion for research at SimLab under the mentorship of Dr. Nurcin Celik at UM. His current research interests are machine learning, optimization, and autonomous microgrids small island developing states (SIDS). He is a strong advocate for STEM education in underrepresented groups and has received the 2021 Topple Graduate Student of the Year Award and 2022 ISE Outstanding Graduate Student Award for Leadership. His email address is jmd437@miami.edu. His website is https://simlab.coe.miami.edu/team/.

TEMITOPE RUNSEWE is a Ph.D. candidate in the department of industrial and systems engineering at the University of Miami (UM). She obtained her bachelor's degree in chemical engineering from covenant university, Nigeria. She then received her master's degree in industrial engineering from Texas State University, San Marcos. She was also awarded the Texas State University graduate college scholarship and student government scholarship. Her research interest includes application of optimization of complex systems, modelling and simulation of various systems and Data analytics. She also received the 2020 REMADE Institute rising Star Spotlight. Her email address is tor9@miami.edu. Her website is https://simlab.coe.miami.edu/team/.

ABDURRAHMAN YAVUZ is a Ph.D. candidate in industrial engineering at the University of Miami (UM). He received the B.Sc. and M.Sc. degrees in industrial engineering from the TOBB University of Economics and Technology, Ankara, Turkey, in 2017 and 2019, respectively. He is currently pursuing a Ph.D. degree in industrial engineering at the University of Miami. He also has carried out the teaching assistantship of the industrial engineering courses during his M.Sc. degree. His current research effort is the development of a computational method to make optimal decisions under uncertain conditions. He also received the ISE Outstanding Graduate Student Award for Research in 2021. His email address is axy286@miami.edu. His website is https://simlab.coe.miami.edu/team/.

NURCIN CELIK, Ph.D., is a Jones Career Development Associate Professor in the Department of Industrial and Systems Engineering at the University of Miami (UM). She received her M.S. and Ph.D. degrees in Systems and Industrial Engineering from the University of Arizona with magna cum laude. Sponsored by AFOSR, DOE, and the City of Coral Gables she has worked on the development of integrated modeling and decision-making methodologies for large-scale, complex, and dynamic systems with a focus on smart grids. Dr. Celik received the Presidential Early Career Award for Scientists and Engineers from the White House in 2017. The PECASE Award is the highest honor bestowed by the U.S. government on outstanding scientists and engineers beginning their independent careers. She has served in the editorial board of academic journals and conference organizations, and has been an ad-hoc reviewer for more than (15) journals. Her email address is celik@miami.edu. Her website is https://simlab.coe.miami.edu/team/.