21, rue d'Artois, F-75008 PARIS
http : //www.cigre.org

CIGRE US National Committee
2022 Grid of the Future Symposium

Performance Analysis of Machine Learning Algorithms for Conservation Voltage Reduction (CVR) Status Detection

A. HAQUE, W. FAN, P. PABST
Commonwealth Edison Company
USA

SUMMARY

Conservation Voltage Reduction (CVR) reduces electrical energy consumption by reducing supply voltage within regulatory limits through Voltage Optimization (VO). To accurately quantify the energy savings through CVR application, a comparative analysis between CVR "On" and "Off" periods is required. Therefore, CVR operational status detection is an important step during the CVR Measurement and Verification (M&V) process. An inaccurate CVR status detection will introduce computational error during energy savings M&V calculation. This paper presents a data-based approach to CVR status detection. Three relevant Machine Learning (ML) algorithms are used as classification algorithms. The selected algorithms are: Artificial Neural Network (ANN), Support Vector Machine (SVM) and K-Nearest Neighbors (KNN). Actual system-level data from a utility in the U.S is utilized for the case study. A detailed performance analysis is presented. Analysis show, the algorithms can detect CVR status with more than 95% accuracy.

KEYWORDS

Conservation Voltage Reduction, Machine Learning, Classification Algorithms

ashraful.haque@comed.com

1. INTRODUCTION

Conservation Voltage Reduction (CVR) is an energy efficiency measure that reduces electrical energy consumption through feeder level voltage reduction. In the grid modernization multi-year program plan outlined by the United States Department of Energy (DoE), efficient energy delivery is highlighted as one of the major attributes of future smart electric grids [1]. CVR is implemented in an electrical grid through voltage control equipment i.e., capacitor banks, voltage regulators. The application reduces grid voltage automatically without any negative impact to customers' regular operation. This is a new application in the power system operation and utilities across the U.S have started introducing CVR in their grid operation because of the net positive impact CVR has in terms of energy efficiency and environmental benefits. Authors in [2] provide an overview of CVR deployments by several electric utilities within the U.S.

A detailed Measurement and Verification (M&V) analysis is required to quantify the energy savings through CVR [3], [4]. Regression, comparison, and simulation-based techniques are some of the popular CVR M&V methodologies [5]. For energy-savings analysis using these methodologies, it is necessary to measure and compare the energy consumed while CVR operation is switched on and off during an interval i.e., 4 days on and 4 days off CVR operation for one year. Network signals and sensors e.g., CVR status flag, capacitor bank operation flag can be used to detect CVR status [6]. However, in the absence of the mentioned data points the practice is to manually detect CVR on/off status. This requires significant effort to complete and is cost-prohibitive to scale up during bulk system-wide CVR operation. Inaccurate status detection negatively impacts the final energy-saving calculation as well [7]. A machine learning (ML)-based CVR status detection technique that can detect CVR on/off status accurately, therefore, will improve the performance and increase efficiency of M&V analysis to quantify energy savings through CVR.

Modern machine learning algorithms have gained much popularity during the recent years for various power system analysis applications e.g., electrical load forecasting [8], energy savings quantification [9], fault detection [10]. These applications mainly focus on regression-based analysis. However, CVR operational status detection is a binary classification problem. The objective is to find whether CVR is "On" or "Off" during a period. Artificial Neural Network (ANN), Support Vector Machine (SVM), Logistic Regression, Decision Tree are some of the popular classification algorithms [11], [12]. The performance of these machine learning algorithms depends on various factors such as: data volume, data quality, relevant input features, optimal network configuration etc [13], [14]. Therefore, a detailed analysis focusing on CVR status detection is required.

An argument can be made that simply by selecting a threshold for the on/off status, CVR status can be detected. However, the voltage profile and the percentage of voltage reduction during CVR is not unique across the feeders and depends on multiple external factors i.e., weather, load mix, customer count etc. Threshold for each of the feeders will need to be manually selected. Hence, the threshold-based detection approach will not be scalable and will be cost prohibitive. If the threshold is not properly defined, it will negatively impact the accuracy to detect CVR status as well. Therefore, for a cost-effective and scalable grid-level CVR implementation, a data-based approach that can compute the threshold automatically and accurately detect CVR status is required.

The major contributions of the paper are:

1. A machine-learning based method to detect CVR on/off status is presented. This method can be utilized to detect CVR status with better efficiency and scalability compared to a manual status detection. The classification algorithms selected for the analysis are:
    a. Artificial Neural Network (ANN)
    b. Support Vector Machine (SVM)
    c. K-Nearest Neighbors (KNN)

2. A detailed performance analysis using actual system level data from a utility in the U.S is used for case study. The validation criteria used for the analysis are: confusion matrix analysis, accuracy, precision, recall and F1 score.

3. The methodology presented is implemented through TensorFlow for the case study. Therefore, this approach can be integrated to detect real-time CVR status for Advanced Distribution Management System (ADMS) applications.

The rest of the paper is organized as follows: methodology and the mathematical background of the algorithms, dataset analysis, case study and result analysis, conclusion.

2. METHODOLOGY FOR CVR STATUS DETECTION

The objective of classification machine learning algorithms is to classify a given set of data into binary or multiple class (output) based on available features (input). CVR status detection is a binary classification that classifies whether CVR status is on or off based on the input features. The algorithms used for this paper are: ANN, SVM and KNN.

<u>Artificial Neural Network (ANN):</u> ANN mimics a human brain to find relationship between the inputs and the outputs. The algorithm can be used for both classification and regression problems. ANN consists of multiple hidden layers between the inputs and the outputs. These hidden layers are consisted of multiple nodes known as the neurons. Fig 1 shows the typical network architecture of an ANN.

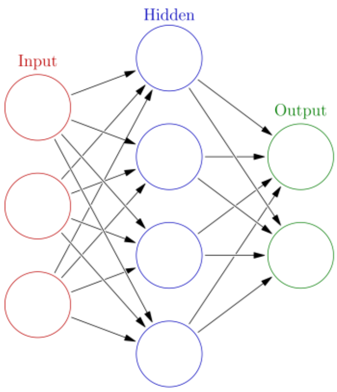


Fig 1: Network architecture of Artificial Neural Network (ANN) [15]

Support Vector Machine (SVM): SVM is a supervised machine learning algorithm that draws a line known as hyperplane separating the data into different classes [16]. The hyperplane can be defined as in Eq.1:

$$w.x + b = e \tag{1}$$

Where, w is the weight vector b is the bias term, and x is the input training data. The objective function of SVM is:

$$\begin{aligned} \min \quad & \frac{1}{2}||\boldsymbol{w}^2|| + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \\ \text{s.t.} \quad & y_i - f(x_i, \boldsymbol{w}) \leq \epsilon + \xi_i^* \\ & f(x_i, \boldsymbol{w}) - y_i \leq \epsilon + \xi_i \\ & \xi_i, \xi_i^* \geq 0, i = 0,1,2,3 \ldots n \end{aligned} \tag{2}$$

Where, C is the cost function and ϵ is the tolerance boundary. Fig 2 shows a representative hyperplane generated through SVM to classify data.

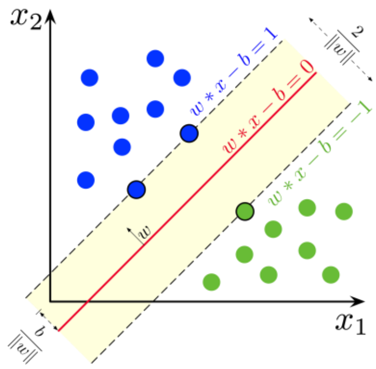


Fig 2: Support Vector Machine (SVM) [17]

K-Nearest Neighbors (KNN): KNN is a machine learning algorithm that classifies data based on proximity with the input features that the model is initially trained with. KNN determines the distance of input test data with the training data set and based on that the algorithm output a class.

The paper follows the following steps to prepare and train these ML models for CVR status detection:

1. Data collection: The first step is to collect data. For the CVR status detection, system level data can be collected from Supervisory Control and Data Acquisition (SCADA) system for real time status detection or from a stored database for offline status detection for M&V.

2. Data preparation: The second step is to prepare the raw data. This step includes processing bad or missing data and splitting data into training, validation and testing dataset.

3. Training ML models: The next step is to train the ML models for pre performance analysis. Initial architecture of the algorithms should be selected based on literature review.

4. Performance analysis: Performance analysis based on selected validation criteria is carried out in this step. Based on the performance analysis, optimal algorithm is selected for further hyperparameter tuning.

5. Hyperparameter tuning: To maximize the performance of ML models hyperparameters are tuned in this step. There is no unique method to optimize the performance and heuristic analysis is needed to find the optimal network architecture.

6. CVR status detection: Once models are optimally trained, they can be used for CVR status detection. Performance of CVR detection should be monitored over the period to figure whether the models need to be re-trained or not.

Fig 3 shows the methodology to detect CVR status detection and the output after each step is completed.

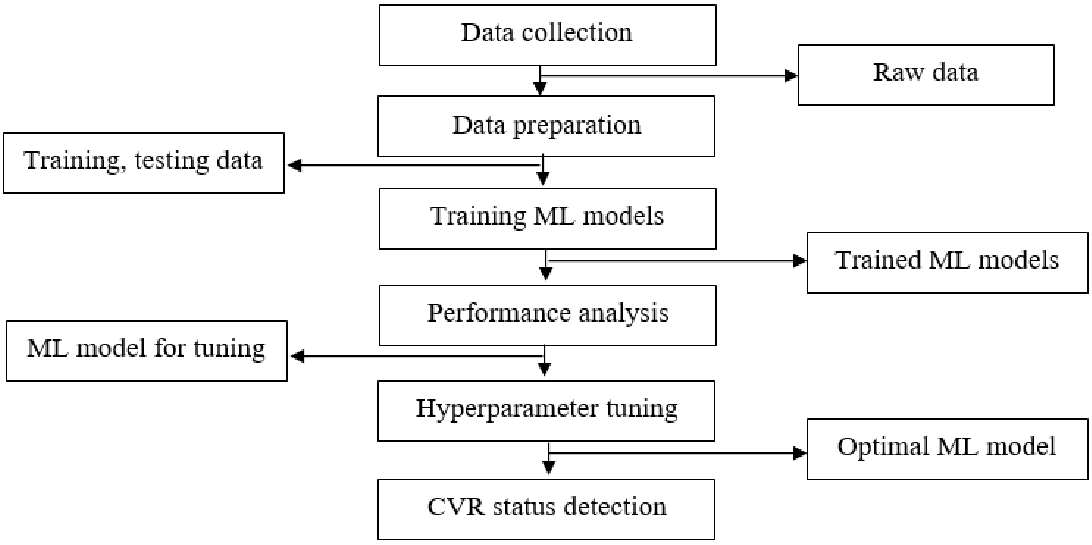


Fig 3: Methodology for CVR status detection

The following validation criteria are used for the performance analysis:

1. Confusion matrix: A confusion matrix shows the number of accurate and inaccurate classification. This matrix can be presented in a tabular format like below and is useful to figure out the type of inaccurate classification.

	Actual 1	**Actual 0**
Predicted 1	True Positives (TP)	False Positives (FP)
Predicted 0	False Negative (FN)	True Negative (TN)

Where, TN is the number of negative cases correctly classified, TP is the number of positive cases correctly classified, FN is the number of positive cases incorrectly classified as negative, and FP is the number of negative cases incorrectly classified as positive.

2. Accuracy: Accuracy is the percentage of cases correctly classified and can be presented using Eq. 3:

$$\text{Accuracy} = (TP+TN) / (TP+TN+FN+FP) \qquad (3)$$

3. Precision: Precision presents how accurately the correct status is detected and can be presented using Eq. 4:

$$\text{Precision} = TP / (TP+FP) \qquad (4)$$

4. Recall: Recall presents the percentage of positive status correctly identified and can be presented using Eq. 5:

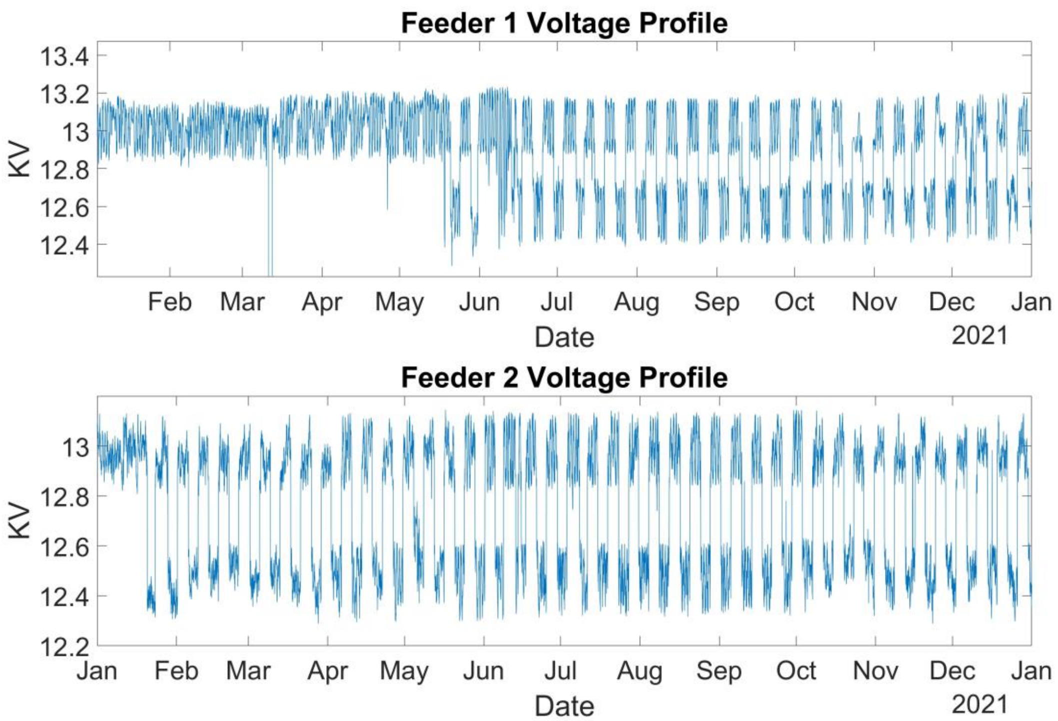
$$\text{Recall} = TP / (TP+FN) \qquad (5)$$

5. F1 score: F1 score is the harmonic mean of precision and recall values and can be presented using Eq.6:

$$\text{F1 score} = 2 * ((\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})) \qquad (6)$$

3. DATASET

The case study uses actual system level voltage data from 5 different feeders from a utility in the U.S. The main objective of CVR is to reduce grid level supply voltage. Therefore, for the classification, only voltage data is utilized, and performance is compared for the selected algorithms. Introducing unrelated variables decrease model performance and is, therefore, not considered. Voltage profile of one year with a 30-min resolution is used for the case study. Fig 4 shows voltage profiles of the 5 different feeders selected for the case study.



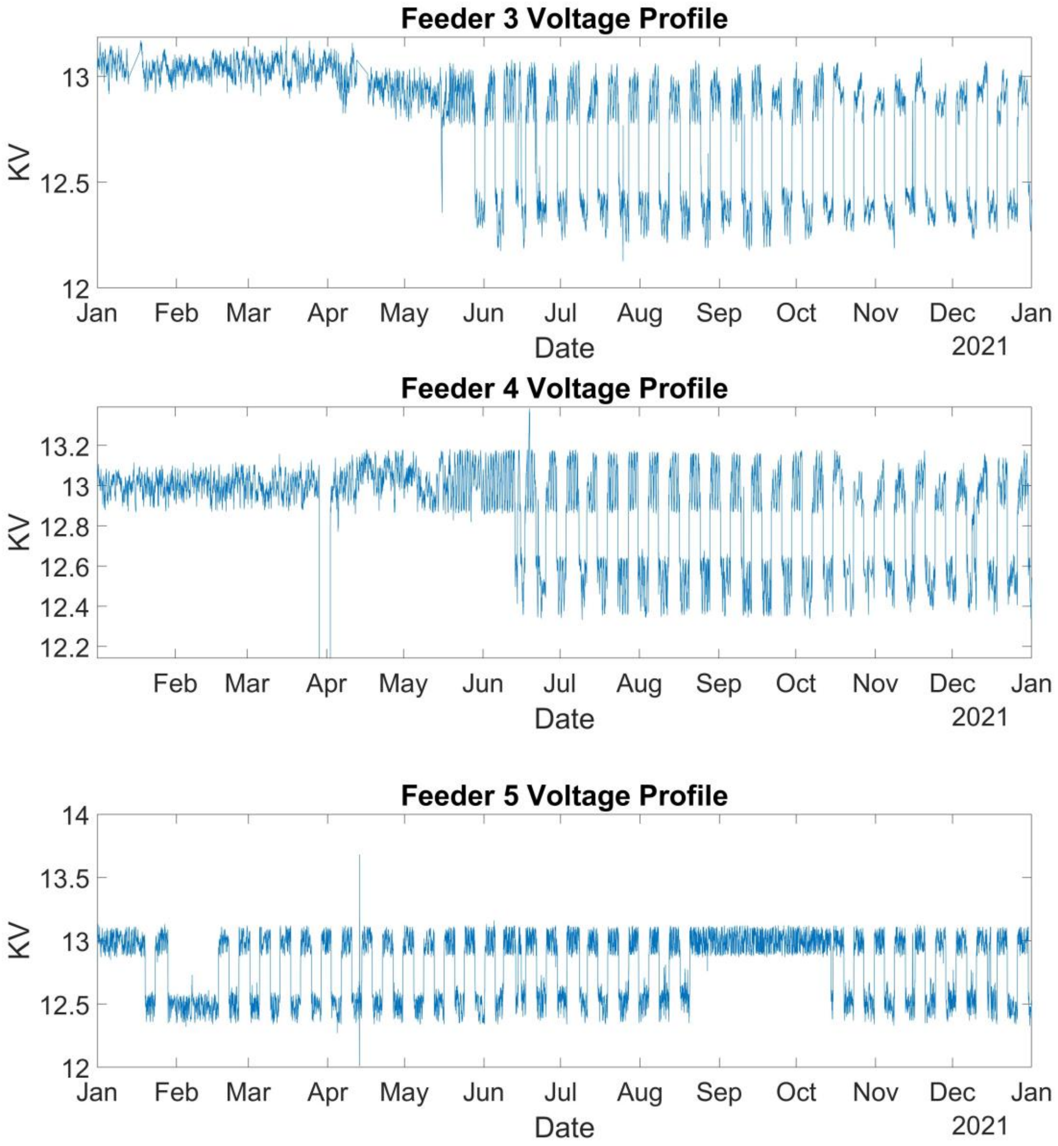


Fig 4: Voltage profile of Feeder 1, 2, 3, 4 and 5

The higher magnitude of voltage is when CVR is "Off", and the lower magnitude of voltage is when CVR is "On". The figure shows, these magnitudes are different for different feeders. The models should be able to detect this difference based on the training data and accurately classify the testing inputs without any prior threshold information. The data is randomly split into a 2:1 ratio for training and testing purpose. The training data is then further randomly split into 4:1 ratio into training and validation dataset.

4. CASE STUDY ANALYSIS

TensorFlow framework is used for the case study. Initial architecture for the ML algorithms is selected based on literature review and then further optimized for better performance. In real life application, optimization should be done for only the algorithm that performs the best with the initial configuration. For ML algorithms, hyperparameter tuning is done heuristically and, therefore, optimizing the best performing algorithm only, ensures optimum use of computational resources.

Table 1 shows the major hyperparameters of the optimal network architecture of ANN, SVM and KNN models developed for this case study.

Table 1: Hyperparameter configuration of ANN, SVM, KNN

Algorithm	Hyperparameter	Value
ANN	Hidden layers	2
	No of neurons in each layer	10
	epochs	200
	Loss function	Binary cross entropy
	Batch size	10
SVM	Kernel	Linear
	Cost (C)	1
	Gamma	20
	Epsilon	0
KNN	No of Neighbors	5
	Metric	Minkowski

Fig 5(a), 5(b) and 5(c) shows the confusion matrix for CVR detection for the 5 selected figures using ANN, SVM and KNN respectively.

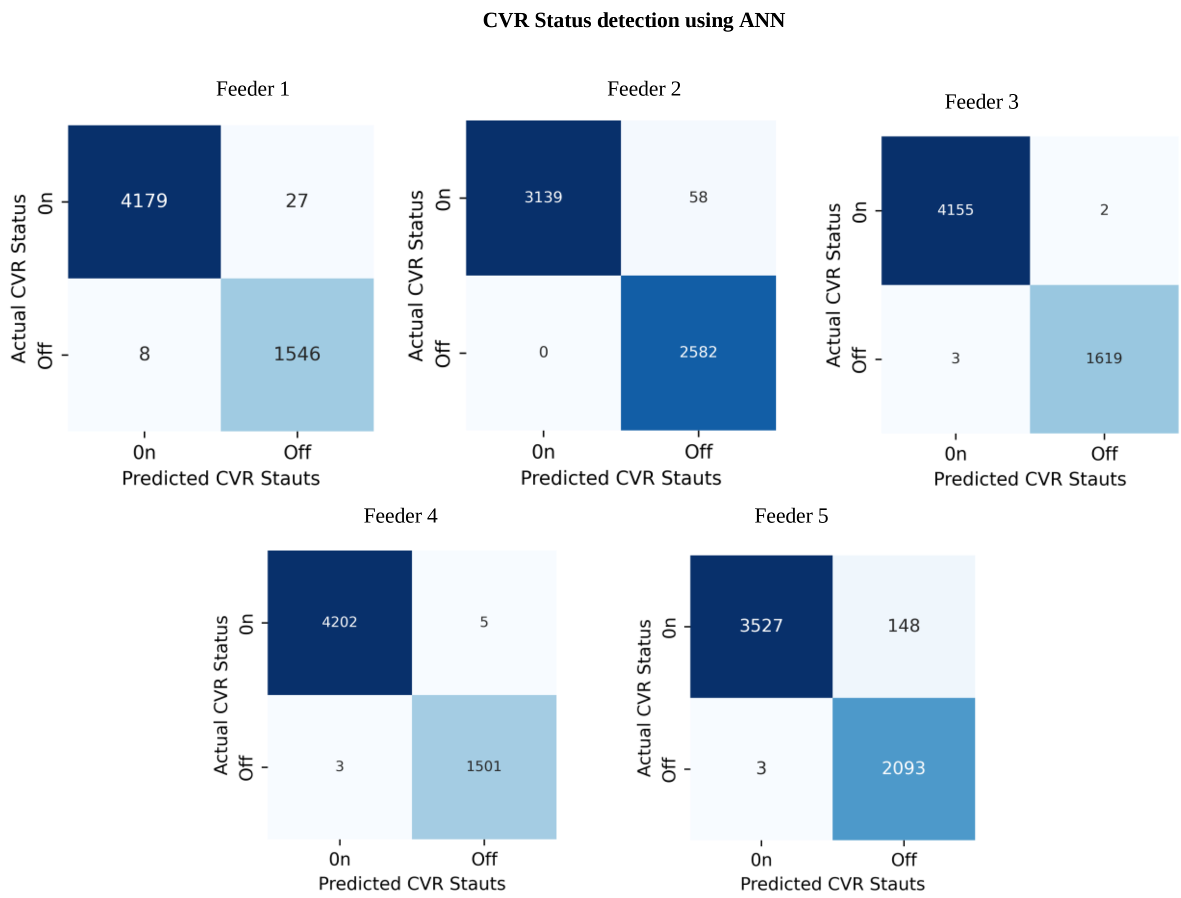


Fig 5(a): Confusion matrix for ANN

CVR Status detection using SVM

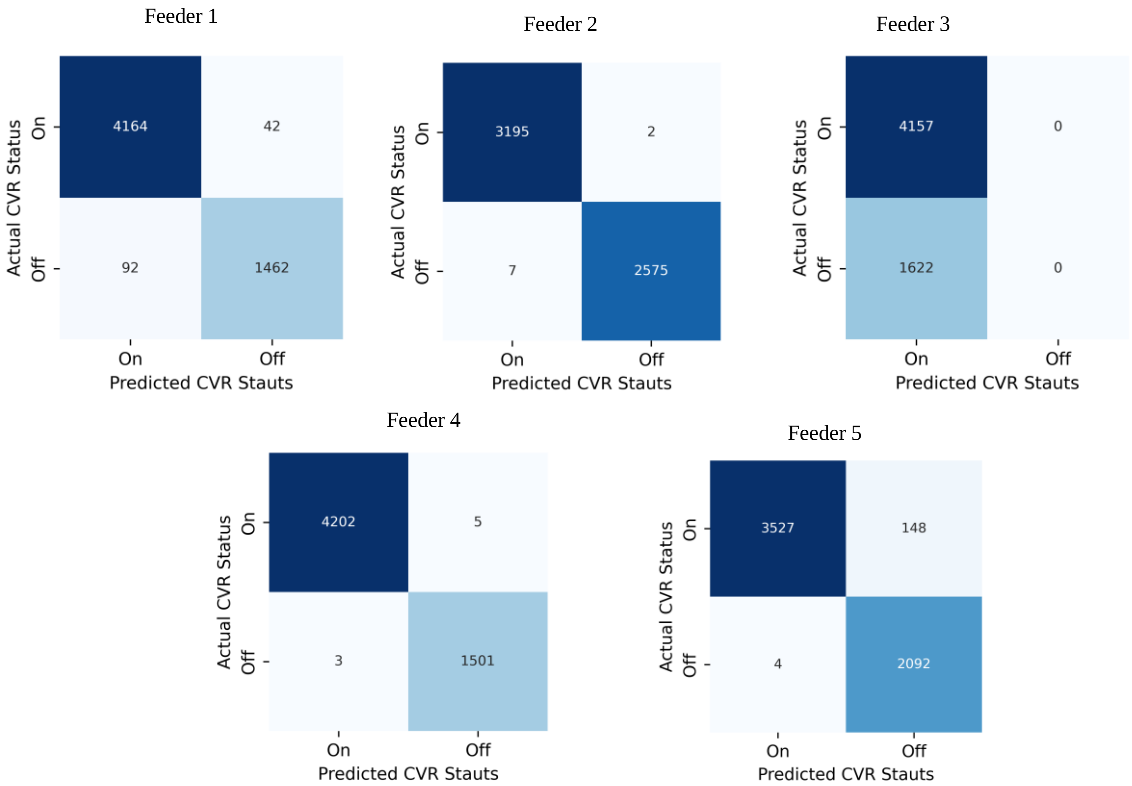


Fig 5(b): Confusion matrix for SVM

CVR Status detection using KNN

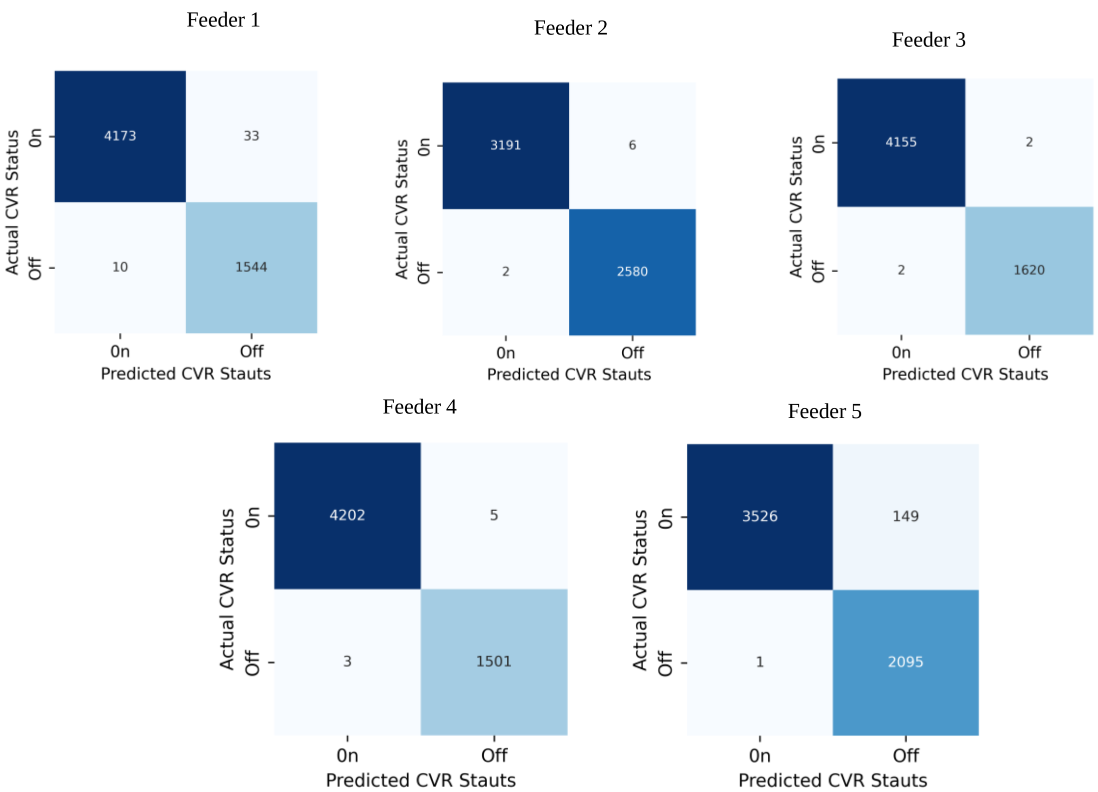


Fig 5(c): Confusion matrix for KNN

Figures show ANN and KNN performs the best in terms of finding true positive and true negative with ANN performing slightly better. Figures also show, performance decreases for Feeder 3 for SVM. Voltage profile of Feeder 3 in Fig 4 shows a voltage shift from the beginning of the year till mid-April. Voltage decreases from a bit higher magnitude to lower magnitude and then settles to a stable level. This voltage shift happens mainly due to shifting loads from one feeder to another. Shift can be both positive and negative. This voltage shift is negatively impacting the performance of SVM since that pushes the linear hyperplane to higher voltage magnitudes. Therefore, the trained model for Feeder 3 through SVM is always predicting CVR is on for the test values. ANN and KNN can learn this non-linear characteristic and, therefore, performs better in this scenario. For SVM, either data can be processed to exclude any information before the shift or radial bias kernel function can be used to improve model performance.

Table 2 shows the Accuracy, Precision, Recall, and F1 score of the selected algorithms. As shown in the table, the algorithms can detect CVR status with high accuracy and precision. Due to voltage shift in Feeder 3, as explained above, performance of the algorithms for Feeder 3 decreases a bit compared to the other feeders for SVM.

Table 2: Performance analysis of ANN, SVM, and KNN

		Accuracy	**Precision**	**Recall**	**F1 score**
Feeder 1	ANN	0.99	0.99	0.99	0.99
	SVM	0.97	0.98	0.97	0.97
	KNN	0.99	0.99	0.99	0.99
Feeder 2	ANN	1	1	1	1
	SVM	1	1	1	1
	KNN	1	1	1	1
Feeder 3	ANN	0.99	0.99	0.99	0.99
	SVM	0.72	1	0.72	0.81
	KNN	0.99	0.99	0.99	0.99
Feeder 4	ANN	1	1	1	1
	SVM	1	1	1	1
	KNN	1	1	1	1
Feeder 5	ANN	0.98	0.98	0.98	0.98
	SVM	0.97	0.97	0.97	0.97
	KNN	0.97	0.98	0.97	0.97

5. CONCLUSION

The paper presents a comprehensive performance analysis of ML-based classification algorithms to detect CVR operational status. The case study uses actual system level feeder data from a utility in the U.S where CVR is implemented. Analysis shows, with the optimal network configuration, the classification algorithms can detect CVR status with high accuracy and precision. This enables an automatic and accurate CVR detection for the M&V process. Furthermore, this approach can be used to detect, real-time, whether CVR is operational in a feeder or not based on historical data. The developed ML models using the proposed methodology can be used to detect CVR status of multiple feeders with high accuracy as demonstrated. This process, therefore, optimizes the use of computational resources and increases the scalability of the application by eliminating the need to manually set threshold for CVR status detection for each individual feeder. The future work of this work includes detecting anomalous data to improve the performance of classification algorithms during CVR status detection.

BIBLIOGRAPHY

[1] Grid Modernization Multi-Year Program Plan, U.S Dept. of energy. Link: https://www.energy.gov/sites/prod/files/2016/01/f28/Grid%20Modernization%20Multi-Year%20Program%20Plan.pdf. Last accessed: July 25,2022

[2] Z. S. Hossein et al., "Conservation Voltage Reduction and Volt-VAR Optimization: Measurement and Verification Benchmarking," in IEEE Access, vol. 8, pp. 50755-50770, 2020, doi: 10.1109/ACCESS.2020.2979242.

[3] W. G. Sunderman, "Conservation voltage reduction system modeling, measurement, and verification," PES T&D 2012, 2012, pp. 1-4, doi: 10.1109/TDC.2012.6281598.

[4] T. L. Wilson, "Measurement and verification of distribution voltage optimization results for the IEEE power & energy society," IEEE PES General Meeting, 2010, pp. 1-9, doi: 10.1109/PES.2010.5589762.

[5] E. Vega-Fuentes, S. León-del Rosario, J. M. Cerezo-Sánchez and A. Vega-Martínez, "Combined comparison-regression method for assessment of CVR effects," 2014 55th International Scientific Conference on Power and Electrical Engineering of Riga Technical University (RTUCON), 2014, pp. 172-176, doi: 10.1109/RTUCON.2014.6998188.

[6] W. Fan et al., "A CVR On/Off Status Detection Algorithm for Measurement and Verification," 2021 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), 2021, pp. 1-5, doi: 10.1109/ISGT49243.2021.9372257.

[7] Z. S. Hosseini et al., "Understanding the Impact of Data Anomalies on Conservation Voltage Reduction Measurement and Verification - A Practical Study," 2022 IEEE/PES Transmission and Distribution Conference and Exposition (T&D), 2022, pp. 1-5, doi: 10.1109/TD43745.2022.9816968.

[8] Haque, Ashraful, and Saifur Rahman. 2022. “Short-Term Electrical Load Forecasting through Heuristic Configuration of Regularized Deep Neural Network.” Applied Soft Computing 122 (June). Elsevier Ltd. doi:10.1016/j.asoc.2022.108877.

[9] A. Haque et al., "An SVR-based Building-level Load Forecasting Method Considering Impact of HVAC Set Points," 2019 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), 2019, pp. 1-5, doi: 10.1109/ISGT.2019.8791649.

[10] H. Livani and C. Y. Evrenosoglu, "A Machine Learning and Wavelet-Based Fault Location Method for Hybrid Transmission Lines," in IEEE Transactions on Smart Grid, vol. 5, no. 1, pp. 51-59, Jan. 2014, doi: 10.1109/TSG.2013.2260421.

[11] Kotsiantis, Sotiris B., Ioannis Zaharakis, and P. Pintelas. "Supervised machine learning: A review of classification techniques." Emerging artificial intelligence applications in computer engineering 160, no. 1 (2007): 3-24.

[12] Osisanwo, F. Y., J. E. T. Akinsola, O. Awodele, J. O. Hinmikaiye, O. Olakanmi, and J. Akinjobi. "Supervised machine learning algorithms: classification and comparison." International Journal of Computer Trends and Technology (IJCTT) 48, no. 3 (2017): 128-138.

[13] Kwon, Ohbyung, and Jae Mun Sim. "Effects of data set features on the performances of classification algorithms." Expert Systems with Applications 40, no. 5 (2013): 1847-1857.

[14] Gudivada, Venkat, Amy Apon, and Junhua Ding. "Data quality considerations for big data and machine learning: Going beyond data cleaning and transformations." International Journal on Advances in Software 10, no. 1 (2017): 1-20.

[15] By Glosser.ca - Own work, Derivative of File:Artificial neural network.svg, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=24913461

[16] Cortes, C., Vapnik, V. Support-vector networks. Mach Learn 20, 273–297 (1995). https://doi.org/10.1007/BF00994018
[17] By Larhmam - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=73710028