Received 5 April 2023; revised 1 December 2023; accepted 22 February 2024. Date of publication 27 February 2024; date of current version 8 March 2024. The review of this article was arranged by Associate Editor Mahesh Illindala.

Digital Object Identifier 10.1109/OJIA.2024.3369993

Machine Learning Techniques to Predict Voltage Unbalance in a Power Transmission System

JONATHAN D. BOYD[1], DONALD R. REISING[1] (Senior Member, IEEE), ANTHONY M. MURPHY[2] (Member, IEEE), JUSTIN D. KUHLERS[2] (Member, IEEE), C. MICHAEL MCAMIS[2] (Member, IEEE), AND JAMES B. ROSSMAN[2] (Senior Member, IEEE)

[1]University of Tennessee at Chattanooga, Chattanooga, TN 37403 USA
[2]Tennessee Valley Authority, Chattanooga, TN 37403 USA

CORRESPONDING AUTHOR: DONALD R. REISING (e-mail:donald-reising@utc.edu)

This work was supported by the Tennessee Valley Authority (TVA), Chattanooga, TN.

ABSTRACT Voltage unbalance is a growing issue that, among other things, can impact three-phase motor and drive loads, result in nuisance tripping of generation units and capacitor banks, and prevent optimization of conservative voltage regulation strategies. This difference between the three phases of voltage delivered to customers can damage the equipment of these customers as well as negatively impact the power system itself. This work presents an approach for predicting voltage unbalance using machine learning. Historical megawatt and megavar data–obtained through a Supervisory Control And Data Acquisition (SCADA) system–are used to train an artificial neural network model as a binary classifier with a portion of the data serving to validate the trained model. Voltage unbalance is predicted at an accuracy above 95% for eight substations within the power utility's extra-high voltage transmission network and over 91% for all 42 substations. The trained model is tested in a manner that would be employed using simulated data generated by state estimation software. This simulated data validates the model's capacity to predict the substation buses that would experience voltage unbalance.

INDEX TERMS Artificial neural network (ANN), automation, classification, prediction, prediction model, supervisory control and data acquisition (SCADA).

I. INTRODUCTION

Voltage unbalance is a measure of the asymmetry between the voltages of a three-phase power system [1]. When power is generated, the resulting voltages on the three phases are equal in magnitude and 120° apart in phase angle [2]. However, voltages on the system itself can become unbalanced in magnitude, phase angle, or harmonic distortion levels. One major driving factor of voltage unbalance is the presence of single-phase loads, which draw power from one phase and not the others. An unequal allocation of these loads will lead to greater unbalance. The fact that many distribution lines are single-phase contributes to these uneven loads. Another known cause of voltage unbalance is the transmission of power across long distances from base load generation to load centers across untransposed transmission lines [1]. The retirement of traditional spinning generation and its replacement with inverter based resources (IBRs) has also contributed to the growing voltage unbalance issue. Voltage unbalance harms the power system, which will incur more losses and heating effects under unbalanced conditions [2]. This is because the system cannot respond to emergency load transfers. The unbalance can also adversely affect large commercial or industrial customers operating large equipment. The effects are particularly severe on induction motors, power electronic converters, and adjustable speed drives (ASDs).

The authors of [3] determined the average voltage unbalance within an operational power transmission system using four seasonal periods for 22 sites and determined that line configuration (e.g., transposed lines) and line length do not significantly influence voltage unbalance. The authors calculated voltage unbalance using Supervisory Control And Data Acquisition (SCADA) and Digital Fault Recorder (DFR) data collection methods. The average voltage unbalance is calculated as 0.59% and 0.64% using the SCADA and DFR data collection methods, respectively. The authors of [3] highlighted that the SCADA system did not detect a 1.5% voltage unbalance condition reported by a large industrial customer; thus, the work in [3] is not predictive and highlights the need for a voltage unbalance prediction method within operational power systems.

Several methods exist for quantifying voltage unbalance, but few exist for predicting it before it occurs. In [4], the authors create a three-phase state estimation framework focused on estimating measurement information due to incomplete or missing information within the power system network and the location, level, and impacts of voltage unbalance within a power distribution network (33 and 11 kV). The authors predict voltage unbalance using statistical estimation using real and reactive power measurements generated by a simulated 24-bus system representing a portion of an operational power distribution network. Due to limited network observability, the authors use pseudomeasurements to “fill in missing data.”

The authors of [5] estimate voltage unbalance using data simulated at the power distribution level. A load flow is performed using the Newton–Raphson method to generate the three-phase voltages. These are then transformed into sequence voltages, which are needed to calculate the voltage unbalance percentage. Probabilistic estimation of voltage unbalance is then performed using Monte Carlo (MC) simulations. The random variation of the power factor–at different buses–highlights which buses are the sources of the voltage unbalance. These methods are then tested on an operational power distribution system, and considering other loading conditions, they are used to determine the expected level of voltage unbalance. Individual voltage unbalance source contributions are summed to determine the total unbalance on any bus. The authors of [6] leverage the findings in [5] to simulate the selection of the optimal locations for a limited number of voltage unbalance monitors that would be placed in the power distribution system.

The authors of [7] developed an algorithm that detects voltage unbalance using the space vector property (SVP), which transforms three voltages into a single complex variable. The algorithm sums the instantaneous values of all three voltages. The authors use a zero-sum to indicate the three-phase voltages are balanced, but adding zero does not guarantee that the voltages are balanced. Thus, the SVP is then compared to a reference space vector to determine whether or not voltage unbalance occurs. The algorithm is tested using five cases, and voltage unbalance is correctly predicted for all five cases.

In [8], the authors use a stochastic approach to predict voltage unbalance within a low-voltage power distribution system due to the presence of single-phase Photovoltaic inverters (PVIs). A stochastic approach is taken because the PVIs’ locations and connected phases are unknown before deployment. The authors show that the random connection of 6 kW PVIs can lead to a voltage unbalance over 1% but is highly improbable of exceeding 2%. The power utility can use this information to quantify the risk associated with the location and concentration of PVI deployments within their system.

The authors of [9] present a voltage unbalance detection approach for three-phase induction motors using an artificial neural network (ANN). A dataset containing one hundred samples–collected over nine days from an operational three-phase induction motor–is used to train the ANN. A feed-forward structure, the most common structure, is used for the ANN. During training, the unbalanced voltages are labeled as “-1” and the balanced voltages are labeled as “1”. The performance of this model is measured using mean squared error (MSE) and root mean squared error (RMSE). The trained ANN correctly detects voltage unbalance with an accuracy of 100%.

Our approach predicts voltage unbalance within an operational, extra-high voltage (EHV) 500 kV transmission network using historical Megawatt (MW), megavolt-ampere reactive (Mvar), or both values collected by a SCADA system [10]. Voltage unbalance prediction is performed within the 500 kV network because it is the “backbone” of the utility’s transmission system. Performing voltage unbalance prediction within an EHV network is advantageous because of the following: 1) it permits corrective action before the unbalance cascades to lower voltage level networks and 2) it is a much simpler network with fewer substations, thus, making modeling and prediction easier. In addition to predicting voltage unbalance within an operational EHV network, our work can be paired with state estimation software. Utility personnel use state estimation software to calculate expected line flows and simulate the impact of taking particular lines out of service. However, state estimation software does not currently predict voltage unbalance. This is important because voltage unbalance is primarily caused by transmission lines operating at their loading limit. This can be exacerbated when other lines are removed from service as part of planned maintenance. Thus, pairing voltage unbalance prediction with state estimation software allows utility personnel to identify potential locations of voltage unbalance and develop and simulate a remediation plan before the planned maintenance is initiated. The contributions of this work are as follows.

1) Voltage unbalance prediction is performed for a 500 kV power transmission network comprised of 42 stations/buses
2) All of the data used in this study are collected from an operational EHV transmission system rather than a simulation.
3) All of the data are obtained by the utility’s SCADA system at a sampling rate of one sample every four seconds. The use of SCADA data make our approach applicable to every transmission utility.
4) This approach shows that voltage unbalance can be predicted using only MW or Mvar data, which is appealing

since three-phase current or voltage data are not always available.

5) Each station's voltage unbalance prediction model is developed and tested using inputs that include data from all lines and substations in the entire EHV network. This approach ensures the interconnection between stations is accounted for within the model.

The rest of this article is organized as follows. Section II presents information on quantifying voltage unbalance and the ANN design. Section III presents the methodology used in gathering the data, ANN training & testing, and a state estimation software line outage study used to assess the trained model's effectiveness in predicting voltage unbalance. Section IV presents the trained model's voltage unbalance prediction performance results obtained using both historical data and the simulated state estimator data. Finally, Section V concludes this article.

II. BACKGROUND

This section explains the voltage unbalance measurement method used in this work–since multiple methods exist–and the ANN.

A. VOLTAGE UNBALANCE MEASUREMENT

The International Electrotechnical Commission (IEC) voltage unbalance measurement method uses a percentage known as the voltage unbalance factor (VUF) that is given by [11]

$$u_2 = \frac{|\boldsymbol{V}_2|}{|\boldsymbol{V}_1|} \times 100\% \tag{1}$$

where u_2 is the percent VUF, $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$ are the positive and negative sequence voltages, respectively. The calculation for $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$ has its basis in the theory of symmetrical components. The Fortescue transformation translates voltages from the phase domain to the sequence domain by

$$\begin{bmatrix} \boldsymbol{V}_0 \\ \boldsymbol{V}_1 \\ \boldsymbol{V}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{A} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{V}_{ab} \\ \boldsymbol{V}_{bc} \\ \boldsymbol{V}_{ca} \end{bmatrix} \tag{2}$$

where

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & \boldsymbol{a} & \boldsymbol{a}^2 \\ 1 & \boldsymbol{a}^2 & \boldsymbol{a} \end{bmatrix}, \ \boldsymbol{a} = -\frac{1}{2} + j\frac{\sqrt{3}}{2} \tag{3}$$

$\boldsymbol{V}_0$, $\boldsymbol{V}_1$, and $\boldsymbol{V}_2$ are the zero, positive, and negative sequence voltages, respectively, and $\boldsymbol{V}_{ab}$, $\boldsymbol{V}_{bc}$, and $\boldsymbol{V}_{ca}$ are the three phase-to-phase voltage phasors, respectively [1]. A drawback to the Fortescue transformation is that the calculation of the sequence voltages requires knowledge of the voltage magnitudes and phases, which are generally unavailable in SCADA data. The Fortescue transformation's drawback is overcome by calculating the VUF using (4), which is equivalent to (1) [11]

$$u_2 = \sqrt{\frac{1 - \sqrt{3 - 6\beta}}{1 + \sqrt{3 - 6\beta}}} \times 100\% \tag{4}$$

where

$$\beta = \frac{|V_{ab}|^4 + |V_{bc}|^4 + |V_{ca}|^4}{\left(|V_{ab}|^2 + |V_{bc}|^2 + |V_{ca}|^2\right)^2}.$$

Equation (4) removes the phase angle component of the VUF calculation to permit the use of SCADA system-measured phase-to-phase quantities.

B. ARTIFICIAL NEURAL NETWORK DESIGN

An ANN performs pattern recognition by mimicking the neurons and synapses of the human brain using a collection of interconnected nodes (a.k.a., artificial neurons). Like the human brain, trained ANNs can recognize complex nonlinear input-output relationships [12]. The ANNs used herein are feed-forward networks–the most popular neural network architecture–trained using supervised learning [13]. In supervised learning, labeled input data are used to train the ANN to learn the nonlinear patterns and relationships between the inputs and a desired output or set of outputs. In this work, the labeled input data are the SCADA collected MW, Mvar, or MW and Mvar values (a.k.a., the predicting features) along with their voltage unbalance status (a.k.a., the labels). The desired output is the prediction of voltage unbalance or voltage unbalance (a.k.a., the target features). All inputs corresponding to voltage unbalance are labeled using a '1', and all others are labeled '0'.

The block diagram in Fig. 1 shows the adopted ANN architecture. N_M data vectors are input into the ANN that is constructed with ten feature extracting, hidden layers, one output layer, and an output vector that contains the two possible class predictions of voltage unbalance or not [12]. The values for the number of input measurements (N_M), number of substations (N_S), and number of ten-minute periods (N_t) are given in Section III-A.

III. METHODOLOGY

This section describes the processes used to construct a dataset, preprocess the dataset, and train the ANN used for voltage unbalance prediction. This section also describes the process of simulating a transmission line outage and how voltage unbalance is predicted.

A. DATA COLLECTION AND PREPROCESSING

Voltage unbalance prediction is performed using MW, Mvar, or both measurements. The use of MW and Mvar is motivated by the fact that they are three-phase measurements that capture the three-phase nature of the power system, making them advantageous for cases in which three-phase currents or voltages are unavailable. The utility's SCADA system collects and transmits all the data via Distributed Network Protocol 3 (DNP3) to a central database to store the data from all

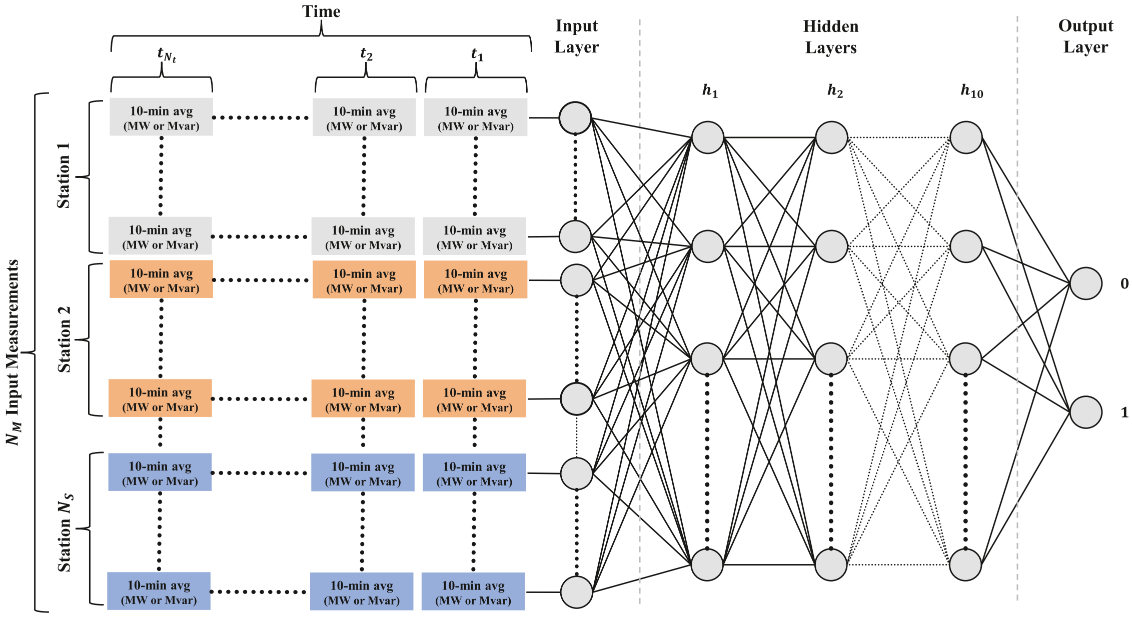

FIGURE 1. Representative diagram of an ANN.

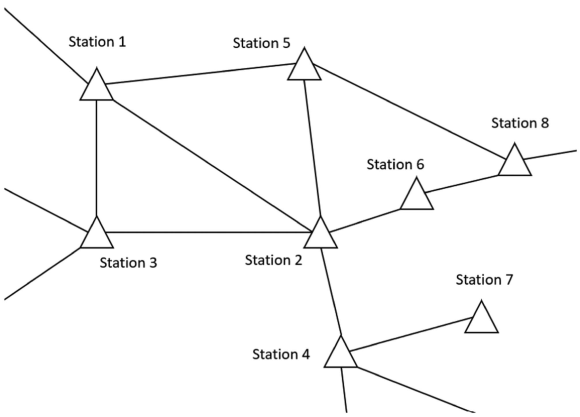

FIGURE 2. Initial, eight station portion of the 500 kV power transmission system studied in this work.

transmission system substations. The SCADA system records each substation's MW and Mvar measurements every 4 s, which results in a sampling rate of fifteen measurements per minute. However, voltage unbalance is not typically observed over short periods; thus, the average of the MW or Mvar measurements is used. One average MW or Mvar measurement is calculated using ten minutes of SCADA-measured MW or Mvar values. This process is repeated for every transmission line and transformer associated with the selected substations over the selected period.

Using the averaged MW and Mvar values, two datasets are constructed. The first dataset is built using SCADA MW and Mvar measurements accumulated over a 28-month period–from January 1, 2020, to April 30, 2022–for $N_S = 8$ substations (see Fig. 2) within the utility's 500 kV transmission system. These eight substations contain measurements from transmission lines or transformers; there are a total of $N_M = 70$, individual, average MW or Mvar values per 10-min period and a total of $N_t = 122\,000$ 10-min periods over the 28-month period. Each 10-min interval–consisting of seventy individual, average MW or Mvar values–constitutes one sample used to train, validate, or blind test the ANN. The averaged MW and Mvar measurements are interleaved, meaning that MW and Mvar measurements are alternated as they are fed into the ANN.

The second dataset spans thirty months–from January 1, 2020, to June 30, 2022–and $N_S = 42$ substations, encompassing the entirety of the utility's EHV transmission system. This second dataset is comprised of $N_t = 131\,000$ 10-min averages. Each 10-min average consists of $N_M = 374$ MW or Mvar values–one per transmission line or transformer–each constitutes one sample used to train, validate, or blind test the ANN.

B. ANN TRAINING AND VALIDATION

During ANN training, each input is assigned a class label corresponding to voltage balance or unbalance. The label assigned is determined by calculating the VUF using (4) for each set of three voltage values measured on the three phases (a.k.a, phases A, B, and C). In cases where the voltage is measured from the line rather than the bus, the percent VUF–for the lines at a given substation–is averaged together. VUF values over a threshold of 1.4% are assigned a class label of '1' (a.k.a., unbalanced), while values below that threshold are assigned the class label of '0' (a.k.a., balanced). This threshold is the same as defined in IEC 61000-3-13 as the planning level for high-voltage (HV) systems [11]. Though the system studied here is at the EHV level, the 1.4% threshold was adopted by the utility.

Both datasets are divided into training and “blind” testing subsets in which 80% of the data are randomly assigned to the training set and the remaining 20% assigned to the testing set. This partitioning can be changed based on individual needs. Each training data are then normalized to ensure that the MW and Mvar values–of all transmission lines and transformers–are in the range of zero and one. The blind test sets are normalized using their corresponding training set’s normalization values.

An ANN is created for each substation or bus using the inputs and labels in the training dataset. The inputs for each substation’s unique model are the 10-min average MW and Mvar measurements for all lines and transformers in the studied region, which accounts for the interconnection of the system. Each station’s model is trained using k =5-fold cross-validation [14], tested using the blind data subset (i.e., data not used during ANN training), and the ANN’s class assignment compared with the “blind” testing set’s known labels. This allows for determining whether the trained ANN correctly predicts balanced voltage (Class 0) or unbalanced voltage (Class 1). Each sample that is not classified correctly is counted as an error. The ANN’s percent correct classification performance is calculated by

$$\%\ \text{Accuracy} = \frac{N_s - N_e}{N_s} \times 100\% \tag{5}$$

where N_s is the number of samples in the blind testing dataset and N_e is the number of classification errors produced.

C. TESTING IN A LINE OUTAGE STUDY

As an additional test of the developed voltage unbalance prediction approach, the trained ANN is tested in an outage study conducted by utility personnel. These outage studies involve state estimation software to simulate what would happen in the transmission system should a line be removed from service. These studies are conducted weeks or months ahead of a scheduled line outage; thus, the utility knows in advance how a line outage will affect the loading of the lines in the system. However, these studies are currently unable to predict voltage unbalance. Voltage unbalance on the EHV system is often a function of the line loading, so removing a critical line from service would cause very high power flows on the remaining lines. The ANN voltage unbalance prediction approach augments these outage studies so that the impact of a line outage on voltage unbalance is included. The state estimation tool–used in this case study–is not inherently capable of predicting voltage unbalance because it only outputs a single-phase voltage reading, thus, preventing the calculation of VUF using (4). However, MW and Mvar readings are produced by the simulation so that these measurements can be used as inputs to the developed prediction model.

The goal is to determine whether voltage unbalance–resulting from a previous line outage–could have been accurately predicted using the trained ANN. Voltage unbalance occurred when the transmission line–connected to Substation 4 in Fig. 2 and connected to a substation not shown–was removed from service on May 4, 2020. Since the voltage unbalance event is present in the first dataset, it is assumed that its associated ANN learned the event’s patterns or features. The simulation case was created on April 27, 2022, to remove the same line from service on May 4, 2022, corresponding to the exact calendar date of the unbalance event on May 4$^{\text{th}}$, two years prior. The system is configured as it is on April 27, 2022, except that a particular line is opened. The MW and Mvar measurements from this case study are then recorded and used as the inputs to the second dataset’s trained prediction model. The outputs of the ANN (either balanced or unbalanced) are then compared with the “true” values of unbalance from the event two years before to gauge the model’s prediction accuracy. The previous event’s unbalance measurements is only used as estimates since the impacts of opening the same line today are not fully known.

TABLE 1. Voltage Unbalance Prediction Results With an Overall Average Accuracy of 99.14% Using the MW and Mvar Values of the First Dataset Associated With the Eight Substations Shown in Fig. 2

Station-Bus	% Above Threshold	% Accuracy
1-1	1.09%	99.95%
1-2	0%	99.99%
2-1	0.19%	99.92%
2-2	3.61%	98.66%
3-1	0%	100%
3-2	0.01%	100%
4-1	0.55%	99.94%
5-1	0.12%	99.93%
6-1	20.90%	96.75%
7-1	2.01%	99.28%
8-1	11.94%	99.41%
8-2	2.71%	98.82%

IV. RESULTS

An ANN is trained for each substation or bus, and the accuracy is calculated using (5). All results are presented using tables in which the first column indicates the substation and bus. For example, a substation number of “2–1” corresponds to Substation 2, Bus 1. The second column shows the percentage of the blind test set entries above the 1.4% voltage unbalance threshold for each substation and bus. In contrast, the third column provides the average percent correct classification performance.

A. RESULTS: FIRST DATASET

The first dataset’s blind test results are displayed in Table 1 for the eight substations in Fig. 2. The results shown in Table I are very accurate, with the lowest accuracy being 96.75% for Substation 6, Bus 1. The issue with some stations is that there are too few or no data points above the 1.4% VUF threshold,

TABLE 2. Voltage Unbalance Prediction Results With an Overall Average Accuracy of 92.65% Using the MW and Mvar Values of the Second Dataset

Station Number	% Above Threshold	Average % Correct	Station Number	% Above Threshold	Average % Correct
1	2.96%	91.80%	22	0.30%	92.49%
2	0.16%	92.28%	23	11.70%	90.54%
3	0.01%	92.42%	24	0.35%	92.28%
4	0.00%	92.42%	25	0.24%	92.41%
5	0.00%	100.00%	26	19.20%	89.88%
6	11.40%	92.16%	27	0.06%	92.41%
7	1.06%	92.44%	28	0.59%	92.34%
8	1.54%	92.02%	29	2.46%	91.66%
9	0.00%	92.47%	30	0.00%	100.00%
10	0.00%	92.32%	31	0.48%	92.28%
11	0.03%	92.31%	32	0.48%	92.33%
12	0.00%	92.58%	33	0.11%	92.46%
13	0.02%	92.42%	34	0.03%	92.48%
14	0.01%	92.38%	35	0.90%	92.40%
15	0.00%	92.43%	36	0.00%	94.00%
16	0.01%	92.48%	37	0.10%	92.29%
17	0.01%	92.44%	38	0.01%	92.34%
18	0.00%	92.39%	39	0.00%	94.02%
19	0.02%	92.40%	40	0.01%	92.41%
20	0.00%	92.53%	41	0.01%	92.46%
21	0.00%	92.35%	42	6.56%	90.72%

TABLE 3. Voltage Unbalance Prediction Results With an Overall Average Accuracy of 92.72% Using Only MW Values of the Second Dataset

Station Number	% Above Threshold	Average % Correct	Station Number	% Above Threshold	Average % Correct
1	2.96%	92.04%	22	0.30%	92.45%
2	0.16%	92.37%	23	11.70%	90.59%
3	0.01%	92.51%	24	0.35%	92.48%
4	0.00%	92.56%	25	0.24%	92.34%
5	0.00%	100.00%	26	19.20%	89.44%
6	11.40%	92.01%	27	0.06%	92.59%
7	1.06%	92.49%	28	0.59%	92.40%
8	1.54%	92.21%	29	2.46%	91.67%
9	0.00%	92.66%	30	0.00%	100.00%
10	0.00%	92.53%	31	0.48%	92.37%
11	0.03%	92.52%	32	0.48%	92.49%
12	0.00%	92.41%	33	0.11%	92.53%
13	0.02%	92.51%	34	0.03%	92.59%
14	0.01%	92.55%	35	0.90%	92.52%
15	0.00%	92.74%	36	0.00%	94.06%
16	0.01%	92.49%	37	0.10%	92.57%
17	0.01%	92.59%	38	0.01%	92.60%
18	0.00%	92.60%	39	0.00%	94.10%
19	0.02%	92.55%	40	0.01%	92.53%
20	0.00%	92.46%	41	0.01%	92.53%
21	0.00%	92.50%	42	6.56%	90.21%

thus, impeding the ANN's ability to predict voltage unbalance at the corresponding substation or bus accurately. A possible solution to this problem would be either of the following: 1) find more data for model training further in the past or 2) lower the threshold for those particular stations or buses to something lower than 1.4% since the thresholds are adaptable for each station or bus. The results are still very accurate for the stations with more data above the threshold. This demonstrates the robustness of the training algorithm and shows the connection between line MW and Mvar loading and voltage unbalance.

B. RESULTS: SECOND DATASET

Voltage unbalance prediction results are presented in Tables 2, 3, and 4 when using *MW and Mvar*, *only MW*, and *only Mvar* values to represent all transmission lines and transformers within the utility's 500 kV system (a.k.a., 42 substations in total), respectively. The results in Tables 2, 3, and 4 allow the voltage unbalance predicting contributions of each measurement and their contribution to be determined. The *MW and Mvar* as well as *only MW* results are very similar and the ANN's voltage unbalance prediction accuracy is poorer than the eight substation results in Section IV-A. This is attributed

TABLE 4. Voltage Unbalance Prediction Results With an Overall Average Accuracy of 95.33% Using Only Mvar Values of the Second Dataset

Station Number	% Above Threshold	Average % Correct	Station Number	% Above Threshold	Average % Correct
1	2.96%	94.79%	22	0.30%	95.28%
2	0.16%	95.13%	23	11.70%	93.30%
3	0.01%	95.32%	24	0.35%	95.21%
4	0.00%	95.33%	25	0.24%	95.20%
5	0.00%	100.00%	26	19.20%	91.75%
6	11.40%	94.77%	27	0.06%	95.36%
7	1.06%	95.24%	28	0.59%	95.13%
8	1.54%	94.94%	29	2.46%	94.36%
9	0.00%	95.36%	30	0.00%	100.00%
10	0.00%	95.28%	31	0.48%	95.12%
11	0.03%	95.32%	32	0.48%	95.23%
12	0.00%	95.24%	33	0.11%	95.28%
13	0.02%	95.27%	34	0.03%	95.38%
14	0.01%	95.38%	35	0.90%	95.33%
15	0.00%	95.47%	36	0.00%	96.31%
16	0.01%	95.29%	37	0.10%	95.34%
17	0.01%	95.33%	38	0.01%	95.31%
18	0.00%	95.35%	39	0.00%	96.29%
19	0.02%	95.36%	40	0.01%	95.31%
20	0.00%	95.31%	41	0.01%	95.32%
21	0.00%	95.31%	42	6.56%	93.23%

to the large amount of data (i.e., 374 MW or Mvar values for each of the 131 000 ten-minute averages), thus, the ANN must discriminate between more cases. Using *only Mvar* measurements results in the highest average voltage unbalance prediction accuracy of 95.33% which is 2.68% higher than the *MW and Mvar* case in Table 2 and 2.61% higher than the *only MW* case in Table 3. One explanation for the improvement in voltage unbalance prediction accuracy–when using *only Mvar* measurements–could be the transmission lines' untransposed nature. When transmission lines are not transposed, the transmission lines' inductance and capacitance are different across the three phases while the resistance is essentially the same on all three. Since inductors consume and capacitors produce Mvar, this would impact Mvar flow. There is also a known relationship between Mvar and voltage. The work in [15] discusses ways to improve the VUF of a power system through different methods of injecting Mvar. Also, the fast decoupled power flow method can separate MW and phase angle from Mvar and voltage magnitude [16]. Thus, it makes sense that Mvar is found to be the main factor in predicting voltage unbalance.

C. RESULTS: LINE OUTAGE STUDY

The ANN trained using the first dataset is further tested using the outage study described in Section III-C, and the corresponding voltage unbalance prediction results are shown in Table 5. The first column contains the same station and bus numbers as Table 1, the second column lists whether or not voltage unbalance occurred during the actual event two years ago, and the third column lists whether or not the ANN predicts the presence of voltage unbalance in the simulated line outage conducted in the present day.

TABLE 5. Voltage Unbalance Prediction Results for a Line Outage Study Using State Estimation Software

Station-Bus	Unbalanced Before?	Unbalanced Now?
1-1	No	Yes
1-2	No	Yes
2-1	Yes	Yes
2-2	Yes	Yes
3-1	No	Yes
3-2	No	No
4-1	Yes	Yes
5-1	Yes	Yes
6-1	Yes	Yes
7-1	Yes	Yes
8-1	Yes	No
8-2	Yes	Yes

The trained ANN is very accurate when used in the line outage study. The outputs from the simulation–balanced or unbalanced–are compared to the historical unbalance values as an approximation of the ANN's performance. One issue with this comparison is that the system is not configured in the same way as it was when the historical event occurred, so it is not known what the actual voltage unbalance values would be if the line were opened today. A future test should conduct an outage study on a line that will be opened in the future, predict whether unbalance occurs, and compare the results with the actual values after the line is opened. That was not feasible in this case as the line known to cause the most voltage unbalance was not scheduled to be removed from service in the near future.

When comparing the line outage study's voltage prediction results to those of the historical event, the ANN performs well overall. Station 1 Bus 1's and Station 3 Bus 1's historical events have VUF values of 1.38% and 1.26%, respectively. Both values are very close to the 1.4% threshold, so the fact that the ANN predicts unbalance is not a serious issue and makes it slightly more secure. In the historical line outage, Bus 2 at Station 1 was de-energized, so the ANN did not have an example within the training data as to whether or not voltage unbalance would occur due to the simulated outage. The misclassification of Station 8 Bus 1 is not immediately apparent. Still, it is attributed to conflicting data within the training data regarding the VUF value(s) when the simulated line is removed from service within the operational system. Another explanation could be that the difference in system configuration between the historical event and the current/simulated event could be enough to cause Station 8 Bus 1 to not result in voltage unbalance. Overall, the ANN is biased toward predicting voltage unbalance even if it does not eventually occur, which is preferable because the opposite bias could cause significant voltage unbalance risks to be missed in outage studies.

V. CONCLUSION

This article presents an approach for voltage unbalance prediction using an ANN and historical SCADA measurements of MW and Mvar, only MW, and only Mvar. An initial investigation focused on eight substations within an operational EHV system. It showed that the trained ANN can predict voltage unbalance at an accuracy of 91% or higher for each of the 70 transmission lines and transformers. The study was expanded to include all 42 substations, and it was determined that using *only Mvar* measurements results in the highest average voltage unbalance prediction accuracy of 95.33% compared to 92.65% and 92.72% when using *MW and Mvar* and *only MW*, respectively. Additionally, the voltage prediction capability of the trained ANN is validated using a line outage study conducted using the power utility's state estimation software. The line outage study removed the same line from service for which there was a historical case of voltage unbalance within the operation 500 kV system. Our voltage unbalance prediction process will save utilities time and money by reducing voltage unbalance-induced damage to power system equipment and increasing customer satisfaction by lowering damage to their equipment. Future work will include the following: 1) gathering more training data to represent all cases of voltage balance or unbalance, 2) performing an outage study before an actual line is removed from service and studying the real-world impact, and 3) incorporating this voltage unbalance prediction approach into the state estimation software used by power utilities to conduct line outage studies.

REFERENCES

[1] T. Seiphetlho and A. Rens, "On the assessment of voltage unbalance," in *Proc. 14th Int. Conf. Harmon. Qual. Power*, 2010, pp. 1–6.
[2] A. von Jouanne and B. Banerjee, "Assessment of voltage unbalance," *IEEE Trans. Power Del.*, vol. 16, no. 4, pp. 782–790, Oct. 2001.
[3] J. Rossman, T. Laughner, A. Murphy, D. Marler, and G. Kobet, "Transmission voltage unbalance evaluation," in *Proc. Future Instrum. Int. Workshop*, 2012, pp. 1–4.
[4] N. Woolley and J. Milanovic, "Statistical estimation of the source and level of voltage unbalance in distribution networks," *IEEE Trans. Power Del.*, vol. 27, no. 3, pp. 1450–1460, Jul. 2012.
[5] Z. Liu and J. Milanović, "Probabilistic estimation of voltage unbalance in MV distribution networks with unbalanced load," *IEEE Trans. Power Del.*, vol. 30, no. 2, pp. 693–703, Apr. 2015.
[6] Z. Liu, H. Liao, J. V. Milanović, T. Guo, and X. Tang, "Optimal monitor placement for voltage unbalance based on distribution network state estimation," in *Proc. IEEE Manchester PowerTech*, 2017, pp. 1–6.
[7] J. Ghaeb and J. Chebil, "Prediction of voltage unbalance employing space vector property," *Int. J. Eng. Res. Develop.*, vol. 12, no. 12, pp. 65–70, 2016.
[8] D. Schwanz, F. Möller, S. Rönnberg, J. Meyer, and M. Bollen, "Stochastic assessment of voltage unbalance due to single-phase-connected solar power," *IEEE Trans. Power Del.*, vol. 32, no. 2, pp. 852–861, Apr. 2017.
[9] M. Okelola and O. Elijah, "Detection of voltage unbalance on three phase induction motor using artificial neural network," *Int. J. Emerg. Trends Eng. Develop.*, vol. 4, no. 8, pp. 18–25, 2018.
[10] *Electric Power Systems and Equipment - Voltage Ratings (60 Hertz)*, ANSI C84.1-2011, American National Standards Institute, Inc., Rosslyn, VA, USA, Dec. 2011.
[11] IEC, "Electromagnetic compatibility (EMC)–Part 3-13: Limits–assessment of emission limits for the connection of unbalanced installations to MV, HV and EHV power systems," Int. Electrotechnical Commission, Geneva, Switzerland, Tech. Rep. IEC/TR 61000-3-13, Feb. 2008. [Online]. Available: http://www.normservis.cz/download/view/iec/info_iec61000-3-13%7Bed1.0%7Den.pdf
[12] J. Basu, D. Bhattacharyya, and T.-H. Kim, "Use of artificial neural network in pattern recognition," *Int. J. Softw. Eng. Appl.*, vol. 4, no. 2, pp. 23–33, 2010.
[13] A. Jain, J. Mao, and K. Mohiuddin, "Artificial neural networks: A tutorial," *Computer*, vol. 29, no. 3, pp. 31–44, 1996.
[14] T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Berlin, Germany: Springer, 2009, vol. 2.
[15] J. Ghaeb, M. Alkayyali, and T. Tutunji, "Wide range reactive power compensation for voltage unbalance mitigation in electrical power systems," *Electric Power Compon. Syst.*, vol. 49, no. 6/7, pp. 715–728, 2022.
[16] R. Portelinha, C. Durce, O. Tortelli, and E. Lourenço, "Fast-decoupled power flow method for integrated analysis of transmission and distribution systems," *Electric Power Syst. Res.*, vol. 196, 2021, Art. no. 107215.