IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 20 January 2023, accepted 14 February 2023, date of publication 22 February 2023, date of current version 28 February 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3247947

RESEARCH ARTICLE

A Bayesian Deep Learning Approach With Convolutional Feature Engineering to Discriminate Cyber-Physical Intrusions in Smart Grid Systems

DEVINDER KAUR[1], ADNAN ANWAR[2], (Member, IEEE), INNOCENT KAMWA[3], (Fellow, IEEE), SHAMA ISLAM[1], (Member, IEEE), S. M. MUYEEN[4], (Senior Member, IEEE), AND NASSER HOSSEINZADEH[1], (Senior Member, IEEE)

[1]School of Engineering, Deakin University, Geelong, VIC 3220, Australia
[2]Centre for Cyber Security Research and Innovation (CSRI), School of Information Technology, Deakin University, Geelong, VIC 3220, Australia
[3]Department of Electrical Engineering and Computer Engineering, Laval University, Quebec City, QC G1V 0A6, Canada
[4]Department of Electrical Engineering, Qatar University, Doha, Qatar

Corresponding authors: Adnan Anwar (adnan.anwar@deakin.edu.au) and S. M. Muyeen (sm.muyeen@qu.edu.qa)

This work was supported in part by the Faculty of Science, Engineering and Built Environment, Deakin University, under the Mini Australian Research Council (ARC) Analog Program; in part by Laval University from the Canada National Sciences and Engineering Research Council (NSERC) under Grant ALLRP 567550—21. Open Access funding provided by the Qatar National Library.

ABSTRACT The emergence of cyber-physical smart grid (CPSG) systems has revolutionized the traditional power grid by enabling the bidirectional energy flow between consumers and utilities. However, due to escalated information exchange between the end-users, it has posed a greater challenge to the cyber security mechanisms for the communication networks at the cyber and physical planes. To address these challenges, we propose a Bayesian approach integrated with deep convolutional neural networks (CNN-Bayesian). While, the Bayesian component is used to discriminate cyber-physical intrusions from the normal events in the binary and multi-class events. CNN layers are utilized to handle the high-dimensional feature space prior to the intrusions classification task. The proposed method is validated using real-time Industrial control systems (ICS) dataset against the standard deep learning-based classification methods such as recurrent neural networks (RNN) and long-short term memory (LSTM). From the experimental results, it can be inferred that the proposed CNN-Bayesian method outperforms the existing benchmark classification methods to discriminate intrusions in CPSG systems using evaluation metrics such as accuracy, precision, recall, and $F1$-score.

INDEX TERMS Bayesian inference, cybersecurity, deep learning, intrusion-detection systems, SCADA, smart grid.

I. INTRODUCTION

The modern power grid widely recognized as smart grid (SG) involves bidirectional exchange of energy and information between the end-users [1]. It comprises of advanced metering and communication technologies having control algorithms implemented at cyber plane of the cyber physical smart grid (CPSG) systems contributing to the grid robustness and digitalization [2]. While, the conventional power grid relies on supervisory control and data acquisition (SCADA) systems for monitoring and control applications, CPSG systems employ advanced technologies such as phasor measurement units (PMUs) for more sophisticated control operations and high resolution monitoring. PMUs send the monitoring information to phasor data concentrators (PDCs) and receive control signals in return, through communication channels,

The associate editor coordinating the review of this manuscript and approving it for publication was Payman Dehghanian.

thus involving the cyber space in control operations. As the network size and different attack types continue to grow, these communication networks have been exposed to greater cyber vulnerabilities more than ever before [3]. Consumers can now directly interact with the grid using smart appliances, thus increasing the probability of cyber intrusions in the CPSG environment [4]. Moreover, the disturbances can occur at both cyber and physical planes caused by either natural or man-made attacks [5]. Therefore, it becomes crucial to discriminate the power system disturbances to support the cyber attacks detection and handling capabilities.

Thus, beside a reliable communication topology such as mesh [6], it is imperative to have a robust classification mechanism incorporated with the control algorithm to detect potential anomalous events in CPSG communication networks. In this regard, an effective intrusion detection system (IDS) can work in the defense of CPSG cyber security. IDS is a dominant security mechanism which can utilize the data availability from advanced metering infrastructure (AMI) and PMUs related to synchrophasor measurements, and relays.

Furthermore, the process of detecting and classifying the cyber intrusions can be automated using data driven intelligent techniques such as machine learning (ML) and deep learning (DL) [7], [8]. In the past decade, traditional ML algorithms such as random forest and decision trees have successfully provided potential solutions for classifying cyber attacks and detecting system disturbances in CPSG systems [9], [10], [11], [12]. However, ML-based techniques can only detect features manually from the network traffic data and do not provide high performance and detection accuracy [13]. In this direction, advanced feature engineering techniques such as convolutional neural layers from the DL domain can be highly effective and need to be further investigated [14]. Moreover, DL methods can be used to generate features automatically without much human intervention and thus, enhancing the IDS performance overall.

Though various traditional and new learning techniques have been proposed so far to provide potential solutions to classify and detect cyber intrusions in CPSG systems, the benchmark DL algorithms are deterministic in nature and fail to quantify the uncertainties present in the DL model parameters [15]. In this regard, probabilistic approaches for data classification may effectively address uncertainty challenges. To be specific, probabilistic distributions integrated with neural network layers can be highly effective to provide probabilistic solutions for various classification tasks in CPSG domain [16].

In this paper, we attempt to provide a holistic solution combining the feature extraction capabilities of convolutional deep neural layers with advanced probabilistic layers. The proposed methodology takes advantage of the Bernoulli distribution implemented with deep neural networks as an integrated method working towards reducing the false positive and false negative rates.

The rest of this paper is organized as follows. Section II discusses related work and research challenges related to IDS methods based upon machine learning and deep learning. Section III elaborates the proposed methodology and pseudocode for the proposed CNN-Bayesian algorithm. Section IV outlines the implementation results and discussions using a case study carries on ICS dataset. Finally, Section V concludes the paper and briefs the future work.

II. RELATED WORK AND RESEARCH CHALLENGES FOR INTRUSION-DETECTION SYSTEMS (IDS) IN CPSG

This section discusses the related work for existing IDS methodologies in CPSG systems using state-of-the-art ML and DL detection algorithms.

The Table 1 outlines the benefits and limitations of conventional IDS methods in CPSG systems.

A. ML-BASED CPSG-IDS

In order to detect and handle cyber intrusions in CPSG, it is imperative to first classify them accurately [2], [17]. In the recent years, various ML-based IDS techniques have been explored to detect malicious events in the traditional security networks [18], [19], [20], [21], [22]. In this regard, the authors in [23] investigated ML-based extreme gradient boosting (XGBoost) classifier integrated with genetic algorithm to detect intrusions and attacks in wireless sensor networks. The authors claimed extreme XGBoost to be more accurate than standalone XGBoost due to its ability to detect minority classes in highly imbalanced traffic data with 99.9% accuracy for normal classes.

Furthermore, the authors in [24] utilized different ML classifiers such as Adaboost, support vector machine (SVM), random forest (RF), etc., to discriminate power system disturbances. The authors demonstrated that RF achieves highest overall accuracy to classify CPSG cyber intrusions from natural events with minimum false positive rate (FPR). Similarly, the authors in [25] utilized an ensemble approach combining random trees (RT) with random subspace (RS) methods to detect cyber attacks in SCADA systems. Although the proposed technique is demonstrated to be more scalable and insusceptible to overfitting, it is not well equipped for datasets with lesser number of features. Given that not all features in SCADA data will be required for identifying attack patterns, the authors in [26] optimize the features using flora optimisation after performing mean shift clustering and then applied the Boltzmann machine learning algorithm to classify attack types based on the optimised features.

A similar approach has been adopted by the authors in [27] where network data is clustered using Markov Chain clustering, features are optimised using rapid probabilistic correlated optimization and then block correlated neural network model classifies the labels to identify attack data. To reduce the false positive and false negative outputs, the authors in [28] utilised an weighted-intrusion based cuckoo search

TABLE 1. Benefits and limitations of conventional IDS methods in CPSG systems.

Sr. no.	Method	Benefits	Limitations
1	Standard machine learning methods (eg., Support vector machines)	Ideal to produce generalized results and with larger number of attributes	Require feature selection while training
2	Hybrid methods (eg., Naive Bayes and decision trees)	Higher performance and optimum results	Increased complexity
3	Ensemble methods (eg., Random forest and gradient boosting algorithms)	Improved prediction accuracy with the help of a voting system	Complicated model architecture
4	Artificial neural networks (ANNs)	Capable of capturing complex relationships between input attributes and classification labels with deep neural layers	Suffer from local minima, time consuming

method and graded neural network to classify the anomalies in SCADA systems.

B. DL-BASED CPSG-IDS

Artificial neural networks (ANN)-based DL algorithms make advantage of huge data generation from AMIs in CPSG systems [29] and, thus can be used for attacks classification and detection tasks effectively. In this direction, CNN has been recently used to extract spatial features in a hybrid approach along with long-short term memory (LSTM) neural networks used to extract temporal nonlinear sequences [30]. This integrated method is trained on three separate datasets reporting significant improvements in accuracy, precision, recall, and F1 scores due to feature extraction capabilities of CNN layers. A recent advancement in this domain includes physics aware graph convolutional neural networks which considers physical configuration of the system to develop the graph neural networks. This method has been implemented for power system state estimation [31] and forecasting the operating states with reduced number of model parameters and reduced computational complexity [32]. However, this method has not been tested yet for intrusion detection problems.

On the other hand, the authors in [33] have autoencoders as the DL algorithm with Gaussian mixture models to detect anomalies in control area networks. To be specific, autoencoders have been used to extract features before they are fed to the Gaussian models to classify normal events from the attack events when sending messages in vehicle intercommunication networks. With this DL approach, authors conducted various experiments on network and vehicular datasets and reported 6.4% of improvement in attacks classification accuracy. Furthermore, multilayer convolutional neural networks (CNNs) are proven very efficient for the feature extraction tasks as they outperform traditional ML algorithms such as SVM and RF to extract features from the high-dimensional network traffic data [34], [35], [36].

C. MOTIVATION

Though CNN can effectively deal with the dimensionality problem in attack classification, it still suffers from uncertainty issues. The uncertainty could be in optimising the weight parameters for CNN layers or the uncertainty involved in the randomness of attack generation. Thus, an advanced technique which can integrate the uncertainty quantification for optimising CNN layers, is highly essential to improve attack detection accuracy and reduce false positive and false negative rates.

In this regard, a CNN incorporated with probabilistic layers in the form of Bayesian neural networks (BNN) can be used to address the problem of parameter uncertainty for IDS applications. The weight parameters in BNNs are represented using probability distributions instead of point estimates, contrary to non-probabilistic neural networks [15]. These distributions define the uncertainty in weights and can further be used to estimate variability in predictions. Bayesian networks are trained using variational inference and instead of learning deterministic weight values directly, distribution parameters are learned [37], [38], [39].

Furthermore, with the help of convolutional layers and filters, automated feature mapping and extraction can be performed. So, there is a need of an integrated IDS method proving competitive performance for various classification evaluation metrics while quantifying the weight uncertainties using a low dimensional feature space.

D. CONTRIBUTIONS

The main contributions of this paper are described as follows:

- We propose a Bayesian probabilistic technique incorporated with deep convolutional neural networks (CNN-Bayesian) to detect and classify malicious events in cyber physical smart grid (CPSG) networks. The proposed method helps to extract features from the multidimensional feature space with the help of convolutional and pooling neural layers and quantify uncertainties in model parameters using Bayesian probabilistic approaches. The proposed method has an advantage over state-of-the-art CNN techniques in terms of a smaller false positive and false negative outcomes, thus improving the detection accuracy.

- The proposed Bayesian scheme utilizes Bernoulli distribution as the prior and posterior distributions to deal with the model uncertainties and provides future probabilities for the potential cyber attacks in CPSG networks.
- We evaluate the proposed Bayesian neural network on binary and multi-class power dataset taken from electric transmission systems [40] using classification evaluation metrics such as accuracy, precision, recall, and $F1$-score. Furthermore, efficacy of the proposed scheme is demonstrated against state-of-the-art deep learning techniques such as vanilla artificial neural networks (ANN), recurrent neural networks (RNN), LSTM, gated recurrent unit (GRU), and standalone CNN algorithms.

III. THE PROPOSED BAYESIAN PROBABILISTIC METHODOLOGY FOR CPSG-IDS

This section mainly involves four subsection dealing with main four units from the proposed methodology as shown in the Fig. 1. The four units namely, data preprocessing, feature extraction, probabilistic layers with Bernoulli distribution, and classification unit are described as following.

A. DATA PREPROCESSING AND FEATURE EXTRACTION

Data preprocessing is one of the initial and important steps in data analytics to improve the input data quality, so that it can further be fed to the neural networks to achieve effective and insightful results. These data preprocessing methods involve data cleaning (such as treating for missing or noisy values) and data transformation (such as scaling and encoding).

Since power system data can be highly dimensional, it is important to select what features can be useful and how we can extract these features to perform the classification. Through this process, it is aimed to combine the information from the original features and transform to a reduced space. In this paper, the convolution layer of CNN is used for feature extraction.

B. BERNOULLI DISTRIBUTION

For classification, BNNs can be expressed as probabilistic model using categorical distribution $p(y|x, w)$ for training data $D = (x_i, y_i)$. Here y is the target variable representing set of classes, x defines the input features and w are the weight parameters. Considering independent and identical probability distributions, the likelihood of classification data as a function of w parameters is defined as:

$$p(D|w) = \prod_i p(y_i|x_i, w) \tag{1}$$

Standard neural networks use categorical cross entropy as a cost function to maximize the likelihood and thus, to achieve high accuracy. In BNN, Bayesian probability is utilized by multiplying Eq. (1) with prior belief $p(w)$ as:

$$p(w|D) \propto p(D|w)p(w) \tag{2}$$

Here $p(w|D)$ represents the posterior distribution of w parameters over D. The marginal probability of the data can be written as:

$$p(D) = \int p(D|w)p(w)dw \tag{3}$$

Using Bayes theorem, eq. (2) can be written as:

$$p(w|D) = \frac{p(D|w)p(w)}{p(D)} \tag{4}$$

Substituting (3) into the denominator, (4) can be represented as:

$$p(w|D) = \frac{p(D|w)p(w)}{\int p(D|w)p(w)dw} \tag{5}$$

Computing posterior in Eq. (5) works as the core concept of BNN. However, it is analytically impossible due to the presence of large integral in the denominator. So, the true posterior is approximated with variational distribution $q(w|\theta)$ over θ parameters using the technique of variational inference (VI) [41]. And, the difference between true and approximated posterior is measured using a distance metric known as kullback-Leibler (KL) divergence as:

$$KL(q_\theta(w)\|p(w)) \tag{6}$$

$$\text{s.t.} \quad KL(q_\theta(w)\|p(w)) \geq 0 \tag{7}$$

where $q_\theta(w)$ denotes the approximated posterior over θ parameters using VI. VI is a mechanism to minimise the gap between true posterior and approximated posterior distributions. Thus, the cost function for the proposed Bayesian scheme is to minimize the KL divergence as:

$$\arg\min_{q_\theta(w)} KL(q_\theta(w)\|p(w)) \tag{8}$$

$$\theta = (W, b) \tag{9}$$

where W and b denote wights and bias in BNN parameter domain. KL divergence can be represented as:

$$KL(q_\theta(w)\|p(w)) = -\mathop{\mathbb{E}}_{q_\theta(w)} \log \frac{p(w)}{q_\theta(w)} \tag{10}$$

Note that the true distribution of weight parameters $p(w)$ are fixed given known input data and thus minimising KL divergence is equivalent to minimising the negative log-likelihood of $q_\theta(w)$.

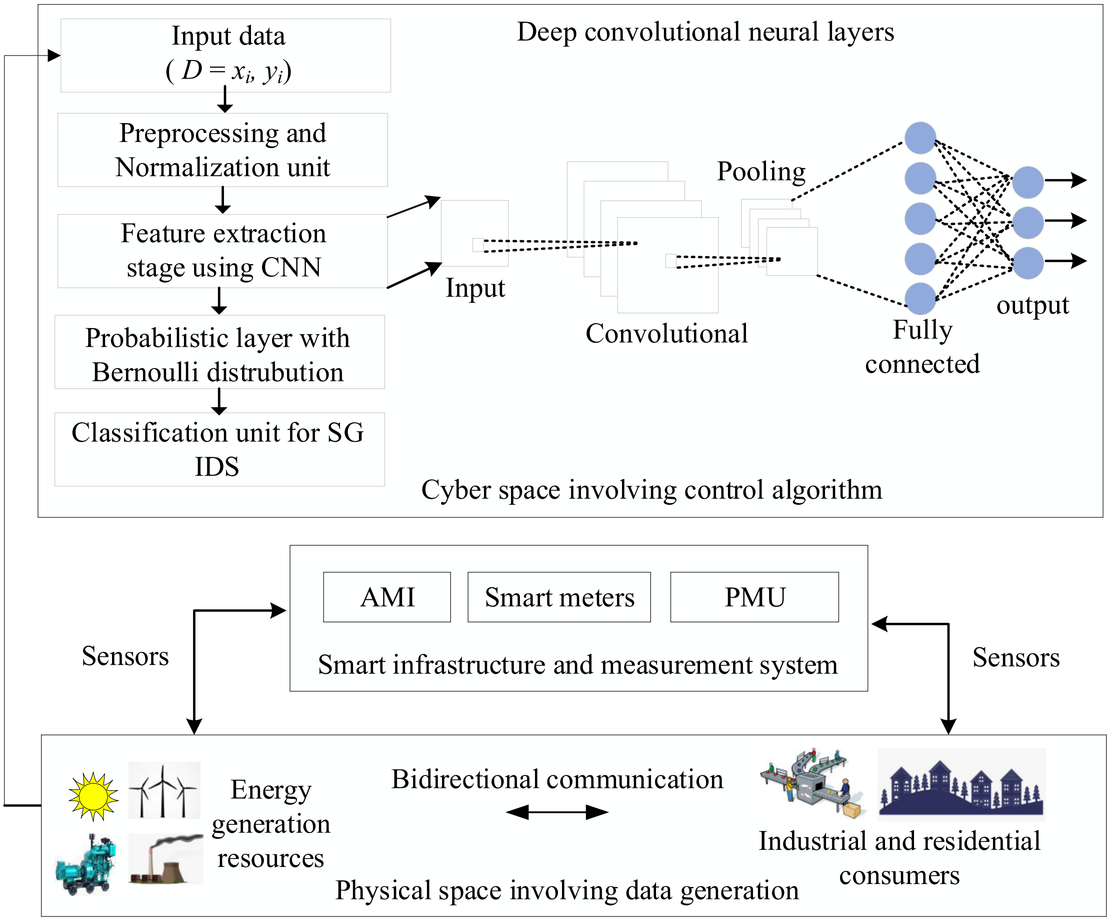
Input data
($D = x_i, y_i$)
Preprocessing and Normalization unit
Feature extraction stage using CNN
Probabilistic layer with Bernoulli distrubution
Classification unit for SG IDS
Deep convolutional neural layers
Input
Convolutional
Pooling
Fully connected
output
Cyber space involving control algorithm
AMI
Smart meters
PMU
Smart infrastructure and measurement system
Sensors
Sensors
Energy generation resources
Bidirectional communication
Industrial and residential consumers
Physical space involving data generation

FIGURE 1. System model of the proposed CNN-Bayesian IDS scheme.

C. ATTACKS CLASSIFICATION

Furthermore, the proposed Bayesian layer utilizes the Bernoulli distribution as prior and posterior distributions for binary classification through the Densevariational layer. Probability mass function (pmf) and cumulative distribution function (cdf) for Bernoulli distribution are given by the following mathematical equations:

$$p(w = k) = \begin{cases} 1 - p & k = 0 \\ p & k = 1 \end{cases} \tag{11}$$

$$p(w \leq k) = \begin{cases} 0 & k \leq 0 \\ 1 - p & 0 \leq k \leq 1 \\ 1 & k \geq 1 \end{cases} \tag{12}$$

Furthermore, the expected value ($E[w]$) and variance ($var[w]$) are obtained as:

$$E[w] = \sum kp(w = k) \tag{13}$$

$$var[w] = E[w^2] - E[w]^2 \tag{14}$$

Fig. 1 presents the system model of the proposed CNN-Bayesian intrusion-detection scheme. Furthermore, the Algorithm 1 presents pseudo-code of the proposed technique. The algorithm is explained using a number of execution steps identified by a number in Algorithm 1. In step 1, the input data is acquired. The feature set in the form of input data is fed to the proposed hybrid method in step 2. Normally, the features extracted from the network traffic involve different range of values and it is crucial to normalize or scale these values in alike range. After splitting the input data into training and testing sets in step 3, training data is fed to the first layer of the proposed CNN-Bayesian method in step 4. CNN convolves the data and a pool of features is extracted in multidimensional form in steps 5 and 6, respectively. The multidimensional output is then converted to 1-dimensional array using flatten operation in step 7. Finally, dense variational layer is used as the Bayesian layer to perform posterior optimization (lines 8-12) to classify the attacks in target variable. For each iteration, the prior distribution on the weight parameters is defined based on Bernoulli distribution as

Algorithm 1 The Proposed CNN-Bayesian Technique for IDS in CPSG Systems

```
Input: Actual observed data, $D = (x_i, y_i)$
Output: Accuracy, precision, recall, $F1$-score
1:  Acquire the input data $D$;
2:  Normalize $D_{scl} \in$ (-1,1);
3:  Split $D_{scl}$ to training ($D_{tr}$) and testing ($D_{test}$) set;
4:  Feed $D_{tr}$ to the input layer of CNN-Bayesian;
5:  Extract features using convolutional layer;
6:  Pool the measure map from the convolved layer;
7:  Flatten the multidimensional output into 1 dimensional array;
8:  Feed the flattened output to Bayesian (DenseVariational) layer;
9:  for (n=1, n $\leq$ m, n++) do
10:     Formulate prior trainable on p(w);
11:     Approximate posterior using (6);
12:     Minimize negative log likelihood loss using (8) and (10);
13:     Validate the training loss using validation split while training;
14: end for
15: Evaluate the proposed model on $D_{test}$;
16: Calculate accuracy, precision, recall, $F1$-score;
17: Obtain confusion matrix and area under curve (AUC);
```

in step 10. The posterior distribution is also assumed to be a Bernoulli distribution parameterized by θ as in step 11. Then the negative log likelihood (NLL) of the KL loss is minimized to obtain optimized posterior distribution using (8) in step 12. The hybrid approach is trained using *Adam* optimization algorithm and tested and validated on the split dataset in step 13. The proposed model is evaluated on the test data in step 15. Step 16 involves the calculation of accuracy, precision, recall and F1-score. To obtain visual representations, the confusion matrix and area under curve is computed in step 17.

IV. RESULTS AND DISCUSSIONS

A. DATASET GENERATION AND DESCRIPTION

This section presents implementation results conducted on the proposed and state-of-the-art DL techniques as a comparative case study using industrial control systems (ICS) cyber attack dataset [40]. Results simulations involve two different use cases with binary and multi-class target values.

In the binary dataset, the target variable involves two types of events, namely, 'Attack' and 'Natural'. On the other hand, from the multi-class data, say three-class dataset includes 'NoEvents', 'Natural', and 'Attacks' as target labels, as depicted in Table 2. While 'NoEvents' is reported in case of no incident, 'Natural' scenario is generated when an usual shortage or fault occurs. The 'Attack' scenario is reported when the system is controlled by an intruder, and it needs to be reported with greater accuracy.

TABLE 2. Dataset description.

Dataset	Events	Features	Train-test size	Validation size
Binary	Natural, attack	129	0.8:0.2	0.2
Multi-class	Natural, attack, NoEvents	131	0.8:0.2	0.2

The binary and multi-class classification problem categorises the outcomes into true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN). For binary dataset, TPs are the outcomes where an 'Attack' is predicted correctly. On the other hand, TNs include the outcomes where or a 'Natural' event is predicted correctly. FPs represent the outcomes where a 'Natural' event is predicted as 'Attack'. Similarly, FNs are the outcomes when an 'Attack' is predicted as a 'Natural' event. For multi-class dataset, TPs are defined as in the binary dataset. TNs define the outcomes when a 'NoEvents' or 'Natural' case is predicted correctly. FPs include the outcomes when a 'NoEvents' or 'Natural' case is predicted as 'Attack'. On the other hand, FNs are the outcomes when an 'Attack' is predicted as a 'NoEvents' or 'Natural' case.

The attack scenarios are generated using several components namely, power generators, electronic devices (IEDs), breakers, etc. The dataset is constructed using 29 sets of measurements from 4 PMUs along with 12 features for control logs and 1 feature for target variable, resulting into 129 features. These PMU measurements have sampling rate of 120 samples/s. The raw data is sampled randomly at one percent and grouped into binary and multi-class datasets.

The implementation results are obtained using python libraries and Tensorflow framework on $i10$ processor and 16 GB RAM.

B. DATA PREPROCESSING AND FEATURE EXTRACTION

Firstly, raw data is loaded and any duplicate and null values are removed using python libraries such as Numpy and Pandas. Then, one-hot encoding is performed to convert the categorical data into numerical form using label-encoder. Furthermore, data normalization is performed to scale the features in one range to achieve better results.

In this paper, convolutional neural layers are utilized to extract features from input ICS data using convolutional functionality from the CNNs. The element-wise multiplication is performed between the array of inputs and array of weights usually known as filters. The filters in the CNNs are applied from left to right and top to bottom resulting into a feature map. Eventually, with multiple training iterations, the network will learn what type of features to extract from the input data. The proposed method uses a set of 256 to 32 filters in parallel to learn the feature set effectively.

Furthermore, first layer of CNN captures simple features and last layer captures complex features from the data. Thus, with more neural layers, the deeper patterns and features can

TABLE 3. Binary classification results on evaluation metrics (percentage values).

Sr. no.	Classifier	Accuracy	Precision	Recall	$F1$-score	AUC
1.	Vanilla-ANN	51.06	49.89	30.10	37.54	0.65
2.	GRU	67.66	59.88	57.70	57.92	0.76
3.	RNN	77.20	72.73	71.43	71.99	0.86
4.	LSTM	88.38	70.58	72.77	71.55	0.89
5.	CNN	83.84	41.92	50.00	45.60	0.85
6.	**CNN-Bayesian**	92.76	88.97	84.74	86.83	0.98

TABLE 4. Multi-class classification results on evaluation metrics (percentage values).

Sr. no.	Classifier	Accuracy	Precision	Recall	$F1$-score	AUC
1.	Vanilla-ANN	69.62	48.77	33.33	39.59	0.58
2.	GRU	70.96	51.78	33.33	40.55	0.625
3.	RNN	71.02	62.87	35.15	45.09	0.69
4.	LSTM	72.56	63.23	35.29	45.29	0.73
5.	CNN	73.23	67.43	38.33	48.87	0.71
6.	**CNN-Bayesian**	84.76	71.97	84.74	77.84	0.84

be learned. The training, testing and validation splits for the binary and multi-class datasets, as well as the number of features are specified in Table 2.

The proposed scheme is validated on a number of classification metrics and parameters as discussed below.

C. PERFORMANCE EVALUATION METRICS

1) ACCURACY

To begin with, accuracy is the accustomed criteria to evaluate the performance of a classifier and it is measured as how many classes in the test dataset are correctly classified with respect to the total test predictions as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{15}$$

However, accuracy metric fails to analyze cases with imbalanced classification with a skewed class distribution.

2) PRECISION AND RECALL

For an imbalanced dataset, relying just on the accuracy to evaluate method performance is not an optimal approach. In order to detect the prediction anomalies, precision and recall are used [42]. Precision is defined as the ratio of positive test predictions with respect to all the predicted positives including FPs, as shown below:

$$Precision = \frac{TP}{TP + FP} \tag{16}$$

As the name suggests, precision traces how accurately the given model has predicted the TPs, out of all true and falsely predicted positive values. Precision is a crucial metric when the cost of FPs is high, eg., classifying a non-spam email as spam and losing of an important email. Furthermore, recall, which is also known as the sensitivity of a classifier, signifies how accurately the actual positives are truly classified taking into account the FN as shown in the mathematical equation below:

TABLE 5. Hyper-parameter settings.

Sr. no.	Name	Value
1.	Optimizer	Adam
2.	Epochs	100
3.	Loss	NLL
4.	Learning rate	0.001
5.	Batch size	128
6.	Hidden units	100
7.	Activation	tanh
8.	Validation split	0.2
9.	Dropout rate	0.5

$$Recall = \frac{TP}{TP + FN} \tag{17}$$

In cyber attacks classification, recall is of prime importance, as classifying an attack (TP) as a normal event (FN) can cause a serious fault in the grid functionality. Although, both the scores are desired to be higher, there is a trade-off involved between the two parameters. And, in our case, we are slightly inclined towards getting a higher recall score with accurate attacks classification.

3) $F1$-SCORE

Finally, $F1$-score also known as F-measure represents the weighted average between precision and recall. It measures the harmonic mean for the aforementioned metrics providing a single overall score, as:

$$F1 = 2 * \frac{precision * recall}{precision + recall} \tag{18}$$

$F1$-score is useful in case of dataset with highly unbalanced classes, for example, when 'normal' and 'non-attack' events are prevalent than the positive 'Attacks' class. The highest value for all the aforementioned metrics is 1.0 and the lowest

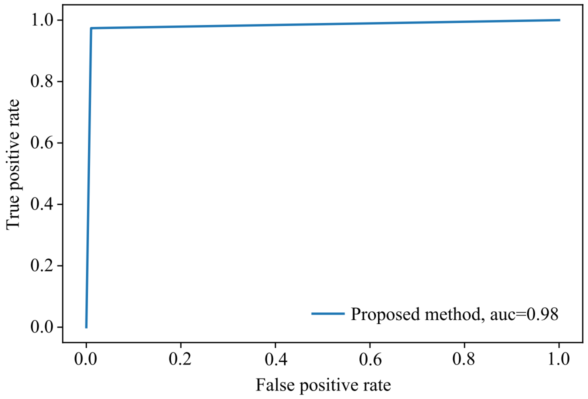

FIGURE 2. **Area under the curve given by the proposed method.**

is 0.0. Note that, the numerical values for the implementation results are considered in percentage form.

4) AREA UNDER THE CURVE - RECEIVER OPERATING CHARACTERISTICS (AUC-ROC)

ROC curve is used to evaluate the quality of a classifier by visualizing how well a classification model works. It features the true positive rate (TPR) on the Y axis and a false positive rate (FPR) on the X axis. Larger the area under the curve (AUC), better the accuracy i.e., a FPR of zero and a TPR of one is highly desirable. The Fig. 2 presents the AUC graph and score given by the proposed method. The proposed method achieves an AUC of 0.98, which illustrates its superiority in terms of the effectiveness of the classification model.

D. IMPLEMENTATION RESULTS

Tables 3 and 4 demonstrate the implementation results over aforementioned evaluation metrics for the proposed and standard DL methods namely, Vanilla-ANN, GRU, RNN, LSTM, and standard CNN to classify anomalous events in CPSG environment. For binary classification, 'Attacks' events are classified as positives and need to be identified with greater accuracy from the 'Normal' operations. From Table 3, it can be seen that LSTM as a recurrent variant provides competitive accuracy similar to proposed CNN-Bayesian method. However, precision and recall values for LSTM confine only to 70.58% and 72.77%. On the other hand, CNN-Bayesian achieves the highest scores, that is, Precision with 88.97% and Recall with 84.75%. Considering the repercussions of an attack scenario falsely classified as a normal event and the fact that there will be more normal scenarios that attack scenarios in a real-life system, it is highly important to train a classifier with grater precision and recall along with competitive accuracy. It is worth noting that RNN, LSTM and GRU are more suitable for handling time-series data but for multidimensional data such as the considered intrusion detection dataset, CNN will be more useful due to its ability to handle the features independently.

Similarly, Table 4 shows numerical results for classifying multiple scenarios from the ICS dataset using the proposed and standard classifiers. It is evident from the numerical values that CNN-Bayesian outperforms state-of-the-art methods for multi-class classification. Furthermore, Table 5 shows the external parameters to tune-in the proposed algorithm. Note that, NLL stands for negative log likelihood which is an important loss function to train probabilistic neural layers. In addition, we make use of dropout [43] to avoid the classifier from overfitting the training data. The hyperparameters are tuned using grid search optimisation so that the performance comparison across the different models is fair. The training and testing process has been repeated 10 times and the evaluation metrics have been averaged over these 10 iterations.

The comparative analysis is further extended and tested for 5 different scenarios of the dataset (S1-S5) for all the above mentioned standard and proposed classifiers. S1-S5 involve binary and multi-class instances sampled randomly from the pool of ICS dataset. Each dataset scenario consists of 5k instances related to electrical transmission system behaviors with 128 independent features.

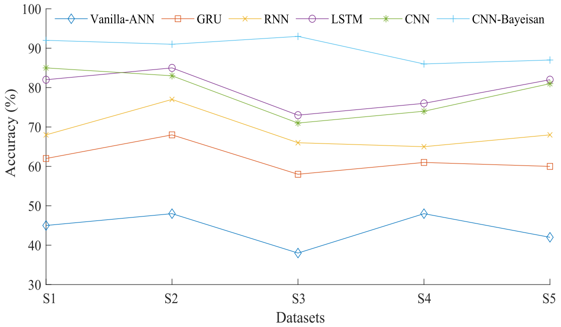

FIGURE 3. **Accuracy scores from binary classification over different dataset scenarios (S1-S5).**

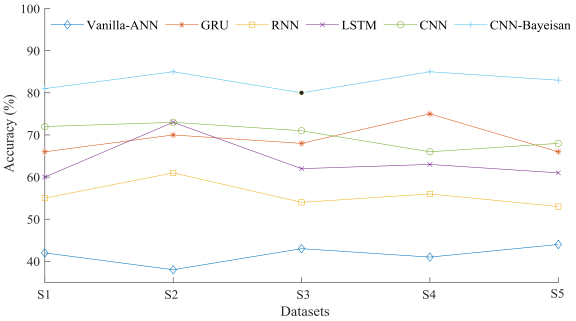

FIGURE 4. **Accuracy scores from multi-class classification over different dataset scenarios (S1-S5).**

The Fig. 3 and Fig. 4 demonstrate overall accuracy values (in percentage) for binary and multi-class scenarios, respectively. It is evident from the visualizations that the recurrent neural layers, such as LSTM and RNN perform better than the vanilla-ANN and GRU. However, these layers can not perform feature engineering and thus, CNN is required. In addition, Fig. 5 illustrates the precision, recall, and weighted

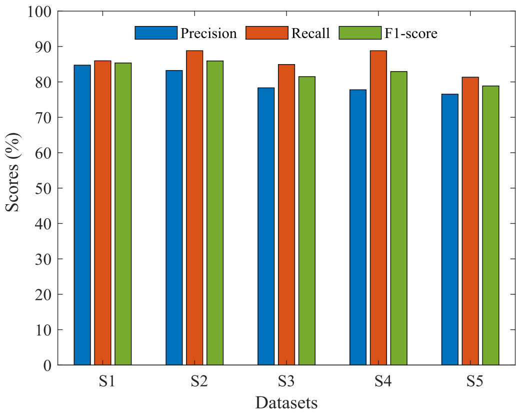

FIGURE 5. Precision, recall, and *F*1-score using proposed CNN-Bayesian method for 5 different dataset scenarios.

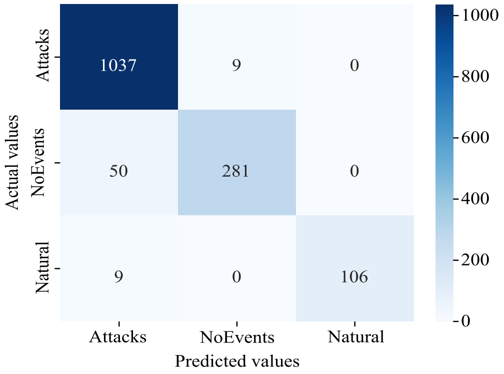

FIGURE 6. Confusion matrix for multi-class classification.

*F*1-score values for S1-S5 dataset scenarios using proposed CNN-Bayesian classifier. It is clear from the bar graph that our proposed method performs most efficiently by yielding the highest precision and recall values over all the dataset scenarios. Fig. 6 presents the confusion matrix for multi-class classification. It can be observed that 99.13% of the attack scenarios have been correctly detected. On the other hand, 0.87% of these scenarios are detected as normal operating conditions. Among the non-attack scenarios, 84.89% cases have been correctly identified, while 15.11% cases have been incorrectly detected as attacks. 92.17% of the scenarios when faults have occurred, have been correctly detected. However, 7.83% of such scenarios have been detected as attacks. Thus, it can be concluded that when there is an attack, the proposed model demonstrates a much higher accuracy. This is essential given the critical nature of the situation if attacks remain undetected.

V. CONCLUSION AND FUTURE WORK

This paper presents a probabilistic deep learning approach integrated with convolutional layers as a feature engineering mechanism for the applications of cyber attack detection in the smart grid systems. We implemented the proposed CNN-Bayesian method on a real-life power systems dataset with multivariate features (128, to be specific) to classify different types of malicious events. Furthermore, a comparative case study is conducted against the standard deep learning methods such as vanilla and recurrent neural networks and its variants. It is inferred from the numerical results that CNN-Bayesian outperform all the other comparative DL classifiers, not just in the terms of accuracy, but precision and recall as well. Though the proposed method can outperform for datasets with more uncertainty, it may not be the best performing model if there is less uncertainty in the dataset. Future work will focus on the application of distributed machine learning framework integrated with Bayesian layers for multi-class event detection. We will also consider the optimisation of the dataset features for improved performance of the proposed algorithm and compare against the more recent machine learning based classification methods.

REFERENCES

[1] A. Usman and S. H. Shami, "Evolution of communication technologies for smart grid applications," *Renew. Sustain. Energy Rev.*, vol. 19, pp. 191–199, Mar. 2013.

[2] C. Peng, H. Sun, M. Yang, and Y.-L. Wang, "A survey on security communication and control for smart grids under malicious cyber attacks," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 49, no. 8, pp. 1554–1569, Aug. 2019.

[3] G. N. Ericsson, "Cyber security and power system communication—Essential parts of a smart grid infrastructure," *IEEE Trans. Power Del.*, vol. 25, no. 3, pp. 1501–1507, Jul. 2010.

[4] R. Vijayanand, D. Devaraj, and B. Kannapiran, "A novel deep learning based intrusion detection system for smart meter communication network," in *Proc. IEEE Int. Conf. Intell. Techn. Control, Optim. Signal Process. (INCOS)*, Apr. 2019, pp. 1–3.

[5] Y. Zhang and J. Yan, "Domain-adversarial transfer learning for robust intrusion detection in the smart grid," in *Proc. IEEE Int. Conf. Commun., Control, Comput. Technol. Smart Grids (SmartGridComm)*, Oct. 2019, pp. 1–6.

[6] P. Yi, Y. Wu, F. Zou, and N. Liu, "A survey on security in wireless mesh networks," *IETE Tech. Rev.*, vol. 27, no. 1, pp. 6–14, 2010.

[7] J. M. Torres, C. I. Comesaña, and P. J. García-Nieto, "Machine learning techniques applied to cybersecurity," *Int. J. Mach. Learn. Cybern.*, vol. 10, no. 10, pp. 2823–2836, 2019.

[8] Y. Zhang, L. Wang, W. Sun, R. C. Green, and M. Alam, "Distributed intrusion detection system in a multi-layer network architecture of smart grids," *IEEE Trans. Smart Grid*, vol. 2, no. 4, pp. 796–808, Dec. 2011.

[9] Y. Wang, Q. Chen, T. Hong, and C. Kang, "Review of smart meter data analytics: Applications, methodologies, and challenges," *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 3125–3148, May 2019.

[10] P. Kumar, Y. Lin, G. Bai, A. Paverd, J. S. Dong, and A. Martin, "Smart grid metering networks: A survey on security, privacy and open research issues," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 3, pp. 2886–2927, 3rd Quart., 2019.

[11] A. S. Eesa, Z. Orman, and A. M. A. Brifcani, "A novel feature-selection approach based on the cuttlefish optimization algorithm for intrusion detection systems," *Exp. Syst. Appl.*, vol. 42, no. 5, pp. 2670–2679, 2015.

[12] A. Javaid, Q. Niyaz, W. Sun, and M. Alam, "A deep learning approach for network intrusion detection system," in *Proc. 9th EAI Int. Conf. Bio-Inspired Inf. Commun. Technol.*, vol. 3, no. 9, 2016, p. e2.

[13] P. Sun, P. Liu, Q. Li, C. Liu, X. Lu, R. Hao, and J. Chen, "DL-IDS: Extracting features using CNN-LSTM hybrid network for intrusion detection system," *Secur. Commun. Netw.*, vol. 2020, pp. 1–11, Aug. 2020.

[14] F. A. Khan and A. Gumaei, "A comparative study of machine learning classifiers for network intrusion detection," in *Proc. Int. Conf. Artif. Intell. Secur.* Cham, Switzerland: Springer, 2019, pp. 75–86.

[15] M. Sun, T. Zhang, Y. Wang, G. Strbac, and C. Kang, "Using Bayesian deep learning to capture uncertainty for residential net load forecasting," *IEEE Trans. Power Syst.*, vol. 35, no. 1, pp. 188–201, Jan. 2019.

[16] A. Kendall and Y. Gal, "What uncertainties do we need in Bayesian deep learning for computer vision?" in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5574–5584.

[17] S. Mane and D. Rao, "Explaining network intrusion detection system using explainable AI framework," 2021, *arXiv:2103.07110*.

[18] W. Hu, Y. Liao, and V. R. Vemuri, "Robust support vector machines for anomaly detection in computer security," in *Proc. ICMLA*, 2003, pp. 168–174.

[19] C. Sinclair, L. Pierce, and S. Matzner, "An application of machine learning to network intrusion detection," in *Proc. 15th Annu. Comput. Secur. Appl. Conf. (ACSAC99)*, 1999, pp. 371–377.

[20] F. A. A. Alseiari and Z. Aung, "Real-time anomaly-based distributed intrusion detection systems for advanced metering infrastructure utilizing stream data mining," in *Proc. Int. Conf. Smart Grid Clean Energy Technol. (ICSGCE)*, Oct. 2015, pp. 148–153.

[21] K.-A. Tait, J. S. Khan, F. Alqahtani, A. A. Shah, F. A. Khan, M. U. Rehman, W. Boulila, and J. Ahmad, "Intrusion detection using machine learning techniques: An experimental comparison," 2021, *arXiv:2105.13435*.

[22] A. Churcher, R. Ullah, J. Ahmad, S. U. Rehman, F. Masood, M. Gogate, F. Alqahtani, B. Nour, and W. J. Buchanan, "An experimental analysis of attack classification using machine learning in IoT networks," *Sensors*, vol. 21, no. 2, p. 446, Jan. 2021.

[23] M. Alqahtani, A. Gumaei, H. Mathkour, and M. M. Ben Ismail, "A genetic-based extreme gradient boosting model for detecting intrusions in wireless sensor networks," *Sensors*, vol. 19, no. 20, p. 4383, Oct. 2019.

[24] R. C. Borges Hink, J. M. Beaver, M. A. Buckner, T. Morris, U. Adhikari, and S. Pan, "Machine learning for power system disturbance and cyber-attack discrimination," in *Proc. 7th Int. Symp. Resilient Control Syst. (ISRCS)*, Aug. 2014, pp. 1–8.

[25] M. M. Hassan, A. Gumaei, S. Huda, and A. Almogren, "Increasing the trustworthiness in the industrial IoT networks through a reliable cyberattack detection model," *IEEE Trans. Ind. Informat.*, vol. 16, no. 9, pp. 6154–6162, Sep. 2020.

[26] S. Selvarajan, M. Shaik, S. Ameerjohn, and S. Kannan, "Mining of intrusion attack in SCADA network using clustering and genetically seeded flora-based optimal classification algorithm," *IET Inf. Secur.*, vol. 14, no. 1, pp. 1–11, 2020. [Online]. Available: https://digital-library.theiet.org/content/journals/10.1049/iet-ifs.2019.0011

[27] S. Shitharth, K. M. Prasad, K. Sangeetha, P. R. Kshirsagar, T. S. Babu, and H. H. Alhelou, "An enriched RPCO-BCNN mechanisms for attack detection and classification in SCADA systems," *IEEE Access*, vol. 9, pp. 156297–156312, 2021.

[28] S. Shitharth, N. Satheesh, B. P. Kumar, and K. Sangeetha, *IDS Detection Based on Optimization Based on WI-CS and GNN Algorithm in SCADA Network*. Singapore: Springer, 2021, pp. 247–265.

[29] T. Ma, F. Wang, J. Cheng, Y. Yu, and X. Chen, "A hybrid spectral clustering and deep neural network ensemble algorithm for intrusion detection in sensor networks," *Sensors*, vol. 16, no. 10, p. 1701, 2016.

[30] A. Halbouni, T. S. Gunawan, M. H. Habaebi, M. Halbouni, M. Kartiwi, and R. Ahmad, "CNN-LSTM: Hybrid deep neural network for network intrusion detection system," *IEEE Access*, vol. 10, pp. 99837–99849, 2022.

[31] A. S. Zamzam and N. D. Sidiropoulos, "Physics-aware neural networks for distribution system state estimation," *IEEE Trans. Power Syst.*, vol. 35, no. 6, pp. 4347–4356, Nov. 2020.

[32] T. Wu, I. L. Carreno, A. Scaglione, and D. Arnold, "Graph convolutional neural networks for physics-aware grid learning algorithms," 2022, *arXiv:2203.16732*.

[33] H. Narasimhan, V. Ravi, and N. Mohammad, "Unsupervised deep learning approach for in-vehicle intrusion detection system," *IEEE Consum. Electron. Mag.*, vol. 12, no. 1, pp. 103–108, Jan. 2021.

[34] J. Kim, H. Kim, M. Shim, and E. Choi, "CNN-based network intrusion detection against denial-of-service attacks," *Electronics*, vol. 9, no. 6, p. 916, Jun. 2020.

[35] R. Vinayakumar, K. P. Soman, and P. Poornachandran, "Applying convolutional neural network for network intrusion detection," in *Proc. Int. Conf. Adv. Comput., Commun. Informat. (ICACCI)*, Sep. 2017, pp. 1222–1228.

[36] M. Elhoseny, M. M. Selim, and K. Shankar, "Optimal deep learning based convolution neural network for digital forensics face sketch synthesis in Internet of Things (IoT)," *Int. J. Mach. Learn. Cybern.*, vol. 12, pp. 3249–3260, Jul. 2020.

[37] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," 2013, *arXiv:1312.6114*.

[38] D. Kaur, S. N. Islam, M. A. Mahmud, M. E. Haque, and A. Anwar, "A VAE-Bayesian deep learning scheme for solar generation forecasting based on dimensionality reduction," 2021, *arXiv:2103.12969*.

[39] D. Kaur, S. N. Islam, and M. A. Mahmud, "A variational autoencoder-based dimensionality reduction technique for generation forecasting in cyber-physical smart grids," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, Jun. 2021, pp. 1–6.

[40] *Industrial Control System (ICS) Cyber Attack Datasets*. Accessed: Oct. 6, 2021. [Online]. Available: https://sites.google.com/a/uah.edu/tommy-morris-uah/ics-data-sets

[41] C. Zhang, J. Bütepage, H. Kjellström, and S. Mandt, "Advances in variational inference," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 8, pp. 2008–2026, Aug. 2018.

[42] J. Brownlee. (2020). *How to Calculate Precision, Recall, and F-Measure for Imbalanced Classification*. Machine Learning Mastery. [Online]. Available: https://machinelearningmastery.com/precision-recall-and-f-measure-for-imbalancedclassification

[43] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 1050–1059.

DEVINDER KAUR received the master's degree from the Department of Computer Science and Engineering, Thapar University, Patiala, Punjab, India. She is currently pursuing the Ph.D. degree with the Faculty of Science, Engineering and Built Environment, Deakin University, Geelong, VIC, Australia. She has been awarded a full Higher Degree Research Scholarship to pursue her doctorate. Previously, she worked as an Assistant Professor with the Computer Science and Engineering Department, Lovely Professional University, Punjab. Her research interests include data-driven intelligent algorithms, energy forecasting applications, probabilistic forecasting, and variational inference.

ADNAN ANWAR (Member, IEEE) received the master's (by research) and Ph.D. degrees from The University of New South Wales (UNSW), Australia. He is currently a Lecturer and the Deputy Course Director of Postgraduate Cyber Security Education with the School of Information Technology, Deakin University, Australia. Previously, he worked as a Data Scientist and an Analytics Team Leader at Flow Power. He has over ten years of industrial, research, and teaching experience in universities and research laboratories, including NICTA (now, Data61 of CSIRO), La Trobe University, Deakin University, and UNSW. He has authored over 70 articles, including journals (mostly in Q1), conference papers, and book chapters in prestigious venues. His research interests include the security research of sensor-connected IoT, cloud and SCADA systems of critical infrastructures, data-driven intelligent techniques, and data science applications in energy systems. He was a recipient of several awards, including UPA Scholarship, UNSW TFR Scholarship, Best Paper Award, industry funding, and several travel grants, including ACM and Postgraduate Research Student Support (PRSS) travel grants. He is active in the IEEE Computer Society Technical Committee on Data Engineering and the IEEE Cybersecurity Committee.

INNOCENT KAMWA (Fellow, IEEE) received the Ph.D. degree in electrical engineering from Laval University in 1989. He was a Full Professor with the Department of Electrical Engineering and Tier 1 Canada Research Chair in decentralized sustainable electricity grids for smart communities, Laval University. He was previously a Researcher at Hydro-Québec's Research Institute, specializing in the dynamic performance and control of power systems. He was also the Chief Scientist for Hydro-Québec's Smart Grid Innovation Program and an International Consultant in power grid simulation and network stability. He is a Fellow of the Canadian Academy of Engineering and Fellow of the IEEE for his innovations in power system control, he is also the 2019 recipient of the IEEE Charles Proteus Steinmetz and Charles Concordia Awards. He was a past Editor-in-Chief of IET Generation, Transmission and Distribution, and he is currently the Editor-in-Chief of IEEE *Power and Energy Magazine* and an Associate Editor of IEEE Transactions on Power Systems.

SHAMA ISLAM (Member, IEEE) received the Ph.D. degree from The Australian National University, in 2015. She is currently a Senior Lecturer in electrical engineering with Deakin University. She is also a leading Researcher in the area of smart grid communication, the IoT for smart energy applications, energy management, and smart grid data analytics. She has successfully attracted internal grants worth U.S. $250,000 and external grants of 1.4 million AUD over the last five years along with other investigators with Deakin University. She has been accredited as a fellow of the U.K.-based Higher Education Academy, in 2021. She has been awarded the Victoria Fellowship 2019 for her contributions to scientific innovations in VIC, Australia.

S. M. MUYEEN (Senior Member, IEEE) received the B.Sc.Eng. degree in electrical and electronic engineering from the Rajshahi University of Engineering and Technology (RUET), Bangladesh, formerly known as the Rajshahi Institute of Technology, in 2000, and the M.Eng. and Ph.D. degrees in electrical and electronic engineering from the Kitami Institute of Technology, Japan, in 2005 and 2008, respectively. He is currently working as a Full Professor with the Electrical Engineering Department, Qatar University. He has been a keynote speaker and an invited speaker at many international conferences, workshops, and universities. He has published more than 250 papers in different journals and international conferences. He has published seven books as the author or editor. His research interests include power system stability and control, electrical machine, FACTS, energy storage systems (ESSs), renewable energy, and HVDC systems. He is a fellow of Engineers Australia. He is serving as an Editor/an Associate Editor for many prestigious journals from IEEE, IET, and other publishers, including the IEEE TRANSACTIONS ON ENERGY CONVERSION, the IEEE POWER ENGINEERING LETTERS, *IET Renewable Power Generation*, and *IET Generation, Transmission & Distribution*.

NASSER HOSSEINZADEH (Senior Member, IEEE) received the B.Sc. degree in electrical engineering from Shiraz University, Shiraz, Iran, in 1986, the M.Sc. degree in electronics engineering from the University of Science and Technology, Tehran, Iran, in 1992, and the Ph.D. degree in electrical power systems from Victoria University, Australia, in 1998. He was the Discipline Leader of electrical engineering, from 2005 to 2006, the Head of the Department of Systems, CQUniversity, Australia, from 2007 to 2008, and the Head of Electrical and Computer Engineering, SQU, Oman, from 2014 to 2018. He is currently with Deakin University, Australia, where he is also the Director of the Centre for Smart Power and Energy Research (CSPER). His research interests include stability assessment of the power grid as impacted by inverter-based generators, microgrids, power system dynamics and control, online monitoring, and real-time control of microgrids. He has been a Regular Reviewer of the IEEE TRANSACTIONS ON SMART GRID, the IEEE TRANSACTIONS ON POWER SYSTEMS, the IEEE TRANSACTIONS ON POWER DELIVERY, the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, the IEEE TRANSACTIONS ON EDUCATION, and the *International Journal of Electrical Power and Energy Systems* (Elsevier).