

Research Article

Very Short-Term Blackout Prediction for Grid-Tied PV Systems Operating in Low Reliability Weak Electric Grids of Developing Countries

Benson H. Mbuya ^{1,2}, Aleksandar Dimovski,³ Marco Merlo,³ and Thomas Kivevele ¹

¹*School of Materials, Energy, Water and Environmental Sciences (MEWES),*

The Nelson Mandela African Institution of Science and Technology, P.O. Box 447, Arusha, Tanzania

²*Electrical Engineering Department, Arusha Technical College, P.O. Box 296, Arusha, Tanzania*

³*Energy Department, Politecnico di Milano, Via la Masa, Milano 20156, Italy*

Correspondence should be addressed to Benson H. Mbuya; mbuyab@nm-aist.ac.tz

Received 22 March 2022; Revised 30 July 2022; Accepted 20 August 2022; Published 8 September 2022

Academic Editor: Lingzhong Guo

Copyright © 2022 Benson H. Mbuya et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Sub-Saharan emerging countries experience electrical shortages resulting in power rationing, which ends up hampering economic activities. This paper proposes an approach for very short-term blackout forecast in grid-tied PV systems operating in low reliability weak electric grids of emerging countries. A pilot project was implemented in Arusha-Tanzania; it mainly comprised of a PV-inverter and a lead-acid battery bank connected to the local electricity utility company, Tanzania Electric Supply Company Limited (TANESCO). A very short-term power outage prediction model framework based on a hybrid random forest (RF) algorithm was developed using open-source Python machine learning libraries and using a dataset generated from the pilot project's experimental microgrid. Input data sampled at a 15-minute interval included day of the month, weekday, hour, supply voltage, utility line frequency, and previous days' blackout profiles. The model was composed of an adaptive similar day (ASD) module that predicts 15 minutes ahead from a sliding window lookup table spanning 2 weeks prior to the prediction target day, after which ASD prediction was fused with RF prediction, giving a final optimised RF-ASD blackout prediction model. Furthermore, the efficacy analysis of the short-term blackout prediction of the formulated RF, ASD, and RF-ASD regression and classification algorithms was compared. Considering the stochastic nature of blackouts, their performance was found to be fair in short-term blackout predictions of the test site's weak grid using limited input data from the point of coupling of the user. The models developed were only able to predict blackouts if they occurred frequently and contiguously, but they performed poorly if they were sparse or dispersed.

1. Introduction

Access to electricity remains an issue yet to be resolved in sub-Saharan Africa (SSA), where 600 million people do not have access to electricity, this being nearly half of the population [1]. Reliable and quality access to electricity boosts livelihood and drives forward the economy [2]. According to a 2019 survey conducted by the Tanzanian Rural Energy Agency (REA), it revealed that only 37.7% of people on the mainland had connections to electricity [3]. This is in part due to expensive connection costs along with

low use of electricity and poverty among countryside communities [4]. Taking Tanzania as an example, even areas with access to electricity at times experience power outages, brownouts, and voltage surges. Many times, electricity demand exceeds supply in the SSA region; Load shedding or rolling blackouts (power rationing) becomes essential to prevent the electric grid from failing. However, end users prefer power cut notifications in advance so that they can plan ahead to mitigate power outage effects [5]. These notifications usually reach a very small percentage of users, especially if the outage will cover a relatively small area.

Rolling blackouts hit the most, residential sector, or poorer neighborhoods than the industrial sector due to both economic and political factors [6, 7].

Studies investigating blackouts are not new. Blackouts are caused by multifaceted interactions of many factors, such as serious line component failure; negligence in handling vegetation along transmission and distribution lines; wildlife; extreme weather; man-made error; and other natural calamities [8]. The problem of blackout prediction can be attempted either by short-term (real-time) prediction or long-term prediction. Nateghi et al. [9, 10] developed a hurricane-induced blackout prediction algorithm based on a random forest algorithm. Hurricane-induced power outage prediction models are complex and involve large expensive datasets. The dataset may comprise the following input variables: electric grid data, data on pre-storm situation of soil moisture, drought, land use data, geographical measures, wind data, and so forth. Gou and Wu [8] classified blackout causes as either deterministic or probabilistic. The study also investigated control strategies with an emphasis on islanding control strategies. In the study by Alkar et al. [11], frequent power outages were found to be due to the inefficiency of the power plant to match loads, malfunction of protection equipment in the transmission lines, poor ability of Supervisory Control and Data Acquisition (SCADA), division of the main power grid to microgrids, which are more prone to oscillations, delayed equipment maintenance, and population increase. The study by Rahman et al. [12] used data from large-scale power outages globally to explore blackout causes, perform a risk analysis, and fault analysis. In Bo et al. [13], worldwide blackout incidents were analyzed, and their causes, mitigation, and restoration measures were investigated. Some blackout causes, protection issues, blackout prevention, and blackout restoration have been investigated by some studies [14, 15]. In the work of Mei et al. [16], two indices for quantitative blackout risk evaluation were developed.

Cheng et al. [17], combined other outage factors such as real-time electric grid system operational data, weather forecast, and geographical data to predict blackout components in the current electric grid system operation condition. They also developed a load behavior forecast model under power outage circumstances using an expert fuzzy system. Kogo et al. [18] proposed 3 heuristics to forecast the start time of the next 24 hrs irregular scheduled power cut, namely, start time of power cut based prediction (SBP), frequency-based prediction (FBP), and a hybrid of SBP and FBP. In Papic and Ciniglio [19], a framework for supporting planners and operators in evaluating multiple outages that lead to cascading outages was developed. Papic et al. [20] identified two indices for power outage reliability, namely, average frequency and average duration of sustained automatic outages. The indices could be employed in forecast based planning, maintenance, and operation activities. According to the study, the foremost causes of outages were found to be weather (rain, snow, ice storms, wind, dust, and so on), equipment failure, and wildfires. The 26th UN climate change conference of the parties (COP26) attests to the fact that climate change is a burning issue globally. Extreme

weather events are on the rise. Unfortunately, they also affect electricity generation and supply infrastructure reliability, sometimes resulting in curtailments and in some cases rolling blackouts [21–23]. Thus, blackouts should continue to be studied.

The main contribution of this work is the study of blackout forecasts from the customer’s perspective—at the end user’s side (premise) of a weak grid. This may be useful in regions where the utility operates in a black box manner, with limited or no grid information available to the end user. Although not addressed in this study, the outcome of the blackout forecast could be used in the implementation of a battery management system strategy that ensures adequate state of charge (SoC) when a blackout is imminent and allows the battery to fully discharge when an occurrence of a blackout is unlikely. Consequently, money could be saved if batteries are not oversized, and the SoC is optimised with respect to blackout forecast output. Blackout forecasts in this study could also be useful to the utility for load shedding and demand-side management applications. In the literature, studies on blackouts have not been fully exhausted. There have been few or no works on blackout forecast studies for emerging countries’ scenarios. Emerging countries power grids give an interesting research study focus because they are weak, still evolving, and typically characterised by frequent disturbances as opposed to the mature grids of developed countries. Moreover, due to scattered settlements in emerging countries like Tanzania, it is not yet feasible for the national main power grid to reach all communities; therefore, most emerging countries embark on rural electrification programs that employ hybrid microgrids that in time get connected to the main power grid. Consequently, microgrids that connect to the main grid have to contend with disturbances in the main grid. This work attempts to address and bridge this gap with a blackout forecast model using a case study from Arusha, Tanzania.

2. Blackout Pilot System Overview

The pilot test site for this work was an office building located in the Levulosi ward of the Arusha municipality in Tanzania. The energy management system at the site is as shown in Figure 1, comprising 200 W PV, 100 Ah lead-acid battery, and a 2.5 kW inverter. Such pico-size systems are typical in SSA, where PV systems are undersized due to financial constraints [24, 25]. The experimental test site already had the PV-battery-inverter system in place before this study was conducted. Therefore, the PV-battery-inverter sizing and setup was not part of our study, rather the goal was to investigate blackout forecast from the customer’s vantage point. The PV-inverter and battery bank at the site are essentially used as backup power for merely small loads and application such as lighting, laptop, phone charging, and a surveillance camera; the backup system additionally was used to power the data logger used in this study during blackouts. Other relatively bigger loads at the site, like photocopy machines, electric fans, microwaves, and electric kettles, are only used when utility power is available.

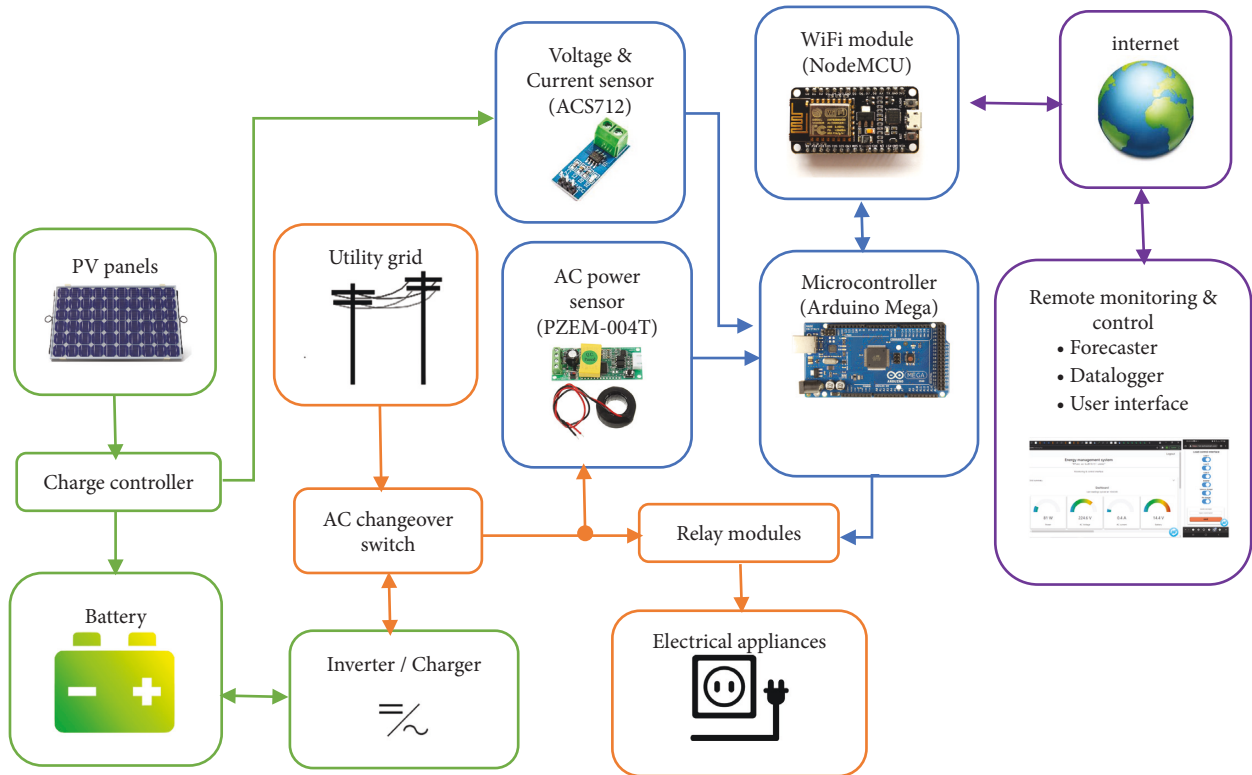


FIGURE 1: Test site energy management system architecture.

The battery bank at the site can be charged either by PV or by utility power through an inverter charger. The main electricity supply to the pilot site comes from the national electricity supply company, TANESCO. Power from both the utility and the inverter is managed by the local smart controller (smart meter). The low-cost smart meter is comprised of a PZEM-004T module used to measure AC voltage, current, and frequency. PV voltage and battery voltage were measured with a voltage divider circuit interfaced to the Arduino mega microcontroller ADC (Analog-to-digital converter) pins whereas PV current was measured by the ACS712 hall effect current sensor module, also interfaced to the ADC port of the Arduino mega board. PV-battery DC voltage and current sensor measurements facilitated PWM charging of the battery using P-channel MOSFET high-side switching controlled by the Arduino mega board. All the measured sensor data were uploaded to the cloud via the WiFi module ESP8266 (NodeMCU). Received data from the pilot test site was stored in a MySQL database on a Linux-based server and also used to perform blackout forecasts. A web page on the server was used as an energy dashboard for monitoring and remote-control operations.

3. Methodology

An outline of the proposed power outage prediction approach is provided, followed by the main individual components that comprise the blackout forecast algorithm, namely, input data preprocessing, forecast models, and

evaluation indexes. This work proposes a procedure for very short-term blackout prediction in emerging country scenarios. Moreover, blackout prediction is realized by a fusion of random forest (RF) and adaptive similar day (ASD) models.

3.1. Proposed Blackout Forecasting Model Approach. Figure 2 shows the power outage prediction approach adopted in this paper. The algorithm was implemented using scikit-learn and other Python libraries. The framework starts at the data preprocessing stage, followed by a moving sliding window that traverses forward through the data during predict and update operations; then, the RF and ASD sections carry out the actual blackout prediction operation; the final stage is performance evaluation and corrective feedback. The cycle then repeats with error performance knowledge at hand to be used to adjust the weights in the subsequent iteration of the short-term blackout prediction. These stages mentioned here briefly are expounded in the preceding sections.

3.2. Input Data Description. The dataset used in this paper covers the period from January 2021 to December 2021, amounting to a total of 8,016 hourly aggregated samples or 30, 144 15-minute interval samples. The subset of the dataset used comprised of 10 relevant input variables from the previous 14 days as predictors. These 10 predictors were selected after performing a correlation test with the blackout indicator variable. The 10 input features used in this work

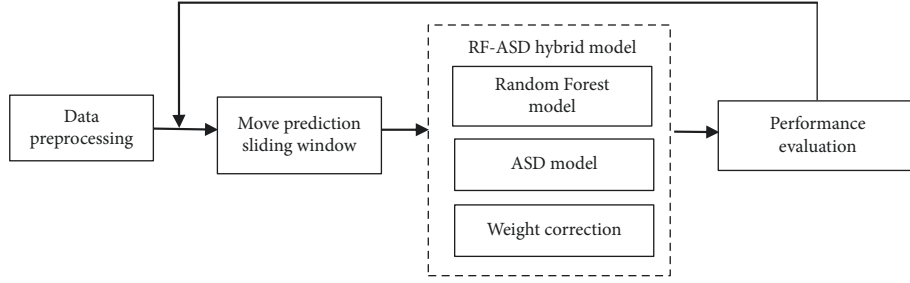


FIGURE 2: Proposed power outage prediction approach.

TABLE 1: Input features details.

#	Input variable	Input variable detail	Evaluation time
1	Day	Day of the month	Day D-1
2	WeekDay	Day of the week	Day D-1
3	Hour	Hour of the day	Day D-1
4	AC_voltage	Average AC supply line voltage from electric supply company	Day D-1
5	Frequency	Average AC supply line frequency	Day D-1
6	B14	Average blackout profile 14 days before	Day D-14
7	B7	Average blackout profile 7 days before	Day D-7
8	B3	Average blackout profile 3 days before	Day D-3
9	B2	Average blackout profile 2 days before	Day D-2
10	B1	Average previous day blackout profile	Day D-1

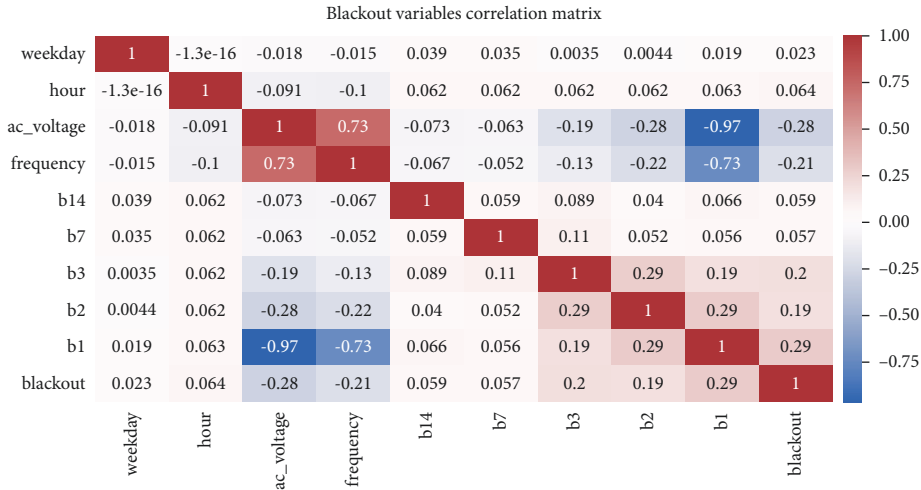


FIGURE 3: Input variables' correlation to blackout variables. A value close to "1" signifies high correlation, whereas a value close to zero shows low correlation to the blackout variable. Negative values show a negative correlation.

are shown in Table 1. These parameters used in the weak grid include time variables such as day of the month, day of the week, and hour of the day; they were significant because blackouts were observed to be cyclical in nature with respect to time. The AC line voltage and frequency parameters were also used to determine the presence or absence of power (blackout). Weather data for the area could not be obtained; thus, it was not used. Figure 3 shows the correlation of the input variables to the blackout variable. The "ac_voltage" and "frequency" variables exhibited negative correlation values because they were inversely related to blackout, in which the presence of blackout (electric power = 0) implies zero AC line volts and zero frequency on the test site power line.

Three different strategies were performed for short-term blackout profile prediction targets or responses, namely, (1) blackout prediction for the next 15 minutes, which is a single scalar value; (2) blackout prediction for the next hour, which is also a single scalar value; and (3) lastly, blackout prediction for the entire next day (24 hours), which is a vector of 24 blackout index values corresponding to the hours of the following day. The blackout prediction of the next 15 minutes is performed 15 minutes prior, whereas prediction of the next hour is performed an hour in advance. Blackout prediction of the entire next day is performed at midnight (00 hrs), which is the start of the next day. For example, for each hour h of the next day, the forecast is based on the values of the 10 predictors at the hour h of the previous day.

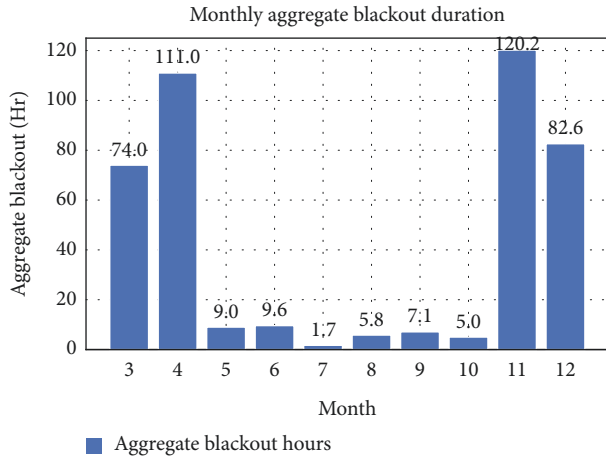


FIGURE 4: Monthly aggregate blackout history. March, April, November, and December had the most blackout events.

Similarly, 15 minutes ahead prediction and 1 hour ahead prediction are performed in the same manner based on the corresponding time step in the past records. During the calibration phase, measurements for important parameters such as voltage, current, and frequency were taken at intervals with a commercial multimeter and corroborated with measurements of the test site smart meter datalogger.

3.3. Input Processing. Initial steps in this work involved outlier removal and checking the correlation of variables related to blackout prediction. To ensure the model’s input data and output quality, dataset preprocessing is indispensable. The dataset was transformed from minute resolution observation entries into 15 minute average observations as well as mean hourly observations. A dataset retimed to 15 minute resolution was used for a 15 minute very short-term prediction, whereas dataset retimed to hourly resolution was used for an hour ahead and day-ahead blackout short-term prediction. Evidently, there could be a risk of misrepresenting the duration of a blackout say if it occurred at the 55th minute of the hour in question. In order to overcome this challenge, the blackout variable represents normalized blackout in an hour using values ranging from “0” to “1” (60 minutes of blackout). For example, in the case of the blackout that occurred in the 55th minute of the hour and lasted for 5 minutes, this is represented by the blackout variable as “0.08” whereas blackout that occurred at the 45th minute of the hour and lasted for 15 minutes would be represented by the blackout variable as “0.25,” and so on. This logic was used to solve blackout prediction as a regression machine learning problem. On the other hand, the short-term blackout prediction was also formulated as a classification machine learning problem. Obviously the shorter the data sampling period the more meaningful and truthful it captures occurred blackout events. In this regard, 15 minutes interval data are better than hourly and so forth. Furthermore, input data were smoothed to remove outliers and noise to optimize the prediction skill of the model.

Power supplied at the pilot test site by the local utility company has been observed to be at times unstable, and suffering from irregular power outages may be experienced at the client-side. In the period between January and December 2021, aggregated blackouts measured at the pilot site amounted to an equivalent of about 16 days. As shown in Figure 4, the months of March, April, November, and December were the worst hit by power outages. Power outages are irregular, and the pattern differs from month to month; for instance, from Figure 5, the month of May suffered fewer power interruptions than April. From the heat map, “1” indicates a complete blackout for an entire hour, whereas “0.5” signifies a blackout for 30 minutes during the respective hour being considered. In some extreme cases, blackout may last more than 24 hours, as was the case on April 18th and 19th.

Power outage notifications are typically sent out in advance through the media if the scheduled power interruption affects or covers a large area, such as an entire city or district. However, few or no power interruption notifications were received in advance at the pilot’s site office buildings because only a small neighborhood was affected. These kinds of localized blackouts are usually due to distribution line faults or maintenance work. Therefore, this work endeavors to predict power outages merely by using less information regarding electric power parameters as observed from the point of coupling at the customer side and without having prior information about scheduled power outages. The energy management system installed at the test site is tasked with predicting blackouts, without access to information about the grid status from TANESCO’s substation control center, or without knowledge of any fault or protection relay that may have tripped upstream.

3.4. Blackout Forecasting Model. This section explains the 3 blackout models investigated in this work, namely, the Random Forest (RF) algorithm, the Adaptive Similar Days (ASD) model, and a hybrid RF-ASD model. Their performance and efficacy are given later in the result section. In this work, blackout forecasting is tackled as both a regression problem as well as a classification problem. The objective of blackout regression is to predict continuous value output that indicates the blackout event and duration. On the other hand, the objective of the blackout classifier is to predict a binary output, which indicates the occurrence of a blackout event.

Adaptive Similar Days (ASD) Blackout Prediction Approach: it has long been established that the past day’s data in a time series can be used to make short-term forecasts. As already observed in Figures 4 and 5, some months suffer from more power outage episodes than others, and the outage trend may evolve dramatically from one month to the next. Due to the stochastic nature of blackouts, a longer moving window could corrupt the training algorithm and yield lower accuracy. Therefore, it was deemed necessary to develop a model that uses fewer training data points for prediction rather than a model that requires a large dataset to gain prediction competency. It is on this premise that past

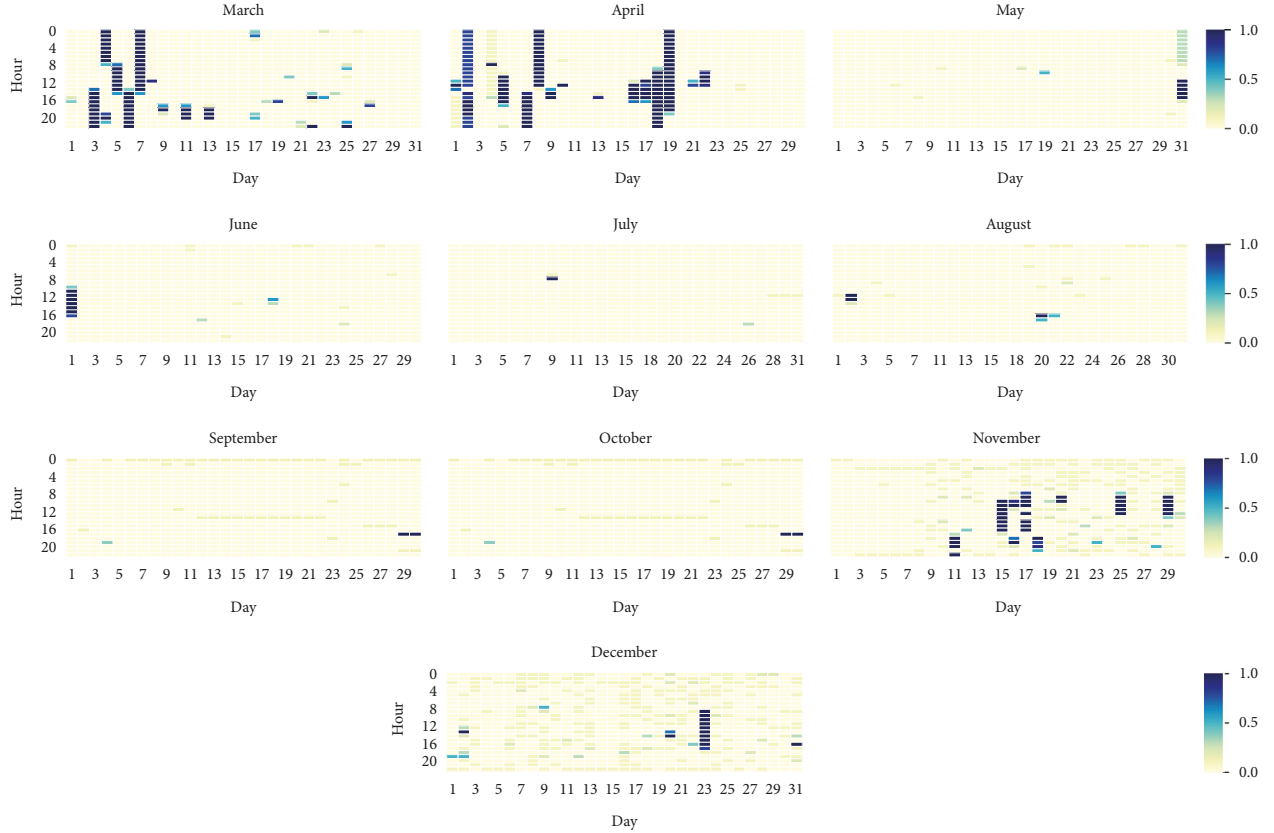


FIGURE 5: Power outage heatmap across different months. The power outage profile changes from month to month.

historical blackout data going 2 weeks back were used in our method for short-term (15 minute-ahead, hour-ahead, and day-ahead horizon) power outage prediction. Given below is the adaptive similar days (ASD) algorithm equation for short-term blackout prediction.

$$Bf_t = \frac{e_t}{N} + \frac{1}{n} \sum_{t=1}^n (b1_t + b2_t + b3_t + b7_t + b14_t). \quad (1)$$

From equation (1), assume short-term (either 15 minutes-ahead, hour-ahead, or 24 hrs ahead) blackout forecast to be represented by Bf_t . n is the number of blackout data points. $b1_t$ is the blackout index value at time t one day before, $b2_t$ is blackout index at time t two days before, $b3_t$ is a blackout index at time t three days before, $b7_t$ is a blackout index at time t seven day before, and $b14_t$ is blackout index at time t fourteen days earlier. e_t is the forecast error, which is obtained as a difference between the prediction value and the observed value. N is a positive number, forming the fraction error term e_t/N . The vectors $b1$ to $b14$ form 2 weeks sliding window. They are chosen because they are closer to the short-term forecast target, and it is assumed that they will capture any blackout pattern nuances that may exist in the time series data, if any cyclical trends exist. For example, in predicting blackout events of the current week's Monday, the previous week Monday's blackout data ($b7_t$) as well as the Monday 2 weeks ago ($b14_t$) are assumed to have some influence in the blackout prediction dynamics.

The short-term blackout forecast Bf_t in this work is computed according to equation (1), using the following algorithm steps:

- (i) Step 1: then will be determined by the mean of the past blackout vectors for $t-1$, $t-2$, and $t-3$. For example, if we wish to predict the second hour of the day, $t=01$ hrs, then will be the blackout index value in the previous 1 day at the corresponding second hour of the day (01hrs), whereas is the blackout index value 2 days before at the corresponding second hour of the day (01hrs), and so on.
- (ii) Step 2: the resulting short-term power outage prediction is compared to the actual observed blackout data for that respective day (prediction target day), and the difference (error), e_t/N , is saved in a lookup table.
- (iii) Step 3: the resulting short-term blackout prediction vector is summed with the fraction of the previous prediction error vector, e_t/N , from the lookup table (memory). After performing sensitivity analysis, $N=2$ was found to increase prediction accuracy. The term e_t/N is used to correct the weights of the day-ahead predicted power outage profile, Bf_t .
- (iv) Step 4: the 2-week sliding window is moved one interval step forward, to make the next short-term power outage prediction.

- (v) Step 5: repeat steps 1 to 4.

Additionally, some general assumptions that govern the mechanism of the ASD and RF algorithms used in this work are as follows:

- (i) Assumption 1: recent records or observations closer to the short-term prediction target have a stronger influence on the intended target period being forecasted. In other words, they have a higher probability of being similar to the forecast target period. Therefore, in the vein of the Pareto principle, past records used in the 2 week sliding window of the ASD algorithm have more importance in predicting the short-term target than past records beyond or before the sliding window. In case there were no blackouts in the 2 weeks' window, the probability of an imminent blackout is assumed to be very low.
- (ii) Assumption 2: the power outage duration for the short-term target period does not exceed the maximum outage event duration of the preceding outages within the sliding window.
- (iii) Assumption 3: by extension to assumption 1, power interruptions and disturbances are cyclical in nature; thus, the interval of the third outage event is assumed to be the same as the interval between the previous two blackout events. For example, if the first blackout event of a day is observed at time t and the second blackout occurs at time $3t$; then, the algorithm will expect the third blackout event to occur at time $5t$. Granted, this may not hold true in all cases due to the stochastic behavior of blackouts. For this reason, assumption 4 is used in effect to reset the learnt interval pattern if an incorrect prediction is made.
- (iv) Assumption 4: under normal conditions, it is assumed that the power line has electricity by default, while power interruption is considered an abnormal condition. At the start of the prediction day, the ASD algorithm is given a temporary prediction token or permit, after which it can make a prediction. This involves asserting a flag variable. However, it incorrectly predicts the occurrence of blackout in the short-term target period contrary to the observed data showing no outage or being under normal powered conditions. In this case, the token is temporarily revoked or withheld by clearing the flag variable, and the power line is now considered to be back to normal operating conditions. Yet again, when a new power outage is detected, a

prediction token is granted back to the ASD algorithm by re-asserting the flag variable, and it continues to participate in subsequent outage predictions.

Random Forest (RF) algorithm is a nonparametric machine learning method that can handle nonlinear regression as well as classification challenges. It is based on decision trees, which individually act as weak learners but overall become strong learners. The RF algorithm is robust and has been found to work well with small datasets, as well as datasets with some missing values. A random forest of 10 bagged ensemble regression trees and 100 classifier trees were grown and used to forecast short-term power outages. For this kind of problem, short-term forecasts depend on past recorded data; therefore, the walk-forward validation method was used instead of cross-validation. After training and fitting the RF model, a single-step forecast is made, followed by error measurements, and then the RF model is updated with observed data for the predicted time (target day) by appending it to the input data ready for the next one-step forecast loop. The model steps through the entire test data in this manner of predicting and updating, until the last test data are reached by the walk-forward validation.

Random Forest Adaptive Similar Days (RF-ASD) Hybrid Model. Output from the ASD module that predicts future short-term time step from a sliding window lookup table spanning 14 days prior to the target day is fused with RF prediction, giving a final optimized RF-ASD blackout prediction via element-wise vector mean of both models. In other words, the output of the ASD model is combined with that of the RF model to obtain the mean of the two models. Thus, the RF-ASD hybrid model is an average of both the RF and ASD models. Figure 6 also describes the RF-ASD blackout prediction algorithm flowchart.

Model Evaluation Indices. To assess blackout classifier model performance, a confusion matrix was used along with classification accuracy score, precision, recall, and $F1$ score. The classification *accuracy score* gives the ratio of correct forecasts to the remaining forecasts; it denotes how accurate the prediction model is. The accuracy score is represented in equation (2). True positive values are those whose observed value and forecasted value are true. A False negative is an error classification where the observed value is correct but the predicted value is false. A false positive is also a misclassification where the forecast value is true, whereas the observed value is false. True negative is the case where the observed value is false and the forecast value is also false:

$$\text{accuracy score} = \frac{\text{true positive} + \text{true negative}}{\text{true positive} + \text{true negative} + \text{false positive} + \text{false negative}}. \quad (2)$$

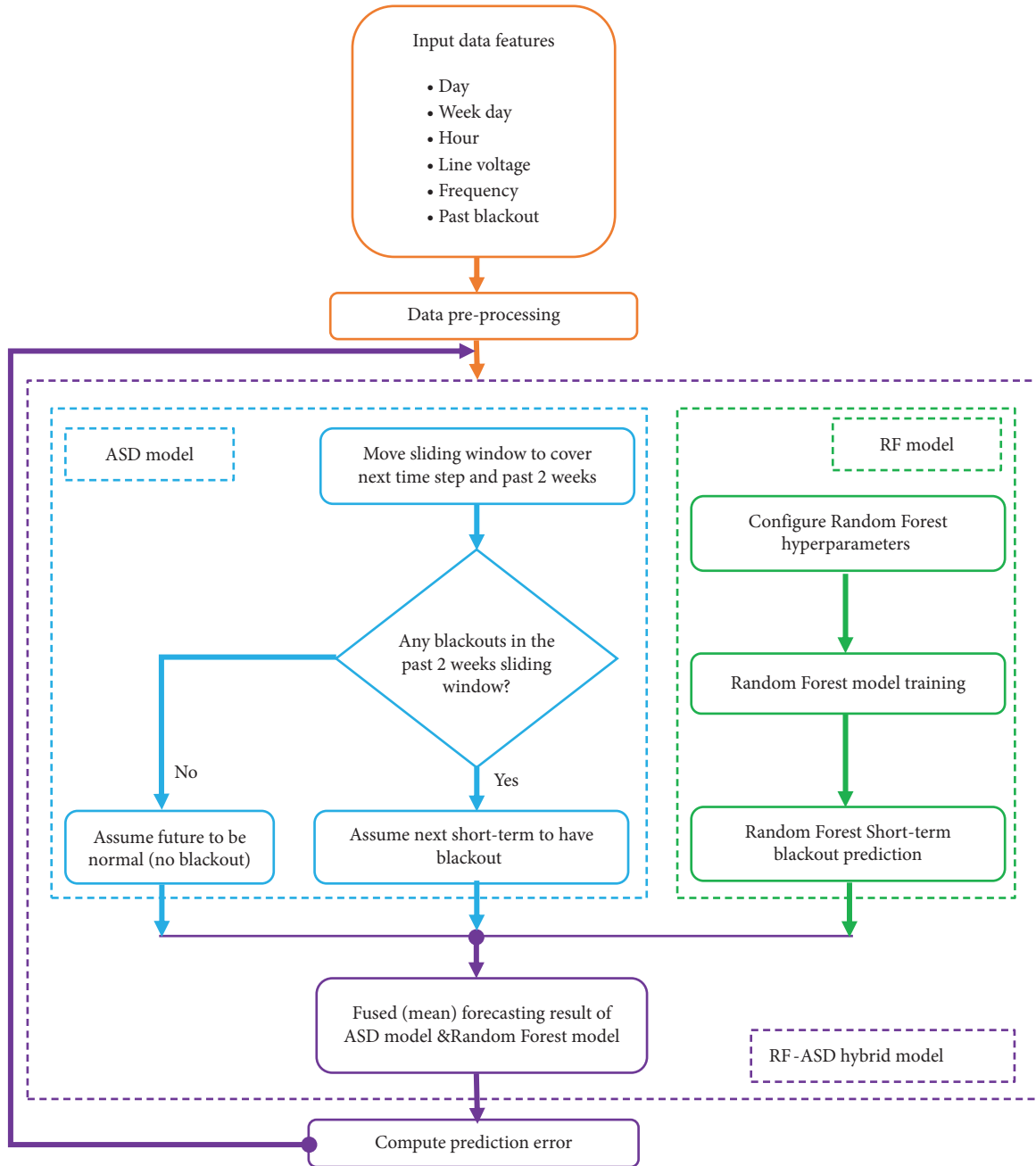


FIGURE 6: RF-ASD hybrid model algorithm flow chart.

The recall metric may be defined as the ratio of correctly classified data divided by the total actual samples of the target class. The Blackout recall used in this work is represented below by equation (3). On the other hand, the precision metric is the ratio of correct positive predictions relative to total positive predictions. It is represented in equation (4). A classifier with good precision will not label as positive a data sample that is negative. Another classification metric, especially useful in this work due to the imbalanced dataset employed, is the *F1* score. *F1* score is the harmonic mean of precision and recall. The best *F1* score is 1, while the worst is 0:

$$\text{blackout sensitivity (recall)} = \frac{\text{correctly predicted blackouts}}{\text{total actual blackout}}, \quad (3)$$

$$\text{precision} = \frac{\text{correctly predicted blackouts}}{\text{total predicted blackout}}, \quad (4)$$

$$F1 \text{ score} = 2 \times \frac{(\text{precision} \times \text{recall})}{(\text{precision} + \text{recall})}, \quad (5)$$

To evaluate the blackout regression algorithm's prediction skill, the mean absolute error (MAE) and root mean square error (RMSE) were employed. Since grid power is ON most of the time, MAE and RMSE of predicted values were measured only against instances when blackout events were actually observed in order to focus and gauge the efficacy of the models in blackout prediction. The MAE metric is more resilient to outliers in the results, whereas RMSE penalizes outliers in the results. They are given by

$$\begin{aligned} \text{MAE} &= \frac{1}{n} \sum_{j=1}^n |y_j - \hat{y}_j|, \\ \text{RMSE} &= \sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2}, \end{aligned} \quad (6)$$

whereby n is the number of blackout data points, y_j is the observed blackout value, and \hat{y}_j is the forecasted short-term blackout value.

4. Results and Discussion

This part discusses the results obtained from investigating the efficacy of RF, ASD, and RF-ASD models in blackout prediction. Figure 7 shows the blackout classification accuracy score for RF, ASD, and RF-ASD models. The results are for a 15-minute-ahead power outage forecast. The three models showed an overall accuracy score of about 90%. However, because most of the time power is available, observed actual blackout events were fewer in the test dataset, inevitably resulting in an imbalanced dataset. Therefore, the machine learning models inadvertently also get a few effective observed blackout samples to train on. It is ideally desired for any classifier model to classify all samples appropriately as True Negatives (TN) and True Positives (TP)—the two diagonal parts of the 2×2 confusion matrix below.

The stand-alone RF model correctly classified 89.4% of the test data samples as the power available (TN), while 1.5% were misclassified as power outages (FP), instead of being predicted as instances where the supply line had power ON (available). The RF model misclassified about 6.2% of the blackout events and mistook them for power-on instances instead of blackouts. The RF model only correctly predicted 672 power outages out of a total of 2085 blackout events. These 2085 blackouts formed only 9.1% of the total samples, whereas actual power was available for only 90.9% of the recorded data. The RF model did relatively better at predicting the presence of power and fared badly in predicting blackout events as compared to the ASD model and RF-ASD hybrid model. The ASD model did better than the RF model in predicting blackouts; it accurately predicted 971 counts of blackout events. The RF-ASD hybrid model predicted accurately almost half of the blackouts (1025 counts), thereby performing slightly better than the RF and ASD models.

Typically, in the emerging countries scenario, the user may want to know in advance if there is going to be any power outage the following day in order to take appropriate actions to alleviate the effects of a lack of electricity supply.

For this, the 24 hrs ahead blackout predictions may be useful. The hour-ahead and 15 minutes ahead blackout predictions may be advantageous to a grid operator. Table 2 summarizes the overall blackout forecast classification of the 3 models RF, ASD, and RF-ASD along 3 forecast horizons, namely, 15 minutes ahead, hour-ahead, and 24 hrs ahead forecast horizon. Considering the first accuracy score metric, it was found to be 92.4%, 90.6%, and 90.2% for RF, ASD, and RF-ASD models, respectively, for the 15 minutes ahead forecast horizon. These are high accuracy scores for a classification model; however, they have been driven up by the majority class (power ON class data) instead of the blackout minority class. The accuracy score is also high for the hour-ahead forecast horizon. The accuracy score was found to be low for the 24-hours ahead forecasts due to the stochastic nature of blackouts. ASD model fared better with an accuracy score of 85%. Attempting to predict blackout many time steps in advance is more prone to forecast errors as was found to be the case in the 24 hours ahead blackout prediction. All in all, an accuracy score in this case does not correctly reflect the performance of the model in predicting blackouts, which is our target.

Looking at the 15 minutes ahead forecast horizon, the recall metric was found to be 32.2%, 46.6%, and 49.2% for RF, ASD, and RF-ASD models, respectively. In our case, the higher the recall value, the better, as it implies that the model was able to predict more blackout events accurately. Combining the RF and ASD models to form RF-ASD had a good effect of increasing blackout recall up to 49.2%. The same effect was observed in the case of hour-ahead and 24 hours ahead forecasts. Just like recall, it is desirable to have a model with a higher precision value. The RF model had higher precision values than the ASD and RF-ASD models in both 15 minutes ahead and hour-ahead forecast horizons. However, the ASD had higher precision in the 24 hours ahead horizon. The $F1$ score is a suitable metric for identifying an overall best performing classifier model for an imbalanced dataset, such as is the case in this work. The RF-ASD model scored slightly higher, with an $F1$ score of 47.7%, with the other models lagging behind. Therefore, the RF-ASD model is made a better candidate for 15 minutes ahead blackout forecasting. In hour-ahead forecasts, the RF model lagged behind in performance compared to ASD and RF-ASD, which both scored 45.7%. The ASD model outperformed RF and RF-ASD models in 24-hours ahead blackout forecasts.

Figure 8 gives insight into the performance of the blackout classifier across different months. The chart below considers the RF-ASD classifier model for 15 minutes ahead predictions. The overall accuracy score is observed to be relatively high from March to October, where it then drops to 64% in November and 57% in December.

The $F1$ score is derived from both recall and precision metrics and is a better indicator of the performance of a classifier. The $F1$ score for March, April, and May performed fairly well above 50%. However, it drops to 18%, 13%, and 34% in June, July, and August. In contrast, the accuracy score remains high, above 90%. With respect to Figure 5 of Section 3, June, July, and August had less consistent

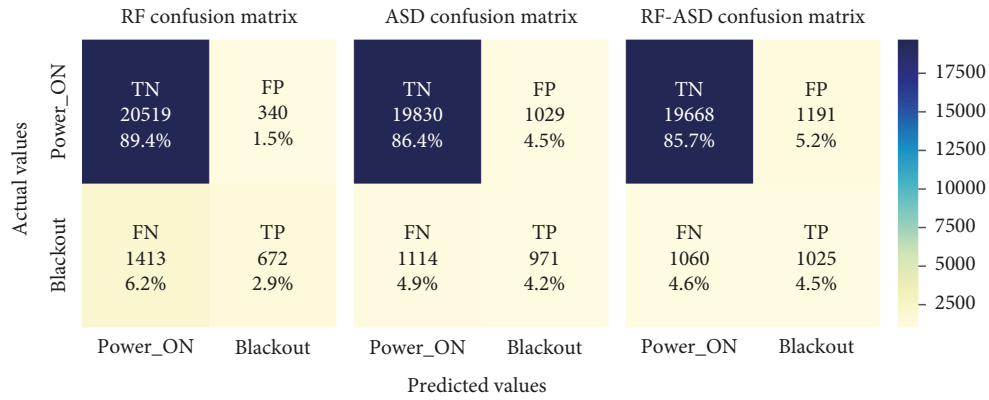


FIGURE 7: Power outage classification accuracy scores for 15 minutes ahead predictions.

TABLE 2: Overall blackout forecast classification performance.

Forecast horizon	Classifier model	Accuracy score (%)	Blackout sensitivity (recall)	Blackout precision (%)	F1 score (%)
15 minutes ahead	RF	92.4	32.2	66.4	43.4
	ASD	90.6	46.6	48.6	47.6
	RF-ASD	90.2	49.2	46.3	47.7
Hour-ahead	RF	88.4	17.0	52.1	25.6
	ASD	87.3	45.7	45.8	45.7
	RF-ASD	86.8	47.2	44.3	45.7
24 hrs ahead	RF	62.0	67.8	19.4	30.2
	ASD	85.0	38.0	36.6	37.3
	RF-ASD	60.9	73.4	19.4	30.7

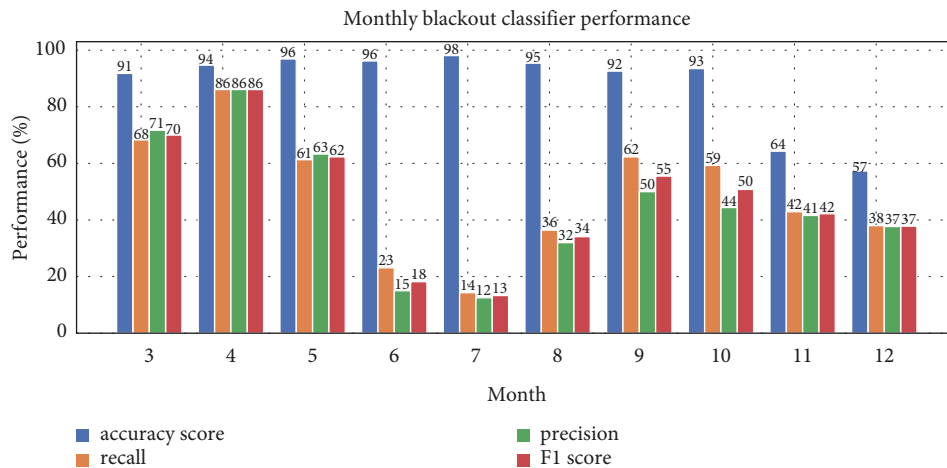


FIGURE 8: Monthly blackout forecast performance for 15-minutes ahead RF-ASD classifier model.

blackouts, which occurred sparsely, thus affecting negatively the performance of the model in predicting blackouts. For this reason, the classifier model performed poorly in recall, precision, and F1 score metrics. November and December had many random blackouts, thus causing the classifier model to perform poorly in accuracy, recall, precision, and F1 scores. Although March and April had about the same level of blackouts as November and December, blackouts in March and April were contiguous and more converged, unlike those of November and December which were

dispersed. Therefore, the classifier had better prediction skills in those earlier months than in November and December.

Table 3 summarizes the performance of the RF, ASD, and RF-ASD regression models. The three models were formulated as regression models in order to tackle the challenge of quantifying the duration of the short-term blackout prediction. The overall prediction skill for the three models had small differences from each other in all the three forecast horizons considered. In the case of 15 minute ahead

TABLE 3: Overall performance for blackout forecast regression models.

Forecast horizon	Regression model	MAE	RMSE
15 minutes ahead	RF	0.253	0.333
	ASD	0.251	0.325
	RF-ASD	0.253	0.329
Hour-ahead	RF	0.169	0.298
	ASD	0.203	0.323
	RF-ASD	0.185	0.296
24 hrs ahead	RF	0.268	0.431
	ASD	0.254	0.409
	RF-ASD	0.258	0.418

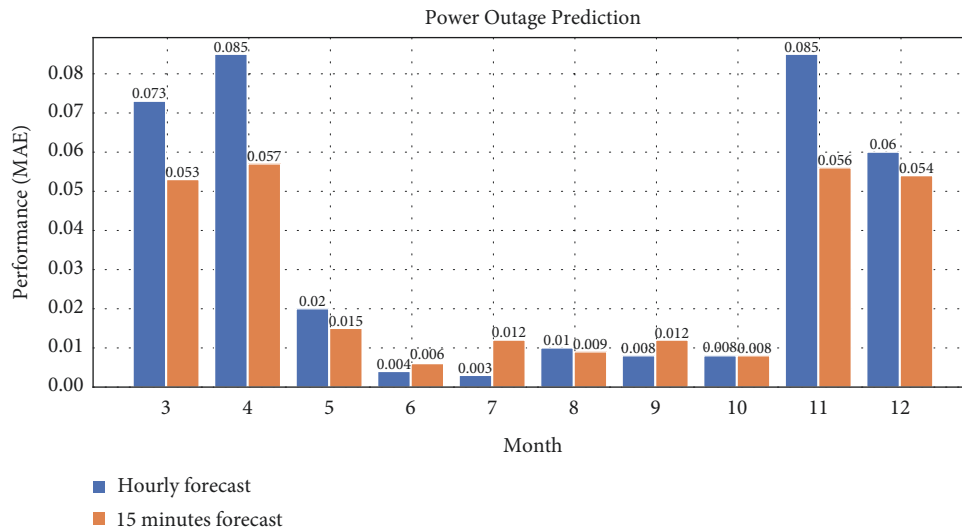


FIGURE 9: Performance of the RF-ASD blackout forecast model on different months' data, considering the hour-ahead and 15-minutes ahead forecast horizons. March, April, November, and December had the most blackout events, thus affecting the performance of the blackout regression models.

forecasts, the ASD model slightly improved with respect to both MAE and RMSE. Although the RF model had a similar MAE score as the RF-ASD model, its RMSE was slightly worse than that of RF-ASD. This implies that the output of the RF model suffered more from outliers. In the case of hour-ahead forecasts, the RF model had a lower MAE value, but the RF-ASD model had the lowest RMSE value, meaning that the RF model suffered from slightly more outlier results than the RF-ASD model. The ASD model had slightly better MAE and RMSE values than the RF and RF-ASD models for the case of 24hr ahead forecasts. In the case of 15 minute ahead forecasts, 96 (24×4) predictions had to be made per day, while the hour-ahead approach requires only 24 predictions per day. Statistically the 15 minutes ahead approach ends up being more prone to forecast errors than the hour-ahead forecast approach. This is visible in Table 3 results.

Figure 9 gives further insight into the average performance of the RF-ASD regression model comparable across different months of the test data. All the models performed relatively poorly in March, April, November, and December whereas the grid was under severe blackout disturbances and

also due to the haphazard nature of the outages. It is worth noting that, in the preceding named months where blackouts were prevalent, 15 minutes ahead forecasts outperformed their counterparts, namely, 24 hour ahead forecasts and hour-ahead forecasts. The RF-ASD 24 hrs ahead prediction produces a blackout prediction for the entire 24 hours of the next day, whereas the RF-ASD hour-ahead blackout prediction algorithm only makes prediction one hour-ahead at a time, and is, therefore, able to update and notice the developing blackout trend of the grid. Thus, the RF-ASD hour-ahead blackout prediction algorithm gains prediction skills and self-learns any prior inaccurate predictions.

The models developed in this work were fairly able to predict power outages using only few information regarding electric power parameters as observed from the point of coupling at the customer side and whilst having neither prior information of scheduled power outages nor weather forecasts regarding any looming extreme weather events that might cause blackout. It is more difficult to predict the duration of a blackout event than the mere occurrence of a blackout. Generally, the shorter the forecast horizon, the

more realistic and practical the prediction result than when a long forecast horizon is used for grid dynamics, which may quickly change between one forecast and the next.

5. Conclusions and Future Work

In this paper, a short-term blackout forecasting model framework has been proposed. Generally, the regression and classification algorithms considered in this work, namely, RF, ASD, and RF-ASD, had about the same performance in blackout prediction. The models developed were only able to predict blackouts if they occurred frequently and contiguously, but they performed poorly if they were sparse or dispersed. The developed models merely make an educated guess on the possible occurrence of a blackout but not the precise time of the outage incidence. Overall, the blackout regression and classification models investigated in this work had fair performance in the power outage prediction challenge along the test data considered months.

The advantage of the prediction models on the PV system and load side reliability is that the blackout forecast models developed have the potential to improve battery management on the consumer side. It is true that the PV-battery and inverter system are meant to supply loads during blackouts, thereby mitigating the effects of blackouts. Since PV systems in developing countries are sometimes pico-sized, they do not offer many hours of autonomy, hence, having blackout prediction in advance may help consumers perform load-shifting by moving energy-intensive tasks (tier 1 big load tasks) outside power outage times and utilizing the mains grid, which can support big loads, unlike pico-sized PV systems typically in place. The output of a blackout forecast could be valuable to the battery management system (BMS) in ensuring that battery charge levels are enough to sustain an imminent blackout episode, especially since the PV systems in this study were very small; blackout forecasts remain indispensable and ensure battery charge levels go a long way in meeting load demand during a blackout. For example, if the output of the blackout prediction model shows an imminent blackout, this can help BMS decide to go into conservation mode, whereby the smart BMS operates the battery storage near full charge. Thus, ensuring the battery is fully charged and ready to withstand the predicted impending blackout thereby increasing load-side reliability. This prepares the PV system to cope with any sudden power outage; however, if the probability of blackout is very low, as predicted by the model, then the system can go into a relaxed mode where the battery SoC level is allowed to discharge to low levels.

Future works could study implementation of BMS strategies using blackout forecast output. This work has endeavored to predict blackouts only from the customer's point of connection to the grid without having other information about the grid as a whole, which could be experiencing disturbances that may result in a blackout. Future work could supply the blackout prediction model with more data about the status of the local electric grid at large; this would give the model a better vantage point in predicting imminent blackouts and increase performance. In contrast

to this work, future works could also study blackout prediction from the utility distributor point of view. Blackout prediction is a challenging task because it is caused by many factors such as weather, various faults in the grid, and their complex interactions. It would also be interesting to investigate blackout prediction models with inputs from weather forecasts, sensor data from grid protection devices, relays, and so forth. Furthermore, an extension to this work could be investigating a blackout prediction model that takes into account not only presence or absence of electric power but rather considers also power quality issues such as mains frequency stability, overvoltage, and undervoltage and more or large PV systems.

Data Availability

The data used to support the findings of this study will be available upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors express their heartfelt gratitude to the Nelson Mandela African Institution of Science and Technology (NM-AIST) for facilitating this research; gratitude also extends to the Politecnico di Milano energy4growth team for their kind assistance. The authors also acknowledge Arusha Technical College for supporting and promoting academic staff development.

References

- [1] International Energy Agency, *Africa energy outlook*, International Energy Agency, Paris, France, 2019.
- [2] A. T. Falentina and B. P. Resosudarmo, "The impact of blackouts on the performance of micro and small enterprises: The impact of blackouts on the performance of micro and small enterprises: Evidence from Indonesia," *World Development*, vol. 124, Article ID 104635, 2019.
- [3] N. B. S. Rea and, "Energy Access and Use Situation Survey II in Tanzania Mainland 2019/20," 2020, <http://www.rea.go.tz/Resources/E-Library/tabid/132/Default.aspx>.
- [4] S. Abdullah and A. Markandya, "Rural electrification programmes in Kenya: policy conclusions from a valuation study," *Energy for Sustainable Development*, vol. 16, no. 1, pp. 103–110, 2012.
- [5] N. P. Nkosi and J. Dikgang, "Pricing electricity blackouts among South African households," *Journal of Commodity Markets*, vol. 11, pp. 37–47, 2018.
- [6] B. G. Pollet, I. Staffell, and K.-A. Adamson, "Current energy landscape in the Republic of South Africa," *International Journal of Hydrogen Energy*, vol. 40, no. 46, Article ID 16685, 2015.
- [7] K. Aidoo and R. C. Briggs, "Underpowered: rolling blackouts in Africa disproportionately hurt the poor," *African Studies Review*, vol. 62, no. 3, pp. 112–131, 2019.
- [8] B. Gou and W. Wu, "Is the prediction of power system blackouts possible?" in *Proceedings of the IEEE Power and*

- Energy Society 2008 General Meeting: Conversion and Delivery of Electrical Energy in the 21st Century*, pp. 1–4, PES, Pittsburgh, PA, USA, July 2008.
- [9] R. Nateghi, S. Guikema, and S. M. Quiring, “Power outage estimation for tropical cyclones: improved accuracy with simpler models,” *Risk Analysis*, vol. 34, no. 6, pp. 1069–1078, 2014.
- [10] R. Nateghi, S. D. Guikema, and S. M. Quiring, “Forecasting hurricane-induced power outage durations,” *Natural Hazards*, vol. 74, no. 3, pp. 1795–1811, 2014.
- [11] K. M. Alkar, M. Meto, M. Jamjum, and K. Amar, “Performance evaluation of the blackout and power outages in Libyan power grid-Al-zawia combined cycle power plant case study,” in *Proceedings of the 2019 1st International Conference on Sustainable Renewable Energy Systems and Applications*, pp. 1–6, (ICSRESA), Tebessa, Algeria, December 2019.
- [12] K. M. J. Rahman, M. M. Munnee, and S. Khan, “Largest blackouts around the world: trends and data analyses,” in *Proceedings of the 2016 IEEE International WIE Conference on Electrical and Computer Engineering*, pp. 155–159, (WIECON-ECE), Pune, India, December 2016.
- [13] Z. Bo, O. Shaojie, Z. Jianhua, S. Hui, W. Geng, and Z. Ming, “An analysis of previous blackouts in the world: lessons for China’s power industry,” *Renewable and Sustainable Energy Reviews*, vol. 42, pp. 1151–1163, 2015.
- [14] M. Sforna and M. Delfanti, “Overview of the Events and Causes of the 2003 Italian Blackout,” in *Proceedings of the 2006 IEEE PES Power Systems Conference and Exposition*, pp. 301–308, Atlanta, GA, USA, November 2006.
- [15] Y. Makarov, V. I. Reshetov, A. Stroeve, and I. Voropai, “Blackout prevention in the United States, Europe, and Russia,” *Proceedings of the IEEE*, vol. 93, no. 11, p. 1942, 2005.
- [16] S. Mei, F. He, X. Zhang, S. Wu, and G. Wang, “An improved OPA model and blackout risk assessment,” *IEEE Transactions on Power Systems*, vol. 24, no. 2, pp. 814–823, 2009.
- [17] L. Cheng, M. Liu, D. Wang, D. Wu, Z. Xu, and Y. Su, “The outage rate model based on the improved evidence theory considering various outage factors,” in *Proceedings of the 2017 IEEE PES Asia-Pacific Power and Energy Engineering Conference*, pp. 1–6, (APPEEC), Bangalore, India, November 2017.
- [18] T. Kogo, S. Nakamura, S. Pravinraj, and B. Arumugam, “A demand side prediction method for persistent scheduled power-cuts in developing countries,” in *Proceedings of the IEEE PES Innovative Smart Grid Technologies*, pp. 1–6, Istanbul, Turkey, October 2014.
- [19] M. Papic and O. Ciniglio, “Prediction and Prevention of Cascading Outages in Idaho Power Network,” in *Proceedings of the 2014 IEEE PES General Meeting| Conference & Exposition*, pp. 1–5, National Harbor, MD, USA, July 2014.
- [20] M. Papic, D. Angell, P. van Patten, and B. Efav, “Historical performance analysis of transmission lines using Idaho power company’s outage database GTORS,” in *Proceedings of the 2018 IEEE International Conference on Probabilistic Methods Applied to Power Systems*, pp. 1–6, (PMAPS), Boise, ID, USA, June 2018.
- [21] H. Chen, S. Liu, Q. Liu et al., “Estimating the impacts of climate change on electricity supply infrastructure: a case study of China,” *Energy Policy*, vol. 150, Article ID 112119, 2021.
- [22] T. Zachariadis and P. Hadjinicolaou, “The effect of climate change on electricity needs—a case study from Mediterranean Europe,” *Energy*, vol. 76, pp. 899–910, 2014.
- [23] S. N. Chandramowli and F. A. Felder, “Impact of climate change on electricity systems and markets—a review of models and forecasts,” *Sustainable Energy Technologies and Assessments*, vol. 5, pp. 62–74, 2014.
- [24] C. Muchunku, K. Ulsrud, D. Palit, and W. Jonker-Klunne, “Diffusion of solar PV in East Africa: what can be learned from private sector delivery models?” *WIREs Energy and Environment*, vol. 7, no. 3, p. e282, 2018.
- [25] A. Abubakar Mas’Ud, A. V. Wirba, F. Muhammad-Sukki et al., “A review on the recent progress made on solar photovoltaic in selected countries of sub-Saharan Africa,” *Renewable and Sustainable Energy Reviews*, vol. 62, pp. 441–452, 2016.