

Article

# Data-Driven Prediction Method for Power Grid State Subjected to Heavy-Rain Hazards

Seongmun Oh <sup>1</sup>, Junhyuk Kong <sup>2</sup>, Minhee Choi <sup>3</sup> and Jaesung Jung <sup>2,\*</sup> 

<sup>1</sup> Energy Conversion Research Laboratory, Korea Electronics Technology Institute, Gwangju 61011, Korea; sm91@keti.re.kr

<sup>2</sup> Department of Energy Systems Research, Ajou University, Suwon 16499, Korea; jhgong8338@ajou.ac.kr

<sup>3</sup> Smart Power Distribution Laboratory, KEPCO Research Institute, Daejeon 34056, Korea; minhee.choi@kepco.co.kr

\* Correspondence: jjung@ajou.ac.kr; Tel.: +82-31-219-2695

Received: 14 May 2020; Accepted: 6 July 2020; Published: 8 July 2020



**Abstract:** This study presents a machine learning-based method for predicting the power grid state subjected to heavy-rain hazards. Machine learning models can recognize key knowledge from a dataset without any preliminary knowledge about the dataset. Hence, machine learning methods have been utilized for solving power grid-related problems. Two sets of historical data were used herein: Local weather data and power grid outage data. First, we investigated the heavy-rain-related outage distribution and analyzed the correlated characteristics between weather and outages to characterize the heavy rain events. The analysis results show that multiple weather effects are significant in causing power outages, even under heavy-rain conditions. Furthermore, this study proposes a cost-sensitive prediction method using a support vector machine (SVM) model. The accuracy of the model was improved by applying a cost-sensitive learning algorithm to the SVM model, which was subsequently used to predict the state of the grid. The developed model was evaluated using G-mean values. The proposed method was verified via actual data of a heavy rain event that occurred in South Korea.

**Keywords:** power outage; heavy rain; machine learning; power big data; grid resilience

## 1. Introduction

The electrical power grid constitutes a vital infrastructural component and serves as an essential foundation for modern life. Severe weather events have been considered a major cause of enormous power-outage events. Moreover, the frequency and intensity of severe weather events have increased over recent years [1–5]. Due to climate change, the frequency, intensity, and duration of extreme weather are expected to increase further, resulting in large-scale power outages [6] and motivating this study on grid resilience, defined as the ability to withstand and rapidly recover from a severe weather event.

A key driver for improving grid resilience is to predict the state of the grid [7]. The grid state demonstrates how the grid, or its components, performs with respect to severe weather events and it can determine the operating conditions of the grid. Identification of the grid state is an essential factor in mitigating the impact of severe weather, preparing the emergency response, and managing the restoration crew within an acceptable time and cost limits. The grid is typically designed to operate under a certain weather intensity. If the weather intensity increases beyond the design criteria, there will be an increase in the probability of a power outage. Therefore, the grid state can be represented by either a normal or outage state, given the weather intensity. Modeling the grid state is very complex; therefore, the formulation of the relationship and its solution is difficult to determine.

Data-driven methods can be utilized to model the complex relationship for the grid state; these methods are time efficient and the components do not require a physical model. Data-driven methods, such as machine learning (ML) models, can recognize key knowledge from a dataset without any prior knowledge about the dataset. Thus, ML methods have been utilized to solve power grid-related problems. ML applications include fault identification and detection, based on the support vector machine or neural network [8,9], and risk analysis [10]. A more prevalent application of ML in the power grid field is in relation to forecasting. Several ML-based forecasting methods have been introduced for short-term load forecasting [11,12], renewable power forecasting [13,14], electricity price forecasting [15], and power grid damage forecasting [16–20], which utilize regression models and tree-based regression. A probabilistic approach based on the Tobit regression model to forecast power outage was proposed by [21]. A Bayesian network-based storm-induced power outage forecasting model for an electrical grid was presented in [22]. Predictive analytics for extreme weather events were presented in [23]. A power system contingency forecasting tool under extreme weather conditions was presented in [24]. These authors proposed a risk-based security method to predict the risky contingencies on the power system.

This study presents an ML-based method for predicting the power grid state subjected to heavy-rain hazards. Two sets of historical data were used: Local weather data and power grid outage data. First, we investigated the heavy-rain-related outage distribution and analyzed the correlated characteristics between weather and outages in order to characterize the heavy-rain events. The analysis results show that multiple weather effects are significant in causing power outages, even under heavy-rain conditions. Furthermore, this study proposes a cost-sensitive prediction method using the support vector machine (SVM) model. The accuracy of the model was improved by applying a cost-sensitive learning algorithm to the SVM model, which was subsequently used to predict the state of the grid. The developed model was evaluated using G-mean values. The proposed method was verified using the actual data of a heavy-rain event that occurred in South Korea.

This study is organized as follows: Section 2 describes the related work, Section 3 describes the used data, Section 4 describes the characteristics of the heavy-rain event, Section 5 presents the method and materials for the development of the power grid state prediction model, Section 6 presents results of the actual data, and Section 7 concludes this study.

## 2. Related Work

Several researchers have studied the grid resilience corresponding to severe weather events. The linear regression model-based prediction approach was presented in [16,17]. They tried to forecast the number of damages caused by strong wind or storm events. Various statistical methods, such as regression tree and multivariate regression, were used to predict power outage durations during a hurricane [18]. Liu et al. [19,20] developed a linear model-based power outage prediction method using hurricane variables and environmental variables at a given location. Han et al. [25,26] presented a power-outage estimation model using a similar approach with Liu et al. [19,20] but with additional exploratory variables. Guikema et al. [27] introduced a generalized hurricane outage prediction model by considering the additional explanatory variables, such as land cover, soil moisture, and hurricane indicators. Recently, the power grid component state prediction method during hurricane events was presented in [28,29]. Logistic regression and support vector machines were used to model the power component state during the hurricane.

Most previous studies focused on hurricane events and their developed models relied on several grid vulnerability variables and specific hurricane indicators. However, not all power companies have detailed data about the power outage caused by weather events and it can be expensive to gather extensive datasets. Recent studies tried to predict the grid component state during the hurricane event based on a few weather variables. However, these researches used artificially generated samples rather than real power outage data. Therefore, the previous studies have limited its scalability to other weather events and real applications.

The objective of this study was to characterize the power outage caused by heavy rain and to predict grid states during a heavy-rain event based on real data. To be scalable and applicable in reality, the common weather variables were used to develop a prediction model. Therefore, it can be easily extended to other disruptive weather events.

### 3. Description of Data

Heavy-rain hazards can be determined by certain weather conditions. For instance, when the precipitation over 6 h is expected to be more than 70 mm, the Korea Meteorological Administration (KMA) issues a special weather report to warn of heavy-rain-related hazards.

The supervised ML approach generally uses previous historical data to build its prediction model. This study considers the SVM approach to predict the grid states using common weather variables. Thus, two sets of historical data, local weather and outage data, were used to analyze the power outage. All the data ranged from January 2008 to March 2018.

In order to evaluate the present power grid state over the subsequent hours, the model required a numerical weather prediction (NWP) forecast. Therefore, this study considered publicly available weather data provided by the KMA, which could be associated with South Korea's NWP system. NWP forecasts were transformed into a grid state via a machine learning model that was trained by the historical data to learn the dependence of the output on explanatory variables.

#### 3.1. Weather Station Data

The local weather data was obtained from an automatic weather station (AWS); the weather variables were recorded at least every hour. The collected data consisted of accumulated rainfall, 1 min average wind speed, 10 min average wind speed (m/s), wind direction ( $^{\circ}$ ), and humidity (%). For this study, the accumulated rainfall and 10 min average wind speed were used to analyze the power outage. These weather variables are the expected cause of power outages under heavy-rain conditions. Furthermore, when several AWSs were active in a region, this study used the maximum/minimum weather values as it is the extreme weather conditions that are associated with power outages.

The standard criteria for issuing advisory or warning information concerning extreme weather conditions in South Korea was investigated to extract the heavy-rain data. The KMA issues a special weather report whenever a sudden and significant change in weather conditions is observed. The standard criteria for heavy rain are divided into the advisory level and warning level. The advisory level means that the precipitation for 6 h is expected to be more than 70 mm or the precipitation for 12 h to be over 110 mm. The warning level means that the precipitation for 6 h is expected to be more than 110 mm or the precipitation for 12 h to be over 180 mm [30]. In this paper, the standard criteria for heavy rain were used to extract the heavy rain condition data to predict the grid state.

#### 3.2. Outage Data

The electrical power company in South Korea records the data regarding power outage events in its branch offices. The outage information includes the details of the head/branch office, time of failure, failure component, cause of failure, number of affected customers, the time required for restoration, and type of failure. The outage is related to the failure components; the location of the failures is not exactly specified in the system topology. Despite this limitation, a wider geographical area managed by each branch office can be evaluated from the data.

#### 3.3. Integration of Weather and Outage Data

Since local weather data is collected by a different institute than outage data, AWS data is required to be mapped with outage data to analyze the relationship between them. Furthermore, the KMA provides a detailed regional short-term weather (STW) forecast every 3 h at a dense grid point by gridding the whole country into  $5 \times 5$  km intervals. The grid state prediction would be performed using STWs; therefore, the available datasets, with different data frames and time periods, required

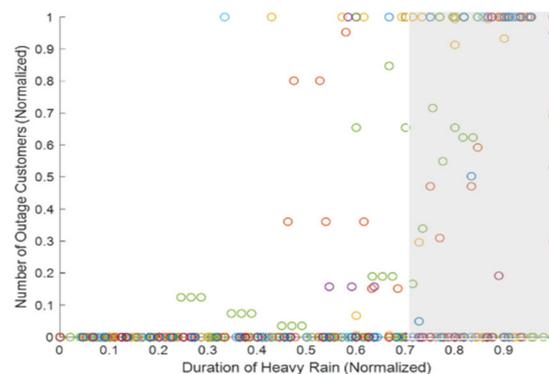
integration into 3-h segments to develop the power grid state prediction model. This data frame is developed to utilize weather forecast data. It allows the proposed method to be extended to other disruptive weather events.

#### 4. Characterizing Heavy Rain Event

##### 4.1. Outage Pattern

The U.S. Department of Energy (DOE) tracked the outage distribution for each power outage caused by major hurricane events [31]. This research reported that the number of outage customers rose within a small number of hours of the event. It suggests that the power outage caused by weather events may be observed to be a general pattern.

Therefore, we can expect that the outage distribution caused by heavy rain may follow a general pattern, similar to hurricane events. Figure 1 illustrates the normalized damage distribution for several heavy rain events. These heavy rain event durations and the number of damages were normalized into between 0 and 1 using its maximum value. Therefore, different heavy rain event durations and the number of damages can be compared over the same axis range. Consequently, Figure 1 shows that the peak number of damages often occurred close to the end of the heavy rain event time. In the case of a hurricane, the system experiences a strong weather attack early in the event. Conversely, heavy rain consistently hits the system with a relatively weak attack, causing the peak outage to occur near the end of the event. Therefore, it can be concluded that the power outage events may occur due to prolonged stress caused by accumulated rainfall under heavy-rain conditions.



**Figure 1.** Number of outage customers for heavy rain events.

##### 4.2. Skewness Distribution of Outage Data

The scarcity of available data on the power grid remains one of the predominant challenges of studying grid resilience. The occurrence of heavy-rain events is relatively rare and does not always cause grid damage. Most points on the scatter plot shown in Figure 1 lie on the  $x$ -axis, i.e., no outage cases. The figure indicates that the grid outage data is not normally distributed, which represents the skewness of the data distribution. Therefore, the grid outage data has an unbalanced label ratio and may lead to a deterioration in the performance of the machine learning model.

Due to these issues, a new method is required to capture the inherently complex characteristics of the imbalanced data and derive knowledge from the data. Innovative research regarding imbalanced learning issues includes oversampling, undersampling, and cost-sensitive learning methods [32]. Although the application of these techniques can help improve the classification results, various empirical studies have been introduced asserting that cost-sensitive learning is superior to sampling methods in certain datasets [33]. To achieve an accurate grid state prediction method, this study considered the SVM model based on a cost-sensitive method for evaluating the grid state under heavy-rain conditions.

### 4.3. Correlated Characteristic between Weather and Outage

Most power components are exposed to weather events that can cause unexpected power outages. Power components are designed to endure normal intensity weather; however, if the weather intensity exceeds the power component’s design threshold point, a power outage occurs. In outage data, wind and rain events are recorded as the causes of power outages and hence it is difficult to specify the exact type of event that induces the power outage. Moreover, heavy-rain events generate a large amount of precipitation and sometimes occur simultaneously to strong wind conditions. Therefore, two types of events must be considered as the cause of power outages under heavy-rain conditions.

There were no historical outage cases predominantly recorded, indicating that the distribution of outage data was significantly distorted. In other words, the dataset has zero-inflated characteristics since the power outage events occur with low probability even under heavy-rain conditions. Therefore, the probability density distribution of this dataset has a right-skewed distribution rather than the normal distribution. A log-transformation was performed to improve interpretability and to reduce the skewness of the data. The zero values of the outage customers were not included when analyzing the relationship between weather and outages. Figures 2 and 3 show that with higher accumulated rainfall and wind speed, the number of outage customers tended to be larger. The Pearson correlation coefficients of rainfall and wind speed with the number of outage customers are calculated as 0.3145 and 0.2354, respectively. The color bar represents the probability density, estimated via the kernel density estimation method, and the red line represents the regression line of the scatter plot. These figures show that major outages occurred with a low probability and minor outages occurred with a relatively high probability.

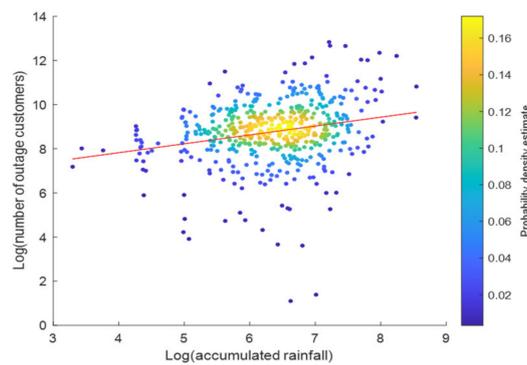


Figure 2. Average accumulated rainfall vs. number of outage customers.

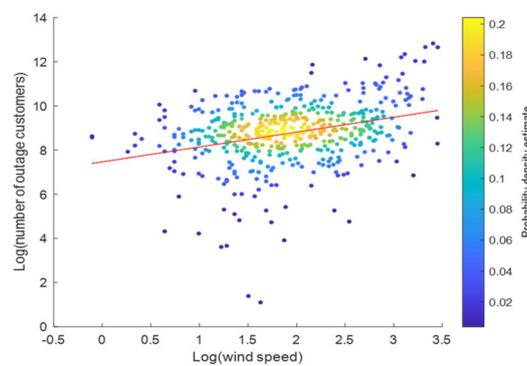


Figure 3. Average wind speed vs. the number of outage customers.

To further verify the relationship between the variables, a multivariate regression model was developed using the two weather variables as inputs. Table 1 shows the R-squared values of each model with various combinations of explanatory variables. A larger R-squared value indicates that the

model can capture a larger portion of the variance. Although all models have a low R-squared value, this analysis suggests that wind speed could help to predict the power outages even under heavy rain conditions. Moreover, the model with wind and rain has the highest R-squared value, accounting for 8.21% of the variance. This suggests that the multiple weather effect could play an important role in causing the power outage event.

**Table 1.** R-squared values of the models.

	Wind + Rain	Only Wind	Only Rain
R-squared	0.0821	0.0728	0.03

### 5. Method and Materials

The grid state can be represented by the normal state or outage state. In this study, we considered the problem of binary classification for predicting the state of the grid, as follows.

Define a binary event, O as follows:

$$O = \begin{cases} -1; & \text{Outage state} \\ +1; & \text{Normal state} \end{cases} \quad (1)$$

A data-driven method can be used to solve the above binary classification problem. SVM models have been widely used in that field, but their model performance is dependent on a dataset. SVM always guarantees the global optimum and is less prone to overfitting than neural network models, which are also widely used [34]. Moreover, SVM could be suited to small and medium datasets. Severe weather events and power outage events occur with low probability and therefore, the used dataset in this study is a relatively smaller dataset than other application field datasets, such as image recognition and natural language. For those reasons, this study utilizes the SVM approach despite SVM not always guaranteeing the most suitable model.

#### 5.1. SVM

SVM is a general approach to supervised learning for classification via the definition of a separating hyperplane, which allows discrimination between classes. It is widely used for classifying binary problems. The optimal hyperplane is considered as the hyperplane with the maximal margin between the classes, which decreases the probability of misclassification and increases the generalization of the model. The following paragraphs briefly demonstrate the SVM model in three-dimensional space for a binary classification problem. The theoretical details of SVM can be found in the literature [35].

Consider  $n$  observations  $x_i \in R^3, i = 1, \dots, n$  with corresponding target values  $t_1, \dots, t_n$ , where  $t_n \in \{-1, 1\}$ , and new data points  $x$  are classified according to the sign of  $y(w^T x + b)$ , where  $w$  is the weight vector that defines a direction perpendicular to the hyperplane of the decision boundary plane, while  $b$  is a bias that moves the hyperplane parallel to itself. The optimal decision boundary plane given by the support vector can be found by solving the optimization problem, as follows:

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \text{ subject to } t_i(w^T x_i + b) \geq 1 - \xi_i, \xi_i \geq 0, i = 1, \dots, n, \quad (2)$$

where  $\xi_i$  is a slack variable that regularizes the weight of misclassified data points inside the margin.  $C$  is a penalty coefficient that controls the trade-off between the margin and classification error. In the aforementioned formulation, increasing  $C$  toward  $\infty$  places more weight on the slack variables  $\xi_i$ , and the optimization attempts to make a stricter separation, whereas reducing  $C$  towards 0 reduces

the importance of misclassification. This ‘soft-margin’ formulation can be re-formulated using the Lagrange multiplier  $\alpha$ , as follows:

$$\max -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j t_i t_j k(x_i, x_j) + \sum_{i=1}^n \alpha_i \text{subject to } 0 \leq \alpha \leq C, \sum_{i=1}^n \alpha_i t_i = 0. \tag{3}$$

The solution of the soft-margin problem is usually found via sequential minimal optimization (SMO) [36]. SMO iteratively selects and optimizes two Lagrange multipliers until convergence at an optimal solution.

Figure 4 shows the support vectors and hyperplane in a separable two-class classification of SVM. Figure 4 shows the support vectors and hyperplane for separating outages from no outages, based on the associated weather variables. The finding of the best hyperplane assumes that the data is linearly separable; however, typical data often has non-linear characteristics. The kernel trick can be used to map the non-linear input space into a linear separable feature space. In this study, we applied the radial basis kernel trick function,  $k(x_i, x_j)$ . This can be replaced as the dot product of  $x$  in another space. The radial basis function kernel (RBF) can be expressed as follows:

$$k(x_i, x_j) = e^{-\frac{1}{2\sigma^2} \|x_i - x_j\|^2}, \tag{4}$$

where  $\sigma^2$  represents the kernel parameter. The best kernel should be found by an experiment while adjusting kernel parameters via a search method to minimize the error of the model.

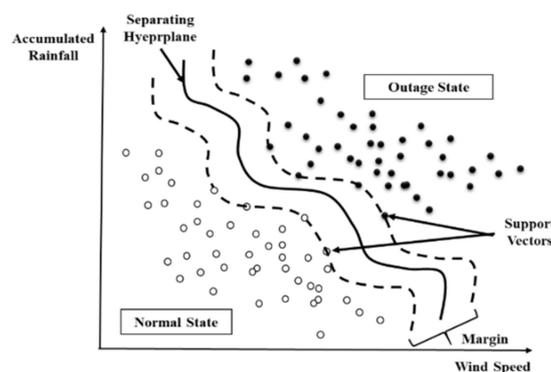


Figure 4. Support vector and margin in support vector machine (SVM).

Furthermore, the performance of the general SVM algorithm tends to be lower for an imbalanced ratio dataset as the separating boundary tends toward the minority class to obtain better classification accuracy. Therefore, in order to obtain an effective classification model, the separating boundary needs to be pushed toward the majority class and away from the minority class. This is realized through penalizing the advantages of the majority class for the separating boundary while maximizing the marginal distance. We define a cost matrix,  $C$ , associated with training observations. For the minority class, the associated misclassification cost,  $C_{ij}$ , is relatively large compared to that of the majority class. The misclassification cost penalizes the favors of the separating boundary for the majority class; therefore, the separating boundary is pushed towards the majority class. Accordingly, the degree to which the separating boundary is pushed must be determined. The element  $C(i, j)$  of this matrix is the cost of classifying an observation into class  $j$ , if the true class is  $i$ . The diagonal elements  $C(i, i)$  of the cost matrix must be 0, as follows:

$$C = \begin{bmatrix} 0 & c \\ 1 & 0 \end{bmatrix}, \tag{5}$$

where  $c > 1$  is the cost of misidentifying an outage as no outage. Cost is relative; multiplying the costs, the separating boundary for the majority class is penalized and the minority class will be preferred.

### 5.2. Evaluation Metric

Classification accuracy can be measured by the number of correct predictions out of all the predictions. However, the accuracy may fail to adequately measure the performance of the classifier due to the presence of imbalanced data. To obtain more accurate classifier information, this study adopts the G-mean evaluation metric, which represents the geometric mean of the accuracies of each class. The G-mean ranges from 0 to 1, where 1 indicates a perfect classifier.

For binary classification problems, the G-mean represents the square root of the product of a positive class accuracy and a negative class accuracy. It can be mathematically expressed as follows:

$$G - \text{mean} = \sqrt{\frac{TP}{TP + FP} \times \frac{TN}{TN + FP'}} \tag{6}$$

where TP, TN, FP, and FN represent the true positive, true negative, false positive, and false negative, respectively.

## 6. Results

The data range used for this study was from 2008 through to 2018; a historical dataset up to mid of 2016 was used to develop the prediction model and the remaining historical dataset was used to evaluate the performance of the prediction model. Furthermore, the parameters (kernel parameter ( $\sigma^2$ ) and penalty coefficient (C)) of the SVM model were tuned using Bayesian optimization. In total, 30 trials were conducted to find the best combination of the parameters with a minimum classification error of 5-fold cross-validation. A search for the C log-scaled in the range between 1e-3 and 1e3 and  $\sigma^2$  log-scaled in the range between 1e-3 and 1e3 was performed to find the best parameters of SVM. The G-mean values were evaluated for each of the 10 available explanatory variables to identify the key explanatory variables as shown in Table 2.

**Table 2.** Available explanatory variables.

Variable Name	
Average wind speed	$x_1$
Maximum wind speed	$x_2$
Minimum wind speed	$x_3$
Average wind direction	$x_4$
Maximum wind direction	$x_5$
Minimum wind direction	$x_6$
Average accumulated rainfall	$x_7$
Maximum accumulated rainfall	$x_8$
Minimum accumulated rainfall	$x_9$
12 h accumulated rainfall	$x_{10}$

The complexity of the model may lead to overfitting and a computational burden and inadequate variables can decrease the performance of the model. Thus, in this study, variables were selected by adding variables sequentially and the model with the highest performance was selected as the final model. Moreover, the degree of misclassification cost was determined to obtain an appropriate separating boundary. First, this study applied the reference ratio of the cost, 0.618:1 (the golden standard), between minority classes and the majority classes, as follows:

$$C = \begin{bmatrix} 0 & 0.618/n(t_j) \\ 1/n(t_i) & 0 \end{bmatrix}, \tag{7}$$

where  $t_j$  and  $t_i$  are the minority and majority class, respectively.

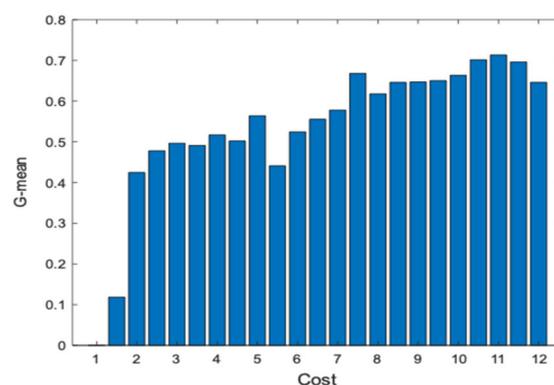
Twenty percent of the historical dataset was used to evaluate the performance of the proposed method. Based on the cost matrix C, the SVM models with a combination of weather variables were

developed and their G-mean values are summarized in Table 3. Based on the average accumulated rainfall,  $x_7$ , other variables were added sequentially. In Table 3, the G-mean values tend to decrease with the addition of the wind direction-related variables,  $x_5, x_6$ . It implies that wind direction is not a substantial factor for power grid outage subjected to heavy rain hazards. Conversely, the G-mean values tend to increase with the addition of the wind speed-related variable,  $x_1, x_2, x_3$ . This means that wind speeds have a substantial impact on the power grid even under heavy-rain conditions. Consequently, the SVM model with  $x_1, x_2, x_3, x_7, x_9$  recorded the highest G-mean values. This model is not significantly different compared to the others; however, it is superior to other models.

**Table 3.** G-mean value with a combination of variables.

Variables	G-mean
$x_1, x_7$	0.64689
$x_1, x_3, x_7$	0.67963
$x_1, x_2, x_3, x_7$	0.69427
$x_1, x_2, x_3, x_7, x_9$	0.70572
$x_1, x_2, x_3, x_6, x_7, x_9$	0.55672
$x_1, x_2, x_3, x_4, x_6, x_7, x_9$	0.60430
$x_1, x_2, x_3, x_4, x_6, x_7, x_9, x_{10}$	0.43253
$x_1, x_2, x_3, x_4, x_6, x_7, x_8, x_9, x_{10}$	0.49495
$x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}$	0.56860

To further improve the SVM model performance and analyze the effect of misclassification cost, the costs were explored with the variables  $x_1, x_2, x_3, x_7, x_9$ . A grid search for the misclassification costs  $c$ ,  $\{1, 1.5, 2, 0, \dots, 11, 11.5\}$  was performed to find the optimal separating boundary that properly classifies the majority class and minority class. Figure 5 shows the G-mean values with cost values. As shown in this figure, when the cost value was 1, the G-mean value was recorded as 0. This implies the general SVM model without the cost matrix so that the separating boundary of the model was placed towards the minority class to obtain better classification accuracy. This results in the separating boundary being pushed toward the majority class by the high misclassification cost. This is realized by penalizing the advantages of the majority class for the separating boundary while maximizing the marginal distance.



**Figure 5.** G-mean values vs. cost values.

Increasing misclassification costs can have the effect of moving the separating boundary away from the minority class. According to the misclassification cost, a biased separating boundary can be obtained. Therefore, the G-mean value, where the true positive rate and true negative rate are similar, is an appropriate result. The G-mean value for a misclassification cost of 11 was calculated as the maximum and it also has the highest true positive rate and true negative rate in the grid search range. Consequently, the highest G-mean value was recorded as 0.7188 with a misclassification cost of 11.

This is superior to the best model shown in Table 3. Therefore, this study used the SVM model with a cost of 11 to demonstrate the overall model performance.

Table 4 shows the confusion matrix of the SVM models, with the variables  $x_1, x_2, x_3, x_7, x_9$  and cost of 11 and Table 5 shows the SVM models' performance. The other typical classification models (SVM with the linear and polynomial kernel, multilayer perceptron, decision tree, random forest, logistic regression) were also trained and tested in the same fashion (i.e., their hyperparameters with misclassification cost were searched and selected to obtain the best performance). The results show that the normal state prediction accuracy was 70.93%, the outage prediction accuracy was 72.83%, and the SVM with RBF had an overall prediction accuracy with a G-mean of 0.7188. Whereas, the random forest can classify with a G-mean of 0.7000, which is the most superior model except for SVM. The results indicate that the obtained decision boundary using the cost-sensitive SVM could be used to adequately predict the grid state for future heavy-rain events.

**Table 4.** Confusion matrix of the proposed SVM model.

	Predicted Normal	Predicted Outage
Actual Normal	1945	797
Actual Outage	72	193

**Table 5.** Comparison of the performance between the SVM and other classification models.

	True Positive Rate	True Negative Rate	G-Mean
SVM (RBF)	0.7093	0.7283	0.7188
SVM (Linear)	0.6402	0.7344	0.6857
SVM (Polynomial)	0.6641	0.6971	0.6804
Multilayer perceptron	0.6757	0.7095	0.6924
Decision Tree	0.7101	0.6722	0.6909
Random Forest	0.7028	0.6971	0.7000
Logistic regression	0.6652	0.7220	0.6930

Predicted results are shown based on administrative areas since the branch office areas are close to the administrative areas and it is a convenient method of displaying the spatial distribution of the results. In order to show the spatial distribution of the proposed method, it was used to predict a new heavy-rain event in October 2016. Figure 6 shows the study area. The actual heavy-rain event lasted 15 h; however, for simplicity, the prediction results show only the final nine hours of heavy rain. Figure 7 shows the spatial distribution of the actual and predicted grid state. The black boundary represents the no heavy-rain event for administrative areas and the red and white boundaries represent the outage and normal grid state, respectively. As shown in Figure 7, the proposed method demonstrated adequate prediction performance, however, it does not represent a perfect classifier. This may be due to a significant number of normal grid states mispredicted as outage grid states.

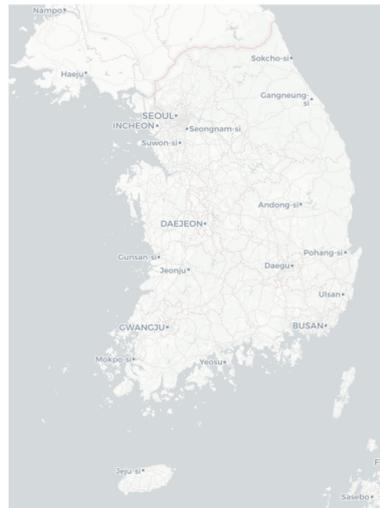


Figure 6. Study area.

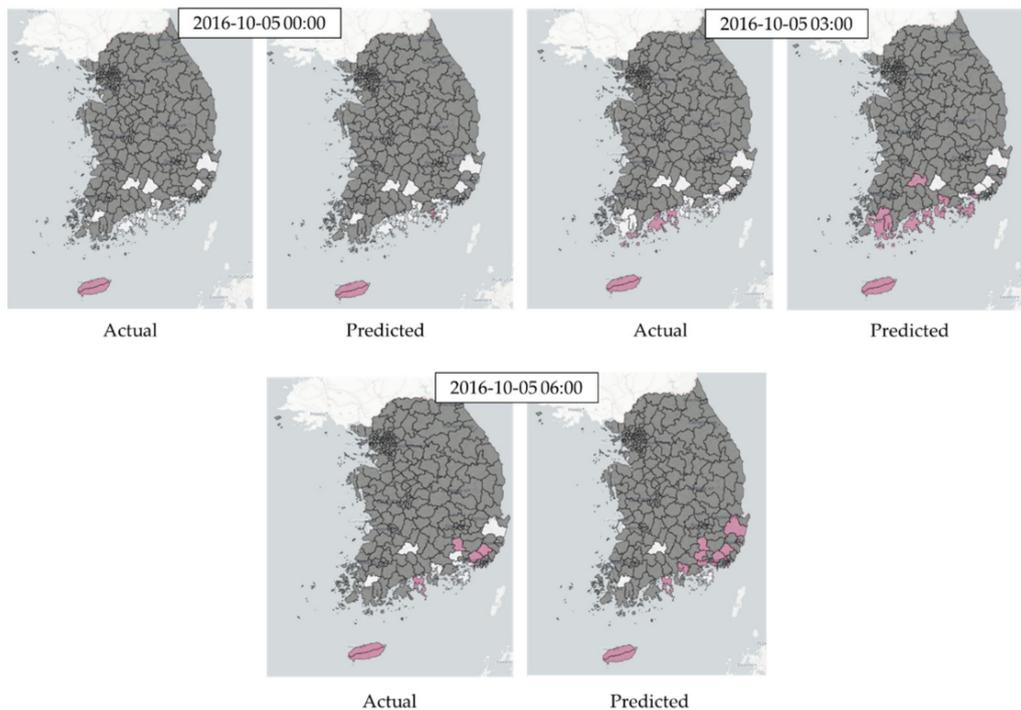


Figure 7. Comparison between the spatial distribution of actual grid state and forecasted grid state for heavy rain in October 2016.

However, the proposed method does predict a grid state similar to the actual grid state and the overall predicted grid states follow a similar spatial pattern to the actual grid states. Hence, it can be concluded that the proposed method can be utilized to mitigate the impact of heavy rain by issuing an alarm for high-risk areas in advance.

### 7. Conclusions

This study presents a machine learning-based method for predicting the power grid state subjected to heavy-rain hazards. This study used two sets of data: Local weather data and power distribution grid damage data, from January 2008 to March 2018. To understand the heavy-rain event, we investigated the heavy-rain-related outage distribution and analyzed the correlated characteristics between weather

and outages. The results show that the effect of wind speed is significant in causing power outages, even under heavy-rain conditions.

Furthermore, this study proposed a grid state prediction method using a cost-sensitive support vector machine (SVM) model. The penalty and kernel parameters of SVM were optimized using Bayesian optimization. The accuracy of the model was improved by applying a cost-sensitive learning algorithm to the SVM model, which was subsequently used to predict the grid state of a region managed by a branch office. The developed models were evaluated using the G-mean values. The proposed method was verified using the actual data of heavy-rain events that occurred in South Korea. The results show that the SVM model with cost-sensitive effects outperforms the general SVM model.

The method proposed in this study is useful in predicting the grid state during heavy-rain events. This study is expected to provide insight into heavy-rain events, as well as damage prediction. Moreover, the proposed method can be extended to other disruptive events since it only relies on the simple weather variable provided by the weather forecast system. Moreover, the condition of power components and weather are changing over time. Thus, future study is needed to develop grid state prediction models to generalize and reflect the changes in the grid conditions over time.

**Author Contributions:** S.O.; writing the original version, methodology, analysis, and interpretation of data, investigation, and conception. J.K.; visualization, analysis, validation. M.C.; sources, data collection. J.J.; conception, supervisory, submitting corresponding and corrections, submit the final approval of the version. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work is financially supported by Korea Ministry of Environment (MOE) as Graduate School specialized in Climate Change. This work was supported by the Korea Institute of Energy Technology Evaluation and Planning (KETEP) granted financial resource from the Ministry of Trade, Industry & Energy, Republic of Korea (No. 20171210000210).

**Acknowledgments:** The authors declare no acknowledgments.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Elsner, J.B.; Kossin, J.P.; Jagger, T.H. The increasing intensity of the strongest tropical cyclones. *Nature* **2008**, *455*, 2–5. [[CrossRef](#)]
2. Emanuel, K. Increasing destructiveness of tropical cyclones over the past 30 years. *Nature* **2005**, *436*, 686–688. [[CrossRef](#)] [[PubMed](#)]
3. Deo, A.A.; Ganer, D.W. Tropical Cyclone Activity over the Indian Ocean in the Warmer Climate. In *Monitoring and Prediction of Tropical Cyclones in the Indian Ocean and Climate Change*; Mohanty, U.C., Mohapatra, M., Singh, O.P., Bandyopadhyay, B.K., Rathore, L.S., Eds.; Springer: Dordrecht, The Netherlands, 2014; pp. 72–73.
4. Venkata, S.S.M.; Hatzigryriou, N. Grid Resilience: Elasticity Is Needed When Facing Catastrophes. *IEEE Power Energy Mag.* **2015**, *13*, 16–23. [[CrossRef](#)]
5. Hanif, F.; Widiputra, V.; Jung, J. State-of-the-art review on power grid resilience to extreme weather events: Definitions, frameworks, quantitative assessment methodologies, and enhancement strategies. *Appl. Energy* **2019**, *239*, 1049–1065.
6. Kezunovic, M.; Dobson, I.; Dong, Y. Impact of Extreme Weather on Power System Blackouts and Forced Outages: New Challenges. In *Proceedings of the 7th Balkan Power Conference, Šibenik, Croatia, 10–12 September 2008*; pp. 1–5.
7. Jufri, F.H.; Kim, J.S.; Jung, J. Analysis of determinants of the impact and the grid capability to evaluate and improve grid resilience from extreme weather event. *Energies* **2017**, *10*, 1779. [[CrossRef](#)]
8. Thukaram, D.; Khincha, H.; Vijaynarasimha, H. Artificial neural network and support vector machine approach for locating faults in radial distribution system. *IEEE Trans. Power Deliv.* **2005**, *20*, 710–721. [[CrossRef](#)]
9. Jufri, F.Z.; Oh, S.; Jung, J. Development of Photovoltaic abnormal condition detection system using combined regression and Support Vector Machine. *Energy* **2019**, *176*, 457–467. [[CrossRef](#)]
10. Guikema, S.D. Natural disaster risk analysis for critical infrastructure systems: An approach based on statistical learning theory. *Reliab. Eng. Syst. Saf.* **2009**, *94*, 855–860. [[CrossRef](#)]

11. Chen, Y.; Chu, Y.; Wu, Y.; Bao, Y.; Xu, P.; Li, W.; Ni, L.; Wang, K. Short-term electrical load forecasting using the Support Vector Regression (SVR) model to calculate the demand response baseline for office buildings. *Appl. Energy* **2017**, *195*, 659–670. [[CrossRef](#)]
12. Lu, C.; Wu, H.; Vemuri, S. Neural network based short term load forecasting. *IEEE Trans. Power Syst.* **1993**, *8*, 336–342. [[CrossRef](#)]
13. Bhaskar, K.; Singh, S.N. AWNN-Assisted Wind Power Forecasting Using Feed-Forward Neural Network. *IEEE Trans. Sustain. Energy* **2012**, *3*, 306–315. [[CrossRef](#)]
14. Jufri, F.Z.; Jung, J. Photovoltaic Generation Forecasting Using Artificial Neural Networks Model with Input Variables and Model Parameters Selection Algorithm in Korea. *Int. J. Mach. Learn. Comput.* **2017**, *7*, 156–161. [[CrossRef](#)]
15. Hanif, F.; Seongmun, J.; Jaesung, O. Day-Ahead System Marginal Price Forecasting Using Artificial Neural Network and Similar Days Information. *J. Electr. Eng. Technol.* **2019**, *14*, 561–568.
16. Oh, S.; Jung, J.; Jufri, F.H.; Choi, M. Predicting the Damage of Distribution System by Strong Wind. In Proceedings of the IEEE Innovative Smart Grid Technologies—Asia (ISGT Asia), Chengdu, China, 21–24 May 2019; pp. 3970–3974.
17. Jufri, F.H.; Oh, S.; Jung, J.; Choi, M.A. Method to Forecast Storm-Caused Distribution Grid Damages Using Cost-Sensitive Regression Algorithm. In Proceedings of the IEEE Innovative Smart Grid Technologies—Asia (ISGT Asia), Chengdu, China, 21–24 May 2019; pp. 3986–3990.
18. Nateghi, R.; Guikema, S.D.; Quiring, M. Comparison and Validation of Statistical Methods for Predicting Power Outage Durations in the Event of Hurricanes. *Risk Anal.* **2011**, *31*, 1897–1906. [[CrossRef](#)] [[PubMed](#)]
19. Liu, H.; Davidson, R.A.; Rosowsky, D.V.; Stedinger, J.R. Negative Binomial Regression of Electric Power Outages in Hurricanes. *J. Infrastruct. Syst.* **2005**, *11*, 258–267. [[CrossRef](#)]
20. Liu, H.; Davidson, R.A.; Apanasovich, T.V. Spatial generalized linear mixed models of electric power outages due to hurricanes and ice storms. *Reliab. Eng. Syst. Saf.* **2008**, *93*, 875–890. [[CrossRef](#)]
21. Mensah, A.F.; Duenas-Osorio, L. Outage predictions of electric power systems under Hurricane winds by Bayesian networks. In Proceedings of the 2014 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS), Durham, UK, 7–10 July 2014; pp. 1–6.
22. Singhee, A.; Wang, H. Probabilistic forecasts of service outage counts from severe weather in a distribution grid. In Proceedings of the IEEE Power & Energy Society General Meeting, Chicago, IL, USA, 16–20 July 2017; pp. 1–5.
23. Kezunovic, M.; Obradovic, Z.; Dokic, T.; Zhang, B.; Stojanovic, J.; Dehghanian, P.; Chen, P.-C. Predicting Spatiotemporal Impacts of Weather on Power Systems Using Big Data Science BT. In *Data Science and Big Data: An Environment of Computational Intelligence*; Pedrycz, W., Chen, S.-M., Eds.; Springer International Publishing: Cham, Switzerland, 2017; Volume 24, pp. 265–299.
24. Ciapessoni, E.; Cirio, D.; Pitto, A.; Sforza, M. A probabilistic risk-based security assessment tool allowing contingency forecasting. In Proceedings of the 2018 IEEE International Conference on Probabilistic Methods Applied to Power Systems (PMAPS), Boise, ID, USA, 24–28 June 2018; pp. 1–6.
25. Han, S.R.; Guikema, S.D.; Quiring, S.M. Improving the predictive accuracy of hurricane power outage forecasts using generalized additive models. *Risk Anal.* **2009**, *29*, 1443–1453. [[CrossRef](#)]
26. Han, S.R.; Guikema, S.D.; Quiring, S.M.; Lee, K.H.; Rosowsky, D.; Davidson, R.A. Estimating the spatial distribution of power outages during hurricanes in the Gulf coast region. *Reliab. Eng. Syst. Saf.* **2009**, *94*, 199–210. [[CrossRef](#)]
27. Guikema, S.D.; Nateghi, R.; Quiring, S.M.; Staid, A.; Reilly, A.C.; Gao, M. Predicting Hurricane Power Outages to Support Storm Response Planning. *IEEE Access* **2014**, *2*, 1364–1373. [[CrossRef](#)]
28. Eskandarpour, R.; Khodaei, A. Machine Learning Based Power Grid Outage Prediction in Response to Extreme Events. *IEEE Trans. Power Syst.* **2017**, *32*, 3315–3316. [[CrossRef](#)]
29. Eskandarpour, R.; Khodaei, A.; Arab, A. Improving power grid resilience through predictive outage estimation. In Proceedings of the North American Power Symposium, Morgantown, WV, USA, 17–19 September 2017; pp. 1–5.
30. Korea Meteorological Administration. Available online: [http://www.kma.go.kr/eng/weather/forecast/standard\\_warning\\_info.jsp](http://www.kma.go.kr/eng/weather/forecast/standard_warning_info.jsp) (accessed on 06 July 2020).

31. EOP (Executive Office of the President). *Economic Benefits of Increasing Electric Grid Resilience to Weather Outages*. Available online: <https://www.energy.gov/downloads/economic-benefits-increasing-electric-grid-resilience-weather-outages> (accessed on 06 July 2020).
32. He, H.; Garcia, E. Learning from imbalanced data. *IEEE Trans. Knowl. Data Eng.* **2009**, *21*, 1263–1284.
33. McCarthy, K.; Zabar, B.; Weiss, G.M. Does Cost-Sensitive Learning Beat Sampling for Classifying Rare Classes? In Proceedings of the 1st International Workshop on Utility-Based Data Mining, Chicago, IL, USA, 21 August 2005; pp. 69–77.
34. Burges, C.J.C. *A Tutorial on Support. Vector Machines for Pattern Recognition*; Kluwer Academic Publishers: Berlin, Germany, 1998; Volume 167, pp. 121–167.
35. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed.; Springer: New York, NY, USA, 2009.
36. Platt, J. Sequential minimal optimization: A fast algorithm for training support vector machines. In *Advances in Kernel Method: Support Vector Learning*; Scholkopf, B., Smola, A., Eds.; MIT Press: Cambridge, MA, USA, 1998; pp. 185–208.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).