



Probabilistic load curtailment estimation using posterior probability model and twin support vector machine

Rozhin ESKANDARPOUR¹, Amin KHODAEI¹



Abstract Estimating the potential load curtailments as a result of hurricane is of great significance in improving the emergency response and recovery of power grid. This paper proposes a three-step sequential method in identifying such load curtailments prior to hurricane. In the first step, a twin support vector machine (TWSVM) model is trained on path/intensity information of previous hurricanes to enable a deterministic outage state assessment of the grid components in response to upcoming events. The TWSVM model is specifically used as it is suitable for handling imbalanced datasets. In the second step, a posterior probability sigmoid model is trained on the obtained results to convert the deterministic results into probabilistic outage states. These outage states enable the formation of probability-weighted contingency scenarios. Finally, the obtained component outages are integrated into a load curtailment estimation model to determine the expected potential load curtailments in the grid. The simulation results, tested on the standard IEEE 118-bus system and based on synthetic datasets, illustrate the high accuracy performance of the proposed method.

Keywords Hurricanes, Machine learning, Power system resilience, Predictive analytics

1 Introduction

The catastrophic consequences of hurricanes, including lost businesses, impeded emergency services, lost output and wages, delayed production, and damaged infrastructure, have continuously and consistently called for investments to enhance power system resilience [1]. Predicting the potential impact of hurricanes on the power system operation is a fundamental step in resource management and scheduling in response to these catastrophic events. This is a challenging task as several factors are involved in hurricane modeling and the behavior and aftermath of the events may not be easily achievable. Historical data on previous hurricanes become invaluable under this circumstance to model and predict the impact of the hurricane on the power system. Given the amount of data that exists on previous hurricanes and the complexity of the system, machine learning can be a viable tool to tackle this problem. Machine learning is an application of artificial intelligence (AI) that provides the system with the ability to learn from historical data and make predictions without being explicitly programmed.

The notion of resilience is originally initiated by Holling [2] in ecological systems. Holling described the concept of resilience as the ability of a system to absorb an external shock and return to the normal condition in the least possible time. The National Academy of Sciences defines resilience as “the ability to plan and prepare for, absorb, recover from, and adapt to adverse events” [3]. In [4], the purpose of resilience engineering is described as predicting the different risks of changes before failures happen. In

CrossCheck Date: 15 January 2019

Received: 16 November 2018 / Accepted: 15 January 2019 / Published online: 21 March 2019
© The Author(s) 2019

✉ Amin KHODAEI
amin.khodaei@du.edu

Rozhin ESKANDARPOUR
rozhin.eskandarpour@du.edu

¹ Department of Electrical and Computer Engineering,
University of Denver, Denver, CO 80208, USA



particular, in power systems, the resilience is defined as the system's capability to sustain external shocks such as hurricanes while ensuring the least possible load supply interruptions and returning to normal operation promptly [5]. Having a precise prediction of the potential impacts of an upcoming hurricane plays a vital role in improving the power system resilience by helping identify the most efficient resource allocation [6].

Resource allocation before and after a hurricane is a well-studied topic in power systems. In [7], a proactive resource allocation model is proposed to repair and recover power system infrastructure located in a hurricane-impacted region, attempting to develop a decision-making tool which ensures the least potential damages in an efficient manner. In [8, 9], a proactive recovery framework of power system components is presented based on a stochastic model for operating the components prior to the event, followed by a deterministic recovery model to manage the available resources after the event. In [10], an optimal restoration model is proposed to minimize the economic loss due to power supply interruptions during the post-disaster phase. In [11], a decision-making model is introduced based on unit commitment constraints and system configuration. The objective of the proposed model is to determine the optimal repair schedule after an extreme event and during the restoration phase.

Pre-hurricane scheduling specifically plays an important role in improving system resilience. A resilience-constrained unit commitment (RCUC) model is proposed in [12] which ensures a resilient supply of loads even in case of multiple component outages. In [5], an event-driven security-constrained unit commitment (E-SCUC) model is presented which finds an optimal schedule of available resources in case of multiple component outages due to an extreme event. In [13], a probabilistic approach is introduced to estimate the effect of standby units, interruptible loads, postponable outages and margin time on the well-being of a generator. The well-being of a system is further quantified in terms of health and margin, as well as several predefined system risk indices. In [14], a two-stage stochastic programming approach to obtain the optimal scheduling of a resilient microgrid is proposed which mitigates damaging impacts of electricity interruptions by effectively exploiting the microgrid capabilities. In [15], the role of networked microgrids as distributed systems is investigated for enhancing the power system resilience in order to respond effectively in emergency conditions. In [16], a framework for analyzing the power system resilience is proposed to illustrate how microgrids can help enhance its resilience in extreme conditions. A machine learning method is proposed in [17] to assist with the prioritization of maintenance and repair work by forecasting the risk of failures for components, which is considered

beneficial for power companies. The main machine learning based forecasting techniques for power system applications are reviewed in [18]. Available forecasting techniques have been discussed with a focus on electricity load and price forecasting in electricity markets in Australia and the U.S.. The neutral online visualization-aided autonomic (NOVA) evaluation framework is further provided for evaluating machine learning algorithms for power systems preventive maintenance to improve reliability [19].

In many engineering problems, including the hurricane modeling problem, a closed-form problem formulation may not be readily available. Machine learning is capable of making forecast through data processing algorithms which can categorize the data for supervised learning (classification), unsupervised learning (clustering) and regression modeling (forecast the output of the system according to its prior actions and historical data) [20]. Machine learning approaches have been used for several applications in power and energy sectors [21]. Security assessment is one of the earliest examples of using machine learning in power systems [22, 23]. Forecasting emerges as another valuable application of machine learning, mainly using decision tree induction and nearest neighbor classifiers [24]. In short-term load forecasting, multitudes of artificial neural network (ANN) models are proposed [25–27]. Wind power is also forecasted by machine learning [28]. Other cases of machine learning applications include risk analysis in power systems using regression models and ANN [29], distribution fault detection using ANN and support vector machine (SVM) [30], and power outage duration estimation using regression trees [31].

In many of the related works on hurricane modeling, the impact of the hurricane on the power system is the input to the model or determined by a stochastic model. Machine learning, however, is recognized as an efficient method in predictive analytics and data analysis to identify the likelihood of future outcomes based on historical data [21]. In particular, SVM is a popular machine learning method for data classification (supervised learning) which is developed on the basis of statistical learning theory and structural risk minimization [32, 33]. SVM has numerous advantages such as providing a global solution for data classification as well as great generalization capability. The achieved results in several studies illustrate SVM as one of the most accurate methods in several applications such as generation forecasting [34, 35], load forecasting [36], fault detection [37], power quality disturbance monitoring [38], and transient stability analysis [39]. SVM has also shown a superior performance in predicting possible outages of power system components in response to extreme events [40]. In [41], a three-dimensional SVM is proposed to predict the outage of power system components in response

to an extreme event, where its accuracy–uncertainty tradeoff is leveraged to achieve more precise results.

Despite the good performance of SVM in several applications, the performance of SVM drops significantly when faced with imbalanced datasets, for example when the number of negative instances far outnumbers the positive instances, or vice versa [42]. Twin support vector machine (TWSVM) is the answer to this, as an efficient machine learning approach which is suitable for complex classification problems. TWSVM classifies the patterns of two classes by using two non-parallel hyperplanes [43]. Since two hyperplanes are defined as representatives of each class, TWSVM can handle imbalanced datasets much better than the traditional SVM [44].

In this paper, a TWSVM classification method is trained to find the operational state of each component by considering the path and the intensity of the hurricane, as well as the distance of each component from the center of the hurricane. A posterior probability model is consequently applied to the output of the TWSVM model to estimate the outage probability of each component. Having an accurate estimation of probable outages plays a vital role in responding to an upcoming hurricane.

Unlike the existing work on outage prediction and extended outage consideration in security-constrained unit commitment (SCUC), including the previous work of authors in [5, 6], this paper considers the probability of outage obtained by a machine learning approach in scheduling. TWSVM is chosen for its performance in complex intertwined classification problems and when dealing with imbalanced datasets. This can be potentially problematic since the data of past hurricanes are imbalanced, i.e., the number of non-operational components is far less than the number of operational components. The merit behind proposed probabilistic load curtailment estimation model is that it considers all contingency scenarios with their probability and hence the most probable scenario or the scenarios with most load curtailment can be recognized. The predicted outage and estimated outage probability can be useful for electric utilities to assess their risk and allocate necessary resources and repair crews to prepare for and recover from hurricanes in a considerably shorter time-frame.

The rest of the paper is organized as follows: Section 2 presents the model outline and formulation of the proposed machine learning method for outage prediction. Section 3 presents simulation results on a test system, and Section 4 concludes the paper.

2 Proposed model

The goal of this section is to determine the probable load curtailments in a power system as a result of hurricane-caused component outages. The considered components include, but are not limited to, transmission lines, generation units, and substations. The problem is solved in three consecutive steps. In first step, a TWSVM model [44, 45] is trained on historical outage data to help classify the operational state of components after the hurricane.

The speed of the hurricane and the distance of each component from the center of the hurricane are used to predict the probability of outage for each component. The output of the TWSVM model will be a list of 0/1 values, showing whether each component is operational or on outage, however it provides no information on the outage probability. To estimate the outage probability for each component, a posterior probability sigmoid model [46] is applied in second step to the output of the first step. The category and the path of the upcoming hurricane in this step are obtained from weather forecasting agencies. In third step, the obtained component outages and their associated probabilities are integrated into a probabilistic load curtailment estimation model to estimate the nodal load curtailments and thus help identify the areas that will potentially be impacted by the hurricane.

2.1 TWSVM

The SVM method has numerous advantages including the ability to provide a global solution for data classification. It generates a unique global hyperplane by solving a quadratic programming problem (QPP) to separate the data samples of different classes rather than local boundaries as compared to other existing data classification approaches. Due to its performance, SVM is one of the most widely-used classification techniques in data mining. One of the main challenges with the traditional SVM, however, is that it solves only one QPP problem to classify the data, which may not be suitable in cases of imbalanced data.

Although SVM often produces effective solutions for balanced datasets, it is sensitive to imbalance in datasets and produces suboptimal results [47]. In other words, the separating hyperplane of an SVM model trained with an imbalanced dataset can be skewed towards the minority class [48], and hence the performance of that model is degraded with respect to the minority class. Several approaches in literature have been proposed to improve the SVM performance when dealing with imbalanced dataset classification [47]. These approaches can be categorized as data processing approaches (such as resampling methods [49] and ensemble learning methods [50]), algorithmic



approaches (such as different error cost [48] or z-SVM [51]), and hybrid approaches (such as hybrid kernel machine ensemble [52]). Despite the performance improvement of these approaches, the suboptimality of the soft-margin is an inherited problem of SVM and majority of these approaches require an expert understanding of data shape and empirical parameter tuning, e.g., setting a proper weight for each class, or finding best ensemble size.

A viable alternative to SVM is TWSVM, as a machine learning approach suitable for complex intertwined classification problems, which classifies the patterns of two classes by using two non-parallel hyperplanes [43]. The biggest advantage of TWSVM, in addition to the training speed, is its ability to handle imbalanced datasets [44]. This is because each class has its own representative hyperplane instead of one hyperplane separating two classes from each other, and therefore TWSVM can classify underrepresented classes better than traditional SVM, especially when the classes are intertwined. Since TWSVM classifies the data using two hyperplanes, it solves a pair of QPPs instead of a single complex QPP as in traditional SVM. Comparing to a traditional SVM over benchmark datasets, TWSVM has shown comparable performance while being approximately four times faster [44, 45]. TWSVM has shown improvement in several practical applications such as classification of biomedical data [53], gesture classification [54] speaker recognition (i.e., personal identity from the speech signal) [55], and image analysis [56], to name a few. Figure 1 illustrates a traditional linear classifier SVM and TWSVM in separating two classes. As shown, traditional SVM does not take the data skewness into account and the separating hyperplane is the one that represents the largest margin between two classes.

The goal of TWSVM in a binary classification problem is to construct two non-parallel planes for each class such that each hyperplane is closer to the data samples of its representative class while distant from the samples of the other class [43]. The distances between the samples and both non-parallel hyperplanes are compared to determine the category of each sample.

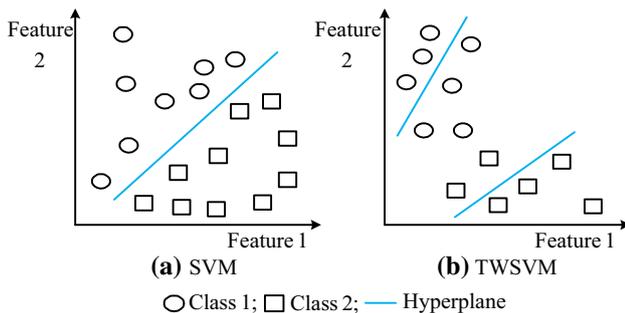


Fig. 1 SVM and TWSVM for imbalanced dataset in two-dimensional feature space

Consider a binary classification problem that classifies m_1 training samples belonging to positive class and m_2 training samples belonging to negative class in an n -dimensional real space \mathbf{R}^n , where $m_1+m_2 = m$. Let matrices A_1 and A_2 represent the training samples of the positive and negative classes respectively. Since a linear TWSVM seeks two non-parallel hyperplanes, two hyperplanes $h_1(x)$ and $h_2(x)$ are defined as:

$$h_i(x) = w_i^T x + d_i = 0 \quad \forall i \in \{1, 2\} \tag{1}$$

where w_i is the normal vector to the hyperplane representing training examples of class i ; and d_i is the bias vector of the separating hyperplanes representing class i . $|d_i|/||w_i||$ is the perpendicular distance from the hyperplanes to the origin, $||\cdot||$ denotes Euclidean distance. To find hyperplanes $h_1(x)$ and $h_2(x)$, such that $h_1(x)$ is closest to the training samples of the positive class and far from the training samples of the negative class, and $h_2(x)$ is closest to the training samples of the negative class and far from the training samples of the positive class, the following QPP is solved for each class:

$$\min_{w_i, d_i, \xi_i} \left(\frac{1}{2} \|A_i w_i + e_i d_i\|^2 + c_i e_j^T \xi_i \right) \tag{2}$$

s.t.

$$-\rho_i (A_j w_i + e_j d_i) + \xi_i \geq e_j \quad \xi_i \geq 0, i \neq j \tag{3}$$

where $c_i > 0$ is the regularization term to control overfitting of class i ; e_i is a vector of ones of appropriate dimension; ξ_i is slack variable of class i ; and ρ_i is the coefficient of each class where $\rho_1=1$ for the positive class and $\rho_2 = -1$ for the negative class. TWSVM solves two QPPs problem (2) and (3) separately for each class. If sample sizes of both classes are approximately equal to $m/2$, the complexity of solving these two QPPs in TWSVM will be $O(2 \times (m/2)^3)$. Comparing with the standard SVM with computational complexity of $O(m^3)$ which solves one QPP problem for both classes at the same time, TWSVM is approximately four times faster [43]. The objective function seeks the distance from the sample to the hyperplane by the square distances (L_2 -norm), and minimizes the distance to ensure the hyperplane is as close as possible to the samples of its own class. The sample x is assigned to class i if:

$$\frac{|w_i^T x + d_i|}{\|w_j\|} + \frac{\|w_j\|}{\|w_i\|} < 1 \quad \forall i \in \{1, 2\}, \forall j \in \{1, 2\}, i \neq j \tag{4}$$

Similar to SVM, kernel method [32] can be applied to TWSVM. The idea of a kernel method (or as sometimes called kernel trick) is to map the input feature vector into a

higher-dimension space where the classes are linearly separable. To apply kernel to TWSVM, the QPP problem of (2) and (3) is formulated as:

$$\min_{\mathbf{w}_i, \mathbf{d}_i, \xi_i} \left(\frac{1}{2} \|K(\mathbf{A}_i, \mathbf{B}^T) \mathbf{w}_i + \mathbf{e}_i \mathbf{d}_i\|^2 + c_i \mathbf{e}_j^T \xi_i \right) \quad (5)$$

s.t.

$$-\rho_i(K(\mathbf{A}_j, \mathbf{B}^T) \mathbf{w}_i + \mathbf{e}_j \mathbf{d}_i) + \xi_i \geq \mathbf{e}_j \quad \xi_i \geq 0, i \neq j \quad (6)$$

where $\mathbf{B}=[\mathbf{A}_1, \mathbf{A}_2]^T$ and K is the kernel function. Finding a proper value of penalty parameter c and the best kernel depends on the shape of classes, which are often found via a search method to minimize the error on the test set.

2.2 Posterior probability estimation

To determine the likelihood of a sample belonging to a specific class, two normalized distances, to each hyperplane \mathbf{h}_i , are defined as:

$$D_i(\mathbf{x}) = \frac{|\mathbf{w}_i^T \mathbf{x} + \mathbf{d}_i|}{\|\mathbf{w}_i\|} \quad \forall i \in \{1, 2\} \quad (7)$$

Given the distance between two representative hyperplanes \mathbf{h}_1 and \mathbf{h}_2 , two new relative distances can be defined as:

$$D_+(\mathbf{x}) = D_1(\mathbf{x}) + D_2(\mathbf{x}) = 0 \quad (8)$$

$$D_-(\mathbf{x}) = D_1(\mathbf{x}) - D_2(\mathbf{x}) = 0 \quad (9)$$

Intuitively, the probability of a sample x belonging to a certain class depends on its relative distance to the positive class D_+ and the negative class D_- . Two relevant quantities $D_{\min}(\mathbf{x})$ and $D_{\max}(\mathbf{x})$ are then defined by:

$$D_{\min}(\mathbf{x}) = \min\{D_+(\mathbf{x}), D_-(\mathbf{x})\} \quad (10)$$

$$D_{\max}(\mathbf{x}) = \max\{D_+(\mathbf{x}), D_-(\mathbf{x})\} \quad (11)$$

Figure 2 shows a sample x and its corresponding relative distances $D_+(\mathbf{x})$ and $D_-(\mathbf{x})$.

As it is shown, the quantities $D_{\min}(\mathbf{x})$ and $D_{\max}(\mathbf{x})$ are the factors influencing the probability of belonging to the positive class. In other words, the probability of belonging to the positive class increases when either $D_{\min}(\mathbf{x})$ or $D_{\min}(\mathbf{x})/D_{\max}(\mathbf{x})$ becomes larger. Hence, a score function $f(\mathbf{x})$ can be define as:

$$f(\mathbf{x}) = \begin{cases} D_{\min}(\mathbf{x}) \left(\frac{D_{\min}(\mathbf{x})}{D_{\max}(\mathbf{x})} \right)^\lambda & D_1(\mathbf{x}) > D_2(\mathbf{x}) \\ 0 & D_1(\mathbf{x}) = D_2(\mathbf{x}) \\ -D_{\min}(\mathbf{x}) \left(\frac{D_{\min}(\mathbf{x})}{D_{\max}(\mathbf{x})} \right)^\lambda & D_1(\mathbf{x}) < D_2(\mathbf{x}) \end{cases} \quad (12)$$

If $D_1(\mathbf{x}) > D_2(\mathbf{x})$, then the sample belongs to the positive class, otherwise to the negative class. If $D_{\min}(\mathbf{x})$ is small

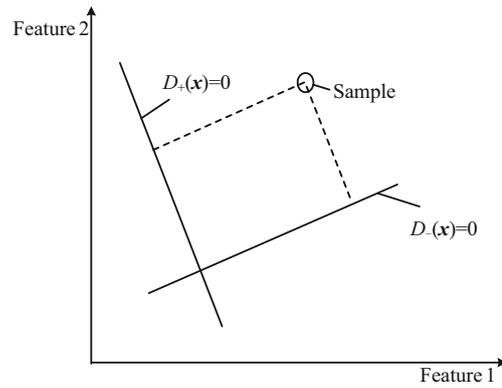


Fig. 2 An example indicating meaning of relative distances of sample x to positive and negative separating hyperplanes in a two-dimensional feature space

and $D_{\max}(\mathbf{x})$ is large, it means that the sample is very close to one of the planes and far away from the other. Hence, the probability is large, i.e., $f(\mathbf{x})$ becomes a very large positive number for the positive class and a very large negative number for the negative class. If $D_{\min}(\mathbf{x}) \approx D_{\max}(\mathbf{x})$, then it means the sample is relatively in the same distance between these classes and the $f(\mathbf{x})$ is small. Constant λ is the weight parameter. This parameter can be determined on a validation set. The data is split into three subsets, training, validation and test. The training set is used to find separating hyperplanes. Then different values of λ in the score functions $f(\mathbf{x})$ will be evaluated on the validation set and the best parameter will be tested on the test subset.

The above formulation can be easily extended to non-linear TWSVM by considering the kernel-generated surfaces instead of the hyperplanes as:

$$D_i(\mathbf{x}) = \frac{|\mathbf{w}_i^T K(\mathbf{x}, \mathbf{B}^T) + \mathbf{d}_i|}{\|\mathbf{w}_i^T K(\mathbf{B}, \mathbf{B}^T) \mathbf{w}_i^T\|} \quad \forall i \in \{1, 2\} \quad (13)$$

Since $D_{\min}(\mathbf{x})$ and $D_{\max}(\mathbf{x})$ can be any arbitrary value, the range of the score function $f(\mathbf{x})$ is $(-\infty, +\infty)$. Platt scaling or Platt calibration is a way of transforming the score of a classification model into a probability distribution over classes [57]. Platt scaling finds the parameters of a sigmoid function which converts the scoring output of $(-\infty, +\infty)$ to a probability of $[0, 1]$. It has been shown that Platt method yields probability estimates that are at least as accurate as ones obtained by training a SVM, while being expedient [58]. Similar to the continuous output in an SVM, the following posterior probability function is constructed over the values of score function $f(\mathbf{x})$ as:

$$P(y = +1|f(\mathbf{x})) = \frac{1}{1 + e^{2f(\mathbf{x})+\beta}} \quad (14)$$

$$P(y = -1|f(\mathbf{x})) = 1 - P(y = +1|f(\mathbf{x})) \tag{15}$$

where α and β are the scaling weights of the sigmoid function calculated using the maximum likelihood estimation (i.e., Platt scaling) [57], by minimizing the following function:

$$\min_{\alpha, \beta} \left\{ - \sum_{k=1}^m [t_k \lg p_k + (1 - t_k) \lg(1 - p_k)] \right\} \tag{16}$$

s.t.

$$p_k = \Pr(y_k = +1|f(\mathbf{x}_k)) \tag{17}$$

$$t_k = \begin{cases} \frac{m_1 + 1}{m_1 + 2} & y_k = +1, k = 1, 2, \dots, m \\ \frac{1}{m_2 + 2} & y_k = -1, k = 1, 2, \dots, m \end{cases} \tag{18}$$

where t_k is the target probability of a particular sample of \mathbf{x}_k and p_k is the predicted probability of that sample.

2.3 Evaluation criteria

1) Evaluation of classifier

To evaluate the performance of the classifier, a cross-fold validation is used. The cross-fold validation splits the data into q subsets, in which the classifier is trained on $q-1$ subsets and evaluated on the subset that is left in the training. This process is performed q times (such that the classifier is evaluated on all samples). The final classification accuracy is the average of classification accuracies on all folds. Reporting the general accuracy of prediction cannot be sufficient as the number of samples may not be balanced in the test set. The F_1 -score is a common and reliable measure of classification performance [20] defined as:

$$F_1 = \frac{2PR}{P + R} \tag{19}$$

where P is the number of correct positive results divided by the number of all positive results returned by the classifier; and R is the number of correct positive results divided by the number of all relevant samples. In case of outage estimation, P is defined as the ratio of number of correctly predicted outages to total number of predicted outages, and R is defined as the ratio of number of correctly predicted outages to total number of actual outages.

A higher value of the F_1 -score, which is a number between 0 and 1, indicates a better classification and justifies the viable performance of the existing decision boundary.

2) Evaluation of posterior probability estimation

A common way to determine how well a posterior probability estimator model fits the data is the area under receiver operating characteristic (ROC) curve [20]. A ROC curve is a graph showing the performance of a classification model at all classification thresholds. The ROC curve is created by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings. In this paper, since the goal is to estimate outage probability, the outage state is considered as positive and the operational state is considered as negative class. The TPR is the number of correctly predicted samples in outage state divided by the total number of samples in outage state, and FPR is the number of incorrectly predicted samples in operational state divided by the total number of samples in operational state.

The area under the ROC curve (AU-ROC) measures the entire two-dimensional area underneath the entire ROC curve as:

$$A_{AU-ROC} = \int_{-\infty}^{+\infty} TPR(\tau) \cdot FPR(\tau) d\tau \tag{20}$$

where τ is a threshold indicating that an instance is classified as positive class if the posterior probability is greater than τ , and negative otherwise. AU-ROC provides an aggregate measure of performance across all possible classification thresholds. It is equal to the probability that a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one [20].

2.4 Probabilistic load curtailment estimation

The objective function of the probabilistic load curtailment estimation problem is defined as:

$$\min \left(\sum_t \sum_g F_g(P_{gt0}, I_{gt}) + \sum_t \sum_s \sum_b \pi_s v_b L_{C,bts} \right) \tag{21}$$

where π_s is the probability of each hurricane scenario, $\sum \pi_s = 1$; $F_g(\cdot)$ is the operation cost function, which includes the generation cost and startup/shutdown costs; P_{gt0} is the real power generation of unit g at time t in scenario zero (i.e., normal operation); I_{gt} is the commitment state of unit g at time t ; v_b is the value of lost load at bus b ; and $L_{C,bts}$ is the amount of nodal load curtailment at bus b at time t in scenario s . The value of lost load is defined as the average cost that each type of customer, i.e., residential, commercial, or industrial, is willing to pay in order to avoid power supply interruptions. Assuming $U_{X,gts}$ and $U_{Y,lts}$ as outage states for generation

units and transmission lines, respectively, the proposed objective function is subject to the following operational constraints:

$$\sum_{g \in B} P_{gts} + \sum_{l \in B} P_{L,lts} + L_{C,bts} = D_{bt} \quad \forall b, \forall t, \forall s \quad (22)$$

$$P_g^{\min} I_{gt} U_{X,gts} \leq P_{gts} \leq P_g^{\max} I_{gt} U_{X,gts} \quad \forall g, \forall t, \forall s \quad (23)$$

$$P_{gts} - P_{g,t-1,s} \leq U_{R,g} \quad \forall g, \forall t, \forall s \quad (24)$$

$$P_{g,t-1,s} - P_{gts} \leq D_{R,g} \quad \forall g, \forall t, \forall s \quad (25)$$

$$T_{gt}^{\text{on}} \geq U_{T,g} (I_{gt} - I_{g,t-1}) \quad \forall g, \forall t \quad (26)$$

$$T_{gt}^{\text{off}} \geq D_{T,g} (I_{g,t-1} - I_{gt}) \quad \forall g, \forall t \quad (27)$$

$$0 \leq L_{C,bts} \leq D_{bt} \quad \forall b, \forall t, \forall s \quad (28)$$

$$-P_{L,l}^{\max} U_{Y,lts} \leq P_{L,lts} \leq P_{L,l}^{\max} U_{Y,lts} \quad \forall l, \forall t, \forall s \quad (29)$$

$$\left| P_{L,lts} - \frac{\sum_b a_{lb} \theta_{bts}}{X_l} \right| \leq M(1 - U_{Y,lts}) \quad \forall l, \forall t, \forall s \quad (30)$$

where P_{gts} is the real power generation of unit g at time t in scenario s ; $P_{L,lts}$ is the real power flow of line l at time t in scenario s ; D_{bt} is the load at bus b at time t ; P_g^{\min} and P_g^{\max} are respectively the minimum and maximum generation capacity of unit g ; $U_{R,g}$ and $D_{R,g}$ are respectively ramp up and ramp down rates of unit g ; T_{gt}^{on} and T_{gt}^{off} are respectively the number of successive on and off hours of unit g at time t ; $U_{T,g}$ and $D_{T,g}$ are respectively the minimum up time and down time of unit g ; $P_{L,l}^{\max}$ is the maximum power flow of line l ; a_{lb} is the element of line l and bus b in line-bus incidence matrix; θ_{bts} is the phase angle of bus b at time t in scenario s ; X_l is the reactance of line l ; and M is a large positive constant.

Load balance equation (22) ensures that the total injected power to each bus from generation units and line flows is equal to the total load at that bus. Load curtailment variable ($L_{C,bts}$) is further added to the load balance equation to ensure a feasible solution when there is not sufficient generation to supply loads (due to component outages). Generation unit output power is limited by its capacity limit and will be set to zero depending on its commitment and outage states (23). Generation units are further subject to prevailing technical constraints including ramp up and down rate limits (24), (25), and minimum up and down time limits (26), (27). The load curtailment at each bus is constrained by the total load on that bus (28). Transmission line capacity limits and power flow constraints are modeled by (29) and (30), respectively, in which the outage state is included to model the line outages in contingency scenarios. Note that (21)-(30) is effectively a SCUC problem with weighted scenarios and simultaneous component outages.

3 Numerical simulations

The standard IEEE 118-bus test system is used for testing the proposed model, by assuming that a hurricane is predicted to pass through the system. The system characteristics, including generation, line, and load data, can be found in [59].

3.1 TWSVM performance

As historical data for the past hurricanes at component level are limited, 550 samples are synthetically generated (500 samples of component in operational state and 50 samples in outage state) following a normal distribution function with a small Gaussian noise. To ensure that these samples fit a practical situation, the models proposed in [5] are used for hurricane modeling and the models in [60] are used for identifying the response of each component to the modeled hurricanes. The features are normalized to [0, 1] range based on the maximum considered values of wind speed and distance. These samples are shown in Fig. 3.

Although several other features can be defined, when the dimension increases, typically a significant amount of training data is required to ensure that the samples cover all combinations of feature values. As gathering component level data is not trivial, a limited number of samples is synthesized in the studied dataset and only the two most important/salient features (i.e., wind speed and distance) are used in the outage estimation problem.

To measure the performance of the proposed method, a series of penalty parameters ($c = 0.01, 0.10, 1.00, 10.00, 100.00$) with various common kernels are examined. In each setting. A weighted soft-margin SVM (wSVM) [48] is used to compare the performance. The wSVM adjusts the

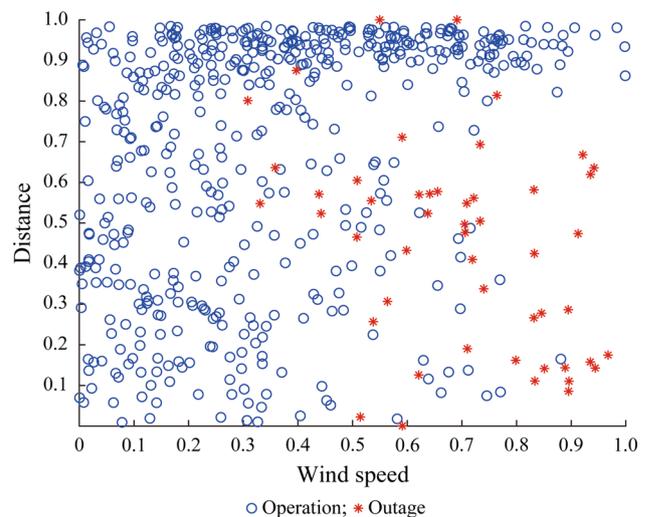


Fig. 3 Generated samples for classes of operation and outage

class sensitivity (penalty of missclassifying) of each class inversely proportional to the frequencies of the class in the training set. In other words, the penalty of missclassifying outage samples is 0.91 (500/550) and the penalty of missclassifying operational samples is 0.09 (50/550). Table 1 shows the average F_1 -score of both wSVM and TWSVM over a 5-fold cross validation. On average, TWSVM took 0.0148 seconds to solve the problem and SVM took 0.0320 seconds to find proper separating hyperplane over 5-fold cross validation.

As it is shown, TWSVM with quadratic kernel and $c = 1.00$ offers the best performance among other settings with the average overall precision of 0.932, recall of 0.912 and F_1 -score of 0.922. The relatively small variance (about 3%) in the F_1 -score of the SVM and TWSVM under various hyper-parameters indicates that both methods are insensitive to hyper-parameters and are not over-fitted to the training data in the studied case. A third order polynomial logistic regression model is also trained and examined in the same fashion (i.e., 5-fold cross validation) to predict the component outages. The logistic regression model has an F_1 -score of 0.856 on the test set which advocates on the superior performance of both SVM and TWSVM in solving this problem.

3.2 Evaluating posterior probability estimation

To determine the likelihood of a sample belonging to each class, a sigmoid posterior probability function is constructed over the values of score function (12) of the trained model with quadratic kernel and penalty parameter $c = 1.00$. The scaling weights of sigmoid function are calculated as $\alpha = -25.93$ and $\beta = 2.12$ by solving (16). The trained model probability weight $\lambda = 0.5$ has overall AU-ROC of 0.89 on the test subset. Other weight parameters ($\lambda = 0, 0.5, 1.0, 1.5$) are further tested on the validation set, however $\lambda = 0.5$ produces the best result in terms of AU-ROC. Figure 4 demonstrates posterior probability for different weight parameters. As shown, by increasing λ the posterior probability function becomes smoother and the classes become less distinguishable. A

small value of weight parameter, e.g., $\lambda = 0$, makes the probabilistic model very sharp where probabilities are either zero or one depending on the predicted class, and hence the model doesn't generalize well for the sample in the area between the two classes.

3.3 Evaluating probabilistic load curtailment estimation

Eight components are considered to be damaged in the path of the upcoming hurricane. The outage probability of these components is calculated based on estimated wind speed and distance from the center of the hurricane and through the proposed posterior probability estimation. Table 2 shows the distance and wind speed of each component, normalized based on the highest wind speed (obtained from the category of the hurricane) and the distance of the furthest impacted component from the center of the hurricane (line 44). The calculated outage probability is also shown in this table for each impacted component. As the results suggest, the components that are closer to the hurricane and experience higher wind speeds, such as line 46, show a very high probability of outage, here as much as 99.5%. On the other hand, the components far from the hurricane and subject to lower wind speeds may show very small chances of outages, such as line 44 which only has a 1.7% outage probability.

The obtained outage probabilities show a promising improvement compared to the existing work in this area which only provide a 0/1 output, i.e., showing whether each component is operational or on outage. Identifying outage probabilities would provide significant opportunities in better managing the available resources as the system response and recovery studies can shift from deterministic models to probabilistic models.

These outage probabilities are used to define $2^8 = 256$ scenarios, where all possible combinations of outage/operational sets of these components are considered. These scenarios are fed into the load curtailment estimation problem which is formulated using mixed-integer linear

Table 1 F_1 -score of classifying system components into two classes of outage and operation with various kernels and penalty parameters

c	Linear kernel		Quadratic kernel		Gaussian kernel	
	wSVM	TWSVM	wSVM	TWSVM	wSVM	TWSVM
0.01	0.871	0.891	0.862	0.892	0.851	0.881
0.10	0.871	0.899	0.871	0.901	0.852	0.891
1.00	0.879	0.915	0.880	0.922	0.851	0.891
10.00	0.881	0.904	0.869	0.912	0.842	0.880
100.00	0.879	0.902	0.869	0.899	0.844	0.872

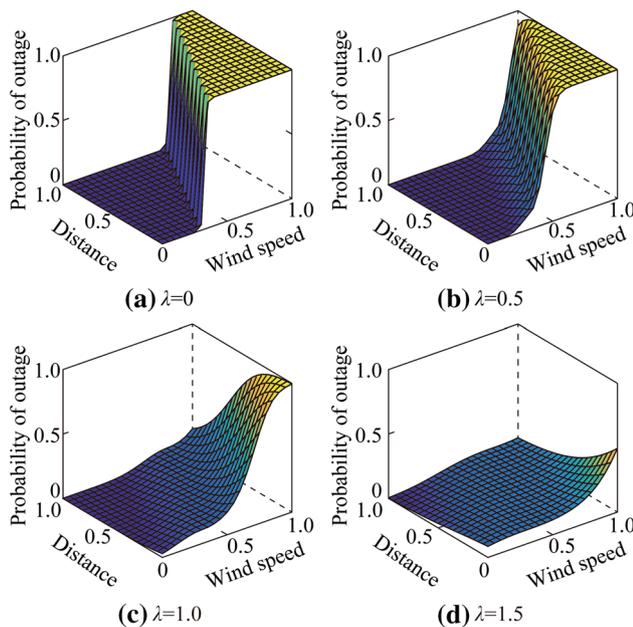


Fig. 4 Posterior probability models for various values of λ

Table 2 Components along hurricane path and their predicted outage probabilities

Component	Wind speed	Distance	Outage probability
Line 44	0.471	1.000	0.017
Line 45	0.471	0.873	0.032
Line 48	0.509	0.571	0.091
Line 50	0.509	0.555	0.077
Line 49	0.509	0.492	0.183
Line 47	0.644	0.444	0.220
Line 30	0.962	0.142	0.971
Line 46	0.994	0.120	0.995

programming (MILP) and solved by CPLEX 12.6 [61]. A value of lost load of 1000 \$/MWh is considered.

The problem objective is calculated as \$1054507 in which \$1024226 is the operation cost and the rest is the aggregated cost of load curtailment in all scenarios. The highest load curtailment is experienced in scenario 129, in which line 30 is in service and all other lines are on outage. The expected load curtailment in this scenario is 434 MWh, however the probability of this scenario is only 1.25×10^{-9} . The highest probability, 0.59, occurs in scenario 112 in which lines 30 and 46 are on outage and other lines are in service. However, there is no load curtailment in this scenario. The focus of this paper is to estimate potential load curtailments in response to imminent hurricanes, however, other probabilistic factors, such as renewable energy generation can be easily formulated and integrated into the proposed model.

4 Conclusion

In this paper, a probabilistic load curtailment estimation model was proposed through a three-step sequential method. At first, to determine a deterministic outage state of the grid components in response to a forecasted hurricane, a machine learning model based on TWSVM was proposed. Then, to convert the deterministic results into probabilistic outage states, a posterior probability sigmoid model was trained on the obtained results from the previous step. Finally, the obtained component outages were integrated into a load curtailment estimation model to determine the potential load curtailments in the system. The simulation results on a standard test system illustrated the high accuracy performance of the proposed method.

The work concludes that the probabilistic load curtailment estimation offers a viable prospect to understand the most impactful outage scenarios in the system, as well as the severity of their impact, in response to an upcoming hurricane, and opens significant opportunities in better planning for those events. In this work, since historical data for hurricanes at component level are limited, a synthetic data is used to show the effectiveness of the proposed method. In future, more detailed historical data for hurricanes will be requested from some of the utility companies affected by hurricanes. In addition, the authors are currently investigating applying the proposed probabilistic outage estimation model for renewable energy integration and accordingly studying the impact of growing renewable penetration on system resilience in response to hurricanes.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- [1] Branscomb LM (2006) Sustainable cities: safety and security. *Technol Soc* 28:225–234
- [2] Holling CS (1973) Resilience and stability of ecological systems. *Ann Rev Ecol Syst* 4:1–23
- [3] Cutter SL, Ahearn JA, Amadei B et al (2013) Disaster resilience: a national imperative. *Environ Sci Policy Sustain Dev* 55:25–29
- [4] Hollnagel E, Woods DD, Leveson N (2007) *Resilience engineering: concepts and precepts*. Ashgate Publishing Ltd., Liverpool
- [5] Eskandarpour R, Khodaei A, Lin J (2016) Event-driven security-constrained unit commitment with component outage estimation based on machine learning method. In: *Proceedings of North American power symposium, Denver, USA, 18–20 September 2016*, 6 pp



- [6] Eskandarpour R, Khodaei A, Lin J (2016) Event-driven security-constrained unit commitment. In: Proceedings of IEEE PES innovative smart grid technologies conference (ISGT), Minneapolis, USA, 6–9 September 2016, 6 pp
- [7] Arab A, Khodaei A, Khator SK et al (2015) Stochastic pre-hurricane restoration planning for electric power systems infrastructure. *IEEE Trans Smart Grid* 6(2):1046–1054
- [8] Arab A, Khodaei A, Han Z et al (2015) Proactive recovery of electric power assets for resiliency enhancement. *IEEE Access* 3:99–109
- [9] Arab A, Khodaei A, Khator SK et al (2013) Post-hurricane transmission network outage management. In: Proceedings of IEEE great lakes symposium on smart grid and the new energy economy, Chicago, USA, 5–8 September 2013, 6 pp
- [10] Arab A, Khodaei A, Khator SK et al (2015) Transmission network restoration considering AC power flow constraints. In: Proceedings of 2015 IEEE international conference on smart grid communications, Miami, USA, 2–5 November 2015, pp 816–821
- [11] Arab A, Khodaei A, Khator S et al (2016) Electric power grid restoration considering disaster economics. *IEEE Access* 4:639–649
- [12] Eskandarpour R, Edwards G, Khodaei A (2016) Resilience-constrained unit commitment considering the impact of microgrids. In: Proceedings of North American power symposium (NAPS), Denver, USA, 18–20 September 2016, 5 pp
- [13] Fotuhi-Firuzabad M, Billinton R, Aboreshaid S (1996) Response health constraints in economic load dispatch considering standby units, interruptible loads and postponable outages. *IEE Proc Gener Transm Distrib* 143(6):599–607
- [14] Gholami A, Shekari T, Aminifar F et al (2016) Microgrid scheduling with uncertainty: the quest for resilience. *IEEE Trans Smart Grid* 7(2):2849–2858
- [15] Li Z, Shahidehpour M, Aminifar F et al (2017) Networked microgrids for enhancing the power system resilience. *Proc IEEE* 105(7):1289–1310
- [16] Liu X, Shahidehpour M, Li Z et al (2017) Microgrids for enhancing the power grid resilience in extreme conditions. *IEEE Trans Smart Grid* 8(2):589–597
- [17] Rudin C, Waltz D, Anderson RN et al (2012) Machine learning for the New York City power grid. *IEEE Trans Pattern Anal Mach Intell* 34(2):328–345
- [18] Negnevitsky M, Mandal P, Srivastava AK (2009) Machine learning applications for load, price and wind power prediction in power systems. In: Proceedings of 2009 15th international conference on intelligent system applications to power systems, Curitiba, Brazil, 8–12 November 2009, 6 pp
- [19] Wu L, Kaiser G, Rudin C et al (2011) Evaluating machine learning for improving power grid reliability. In: Proceedings of ICML 2011 workshop on machine learning for global challenges, Bellevue, USA, 11–13 January 2011, 4 pp
- [20] Bishop CM (2006) Pattern recognition. *Mach Learn* 128:1–58
- [21] Hatziaargyriou N (2001) Machine learning applications to power systems. In: Paliouras G, Karkaletsis V, Spyropoulos CD (eds) *Machine learning and its applications: advanced lectures*. Springer, Heidelberg, pp 308–317
- [22] Pang CK, Prabhakara FS, El-abiad AH et al (1974) Security evaluation in power systems using pattern recognition. *IEEE Trans Power Appar Syst PAS* 93(3):969–976
- [23] Wehenkel L (1997) Machine learning approaches to power-system security assessment. *IEEE Expert* 12(5):60–72
- [24] Sobajic DJ, Pao Y-H (1989) Artificial neural-net based dynamic security assessment for electric power systems. *IEEE Trans Power Syst* 4(1):220–228
- [25] Quan H, Srinivasan D, Khosravi A (2014) Short-term load and wind power forecasting using neural network-based prediction intervals. *IEEE Trans Neural Netw Learn Syst* 25(2):303–315
- [26] Hernández L, Baladrón C, Aguiar JM et al (2014) Artificial neural networks for short-term load forecasting in microgrids environment. *Energy* 75:252–264
- [27] Papalexopoulos AD, Hao S, Peng T-M (1994) An implementation of a neural network based load forecasting model for the EMS. *IEEE Trans Power Syst* 9(4):1956–1962
- [28] Kariniotakis GN, Stavrakakis GS, Nogaret EF (1996) Wind power forecasting using advanced neural networks models. *IEEE Trans Energy Convers* 11(4):762–767
- [29] Guikema SD (2009) Natural disaster risk analysis for critical infrastructure systems: an approach based on statistical learning theory. *Reliab Eng Syst Saf* 94:855–860
- [30] Thukaram D, Khincha HP, Vijaynarasimha HP (2005) Artificial neural network and support vector machine approach for locating faults in radial distribution systems. *IEEE Trans Power Deliv* 20(2):710–721
- [31] Nateghi R, Guikema SD, Quiring SM (2011) Comparison and validation of statistical methods for predicting power outage durations in the event of hurricanes. *Risk Anal* 31(12):1897–1906
- [32] Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20:273–297
- [33] Guenther N, Schonlau M (2016) Support vector machines. *Stata J* 16(4):917–937
- [34] Bae KY, Jang HS, Sung DK (2017) Hourly solar irradiance prediction based on support vector machine and its error analysis. *IEEE Trans Power Syst* 32(2):935–945
- [35] Shi J, Lee W-J, Liu Y et al (2012) Forecasting power output of photovoltaic systems based on weather classification and support vector machines. *IEEE Trans Ind Appl* 48(3):1064–1069
- [36] Zhang X, Wang J, Zhang K (2017) Short-term electric load forecasting based on singular spectrum analysis and support vector machine optimized by Cuckoo search algorithm. *Electr Power Syst Res* 146:270–285
- [37] Alam MR, Muttaqi K, Bouzerdoum A (2017) Evaluating the effectiveness of a machine learning approach based on response time and reliability for islanding detection of distributed generation. *IET Renew Power Gener* 11(11):1392–1400
- [38] Liu Z, Cui Y, Li W (2015) A classification method for complex power quality disturbances using EEMD and rank wavelet SVM. *IEEE Trans Smart Grid* 6(4):1678–1685
- [39] Moulin LS, Da Silva AA, El-Sharkawi MA et al (2004) Support vector machines for transient stability analysis of large-scale power systems. *IEEE Trans Power Syst* 19(2):818–825
- [40] Eskandarpour R, Khodaei A (2017) Machine learning based power grid outage prediction in response to extreme events. *IEEE Trans Power Syst* 32(4):3315–3316
- [41] Eskandarpour R, Khodaei A (2018) Leveraging accuracy-uncertainty tradeoff in SVM to achieve highly accurate outage predictions. *IEEE Trans Power Syst* 33(1):1139–1141
- [42] Wu G, Chang EY (2003) Class-boundary alignment for imbalanced dataset learning. In: Proceedings of ICML 2003 workshop on learning from imbalanced data sets II, Washington DC, USA, 8–9 June 2003, pp 49–56
- [43] Khemchandani R, Chandra S (2007) Twin support vector machines for pattern classification. *IEEE Trans Pattern Anal Mach Intell* 29(5):905–910
- [44] Ding S, Yu J, Qi B et al (2014) An overview on twin support vector machines. *Artif Intell Rev* 42(2):245–252
- [45] Khemchandani R, Chandra S (2017) *Twin support vector machines (TWSVM) for classification*. Springer, Heidelberg, pp 43–62

- [46] Shao Y-H, Deng N-Y, Yang Z-M et al (2012) Probabilistic outputs for twin support vector machines. *Knowl Based Syst* 33:145–151
- [47] Batuwita R, Palade V (2013) Class imbalance learning methods for support vector machines. In: He H, Ma Y (eds) *Imbalanced learning: foundations, algorithms, and applications*. Wiley, Hoboken
- [48] Veropoulos K, Campbell C, Cristianini N et al (1999) Controlling the sensitivity of support vector machines. In: *Proceedings of the international joint conference on artificial intelligence*, Stockholm, Sweden, 31 July–6 August 1999, pp 55–60
- [49] Chawla NV, Bowyer KW, Hall LO et al (2002) SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res* 16(1):321–357
- [50] Lin Z, Hao Z, Yang X et al (2009) Several SVM ensemble methods integrated with under-sampling for imbalanced data learning. In: *International conference on advanced data mining and applications*, Nanjing, China, 16–18 November 2009, pp 536–544
- [51] Imam T, Ting KM, Kamruzzaman J (2006) z-SVM: an SVM for improved classification of imbalanced data. In: *Proceedings of the 19th Australian joint conference on artificial intelligence: advances in artificial intelligence*, Hobart, Australia, 4–8 December 2006, pp 264–273
- [52] Li P, Chan KL, Fang W (2006) Hybrid kernel machine ensemble for imbalanced data sets. In: *18th international conference on pattern recognition*, Hong Kong, China, 20–24 August 2006, pp 1108–1111
- [53] Arjunan SP, Kumar DK, Naik GR (2010) A machine learning based method for classification of fractal features of forearm sEMG using twin support vector machines. In: *Proceedings of 2010 annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, Buenos Aires, Argentina, 31 August–4 September 2010, pp 4821–4824
- [54] Naik GR, Kumar DK (2010) Twin SVM for gesture classification using the surface electromyogram. *IEEE Trans Inf Technol Biomed* 14(2):301–308
- [55] Cong H, Yang C, Pu X (2008) Efficient speaker recognition based on multi-class twin support vector machines and GMMs. In: *Proceedings of 2008 IEEE conference on robotics, automation and mechatronics*, Chengdu, China, 21–24 September 2008, pp 348–352
- [56] Zhang X, Gao X, Wang Y (2009) Twin support tensor machines for MCS detection. *J Electron* 26(3):318–325
- [57] Platt JC (1999) Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Adv Large Margin Classif* 10:61–74
- [58] Zadrozny B, Elkan C (2002) Transforming classifier scores into accurate multiclass probability estimates. In: *Proceedings of the eighth ACM SIGKDD international conference on knowledge discovery and data mining*, Edmonton, Canada, 23–26 July 2002, pp 694–699
- [59] Illinois Center for a Smarter Electric Grid (2018) IEEE 118-bus system. <http://publish.illinois.edu/smartergrid/ieee-118-bus-system/>. Accessed 7 June 2018
- [60] Arab A, Tekin E, Khodaei A et al (2014) Dynamic maintenance scheduling for power systems incorporating hurricane effects. In: *Proceedings of 2014 IEEE international conference on smart grid communications*, Venice, Italy, 3–6 November 2014, pp 85–90
- [61] GAME (2018) CPLEX 12. https://www.gams.com/latest/docs/S_CPLEX.html. Accessed 10 August 2018

Rozhin ESKANDARPOUR received the M.S. degree in electrical engineering from the Department of Electrical and Computer Engineering at University of Denver, Denver, USA in 2016. She is currently pursuing her Ph.D. degree and working as a graduate research assistant in the same department. Her primary research focus is on improving power system resilience using advanced Artificial Intelligence techniques.

Amin KHODAEI received the Ph.D. degree in electrical engineering from the Illinois Institute of Technology, Chicago, in 2010. He was a visiting faculty (2010–2012) in the Robert W. Galvin Center for Electricity Innovation at Illinois Institute of Technology. He joined University of Denver, Denver, USA, in 2013 as an Assistant Professor. He is currently an Associate Professor and the Chair of the Electrical and Computer Engineering department in the Daniel Felix Ritchie School of Engineering & Computer Science, University of Denver. His research interests include power system operation, planning, computational economics, microgrids, smart electricity grids, and power system resilience.

