A Machine Learning Approach for Collusion Detection in Electricity Markets Based on Nash Equilibrium Theory

Peyman Razmi, Majid Oloomi Buygi, and Mohammad Esmalifalak

***Abstract*—We aim to provide a tool for independent system operators to detect the collusion and identify the colluding firms by using day-ahead data. In this paper, an approach based on supervised machine learning is presented for collusion detection in electricity markets. The possible scenarios of the collusion among generation firms are firstly identified. Then, for each scenario and possible load demand, market equilibrium is computed. Market equilibrium points under different collusions and their peripheral points are used to train the collusion detection machine using supervised learning approaches such as classification and regression tree (CART) and support vector machine (SVM) algorithms. By applying the proposed approach to a four-firm and ten-generator test system, the accuracy of the proposed approach is evaluated and the efficiency of SVM and CART algorithms in collusion detection are compared with other supervised learning and statistical techniques.**

***Index Terms*—Market power, collusion detection, machine learning, support vector machine (SVM), classification and regression tree (CART), statistical method.**

I. INTRODUCTION

In a competitive electricity market, producers cannot affect the market price. They can only determine the amount of their own production according to the market price. Therefore, the optimum production level will be calculated by the intersection of the marginal cost curve with market price. The limitations of power systems have pushed the electricity markets from competitive environment toward oligopoly market. In an oligopoly market, there is a possibility for producers to affect the market price. The producers who are interested in raising the market price may increase their bids (economic withholding) or reduce the production (physical withholding) to influence the market price in their favorable direction. The producers, who can affect the market with any of the above mentioned methods, have the so called "market power" [1], [2]. In addition to raising prices higher than the competitive level, the ability of maintaining the market price for a long period is important.

After the liberalization of the electricity market, price manipulation and market power of individual producers [3]-[5] can result from the joint decision and the alliance of two or several producers. In such cases, the collusion is done in the form of implicit and explicit [6], [7]. A long-term alliance between several firms to inflate the market price out of its competitive range is called collusion. Generation firms, which cannot obtain their desired profit through fair competition in oligopoly market, may collude and set their bid and market prices to a value abnormally higher than what oligopoly competition commands. Explicit collusion is a hidden agreement for the interaction among electricity power producers who share confidential information in order to control the market price. The goal of a coalition is to adopt or enforce unified strategies to increase the profit of its members. However, the influence of a coalition does not remain limited to the bids of its members and will alter the bids of its competitors. Stakeholder's choice of coalition is based on the projected profit after the formation of the coalition. In other words, each agent selects the coalition that provides the maximum profit. Members of an explicit coalition share the profit that is obtained through price manipulation [8], [9].

With the advent of modern wholesale electricity markets in developed countries, these markets have become the subject of many discussions. These discussions often originate from the nature of the electricity markets, e.g., they are often controlled by a few number of firms, the traded electricity cannot be stored, the demand has an inelastic nature, and technical limitations such as the congestion of transmission line can lead to the isolation of sub-markets. These characteristics provide an enabling environment for coalitions, which can be regarded as the collusion [10]. The absence of the competition in a market where prices are set by the collusion leads to the violation of the consumers' rights and the reduction of producer's efficiency. In light of these adverse effects, regulatory organizations often prohibit such coalitions to protect the competition [11], [12]. This paper focuses on the explicit collusion and proposes an approach to reveal and detect such collusive behaviors. However, it is not

---

Manuscript received: August 27, 2018; accepted: October 12, 2019. Date of CrossCheck: October 12, 2019. Date of online publication: June 2, 2020.

P. Razmi is with the Department of Electrical Engineering, Ferdowsi University of Mashhad, Mashhad, Iran (e-mail: Peyman.razmi65@gmail.com).

M. O. Buygi is with the Department of Electrical Engineering, Ferdowsi University of Mashhad, Mashhad, Iran (e-mail: m_oloomi@yahoo.com).

M. Esmalifalak is with the Predictive Modeling Department of Fiix Software, Toronto, Canada (e-mail: esmalifalak.mohammad@gmail.com).

DOI: 10.35833/MPCE.2018.000566

easy to automatically detect such collusive behaviors.

The basic approach is to use supervised learning, but the major problem of this approach is the unavailability of data specifically referring to collusion. The alternative is to use unsupervised learning techniques for this purpose [13], [14]. The detection mechanism introduced in [13] focuses on circular trading, in which a ring of colluder trades a certain share repeatedly to raise their price. The proposed method is based on the interpretation of stock flow graph with Markov clustering algorithm. Reference [14] focuses on the detection of cross trading and investigates the aptitude of different clustering algorithms for collusion detection in electricity markets. Reference [15] detects the collusion using a two-step process in an energy market. Firstly, the behavioral pattern of the collusion is investigated. Then, by using a change-point analysis, the behavior of each agent is studied to reveal the possible structural breakpoints. The existence of such breakpoints could be a sign for the collusion and has to be analyzed more in a verification phase. In the verification phase, behavioral similarities of candidates are checked using statistical methods. The execution of market power and the inefficiencies of restructured electricity market in Alberta, Canada, are studied in [16]. The hourly wholesale market data set from 2008 to 2014 is used to find that firms conduct considerable power in the highest demand hours with bounded excess generation capacity. A game-theoretic model is used in [17] to analyze the behaviours of independent system operator (ISO) and the generator leading to collusive transactions. Reference [18] demonstrates the relationship between market concentration and collusion sustainability and their dependence to the strength of the network externalities. In [19], from the regulatory body's viewpoint, a model powered by game theory and agent-based systems is used to investigate the behaviours of each market player. The proposed method looks at the tacit collusive behaviour of generation firms as well as the possibility of explicit collusion. Reference [20] intends to detect the possibility of collusive behavior in the liberalized retail electricity market by testing the validity of the Cournot model of imperfect competition. The monthly data on the Japanese electric power industry are applied from 2005 to 2010. It indicates that the possibility of collusive behavior among the incumbents cannot be excluded. Moreover, the test results imply that larger market size and firm size might reduce the possibility of collusion among incumbent retail electricity suppliers. Reference [21] provides an approach to analyze the development of tacit collusion between a generator company (Genco) and a distributed company (Disco) in a simulated electricity market. Behaviors of Gencos and Discos are modeled using state-action-reward-state-action (SARSA) learning algorithm, and the model is used to tune continual exploration and make the tradeoff between the exploration and exploitation. A typical market with three Gencos and two Discos to prove the possibility of tacit collusion between a Genco and a Disco is analyzed, which shows that the development of this type of collusion increases their utilities. Reference [22] studies the collusion during bilateral agreement in the direct electricity purchase market. Firstly, a multiple-to-multiple (MTM) model of bilateral negotiation in direct electricity purchase market is provided based on the traditional Rubinstein negotiation model. Then, the collusion is divided into two categories.

Specifically, the collusion can be divided into two categories. One is based on the pre-specified information and the other is based on process communication. Then, the bidding strategies under the above two collusions are studied in detail. Finally, the impact of the collusion on the market is analyzed through simulations, and regulatory proposals are proposed. Reference [23] investigates the competition on local electricity markets to avoid the effects of market power and analyze the impact of tacit collusion on the possibility of exercising market power. It shows that more competitions would diminish the effects of market power, especially when a high-level competition with market shares of no more than 20%, because each competitor is needed for peak production. They also indicate that the overall effect of tacit collusion seems to be small in order to manipulate the market price by the colluder. Furthermore, tacit collusion in the form of signaling has no impact on base agent markups. This paper uses supervised learning methods to detect the collusion in electricity markets. Market equilibrium is firstly computed for different loads and collusions. Market equilibrium data and their neighboring operation points are used to train the collusion detection machine by applying support vector machine (SVM) and decision tree algorithms. The efficiencies of these algorithms are compared with other supervised learning and statistical method.

The rest of this paper is organized as follows. Section II describes the process through which Nash equilibrium is modeled and market equilibrium is computed. Section III overviews the machine learning approach used to detect the collusion. In Section IV, the method of finding quasi-actual system operation points is described. In Section V, the developed collusion detection method is applied to the model of a typical electricity market, and then the results are discussed and analyzed in Section VI.

II. MODELING OF EQUILIBRIUM POINT

In practice, there is no enough data from different colluded scenarios to train the collusion detection machine. The purpose of equilibrium point modeling is to simulate and collect data for the training process in subsequent steps. Normal operation points are usually fluctuating around Nash equilibrium. Nash equilibrium is the point where no firm is better by changing its strategy unilaterally [24]. In other words, the deviation from the Nash equilibrium point by participants will not increase their profits. For the modeling of equilibrium point, it is necessary to define the assumptions and conditions of market operation in the restructured power industry. For this purpose, we assume that the day-ahead market is a pool-based market with uniform pricing. Assume that the cost of generating Qs_i by unit i is:

$$C(Qs_i)=a_i\cdot Qs_i+\frac{1}{2}b_i\cdot Qs_i^2 \tag{1}$$

where a_i and b_i are the cost function coefficients of unit i. These coefficients reflect the operation cost of unit i when it generates Qs_i. Also, the utility of consuming Q_{Dj} by consumer j is:

$$C(Q_{Dj}) = c_j Q_{Dj} - \frac{1}{2} d_j Q_{Dj}^2 \tag{2}$$

where c_j and d_j are the demand function coefficients of consumer j. These coefficients reflect the utility of consumer j when he or she consumes Q_{Dj}. ρ_{true} denotes the true linear supply function of generation unit i, which is also the real marginal cost of unit i:

$$\rho_{true}(Qs_i) = a_i + b_i \cdot Qs_i \tag{3}$$

As strategic data, these marginal costs are confidential for the producers. On the supply side, a set of firms is defined as:

$$F = \{f_1, f_2, ..., f_k\} \tag{4}$$

The goal of firm $f_i \in F$ is to maximize its profit by determining the optimal parameters of a similar linear bid function as:

$$\rho_{bid}(Qs_i) = \alpha_i + \beta_i \cdot Qs_i \tag{5}$$

These coefficients reflect the purchasing cost from unit i when it generates Qs_i. To obtain a unique solution, one must avoid changing both α and β. Instead, one parameter should be kept constant, e. g., $\beta_i = b_i$, and the other parameter α_i should be altered [20]. ISO supervises the schedule of generators and market clearing price (MCP) with the aim of maximizing social welfare without violating the technical constraints. The objective of ISO can therefore be modeled as follows:

$$\max J_{ISO} = \sum_{j \in D} \left(c_j Q_{Dj} - \frac{1}{2} d_j Q_{Dj}^2 \right) - \sum_{i \in S} \left(a_i Qs_i + \frac{1}{2} b_i Qs_i^2 \right) \tag{6}$$

s.t.

$$\sum_{i \in S} Qs_i - \sum_{i \in D} Q_{Di} = 0 \tag{7}$$

$$Qs_i^{\min} \le Qs_i \le Qs_i^{\max} \tag{8}$$

where J_{ISO} is the social welfare; S is the set of generation units; D is the set of consumers; and $Qs_i^{\min}$ and $Qs_i^{\max}$ are the capacity limits of unit i. Firm f maximizes its profit by determining the optimal bid for its units using the following bi-level optimization.

$$\begin{cases} \max \pi_f = \sum_{i \in f} \left(\lambda \cdot Qs_i - a_i \cdot Qs_i - \frac{1}{2} Qs_i^2 \right) \\ \text{s.t.} \;\; (6)\text{-}(8) \\ \qquad \alpha_i^{\min} \le \alpha_i \le \alpha_i^{\max} \end{cases} \tag{9}$$

where π_f is the profit of firm f; $\alpha_i^{\min}$ and $\alpha_i^{\max}$ are the lower and upper limits of α_i, respectively; and λ is the power balance constraint (7). The strategy of firm f to outbid other firms is obtained by solving a mathematical program with equilibrium constraints (MPECs) consisting (9). Decision variable of generation firm i is α_i. Market equilibrium is obtained by solving equilibrium problem with equilibrium constraint (EPEC) that consists of the sets of MPECs for all generation firms. To solve the EPEC, KKT conditions for optimization problem of each generation firm are proposed. Then, KKT conditions of all generation firms are solved simultaneously using dual-variables-based algorithm [25], which is an iterative algorithm. In each iteration, the first active constraint with the biggest dual variable is identified and KKT conditions are revised. The algorithm continues until all active constraints at market equilibrium are identified and KKT conditions are revised based on active constraints. The remaining KKT conditions are linear equations with unique market equilibrium [25].

III. MACHINE LEARNING

The concept of machine learning originates from pattern recognition and computation learning theory in artificial intelligence [26]. The algorithms developed with machine learning concepts are expected to do their tasks for unspecified data [27]. Machine learning has a multitude of tasks, the most important of which may be the supervised learning [28]. In the supervised learning, the algorithm with a certain input, a certain output, and a group of labeled training data is provided. The algorithm then uses machine inference to develop a function capable of emulating the process and mapping the new data.

A. SVM Algorithm

SVM algorithm [29], [30] is one of the most widely used supervised algorithms in data classification. In fact, SVM algorithm is a binary classifier that defines the optimal boundary between two (or more) classes. Mathematically, SVM uses a linear hyper-plane to separate two classes by maximizing the distance of each class from this plan. The algorithm uses a quadratic programming to find data samples that are in the border of each class, and these samples are known to support vectors. Then, the algorithm searches for the hyper-plane which firstly has the maximum distance from each support vector. Secondly, the distances between the hyper-plane and each support vector are equal. As mentioned above, the support vector algorithm is used only to separate the two classes from each other. But for collusion detection and diagnosis of the firms involved in the collusion, the number of different classes may depend on the type of the collusion, and the number of firms is more than those of the two classes. Hence, for the category and classification of different types of the collusion, specific methods of multi-class SVMs must be used, which include one against one and one against all methods [31], [32]. If the data are inseparable by a linear hyper-plane, the data are mapped to a higher dimensional space which is called feature space by kernel function. Through this procedure, the algorithm becomes capable of fitting the maximum-margin hyper-plane in a transformed feature space. This transformation could be linear or nonlinear and the transformed feature space could be of a higher dimension. Note that in the transformed feature space, the classifier is a hyper-plane, but in the original input space, it could be nonlinear [33]. Factors deciding the SVM performance are the chosen kernel function and its parameters as well as the soft margin parameter C. Gaussian kernel, which has a single γ, is a typical choice for SVM [34]. The common practice for finding the best values of C and γ is to conduct a grid search, i.e., to repeat the calculations with different C and γ combinations and determine the values yielding the best accuracy through cross-validation [35].

B. Classification and Regression Tree (CART) Algorithm

As a predictive model, decision tree is one of the supervised learning and non-parametric methods with the hierarchical structure which is used to classify different types of data, and its results are delivered in a flowchart-like tree structure. According to the dependent variable type, this algorithm is divided into two categories: classification trees for discrete variables and regression trees for continuous variables. In this algorithm, the fragmentation of the data is implemented by using the features as a tree. For better understanding, it is written by using as if-then rules [36]. In each stage of this method, a feature for the data is selected based on the training data and the data set is divided into a class grouping based on the chosen features. This continues until all the data in a category has a single label. One of the most famous decision tree algorithms is CART algorithm, which leads to the creation of a tree with binary division introduced in [37]. CART algorithm is a non-parametric decision tree learning technique, which is used to classify any type of the data. In other words, the CART algorithm produces either the classification or regression trees, depending on whether the dependent variable is categorical or numerical, respectively. At each stage, the Gini index is used to select the best attribute for the data used. One important application of machine learning is to model costumer behaviors in financial markets, which comes from the fact that finding mathematical rules to model the behaviors of individual costumers is a very complex process, e.g., in credit card fraud detection. It would be extremely hard to find reasons and rules for justifying safe versus fraudulent behaviors without the help of machine learning approaches. This research is another attempt to model the behavior of market participants without the detailed knowledge about their incentives. In creating synthetic data, it is assumed that we know safe or fraudulent behaviors, but we let machine learning approach distinguish those behaviors without access to any of rules used to create synthetic data.

IV. COLLUSION DETECTION AND CLASSIFICATION

A. Data Generation

The objective of this paper is to provide a tool that would allow ISO to detect collusion and identify the colluding firms using the day-ahead data, which is pursued via machine learning paradigm. One problem that obstructs the use of trainable machines in collusion detection and classification is the lack of adequate knowledge about the data associated with different collusion among participants. According to Section II, equilibrium points are obtained for different load demands and different kinds of collusion scenarios. To achieve a proper collusion detection technique, the modeling should go beyond the limits of a single period or a certain class of load demand. An exact match between market operation point and computed market equilibrium is rare. But mature electricity markets are often close enough to the equilibrium to allow the peripheral points of the equilibrium to be assumed as quasi-actual operation points [25], [38].

Moreover, due to different changes that occur every day in power systems, and the fact that power producers do not know these changes, market players are not faced with a static environment. Hence, electricity markets may not work exactly on their Nash equilibrium. However, since daily changes in game environment is small in comparison to the whole environment, and all producers are rational and smart, it is reasonable to assume that electricity markets work at a point close to their Nash equilibrium. Thus we consider peripheral equilibrium points for creating synthetic data. Nash equilibrium point is located in a ball with the center of Nash equilibrium and a specified radius. To implement this assumption, the intercept of supply curve of each generation unit is altered around its value at the equilibrium point. And the resulting points are considered as quasi-actual operation points. To determine quasi-actual data, suppose α_p is the set that contains the intercept of bid of all generation units at the p^{th} equilibrium point. Balls B $(\alpha_p,\ h\alpha_p)$ are defined for equilibrium points $p=1,2,...,P$, where α_p and $h\alpha_p$ are the center and radius of the ball p, respectively; P is the number of equilibrium points computed for different load levels and collusion scenarios; and h is the deviation percentage that represents the extent of deviation of quasi-actual data from the considered equilibrium points. Quasi-actual data are selected randomly within these balls. To create the quasi-actual operation points for collusion states, we need to model electricity market and collusion. To model the collusion, the firms participating in the collusion are considered as a single entity, which means that generation firms that participate in the collusion seek to optimize the sum of profits of all firms participating in the collusion, instead of optimizing their profits separately and independently. In other words, in electricity markets, the collusion means that two or more power producers collaborate to increase electricity price for a long time. The producers' bid is colluded so that sum of profits of all colluding firms is maximized. To generate the data that reflect collusion behavior, equilibrium points for all possible collusion scenarios are computed. To simulate the collusion, in each scenario, all colluding power producers are considered as one firm that maximizes the profit of the coalition. Equilibrium points and their peripherals for different collusion scenarios at different load levels are considered as synthetic data for machine learning.

B. Training Machine for Collusion Detection

According to Section IV-A, the equilibrium points and quasi-actual operation points are used for training the machines. In order to detect the collusion and the firms participating in it, some attributes or criteria are required to distinguish the operation points related to the collusion from those without the collusion, which are used to train the trainable machines. After training, the data of a real operation point are given to the trained machine, and the machine detects the occurrence of the collusion and firms participating in this violation. The attributes or criteria used for training the collusion detection machine are as follows.

1) Marginal cost of generators (*MC)*.

2) Bid price of generators (*BP*).

3) Lerner index for different generators (*LER)*.

4) Market share of each producer with load demand (*MS*).

5) Herfindahl Hirschman index of the market with each load demand (*HHI*).

Privately-owned generators naturally refuse to publicly report their marginal cost functions. It is assumed that ISO estimates marginal cost functions of all generators using the available model in [39], [40]. The Lerner index measures the extent to which the bid price of a given firm exceeds its real marginal costs. In other words, the Lerner index measures a firm's level of electricity power and describes the relationship between the elasticity and price margins. The Herfindahl Hirschman index is accepted as an indicator of market competition that measures the level of the concentration in a given market [41], [42]. The market simulation performed in this paper aims to calculate the above-mentioned criteria at the quasi-actual operation points under different collusive conditions. In this paper, for model assessment and error estimation of classifier, *k*-fold cross-validation estimator is used as a randomized method for error estimation [43]. Algorithm 1 shows the procedure of the collusion detection method, where *SC* is the collusion scenario.

```
Algorithm 1: collusion detection algorithm
for $h=h_1:\Delta h:h_{\max}$ do
    1. create historical data
    for $SC=1:SC_{\max}$ (collusion scenarios) do
        for $LD=Q_{D1}:Q_D:Q_{D\max}$ do
            Compute:
            a) Market equilibrium (6)-(11)
            b) Generation power at equilibrium
            c) Collusion criteria
            $\boldsymbol{x}\leftarrow[MC, BP, LER, MS, HHI]$
        end
        Save criteria and label for each scenario
        $\boldsymbol{X}\leftarrow\boldsymbol{x}$
        $\boldsymbol{Y}\leftarrow SC$ (label)
    end
    Data$\leftarrow[\boldsymbol{X}, \boldsymbol{Y}]$
    2. Train machine learning algorithm
    3. Analyze colluding behaviour as the loop below:
    for each new sample do
        new sample belongs to anomaly cluster
                            or
        is not statistically similar to normal operation
        Flag colluding agents ⚑
        Otherwise
            Collusion==No
    end
end
```

V. Simulation Results and Analysis

In this section, the proposed approach is applied to an electricity market with four firms and ten generators. Table I shows the list of generators owned by each producer and Table II shows the characteristics of each generator. Parameters in Table II are coefficients of marginal cost function ($MC_i = a_i + b_i \cdot Qs_i$) and the capacity limits of units. In this paper, we assume that the auction in electricity markets is a static and incomplete information game.

TABLE I
Unit Owned by Each Firm

Firm	Unit	Firm	Unit
Firm 1	1, 2, 3	Firm 3	6, 7
Firm 2	4, 5	Firm 4	8, 9, 10

TABLE II
Parameter of Generation Unit

Firm	Unit	a_i	b_i	$Qs_i^{\min}$	$Qs_i^{\max}$
1	1	25	0.032	0	800
	2	20	0.050	0	600
	3	30	0.038	0	600
2	4	30	0.042	0	400
	5	26	0.060	0	650
3	6	32	0.040	0	700
	7	22	0.055	0	700
4	8	35	0.036	0	600
	9	25	0.045	0	400
	10	20	0.030	0	600

Table III shows the list of bilateral and multi-lateral collusion scenarios for the studied market. There are 10 different scenarios for bilateral and multi-lateral collusion. The goal is to distinguish collusion scenarios from collusion-free scenarios and identify the colluding firms. To compute the equilibrium point for no collusion scenario, we have four coupled bi-level optimization problems, each optimization for one generation firm. To simulate the collusion between Firm 1 and Firm 2, we have three coupled bi-level optimization problems, an optimization for Firm 1 and Firm 2, an optimization for Firm 3, and an optimization for Firm 4. We assume the load demand varies between 3000 MW and 5000 MW ($Q_D^{\max}=5000$) with 50 MW steps ($\Delta Q_D=50$). Hence, we have 41 different load levels. To create quasi-actual data for machine training, equilibrium points for 41 different load levels and different collusion scenarios are computed. Operation points located inside balls B (α_p, $h\alpha_p/100$) are considered as quasi-actual points. Collusion detection attributes or criteria are computed for each quasi-actual operation point.

In this paper, the range of variation applied on the intercept of linear bid function is considered ±10% of the intercept value at the equilibrium $h_{\max}=\pm10$. Therefore, the simulation is performed for different data sets. As shown in Fig. 1, five distinct data sets are defined, for h equal to 2, 4, 6, 8, 10, respectively.

Each data set is specified with its deviation percentage and consists of all quasi-actual operation points located at balls B (α_p, $h\alpha_p/100$) for $p=1,2,\dots,P$. The set of all above-mentioned criteria for all units at any given scenario and load is called a data sample. Each data sample forms a 10×5 matrix. Row i indicates unit i and column j indicates criterion j.

TABLE III
DIFFERENT SCENARIOS OF COLLUSION

Scenario	Collusion type	Label
Fair competition	Collusion free	1
Firm 1 and Firm 2	Bilateral collusion	2
Firm 1 and Firm 3		3
Firm 1 and Firm 4		4
Firm 2 and Firm 3		5
Firm 2 and Firm 4		6
Firm 3 and Firm 4		7
Firm 1, Firm 2, and Firm3	Multi-lateral collusion	8
Firm 1, Firm 2, and Firm4		9
Firm 1, Firm 3, and Firm4		10
Firm 2, Firm 3, and Firm 4		11

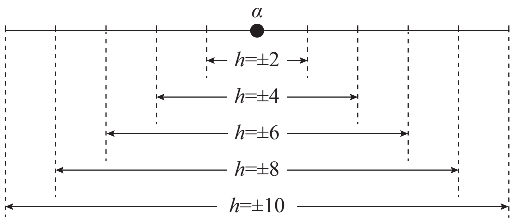


Fig. 1. Percentage of equilibrium point deviation to create peripheral point.

A. Collusion Detection Using Equilibrium Point Data for Machine Training

In this subsection, we only use operation points related to market equilibrium points for machine training and validation. It is also supposed that peripheral equilibrium points, which are selected randomly as quasi-actual operation points, are only used for the testing machine. For machine training, there is a data sample, or a 10×5 matrix of detection criteria for each load level in each of the collusion scenario. A data sample consists of 10 sub-samples and each sub-sample is related to each generation unit. In other words, each sub-sample indicates the behaviors and features of each generator.

Considering 41 different load levels and all collusion scenarios given in Table III , there are 451 data sample, in which 41 data samples are related to fair competition and the rest of the samples are referring to the collusion (41 data samples for each collusion scenario). There are overall 4510 sub-samples, and all of these instances are used in training process, as shown in Fig. 2.

For any equilibrium point, only one peripheral point is considered. All of the peripheral points are used as test samples. Figure 3 show the confusion matrix for SVM and CART algorithms in the test process for data set related to $h = \pm 2$. A confusion matrix contains the information about actual and predicted classifications done by the related classification algorithm [23]. The performance or accuracy of a learning algorithm is commonly evaluated using the confusion matrix. Each row represents the instances in an actual class (target class) while each column of the matrix represents the instances in a predicted class (output class).

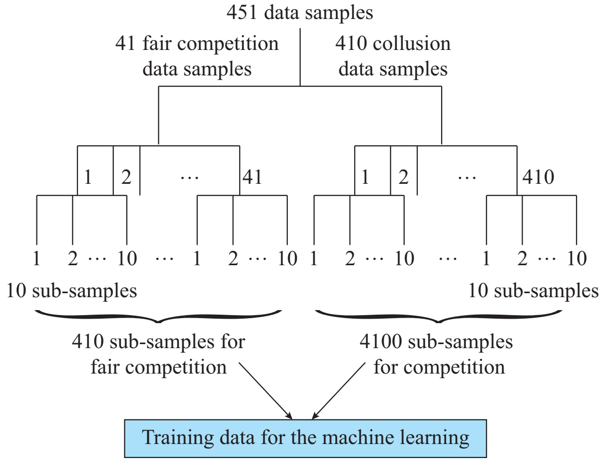
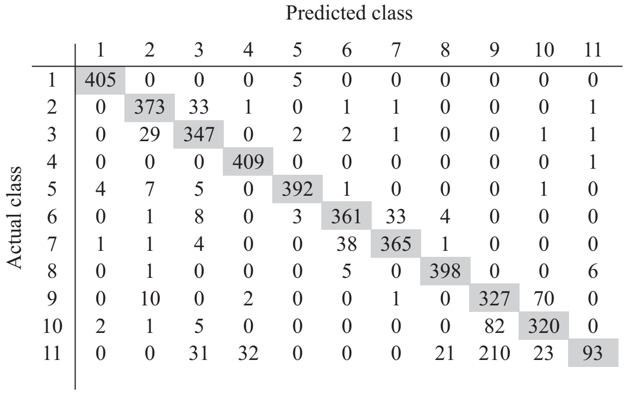


Fig. 2. Samples and sub-samples in training process.



		Predicted class										
		1	2	3	4	5	6	7	8	9	10	11
Actual class	1	405	0	0	0	5	0	0	0	0	0	0
	2	0	373	33	1	0	1	1	0	0	0	1
	3	0	29	347	0	2	2	1	0	0	1	1
	4	0	0	0	409	0	0	0	0	0	0	1
	5	4	7	5	0	392	1	0	0	0	1	0
	6	0	1	8	0	3	361	33	4	0	0	0
	7	1	1	4	0	0	38	365	1	0	0	0
	8	0	1	0	0	0	5	0	398	0	0	6
	9	0	10	0	2	0	0	1	0	327	70	0
	10	2	1	5	0	0	0	0	0	82	320	0
	11	0	0	31	32	0	0	0	21	210	23	93

Fig. 3. SVM confusion matrix in test process.

For example, in Fig. 3, in the whole 410 sub-samples of class 11, only 93 sub-samples are correctly predicted to belong to this class, and the rest instances are incorrectly predicted to belong to other classes. According to Figs. 3 and 4, in the whole 4510 sub-samples related to the all classes in the test data, SVM and CART algorithms correctly predict 3817 and 3238 sub-samples of the actual classes (target classes), respectively. Based on these tables, machine learning has a less accuracy for the separation of samples of all classes. Specially, the error of the machine is more likely to be used to detect the instances of class 11, and class 11 has the highest error rate. Table IV and Table V show the error percentage of collusion detection and the classification for different data sets in training, validation, and test processes using SVM and CART algorithms, respectively. For all deviation percentage in α and different data sets, optimal Gaussian kernel and soft margin parameters $C = 100$ and $\gamma = 5$ are selected using k-fold cross- validation method.

The percentage error and σ in the training and validation process refer to the average percentage and standard deviation of error in 10-fold cross-validation method (10 times training), respectively. Table IV and Table V show that the machine learning does not have a proper performance when the learning and testing data are separated from each other.

	Predicted class										
Actual class	1	2	3	4	5	6	7	8	9	10	11
1	388	2	5	0	9	0	3	3	0	0	0
2	1	290	42	0	16	7	27	14	1	1	11
3	0	31	299	0	3	8	17	24	2	8	18
4	0	5	27	321	0	0	6	17	8	16	10
5	46	11	10	0	322	5	2	7	0	7	0
6	7	19	34	0	21	246	37	42	0	2	2
7	4	10	16	0	5	46	313	16	0	0	0
8	31	2	30	0	0	6	4	325	0	0	12
9	0	0	0	14	0	0	0	9	311	69	7
10	4	0	3	7	0	0	0	2	97	297	0
11	0	11	4	13	0	1	0	179	15	61	126

Fig. 4. CART confusion matrix in test process.

TABLE IV
PERCENTAGE ERROR OF SVM ALGORITHM

Deviation percentage of data set	Train	Validation	Test
$\alpha \pm 0.02\alpha$	0.0225 ± 0.001	0.0285 ± 0.020	0.1596 ± 0.010
$\alpha \pm 0.04\alpha$	0.0234 ± 0.004	0.0276 ± 0.030	0.2288 ± 0.021
$\alpha \pm 0.06\alpha$	0.0244 ± 0.003	0.0256 ± 0.015	0.2996 ± 0.010
$\alpha \pm 0.08\alpha$	0.0222 ± 0.005	0.0264 ± 0.025	0.3470 ± 0.030
$\alpha \pm 0.10\alpha$	0.0231 ± 0.003	0.0291 ± 0.030	0.3860 ± 0.002

TABLE V
ERROR PERCENTAGE OF CART ALGORITHM

Deviation percentage of data set	Train	Validation	Test
$\alpha \pm 0.02\alpha$	0.0152 ± 0.005	0.0632 ± 0.05	0.282 ± 0.04
$\alpha \pm 0.04\alpha$	0.0232 ± 0.008	0.0563 ± 0.05	0.301 ± 0.03
$\alpha \pm 0.06\alpha$	0.0161 ± 0.003	0.0688 ± 0.04	0.322 ± 0.06
$\alpha \pm 0.08\alpha$	0.0184 ± 0.005	0.0612 ± 0.02	0.394 ± 0.04
$\alpha \pm 0.10\alpha$	0.0175 ± 0.004	0.0644 ± 0.03	0.409 ± 0.01

In other words, we consider equilibrium points as training data which refer to specific points. When we use the peripheral equilibrium point in the test process which refers to a space around the equilibrium point, by applying this approach to create historical data, the machine will not have a proper performance. Therefore, the utilization of equilibrium points alone in the train process is not sufficient, and peripheral points of the equilibrium should be utilized in the machine training process to provide a proper tool for collusion detection.

B. Collusion Detection Using Equilibrium Points and Peripheral Points For Training of Supervised Learning

In this subsection, equilibrium points and their peripheral points are used in all processes. In order to create a more comprehensive model of the detector machine, we investigate more points around the equilibrium points. Therefore, five peripheral points are considered in each ball B (α_p, $h\alpha_p/100$). Hence, there are 2255 data samples in each data set, and each data sample consists of 10 sub-samples. Machine learning approaches randomly select 80% of data samples for the learning and validation process, and the remaining data samples are allocated to the test process. For data set related to $h=\pm 2$, Fig. 5 shows the SVM confusion matrix for categorizing 11 classes in the test process. The error percentag for the relevant data set is equal to 3.77%. Table VI and Table VII show the satisfying error percentage of algorithms for different data sets in training, validation, and test processes. By utilizing the equilibrium points and their peripheral points in training and validation process, the machine has a suitable performance in collusion detection and classification, considering collusion-free state as a class and all collusion states defined in Table III as the other class.

	Predicted class										
Actual class	1	2	3	4	5	6	7	8	9	10	11
1	410	0	0	0	0	0	0	0	0	0	0
2	0	385	14	0	11	0	0	0	0	0	0
3	0	21	386	0	1	0	0	0	0	2	0
4	0	0	0	410	0	0	0	0	0	0	0
5	0	13	5	0	392	0	0	0	0	0	0
6	0	0	8	0	0	385	20	5	0	0	0
7	0	0	0	0	0	17	392	1	0	0	0
8	0	0	0	0	0	5	0	404	0	1	0
9	0	0	0	0	0	0	0	0	390	20	0
10	0	0	0	0	0	0	0	0	34	376	0
11	0	0	0	0	0	1	0	0	0	0	410

Fig. 5. SVM confusion matrix in test process.

TABLE VI
SATISFYING ERROR PERCENTAGE OF SVM ALGORITHM

Deviation percentage of data set	Train	Validation	Test	C	γ
$\alpha \pm 0.02\alpha$	0.0076 ± 0.002	0.0356 ± 0.002	0.0377 ± 0.001	50	0.08
$\alpha \pm 0.04\alpha$	0.0854 ± 0.001	0.1054 ± 0.015	0.0936 ± 0.003	30	0.70
$\alpha \pm 0.06\alpha$	0.1230 ± 0.002	0.1570 ± 0.022	0.1460 ± 0.005	50	0.80
$\alpha \pm 0.08\alpha$	0.1560 ± 0.002	0.1910 ± 0.036	0.1960 ± 0.008	30	0.80
$\alpha \pm 0.010\alpha$	0.1880 ± 0.002	0.2140 ± 0.041	0.2240 ± 0.005	25	0.60

Figure 6 shows that quasi actual operation points of the market can be classified into two classes, collusion and collusion-free for $h=\pm 6$. Only marginal cost and Herfindahl Hirschman index are used as detection attributes. Considering $\gamma=0.6$ and $C=100$, SVM is able to separate the training and test samples correctly into two classes with the accuracy of 94% and 93%, respectively. In this subsection, we evaluate the accuracy and efficiency of SVM and CART algorithms together and with other algorithms in collusion detection and classification. According to Table VII and Table VIII, the efficiency of SVM in collusion detection and identifying the collaborator firms for all different data set is higher than that of CART algorithm. SVM and CART algorithms are used to identify colluding firms. SVM algorithm has a proper performance and higher accuracy in identifying colluding firms than CART algorithm.

TABLE VII
SATISFYING ERROR PERCENTAGE OF CART ALGORITHM

Deviation percentage of data set	Train	Validation	Test
$\alpha \pm 0.02\alpha$	0.0145 ± 0.003	0.0448 ± 0.002	4.75 ± 0.003
$\alpha \pm 0.04\alpha$	0.0515 ± 0.004	0.1032 ± 0.030	11.51 ± 0.03
$\alpha \pm 0.06\alpha$	0.1080 ± 0.001	0.1671 ± 0.050	17.80 ± 0.04
$\alpha \pm 0.08\alpha$	0.1476 ± 0.004	0.2030 ± 0.030	21.14 ± 0.02
$\alpha \pm 0.10\alpha$	0.1849 ± 0.004	0.2380 ± 0.030	24.32 ± 0.03

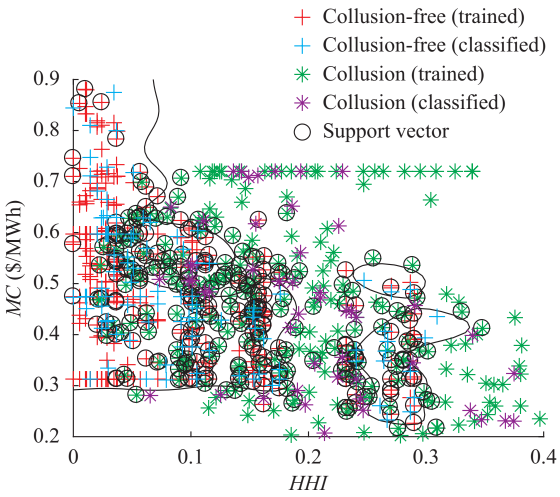


Fig. 6. Classification of created operation points in 11 scenarios into collusion and collusion-free classes.

TABLE VIII
ERROR PERCENTAGE OF DIFFERENT ALGORITHMS

Deviation percentage of data set	Error (%)					
	MLP	RBF	KNN	Bayesian	CART	SVM
$\alpha \pm 0.02\alpha$	25.6	24.3	15.4	52.4	4.75	3.77
$\alpha \pm 0.04\alpha$	30.2	25.2	25.3	54.3	11.51	9.39
$\alpha \pm 0.06\alpha$	32.5	34.4	32.3	57.7	17.80	14.00
$\alpha \pm 0.08\alpha$	36.2	36.5	35.6	57.6	21.14	19.60
$\alpha \pm 0.10\alpha$	37.6	38.7	36.5	59.5	24.32	22.40

Furthermore, the results obtained by SVM and CART algorithms are compared with those of other algorithms such as multi-layer perception (MLP) neural networks, radial basis function (RBF) neural network [46], [47], *K*-nearest neighbor (KNN) [48] and Bayesian classifier [49].

Table VIII shows that SVM outperforms other methods due to its capability in providing a good out-of-sample generalization with proper parameter tuning for C and γ [34]. In other words, by choosing appropriate generalization grades, SVM can be robust even when the training sample has some biases. Furthermore, SVM is a useful tool for insolvency analysis when the data are not regularly distributed or have an unknown distribution as the data of Figs. 3 and 5. In fact, in many real-world problems such as electricity markets, collocated data are not linearly separable like the data of Fig. 6. Therefore, SVM can work well with higher accuracies. As shown in Table VIII, CART algorithm also has a good performance, because it is inherently non-parametric. In other words, no assumptions are made regarding the distribution of predictor variables. Thus, CART algorithm can handle numerical data that are highly skewed or multi-modal as well as categorical predictors with either ordinal or nonordinal structure [37].

C. Collusion Detection Using Equilibrium Points and Peripheral Points in Statistical Method

In order to compare the obtained results from supervised algorithm with other methods, we analyze the statistical method in collusion detection and anomaly detection in electricity market. In this subsection, we divide all original data into two groups, normal data and anomaly data. Then, collusion detection algorithm is examined for $0 \le h \le \pm 18$. For each deviation, we calculate F_1 score as a precision criterion which can be calculated as:

$$F_1 = \frac{2P_r R_e}{P_r + R_e} \tag{10}$$

$$R_e = \frac{T_r}{A_c} \tag{11}$$

where F_1 score is a well-known method of evaluating predictive models on skewed data sets [50], which considers the contribution of both precision and recall as an instance, and models with lower precision or recall will lead to lower F_1 score; T_r corresponds to the points that the algorithm detects as positive samples; A_c is the positive point in the data sets; and P_r and R_e are the precision and recall, respectively. P_r is calculated using:

$$P_r = \frac{T_r}{P_o} \tag{12}$$

where P_o is the point that the algorithm detects as positive point but it may have errors. Since our data set could have fewer observations in one of the classes, F_1 score is an effective way to evaluate the predictive model. F_1 score can never be higher than 1. Moreover, the bigger value of F_1 represents the more accurate classifier in general. Based on the obtained result of Fig. 7, it is obvious that for $h < 15$, the statistical method has a better performance than the supervised methods in detecting anomaly data in electricity market. The detection of collusion occurrence and identification of colluding firms are the main issues of collusion detection. Figure 7 shows that statistical methods outperform the supervised methods in collusion detection. On the other hand, since the statistical method cannot identify colluding firms, they cannot be appropriate tools for collusion detection. Therefore, supervised learning algorithms are more suitable for collusion detection. In order to determine the importance of the above-mentioned criteria in collusion detection and classification, we eliminate one of the criteria from the set of criteria and then analyze the collusion detection algorithm for all data sets. Table IX shows the decrease in accuracy of the algorithms after eliminating different criteria. According to Table IX, the importance of *HHI* index in collusion classification is greater than other criteria. Also, the elimination of *MC* and *LER* at the same time strongly reduces the accuracy of the algorithms, which indicates that *MC* or *LER* is important and necessary as a major criterion in collusion detection and classification.

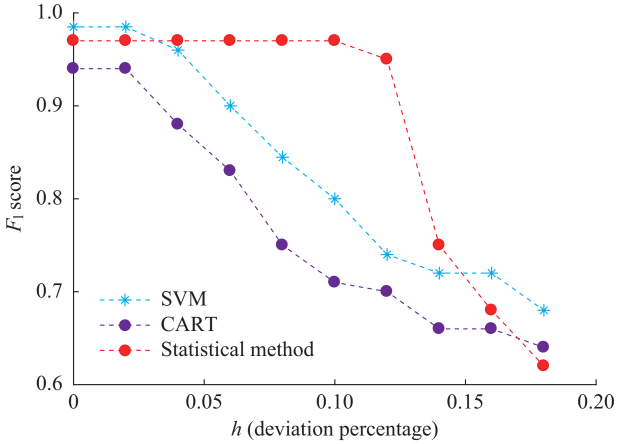


Fig. 7. F_1 score curve in collision detection data.

TABLE IX
ACCURACY DECLINE OF ALGORITHMS IN VALIDATION AND TEST PROCESS FOR 11 CLASSES AFTER ELIMINATING DIFFERENT CRITERIA

Eliminated criteria	Accuracy decline (%)		
	ID3	CART	SVM
MC	0-3	0-3	0-3
LER	0-2	0-2	0-2
HHI	15-20	15-20	10-15
MS	0-2	0-2	4-6
BP	0-2	0-2	0-2
MC and *LER*	5-20	5-20	7-20

Table IX shows a downward trend in the performance of the algorithms after eliminating different criteria. These results will be true until the designed algorithms could be used for categorizing 11 classes. In other words, if the number of scenarios in a new case study is less than 11, it will be possible that the accuracy of algorithms is less after removing the indexes due to the sensitivity of SVM and CART algorithms to a large number of classes. Consider the bilateral collusion states as a class and multi-lateral collusion states defined in Table III as the other class, Fig. 8 shows the classification of generated data set into two classes, if $h=\pm 2$, and only bid and market share are used as detection attributes.

As shown in Fig. 8, considering $\gamma=0.2$ and $C=100$, after eliminating several criteria such as *MC* and *LER*, SVM has appropriate accuracy and discriminates the training and test samples related to the two classes with the accuracy of 93% and 92%, respectively. The red and blue samples (training and test samples of the bilateral class) among multi-lateral collusion samples are related to the bilateral collusion samples created at the high load levels. Thus, bid price in these samples is higher than that in other same samples. The bid price in multi-lateral collusion is much higher than that in bilateral collusion at lower load levels close to 3000 MW.

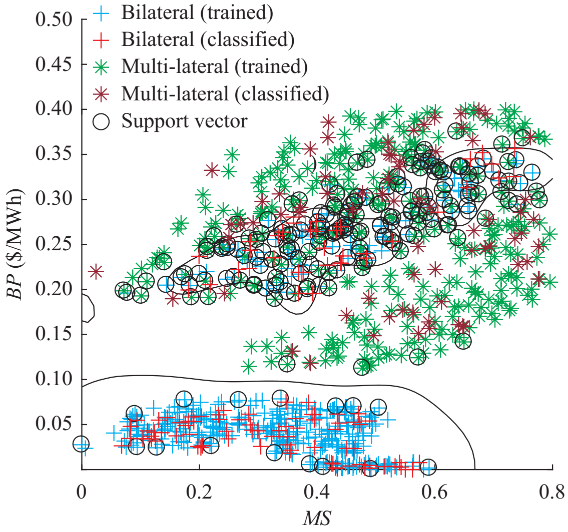


Fig. 8. Classification of created operation points for 10 scenarios in bilateral and multi-lateral collusion classes.

The collusion detection framework is also examined under unconstrained condition, i.e., if generation constraints are ignored. The results of this part are not discussed in this paper. In the unconstrained case, the accuracy and performance of the collusion detection algorithm based on the mentioned approach are slightly better than that under the constrained condition. In the constrained model, the obtained data have some bias and have been distributed erratically. It means that all of the learning algorithms have a relative relevance to the data analyzed and evaluated. Despite a few problems with the generated data in the constrained model, e.g., its non-regularity distribution, SVM and CART algorithms outperform other algorithms and have an appropriate performance in the constrained model. Generally, the existing electricity markets are oligopoly markets, which means that the number of generation firms is limited although each generation firm may have several generation units in practice. A test system with 10 generation units classified in 4 generation firms is considered in this paper. The total running time of the computer is used as a criterion for computation complexity. The total running time for collusion detection in a 2.5 GHz laptop is nearly 20 s. Although the number of generation units is much more than 10 in practice, the number of generation firms is not much more than 4. Hence, the number of the players is not much more than 4 in practice. Therefore, the complexity and scalability of the computation are not big issues.

VI. CONCLUSION

In this paper, an approach for collusion detection in electricity market is proposed based on machine learning. Firstly, the possible scenarios of collusion among generation firms are identified. Then, for each load level and possible collusion scenario, the market equilibrium is computed and peripheral points of the equilibrium are determined. Collusion detection criteria are computed for the equilibrium and their peripheral points. The computed criteria are used to train the learning machines using SVM and decision tree algorithms. Simulation results show that the accuracy of the used machines is acceptable in collusion detection. As the selected peripheral points deviate from the equilibrium points, a downward trend in the accuracy of the collusion detection algorithm is observed. SVM and decision tree algorithms are compared with other machine learning approaches, and it is observed that SVM and decision tree algorithms are the most appropriate techniques to detect the collusion. Finally, the impact of each criterion on the collusion detection accuracy is evaluated and the *MC* and *LER* indices play more important roles in collusion classification performance than other criteria. ISO runs the collusion detection program at any hour. If the results show the collusion of two or more firms for a long time period, the collusion is valid. However, if the results show the collusion occurs only in an hour, it cannot be considered as a collusion.

REFERENCES

[1] S. Stoft, "Power system economics," *Journal of Energy Literature*, vol. 8, pp. 94-99, Feb. 2002.

[2] U. Helman, "Market power monitoring and mitigation in the US wholesale power markets," *Energy Journal*, vol. 31, no. 6, pp. 877-904, May 2006.

[3] E. S. Amundsen and L. Bergman, "Green certificates and market power on the Nordic power market," *Energy Journal*, vol. 33, no. 2, pp. 101-117, May 2012.

[4] D. P. Brown and D. E. Olmstead, "Measuring market power and the efficiency of Alberta's restructured electricity market: an energy-only market design," *Canadian Journal of Economics*, vol. 50, no. 3, pp. 838-870, Aug. 2017.

[5] F. M. Mirza and O. Bergland, "Market power in Norwegian electricity market: are the transmission bottlenecks truly exogenous?" *The Energy Journal*, vol. 36, no. 4, pp. 27-45, Oct. 2015.

[6] N. Fabra and J. Toro, "Price wars and collusion in the Spanish electricity market," *International Journal of Industrial Organization*, vol. 23, no. 3, pp. 155-181, Apr. 2005.

[7] USAEE Working Paper. (2018, Aug.). Tacit collusion in real-time US electricity auctions. [Online]. Available: https://ssrn. com/abstract=193875USAEEWorkingPaper7

[8] M. Shafie-Khah, M. P. Moghaddam, and M. K. Sheikh-El-Eslami, "Development of a virtual power market model to investigate strategic and collusive behavior of market players," *Energy Policy*, vol. 61, pp. 717-728, Oct. 2013.

[9] M. Shafie-Khah, M. P. Moghaddam, and M. K. Sheikh-El-Eslami, "Ex-ante evaluation and optimal mitigation of market power in electricity markets including renewable energy resource," *IET Generation, Transmission & Distribution*, vol. 10, no. 8, pp. 1842-1852, May 2016.

[10] A. Liu and B. F. Hobbs, "Tacit collusion games in pool-based electricity markets under transmission constraints," *Mathematical Programming*, vol. 140, no. 2, pp. 351-379, Sept. 2013.

[11] G. J. Werden, "Consumer welfare and competition policy," in *Competition Policy and the Economic Approach.* Cheltenham: Edward Elgar Publishing, 2011.

[12] K. Dawar and P. Holmes, "Trade and competition policy," in *The Ashgate Research Companion to International Trade Policy.* Berlin: Springer, 2016.

[13] M. Nazrullslam, S. R. Haque, K. M. Alam *et al.*, "An approach to improve collusion set detection using MCL algorithm," in *Proceedings of International Conference on Computers and Information Technology*, Dhaka, Bangladesh, Dec. 2009, pp. 237-242.

[14] G. K. Palshikar and M. M. Apte, "Collusion set detection using graph clustering," *Data Mining and Knowledge Discovery*, vol. 16, no. 2, pp. 135-164, Apr. 2008.

[15] R. C. Mihailescu and S. Ossowski, "A framework for fraud discovery via illicit agreements in energy markets," *AI Communications*, vol. 28, no. 4, pp. 607-616, Jan. 2015.

[16] D. P. Brown and D. E. H. Olmstead, "Measuring market power and the efficiency of Alberta's restructured electricity market: an energy-only market design," *Canadian Journal of Economics/Revue canadienne d' economique*, vol. 50, no. 3, pp. 838-870, Aug. 2017.

[17] D. E. Aliabadi, M. Kaya, and G. Jahin, "Determining collusion opportunities in deregulated electricity markets," *Electric Power Systems Research*, vol. 141, pp. 432-441, Dec. 2016.

[18] R. Pal and M. Scrimitore, "Tacit collusion and market concentration under network effects," *Economics Letters*, vol. 145, pp. 266-269, Aug. 2016.

[19] N. Hajibandeh, M. Shafie-khah, G. J. Osorio *et al.*, "A new approach for market power detection in renewable-based electricity markets," in *Proceedings of IEEE Environment and Electrical Engineering*, Milan, Italy, Jun. 2017, pp. 1-5.

[20] I. Matsukawa, "Detecting collusion in retail electricity markets: results from Japan for 2005 to 2010," *Utilities Policy*, vol. 57, pp. 16-23, Apr. 2019.

[21] M. J. Zideh and S. Mohtavipour, "Two-sided tacit collusion: another step towards the role of demand-side," *Energies*, vol. 10, no. 12, p. 2045, Dec. 2017.

[22] K. Weng, K. Zhang, J. Geng *et al.*, "The impact of collusion on bilateral negotiation in direct electricity purchase market," in *Proceedings of IEEE International Conference on Power System Technology*, Guangzhou, China, Nov. 2018, pp. 685-692.

[23] Tagungsband Multikonferenz Wirtschaftsinformatik. (2018, Mar.). Assessment of market power in local electricity markets with regards to competition and tacit collusion. [Online]. Available: https://pdfs.semanticscholar.org/3bb3/23edb38c4fed9543e1734246e57e3138127a.pdf

[24] B. F. Hobbs, C. B. Metzler, and J. S. Pang, "Strategic gaming analysis for electric power systems: an MPEC approach," *IEEE Transactions on Power Systems*, vol. 15, no. 2, pp. 638-645, May 2000.

[25] M. O. Buygi, H. Zareipour, and W. D. Rosehart, "Impacts of large-scale integration of intermittent resources on electricity markets: a supply function equilibrium approach," *IEEE Systems Journal*, vol. 6, no. 2, pp. 220-232, Aug. 2012.

[26] P. Simon, *Too Big to Ignore: the Business Case for Big Data.* New York: John Wiley & Sons, Mar. 2013.

[27] R. Kohavi and F. Provost, "Glossary of terms," *Machine Learning*, vol. 30, no. 2, pp. 271-274, Feb. 1998.

[28] S. J. Russell, P. Norvig, J. F. Canny *et al.*, "Artificial intelligence: a modern approach," *The Knowledge Engineering Review*, vol. 11, no. 1, pp. 78-79, Mar. 1996.

[29] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273-297, Sept. 1995.

[30] V. N. Vapnik and V. Vapnik, "An review of statistical learning theory," *IEEE Transactions on Neural Networks*, vol. 11, no. 1, pp. 78-79, Mar. 1996.

[31] K. B. Duan and S. S. Keerthi, "Which is the best multiclass SVM method? an empirical study," in *Multiple Classifier Systems*. Berlin: Springer, 2005.

[32] C. Hsu and C. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Transactions on Neural Networks*, vol. 13, no. 2, pp. 415-425, Aug. 2002.

[33] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, Pittsburgh, USA, Jul. 1992, pp. 144-152.

[34] A. Ben-Hur and J. Weston, "A user's guide to support vector machines," *Data Mining Techniques for the Life Sciences*, vol. 609, pp. 223-239, Oct. 2009.

[35] C. Hsu, C. Chang, and C. J. Lin. (2016, May). A practical guide to support vector classification. [Online]. Available: http://www.csie. ntu. edu.tw/~cjlin/papers/guide/guide.Pdf

[36] L. Rokach and O. Maimon, *Data Mining with Decision Trees: Theory and Applications.* Singapore: World Scientific, Sept. 2014.

[37] L. Breiman, J. H. Friedman, R. A. Olshen *et al*., *Classification and Regression Trees*, *Cole Advanced Books and Software.* Monterey: Wadsworth & Books, 1984.

[38] B. E. Hobbs, "Linear complementarity models of Nash-Cournot competition in bilateral and POOLCO power markets," *IEEE Transactions on Power Systems*, vol. 16, no. 2, pp. 194-202, May 2001.

[39] H. Mardani and D. Souri, "Estimation of marginal cost function using bid price and quantity to power market," in *Proceedings of Electric Power Conference*, Vancouver, Canada, Oct. 2008, pp. 1-6.

[40] E. Maghool, A. Naghavi, and S. F. Ghaderi, "Estimating cost function in power markets under pay-as-bid pricing rules using observed bid data," in *Proceedings of 2008 Electric Power Conference*, Vancouver, Canada, pp. 1-7, Oct. 2008.

[41] P. Twomey, R. J. Green, K. Neuhoff *et al*. (2006, Mar.). A review of the monitoring of market power, the possible roles of TSOS in monitoring for market power issues in congested transmission systems. [Online]. Available: https://doi.org/10.17863/CAM.5068

[42] S. P. Karthikeyan, I. J. Raglend, and D. P. Kothari, "A review on market power in deregulated electricity market," *International Journal of Electrical Power and Energy Systems*, vol. 48, pp. 139-147, Jun. 2013.

[43] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proceedings of International Joint Conference on Artificial Intelligence*, Anchorage, USA, Aug. 1995, pp. 1137-1145.

[44] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: a survey," *ACM Computing Surveys*, vol. 41, no. 3, pp. 137-195, Jul. 2009.

[45] M. Esmalifalak, L. Liu, N. Nguyen *et al*., "Detecting stealthy false data injection using machine learning in smart grid," *IEEE Systems Journal*, vol. 11, no. 3, pp. 1644-1652, Aug. 2017.

[46] F. Murtagh, "Multilayer perceptrons for classification and regression," *Neurocomputing*, vol. 2, no. 5, pp. 183-197, Jul. 1991.

[47] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. New York: John Wiley and Sons, 2012.

[48] D. T. Larose and C. P. Laros, *Discovery Knowledge in Data: an Introduction to Data Mining*. New York: John Wiley & Sons, 2005, pp. 90-106.

[49] N. Friedman, D. Geiger, and M. Goldszmidt, "Bayesian network classifiers," *Machine Learning*, vol. 29, no. 2, pp. 131-163, Nov. 1997.

[50] P. D. Martin, "Evaluation: from precision, recall and f-measure to ROC, informedness, markedness and correlation," *Journal of Machine Learning Technologies*, vol. 2, no. 1, pp. 37-63, Dec. 2011.

Peyman Razmi received the M. Sc. degree in electrical engineering from Ferdowsi University of Mashhad, Mashhad, Iran, in 2015. His main research interests include electricity market, machine learning, convex and stochastic optimization and application of game theory in deregulated power market.

Majid Oloomi Buygi received the Ph. D. degree in electrical engineering from the Darmstadt University of Technology, Darmstadt, Germany, in 2004. From 2005 to 2008, he was the Director of the Electric Power Engineering Department, Shahrood University of Technology, Shahrood, Iran. In 2009, he was with the University of Calgary, Calgary, Canada, as a Postdoctoral Fellow. In 2010, he joined the Ferdowsi University of Mashhad, Mashhad, Iran. He was with the University of Calgary and University of Saskatchewan, Saskatoon, Canada, as a Visiting Professor in 2014 and 2017, respectively. His research interests include power system operation and planning, physical and financial electricity markets, and micro and smart grids.

Mohammad Esmalifalak received the Ph.D. degree from the University of Houston, Houston, USA, in August 2013 and later, worked as a Data Scientist for different industries such as manufacturing or oil and gas. His specialty is predictive maintenance where he is building predictive models for such industries.