

## RESEARCH ARTICLE

# Self-Training Enabled Efficient Classification Algorithm: An Application to Charging Pile Risk Assessment

WEN WANG<sup>1</sup>, XIAOFENG PENG<sup>1</sup>, YE YANG<sup>1</sup>, CHUN XIAO<sup>2</sup>, SHUAI YANG<sup>2</sup>, MINGCAI WANG<sup>1</sup>, LINGFEI WANG<sup>1</sup>, YANLING WANG<sup>3</sup>, LIN LI<sup>1,3</sup>, AND XIAOLIN CHANG<sup>1,3</sup>, (Senior Member, IEEE)

<sup>1</sup>State Grid Electric Vehicle Service Company Ltd., Beijing 100031, China

<sup>2</sup>State Grid Shanxi Marketing Service Center, Taiyuan, Shanxi 030001, China

<sup>3</sup>Beijing Key Laboratory of Security and Privacy in Intelligent Transportation, Beijing Jiaotong University, Beijing 100044, China

Corresponding authors: Wen Wang (wangwen@evs.sgcc.com.cn), Lin Li (lilin@bjtu.edu.cn), and Xiaolin Chang (xlchang@bjtu.edu.cn)

This work was supported by the State Grid Technology Project “Research on Interaction between Large-Scale Electric Vehicles and Power Grid and Charging Safety Protection Technology” from State Grid Corporation of China under Grant 5418-202071490A-0-0-00.

**ABSTRACT** With the continuous development of electric vehicles (EV), large-scale distributed charging piles have been deployed in the wild. Therefore, it is extremely essential to evaluate the risk state of EV charging piles efficiently and effectively. This paper aims to measure the capability of supervised and semi-supervised machine learning techniques in assessing the risk state of EV charging piles. We investigate 8 algorithms, including Support Vector Machine (SVM), Random Forest (RF), Adaptive Boosting (AdaBoost), Gradient Boosting Decision Tree (GBDT), Self-Training based on SVM (ST-SVM), Self-Training based on RF (ST-RF), Self-Training based on AdaBoost (ST-AdaBoost) and Self-Training based on GBDT (ST-GBDT). We first collect data on normal and abnormal termination of charging services from an actual Internet of Vehicles platform. The dataset consists of 17,773 recordings and 7 features generated from the records, which are used for classification. According to the statistical times of 7 features, 20% of recordings are labeled by knowledgeable experts into three classes: low-risk, medium-risk and high-risk. Experimental results indicate that ST-AdaBoost and ST-GBDT show more excellent overall classification performance, compared with the other traditional supervised methods. We also apply ST-GBDT to predict the risk state of the unclassified piles and produce the statistic of piles from different manufacturers.

**INDEX TERMS** AdaBoost, charging piles, gradient boosting decision tree, risk assessment, self-training.


## I. INTRODUCTION

To control the global temperature increasing range below 2°C, annual energy-related CO<sub>2</sub> emissions still need to decline by 2050 from 35 Gt to 9.7 Gt, a fall of more than 70%. The imperative to reduce the emission of carbon dioxide and achieve sustainable growth is strengthening the momentum of the global energy transition. Renewable energy and energy efficiency are the main two pillars of energy transition [1]. As a vehicle driven by renewable electric energy, electric vehicles (EV) with the advantage of environment-friendliness

and energy efficiency is considered to replace traditional fuel vehicles [2].

With the increasing number of EVs, a large number of distributed charging piles are being one of the most essential infrastructures [3]. Charging piles are mostly deployed in the wild with uncontrollable environmental factors, causing frequent charging faults. Therefore, effective analysis of charging safety and comprehensive assessment of charging piles have become practical problems [4], [5].

For charging safety, existing researches focus on evaluating the state of charging piles according to specific indexes but fail to provide a feasible index system to evaluate the long-term operation of charging service providers. In the assessment-index system for the electrical safety

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

performance of EV charging equipment proposed in [6], many internal factors are considered like the contact current, insulation resistance, and impulse withstand voltage. However, it is difficult, if it is not impossible, to distinguish between charging equipment that has the same internal factors but has different actual performance. Since charging service providers always purchase large quantities of charging piles from one manufacturer, the internal factors of these piles are much the same. Furthermore, the raw data on internal factors is hard to obtain for the providers. Li *et al.* [7] took the failure rate of charging piles into account when establishing the integrated safety-assessment-index system. The failure rate can reflect the actual operation of charging piles, but it ignores the different frequency and risk degree of each type of failure because failure rate takes the total failure rate as an evaluation metric. It is critical to establish an effective evaluation system focusing on the long-term safety performance of charging piles, which can provide suggestions on purchasing, maintaining and managing charge piles,

As for assessment methods, Wei *et al.* [5], [6], [7] used the analytic hierarchy process (AHP [8]) to calculate the evaluation metric weights depending on expert experience. However, when there are many elements in the same hierarchy, subjective evaluation becomes vague and untrustworthy, and then the judgment matrix is prone to serious inconsistency. In recent years, Machine Learning (ML) methods have been becoming a research hotspot to state assessments in many fields. Mangalathu *et al.* [9] proposed a methodology for the rapid damage state assessment (green, yellow, or red) of bridges utilizing various classification algorithms such as K-nearest neighbors, random forests, and naïve Bayes. Chen *et al.* [10] applied SVM to classify the risk levels of customers. The inputs of the SVM classifier include nine assessment features and the output is the risk level for customers. Although there are various applications of ML technology in risk state assessment, there is no such solution for the risk state of charging piles.

In this paper, we propose to transform the risk assessment task of charging piles into a classification task in order to provide an effective approach for evaluating the long-term risk state of charging piles. We investigate the capability of various supervised and semi-supervised ML algorithms in discriminating risk states into three categories (low-risk, medium-risk, and high-risk).

FIGURE 1 presents a simplified workflow of our paper, from which we summarize the following four main contributions:

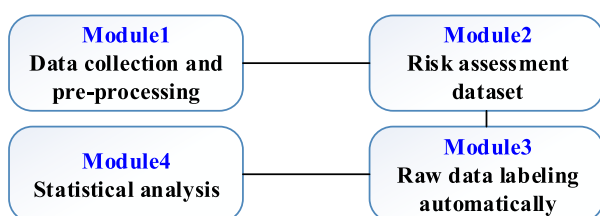


FIGURE 1. Workflow of this paper.

- We collect original long-term operation data of an actual Internet of Vehicles platform from June to December 2021 and make pre-processing by removing data records with missing critical information (**Module1** in FIGURE 1).
- We establish an assessment-index system by defining risk levels of charging faults and construct the structured dataset of risk assessment with manual expert knowledge (**Module2** in FIGURE 1). **Section III** shows the details of building the dataset.
- We investigate the capability of ML algorithms in automatically labeling a large amount of raw data with little expert knowledge (**Module3** in FIGURE 1). Specifically, we compare the performance of 8 representative algorithms, including Support Vector Machine (SVM), Random Forest (RF), Gradient Boosting Decision Tree (GBDT), Adaptive Boosting (AdaBoost), self-training based on SVM (ST-SVM), self-training based on RF (ST-RF), self-training based on GBDT (ST-GBDT), and self-training based on Ada-Boost (ST-AdaBoost). We compare the classification performance of each model, and we observe that ST-AdaBoost and ST-GBDT perform the best. Hence, we apply ST-GBDT to predict the risk state of the unclassified piles (see **Section IV.C** and **D**).
- We carry out extensive statistical analysis to evaluate the overall long-term risk state of charging piles manufactured by various companies (**Module4** in FIGURE 1). To the best of our knowledge, we are the first to analyze the overall long-term risk state and fault rate of charging piles (see **Section IV.E**). Notably, the existing researches only focus on internal factors, such as contact current, insulation resistance, and impulse withstand voltage. They fail to reflect the actual long-term performance such as risk state and fault rate.

The rest of the paper is organized as follows. Section II presents the overview of classification algorithms. Section III presents the construction of the dataset and the labeling. Experiments and discussions are given in Section IV and V. Section VI gives the conclusion.

## II. OVERVIEW OF CLASSIFICATION ALGORITHMS

ML techniques can be classified into supervised learning, semi-supervised learning, and unsupervised learning. Semi-supervised learning is a method combining supervised learning and unsupervised learning. Its main idea is to use a small amount of labeled data to predict the class of unlabeled data, and merge it into a labeled dataset [11]. This paper adopted SVM, RF, GBDT, AdaBoost, and a semi-supervised approach using self-training. A brief overview of various algorithms is provided in this section.

### A. SUPPORT VECTOR MACHINE

Based on structural risk minimization principles, SVM can handle problems of multi-category [12]. It aims at

maximizing the margin between the two sides of a separating hyperplane, of which either side separates two data classes. When dealing with linearly separable data, SVM is unaffected by the number of features. Nevertheless, SVM may not be able to find a hyperplane successfully when the training set involves non-separable instances. The problem can be addressed by mapping the data onto a feature space with higher dimensions, where the optimum separating hyperspace can be found [13].

**B. RANDOM FOREST**

RF is an ML algorithm that integrates multiple decision trees based on the idea of ensemble learning. It takes advantage of bagging and random feature selection. Random forest uses bootstrap to extract multiple samples from the original data set, trains the extracted samples with a decision tree, and then combines these decision trees to obtain the final prediction results through majority voting [14]. The steps involved in RF are as follows:

- (1) Generate  $N$  samples by bootstrap from the training set.
- (2) Randomly select a subset of all features and obtain the best split point by generating a decision tree from  $N$  bootstrap samples.
- (3) Repeat the two steps above  $M$  times to generate  $M$  decision trees.
- (4) Combine the predicted output of each decision tree and predict the output of the testing set.

**C. AdaBoost**

AdaBoost is one of the most excellent ensemble methods. It has a solid theoretical basis and has made great success in practical applications. In the iteration of Ada-Boost, a new weak classifier is added in each round until a predetermined small error rate is reached. Each training sample is assigned a weight indicating the probability that it is selected into the training set by a classifier. If a sample has been accurately classified, its probability of being selected in the construction of the next training set will be reduced. Conversely, if a sample point is not accurately classified, its weight is increased. In this way, Ada-Boost concentrates on samples that are difficult to classify. Ada-Boost is sensitive to noisy data and abnormal data, it rarely over-fits compared with most other classification algorithms [15].

**D. GBDT**

GBDT is a Boosting algorithm proposed by Friedman in 2001 [16]. Composed of multiple decision trees, GBDT is trained in sequence, and the conclusion of all trees adds up to the final answer. The diagram of GBDT is shown in FIGURE 2. It can be noticed that the residual of the previous decision tree is taken as the input for the next decision tree, which is trained by following the negative gradient direction of the previous decision tree.

**E. SELF-TRAINING**

Self-training is a widely used method of semi-supervised learning. The training process is shown in FIGURE 3. First,

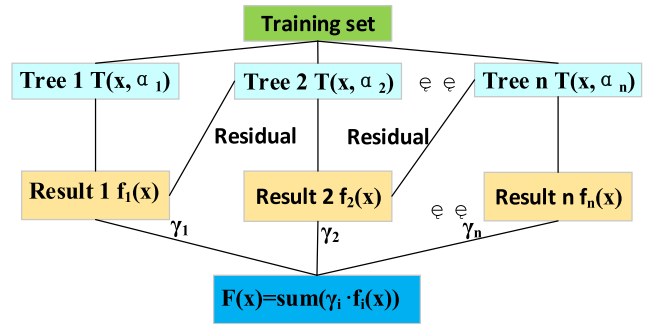


FIGURE 2. Diagram of GBDT algorithm.

a small amount of labeled data samples is applied to train an original classifier. Then the original classifier is used to constantly predict labels for unlabeled data samples. Next, the self-training model selects the most accurate unlabeled data samples according to a threshold and merges them into the training set. The training set is constantly updated and the classifier is retrained until the iteration termination condition is satisfied. Finally, a final classifier with high classification accuracy and strong generalization is obtained [11].

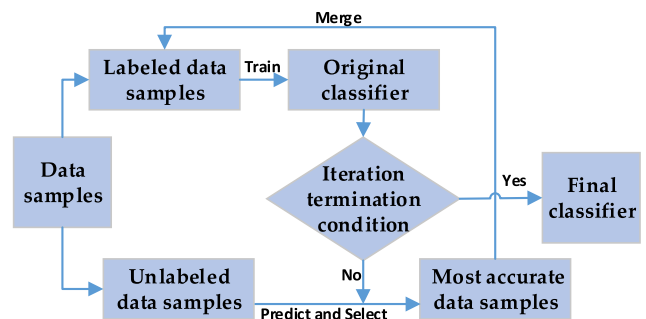


FIGURE 3. The training process of the self-training model.

**III. BUILDING A RISK ASSESSMENT DATASET**

In this section, a dataset consisting of 17,773 recordings and 7 features is generated from records of normal and abnormal termination of charging services. When a charging pile stops providing service normally or abnormally, the reason for terminating the charging service is recorded, which can reflect the state of the piles. We analyze and process the records collected from an actual Internet of Vehicles platform from June to December 2021.

A flowchart of building the risk assessment dataset is shown in FIGURE 4. The outline is given as follows.

*Step 1:* Separate the records with normal termination reasons from abnormal reasons and count the number of records for each pile as effective service times.

*Step 2:* Define three risk levels of faults from various abnormal reasons. We remove the records caused by low-risk faults and concentrate on medium-risk and high-risk faults.

*Step 3:* Classify the medium-risk and high-risk faults into mechanical faults, electrical faults, and other faults.

*Step 4:* Define six assessment indexes of charging piles through the combination of three fault categories and two

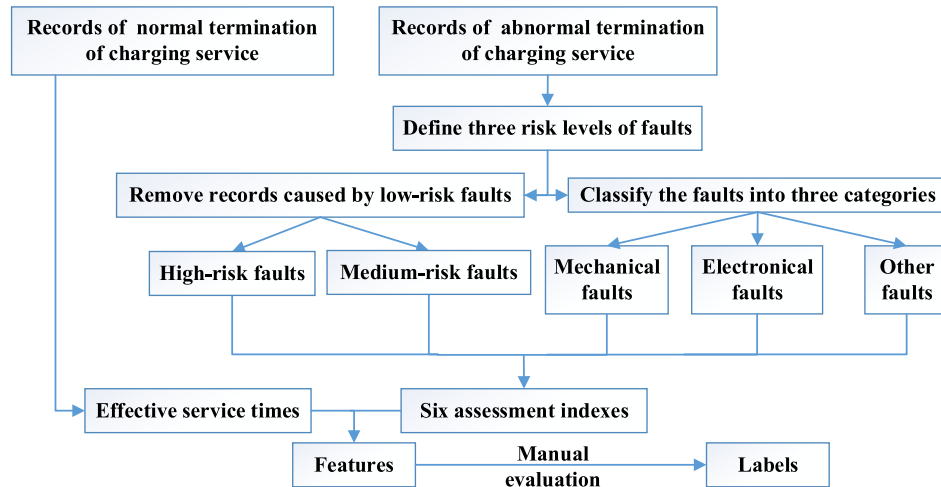


FIGURE 4. Flowchart of building the risk assessment dataset.

risk levels. In addition, we add effective service times into assessment-index system. Hence, we establish a risk-assessment-index system consisting of seven features.

Step 5: By counting the features of each charging pile, a structured dataset is obtained and 20% of recordings in the dataset are selected for being labelled by knowledgeable expert, denoted as manual evaluation in this paper.

TABLE 1. Risk level definition of charging pile faults.

Risk Level	Definition
Low-risk (Level I)	The fault has small or no impact on charging safety and can always be automatically recovered.
Medium-risk (Level II)	The fault has some influence on charging safety, resulting in failure to provide charging service and some maintenance costs.
High-risk (Level III)	The fault has a great impact on charging safety and cannot provide charging service. Maintenance costs are high and timely maintenance is required.

A. RISK ASSESSMENT INDEXES OF CHARGING PILE

To establish a risk-assessment-index system of charging piles, first, the number of records for normal termination reasons is counted as a metric, namely effective service times. Then, 34 types of faults in all records are divided into three risk degrees for refining evaluation metrics. Due to the lack of fault risk levels definition standards for these faults, three fault risk levels are defined based on the impact degree of faults on charging safety and the cost of maintenance. The risk level definitions are shown in TABLE 1.

Owing to the small impact of low-risk faults, we remove records caused by 7 types of low-risk faults, including offline shutdown conditions are reached, the charging gun is not properly inserted, etc. The evaluation metrics mainly originate from medium-risk and high-risk faults.

Among the remaining 27 types of faults, each type of fault may occur in the long-term operation of charging piles, and the occurrence frequency of some high-risk faults is relatively low. For instance, AC circuit breaker failure occurred

TABLE 2. Common faults were classified according to three categories and three risk levels.

Fault Category	Risk Level	The Fault of Charging Pile
Common Faults	Level I	Offline shutdown conditions are reached; The charging gun is not properly inserted; The device is not connected to the vehicle; Web view stops loading; Timeout happens when calibrating the clock; Time inconsistency happens when inserting the charging gun; The device is providing charging service (7 types)
	Level II	Charging connection failure; Earth wire fault; Emergency stop button is pressed; Output contactor fault; Abnormal charging port; Arrester fault; Frequent switch of S2 (6 types)
Mechanical Faults	Level III	AC contactor fault (1 type)
	Level II	CP circuit overvoltage; AC input voltage is under-voltage (2 types)
	Level III	Input power failure (Overvoltage, overcurrent, under-voltage, tripping); Output short circuit fault; AC circuit breaker failure; AC input voltage is overvoltage (4 types)
Electrical Faults	Level II	Startup charge timed out; Leakage protection failure; Communication failure of electricity meter; Abnormal electricity meter data; Device self-check failure; PWM control output overcurrent fault; Output current overrated fault; Charging device over-temperature alarm; Failed to execute the remote power allocation policy; Internal communication fault (10 types)
	Level III	Unavailable device; Power failure of control loop; Charging port over-temperature; Charging device over-temperature fault (4 types)

284 times and power failure of the control loop occurred 53 times for all charging piles during 7 months. That is, their influence on charging safety cannot be ignored, and the frequency of these failures is essential for evaluating the risk state of charging piles. Although the frequency of these faults is not suitable as a single input to the ML model, the

problem can be solved by classifying 27 types of faults into three categories: mechanical faults, electrical faults, and other faults. Namely, as shown in TABLE 2, 27 types of common faults were classified into level II mechanical fault, level III mechanical fault, level II electrical fault, level III electrical fault, level II other faults, and level III other faults.

**B. MANUAL EVALUATION**

By counting 7 features of each charging pile, a structured dataset is obtained. The dataset is sufficiently shuffled, and 20% of data records are selected randomly for manual evaluation.

Manual comprehensive analysis is conducted to judge the risk grade of charging piles by experienced professionals. In general, there are three risk grades for charging piles:

- The data samples of charging piles at low-risk grade, with health status and having consumed few maintenance costs, were labeled as 2.
- The data samples of charging piles at medium-risk grade, with unhealthy status and having caused some maintenance costs, are labeled as 1.
- The data samples of charging piles at high-risk grade, with poor status and having consumed large maintenance costs, are labeled as 0.

Statistical results of manual labels are shown in TABLE 3. And in TABLE 4, we present one example of a labeled data sample in the risk assessment dataset. Pile number, 7 features, and the label are used to train the ML classifiers. The work in this section provided sufficient data samples for the learning task to train and test the risk assessment models.

**TABLE 3. Statistical results of three types of labels.**

Label	Risk Grade	Number of Data Sample
0	High-risk	35
1	Medium-risk	332
2	Low-risk	3247

**TABLE 4. One example of labeled data sample in the risk assessment dataset.**

Data item	Data sample
Pile number	001050101021000000127516
Effective service times	75
Mechanical fault II times	1
Mechanical fault III times	0
Electrical fault II times	4
Electrical fault III times	0
Other fault II times	4
Other fault III times	2
Label	2

**IV. EXPERIMENTAL RESULTS & DISCUSSION**

In this section, the performance measurement metrics are given. Then, the dataset splitting and oversampling are presented. After then, the environment and hyper-parameters are given. At last, we conduct extensive experiments to provide statistical and experimental support for analysis.

**A. PERFORMANCE MEASUREMENT**

In the field of ML algorithms, researchers often use precision, recall, accuracy, F1-score, and AUROC (area under the ROC curve) as performance metrics in classification experiments [17], [18]. In this experiment, there is a big difference in the number of data samples for three classes, which results in an unbalanced dataset. Therefore, macro averaged scores across categories are calculated to average the scores of all three binary tasks. The macro averaged scores are called macro averaged recall, precision, and F1-score respectively. Besides, AUC was introduced which stands for ‘‘Area under the precision-recall (PR) curve’’, for the PR curve is more informative than ROC when evaluating classifiers on unbalanced datasets. The ROC curve can reflect the comprehensive performance of the classifier. However, we pay attention to the model classification performance on unbalanced data. It’s more appropriate to use the PR curve which is more sensitive to minority classes [19].

Regard class  $i$  ( $i = 0, 1, 2$ ) as a positive category and all other classes as a negative category. For each sample in the dataset, there are four possible partitioning outcomes:

TP <sub>$i$</sub>  (True Positive): Number of samples belonging to and classified as a positive class;

FP <sub>$i$</sub>  (False Positive): Number of samples belonging to a negative category and classified as a positive category;

FN <sub>$i$</sub>  (False Negative): Number of samples belonging to a positive category and classified as a negative category;

TN <sub>$i$</sub>  (True Negative): Number of samples in the negative category and classified as negative.

The subscript in the formulas above represents the label, taking 0, 1 or 2. And then the precision, recall, and F1-score of each class can be calculated respectively as follows:

$$\text{precision}_i = \frac{TP_i}{TP_i + FP_i} \tag{1}$$

$$\text{recall}_i = \frac{TP_i}{TP_i + FN_i} \tag{2}$$

$$F1_i = \frac{2 \cdot \text{precision}_i \cdot \text{recall}_i}{\text{precision}_i + \text{recall}_i} \tag{3}$$

Macro averaged recall, precision and F1-score are defined as follows:

$$\text{precision}_{\text{macro}} = \frac{\sum_{i=0}^{n-1} \text{precision}_i}{n} \tag{4}$$

$$\text{recall}_{\text{macro}} = \frac{\sum_{i=0}^{n-1} \text{recall}_i}{n} \tag{5}$$

$$F1_{\text{macro}} = \frac{2 \cdot \text{precision}_{\text{macro}} \cdot \text{recall}_{\text{macro}}}{\text{precision}_{\text{macro}} + \text{recall}_{\text{macro}}} \tag{6}$$

**B. DATASET SPLITTING AND OVERSAMPLING**

In this experiment, 3614 labeled data samples are randomly shuffled, and divided into the training set and testing set on a scale of 1:1, which is shown in TABLE 5. To compare the performance of four supervised learning models and four self-training models, every model is tested on the same testing set. Only the training set is applied to

**TABLE 5. Splitting result of dataset.**

	0	1	2	Sum
Training set	16	161	1630	1807
Testing set	19	171	1617	1807
Sum	35	332	3247	3614

**TABLE 6. Environment and corresponding libraries.**

Environment	Libraries
Python 3.7.5	Pandas 1.0.1 Imbalanced-learn 0.9.0 Scikit-learn 1.0.2

**TABLE 7. Hyper parameters.**

Classifier	Parameters
SVM	C = 10, kernel = 'rbf', gamma='auto', tol=1e <sup>-3</sup> , and probability = True
RF	n_estimators = 20, criterion='gini' and max_depth = 20
GBDT	n_estimators = 88, learning_rate = 0.8, and max_depth = 20
AdaBoost	base_estimator = DecisionTreeClassifier (max_depth = 20), algorithm = "SAMME", learning_rate = 1, and n_estimators = 200
Self-Training	criterion='threshold', threshold = 0.7, and max_iter = 100

**TABLE 8. Performance metrics across various models.**

Algorithms	Accuracy	Precision <sub>macro</sub>	Recall <sub>macro</sub>	F1 <sub>macro</sub>
SVM	0.872	0.623	0.760	0.664
ST-SVM	0.901	0.701	0.836	0.753
RF	0.889	0.675	0.674	0.671
ST-RF	0.987	0.979	0.971	0.975
GBDT	0.893	0.652	0.665	0.658
<b>ST-GBDT</b>	<b>0.995</b>	<b>0.985</b>	<b>0.996</b>	<b>0.990</b>
AdaBoost	0.893	0.671	0.653	0.661
<b>ST-AdaBoost</b>	<b>0.995</b>	<b>0.993</b>	<b>0.988</b>	<b>0.990</b>

the supervised learning model. As for the semi-supervised learning model, more than 10,000 data samples are used as unlabeled data in the process of training, in addition to the training set.

An unbalanced dataset may be able to cause the prediction of the minority class to be difficult and imprecise [20]. To alleviate this problem, the data samples of minority class is up-sampled by random oversampling, which can make the training set more balanced, and improve the model performance on minority classes.

### C. EXPERIMENTAL ENVIRONMENT AND PARAMETERS

In experiments, we use python of version 3.7.5 as the experimental platform, and several essential libraries are utilized. More details about the dependency of the experiment and hyper-parameter of classifiers are presented in TABLE 6 and TABLE 7. In fact, four self-training classifiers are respectively built just on the four supervised models with the same additional parameters, including criterion, threshold, and max\_iter.

### D. RESULTS OF CLASSIFICATION EXPERIMENTS

To compare the performance of supervised and semi-supervised models, first, we apply oversampled training set to supervised classifiers namely SVM, RF, GBDT, and Ada-Boost. Then we add 14,159 data samples without labels to train four semi-supervised models based on different original classifiers, including ST-SVM, ST-RF, ST-AdaBoost, and ST-GBDT.

TABLE 8 depicts the accuracy and macro averaged scores achieved by using each algorithm. The following conclusion can be drawn as follows:

- The accuracy of SVM, RF, GBDT, and AdaBoost is about 0.90, but the macro average precision, recall, and F1-score are poor. It is the abundant data of low-risk charging piles in the training set that makes the model properly classify the low-risk charging piles in the testing set. However, randomly oversampling the minority class hasn't effectively improved the classification performance of the supervised model on medium-risk and high-risk data samples, leading to the low macro average metrics.
- Self-training outperforms the supervised model in terms of accuracy, macro averaged precision, recall, and F1-score. It is because unlabeled data samples provide some insights about charging piles at different risk grades which are exploited during training.

FIGURE 5 compares the PR curve and AUC obtained with the semi-supervised approach using self-training and supervise approaches using SVM, RF, GBDT, and AdaBoost. It can be observed in the figure that the self-training model takes advantage of unlabeled data samples to improve AUC greatly in classifying charging piles at medium-risk grades and high-risk grades. It is concluded that ST-GBDT and ST-AdaBoost algorithms perform better as compared to all other algorithms.

### E. COMPARING STATISTICS OF MANUFACTURERS

We apply ST-GBDT to predict the risk grade of unlabeled data samples and then combine the predicted pseudo labels with real labels to obtain a mixed dataset. Based on the dataset, we analyze the statistics of risk assessment results of charging piles produced by different manufacturers.

FIGURE 6 shows the percentage of charging piles from different manufacturers in the dataset, and the statistic of charging piles at three risk grades. The piles come from five manufacturers, denoted by C1, C2, C3, C4, and C5. We can observe that both high-risk and medium-risk charging piles assessed are mainly produced by C4. Specifically, 63% of charging piles at high-risk grades are produced by C4, and charging piles produced by C4 account for 81% of all medium-risk piles, both ranking first. Nevertheless, charging piles produced by C4 account for 43%, much smaller than the two percentages.

$$\text{Fault Rate} = \frac{\#\text{Fault Times}}{\#\text{Effective Service Times} + \text{FaultTimes}} \quad (7)$$

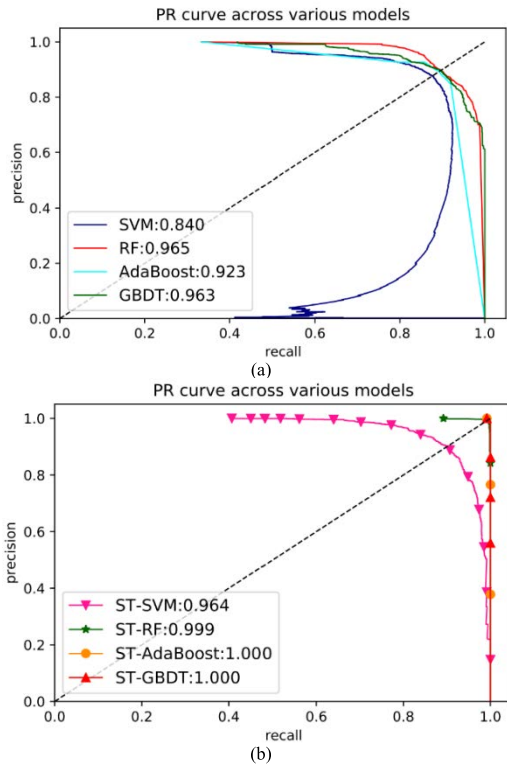


FIGURE 5. AUC and PR curve comparison of self-training and supervised classifiers. (a) compares the AUC and PR curve of various supervised approaches; (b) compares the AUC and PR curve of self-training models based on various original classifiers.

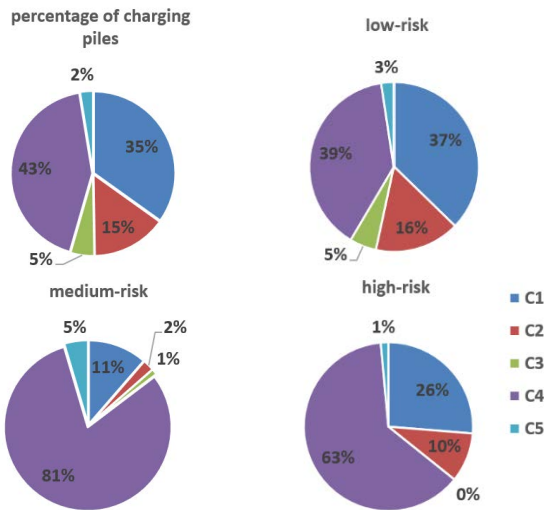


FIGURE 6. Percentage of charging piles produced by five manufacturers.

To further compare the overall state of different manufacturers, we calculate the average effective service times and fault times for every manufacturer. And the fault rate of each charging pile is calculated by equation (7). Then the average fault rate of charging piles for every manufacturer can be obtained. All the results are shown in TABLE 9, and it is presented that the average mechanical fault times for C4 is about 31.834, obviously exceeding other manufacturers, even

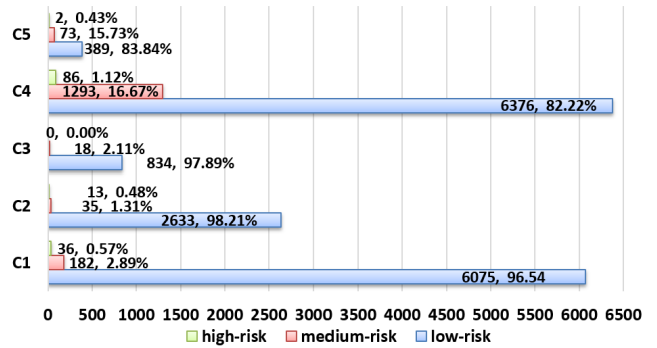


FIGURE 7. Percentage of charging piles at different risk grades produced by five manufacturers.

TABLE 9. Average effective service times, fault times, and fault rate.

Avg	C1	C2	C3	C4	C5
Electrical fault II times	0.001	0.040	0.918	0.002	0
Electrical fault III times	4.932	2.519	0.381	3.577	6.155
Other fault II times	0.384	3.560	1.771	0.167	1.097
Other fault III times	0.004	0.644	0	0.001	0
Mechanical fault II times	8.158	1.459	5.661	31.834	22.899
Mechanical fault III times	0	0	0.034	0	0
Fault time	13.480	<b>8.256</b>	8.731	35.581	30.151
Fault rate	0.161	<b>0.116</b>	0.120	0.400	0.493
Effective service times	<b>90.144</b>	73.382	65.281	35.870	28.052

\*C1-C5 denote five cities.

matching its average effective service times, and resulting in a high average fault rate. We speculate that frequent mechanical failures are responsible for the relatively worse state of piles from C4.

Among all manufacturers, the majority of charging piles produced by C1, C2, and C3 are at the low-risk grade, accounting for more than 95%, which is shown in FIGURE 7. It is shown in TABLE 9 that the average effective service times of the three manufacturers are high, exceeding 65, and especially C1 reaches about 90. The average fault times and fault rate of the three manufacturers are low, among them, Evergrande has the smallest values, which reflects the satisfactory safety state of the charging piles produced by these three manufacturers from another perspective.

### V. ANALYSIS

Different from existing works, the data in our work is not internal factor data of charging piles used for factory inspection, nor real-time operation data for real-time risk monitoring. To improve economic benefits, charging service companies need to know the long-term operation and performance of charging piles. Therefore, our work is based on the charging service records accumulated by charging piles for more than half a year, which is helpful to evaluate the long-term operation status and performance of charging piles.

On the one hand, our results can provide advice for charging service companies to purchase and maintain charging piles. In detail, when replacing or repairing old charging piles in large quantities, the scope of investigation and troubleshooting can be greatly reduced, and when purchasing new charging piles in large quantities, some manufacturers are preferable.

On the other hand, there're several limitations to our work. From the perspective of data, our risk assessment-index system relies solely on termination records of charging service, without considering other possible accumulated risk factors and ignoring the different aging degrees of charging piles. Moreover, due to the lack of dates in these records, we fail to process multiple abnormal termination records caused by the same fault in a short period. From the perspective of assessment methods, our work only discusses the applicability of a few machine learning algorithms on the established dataset. Actually, providing a more complicated assessment dataset including unstructured data, the current optimal method may not be applicable. Another drawback is that our approach must rely on expert evaluation, and thus it's hard to carry out the work without expert knowledge.

## VI. CONCLUSION

This paper mainly studies the applicability of ML algorithms in the evaluation system of electric vehicle (EV) charging piles. For this purpose, based on establishing a feasible risk assessment-index system for long-term operation and performance, we build a risk assessment dataset and select a small part of data samples to be labeled. One-half of artificial comprehensive evaluation results are used to train various supervised classification models and semi-supervised models, including SVM, RF, GBDT, AdaBoost, ST-SVM, ST-RF, ST-GBDT, and ST-AdaBoost. Among these classification models, ST-GBDT and ST-AdaBoost classify charging piles at different risk grades with the highest macro averaged recall, precision, and F1-score, which show ideal performance for assessing risk state. Self-training algorithm performs better than representative supervised algorithms, especially in classifying high-risk and medium-risk charging piles, due to taking advantage of unlabeled data samples.

After predicting pseudo labels for unlabeled data samples using ST-GBDT, we combine the pseudo labels with the artificial evaluation results and conducted statistical analysis for charging piles produced by different manufacturers. In our statistical results, there exist obvious differences in the overall state of charging piles among different manufacturers.

## REFERENCES

- [1] *Global Energy Transformation A Roadmap to 2050*, Int. Renew. Energy Agency, IRENA, Abu Dhabi, UAE, 2018.
- [2] Y. Yu, Q. Li, W. Chen, B. Su, J. Liu, and L. Ma, "Optimal energy management and control in multimode equivalent energy consumption of fuel cell/supercapacitor of hybrid electric trabisbite," *IEEE Trans. Ind. Electron.*, vol. 66, no. 8, pp. 6065–6076, Aug. 2019.
- [3] *National Development and Reform Commission, Electric Vehicle Charging Infrastructure Development Guide (2015-2020)*, China, 2015. [Online]. Available: [http://www.gov.cn/zhengce/2015-10/09/content\\_5076250.htm](http://www.gov.cn/zhengce/2015-10/09/content_5076250.htm)

- [4] L. Zhang, T. Gao, G. Cai, and K. L. Hai, "Research on electric vehicle charging safety warning model based on back propagation neural network optimized by improved gray wolf algorithm," *J. Energy Storage*, vol. 49, May 2022, Art. no. 104092.
- [5] S.-Y. Wei, Q. Zhu, X.-M. Li, and X.-H. Meng, "Research on comprehensive evaluation of electric vehicle charging failure," in *Proc. 6th Int. Conf. Intell. Comput. Signal Process. (ICSP)*, Apr. 2021, pp. 1255–1259.
- [6] K. Zhang, Z. Yin, X. Yang, Z. Yan, and Y. Huang, "Quantitative assessment of electric safety protection for electric vehicle charging equipment," in *Proc. Int. Conf. Circuits, Devices Syst. (ICDS)*, Sep. 2017, pp. 89–94, doi: [10.1109/ICDS.2017.8120457](https://doi.org/10.1109/ICDS.2017.8120457).
- [7] P.-C. Li, Y.-L. Zhao, J. Yang, R. Xu, Q.-Y. Zhang, and L.-Y. Ying, "A design method of EV charging security early warning mode," in *Proc. 36th Chin. Control Conf. (CCC)*, Jul. 2017, pp. 9326–9330.
- [8] T. L. Saaty, "Axiomatic foundation of the analytic hierarchy proces," *Manage. Sci.*, vol. 32, no. 7, pp. 841–855, Jul. 1986.
- [9] S. Mangalathu, S.-H. Hwang, E. Choi, and J.-S. Jeon, "Rapid seismic damage evaluation of bridge portfolios using machine learning technique," *Eng. Struct.*, vol. 201, Dec. 2019, Art. no. 109785.
- [10] T.-H. Chen, "Do you know your customer? Bank risk assessment based on machine learning," *Appl. Soft Comput.*, vol. 86, Jan. 2020, Art. no. 105779.
- [11] A. E. Mehyadin and A. M. Abdulazeez, "Classification based on semi-supervised learning: A review," *Iraqi J. Comput. Inform.*, vol. 47, no. 1, pp. 1–11, 2021.
- [12] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995, doi: [10.1007/BF00994018](https://doi.org/10.1007/BF00994018).
- [13] S. B. Kotsiantis and I. P. Z. Pintelas, "Supervised machine learning: A review of classification techniques," *Emerg. Artif. Intell. Appl. Comput. Eng.*, vol. 160, no. 1, pp. 3–24, 2007.
- [14] W.-Y. Loh, "Classification and regression trees," *Wiley Interdiscipl. Rev., Data Mining Knowl. Discovery*, vol. 1, no. 1, pp. 14–23, 2011.
- [15] G. Rätsch, T. Onoda, and K.-R. Müller, "Soft margins for AdaBoost," *Mach. Learn.*, vol. 42, no. 3, pp. 287–320, 2001.
- [16] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Statist.*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001.
- [17] M. Sokolova, N. Japkowicz, and S. Szpakowicz, "Beyond accuracy, F-score and ROC: A family of discriminant measures for performance evaluation," in *Proc. Australas. Joint Conf. Artif. Intell.* Berlin, Germany: Springer, Dec. 2006, pp. 1015–1021.
- [18] A. P. Bradley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms," *Pattern Recognit.*, vol. 30, no. 7, pp. 1145–1159, 1997.
- [19] J. Davis and M. Goadrich, "The relationship between precision-recall and ROC curves," in *Proc. 23rd Int. Conf. Mach. Learn. (ICML)*, 2006, pp. 233–240.
- [20] G. E. Batista, A. C. Carvalho, and M. C. Monard, "Applying one-sided selection to unbalanced datasets," in *Proc. Mexican Int. Conf. Artif. Intell.* Berlin, Germany: Springer, Apr. 2000, pp. 315–325.
- [21] Q. Ge, H. Qiao, C. Li, Q. Yang, and H. Jiang, "Real-time charging risk assessment for electric vehicles based on improved broad BP-AHP," *IEEE Trans. Ind. Electron.*, vol. 69, no. 9, pp. 9472–9482, Sep. 2022.
- [22] F.-Y. Qin, Z.-Q. Lv, D.-N. Wang, B. Hu, and C. Wu, "Health status prediction for the elderly based on machine learning," *Arch. Gerontol. Geriatrics*, vol. 90, Sep. 2020, Art. no. 104121.
- [23] I. Shcherbatov, E. Lisin, A. Rogalev, G. Tsurikov, M. Dvořák, and W. Strielkowski, "Power equipment defects prediction based on the joint solution of classification and regression problems using machine learning methods," *Electronics*, vol. 10, no. 24, p. 3145, Dec. 2021.
- [24] H. Gao, B. Zang, L. Sun, and L. Chen, "Evaluation of electric vehicle integrated charging safety state based on fuzzy neural network," *Appl. Sci.*, vol. 12, no. 1, p. 461, Jan. 2022.



**WEN WANG** is currently the Deputy General Manager of State Grid Electric Vehicle Service Company. He is a Professorial Senior Engineer; the Expert of China National Key Research and Development Program Review Group; the Secretary-General of IEEE PES EV Satellite Committee, China; a Senior Member of the National Electrical System Security Protection Expert Group; and a Specialist in the field of power system automation, power trading, and information security.



**XIAOFENG PENG** is currently a Senior Engineer and the Head of the V2G Department with State Grid EV Service Company Ltd. His research interests include V2G and load aggregation technology.



**YE YANG** is currently a Senior Engineer and a Research and Development Scientist with State Grid EV Service Company Ltd. His research interests include AI, block-chain, and smart grid control technology.



**YANLING WANG** is currently pursuing the master's degree with the School of Computer and Information Technology, Beijing Jiaotong University. Her research interests include machine learning and differential privacy.

**CHUN XIAO** is currently a Senior Engineer with State Grid Shanxi Marketing Service Center and the Specialist in marketing service.

**SHUAI YANG** is currently a Senior Engineer with State Grid Shanxi Marketing Service Center and the Specialist in marketing service.

**MINGCAI WANG** is currently a Senior Engineer with State Grid Electric Vehicle Service Company Ltd. He is the Specialist in the field of power system automation.

**LINGFEI WANG** is currently a Senior Engineer with State Grid Electric Vehicle Service Company Ltd. He is the Specialist in power trading and control of electric power systems.



**LIN LI** is currently an Associate Professor with the School of Computer and Information Technology, Beijing Jiaotong University. Her current research interests include cryptographic protocols, privacy preserving, and federated learning.



**XIAOLIN CHANG** (Senior Member, IEEE) is a Professor with the School of Computer and Information Technology, Beijing Jiaotong University. Her current research interests include edge/cloud computing, network security, and security and privacy in machine learning.

...