

Generative adversarial networks and convolutional neural networks based weather classification model for day ahead short-term photovoltaic power forecasting

Fei Wang^{1,2,3}, Zhanyao Zhang², Chun Liu⁴, Yili Yu², Songling Pang⁵, Neven Duić⁶, Miadreza Shafie-khah⁷, João P. S. Catalão^{8,9,10}

1. State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources (North China Electric Power University), Baoding 071003, China
2. Department of Electrical Engineering, North China Electric Power University, Baoding 071003, China
3. Hebei Key Laboratory of Distributed Energy Storage and Micro-grid, North China Electric Power University, Baoding 071003, China
4. State Key Laboratory of Operation and Control of Renewable Energy & Storage Systems, China Electric Power Research Institute, Beijing 100192, China
5. Electric Power Research Institute, Hainan Power Grid Co., Ltd, Haikou 570311, China
6. University of Zagreb, Faculty of Mechanical Engineering and Navala Architecture, Ivana Lučića 5, 10000 Zagreb, Croatia
7. INESC TEC, Porto 4200-465, Portugal
8. Faculty of Engineering of the University of Porto and INESC TEC, Porto 4200-465, Portugal
9. INESC-ID, Instituto Superior Técnico, University of Lisbon, Lisbon 1049-001, Portugal
10. C-MAST, University of Beira Interior, Covilhã 6201-001, Portugal

Abstract: Accurate solar photovoltaic power forecasting can help mitigate the potential risk caused by the uncertainty of photovoltaic out power in systems with high penetration levels of solar photovoltaic generation. Weather classification based photovoltaic power forecasting modeling is an effective method to enhance its forecasting precision because photovoltaic output power strongly depends on the specific weather statuses in a given time period. However, the most intractable problems in weather classification models are the insufficiency of training dataset (especially for the extreme weather types) and the selection of applied classifiers. Given the above considerations, a generative adversarial networks and convolutional neural networks-based weather classification model is proposed in this paper. First, 33 meteorological weather types are reclassified into 10 weather types by putting several single weather types together to constitute a new weather type. Then a data-driven generative model named generative adversarial networks is employed to augment the training dataset for each weather types. Finally, the convolutional neural networks-based weather classification model was trained by the augmented dataset that consists of both original and generated solar irradiance data. In the case study, we evaluated the quality of generative adversarial networks-generated data, compared the performance of convolutional neural networks classification models with traditional machine learning classification models such as support vector machine, multilayer perceptron, and k-nearest neighbors algorithm, investigated the precision improvement of different classification models achieved by generative adversarial networks, and applied the weather classification models in solar irradiance forecasting. The simulation results illustrate that generative adversarial networks can generate new samples with high quality that capture the intrinsic features of the original data, but not to simply memorize the training data. Furthermore, convolutional neural networks classification models show better classification performance than traditional machine learning models. And the performance of all these classification models is indeed improved to the different extent via the generative adversarial networks-based data augment. In addition, weather classification model plays a significant role in determining the most suitable and precise day-ahead photovoltaic power forecasting model with high efficiency.

Keywords: photovoltaic power forecasting; weather classification; generative adversarial networks; convolutional neural networks

1. Introduction

Among developing countries, the growing demand for energy that is expected to double from 2015 levels by 2020 has posed a great challenge in the additional generation capacity in power system [1]. This severe fact undoubtedly promotes the rapid development and integration of renewable energy technologies such as energies solar photovoltaic (PV) power generation in recent years in order to mitigate the various pressures of climate change, environmental pollution, and fossil

energy resource depletion [2]. The world energy outlook newly published in 2017 has revealed some striking change of the current state-of-play for different clean energy technologies. As it was reported, deployment of solar power had another very strong year in 2016. In addition, solar PV annual additions surpassed those of wind power for the first time, with more than 70 GW coming online, 50% higher than the previous year [3]. At the end of 2016, global solar PV capacity reached at least 303 GW (up 48% over 2015), which is more than the cumulative world capacity five years earlier [4]. However, with the high penetration of grid-connected PV systems, the intermittent and stochastic of its output power caused by various weather conditions have contributed to great difficulties in the operation, scheduling, and planning of power systems [5]. Therefore, PV Power forecasting (PVPF) that involves various machine learning algorithms is one of the indispensable measures for regional power system and local energy management system (EMS) to address this issue. For example, demand response (DR) can shift electricity consumptions in DR event time to improve the flexibility of system operation [6], leverage the demand-side resources through a variety of electricity market mechanisms to balance the active power and enhance system reliability [7], participate the operation optimization of micro-grid [8], and ultra-short-term solar PV power forecasting [9].

For a certain PV system, the solar spectral irradiance received by PV array is generally impacted by many other meteorological impact factors (MIF), such as aerosol changes, wind speed, wind direction, cloud cover, etc [10-14]. Diverse weather conditions have an important influence in fluctuating output power of solar PV systems [15]. This means that the PVPF accuracy depends on not only the chosen forecasting models, but also the weather statuses. Therefore, researchers show increasing interest in the studies that combine weather classification and PV forecasting models in order to accurately predict the uncertainty and fluctuation of solar output power due to various weather patterns [16-19]. According to the existing achievements in relevant studies, weather classification has been regarded as an effective pre-processing step in order to enhance the prediction accuracy for short-term solar forecasting [20-22], especially for day-ahead forecasting. Specifically speaking, multiple forecasting models that fit for different weather conditions can be more precise and efficient than using only one single uniform forecasting model for all kinds of weather statuses. The reason behind this is that the mapping relationship between the input and output of the forecast model always varies under different weather statuses. So, it is very difficult for a single model to accurately describe the entire mapping relationship of PVPF model under all kinds of weather statuses. In sum, weather classification model plays a significant role in determining the most suitable and precise day-ahead PVPF model with high efficiency.

However, few studies have completely focused on the specific establishment process of an accurate weather classification model. Majority researchers just regard the weather classification as a small part of PVPF work and simply choose one kind of classifier to fill the classification task without the knowledge of whether this classifier well fits for the classification task based on the collected weather data [20-22]. In literature [23], index of weather type is only considered as one of the inputs of the short-term PVPF model rather than the benchmark of selecting the forecasting model under specific weather conditions. Yang et al. applied self-organizing map (SOM) and learning vector quantization (LVQ) to classify the collected historical data of PV power output during the classification stage [20]. Chen et al. also adopted SOM to classify the local weather type of 24h ahead provided by the online meteorological services [22]. Wang et al. proposed a solar irradiance feature extraction and support vector machines (SVM) based weather status pattern recognition model for short-term PVPF [24]. They further assessed the performance of SVM and K-nearest neighbors (KNN) approaches, and then investigated the influences of sample scale, the number of categories, and the data distribution in different categories on the daily weather classification results [25]. However, the primary demerit of these abovementioned traditional classification methods is their shallow learning models. That is, these methods are unable to achieve ideal classification accuracy because they may be insufficient to extract the deep nonlinear traits of the input data especially when feeding large amounts of complex data. One effective way to address the shallow model issue is the use of convolutional neural network (CNN), due to its ability to discover the inherent abstract features and hidden high-level invariant structures in data [26]. Indeed, it has been proved in previous studies that CNN based classification models start emerging as the best performance achievers in various applications, outperforming classical intelligence methods like SVM [27].

As a typical branch of deep learning (DL) algorithm, CNN is a novel approach where the connectivity principles between the neurons are inspired by the organization of animal visual cortex. Besides, it learns to recognize patterns usually through highlighting the edges and pixel behaviors that are generally observed in various images in its layers [28]. Compared with the traditional fully-connected neural networks, the obvious advantage of CNN is the reduced number of parameters to be estimated due to the weight sharing technique. Furthermore, convolutional layers that consist of small size kernels provide an effective way of extracting the hidden structure and inherent features. In addition to these structural performance advantages, the practical application conditions of CNN should also be taken into consideration. Specifically speaking, CNN is good at processing data with grid topology features [29]. Time-series data can be viewed as one-dimensional grid data sampled at fixed time intervals. In our weather classification model, its input is the daily solar irradiance data and output is the weather type of that day. The solar irradiance data used in the model is exactly time series data. Therefore, the merits of

CNN in automatic feature extraction and its suitability for processing time series data inspire us to rethink the weather classification problem based on CNN rather than other traditional classification approaches.

Nevertheless, it is of great difficulty to train a large high-capacity CNN without overfitting when the training data is not enough and imbalanced. Because the accuracy of classification model not only depends on the applied classifiers but also the training data [24,25]. Literature [25] reveals that the decreasing amount of training data will certainly worsen the classification performance to some extent. Moreover, a standard classification model probably performs poorly in imbalanced training data because they induce a bias in favor of the majority class [30]. Indeed, there exists a common problem of lacking massive historical data, especially the data in rare weather circumstances, for most PV plants. This problem results in a considerable challenge of weather classification because the training data is not sufficient and balanced. Thus it is important to figure out a valid and reliable generative model for mimicking real historical records of different weather statuses in order to augment and balance the training dataset in classification.

The core of generative model is to understand the different distribution of historical solar irradiance data under diverse weather status. Many advanced generative models have been presented in prior works. As a simple generative model, Random Oversampling (RO) creates new data via copying random minority class examples [30]. Although the mechanism behind RO is simple, its exact replication of training examples can lead to overfitting since the classifier is exposed to the same information. An alternative approach, known as Synthetic Minority Oversampling Technique (SMOTE) [31], aims to avoid the overfitting problem through generate synthetic data along the line segment that joins minority class samples. However, SMOTE may generate noisy samples because the separation between majority and minority class clusters is not often clear [32]. Some existing works applied the Markov chains or agent-based schemes to established generative model, but this kind of model requires the data to follow specific assumptions such as first-order Markov property [33]. Auto-encoder is also kinds of generative model composed of an encoder and decoder, which are widely used in image processing to reconstruct an image that is as close as possible to the original [34]. The application of auto-encoder may lead to the same problem of RO because the newly generated data merely remember the distribution of the original data but failed to remain the diversity of generated samples.

Recently, a new powerful generative model named generative adversarial network (GAN) has been developed, which is exactly a promising solution to overcome the problems in the above-mentioned approaches. Inspired by two-player zero-sum game, GAN simultaneously trains two adversarial models: the generator network produces artificial data from noise through capturing the original data distribution, while the discriminator network is trained to distinguish generated data from original data [35]. Since their initiation, GAN has been applied to various research field [36], including image and vision computing, speech and language processing, etc. Owing to the powerful modeling ability, GAN is able to create the new data that has the similar distribution of the original data but also remain the diversity of generated data at the same time [36]. And it does not require the data to follow any specific assumptions when modeling the complex data even with some implicit distributions [33]. Therefore, GAN is naturally suitable for the task of data generation in order to augment and balance the training dataset in classification. However, almost no studies have applied GAN for generating time series weather data (i.e. daily solar irradiance data). Indeed, enough training data, either from real dataset or augmentations, is in great demand for weather classification models.

In this paper, we focus on the establishment of weather classification model for day ahead short-term PVPPF. The technical specifications of State Grid Corporation of China (SGCC) specify that the time range of short-term power forecasting is 24 h ahead from 00:00 of the next day [24]. Thus, in current research, weather statuses refer to the atmospheric conditions in one day (24 h) at the location of certain PV plant. According to the national standard formulated by the China Meteorological Administration, weather statuses are divided into 33 weather types [37]. Although the separate establishment of the forecasting model for each weather type can meticulously and accurately describe the mapping relation, such task is hard to be carried out due to the heavy modeling workload and the limited historical data of rare weather types [24]. In our studied dataset, the information of weather status is provided by meteorological services institution and it gives the specific label for each day. We reclassify 33 meteorological weather types into 10 weather types by putting several single weather types together to constitute a new weather type based on the given label, which can considerably reduce the total number of classification models that needed to be trained. Then considering the superiority of improved wasserstein-GAN (WGAN) over the original GAN, wasserstein generative adversarial networks with gradient penalty (WGAN-GP) is introduced to augment the training dataset of the above ten weather types. For certain dataset of weather types, the goal of WGAN-GP is to generate new and distinct data that captures the intrinsic features and distributions of the historical solar irradiance data, but not to simply memorize the input data. Further, the CNN based weather classification model is trained by the augmented solar irradiance dataset that consists of both historical and generated data. Finally, the case study is carried out based on the historical irradiance data collected by Earth System Research Laboratory in Boulder city of USA. Therefore, compared with the existing studies, the contributions of this paper can be summarized as follows:

- First, 33 meteorological weather types are reclassified into 10 weather types by putting several single weather types together to constitute a new weather type, while majority previous studies simply classify all kinds of weather types specified by meteorological institution into several generalized weather classes usually less than four kinds [21,22]. The separate establishment of the PVPF model for each weather type can more meticulously and accurately describe the mapping relation [24,25].
- Second, the augment of solar irradiance dataset is achieved via the application of newly developed WGANGP. This data augment aims to enhance the accuracy of weather classification model. Further, three effective evaluation indexes are introduced in this paper to assess the quality of generated data. The indexes are listed as follows: 1) Standard deviation (STD), 2) Euclidean distance (EDD), 3) Cumulative distribution function (CDF).
- Third, as a state-of-the-art deep learning method, CNN that has been widely used in computer image area is turned to be applied in weather classification. This means that we successfully transform the weather classification problem into an image processing task, which is also helpful for the improvement of classification accuracy. Specifically, 1D irradiance data should be first converted into a 2D image for feature extraction, and then reconverted to a 1D vector for classification.
- Fourth, this research focuses on the establishment of weather classification models taking into account the issue of insufficient training data simultaneously.
- Fifth, the above proposed method is then verified by the case study based on the solar irradiance dataset collected by Earth System Research Laboratory in Boulder city of USA.

The rest of the paper is organized as follows. Section 2 gives a detailed description of the GANs and CNN related theories, as well as the integrated framework of weather classification based PVPF model. In section 3, a case study using laboratory data firstly introduced the data source, experimental setup, model training and hyperparameters selection. Then the study evaluated the quality of GAN-generated data, compared the performance of CNN classification models and traditional machine learning classification models (MLP, SVM, and KNN), investigated the precision improvement of different classification models achieved by GAN, and applied the weather classification models in solar irradiance forecasting. Section 4 highlights the concluding remarks and future work directions.

2. Method

2.1. Generative adversarial networks

As a powerful class of generative models, GAN achieves its function through implicitly modeling high-dimensional distributions of data [29]. In the image processing community, GAN has the ability to synthesize realistic images with better quality when compared to other generative methods [34,38,39]. In terms of the application in our research work, GAN is used to learn the statistical distribution of historical solar irradiance data under ten types of weather status respectively. This allows us to synthesize new samples of daily irradiance data from the learned distribution, which is as one of downstream tasks of GAN namely data augmentation [40]. Here, we adopt an advanced GAN variant called WGANGP in consideration of the intractable training problems in the original GAN. In this section, we introduce the GANs related theories in the proposed weather classification model. We first introduce the original GAN in subsection 2.1.1, and then respectively describe its improved versions (i.e. WGAN and WGANGP) in subsection 2.1.2 and 2.1.3.

2.1.1. Original generative adversarial networks (GAN)

Inspired by two-player zero-sum game, Goodfellow et al. [35] proposed a new framework of generative model named the original GAN. Such generative models are established via simultaneously training two neural networks (i.e. generator and discriminator) under the adversarial learning idea (see Figure 1). The generator generates new samples from noise based on the learned distribution of the real data, while the discriminator aims to distinguish whether the input data comes from the generator or real samples. In order to help readers better understand the main ideas behind GAN, a vivid analogous given here is to think of the generator as a team of counterfeiters and the discriminator as the police. The former creates fake currency, with the aim of making realistic currency. The later receives both fake and authentic currency and aims to tell them apart. Both teams are driven by the mutual competition to further improve their methods until the counterfeits are indistinguishable from the genuine articles. The training details and related mathematical theories of GAN are elaborated as follows.

Computation procedure and structure of GAN

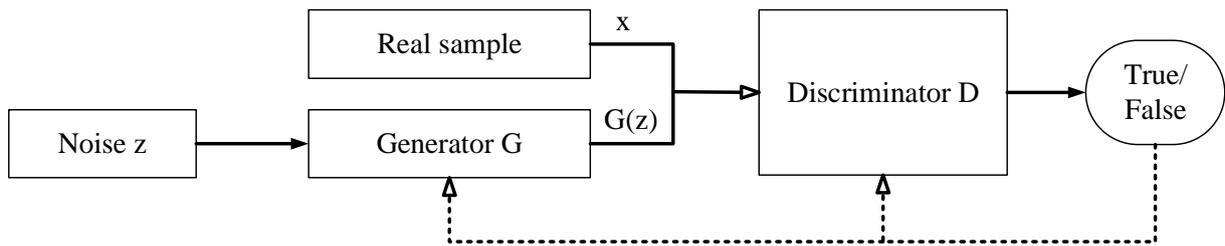


Figure 1. The integrated framework of generative adversarial networks (GAN). The two models that are learned during the training process for a GAN are the discriminator D and the generator G . These are typically implemented with deep neural networks, but they could be implemented by any form of differentiable system that maps data from one space to another.

In the field of signal processing, vectors are usually represented by bold, lowercase symbols, and we adopt this convention to emphasize the multidimensional nature of variables. Accordingly, the distribution of the real data is represented by the probability density function $p_{data}(\mathbf{x})$, where \mathbf{x} is a random vector that lies in $\mathbf{R}^{|\mathbf{x}|}$, $|\mathbf{x}|$ denotes the number of dimensions. We use $p_g(\mathbf{x})$ to denote the distribution of the vectors produced by the generator. A noise vector \mathbf{z} can be easily obtained from the given distribution $p_z(\mathbf{z})$ (p.s. this paper adopts the uniform distribution that is in the range of $[-1, 1]$). Let $D(g\theta^{(D)})$ denote the discriminator function parametrized by $\theta^{(D)}$. Let $G(g\theta^{(G)})$ denote the generator function parametrized by $\theta^{(G)}$. Here, $\theta^{(D)}$ and $\theta^{(G)}$ are the weights of two neural networks, respectively.

- *Discriminator (abbr. D)*

The discriminator, D , takes input samples either coming from real data or coming from generator. As for the neural network of D , its input data vector is mapped to a probability, P_{real} , that such data vector is from the real data distribution, rather than the generator distribution: $P_{real} = D(x; \theta^{(D)}) \rightarrow (0, 1)$. For a fixed G , the D is trained to accurately classify the input data as either being from the real data ($P_{real} \rightarrow 1$) or from a fixed G ($P_{real} \rightarrow 0$). The weights of D 's neural network are updated based on the loss function L_D defined by Equation (1). A smaller L_D can be attained via maximizing $D(\mathbf{x})$ and minimizing $D(G(\mathbf{z}))$, which indicates that D is good at telling the difference between $p_{data}(\mathbf{x})$ and $p_g(\mathbf{x})$.

$$L_D = -\mathbf{E}_{\mathbf{x}: p_{data}} [\log D(\mathbf{x})] - \mathbf{E}_{\mathbf{z}: p_z} [\log(1 - D(G(\mathbf{z})))] \quad (1)$$

- *Generator (abbr. G)*

When the discriminator is optimal, it may be frozen, and the generator, G , may continue to be trained so as to lower the accuracy of the D . In other words, for a fixed D , the training goal of G is to make the generated distribution $p_g(\mathbf{x})$ as close as to the target data distribution $p_{data}(\mathbf{x})$ as possible. During the training, a batch of samples drawn with distribution $p_z(\mathbf{z})$ are fed into G , then G outputs newly generated data that obeys $p_g(\mathbf{x})$ distribution. The weights of G 's neural network are updated based on the loss function L_G defined by Equation (2). A small L_G can be achieved by maximizing $D(G(\mathbf{z}))$, which reflects the generated samples are as realistic as possible from the discriminator's perspective. Further, Goodfellow et al. [35] also proposed an alternate, non-saturating training criterion for G (see Equation (3)) via maximizing $\log D(G(\mathbf{z}))$ rather than minimizing $\log(1 - D(G(\mathbf{z})))$.

$$L_{G(1)} = \mathbf{E}_{\mathbf{z}: p_z} [\log(1 - D(G(\mathbf{z})))] \quad (2)$$

$$L_{G(2)} = \mathbf{E}_{\mathbf{z}: p_z} [-\log D(G(\mathbf{z}))] \quad (3)$$

Then the two loss functions L_D and L_G can be combined to form a two-player minimax game with the value function $V(D, G)$:

$$\min_G \max_D V(D, G) = \mathbb{E}_{\mathbf{x}: p_{data}} [\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z}: p_z} [\log(1 - D(G(\mathbf{z})))] \quad (4)$$

In spite of the great success achieved by the original GAN, the above mentioned two loss functions for the generator, $L_{G(1)}$ and $L_{G(2)}$, can cause fatal issues that make GAN's training challenging and difficult. These issues are illustrated as follows. The more details about the relevant mathematical derivation and proof can be found in literature [41].

- Issues caused by the original loss function $L_{G(1)}$

This kind of loss function can lead to gradient vanishing problem in the generator, especially if D is trained to be very strong. When D is optimal, minimizing $L_{G(1)}$ is equivalent to minimizing the *Jensen-Shannon* (JS) divergence between distributions $p_{data}(\mathbf{x})$ and $p_g(\mathbf{x})$ as shown in Equation (5). These two distributions have support contained in two closed manifolds that don't perfectly align and don't have full dimension. Due to this reason, the JS divergence will always be a constant $\log 2$, which renders the gradient of the generator loss to be 0. This points us to an important training trick: the D shouldn't be trained too well or the gradient of G will vanish. Thus this loss function can make GAN training extremely hard because the user has to decide the precise amount of training dedicated to the discriminator.

$$L_{G(1)} = 2JS(p_{data}(\mathbf{x}) \| p_g(\mathbf{x})) - 2\log 2 \quad (5)$$

- Issues caused by the $-\log D$ alternative $L_{G(2)}$

This kind of loss function can lead to instability of generator gradient updates, as well as the mode collapse (i.e. generator produces very similar samples for different inputs so as to lack of diversity in the generated samples). When D is optimal, minimizing $L_{G(2)}$ is equivalent to minimizing the Equation (6). This equivalent loss function aims to minimize the *Kullback-Leibler* (KL) divergence between distributions $p_{data}(\mathbf{x})$ and $p_g(\mathbf{x})$, and simultaneously maximize the JS divergence between them. This is intuitively ridiculous and will cause instability in the gradient.

$$L_{G(2)} = KL(p_g(\mathbf{x}) \| p_{data}(\mathbf{x})) - JS(p_{data}(\mathbf{x}) \| p_g(\mathbf{x})) \quad (6)$$

$$KL(p_g(\mathbf{x}) \| p_{data}(\mathbf{x})) = \int \log \left(\frac{p_g(\mathbf{x})}{p_{data}(\mathbf{x})} \right) p_g(\mathbf{x}) d\mu(\mathbf{x}) \quad (7)$$

In addition, even the KL term by itself has some issues (see Equation (7), where both $p_{data}(\mathbf{x})$ and $p_g(\mathbf{x})$ are assumed to be absolutely continuous, and therefore admit densities, with respect to the same measure μ defined on \mathbf{x}). Due to its asymmetry, the penalty for different types of errors is quite different. For example, when $p_g(\mathbf{x}) \rightarrow 0$ and $p_{data}(\mathbf{x}) \rightarrow 1$, we have $KL(p_g(\mathbf{x}) \| p_{data}(\mathbf{x})) \rightarrow 0$, which has almost 0 contribution to $KL(p_g(\mathbf{x}) \| p_{data}(\mathbf{x}))$. But when $p_g(\mathbf{x}) \rightarrow 1$ and $p_{data}(\mathbf{x}) \rightarrow 0$, we have $KL(p_g(\mathbf{x}) \| p_{data}(\mathbf{x})) \rightarrow +\infty$, which has gigantic contribution to $KL(p_g(\mathbf{x}) \| p_{data}(\mathbf{x}))$. So the tiny penalty caused by the first type of error can let the G fail to produce realistic samples, leading to a lack of diversity. While the enormous penalty caused by the second type of error can let G produces unrealistic samples, leading to a lack of accuracy. Therefore, the generator would rather produce repetitive and 'safe' samples, than the samples with high diversity and the risk of triggering the second type of error. This causes the infamous mode collapse.

2.1.2. Wasserstein generative adversarial networks (WGAN)

Arjovsky and collaborators [41,42] argued that the common divergences, such as JS and KL divergences, are potentially not continuous and thus do not provide a usable gradient for G. Therefore, a new distance metric, so-called Wasserstein distance (Earth-Mover distance), is introduced by the authors to transform the training objectives of the original GAN. Such change can provide more smooth gradients for G and thus successfully avoid the issues of vanished and instable gradient as well as mode collapse. The equation of Wasserstein distance is shown as follows.

$$W(p_{data}, p_g) = \inf_{\gamma \in \Pi(p_{data}, p_g)} \mathbb{E}_{(x,y): \gamma} [\|\mathbf{x} - \mathbf{y}\|] \quad (8)$$

Where $\Pi(p_{data}, p_g)$ denotes the set of all joint distributions $\Upsilon(x, y)$ whose marginal distributions are respectively p_{data} and p_g . Intuitively, $\Upsilon(x, y)$ calculates how much “mass” need to be transported from x to y in order to transform the distribution p_g into the distribution p_{data} . The Earth-Mover distance then is the “cost” of the optimal transport plan. Though the Wasserstein distance has nicer properties than JS and KL divergences [42], the infimum $\inf_{\Upsilon \in \Pi(p_{data}, p_g)}$ is highly intractable. Thanks to the Kantorovich-Rubinstein duality [43], the Wasserstein distance becomes:

$$W(p_{data}, p_g) = \sup_{\|f\|_L \leq 1} \mathbb{E}_{x: p_{data}} [f(x)] - \mathbb{E}_{x: p_g} [f(x)] \quad (9)$$

Where the supremum is over all the 1-Lipschitz functions $f: x \rightarrow \mathbf{R}$. Therefore, we can have a parameterized family of functions $\{f_w(x)\}_w \in W$ that are K -Lipschitz for some K , the problem we are solving now becomes:

$$KW(p_{data}, p_g) \approx \max_{w \in W: \|f_w\|_L \leq K} \mathbb{E}_{x: p_{data}} [f_w(x)] - \mathbb{E}_{x: p_g} [f_w(x)] \quad (10)$$

Let the neural network of the discriminator (D) with weights w denote the K -Lipschitz function $f_w(x)$, and maximize $L = \mathbb{E}_{x: p_{data}} [f_w(x)] - \mathbb{E}_{x: p_g} [f_w(x)]$ as much as possible so that D can well approximate the actual Wasserstein distance between distribution p_{data} and p_g (i.e. $D(x) \approx f_w(x)$). Thus the loss function of WGAN’s discriminator is defined by Equation (11). Furthermore, G will try to minimize L , and since the first term in L does not concern the generator, its loss function is to minimize $-\mathbb{E}_{x: p_g} [f_w(x)]$ as shown in Equation (12).

$$L_D = \mathbb{E}_{x: p_g} [f_w(x)] - \mathbb{E}_{x: p_{data}} [f_w(x)] \quad (11)$$

$$L_G = -\mathbb{E}_{x: p_g} [f_w(x)] \quad (12)$$

To implement Wasserstein GAN, Arjovsky et al. [35] suggested the following modifications to the original GAN:

- Sigmoid or softmax activation is removed in the output layer of the discriminator since WGAN’s discriminator is trained for regression rather than classify.
- Log function is no longer adopted in the loss function of WGAN’s discriminator and generator.
- Weight clipping is conducted to enforce a Lipschitz constraint on $f_w(x)$ denoted by the discriminator, confining discriminator weights w to range $(-c, c)$, i.e $w \leftarrow clip(w, -c, c)$.
- Using momentum based optimizer such as Adam is avoided. Use RMSProp or SGD.

Due to the above modifications, the success achieved by WGAN is listed as follows [42]:

- The issues of training instability in the original GAN have been solved, and a careful balance in training of the discriminator and the generator is not required during the training of WGAN.
- The issues of vanished and instable gradient as well as mode collapse are avoided, which ensures the diversity of generated samples.
- WGAN can continuously estimate the EM distance by training the discriminator to optimality.
- The achievement of the abovementioned three benefits doesn’t require a careful design of the network architecture either. This can be done by the simplest multilayer perceptron network.

2.1.3. Wasserstein generative adversarial networks with gradient penalty (WGANGP)

The recently proposed WGAN has attracted much attention due to its significant progress of stable GAN training. But WGAN sometimes can still generate low-quality samples or fail to converge in some settings. Gulrajani et al. [44] argued that the use of weight clipping leads to WGAN’s pathological behavior:

- Capacity underuse

The discriminator network tends to fit much simpler functions when weight clipping is used to enforce a K -Lipschitz constraint on the discriminator. As a deep neural network, this is really a huge waste of discriminator’s own powerful fitting

ability. The discriminator fails to make full use of its own model capabilities, and the gradient it returns back to the generator will also deteriorate.

- Exploding and vanishing gradients

The WGAN's optimization process is difficult due to the interactions between the weight constraint and the loss function. Thus, the use of weight clipping can also result in either vanishing or exploding gradients if the clipping parameter c is not carefully tuned.

Therefore literature [37] proposed penalizing the norm of the gradient of the discriminator with respect to its input as an alternative method for enforcing the Lipschitz constraint. The loss function of the discriminator is redefined by Equation (13) while the loss function of the generator remains the same as WGAN's.

$$L_D = \mathbb{E}_{x: p_g} [f_w(x)] - \mathbb{E}_{x: p_{data}} [f_w(x)] + \lambda \mathbb{E}_{\hat{x}: p_x} \left[\left(\|\nabla_{\hat{x}} f_w(\hat{x})\|_2 - 1 \right)^2 \right] \quad (13)$$

Where we implicitly define p_x sampling uniformly along straight lines between pairs of points sampled from the real data distribution $p_{data}(\mathbf{x})$ and the generator distribution $p_g(\mathbf{x})$. The empirical value of the penalty coefficient λ is 10.

Such modified WGANP loss function stabilizes GAN training over a wide range of architectures with almost no hyper-parameter tuning, which make WGANP performs better than standard WGAN in terms of training speed and sample quality. Moreover, WGANP simultaneously preserves the WGAN's property that WGAN's loss correlates with sample quality and converges toward a minimum.

2.1.4. Quality comparison of the sample data generated by different generative models

In the above sections, the disadvantages of the original GAN [35] and WGAN [42], as well as the superiorities of WGANP [44] are discussed in details. In order to further validate the abovementioned conclusions, we carried out related simulations to compare the quality of sample data generated by different generative models (i.e. GAN, WGAN, and WGANP). It should be mentioned that we train these three kinds of generation adversarial networks with the same network configuration (two hidden layers with 124 neurons). And the simulation results are shown in Figure 2-4.

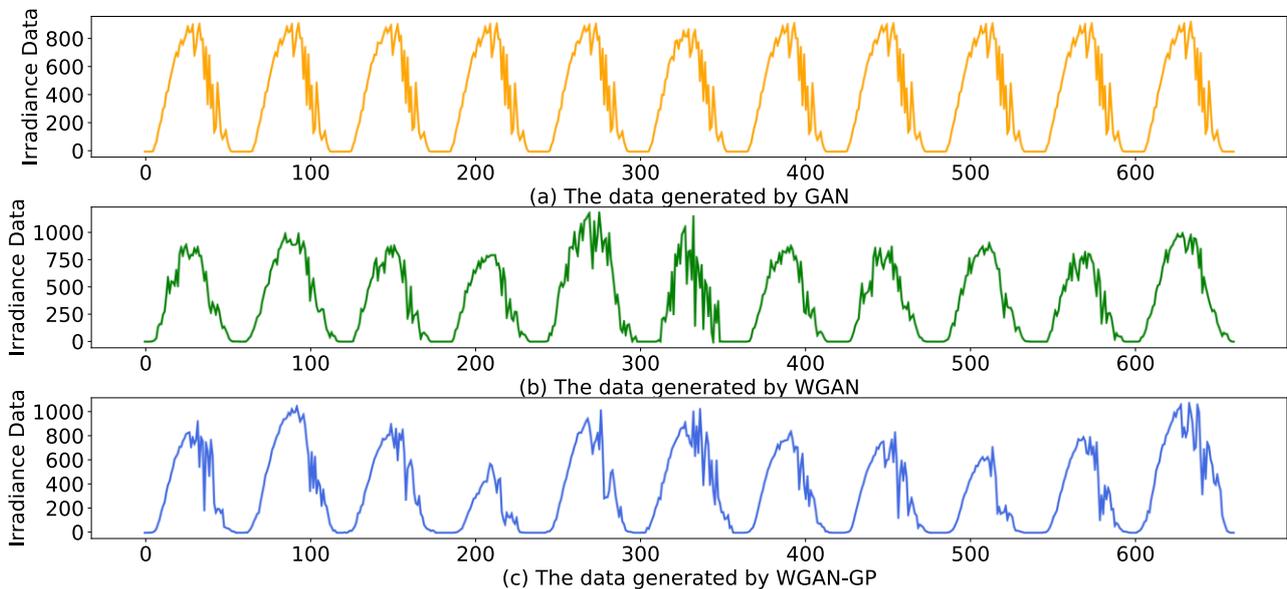


Figure 2. (a) The data generated by the original GAN; (b) The data generated by WGAN; (c) the data generated by WGANP; All the data used to train the generative models is from the solar irradiance data of weather type ‘‘Class 3’’ (i.e. morning is sunny and afternoon is rainy).

First, during the training of GAN's network, we use the $L_{G(1)}$ (refer to Equation (2)) as loss function. The relevant simulation shows that when the model is trained to 5000 iterations, the gradient vanishing problem appears. If we choose $L_{G(2)}$ (refer to the Equation 3 in our paper) as loss function, the generator would rather produce repetitive and ‘safe’ samples, than the samples with high diversity and the risk of triggering the second type of error. This causes the infamous mode collapse. So the training process of GAN is relatively difficult. The game process of generator and discriminator needs to be carefully

adjusted. The above discussions are in accordance with the results shown in the subplot (a) of Figure 2 where all the generated data are highly similar. That is, the data generated by the original GAN always lack diversity, which is not good for the training of classification models.

Second, the data generated by WGAN is shown in the subplot (b) of Figure 2. As you can see, the primary merit of WGAN is that it can avoid the issues of vanished and instable gradient as well as mode collapse. This can ensure the diversity of generated samples. But WGAN also has a severe problem that the unstable training caused by weight clipping can lead to poor quality of generated samples. Thus the clipping parameter should be carefully tuned. In our simulation, the value of clipping parameter ranges from -0.01 to 0.01. The above discussions are in accordance with the results shown in the subplot (b) of Figure 2 that some noises exist in the generated data, which is caused by the unstable training.

Third, the subplot (c) of Figure 2 shows that the samples generated by WGAN-GP are highly diverse. And the sample noise is also very small. Almost no hyper-parameter tuning is needed here. As for the WGAN applied in literature [45], its structures of the neural networks in the generator and the discriminator are very complex. Both the neural networks are consisted of multiple convolutional layers and fully connected layers. Additionally, the number of neurons in these layers is also large. But the WGAN-GP adopted in our paper is merely made by two fully connected layers. Although the structure of WGAN-GP is very simple, the WGAN-GP can still generate diverse sample data with high quality. And the computational cost is also obviously reduced because of its simple structure.

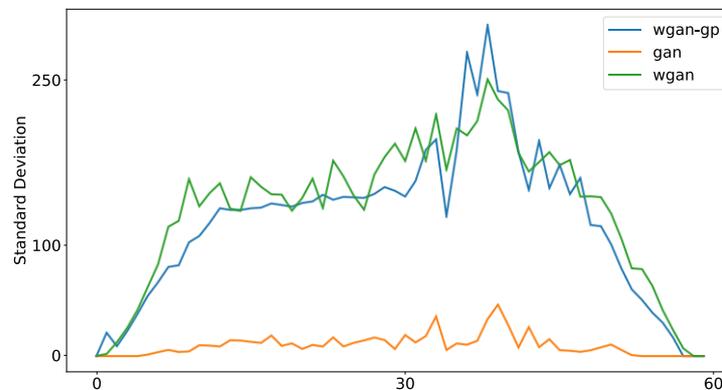


Figure 3. For weather types of “Class 3”, the standard deviation (STD) of solar irradiance data corresponding to each time point is compared among three dataset: 1) the solar irradiance dataset generated by the original GAN (denoted by orange color), 2) the solar irradiance dataset generated by WGAN (denoted by green color), 3) the solar irradiance dataset generated by WGAN-GP (denoted by blue color). The abscissa denotes the time point in one day with the temporal resolution of 15 minutes while the ordinate denotes the STD.

To more objectively evaluate the performance differences among the GAN, WGAN, and WGAN-GP, two indexes namely STD and EDD are applied to further conduct a host of tests. For more details about the definition of STD and EDD, you can refer to the content in Section 3.3.1 and 3.3.2. The STD is used to evaluate the pattern diversity and noisy of generated samples. The smaller STD value means that the generated sample data is not diverse, and the frequently fluctuated STD value indicates that there are lots of noises in the sample data. The EDD is used to evaluate the statistical resemblance of generated samples. The sample size of all the generated dataset is controlled as 1000. In Figure 3, the STD of solar irradiance data (belongs to weather types of “Class 3”) corresponding to each time point is compared among three datasets generated by GAN, WGAN, and WGAN-GP. During the whole time period, all the STD values of the yellow curve are very small, which indicates that the original GAN always creates repetitive sample data that lack of diversity. In addition, the green curve fluctuates more frequently than the blue curve during the first half time period. This phenomenon can be attributed to the noise in the data generated by WGAN. In fact, the STD’s curve during the first half time period should be relatively gentle because the weather type of the morning is sunny. Fortunately, our adopted generative model named WGAN-GP well capture this gentle trend of STD curve during the first half time period.

Figure 4 shows the EDDs between averaged original daily solar irradiance (DSI) curve and averaged generated DSI curves. For ease of drawing, the ordinate is the logarithm of the EDDs, which is simply abbreviated as $\log(\text{EDDs})$. It is obvious that the $\log(\text{EDDs})$ between averaged original daily solar irradiance (DSI) curves and GAN-generated DSI curve is the largest. This further verifies that the distribution of GAN-generated data is not much similar to the original one. The shortest blue histogram represents the $\log(\text{EDDs})$ between averaged original daily solar irradiance (DSI) curves and WGAN-GP-generated DSI curve. This result validates that WGAN-GP can well create the new sample with the resemble distribution of the original dataset.

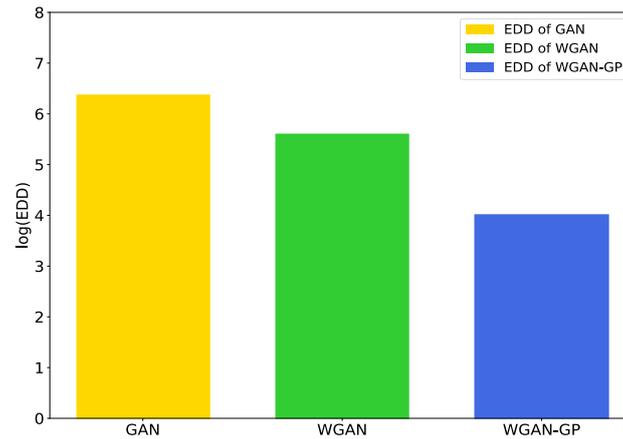


Figure 4. “Euclidean distances (EDDs) between averaged original daily solar irradiance (DSI) curve and averaged GAN-generated DSI curve” versus “EDDs between averaged original DSI curve and averaged WGAN-generated DSI curve” versus “EDDs between averaged original DSI curve and averaged WGAN-GP-generated DSI curve”. The ordinate is the logarithm of the Euclidean distance.

2.2. Convolutional neural networks (CNN)

As a well-known deep learning architecture, Convolutional Neural Networks (CNN) has been widely applied in the fields such as image recognition and classification due to its powerful ability of feature extraction. In general, CNN based classification models consist of four types of layers [46], namely convolutional layers, pooling layers, fully connected layers and logistic regression layers. The structure of the CNN is illustrated in Figure 5.

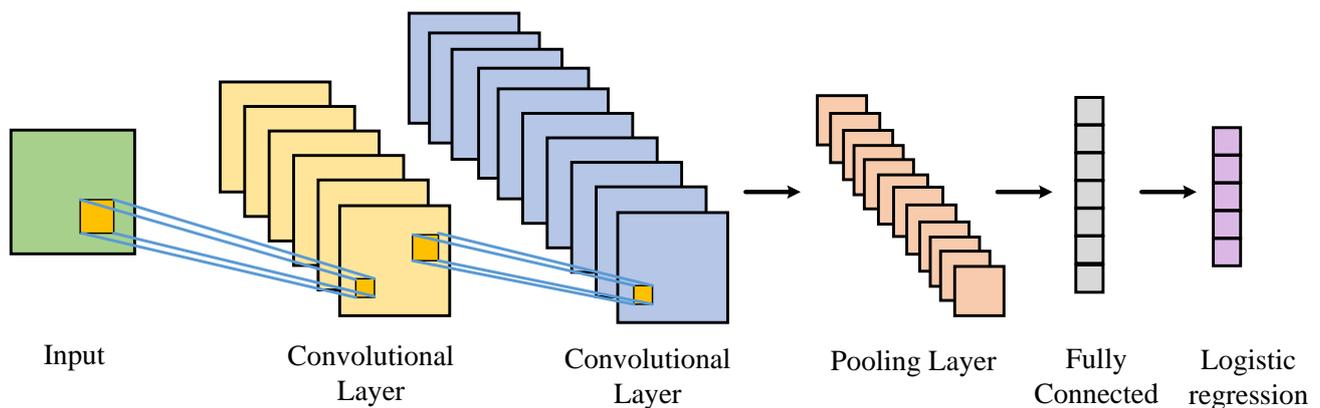


Figure 5. The structure of the CNN.

2.2.1. Convolutional Layer

The purpose of the convolutional layer is to extract feature representations of the inputs. The convolutional layer is composed of multiple convolution kernels, each of which is used to calculate one feature map. Each neuron of the feature map is connected to the region of an adjacent neuron in the previous layer. This region is called the receptive field in the previous layer. As for a new feature map, it can be obtained via the following two steps. First, convolution operation is conducted on the inputs based on a convolution kernel. Second, an element-wise nonlinear activation function is then applied to the convolution results in the first step. In addition, one convolution kernel is shared by all spatial locations of the input to generate a feature map. In practice, multiple different kernels will be adopted to generate multiple feature maps. For example, as shown in Figure 5, the first convolutional layer consists of six kernels and then six corresponding feature maps are acquired. The equation of convolutional layer is shown below:

$$y_{i,j,k}^l = F((w_k^l)^T x_{i,j}^l + b_k^l) \quad (14)$$

Where w_k^l and b_k^l are the weight and bias of the k -th convolution kernel in the l th layer. $x_{i,j}^l$ denotes the input region at location (i, j) of the l -th layer. Note that the weight w_k^l is shared to all regions of input. The weights sharing method can effectively reduce the parameters which model needs to learn and hence make it easier to train the model. $F(g)$ denotes to the activation function which is applied to improve the fitting abilities of model. Here, we adopt RELU [47] activation function that has the advantage of reducing the appearance of overfitting.

2.2.2. Pooling Layer

The pooling layer is usually constructed to connect the previous convolutional layer. The aim of pooling layer is to aggregate the input features by reducing the resolution of feature maps. And the equation of pooling operation is shown below:

$$P_{i,j,k}^l = \text{pool}(y_{m,n,k}^l) \quad (15)$$

Where $(m,n) \in R_{i,j}$, and $R_{i,j}$ is the region around location (i, j) . The traditional pooling operations are max pooling and average pooling.

2.2.3. Fully Connected Layer

Fully Connected layer, a typical layer of artificial neural networks, is usually set between pooling layer and logistic regression layer. The aim of fully connected layer is to transport the learned distributed feature representation to one space in order to perform high-level reasoning. All the neurons of previous layer are connected to every single neuron of current layer.

2.2.4. Logistic Regression Layer

Regression layer is the last layer of CNN classification model. As a widely used function in various multiclass classification methods, softmax function is usually employed as the activation function of logistic regression layer. The equation of softmax function is shown below:

$$P(y = j) = \frac{e^{x^T w_j}}{\sum_{k=1}^K e^{x^T w_k}} \quad (16)$$

Where the meaning of this equation is the predicted probability of the j -th class with a given sample vector x and a weighting vector w . And k is the number of overall class. During the testing process, each sample obtains a corresponding probability value for each class. In the end, each sample is classified to be the certain class with the largest probability value.

2.2.5. Description of one-dimensional (1D) CNN and 2D CNN

One-dimensional (1D) CNN and 2D CNN are two branches of CNN. In our research, both of them are chosen to establish the weather classification models for day ahead short-term PVPF. Their corresponding classification performance are analyzed and compared in subsequent section. It should be mentioned that the input of classification model is the daily solar irradiance data (i.e. time series data) and the output is the weather type of that day. One-dimensional (1D) CNN and 2D CNN are respectively abbreviated as CNN1D and CNN2D in our paper.

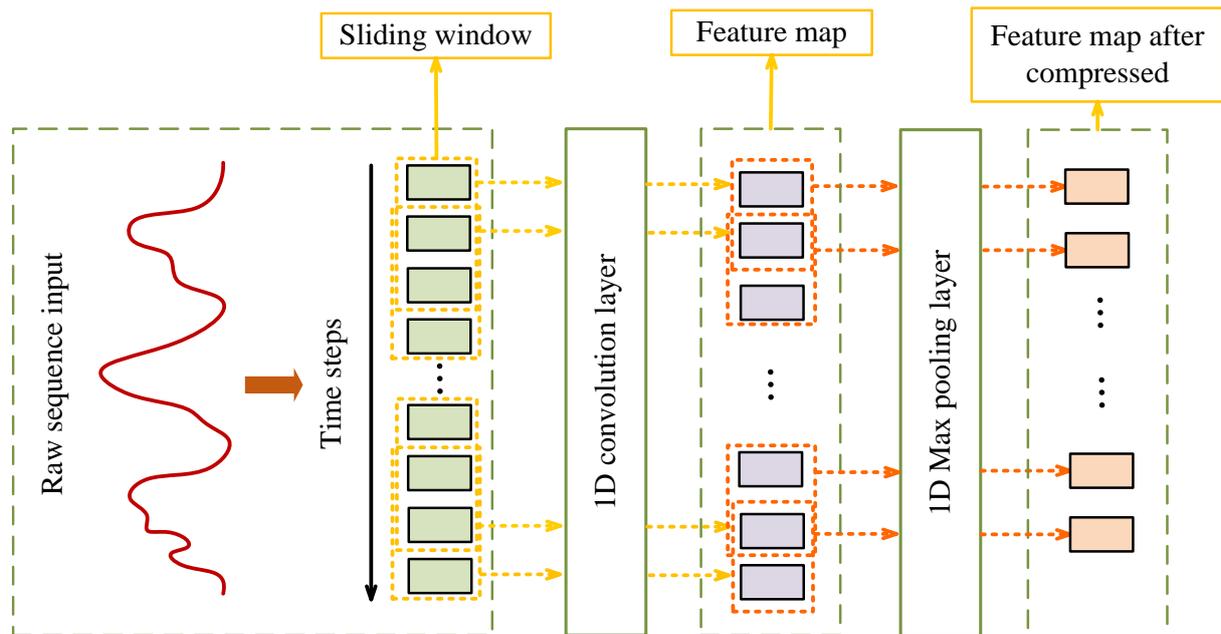


Figure 6 Illustration of 1D convolution and pooling. In the 1D convolution layer, the size of kernel is 3 and the corresponding stride is 1. In the 1D pooling layer, the size of kernel is 2 and the corresponding stride is also 1.

The common characteristic of CNN1D and CNN2D is that they are built by stacking multiple different layers, such as a convolutional layer, pooling layer, fully connected layer and etc. The convolution layer and the pooling layer are the core layers of CNN. The convolution layer plays the role of feature extractors and learns the local features which restrict the receptive fields of the hidden layers to be local. The pooling layer is usually constructed to connect the previous convolutional layer. The aim of pooling layer is to aggregate the input features by reducing the resolution of feature maps. The detailed structures of CNN1D and CNN2D are respectively shown in Figure 6 and Figure 7. According to these two figures, it can be seen that the main difference between CNN1D and CNN2D is their different sliding ways of the sliding windows during the convolutional process. Regarding the CNN1D, the window merely slides along the single direction (i.e. time steps). As for the CNN2D, the window moves horizontally and vertically on the input matrix. That is, the window has two sliding directions. This is the reason behind the definition of two-dimensional CNN. The biggest advantage of such sliding ways is that CNN2D can better capture the global deep features of the input data from the perspective of time and space.

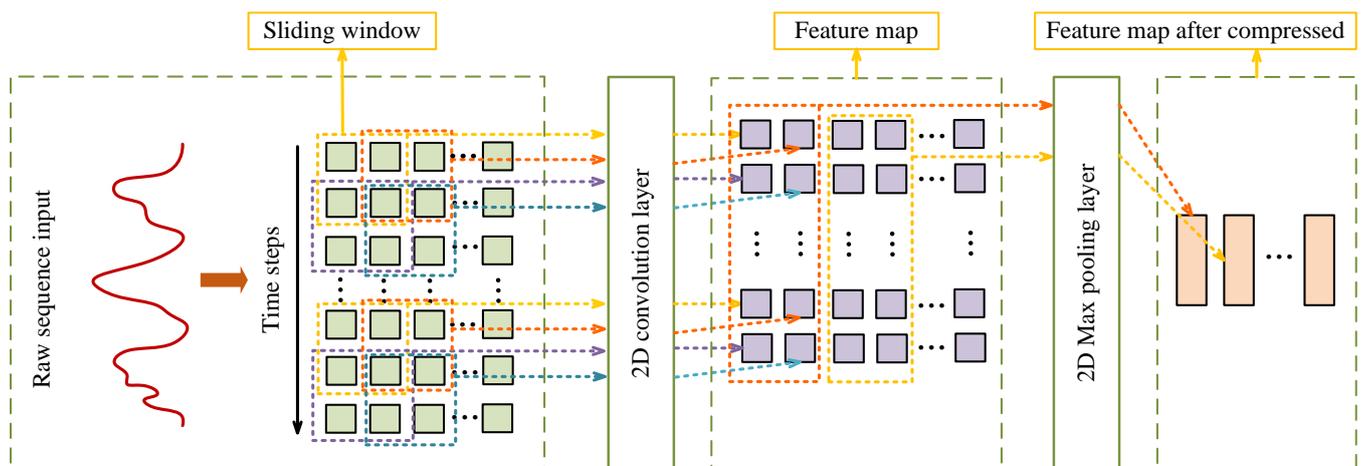


Figure 7 Illustration of 2D convolution and pooling. In the 2D convolution layer, the size of kernel is 2×2 and the corresponding stride is 1×1 . In the 2D pooling layer, the size of kernel is $n(\text{row}) \times 2(\text{column})$ and the corresponding stride is also $n \times 2$.

Table 1. Corresponding relation between 33 meteorological weather types and four general weather types (GWT).

GWT	Meteorological weather types
Sunny	Sunny, sunny interval cloudy, cloudy interval sunny
Cloudy	Cloudy, overcast, overcast interval cloudy, cloudy interval overcast, fog
Rainy	Shower, thunder shower, thunder shower and hail, sleet, light rain, showery snow, light snow, freezing rain, light to moderate rain, light to moderate snow
Heavy rainy	Moderate rain, heavy rain, torrential rain, big torrential rain, extra torrential rain, moderate snow, heavy snow, torrential snow, moderate to heavy rain, heavy to torrential rain, torrential to big torrential rain, big torrential to extra torrential rain, moderate to heavy snow, heavy to torrential snow, sand storm

Table 2. Definition and description of weather types.

Weather types	Description		Weather types	Description	
	Morning	Afternoon		Morning	Afternoon
Class 1	Sunny	Sunny	Class 6	Cloudy	Rainy
Class 2	Sunny	Cloudy	Class 7	Rainy	Sunny
Class 3	Sunny	Rainy	Class 8	Sunny	Cloudy
Class 4	Cloudy	Sunny	Class 9	Rainy	Rainy
Class 5	Cloudy	Cloudy	Class 10	Heavy rainy	Heavy rainy

2.3. Proposed weather classification model based on WGANGP and CNN

WGANGP and CNN based weather classification model for day ahead short-term PVPF is proposed in this paper. First of all, the specific reclassification process of the abovementioned 33 meteorological weather types is elaborated as follows. According to the solar irradiance dataset collected by Earth System Research Laboratory in Boulder city of USA, the information of weather status is provided by meteorological services institution and it gives the specific label for every day, such as sunny, cloudy, overcast, light rain, moderate rain, heavy rain, torrential rain, shower, thunder shower, fog, light snow, etc. As shown in Table 1, all the 33 weather types are generally classified into four general weather types (GWTs) namely sunny, cloudy, rainy and heavy rainy [24,25]. However, such rough classification still cannot meet the requirement of meticulously and accurately describing the mapping relation in the PVPF model. Therefore, finer reclassification of the weather types is necessary. First, the whole day is divided into two time periods that is morning and afternoon. And all the GWTs are likely to occur in these two periods. Thus the 33 meteorological weather types are reclassified into 16 classes at the beginning. Then considering the fact that heavy rainy condition is rare, the heavy rainy type is merged into rainy type, and a total of ten weather types are obtained. For example, if the weather status on the morning is heavy rainy, and the weather status on the afternoon is sunny, then weather type of that day is classified as Class 7. But it should be noted that the (heavy rain, heavy rain) weather status is reserved in this classification, which is defined as Class 10. In summary, we reclassify 33 meteorological weather types into 10 weather types by putting several single weather types together to constitute a new weather type based on the given label in studied dataset, which can considerably reduce the total number of classification models that needed to be trained. The definition and description of the ten weather types are shown in Table 2.

The integrated framework of the proposed weather classification based PVPF model is presented in Figure 8. We first reclassify the 33 meteorological weather types into 10 classes. Then, the daily solar irradiance dataset with labeled weather types is divided into the training set and testing set. Next, WGANGP is applied to generate new samples to achieve the augment of the training dataset for each weather type. The CNN classification model is trained by the augmented training dataset that consists of the original and newly generated data. At last, a weather classification based PVPF model is established.

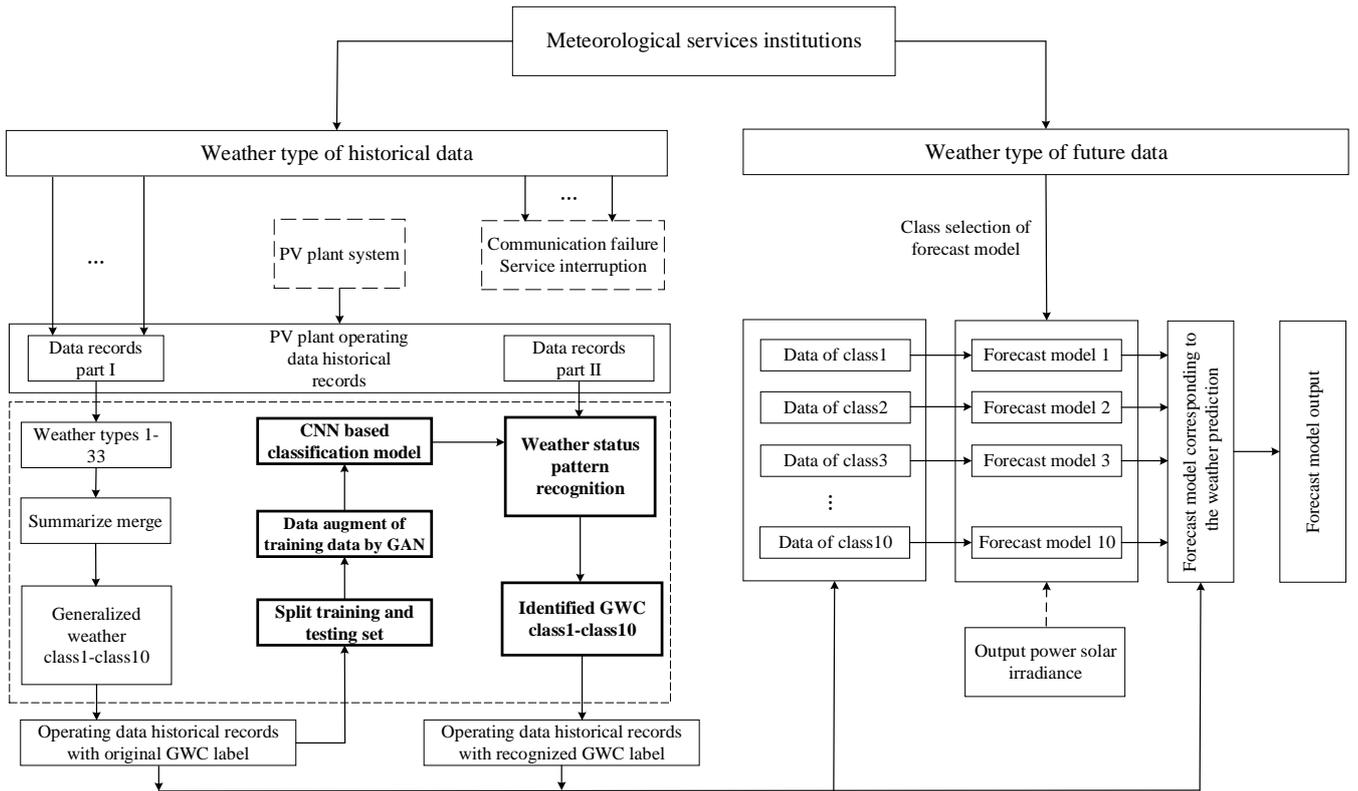


Figure 8. The integrated framework of weather classification based PVPF model.

3. Case study

3.1. Data source and experimental setup

The data used in all the simulations of this paper are collected by the National Oceanic & Atmospheric Administration (NOAA) Earth System Research Laboratory website via the measuring device in the Surface Radiation (SURFRAD) station during 2014 to 2015 at Desert Rock. There are totally 650 days' available irradiance data with 1min time resolution. To meet the international standard of a short-period solar irradiance forecasting, the irradiance data should be further transformed to be the data with 15min time resolution through taking the average of 15 points data in the span of every 15 minutes. So, there are total 96 irradiance data points in one day. Considering the earliest sunrise time and the latest sunset time, we only use the 18th to 78th data points during the whole day. We set the first 350 days of irradiance data as training samples and the last 300 days of irradiance data as testing samples. The specific training sample size and the corresponding ratio of each weather class are shown in Table 3. The specific testing sample size and corresponding ratio of each weather class are shown in Table 4.

All experimental platforms are built on high-performance Lenovo desktop computer equipped with the Win10 operating system, Intel(R)Core(TM) i5-6300HQ CPU @2.30GHz, 8.00GB RAM, and NVIDIA GeForce GTX 960M GPU. We use Python 3.6.1 with Tensorflow and scikit-learn to perform all the simulations.

Table 3. The number and distribution of ten weather types in training dataset.

Training dataset	Class1	Class2	Class3	Class4	Class5	Class6	Class7	Class8	Class9	Class10
Samples number	119	45	18	17	36	9	8	4	78	16
ratio	34%	12.85%	5.14%	4.857%	10.29%	2.57%	2.29%	1.143%	22.28%	4.57%

Table 4. The number and distribution of ten weather types in testing dataset.

Testing dataset	Class1	Class2	Class3	Class4	Class5	Class6	Class7	Class8	Class9	Class10
Samples number	98	40	13	19	27	5	5	7	65	12
ratio	33.67%	13.7%	4.33%	6.33%	9%	1.66%	1.66%	2.33%	21.67%	4%

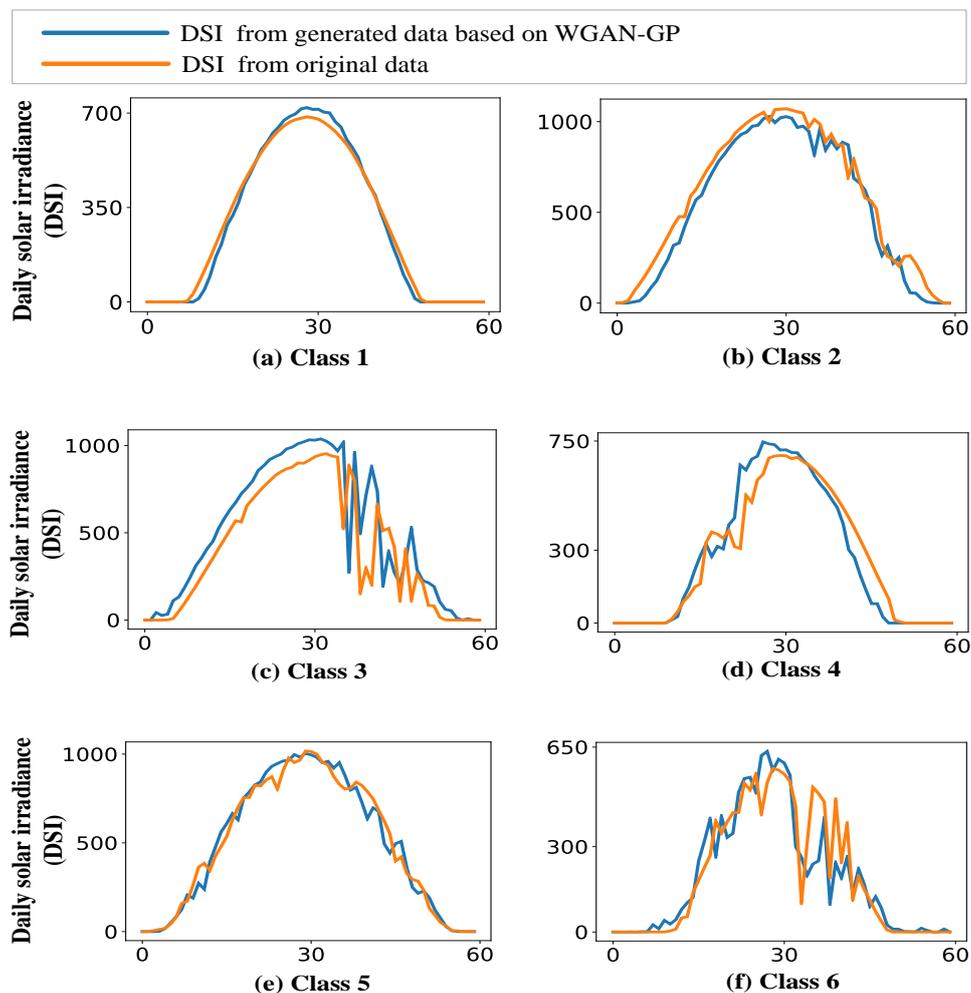
3.2. Model training and hyperparameters selection

In the generative adversarial network training process, both the generator and the discriminator utilize a neural network that consists of two hidden layers with 124 neurons. In order to accelerate the convergence of neural networks, we first make the irradiance data normalized to the range of $[0,1]$ by minmaxscaler. The generated data is then anti-normalized to normal range after the model training process is completed. The input noise of the generator is 60 dimensional uniform distribution data which is in the range of $[-1, 1]$. This paper adopts Adam as the optimizer of generator and discriminator. ReLU activation is used in both the generator and discriminator, except for the output layer. In the actual summation of WGANGP, we adopt the training trick that the model is training alternatively between 15 steps of optimizing discriminator and 1 step of generator.

In the classification stage, the CNN2D model consists of 3 convolutional layers and their kernel size are 1×1 , 2×1 and 3×2 respectively. The kernel size of a max pooling layer in CNN2D is 2×2 . The CNN1D model consists of 3 convolutional layers and their kernel size are 3, 5 and 8 respectively. The kernel size of a max pooling layer in CNN1D is 2. As for each convolutional layer in both CNN1D and CNN2D, the number of kernel of is 64. The MLP classification model is made up by 2 fully connected hidden layers. There are 100 neurons in each fully connected hidden layer. As a loss function, the cross entropy is applied to CNN1D, CNN2D and MLP classification models. In addition, the grid-search method is used to find the optimal parameters of SVM and KNN classification models.

3.3. Evaluation of generated data's quality

The in-depth evaluation of generated data's quality is elaborated in this section. Figure 9 clearly shows the different variation trend and amplitude fluctuation of daily solar irradiance (DSI) under ten weather types defined in section 3.1. It should be noted that the solar irradiance curve of a sunny day is much smoother than other weather status. Intuitively, as the comparison shown in Figure 9, our trained WGANGPs under all the ten weather types are able to create samples with similar behaviors and exhibit a diverse range of patterns. In other words, the goal of WGANGP is to generate new and distinct samples that capture the intrinsic features of the original data, but not to simply memorize the training data.



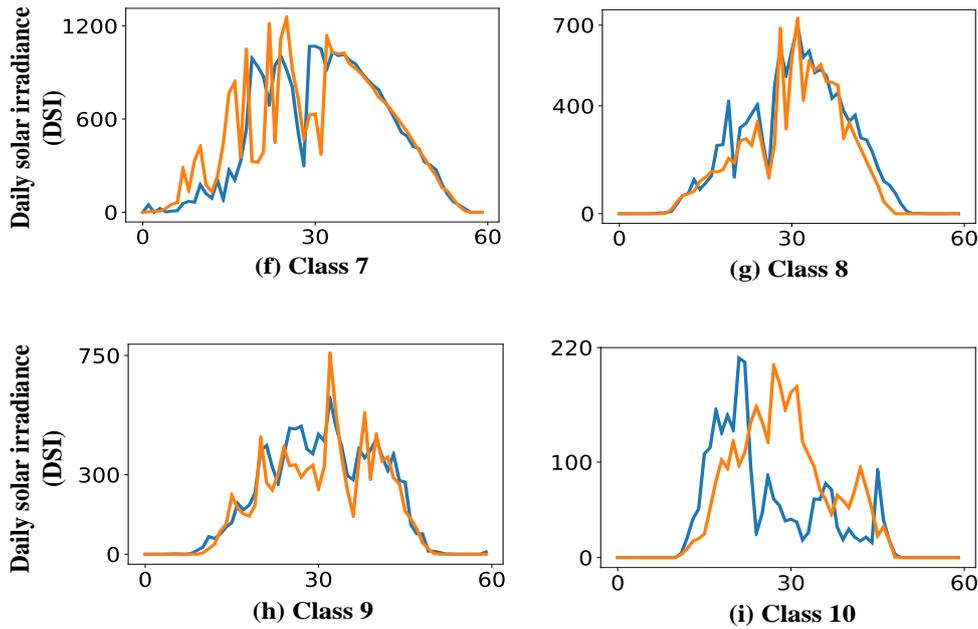


Figure 9. Selected samples from the original daily solar irradiance (DSI) dataset versus newly generated samples from our trained WGANGPs for all the ten weather types. The pair of samples is selected using Euclidean distance based search. In each subplot, such as figure (a), the abscissa denotes the time point in one day with the temporal resolution of 15 minutes while the ordinate denotes the solar irradiance values at a certain time. Here, the orange curve is the original data, and the blue curve is the generated data based on WGANGP.

For example, the subplot (c) in Figure 9 shows the comparison between generated curve and real curve under weather type of “Class 3” (i.e. morning is sunny while afternoon is rainy). The generated curve correctly captures the smooth trend during the morning period as well as the frequent fluctuation during the afternoon period, simultaneously showing the strong diurnal cycles. Moreover, the generated curve not completely match the real curve from the original data, but slightly exhibit a change on the whole. This phenomenon further verifies the two important properties of the WGANGP that is pattern diversity and statistical resemblance.

To more objectively assess the quality of WGANGP’s generated samples, we conduct a host of tests with the three indexes (i.e. STD, EDD and CDF). The first index is used to evaluate the pattern diversity of generated samples. The last two indexes are used to evaluate the statistical resemblance of generated samples.

3.3.1. Standard deviation (STD)

In statistics, the standard deviation (STD, also represented by the Greek letter sigma σ) is a measure that is used to quantify the amount of variation or dispersion of a set of data values [48]. A low STD indicates that the data points tend to be close to the mean (also called the expected value) of the set, while a high STD indicates that the data points are spread out over a wider range of values. The calculation of STD is defined by Equation (17).

$$\sigma_t = \sqrt{\frac{1}{N_t} \sum_{i=1}^{N_t} (x_{ti} - \mu_t)^2} \quad (t = 1, 2, \dots, 60) \quad (17)$$

Where σ_t and μ_t are respectively the SID and expected value of a set of solar irradiance data $\{x_{ti}\}$ at the certain time point t , and N_t is the total sample size of that set.

In practice, this statistical index is applied here to evaluate the pattern diversity of generated samples. As for a set of generated solar irradiance data at a certain time point, a larger STD can reflect the higher dispersion degree of the set, which verifies the diversity of solar irradiance values at that time point. Nevertheless, a small STD value denotes that the corresponding generative model can only simply remember the training data but fail to present the diversity.

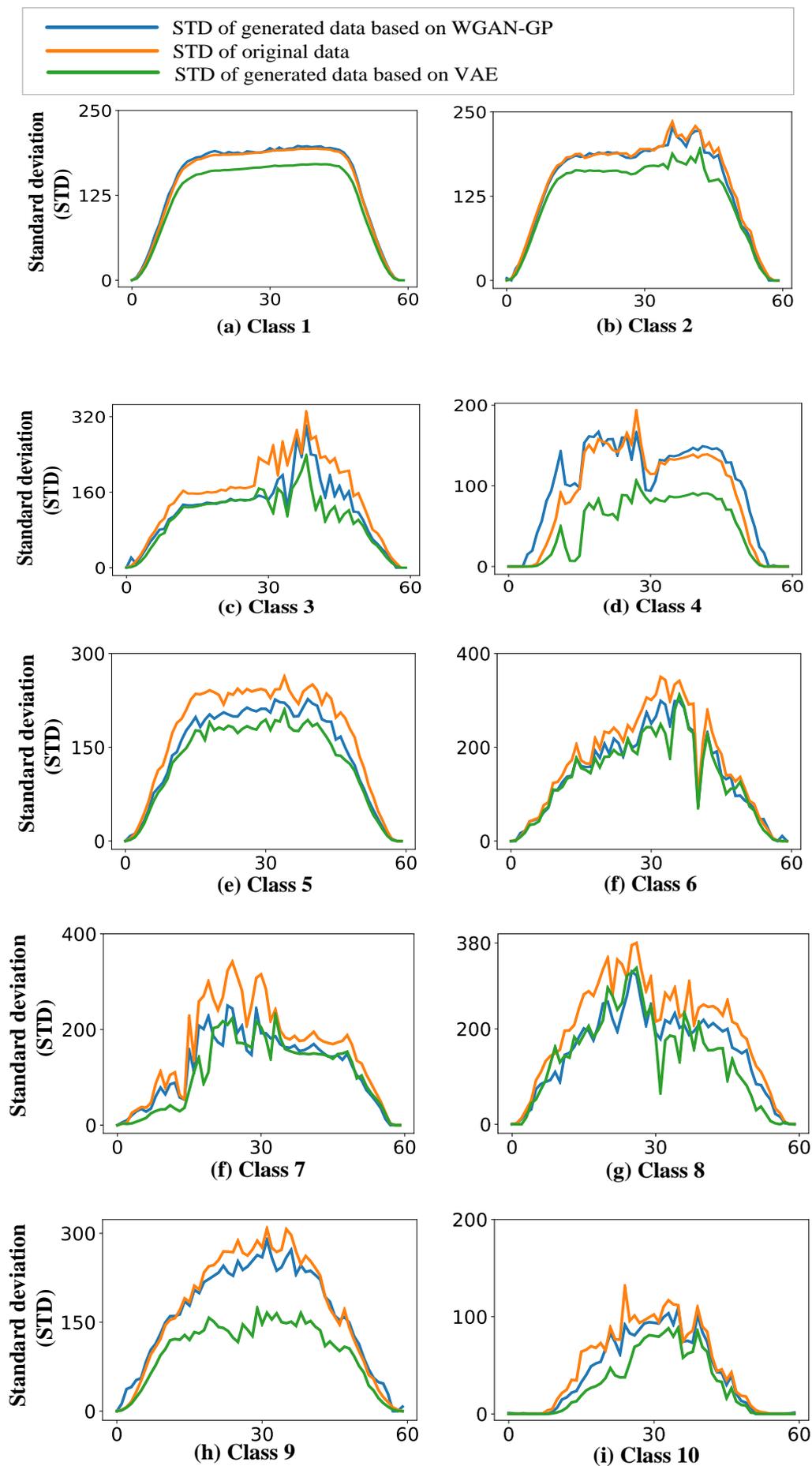


Figure 10. For all the ten weather types, the standard deviation (STD) of solar irradiance data corresponding to each time point is compared among three dataset: 1) the original solar irradiance dataset (denoted by orange color), 2) the generated dataset based on WGANGP (denoted by blue color), 3) the generated dataset based on VAE (denoted by green color). In each subplot, such as figure (a), the abscissa denotes the time point in one day with the temporal resolution of 15 minutes while the ordinate denotes the STD.

Here, another kind of generative model, named variational auto-encoder (VAE) [49], is used to compare with WGANGP in order to highlight the superior properties of WGANGP. The comparison results of pattern diversity among three datasets (i.e. the original solar irradiance dataset, the generated dataset based on WGANGP, and the generated dataset based on VAE), are clearly illustrated in Figure 10. As for the ten subplots in that figure, it can be found that most blue curves are much closer to the yellow curves than the green curves. Even the blue curves in Figure (a) and (b) perfectly match the orange curves. This fact indicates that WGANGPs for ten weather types are able to maximally remain the pattern diversity of the samples in the original dataset. But the pattern diversity of the generated samples via VAE are reduced a lot compared to whether the dataset generated by WGANGPs or the original dataset. In addition, part blue curves are slightly higher than the orange curves during some time period (e.g. the time period of “t=0” to “t=15” in Figure (d)). This shows that sometimes the generated samples based on WGANGPs spread out over a wider range of values than the original samples. In other words, the WGANGPs can also generate new samples that are not contained in the original training data but obey the same statistical distribution of the original samples. For example, as shown in Figure 10, the generated DSI curves are resembled to the original one but are not completely the same. In summary, WGANGP has the good properties of pattern diversity.

3.3.2. Euclidean distance (EDD)

Euclidean distance (EDD) is one of the most generally used distances in research work. In mathematics, the EDD is the "ordinary" straight-line distance between two points in Euclidean space [50]. With this distance, Euclidean space becomes a metric space. The associated norm is called the Euclidean norm. In n -dimensional Euclidean space, the distance is calculated by Equation (18).

$$d(\mathbf{p}, \mathbf{q}) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_i - q_i)^2 + (p_n - q_n)^2} \quad (18)$$

Where $\mathbf{p} = (p_1, p_2, \dots, p_n)$ and $\mathbf{q} = (q_1, q_2, \dots, q_n)$ are two points in n -dimensional Euclidean space. The smaller the EDD is, the closer the two points are. In practice, this distance measure is used to test the statistical resemblance between generated DSI curves and original DSI curves from the direct perspective of DSI curve's distribution. Thus \mathbf{p} and \mathbf{q} are the vectors that both are composed by 60 average values of solar irradiance data corresponding to the specific time points ($n = 60$).

Table 5. The EDD between averaged original DSI curves and averaged generated DSI curves.

	Class1	Class2	Class3	Class4	Class5	Class6	Class7	Class8	Class9	Class10
WGANGP ¹	34.41	146.29	55.65	109.32	49.55	221.96	99.49	385.00	149.43	60.22
VAE ²	26.63	161.70	143.93	156.68	146.81	533.78	234.12	501.56	131.64	61.66

¹ Values in the “WGANGP¹” row is the Euclidean distances (EDDs) between averaged original DSI curves and averaged WGANGPs-generated DSI curves for ten weather types; ² Values in the “VAE²” row is the Euclidean distances (EDDs) between averaged original DSI curves and averaged VAEs-generated DSI curves for ten weather types. The columns are corresponding to ten weather types.

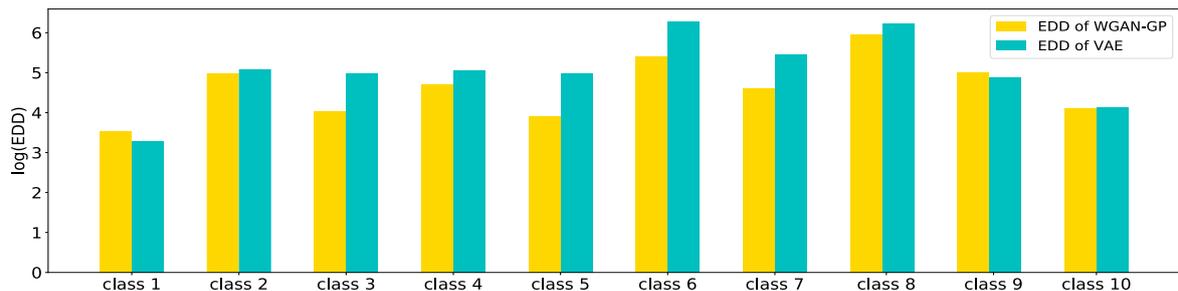


Figure 11. “Euclidean distances (EDDs) between averaged original DSI curves and averaged WGAN-GPs-generated DSI curves for ten weather types” versus “Euclidean distances (EDDs) between averaged original DSI curves and averaged

VAEs-generated DSI curves for ten weather types”. The former is marked by yellow bars while the latter is marked by green bars. In this histogram, the abscissa denotes the weather types while the ordinate denotes the logarithm value of EDD.

As for all the ten weather types, “the EDDs between averaged original DSI curves and averaged WGANGP-generated DSI curves” are compared with “the EDDs between averaged original DSI curves and averaged VAE-generated DSI curves”. The relevant results are shown in Table 5 and illustrated in Figure 11. The smaller EDD indicates the higher degree of the statistical resemblance between two curves (i.e. one curves is very close to another curve). The following results show that majority “EDDs between averaged original DSI curves and averaged WGANGP-generated DSI curves” are much lower than the “the EDDs between averaged original DSI curves and averaged VAE-generated DSI curves” (See the values in column “Class 2”-“Class 8” of Table 5). This reveals that WGANGP can better imitate the distribution of the original data in most cases when compared to VAE.

3.3.3. Cumulative distribution function (CDF)

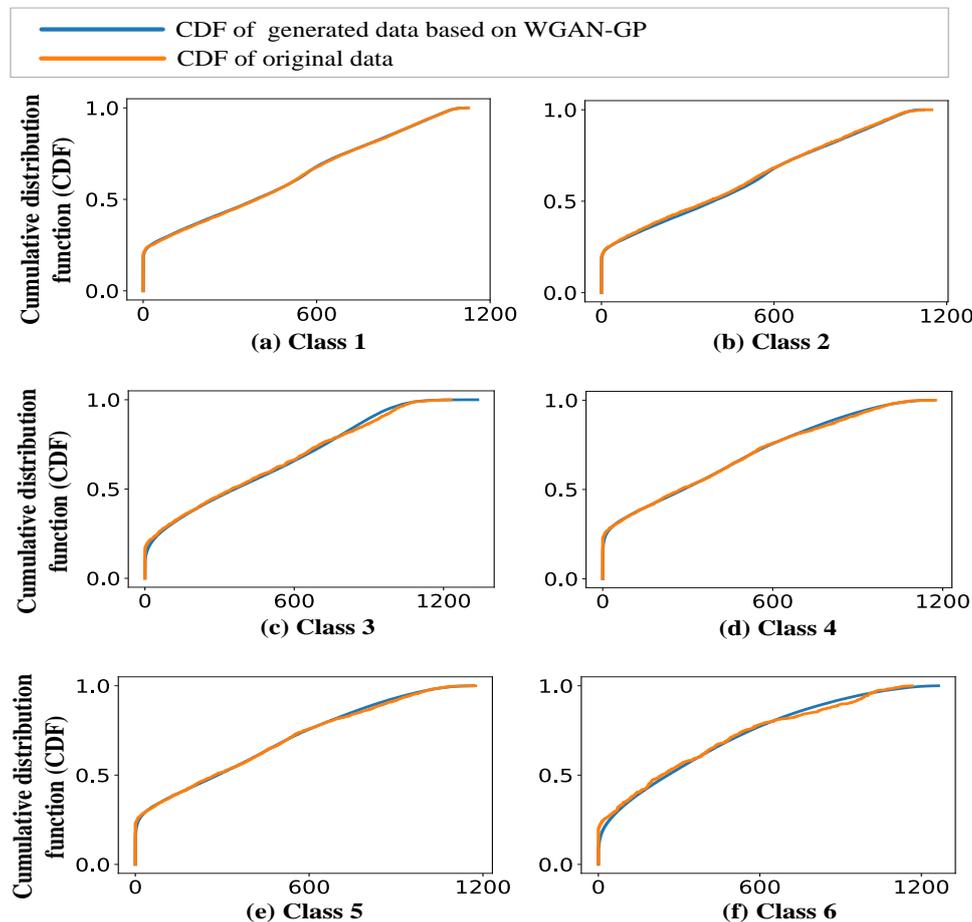
In probability theory and statistics, the cumulative distribution function (CDF, also cumulative density function) of a real-valued random variable X , or just distribution function of X , evaluated at x , is the probability that X will take a value less than or equal to x [51]. In the case of a continuous distribution, it gives the area under the probability density function from minus infinity to x . The CDF of a continuous random variable X can be expressed as the integral of its probability density function f_X as follows:

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f_X(t) dt \quad (19)$$

Different to the index of EDDs, the CDF is applied here to measure the statistical resemblance between the original samples and WGANGP based generated samples from another indirect perspective of DSI data’s probability distribution. In our paper, it is assumed that the probability density functions (PDFs) of original data and generated data are both continuous, which are denoted by $p_{data}(\mathbf{x})$ and $p_g(\mathbf{x})$ respectively. Therefore, the corresponding CDFs can be rewritten as follows:

$$F_{data}(x) = P(X \leq x) = \int_{-\infty}^x p_{data}(t) dt \quad (20)$$

$$F_g(x) = P(X \leq x) = \int_{-\infty}^x p_g(t) dt \quad (21)$$



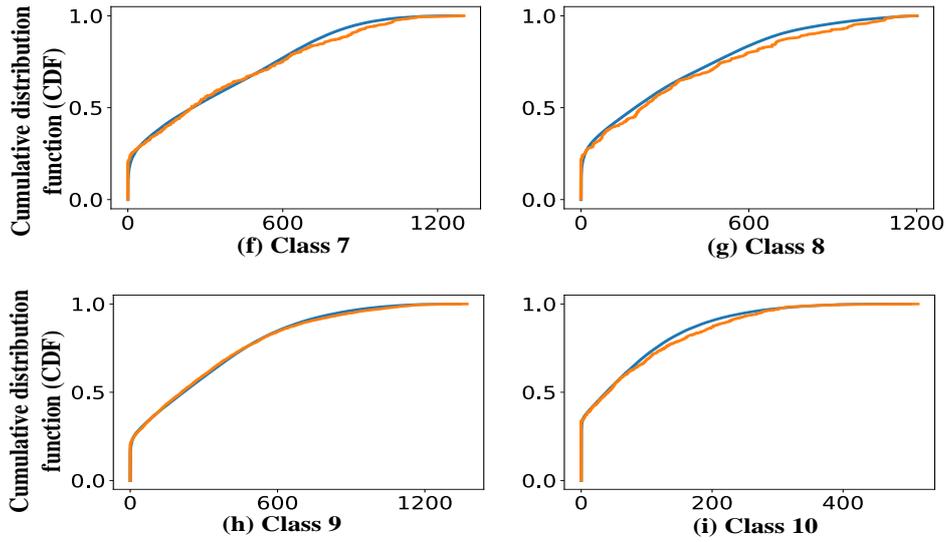


Figure 12. Cumulative distribution function (CDF) of the original solar irradiance dataset versus CDF of generated dataset from our trained WGANGPs for all the ten weather types. In each subplot, such as figure (a), the abscissa denotes solar irradiance value while the ordinate denotes the CDF. Here, the orange curve is the CDF of the original dataset, and the blue curve is the CDF of generated dataset based on WGANGP.

The above Figure 12 clearly shows the comparison results of CDFs between original solar irradiance dataset and the WGANGP generated dataset. It is worthy to note that two CDFs nearly lie on top of each other under all the weather types. This indicates that our trained WGANGPs for different weather status have the capability to generate samples with the correct marginal distributions that is almost the same as the original one. These results further verify the WGANGPs second properties of statistical resemblance.

3.4. Accuracy comparison of different weather classification models

In the problem of statistical classification, a confusion matrix (also known as an error matrix [52]), is a specific table layout that allows visualization of the performance of classification algorithm. The name stems from the fact that it makes it easy to see if the classification model is confusing two classes (i.e. commonly mislabeling one as another). Confusion matrix defined in Equation (22) contains all the information about actual and predicted classes produced by a classification model.

$$M = \begin{bmatrix} m_{11} & \dots & m_{1n} \\ \dots & \dots & \dots \\ m_{n1} & \dots & m_{nn} \end{bmatrix} = [m_{ij}] (i, j = 1, 2, \dots, n) \quad (22)$$

Where m_{ij} is the amount of samples that belong to the class i and classified to the class j , n is the number of total categories.

According to the elements in the confusion matrix, the three indexes defined by Equation (23) to (25) are the commonly used criteria to evaluate the performance of the classification model.

$$PA_i = \frac{m_{ii}}{\sum_{j=1}^n m_{ij}} (i = 1, 2, \dots, n) \quad (23)$$

$$UA_i = \frac{m_{ii}}{\sum_{j=1}^n m_{ji}} (i = 1, 2, \dots, n) \quad (24)$$

$$OA = \frac{\sum_{i=1}^n m_{ii}}{\sum_{j=1}^n \sum_{i=1}^n m_{ij}} \quad (i = 1, 2, \dots, n) \quad (25)$$

Where PA is product's accuracy, UA is user's accuracy, OA is overall accuracy. The OA is the indicator to describe the classification accuracy of all the outputs. The PA and UA evaluate the performance of classification models from different perspectives respectively from the tester and user. For a tester, the actual class of a sample is known, and the PA (also called recall rate) is the indicator to describe the classification accuracy of this specific actual class. For users, they are more concerned about if the given classification result is correct, and the UA (also called precision) is the indicator to characterize the credibility of the output (i.e. the indicator to describe the correctly identified ratio of one specific output class).

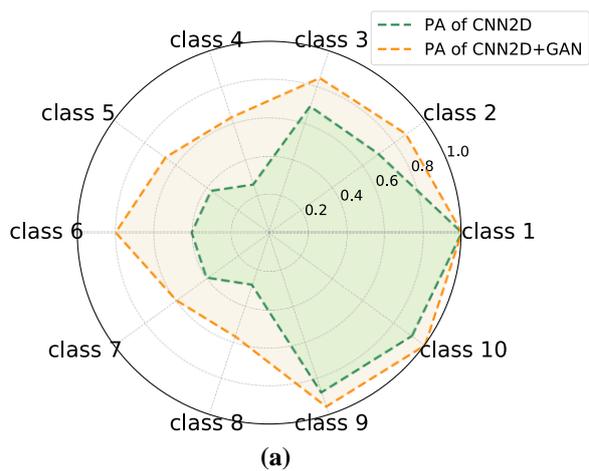
According to the above three indexes, a comprehensive comparison between five commonly used classification models is shown in Table 6 in order to better understand their classification performance. These five classification models are listed as follows: CNN with 1-dimensional convolution layer (CNN1D), CNN with 2-dimensional convolution layer (CNN2D), multilayer perceptron (MLP), support vector machine (SVM), and k-nearest neighbor (KNN). The improvement of their classification performance achieved by WGANGP is also reflected in this Table. The comparison results shown in Table 6 are discussed from three aspects: 1) the performance difference among these five classification models when controlling the consistency of the other conditions, 2) the changes in the performance of classification models when the training dataset is augmented by WGANGP, 3) the differences in the classification results for various weather types. In addition, all the information contained in Table 6 is vividly illustrated in Figure 13 and 14 respectively.

As for the first aspect, we primarily focused on the data in the green marked rows. It can be concluded that CNN2D trained by current dataset shows its great potential to achieve a higher classification performance than other classification models. Specifically speaking, the OA value of CNN2D is high to 76.9% while the worst OA among other classification models can low to 48.1% (KNN). For most weather types, CNN2D can also reach highest PAs and UAs compared to others. In a few cases, such as the UA_4 of Class 4 and UA_6 of Class 6, its OAs and PAs are lower than the CNN1D. Such differences in the classification performance of CNN2D under various weather types are probably caused by the imbalance of data distribution in the ten weather types (See Table 3 and 4).

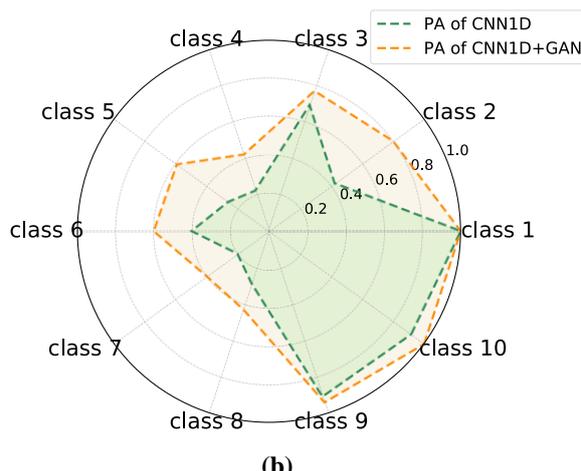
Regarding the second aspect, it can be reflected by Table 6 that the performance of various classification models is indeed improved a lot through the WGANGP based augment of training data. First of all, all the OAs of these five classification models have increased. The increment of the classification accuracy ranges from 4.2% (MLP) to 21.3% (KNN). In terms of the PAs and UAs of classification models under various weather types, the results in Table 6 strongly verify that the WGANGP based data augment has large potential in enhancing the classification accuracy for the weather types with small sample size. As shown in Table 3, the sample sizes of Class 6-8 are very small compared to others, which causes great difficulty in the achievement of high classification accuracy for such weather types. But this difficulty has been relieved via the data augment. Specifically, for Class 6, the PA of KNN (i.e. 20%) has been increased by 60% through the combination with WGANGP. Moreover, for Class 7 and 8, the PAs and UAs of KNN are both low to zero. Their values are then increased, ranging from 14.3% to 50%.

Considering the third aspect, the data in Table 6 does show great differences of the classification results among various weather types. As for the "Class 1" column, the PA_{1S} among all the classification models reach 100% and part of UA_{1S} are also highly close to 100%. This result is likely to be explained by the fact that the sample size of Class 1 is the highest. Besides, the classification results of Class 10 among different classification models are also not bad. This is because the solar irradiance curve of that weather type (i.e. Class 10) is obviously distinguishable from others (See subplot (i) in Figure 9), which makes the Class 10 easy to be correctly identified by the classification models. However, the classification results of Class 7 and 8 are not very ideal. For example, the PA_7 , UA_7 , PA_8 , UA_8 of KNN are all low to zero. This is resulted by the small training sample size of Class 7 and 8 (See Table 3).

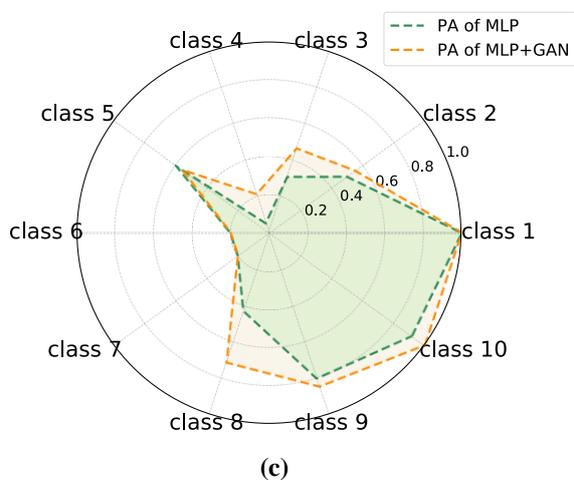
In summary, based on the current dataset, CNN2D shows the superior properties in the classification of weather types compared to other classification models. Furthermore, the application of WGANGP does improve the performance of various classification models to different extent. Also, the classification results of various weather types present obvious differences due to the data's imbalance distribution for the ten weather types.



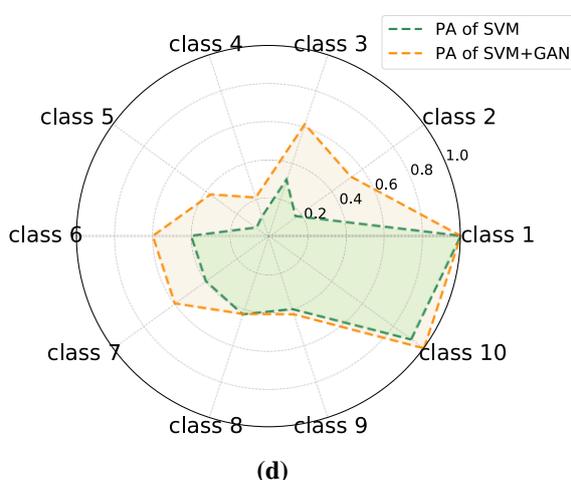
(a)



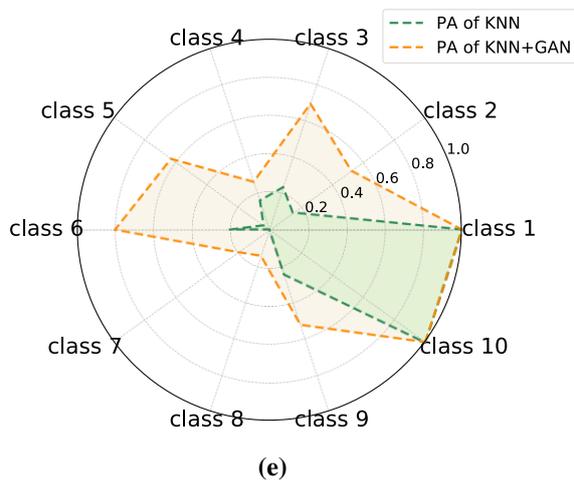
(b)



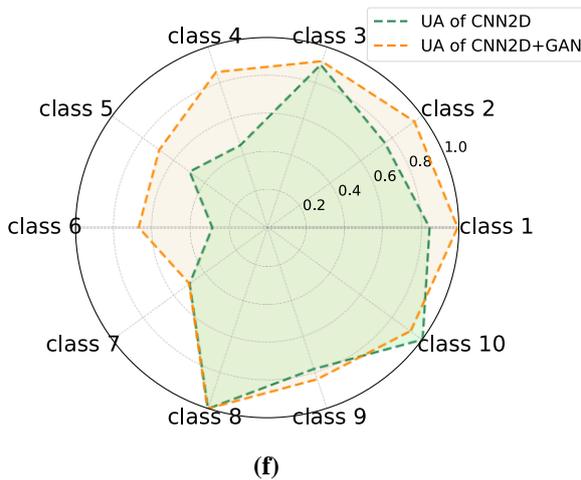
(c)



(d)



(e)



(f)

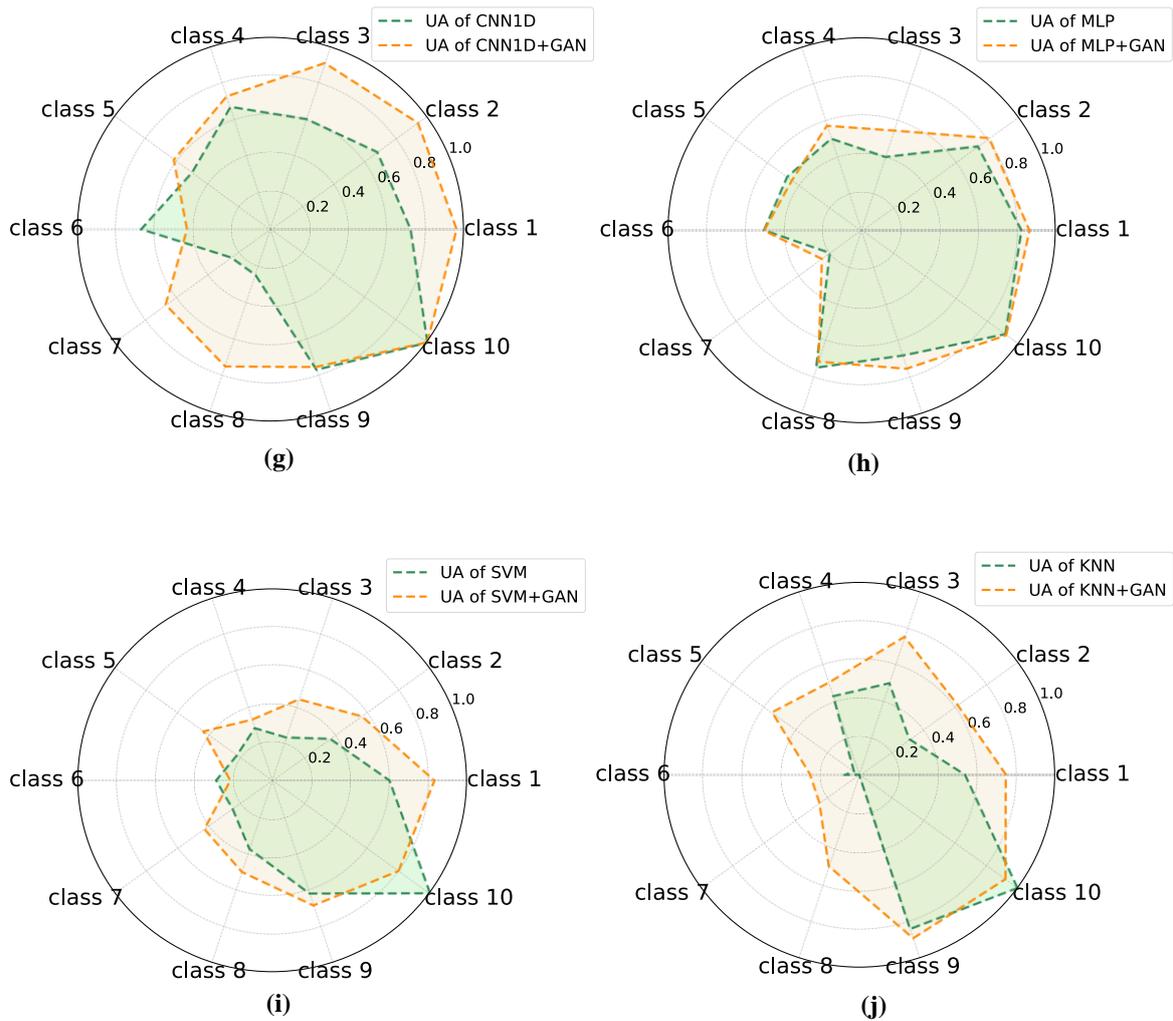


Figure 13. Classification results of all classification models for 1-10 classes with or without data augment. Figure (a)-(e) show the PA score distribution of all classification models. Figure (f)-(j) show the UA score distribution of all classification models. In each subplot, the green curve illustrates the classification performance of classification models trained by original data; the orange curve illustrates the classification performance of classification models trained by augmented data.

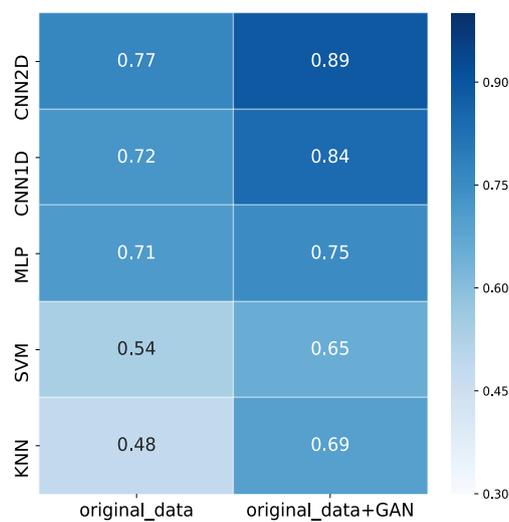


Figure 14. The OA scores' heatmap of all classification models trained by dataset with or without WGANGP based data augment.

3.5. The application of weather classification models in solar irradiance forecasting

For a certain PV system, the solar irradiance is the most important influence factor of the PV power generation [24]. Therefore, the proposed WGANGP and CNN based weather classification models are further applied to the solar irradiance forecasting. This step is to verify the effectiveness of accurate weather classification for improving the precision of subsequent solar irradiance forecasting. Figure 15 clearly shows one week's solar irradiance curves predicted by different forecasting methods. A total of three forecasting methods are compared here. The first method is our proposed model shown in Figure 8. The core idea of the second method is similar to the first one. The only difference between them is that the 33 weather types are roughly classified into four GWTs. As for the third method, there is only single forecasting model where the influence of different weather status is not taken into considerations. For the convenience of subsequent discussion, these three methods are abbreviated as follows: "Ten weather type", "Four weather type" and "Without weather classification".

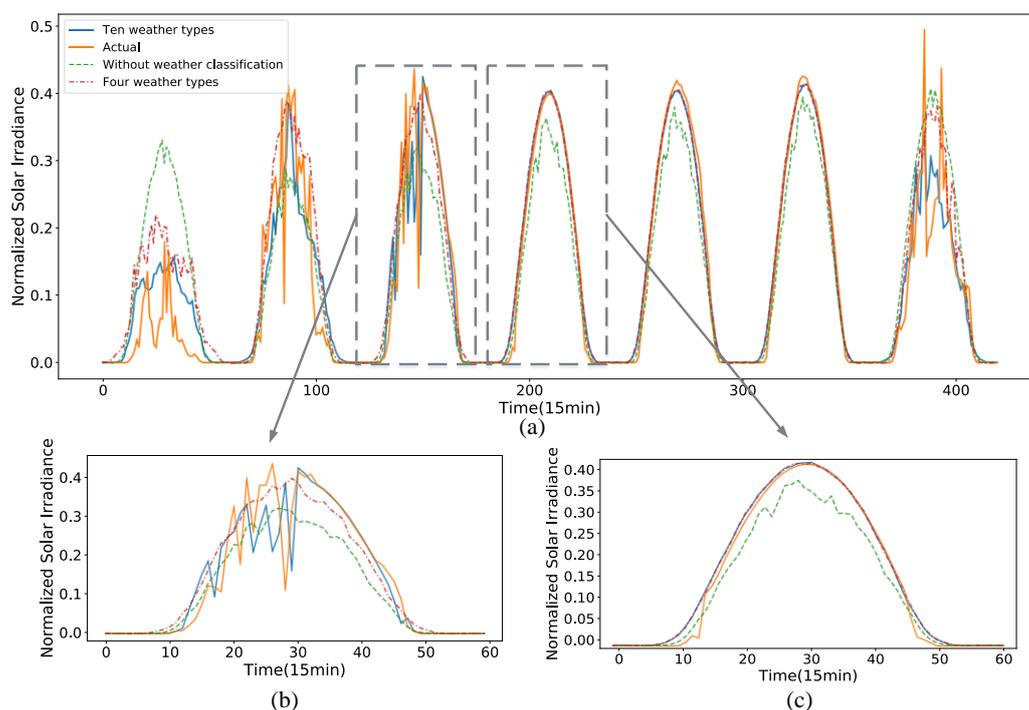


Figure 15 The comparison results of solar irradiance forecasting performance among different forecasting models. In the above three subplots, the orange solid line represents the actual solar irradiance curve. The blue solid line is the solar irradiance curve that is predicted by weather classification based forecasting models (ten forecast models are respectively trained under ten different weather types). The red dotted line is the solar irradiance curve that is predicted by weather classification based forecasting models (four forecast models are respectively trained under four different weather types). The green dotted line is the solar irradiance curve that is predicted by single forecasting models where the influence of weather status is not into consideration.

The subplot (b) is the comparison results under the extreme weather status where the actual solar irradiance curve frequently fluctuates. It is obvious that "Ten weather type" can almost perfectly depicting the true solar irradiance curve. That is, our proposed method is able to accurately predict the uncertainty and fluctuation of solar irradiance curve. But "Four weather type" and "Without weather classification" merely predict the approximate trend of curve so the deviation is large. This further verifies that the weather classification has been proven to be an effective pre-processing step in order to enhance prediction accuracy for short-term solar forecasting [20-22]. Moreover, building more forecasting models that fit for different weather conditions (e.g. ten weather types) can achieve higher precision than using four or one single uniform forecasting model. The reason behind this is that the mapping relationship between the input and output of the forecast model always varies to some extent under different weather statuses. So it is very difficult for a single model to accurately describe the entire mapping relationship between input and output.

34 The subplot (c) is the comparison results under the non-extreme weather status where the actual solar
35 irradiance curve is very smooth. The curves predicted by “Ten weather type” and “Four weather type” both
36 can well match the actual solar irradiance curve, while the curve forecasted by “Without weather
37 classification” is much lower than the true values. In addition, the solar irradiance curve predicted by the
38 traditional single forecasting model (i.e. “Without weather classification”) is easily affected by the previous
39 three days of historical data. That is, “Without weather classification” lacks the ability to identify the weather
40 type of that day. The simulation results also indicate that the introduction of weather classification can
41 obviously improve the forecasting precision.

42 In sum, weather classification model plays a significant role in determining the most suitable and precise
43 day-ahead PVPF model with high efficiency. The accurate classification can significantly improve the
44 forecasting accuracy. In addition, building a suitable number of forecasting models that fit for different
45 weather conditions (e.g. ten weather types) can achieve higher precision than using one single uniform
46 forecasting model, which means the proposed model can be utilized not only for day-ahead PVPF [53-56], but
47 also for some other application scenarios such as the automatic demand response strategy of PV-Assisted EV
48 charging station [57], day-ahead electricity price forecasting [58], optimal dispatch of building energy
49 management system, electricity consumption patterns analysis and combined cooling-heating-power (CCHP)
50 [59-62].

51 4. Conclusions

52 Weather classification model is considered as an effective tool to enhance the PV power forecasting
53 precision, which is important for the operation and planning of the system with the high penetration level of
54 solar PV generation. Moreover, the finer classification of the weather types, the higher the precision of the
55 corresponding PV power forecasting model. Therefore, based on the given label in studied dataset, we
56 reclassify 33 meteorological weather types into 10 weather types by putting several single weather types
57 together to constitute a new weather type. However, the insufficient training data, especially for the extreme
58 weather types, pose great challenges on the weather classification modeling. So given the influence of the
59 training data and applied classifiers, the WGANGP and CNN based weather classification model is proposed
60 to improve the classification performance. The WGANGP is a powerful generative model that is applied to
61 synthesize new and realistic training data samples by mimicking the input samples (i.e. the original training
62 samples). Based on the augmented training dataset, CNN is trained to achieve the accurate classification of
63 weather types subsequently.

64 According to the actual daily solar irradiance data collected by the National Oceanic & Atmospheric
65 Administration (NOAA) Earth System Research Laboratory website, a case study is carried out to validate the
66 effectiveness of our proposed model for weather classification. In this case study, the quality of WGANGP
67 generated data is evaluated via three indexes (i.e. STD, EDD and CDF) and compared with another deep
68 learning generative model named VAE. Furthermore, the comparison of classification accuracy is also
69 conducted among traditional machine learning classification models (MLP, SVM, and KNN) and deep
70 learning classification models (CNN1D and CNN2D). And the accuracy change of these classification models
71 after the application of WGANGP is also investigated. Finally, the weather classification models are applied
72 in solar irradiance forecasting. The simulation results indicate that WGANGP can generate new and distinct
73 samples that capture the intrinsic features of the original data, but not to simply memorize the training data. In
74 addition, deep learning models, especially the CNN2D, show better classification performance than traditional
75 machine learning models. And the great improvement of the performance has been achieved for these
76 classification models via the WGANGP based data augment. Moreover, the application of weather
77 classification model validates that the accurate classification can significantly improve the forecasting
78 accuracy.

79 In summary, the proposed model provides a feasible approach to address the difficulty caused by the
80 small sample size of some certain classes during the establishment of classification model. As for the future
81 work, we will pay efforts to the combination of the GAN and manual feature extraction methods in order to
82 further improve the performance of weather classification models. At the same time, we want to investigate
83 how the training sample size affects the classification accuracy of the models, as well as how the imbalanced
84 distribution of original training data in the ten weather types influence the quality of WGANGP generated
85 data and the performance improvement of classification models.

86 **Acknowledgments:** This work was supported by the National Key R&D Program of China (grant No.
 87 2018YFB0904200), the National Natural Science Foundation of China (grant No. 51577067), the Beijing Natural Science
 88 Foundation of China (grant No. 3162033), the State Key Laboratory of Alternate Electrical Power System with Renewable
 89 Energy Sources (grant No. LAPS18008), the Fundamental Research Funds for the Central Universities (grant No.
 90 2018QN077) and the Science and Technology Project of State Grid Corporation of China (SGCC) (grant No. NY7116021,
 91 kjgw2018-014). J.P.S. Catalão acknowledges the support by FEDER funds through COMPETE 2020 and by Portuguese
 92 funds through FCT, under Projects SAICT-PAC/0004/2015 - POCI-01-0145-FEDER-016434,
 93 POCI-01-0145-FEDER-006961, UID/EEA/50014/2013, UID/CEC/50021/2013, UID/EMS/00151/2013, and
 94 02/SAICT/2017 - POCI-01-0145-FEDER-029803.309048.

95 References

- 96 [1] Pandey AK, Tyagi V V., Selvaraj JA, Rahim NA, Tyagi SK. Recent advances in solar photovoltaic systems for
 97 emerging trends and advanced applications. *Renew Sustain Energy Rev* 2016;53:859–84.
 98 doi:10.1016/j.rser.2015.09.043.
- 99 [2] Wang F, Mi Z, Su S, Zhao H. Short-term solar irradiance forecasting model based on artificial neural network using
 100 statistical feature parameters. *Energies* 2012;5:1355–70. doi:10.3390/en5051355.
- 101 [3] International Energy Agency. *World Energy Outlook 2017*. OECD IEA; 2017.
- 102 [4] REN21. *Renewables 2017 Global Status Report*. 2017.
- 103 [5] Peng J, Lu L. Investigation on the development potential of rooftop PV system in Hong Kong and its environmental
 104 benefits. *Renew Sustain Energy Rev* 2013;27:149–62. doi:10.1016/j.rser.2013.06.030.
- 105 [6] Wang F, Li K, Liu C, Mi Z, Shafie-khah M, Catalao JPS. Synchronous Pattern Matching Principle Based Residential
 106 Demand Response Baseline Estimation: Mechanism Analysis and Approach Description. *IEEE Trans Smart Grid*
 107 2018; 9(6): 6972-6985. doi:10.1109/TSG.2018.2824842.
- 108 [7] Wang F, Xu H, Xu T, Li K, Shafie-khah M, Catalão JPS. The values of market-based demand response on improving
 109 power system reliability under extreme circumstances. *Appl Energy* 2017;193:220–31.
 110 doi:10.1016/j.apenergy.2017.01.103.
- 111 [8] Wang F, Zhou L, Ren H, Liu X, Shafie-khah M, Catalão JPS. Multi-objective optimization model of
 112 source-load-storage synergetic dispatch for building energy system based on tou price demand response. 2017 IEEE
 113 Ind Appl Soc Annu Meet IAS 2017 2017;2017–January:1–10. doi:10.1109/IAS.2017.8101713.
- 114 [9] Wang F, Zhen Z, Liu C, Mi Z, Hodge BM, Shafie-khah M, et al. Image phase shift invariance based cloud motion
 115 displacement vector calculation method for ultra-short-term solar PV power forecasting. *Energy Convers Manag*
 116 2018;157:123–35. doi:10.1016/j.enconman.2017.11.080.
- 117 [10] Kasten F, Czeplak G. Solar and terrestrial radiation dependent on the amount and type of cloud. *Sol Energy*
 118 1980;24:177–89. doi:https://doi.org/10.1016/0038-092X(80)90391-6.
- 119 [11] Nann S, Riordan C. Solar spectral irradiance under clear and cloudy skies: Measurements and a semiempirical model.
 120 *J Appl Meteorol* 1991;30:447–62. doi:10.1175/1520-0450(1991)030<0447:SSIUCA>2.0.CO;2.
- 121 [12] Kaskaoutis DG, Kambezidis HD, Jacovides CP, Steven MD. Modification of solar radiation components under
 122 different atmospheric conditions in the Greater Athens Area, Greece. *J Atmos Solar-Terrestrial Phys* 2006;68:1043–
 123 52. doi:10.1016/j.jastp.2005.05.002.
- 124 [13] Badarinath KVS, Kharol SK, Kaskaoutis DG, Kambezidis HD. Influence of atmospheric aerosols on solar spectral
 125 irradiance in an urban area. *J Atmos Solar-Terrestrial Phys* 2007;69:589–99. doi:10.1016/j.jastp.2006.10.010.
- 126 [14] Sun Y, Wang F, Wang B, Chen Q, Engerer NA, Mi Z. Correlation feature selection and mutual information theory
 127 based quantitative research on meteorological impact factors of module temperature for solar photovoltaic systems.
 128 *Energies* 2017;10. doi:10.3390/en10010007.
- 129 [15] Ishii T, Otani K, Takashima T, Xue Y. Solar spectral influence on the performance of photovoltaic (PV) modules
 130 under fine weather and cloudy weather conditions. *Prog Photovoltaics Res Appl* 2013;21:481–489.

- 131 [16] Lima FJL, Martins FR, Pereira EB, Lorenz E, Heinemann D. Forecast for surface solar irradiance at the Brazilian
132 Northeastern region using NWP model and artificial neural networks. *Renew Energy* 2016;87:807–18.
133 doi:10.1016/j.renene.2015.11.005.
- 134 [17] Shah ASBM, Yokoyama H, Kakimoto N. High-Precision Forecasting Model of Solar Irradiance Based on Grid
135 Point Value Data Analysis for an Efficient Photovoltaic System. *IEEE Trans Sustain Energy* 2015;6:474–81.
136 doi:10.1109/TSTE.2014.2383398.
- 137 [18] Engerer NA. Minute resolution estimates of the diffuse fraction of global irradiance for southeastern Australia. *Sol*
138 *Energy* 2015;116:215–37. doi:10.1016/j.solener.2015.04.012.
- 139 [19] Wang F, Zhen Z, Liu C, Mi Z, Shafie-Khah M, Catalão JPS. Time-section fusion pattern classification based
140 day-ahead solar irradiance ensemble forecasting model using mutual iterative optimization. *Energies* 2018;11.
141 doi:10.3390/en11010184.
- 142 [20] Huang CMT, Huang YC, Huang KY. A weather-based hybrid method for 1-day ahead hourly forecasting of PV
143 power output. *IEEE Trans Sustain ENERGY* 2014;5:917–26. doi:10.1109/TSTE.2014.2313600.
- 144 [21] Shi J, Lee WJ, Liu Y, Yang Y, Wang P. Forecasting power output of photovoltaic systems based on weather
145 classification and support vector machines. *IEEE Trans Ind Appl* 2012;48:1064–9. doi:10.1109/TIA.2012.2190816.
- 146 [22] Chen C, Duan S, Cai T, Liu B. Online 24-h solar power forecasting based on weather type classification using
147 artificial neural network. *Sol Energy* 2011;85:2856–2870. doi:10.1016/j.solener.2011.08.027.
- 148 [23] Changsong C, Shanxu D, Jinjun Y. Design of photovoltaic array power forecasting model based on neural network.
149 *Trans China Electrotech Soc* 2009;24:153–158. doi:10.19595/j.cnki.1000-6753.tces.2009.09.023.
- 150 [24] Wang F, Zhen Z, Mi Z, Sun H, Su S, Yang G. Solar irradiance feature extraction and support vector machines based
151 weather status pattern recognition model for short-term photovoltaic power forecasting. *Energy Build* 2015;86:427–
152 38. doi:10.1016/j.enbuild.2014.10.002.
- 153 [25] Wang F, Zhen Z, Wang B, Mi Z. Comparative Study on KNN and SVM Based Weather Classification Models for
154 Day Ahead Short Term Solar PV Power Forecasting. *Appl Sci* 2017;8:28. doi:10.3390/app8010028.
- 155 [26] He K, Zhang X, Ren S, Sun J. Delving deep into rectifiers: Surpassing human-level performance on imagenet
156 classification. *Proc IEEE Int Conf Comput Vis* 2015:1026–34. doi:10.1109/ICCV.2015.123.
- 157 [27] Sezer OB, Ozbayoglu AM. Algorithmic financial trading with deep convolutional neural networks : Time series to
158 image conversion approach. *Appl Soft Comput* 2018;70:525–38. doi:10.1016/j.asoc.2018.04.024.
- 159 [28] Ferreira A, Giraldi G. Convolutional Neural Network approaches to granite tiles classification. *Expert Syst Appl*
160 2017;84:1–11. doi:10.1016/j.eswa.2017.04.053.
- 161 [29] Wang Y, Li H. A novel intelligent modeling framework integrating convolutional neural network with an adaptive
162 time-series window and its application to industrial process operational optimization. *Chemom Intell Lab Syst*
163 2018;179:64–72. doi:10.1016/j.chemolab.2018.06.008.
- 164 [30] Douzas G, Bacao F. Effective data generation for imbalanced learning using conditional generative adversarial
165 networks. *Expert Syst Appl* 2018;91:464–71. doi:10.1016/j.eswa.2017.09.030.
- 166 [31] Chawla N V, Bowyer KW, Hall LO. SMOTE: Synthetic Minority Over-sampling Technique. *J Artif Intell Res*
167 2002;16:321–57.
- 168 [32] He H, Garcia EA. Learning from Imbalanced Data. *IEEE Trans Knowl Data Eng* 2009;21:1263–84.
- 169 [33] Chen Z, Jiang C. Building occupancy modeling using generative adversarial network. *Energy Build* 2018;174:372–9.
170 doi:10.1016/j.enbuild.2018.06.029.
- 171 [34] Creswell A, White T, Dumoulin V, Kai A, Sengupta B, Bharath AA. Generative Adversarial Networks: An
172 Overview. *IEEE Signal Process Mag* 2018;35:53–65.

- 173 [35] Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative Adversarial Networks,
174 2014, p. 2672–80. doi:10.1001/jamainternmed.2016.8245.
- 175 [36] Wang K, Gou C, Duan Y, Lin Y, Zheng X, Wang FY. Generative Adversarial Networks: Introduction and Outlook.
176 IEEE/CAA J Autom Sin 2017;4:588–98.
- 177 [37] China Meteorological Administration, Public Meteorological Service-Weather Graphic Symbols, China
178 Meteorological Administration, Beijing, 2008, GB/T22164-2008.
- 179 [38] Mirza M, Osindero S. Conditional Generative Adversarial Nets. Comput Sci 2014:2672–80.
- 180 [39] Radford A, Metz L, Chintala S. Unsupervised Representation Learning with Deep Convolutional Generative
181 Adversarial Networks. Comput Sci 2015.
- 182 [40] Bousmalis K, Silberman N, Dohan D, Erhan D, Krishnan D. Unsupervised Pixel-Level Domain Adaptation with
183 Generative Adversarial Networks 2016:95–104.
- 184 [41] Arjovsky M, Bottou L. Towards Principled Methods for Training Generative Adversarial Networks 2017.
- 185 [42] Arjovsky M, Chintala S, Bottou L. Wasserstein GAN 2017.
- 186 [43] Cédric Villani. Optimal Transport: Old and New. Grundlehren der mathematischen Wissenschaften. Springer,
187 Berlin 2009.
- 188 [44] Gulrajani I, Ahmed F, Arjovsky M, Dumoulin V, Courville A. Improved Training of Wasserstein GANs 2017.
- 189 [45] Chen, Y.; Wang, Y.; Kirschen, D.; Zhang, B. Model-Free Renewable Scenario Generation Using Generative
190 Adversarial Networks. IEEE Trans Power Systems 2018, 33, 3265–3275, doi:10.1109/TPWRS.2018.2794541.
- 191 [46] Gu J, Wang Z, Kuen J, Ma L, Shahroudy A, Shuai B, et al. Recent Advances in Convolutional Neural Networks.
192 Comput Sci 2015.
- 193 [47] Nair V, Hinton GE. Rectified Linear Units Improve Restricted Boltzmann Machines. Proc 27th Int Conf Mach
194 Learn 2010:807–14. doi:10.1.1.165.6419.
- 195 [48] Bland JM, Altman DG. Statistics notes: measurement error. Bmj 1996;313:744. doi:10.1136/bmj.313.7059.744.
- 196 [49] Kingma DP, Welling M. Auto-Encoding Variational Bayes 2013.
- 197 [50] Deza MM, Deza E. Encyclopedia of distances. 2009. doi:10.1007/978-3-642-00234-2.
- 198 [51] Miller FP, Vandome AF, Mcbrewster J. Cumulative Distribution Function. Springer US; 2006.
- 199 [52] Stehman S V. Selecting and interpreting measures of thematic classification accuracy. Remote Sens Environ
200 1997;62:77–89. doi:10.1016/S0034-4257(97)00083-7.
- 201 [53] Wang Z, Wang F, Su S. Solar Irradiance Short-Term Prediction Model Based on BP Neural Network. Energy
202 Procedia 2011;12: 488-494. doi:10.1016/j.egypro.2011.10.065.
- 203 [54] Wang F, Mi Z, Su S, Zhang C. A practical model for single-step power prediction of grid-connected PV plant using
204 artificial neural network. 2011 IEEE PES Innovative Smart Grid Technologies (2011). doi:
205 10.1109/ISGT-Asia.2011.6167097.
- 206 [55] Sun Y, Wang F, Zhen Z, Mi Z, Liu C, Wang B, Lu J. Research on short-term module temperature prediction model
207 based on BP neural network for photovoltaic power forecasting. 2015 IEEE Power & Energy Society General
208 Meeting (2015). doi: 10.1109/PESGM.2015.7286350.
- 209 [56] Wang F, Mi Z, Zhen Z, Yang G, Zhao H. A classified forecasting approach of power generation for photovoltaic
210 plants based on weather condition pattern recognition. Proc Chin Soc Electri Eng 2013;33(34):75–82.
211 <https://doi.org/10.13334/j.0258-8013.pcsee.2013.34.013>.
- 212 [57] Chen Q, Wang F, Hodge B.M., Zhang J, Li Z, Shafie-Khah M, Catalão J.P.S. Dynamic price vector formation
213 model-based automatic demand response strategy for PV-assisted EV charging stations. IEEE Trans Smart Grid
214 2017, 8(6): 2903-2915. doi: 10.1109/TSG.2017.2693121.

- 215 [58] Wang F, Li K, Zhou L, Ren H, Contreras J, Shafie-Khah M, Catalão J.P.S. Daily pattern prediction based
216 classification modeling approach for day-ahead electricity price forecasting. *Int J Electr Power Energy Syst* 2019,
217 105: 529-540.doi: 10.1016/j.ijepes.2018.08.039.
- 218 [59] Wang F, Zhou L, Ren H, Liu X, Talari S, Shafie-khah M, Catalão J.P.S. Multi-objective optimization model of
219 source-load-storage synergetic dispatch for building energy system based on TOU price demand response. *IEEE*
220 *Trans Industry Appl* 2017, 54(2): 1017-1028. doi: 10.1109/TIA.2017.2781639.
- 221 [60] Wang F, Zhou L, Ren H, Liu X. Search Improvement Process-Chaotic Optimization-Particle Swarm
222 Optimization-Elite Retention Strategy and Improved Combined Cooling-Heating-Power Strategy Based Two-Time
223 Scale Multi-Objective Optimization Model for Stand-Alone Microgrid Operation. *Energies* 2017;10. doi:
224 10.3390/en10121936.
- 225 [61] Wang F, Li K, Duić N, Mi Z, Hodge BM, Shafie-khah M, Catalão J.P.S. Association rule mining based quantitative
226 analysis approach of household characteristics impacts on residential electricity consumption patterns. *Energy*
227 *Convers Manag* 2018;171:839–854. doi: 10.1016/j.enconman.2018.06.017.
- 228 [62] Wang F, Zhou L, Wang B, Wang Z, Shafie-Khah M, Catalão J.P.S. Modified Chaos Particle Swarm
229 Optimization-Based Optimized Operation Model for Stand-Alone CCHP Microgrid. *Appl Sci* 2017; 7(8). doi:
230 10.3390/app7080754.