On the classification of weather based on the production of photovoltaic installations

Paweł Parczyk[a,*], **Tobiasz Puślecki**[a] **and Robert Burduk**[a]

[a]Wrocław University of Science and Technology, Department of Systems and Computer Networks, Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland
ORCID (Paweł Parczyk): https://orcid.org/0009-0004-3287-9520, ORCID (Tobiasz Puślecki): https://orcid.org/0000-0002-4665-3301, ORCID (Robert Burduk): https://orcid.org/0000-0002-3506-6611

Abstract. In recent years, there has been an energy transition in which fossil fuel-fired power plants are being replaced by renewable energy sources such as photovoltaics. While the influence of factors such as location and method of installation or the sun's position above the horizon is deterministic, the weather factor is rather random. Weather data can be easily obtained from nearby weather stations, but some weather phenomena are so dynamic that data obtained in this way can become useless. When considering the problem of classifying weather conditions based on production data of photovoltaic installations, we propose two new features extracted from the production data that describe the overall level of sunshine and the variability of sunlight. Three supervised learning methods were used to classify the dataset: CNN, Random Forest classifier and decision tree classifier. Reference values were obtained by generating classification results from the raw output and compared with those generated from the two new features. The experiments showed a statistically significant advantage for the models fitted to the data extended with the new features.

1 Introduction

A time series is a set of observations ordered in time, usually made at a constant interval between subsequent samples. They are often used, among others, to analyze changes in stock market indices, speech, and physical or biological phenomena. Consequently, many methods have been developed over the years for analyzing, classifying, and forecasting time series [6].

Nevertheless, working with time series is a non-trivial task. The key aspect of the analysis leading to satisfactory results is the methods of extracting features from raw data. Moreover, each field in which time series are used requires specialized expert knowledge necessary to verify the obtained results.

For some time series, e.g.: stock market indices, the main task is forecasting, because knowing future values allows better decisions. In contrast, for other series e.g.: ECG, the task of classification is crucial, for the reason that it allows the earlier reaction to the occurrence of alarming conditions of the patient. The electrocardiogram, in addition to the heart rate and rhythm information visible at a glance, provides a wealth of information about the electrical activity of the heart and can help cardiologists diagnose and monitor various heart conditions (e.g., ischemia, myocardial infarction, pericarditis or structural abnormalities) [7].

The same is true for data from photovoltaic installations. Predicting power values is crucial for many different reasons. Appropriate management gives greater control over the utility grid, allowing for more precise scheduling of backup energy sources (e.g., fossil fuel power plants). Therefore, this issue is frequently discussed in the literature. The authors use models of varying complexity to generate forecasts. Models based on artificial neural networks are often used, in particular, LSTM [16, 8, 2, 4]. Another aspect is the set of features. The use of models for which the only input parameter is past production is fraught with errors already at the planning stage - Past data does not contain information about the weather.

Some authors take this phenomenon into account. One solution to this problem is to build several predictors for different weather conditions, as shown in [10]. Some studies in which models were trained on extended data sets can be found in the literature. In particular, on the weather data such as temperature, humidity, or wind direction [12], [9]. [1] also uses solar irradiance information. In addition to historical weather, weather forecasts (NWP) [15], [14] can be taken into account. In the case of building many predictors, precise weather classification is necessary, which allows the division of training data among the models.

However, good quality weather data is rarely available along with installation production data. This is due to the fact that the residential installations are not equipped with weather measurement equipment, so the only method of obtaining weather data is to download it from a nearby weather station. However, due to the distance, this data is susceptible to errors. To meet this problem, two novel features, `Overall Cloudiness` and `Variability Cloudiness`, that allow accurate weather classification based on the production of a photovoltaic installation are proposed in this work.

The main contribution of this paper is as follows:

1. Proposing an algorithm for extracting two novel features, `Overall Cloudiness` and `Variability Cloudiness`, that increase classification accuracy during the weather type classification.
2. Evaluation of the impact of the proposed `Overall Cloudiness` and `Variability Cloudiness` features on the quality of classifications performed by a selected set of classifiers using cross-validation and statistical tests.

* Corresponding Author. Email: pawel.parczyk@pwr.edu.pl

3. Providing dataset and manual annotations of weather data prepared by experts.

The rest of this article is organized as follows. Section 1 contains the introduction. The considered problem is defined in Section 2. The dataset used in this work is shown and discussed in Section 3. Section 4 presents the assumptions of the method proposed in the article. Section 5 is devoted to presenting and discussing the results of the conducted experiments, which are summarized in Section 6.

2 Problem description

The main issue addressed in this work is the classification of photovoltaic installation production data to classes describing the weather. In other words, the task is to determine which periods were sunny, cloudy, completely cloudy, etc., using only the production information.

This issue is worth raising because access to precise weather data provides great opportunities for forecasting. Unfortunately, access to such data is very limited because most photovoltaic installations are not equipped with devices enabling weather analysis, so there is a shortage of good data but the nearby weather station data can be used instead.

However, the impact of weather on photovoltaic installations is an extremely complex phenomenon. Small changes in the general weather are enough to observe a large change in production. For example, the appearance of a small but dense (and therefore not transparent) cloud in the sky, which will cast a shadow directly on the photovoltaic installation, may result in a significant power reduction. In addition, there are weather conditions in which objects located at a very short distance from each other (several hundred meters) may receive completely different doses of solar radiation.

Figure 1. A photo showing the impact of a single cloud during a sunny day on the amount of light reaching the earth's surface. On the left one can see a shadow falling on the mountains, and on the right there is a very brightly illuminated surface. In the central part, a sharp border between the shadow zone and the shadowless zone can be observed. Location: 49.24668N, 19.88683E

This situation is presented in the Figure 1, which shows a mountain landscape on a sunny day with single clouds in the sky. So, the weather may be defined as sunny or even clear sunny weather. As can be seen on the left-hand side of the Figure, the mountains are strongly shaded, while on the right-hand side, the area is strongly illuminated. Observation of this landscape leads to the conclusion that there may be two areas with completely different levels of sunlight close to each other.

Assuming the location of the weather station used to mark the time series obtained from the photovoltaic installation at the right-hand side of the Figure or the camera position, and the installation location a few kilometers away in the shaded part of the image. The assigned weather (the ground truth) will be subject to an error.

This situation can be observed in real data, as shown in Figure 2. The too-frequent marking of weather as cloudy is clearly visible, as both 14/07/22 and 18/07/22 were classified almost entirely as cloudy days despite visible differences. Moreover, observing the production shape during the morning of 18/07/22 and 19/07/22, it can be assumed that the weather was similar. However, automatically marked data contradicts this.

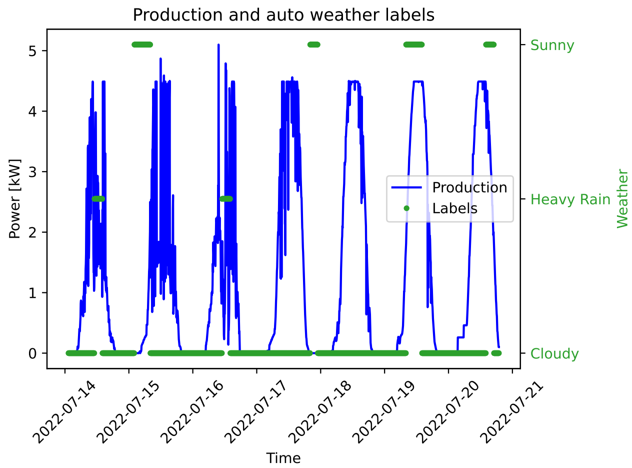

Figure 2. Automatically assigned weather condition

Enriching the model with additional parameters may be strongly limited by significant errors in data acquisition and may not bring the expected increases in the quality of solutions.

Considering the described limitations of the weather data, this paper presents algorithms for extracting `Variability Cloudiness` and `Overall Cloudiness` features, allowing for weather classification based on raw production data. The usefulness of the proposed features was verified by performing classification on the prepared data set. Models based on three supervised learning methods were used: CNN-ANN, Random Forest Classifier (RFC) and Decision Tree Classifier (DTC). First, all models were fitted to raw production data. The classification results obtained in this way were treated as reference results. Then, the same supervised learning models were fitted to the data by previously extracting the `Variability Cloudiness` and `Overall Cloudiness` features from them. The obtained results were compared with reference values.

3 Dataset

To conduct the experiments described in this article, data was collected from a 6500W photovoltaic installation located in Poland in the Lower Silesian Voivodeship near Wrocław. The dataset covers the period from 30/06/2022 to 30/03/2024. Samples were collected every 5 minutes. The total size of the dataset is 184052 samples.

In order to avoid errors in the weather data, a team of photovoltaics experts manually marked the set containing observations from the photovoltaic installation. When marking, experts used raw production data and information about the observed weather during the following days. This way, each sample that exists in the dataset was assigned to one of 10 categories. The meaning of each category is described in the Table 1. An example period is presented in the figure 3.

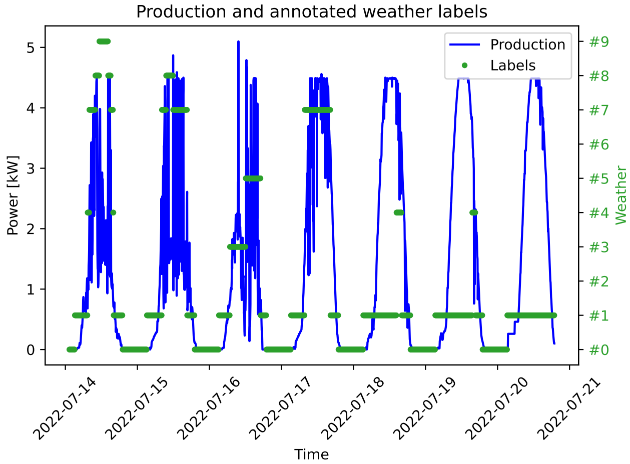

Figure 3. Manually assigned weather condition

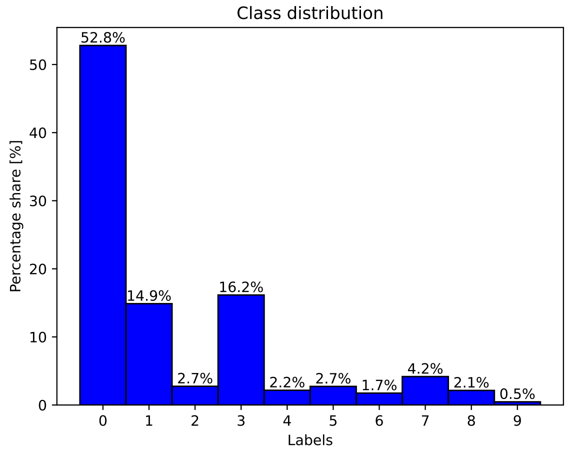

Figure 4. Class distribution in the dataset.

Table 1. List of classes and their meanings used in annotation process

Class no.	Description
0	Night
1	Clear Sunny
2	Semi-transparent clouds
3	Full Cloud Cover (Overcast)
4	Sunny with long-term clouds
5	Semi-transparent clouds with long-term clearings
6	Full Cloud Cover with long-term clearings
7	Sunny with short-term clouds
8	Semi-transparent clouds with short-term clearings
9	Full Cloud Cover with short-term clearings

The share of individual classes in the set is presented in the Figure 4. It can be observed that the data set is highly unbalanced. The largest number of elements is class 0, which represents over 52% of the set. However, this is a class denoting night, so it can be easily filtered out using the hours of sunrise and sunset. Figure 5 shows the distribution of labels after filtering out the night using the proposed solution.

After filtering out the night data, it can be observed that the most elements are class 3 (Full Cloud Cover) - 32.6%, then class 1 (Clear Sunny) - 30.6% and class 7 (Sunny with short-term clouds) - 8.5%. Together, these three classes constitute 71.7% of the dataset. In addition, some data is still marked with 0 (Night), but it only constitutes 3.3%. It can result from observing very cloudy and dark early morning, which was marked as night even if theoretically the sun had already risen.

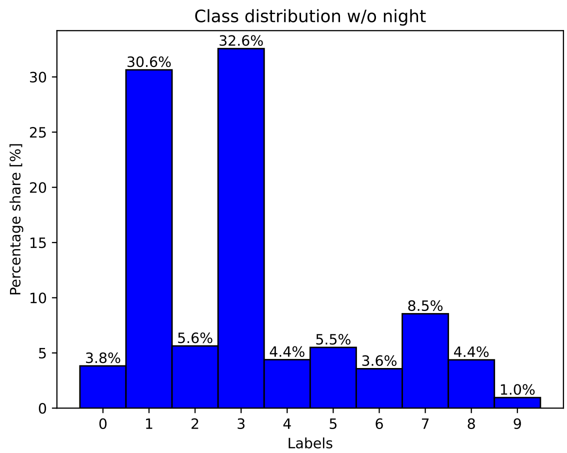

Figure 5. Class distribution in the dataset omitting the night observations

4 Model

When working with data on electricity production from a photovoltaic installation, presented in the form of a time series, one should bear in mind the strong dependence of these data on three factors. The first factor - the location and method of installation of the installation - is unchanged throughout the entire life of the installation, its influence can be observed in the constant nature of production (e.g. installations facing west reach their peak power at a later hour).

Due to this, each installation has its own characteristic shape of daily production, and the differences can be observed by comparing two different installations. The second factor is the position of the sun above the horizon [13]. It is expressed by the elevation angle or altitude value and has a significant impact on the instantaneous power, as shown in the Figure 6. The last factor is the weather, which is rather a random factor without using exogenous data.

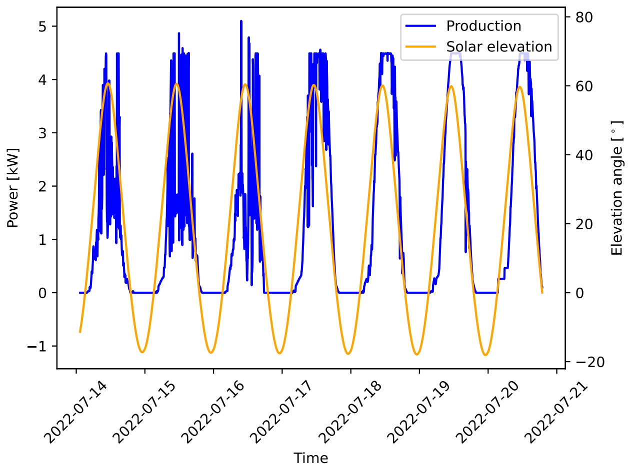

Figure 6. The relation between observer power generation and elevation angle.

4.1 New features

Two characteristic features can be noticed by analyzing the data on electricity production from a photovoltaic installation. The first of these features is weather variability, denoted by `Variability Cloudiness`. One can distinguish days during which power changes occur gently. On the other hand, there are also days when changes occur very dynamically, i.e. large fluctuations are visible - (Compare 3 days from the left and 3 days from the right in the Figure 6). The second characteristic is the average level of sunlight, denoted by `Overall Cloudiness`. If the observed day was darker (with more cloud cover), the average amount of solar energy reaching the earth was limited, and therefore, the average daily electricity production was lower. Since these features are independent of each other, there are 4 basic weather classes that can be recognized using the described features (see Table 2.

Table 2. Basic weather classification classes according to the `Overall Cloudiness` and `Variability Cloudiness` features.

		Variability Index	
		Low	High
Overall Index	Low	Sunny weather	Sunny weather with short clouds
	High	Cloudy weather	Cloudy weather with short showers

Despite the ability to distinguish some particular combinations, feature values remain linear (rather than discrete). The implementation of the features in an example 7-day production window can be seen in the Figure 7.

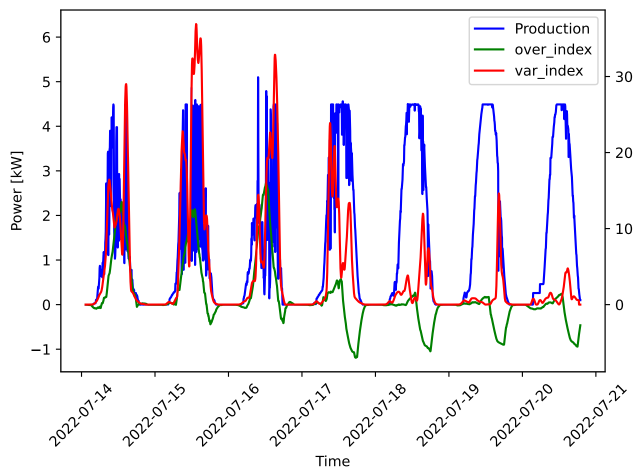

Figure 7. Production with calculated feature values. `Overall Cloudiness` is marked in green and `Variability Cloudiness` is marked in red. The `Overall Cloudiness` feature has a power dimension and is therefore placed on the left axis (Power), the `Variability Cloudiness` feature is dimensionless and is marked on the right axis.

4.2 Overall cloudiness feature

`Overall Cloudiness` feature is an indicator that reflects the degree of average cloud cover during a particular timestamp ts. This means that its value is directly proportional to the difference between the average production (production is marked P) value in a given window w (convolution operation using a uniform filter) $MA(P, w)$ and the expected production value determined from the model $M(P, \phi)$ [13]. The formula is presented in Equation 1.

$$OC(ts, w) = (P \circledast \underbrace{[\frac{1}{w}, \frac{1}{w}, ..., \frac{1}{w}]}_{1\times w})(ts) - M(ts, \phi) \quad (1)$$

Figure 8 presents the course of the feature value and other variables (in the form of time series) that are used in the process of calculating its value. The average production value (after applying the convolution) is marked in orange, while the expected value of the model is marked in red. The subtraction of both values is marked with a pink line. Observing both the upper and lower graphs in Figure 8, it can be seen that as the production value decreases (blue line), the value of the feature increases. However, the expected changes in production during the daytime are considered. This can be seen by comparing the production value in the morning and in the afternoon (bottom graph), where although production is lower in the morning, the feature's value is low, indicating little cloud cover.

4.3 Variablility cloudiness feature

`Variability Cloudiness` feature is an indicator that reflects the rate of change in production. It corresponds to the dynamics of changes in cloud cover in the sky. Frequent and short-term changes correspond to high values, while expected changes, such as an increase in production in the morning and a decrease in the evening, are omitted. Similarly to `Overall Cloudiness`, `Variability Cloudiness` is calculated for each time sample separately ts. However, to calculate it it is necessary to provide the width of the period that is to affect the current timestamp data ts. That value is

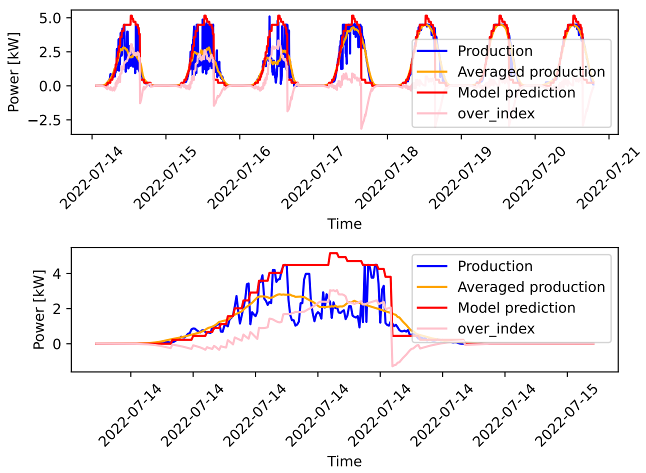

Figure 8. The course of `Overall Cloudiness` feature (pink), as well as its components: averaged production value (orange) and the expected value from the model (red) over 7-day period (top chart) and within 1-day period (bottom chart).

named w_{vc}. The feature value is calculated as the area between the maximums and minimums values in the w_{vc} window. While the maximums and minimums are computed using The max pooling and min pooling functions from the first-order derivative (value of increases) of production. The Equation 2 describes the above procedure.

$$VC(ts, w_{vc}, w) = \sum_{i=-\frac{w_{vc}}{2}}^{\frac{w_{vc}}{2}} MaxPool(P', w)(ts + i) \\ - MinPool(P', w)(ts + i) \tag{2}$$

The operation of the equation is presented in Figure 9, which shows the graphical representation of the metric. The metric value is indicated by the gray box in the bottom chart. This is the area between the two red lines, which result from the max pool and min pool operations on the calculated production first-order derivative (orange line). Looking at the top chart, one can see how the metric works. The first 3 days from the left (top graph) show sunny days with passing clouds. One can see high values of the feature compared to sunny days without clouds (2 days from the right), where the value of the feature is very close to zero.

4.4 *Feature plane*

The use of two independent features allows one to define the plane on which individual observations are located. Figure 10 presents the set of all observations located on a plane with the values of the `Overall Cloudiness` feature on the X axis and the values of the `Variability Cloudiness` feature on the Y axis. Each of the manually marked classes was presented in a different color with a partial transparency set.

Class 1 (clear sunny) and class 3 (Full cloud cover) are clearly visible. They occur on both sides of the Y axis (negative and positive values) and low values `Variability Cloudiness` features. It can also be observed that as the value of the `Variability Cloudiness` feature increases, the samples are closer to 0 on the `Overall Cloudiness` feature axis, which may mean that these

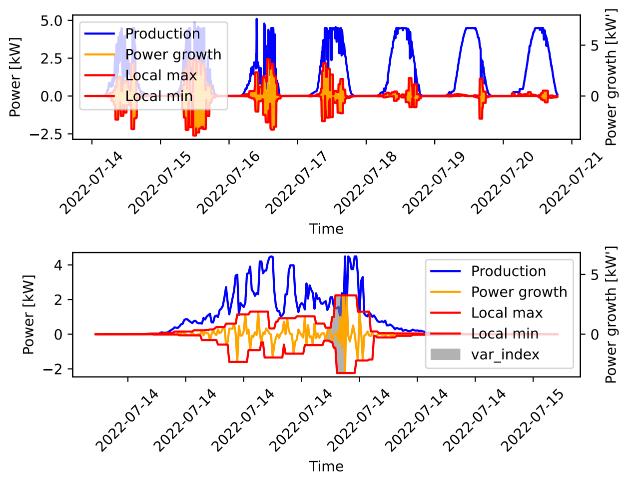

Figure 9. The course of the `Variability Cloudiness` feature (gray field), as well as its components: the value of the production first-order derivative (orange) and the local maximum and minimum values of the derivative (red) over a 7-day period (top chart) and within a 1-day period (bottom chart).

features are not completely independent of each other. The third distinctive class is class 7 (Sunny with short-term clouds), located in the plane's central part.

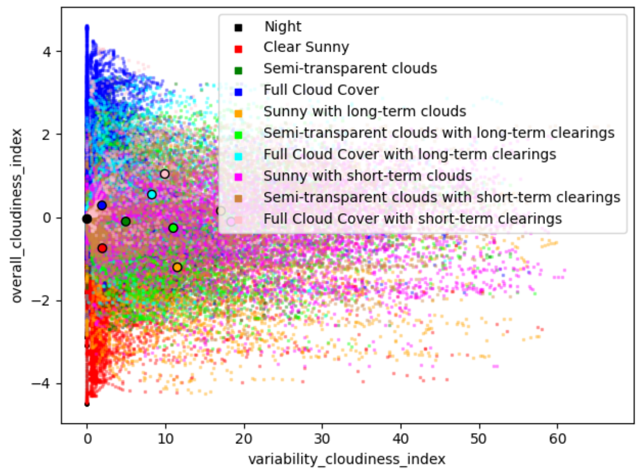

Figure 10. Feature space with labeled samples. Centroids for each class are marked with solid circles.

Moreover, the centroid was calculated for each class, which is the arithmetic mean of all samples from a given class. Centroids are marked with opaque points with a black border. They allow one to determine the expected location of each class.

Due to the large number of samples in the figure discussed, the classes overlap, which does not give full information on the distribution of individual classes on the plane. Therefore, Figure 11 portrays nine planes, one for each class in the database. Class 0 (Night) was omitted because it only occurs near position (0,0). The figure shows the relationship between class 1 (clear sunny) and class 3 (Full cloud cover), and it is also visible that class 2 (Semi-transparent clouds) is located between classes 1 and 3. As the number of classes increases

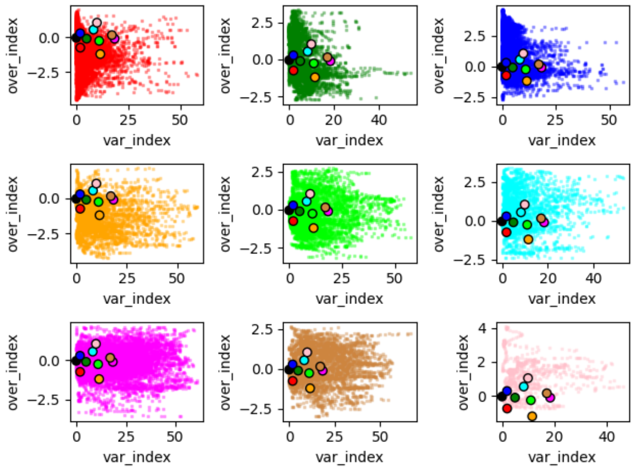

Figure 11. Feature space with marked samples - separate projection for all classes, excluding class 0.

(4-6, 7-9), a distance of the samples from the Y axis can be observed (an increase in the `Variability Cloudiness` value of the feature).

5 Evaluation

The aim of this work is to test the usefulness of the proposed `Overall Cloudiness` and `Variability Cloudiness` features in the task of weather classification based on data from a photovoltaic installation. Three models known from the literature will be used as classifiers: CNN-ANN, Random Forest Classifier, and Decision Tree Classifier. All models will be trained on production data, taking subsequent samples as input features. The classification results obtained in this way will be used as reference results. The same models will then be trained on data containing the generated `Overall Cloudiness` and `Variability Cloudiness` features and compared with reference methods.

All experiments were implemented in Python language, using scikit-learn [11], numpy [5] and Keras [3] libraries. When conducting experiments, the models will be cross-validated 5 times using StratifiedKFold. The results from each cross-validation fold will be averaged. Due to the large disproportion in the share of individual classes in the data set, this should be considered when selecting training and testing subsets.

Finally, the results obtained by each model trained on the production itself will be tested for statistically significant differences between the same model trained on the same data preprocessed by `Overall Cloudiness` and `Variability Cloudiness` features.

5.1 Reference results

Classifiers were evaluated based on raw production data according to the experimental protocol described above. Table 3 presents the average values of selected metrics (along with standard deviations) for selected classifiers. The best results for all three metrics were achieved by the Random Forest Classifier. The worst results in terms of accuracy and F1 metrics were achieved by the Decision Tree Classifier, and for the balanced accuracy metric, the CNN network performed the worst.

Table 3. The values of selected metrics [%] (std) for selected methods on reference dataset

	accuracy	balanced accuracy	f1 score
CNN	0.730 (0.004)	0.289 (0.008)	0.693 (0.005)
RFC	0.745 (0.003)	0.317 (0.003)	0.714 (0.003)
DTC	0.696 (0.002)	0.303 (0.004)	0.681 (0.002)

5.2 Experimental results

The classifiers were evaluated based on production data extended with the extracted features `Variability Cloudiness` and `Overall Cloudiness`, in accordance with the experimental protocol described above. The Table 4 presents the average values of selected metrics (along with standard deviations) for selected classifiers. The best results for all three metrics, similarly to the reference set, were obtained by the Random Forest Classifier. The worst results, taking into account all metrics, were achieved by CNN. The optimal hyperparameter values of the selected models were determined during previous tests. Table 5 presents the average values of selected metrics for selected classifiers for each class. Analyzing the classification efficiency of classifiers in individual classes, it can be seen that the best results were achieved for the classes with the largest representation in the set (i.e., 0, 1, 3, and 7).

Table 4. The values of selected metrics [%] (std) for selected methods on extended dataset

	accuracy	balanced accuracy	f1 score
CNN	0.901 (0.002)	0.580 (0.011)	0.894 (0.002)
RFC	0.939 (0.001)	0.760 (0.006)	0.937 (0.001)
DTC	0.918 (0.001)	0.736 (0.006)	0.917 (0.001)

Table 5. The values of selected metrics [%] per class for selected methods on extended dataset

	Accuracy			Balanced Accuracy			F1 Score		
	CNN	RFC	DTC	CNN	RFC	DTC	CNN	RFC	DTC
0	0.95	0.97	0.96	0.95	0.97	0.96	0.96	0.97	0.97
1	0.93	0.96	0.94	0.93	0.96	0.94	0.90	0.94	0.92
2	1.00	1.00	1.00	0.67	0.83	0.82	0.48	0.71	0.64
3	1.00	1.00	1.00	0.85	0.91	0.93	0.79	0.89	0.87
4	0.97	0.98	0.98	0.60	0.76	0.77	0.28	0.61	0.56
5	0.99	1.00	1.00	0.71	0.80	0.78	0.41	0.63	0.55
6	1.00	1.00	1.00	0.51	0.74	0.77	0.03	0.59	0.55
7	0.98	0.99	0.98	0.87	0.94	0.90	0.76	0.87	0.80
8	0.99	0.99	0.99	0.82	0.90	0.84	0.64	0.78	0.68
9	0.99	0.99	0.99	0.94	0.95	0.91	0.78	0.88	0.83

5.3 Results

A statistical evaluation of the results was performed to obtain information about statistically significant differences. Tables 6 7 8 present statistically significant differences for individual metrics. If there is an check mark at the intersection of a column and a row, then the column classifier is statistically significantly better than the row classifier. It is noteworthy that in each case considered, the methods evaluated on the extended set performed statistically significantly better than the same methods evaluated on the reference set. Moreover, for the augmented set and for each metric, the CNN network and Decision Tree Classifier were better than Random Forest Classifier. For the accuracy and F1 Score metrics, Decision Tree Classifier performed better than the CNN network. For the balanced accuracy metric, the CNN network was statistically significantly better than Decision Tree Classifier. The optimal hyperparameter values of the selected models were determined during previous tests.

Table 6. Statistical significant differences for accuracy metric.

	CNN	RFC	DTC	CNN_e	RFC_e	DTC_e
CNN				✓	✓	✓
RFC	✓		✓	✓	✓	✓
DTC	✓			✓	✓	✓
CNN_e						✓
RFC_e				✓	✓	✓
DTC_e						✓

	CNN	RFC	DTC	CNN_e	RFC_e	DTC_e
CNN				✓	✓	✓
RFC	✓		✓	✓	✓	✓
DTC	✓			✓	✓	✓
CNN_e						
RFC_e				✓		✓
DTC_e				✓		

Table 7. Statistical significant differences for balanced accuracy metric

Table 8. Statistical significant differences for F1 Score metric

	CNN	RFC	DTC	CNN_e	RFC_e	DTC_e
CNN				✓	✓	✓
RFC	✓		✓	✓	✓	✓
DTC	✓			✓	✓	✓
CNN_e						✓
RFC_e				✓		✓
DTC_e						

The experiments performed showed the existence of statistically significant differences in the accuracy of the classification performed by the model, which was prepared on data extended with the `Variability Cloudiness` and `Overall Cloudiness` features proposed in this work. A good indicator of the quality of the obtained classifications is the balanced accuracy metric, which takes into account the unbalanced nature of the data set. For reference models, the best result was achieved by the Random Forest Classifier and amounted to (0.317 ± 0.003), which is an unsatisfactory result. The same classifier trained on the extended set achieved a much better result (as confirmed by the statistical analysis) (0.76 ± 0.006).

The experiments conducted in this article confirm the statistically significant impact of the proposed `Variability Cloudiness` and `Overall Cloudiness` features on the quality of classifications made by state-of-the-art classifiers.

6 Conclusion

This paper considered the problem of classifying weather conditions based on energy production data generated by a photovoltaic installation. Weather data can be easily obtained from nearby (for installation) weather stations, but some weather phenomena are so dynamic that the data obtained in this way may become useless.

This work proposes two novel features extracted from the production data `Overall Cloudiness` and `Variability Cloudiness`, which describe the general level of sunlight and the variability of sunlight. Then, three supervised learning methods were used to classify the dataset: CNN, Random Forest Classifier, and Decision Tree Classifier. Reference values were obtained by generating classification results based on raw production and compared with results generated based on the `Overall Cloudiness` and `Variability Cloudiness` features.

The experiments showed a statistically significant advantage of models fitted to data extended with `Overall Cloudiness` and `Variability Cloudiness` features. The best was Random Forest Classifier with a balanced accuracy result of (0.76 ± 0.006), compared to the result (0.317 ± 0.003) the same model trained on raw production data.

The features proposed in this work, extracted from the raw production data of a photovoltaic installation, can be used in models used to forecast future production values as filters helping to clean the data or simply as features. Moreover, the proposal in this work mechanisms can be used to assess the quality of datasets marked by weather stations.

Further research will mainly focus on eliminating the weak points of the proposed methods, searching for the best-fitting classifiers, and finally implementing the features into the forecasting model.

References

[1] B. Ankit *et al.* Sliding window approach with first-order differencing for very short-term solar irradiance forecasting using deep learning models. *Sustainable Energy Technologies and Assessments*, 50:101864, 2022. ISSN 2213-1388. doi: https://doi.org/10.1016/j.seta.2021.101864.

[2] Y. Cheng *et al.* Short-term photovoltaic power forecasting based on similar day selection and deep neural network. In *2023 42nd Chinese Control Conference (CCC)*, pages 6323–6328, 2023. doi: 10.23919/CCC58697.2023.10240979.

[3] F. Chollet *et al.* Keras. https://keras.io, 2015.

[4] G. L. *et al.* Short-term forecasting approach based on bidirectional long short-term memory and convolutional neural network for regional photovoltaic power plants. *Sustainable Energy, Grids and Networks*, 34: 101019, 2023. ISSN 2352-4677. doi: https://doi.org/10.1016/j.segan.2023.101019.

[5] C. Harris *et al.* Array programming with NumPy. *Nature*, 585(7825): 357–362, Sept. 2020. doi: 10.1038/s41586-020-2649-2.

[6] X. He. A survey on time series forecasting. In S. Patnaik, R. Kountchev, Y. Tai, and R. Kountcheva, editors, *3D Imaging—Multidimensional Signal Processing and Deep Learning*, pages 13–23, Singapore, 2023. Springer Nature Singapore.

[7] N. Herring and D. Paterson. ECG diagnosis of acute ischaemia and infarction: past, present and future. *QJM: An International Journal of Medicine*, 99(4):219–230, 02 2006. ISSN 1460-2725. doi: 10.1093/qjmed/hcl025.

[8] D. Lee and K. Kim. Recurrent neural network-based hourly prediction of photovoltaic power output using meteorological information. *Energies*, 12(2), 2019. ISSN 1996-1073. doi: 10.3390/en12020215.

[9] Y. Li *et al.* A short-term photovoltaic power generation forecast method based on lstm. *Mathematical Problems in Engineering*, 2021:6613123, Jan 2021. ISSN 1024-123X. doi: 10.1155/2021/6613123.

[10] G. Mingming *et al.* Day-ahead power forecasting in a large-scale photovoltaic plant based on weather classification using lstm. *Energy*, 187: 115838, 2019. ISSN 0360-5442. doi: https://doi.org/10.1016/j.energy.2019.07.168.

[11] F. Pedregosa *et al.* Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[12] L. Pengtao *et al.* A hybrid deep learning model for short-term pv power forecasting. *Applied Energy*, 259:114216, 2020. ISSN 0306-2619. doi: https://doi.org/10.1016/j.apenergy.2019.114216.

[13] P.Parczyk, R. Burduk. Transition into solar elevation angle domain for photovoltaic power generation forecasting. *Progress in Polish Artificial Intelligence Research 5*, In press.

[14] X. Qing and Y. Niu. Hourly day-ahead solar irradiance prediction using weather forecasts by lstm. *Energy*, 148:461–468, 2018. ISSN 0360-5442. doi: https://doi.org/10.1016/j.energy.2018.01.177.

[15] C. Tao *et al.* Short-term forecasting of photovoltaic power generation based on feature selection and bias compensation–lstm network. *Energies*, 14(11), 2021. ISSN 1996-1073. doi: 10.3390/en14113086.

[16] J. Yoonhwa *et al.* Long short-term memory recurrent neural network for modeling temporal patterns in long-term power forecasting for solar pv facilities: Case study of south korea. *Journal of Cleaner Production*, 250:119476, 2020. ISSN 0959-6526. doi: https://doi.org/10.1016/j.jclepro.2019.119476.