Detection of Electric Vehicles and Photovoltaic Systems in Smart Meter Data

Martin Neubert, Oliver Gnepper*, Oliver Mey and André Schneider

Fraunhofer IIS/EAS, Fraunhofer Institute for Integrated Circuits, Division Engineering of Adaptive Systems, 01187 Dresden, Germany; martin.neubert@eas.iis.fraunhofer.de (M.N.); oliver.mey@eas.iis.fraunhofer.de (O.M.); andre.schneider@eas.iis.fraunhofer.de (A.S.)

* Correspondence: oliver.gnepper@eas.iis.fraunhofer.de

Abstract: In the course of the switch to renewable energy sources, there is a shift from a few large energy sources (power plants) to a large number of small, distributed energy sources (e.g., photovoltaic systems) and energy storage devices (e.g., electric vehicles). This results in the need to know and identify these energy sources and sinks as soon as new devices are installed, in order to ensure grid stability. This paper presents an approach to identify energy sources and energy storage in smart meter data, using photovoltaic systems and electric vehicles as examples. For this purpose, the Pecan Street dataset is used, which has been extended by charging processes from the ACN dataset. The presented approach comprises a combination of a Convolutional Neural Network and a Multilayer Perceptron, which decides separately, on the basis of the smart meter data of a household, whether an electric vehicle and a photovoltaic system are present. It is shown that the combination of both classifiers achieves accuracy of 90.50% in the case of electric vehicle detection and 96.37% in the case of photovoltaic systems. It is also shown that the power levels lower than 0 kW in the case of the photovoltaic system and higher than 5 kW in the case of the electric vehicles have the largest influence on the output of the Multilayer Perceptron branch, which uses the power balance distribution as input.

Keywords: classification; machine learning; smart meter data; data fusion; electric vehicle; photovoltaic system

1. Introduction

In its latest climate report [1] published in 2022, the Intergovernmental Panel on Climate Change (IPCC) once again drew attention to the fact that a worldwide shift from fossil fuels to renewable energy sources is essential in order to limit global warming to 1.5 degrees. In contrast to fossil energy sources, renewable energy sources such as wind energy or photovoltaics cannot flexibly adjust electricity production. Renewable energies are basically dependent on their associated energy source. Meteorological weather conditions such as cloud cover, wind speed or sunshine duration lead to a generation profile that is not constant but characterized by individual generation peaks.

The consequence of unsteady power generation is that grid stabilization is at risk. Electric vehicles can be part of the solution to this problem, as the study “Distribution grid expansion for the energy transition-electromobility in focus” shows as a key finding [2]. These can reduce load peaks by simultaneous charging or avoid self-produced load peaks by delayed charging. For electric vehicles to be a part of the solution to grid stabilization, they must be present across Europe in terms of area. In 2020, 1.325 million electric vehicles were already registered in Europe, representing 11% of total new car registrations. In 2019, the proportion was still 3.5% [3]. In addition to electric vehicles, photovoltaic systems in conjunction with their buffer storage units represent a second factor in grid stabilization. In this case, the household can be powered autonomously via the buffer storage units when load peaks occur and not place an additional load on the distribution network. Photovoltaic systems without buffer storage, on the other hand, feed directly into the...
distribution network. On the one hand, they represent a variant of decentralized energy sources. On the other hand, there is the risk that short-term feed-in peaks cannot be absorbed by the distribution network due to a lack of consumers at the time in question. Depending on the size of the oversupply, the distribution network infrastructure may be overloaded for a short time and, in the worst case, the power supply in the affected area may be interrupted. At this point, distribution network operators can counteract the problem by taking short-term measures to protect their infrastructure or adapt the infrastructure to the existing photovoltaic systems in the affected section [4].

With the information on whether an electric vehicle or a photovoltaic system is available, a flexible consumption control is possible at high peak loads. The charging processes of electric vehicles using networked wallboxes could be switched in a way that dynamically adjusts the charging times. The customer would only have to inform the electricity provider, by which time the electric vehicle should be fully charged. Up to this point, it would even be possible to discharge the electric vehicle in the meantime to stabilize the distribution network. In the case of an obligation to report electric vehicles or photovoltaic systems, it would be interesting for the distribution network operator to have control authority in this respect. Independently, the knowledge about installed photovoltaic systems as well as existing electric vehicles opens up the possibility for the electricity provider to offer more flexible tariffs and to be able to use the existing infrastructure of the customer more flexibly for electricity trading. In Germany, wallboxes for charging electric vehicles (§ 19 para. 2 NAV Niederspannungsanschlussverordnung of 1 November 2006 (BGBl. I S. 2477) which was last amended by Article 35 of the law of 23 June 2021 (BGBl. I S. 1858)) and photovoltaic systems (§ 74a para. 1 EEG Erneuerbare-Energien-Gesetz of 21 July 2014 (BGBl. I S. 1066), which was last amended by Article 11 of the law of 16 July 2021 (BGBl. I S. 3026)) are subject to registration. However, this obligation only applies to the network operator and not to the electricity provider. The latter has an increasing amount of data from installed smart meters, on the basis of which a more precise assessment of the feed-in profile of photovoltaic systems and the consumption profile of electric vehicles is possible. Furthermore, these data provide information on whether the reported data match the existing devices or if new devices are installed. Therefore, both the distribution network operator and the electricity provider have an interest in identifying photovoltaic systems and electric cars in smart meter data.

Several papers have already addressed the problem of detecting the charging of electric vehicles and detecting photovoltaic systems from smart meter power consumption data. In this regard, it can be seen that the majority of the papers use the Pecan Street dataset as the data basis for their methodologies [5–8]. Other papers use datasets that are not available for public use [9–11]. This makes it difficult to reproduce the results of the elaborations. In addition to the datasets used, it can be seen that publications on electric vehicle charging detection use a minute sampling rate [5,6,8] or an hourly sampling rate [7,9–11] of smart meters as the basis for their methodologies. In addition to an hourly sampling rate [12], a half-hourly resolution [13,14] is also used as a database for the detection of photovoltaic systems. Independent of the object to be detected, supervised learning methods such as Random Forest [9], CNN [7], RNN [7] or kNN [10,12] are used in the papers. CHAID [10] and CART [10] are used as classification methods to create decision trees. Furthermore, cluster analyses are performed using the GMM [11], k-means [12], agglomerative clustering [12] and SOM [12] algorithms. For data preprocessing, ICA [8], PAA [14] and SAX [14] methods are applied for dimension reduction. The papers [7,10] achieve the highest detection rate of 90% in electric vehicle charging detection with a one-hour sampling rate. In comparison, the paper [13] detects 100% of photovoltaic systems, but a half-hourly sampling rate is used as a basis.

This paper also uses the Pecan Street dataset, described in Section 2.1.1, for its approach. This will be extended by the ACN dataset, explained in Section 2.1.2. The actual synthesis of the datasets is described in Section 2.2. The database is first preprocessed, as described in Section 2.3, by transforming it into different input time spans and applying a kernel density estimation to it. A combined neural network consisting of a Convolutional Neural
Network (CNN) and a Multilayer Perceptron (MLP) branch is used as a machine learning method. This is explained in Section 2.4. Moreover, studies have been conducted to analyze the detection accuracy as a function of the time duration of the input consumption data. Furthermore, it was analyzed which power levels have the largest influence on the output of the MLP branch, which uses the power balance distribution as its input. For this purpose, the methodology is described in Section 2.5. The source code used for this work is available at Github [15].

2. Methods

This section outlines the process of developing a classifier to detect electric vehicles and photovoltaic systems from smart meter power consumption data. First, the datasets used are introduced and the subsequent data preprocessing is described.

In the following part, the used Combined Classifier is introduced. Finally, it is described how the information is identified that is considered by the classifier to be particularly relevant with regard to the classification.

2.1. Dataset

2.1.1. Pecan Street Dataset

The Pecan Street [16] dataset is freely available for university or research use and includes the overall power consumption of each household. Furthermore, the power consumption of each household is disaggregated into the share of each consumer (e.g., electric vehicles) and source (e.g., photovoltaic systems). Table 1 below provides a more detailed breakdown of each part of the Pecan Street dataset.

| Region   | Number of Households | Time Period | Sampling Rate |
|----------|----------------------|-------------|---------------|
| New York | 25 households        | 6 months    | 1 s           |
| California | 23 households    | 12 months   | 60 s          |
| Austin   | 25 households        | 12 months   | 1 s           |

The Pecan Street dataset includes electricity consumption data from a total of 73 households located in New York, California and Austin. For each household, the energy consumption of various consumers was measured over a period of 6 months or 12 months, respectively. Smart meters were installed at the individual consumer devices, such as the microwave, the refrigerator or the dryer, which measured the energy consumption at intervals of 1 s or 1 min, respectively.

In the following, the time series of customer number, timestamp, electric vehicle charging power consumption, solar power generation and total household energy balance are extracted from the Pecan Street dataset. As shown in Table 1, the time resolution of the parts of the datasets differs. As described in BSI TR-03109-1 [17], the sampling rate is not smaller than 15 min, but can be chosen freely. At this point, it is assumed that the sampling rate is set to 1 h. This assumption is a compromise between the customer’s interest in protecting their data and not disclosing their behavior and the interest of the electricity provider, which seeks the most accurate information possible about the customer’s consumption behavior in order to optimally adapt the products offered. Furthermore, it is easier to transfer the presented approach to other locations even if the smart meter data at these locations align with the 15 min or 30 min sampling rate, which is common in the EPEX SPOT intraday trading. Considering the 60 min trading packages in Switzerland, which is the largest span of time in the EPEX SPOT intraday trading, the presented approach is applicable for all countries that are part of this market [18].

The following Figure 1 shows an example of the total energy balance of household 27 in New York over 72 h. The charging process of the electric vehicle and the power generation of the photovoltaic system are shown proportionally. It can be seen that the
detection of charging processes of electric vehicles is more difficult if the photovoltaic system is producing electricity at the same time.

![Figure 1. Comparison of the total energy balance with the charging curve of the electric vehicle and the electricity produced by the photovoltaic system for household number 27 over 72 h.](image)

In Germany, according to § 74a para. 1 EEG Erneuerbare-Energien-Gesetz of 21 July 2014 (BGBl. I S. 1066), which was last amended by Article 11 of the law of 16 July 2021 (BGBl. I S. 3026), there is an obligation to install a bidirectional smart meter if the peak power of the photovoltaic system is greater than 7 kW. The installed bidirectional smart meter ensures that the electricity generation and the electricity consumption of the household are counted separately. If the photovoltaic system has a peak power of less than 7 kW, an aggregated total energy balance is formed. Seen across Europe, the regulations for recording the generated or consumed electricity vary. This leads to the challenge that the chosen approach must also guarantee reliable results if there is no separate recording of the power fed into the grid by the photovoltaic system. In order to simulate this situation, the total power consumption of the respective household is analyzed in the following without excluding the share of an existing photovoltaic system from the time series.

The Pecan Street dataset, with its energy consumption data, provides a good basis for solving the classification problem. Analysis of the partial datasets from Table 1 shows that the California households do not include electric vehicle charging. In order to use the partial dataset from California, it is necessary to search for a dataset that contains charging cycles of electric vehicles and add these artificially to the dataset.

### 2.1.2. ACN Dataset

The Adaptive Charging Network (ACN) dataset [19] is freely available for university and research use and includes electric vehicle charging curves. The data were collected at three different locations, which are listed in Table 2 below.

| Location  | Description                        | Number of Charging Stations | Sampling Rate |
|-----------|------------------------------------|----------------------------|---------------|
| Caltech   | University located in Pasadena, CA | 54 charging stations        | 4 s           |
| JPL       | Research lab located in La Canada, CA | 50 charging stations        | 4 s           |
| Office 1  | Office building located in Silicon Valley | 8 charging stations        | 4 s           |

The respective electric vehicle charging events can be freely downloaded, sorted by location and date. Each charging process of the ACN dataset has a Session_id, which
uniquely identifies the charging load curve. Figure 2 below shows a representative electric vehicle charging event of all charging events in the ACN dataset.

![Charging curve of an electric vehicle on 7 November 2018 at the Caltech location.](image)

**Figure 2.** Charging curve of an electric vehicle on 7 November 2018 at the Caltech location.

Visually, it can be seen that the charging process of an electric vehicle at a wallbox is approximately constant. Not all charging stations of the ACN dataset charge with the same charging power. To analyze the different charging powers, a histogram of the three locations of the ACN dataset is shown in the following Figure 3.

![Histogram of the charging capacities of all charging processes of the ACN dataset.](image)

**Figure 3.** Histogram of the charging capacities of all charging processes of the ACN dataset.

The histogram shows the frequency of charging power of individual charging processes of electric vehicles. It can be seen that the charging processes can be subdivided according to their charging power. For the three locations, it can be seen that electric vehicles are primarily charged with a charging power of approx. 2 kW, approx. 8 kW, approx. 16 kW, approx. 22 kW or approx. 30 kW. It can also be assumed that there will be charging stations with a charging power of approx. 11 kW. However, these do not represent an essential part of the total charging stations. Since the ACN dataset in this paper is primarily intended to be used for the European market, only charging processes with a charging power of less than 23 kW are used to meet the conditions of normal power recharging units in the EU [20].
To ensure that the sampling rates of the Pecan Street dataset and the ACN dataset are consistent, the loading curves of the ACN dataset are aggregated to an hourly resolution by computing the arithmetic mean. At this point, both datasets are aggregated to the same resolution, but are not combined. Subsequently, a methodology must be developed to establish a link between the explained datasets.

2.2. Data Synthesis

In this paper, the Pecan Street dataset is used as the basis for the Combined Classifier. To ensure that the power consumption data from households in California can also be used for machine learning, charging events from electric vehicles are taken from the ACN dataset and added to them. In order to have a balanced ratio of households with and without electric vehicle charging events, the first 12 out of a total of 23 households are used for the synthetic dataset.

The procedure for generating the synthetic dataset is shown in the structure chart below (Figure 4) and then explained step by step.

![Figure 4. Structure chart of the generation of the synthetic dataset.](image)

To generate the synthetic dataset, the first day of the time series and the length of the time series are first determined and stored as variables. This is followed by the day-by-day summation of the charging events to the time series of households in California. For summation, two parameters are randomly generated per day. First, the time of day at which the charging process is to start is determined randomly. Second, a random charging process is selected from the ACN dataset. This charging process is added from the assigned start time of the respective day, hour by hour, to the time series of the total energy consumption of the present household. If the selected charging process from the start time is longer than the present day, the overlapping charging hours are added up to the following day. Overlapping charging processes are deliberately allowed so that the case of a household owning several electric vehicles can also be simulated. If the charging event is successfully added to the respective day, the next day of the time series is selected. The loop is executed until an electric vehicle charging event has been added to each day of the selected time series of a household from California.

2.3. Data Preprocessing

The goal is to obtain a steady estimate of the distribution of energy consumption values. For this purpose, kernel density estimation is applied to the formed datasets. The KernelDensity method from the scikit-learn package \[21\] is used for this. As parameters, the kernel = ‘gaussian’ and the bandwidth = 0.75 are chosen. The bandwidth is assigned a value of 0.75, which is a compromise between an overly smoothed and an overly jagged curve of the estimated distribution. The following Figure 5 shows two exemplary kernel density estimates of the synthetic dataset from California. Both kernel density estimates represent the same household, to illustrate the impact that an artificially added electric vehicle has on the distribution.
Figure 5. Comparison of kernel density estimates of the synthetic dataset from California with and without EV charging.

The two kernel density estimates are based on a total sample size of 744 hourly energy consumption data over a period from 1 January 2015 to 31 January 2015. In Figure 5, the individual samples per distribution are color-coded as plus signs. The blue kernel density estimate is based on the California synthetic dataset with loading events, and the kernel density estimate shown in orange is based on those without loading events.

If the two kernel density estimates are compared with each other, it can be seen which distribution contains charging processes. The orange distribution with no charging events has a single peak, while the rest of the distribution has a constant density of 0. In comparison, the blue kernel density estimate has other small peaks throughout the distribution, in addition to the initial peak with a density of approx. 0.65.

As a result of this comparison, it can be seen that even a visual inspection of a kernel density estimate based on energy consumption data can give an indication of whether the charging processes of electric vehicles are included in the data.

The time series datasets and the kernel density datasets based on them, of different input time spans, are then divided into training, testing and validation datasets. Each dataset is split with a ratio of 70% training data, 10% test data and 20% validation data.

2.4. Classification

As proposed in [22], we combine a classifier from a CNN and an MLP into an entire neural network and use it as a Combined Classifier. This allows us to combine the described domain knowledge-based feature extraction method with the automated feature extraction capabilities of a CNN. By using both feature spaces, the Combined Classifier is capable of combining the information of both domains and thus has the potential to achieve higher accuracy compared to the sole use of one of the branches [23]. The structural design of the Combined Classifier is shown in the following Figure 6. The values in the brackets of each layer correspond to the respective dimensionality of the layer.

The TensorFlow package [24] is used to create the Combined Classifier. In the input layer in Figure 6, the CNN branch is given the hourly time series data of the monthly input periods, which correspond to a total of 720 features. It is assumed that a month has 30 days. In the case of the weekly input time span, 168 features, and in the case of the daily input time span, 24 features are passed to the input layer. Building on the input layer of the CNN, one-dimensional convolutional (Conv1D) layers and one-dimensional maximum pooling (MaxPooling1D) layers take place alternately. The dimensionality of the Conv1D layers and the MaxPooling1D layers is derived from the dimensionality of the input layer. The Conv1D layer performs kernel-based convolutional operations on the input layer and
the MaxPooling1D layer aggregates the information. In this case, the maximum is used as aggregation. The Conv1D layer always has an output shape that is decimated by one feature. The reason for this is that the filter overhangs the edge in the first feature due to its size. Thus, the first feature is skipped and the second is started. The MaxPooling1D layer halves the respective input shapes by the parameters kernel_size = 2 and strides = 2. The Flatten layer transforms the output matrix into a vector. Here, the 64 kernels of the fourth Conv1D layer with a respective dimension of 88 are lined up to form a vector of size 5632. The Convolutional Neural Network is completed by a Dense layer, which is given 200 neurons, so that it has the same dimensionality as the output info of the MLP.

Figure 6. Structure of the Combined Classifier supplemented by the output shapes of the layers based on the time series data/kernel density estimation with the monthly input time span.

The MLP branch is given the kernel density estimate of the hourly time series data of the monthly input time period in the input layer, which corresponds to a total of 100 features. This layer is followed by a total number of three Dense layers with 200 neurons each.

The two branches of the CNN and the MLP are combined in the Concatenate layer with 400 neurons. This is followed by a Dense layer with 200 neurons, which outputs the classification result. The activation functions used are the ReLU function for the hidden layer and the sigmoid function for the output layer.

The implemented Combined Classifier is subjected to a two-stage training phase. In the first phase, the two preclassifiers, MLP and CNN, are trained with an extra Dense layer with one neuron separately from each other and outside the Combined Classifier with the training dataset. After the training phase of the preclassifiers is completed, the formed classification models are passed to the Combined Classifier from the respective input to the activation function of the last dense layer with 200 neurons. The two classification models are then assigned the attribute trainable = False, which specifies that the passed models do not go through a training phase within the Combined Classifier and that the passed weight vectors remain constant. The two classification models are combined at the Concatenate layer. On this basis, a combined model is created in the form of an MLP with a Dense layer and an output layer. The full Combined Classifier is then trained with the test dataset. During the training phase, based on the unknown test data, the combined model learns which preclassifier provides the more accurate classification results for which input samples. For the combined model to learn this distinction, the weight vectors of the preclassifiers must remain constant. Finally, the validation dataset is passed to the Combined Classifier, and it is checked to what extent the model is generally valid.
2.5. Feature Importance

By analyzing the feature importances, the influence of the features on the classification is measured. This allows us to verify to what extent the MLP is based on plausible features. To determine the power consumption range that is focused by the MLP branch, the Permutation Feature Importance [25] is used. The latter is calculated by performing the following steps:

1. Evaluation of the accuracy of the trained model on a held-out dataset;
2. Random shuffling of the column values of the feature to be examined;
3. Reclassification of the dataset, which includes the mixed feature column, based on the trained machine learning procedure;
4. Determination of the difference in classification accuracy from step 1 and step 3;
5. Repetition of this process for each individual feature column of the original dataset.

Shuffling of the feature column is performed to check to what extent the classification accuracy deviates compared to the original dataset. If the feature column has only a small impact on the classification, the classification accuracy deviates only minimally. On the other hand, if the difference is high, it can be assumed that the machine learning procedure assigns a high influence to the feature.

For the calculation of the feature importances of the MLP, the method permutation_importance from the scikit-learn package is used. Here, the additional parameter n_repeats is set, which specifies the number of mixing operations of the respective feature column. In order to achieve good mixing of the feature column, the parameter is set to a value of 30. For reproducibility of the results, the parameter random_state is also set to 0, so that the results are comparable for multiple executions. The method permutation_importance returns as a list the feature importances of the selected machine learning method.

3. Results

In the following section, the results of the Combined Classifier in the field of electric vehicle charging detection and photovoltaic system detection are shown. The classification accuracies of the Combined Classifier are compared on the basis of different input time spans of the synthetic dataset. In addition to classification accuracies, the feature importances are compared to evaluate which input features the machine learning process relies on during classification.

3.1. Detecting Charging Processes of Electric Vehicles

The detection of the charging processes of electric vehicles from smart meter power consumption data is carried out using the Combined Classifier. In this case, the branch of the MLP is trained and validated with the kernel density data and the branch of the CNN is trained and validated with the time series data. The reason for the allocation of the time series and the kernel density data based on them to the branches of the Combined Classifier is that the MLP tries to find features that are essential for the present classification. In the case of time series data, it is possible that a loading event is in the first time stamp for one input example and in the last time stamp for another input example. This makes it difficult for the MLP to determine to what extent the feature is an argument for a particular class. In comparison, each input sample of the kernel density data consists of a probability distribution with 100 feature values in each case. In the CNN branch, the time series data are more suitable, as compared to the MLP, using the convolutional layers to search for patterns in the input data and extract features from them.

The time series and the kernel density data based on them are passed to the Combined Classifier with different input time spans. In this case, the Combined Classifier is run 15 times and the resulting validation results are averaged. The validation results are shown in the following Table 3.
Table 3. Comparison of the classification accuracies of electric vehicle charging of the Combined Classifier for different input periods.

| Model              | Dataset | Monthly | Weekly | Daily |
|--------------------|---------|---------|--------|-------|
| Combined Classifier| TS+KDE  | 90.50%  | 86.71% | 73.35%|

This shows that the classification accuracy decreases with a smaller input time span. The Combined Classifier achieves the highest classification accuracy of 90.50% with a standard deviation of 1.75% based on the monthly input time span. The classification model with a classification accuracy of 90.50% is shown again in the following confusion matrix in Figure 7.

![Confusion Matrix](image)

Figure 7. Confusion matrix of the Combined Classifier with the highest classification accuracy in electric vehicle charging detection.

There is a balance in the detection of households with and without electric vehicle charging in this case. The Combined Classifier detects 91.67% of the time series of households that charge an electric vehicle and also 91.67% of households that do not charge an electric vehicle.

The following Figure 8 shows the feature importance diagram of the MLP branch of the Combined Classifier, with the highest detection rate based on the kernel density data with the monthly input time span. The feature importances of a respective classification method clarify which features have a high impact on the model and thus on the values to be classified. The goal of a feature importance diagram is to increase the interpretability of the model.

![Feature Importance Diagram](image)

Figure 8. Feature importance of the MLP branch of the Combined Classifier based on the kernel density data of the monthly input time span for detecting electric vehicle charging events.
When analyzing the feature importance using the MLP branch of the Combined Classifier, it can be seen that the classification procedure is largely based on the features of greater than 5 kW, which represent the charging processes of electric vehicles. This is an indication that the Combined Classifier can also satisfactorily classify unknown data. At the same time, there are peaks at around 0 kW and in the negative range in Figure 8.

### 3.2. Detecting Power Generation of a Photovoltaic System

Additionally, the Combined Classifier is applied to the smart meter power consumption data to detect photovoltaic systems. Once again, for the reasons explained in Section 3.1, the MLP branch is given the kernel density data and the CNN branch is given the time series data for the training and validation phase.

The time series and their kernel density data are passed to the Combined Classifier with different input time spans. The Combined Classifier is run 15 times and the resulting validation results are averaged. The validation results are shown in the following Table 4.

Table 4. Comparison of the classification accuracies of the power generation of photovoltaic systems of the Combined Classifier for different input spans.

| Model             | Dataset   | Monthly   | Weekly   | Daily   |
|-------------------|-----------|-----------|----------|---------|
| Combined Classifier| TS + KDE  | 95.25%    | 96.37%   | 91.72%  |

Compared to the detection of electric vehicle charging, an increase in validation results can be seen in the detection of photovoltaic systems based on the synthetic dataset, independent of the input span. However, the validation results in the case of the detection of photovoltaic systems do not become smaller with a decrease in input span. Instead, a similar validation rate can be seen for the monthly input period at 95.25% ± 0.98% and for the weekly input span at 96.37% ± 0.44%.

For a detailed analysis of the reasons for this, the following Figure 9 compares the confusion matrix of the Combined Classifier with the monthly input time span and the weekly input time span.

Figure 9. Comparison of the Combined Classifier confusion matrix with the monthly input time period and the weekly input time period for detecting the power generation of a photovoltaic system.

The comparison of the confusion matrix shows that the Combined Classifier correctly classifies all input samples from households without photovoltaic systems, independent of the input time span. Mistakes in the classification of households with photovoltaic systems occur with the monthly input time period as well as the weekly input time period. In the monthly input span, 9.72% of the households with a photovoltaic system are incorrectly classified, and in the weekly input period, this is true for a total of 7.01%. Here, it is shown
that a shorter input time span with a simultaneous increase in the amount of samples leads to the slightly increased classification accuracy of the model.

The following Figure 10 shows the feature importances of the MLP branch of the Combined Classifier based on the kernel density data with the monthly input time span for the detection of photovoltaic systems. The Combined Classifier with the monthly input time span is used here, although the weekly input time span has slightly increased classification accuracy. Since the classification accuracies of the monthly input time span are not significantly smaller than those of the weekly input time span, the feature importances of the monthly input time span are analyzed. This allows a direct comparison to the feature importances from Figure 8 for the detection of the charging processes of electric vehicles.

In the feature importances of the Combined Classifier for photovoltaic system detection, it can be seen that the classification model is mainly based on the features at approximately $-2$ kW and approximately 0 kW. The peak at approximately $-2$ kW corresponds to household solar installations and thus represents a plausible feature. The second peak is at approximately 1 kW, which the Combined Classifier has already referred to for the detection of the charging processes of electric vehicles, as shown in Figure 8.

4. Discussion

Currently, only a small amount of smart meter power consumption data exist in terms of availability and scope, which include consumption data from electric vehicle charging and photovoltaic systems. Since data with electric vehicles are not available for all regions in the Pecan Street dataset, a synthetic dataset is generated. For this purpose, the Pecan Street dataset is used as a basis and extended by the charging processes of electric vehicles from the ACN dataset. The charging processes of the electric vehicles added to the Pecan Street dataset correspond to European conditions in terms of their charging speeds. If the Combined Classifier is applied to energy consumption data from non-European countries, the classification accuracies could decrease due to deviating country-specific charging rates. In order to be able to individualize the dataset to country-specific characteristics, the synthesized charging consumption curves would have to be scaled according to the country-specific conditions in a study following this paper. This would allow the quick adaptation of the synthetic dataset to different countries or changing legislation.

When comparing the classification accuracies, especially in identifying electric vehicle charging from electricity consumption data, it can be seen that the classification accuracy becomes lower with a smaller input time span. The reason for the reduction in classification
accuracy with shorter input time spans in the area of electric vehicle charging detection is that machine learning methods have more information available for analysis with a longer input time span than with a short input time span. The greater the input time span, the easier it is for the Combined Classifier to recognize features in it. With smaller input time spans, these charging processes could be cut off, non-existent or balanced out by the power generation of the photovoltaic system in the overall energy balance. When comparing the classification accuracies for the detection of photovoltaic systems from power consumption data in Table 4, a similar picture can be seen. The classification accuracies for the monthly and weekly input periods are comparable. The classification accuracies for the daily input period are significantly lower. The slight increase in classification accuracy for the weekly input period is due to the increased amount of samples. A misclassified month is more significant than a misclassified week because of the smaller sample amount. One potential optimization for the model would be to measure the solar radiation for the area being classified. In the event of high solar radiation, it is ensured that the photovoltaic systems of the houses to be classified produce electricity during the measurement period. During this period, the Combined Classifier can be applied to the smart meter data to investigate whether the households have a photovoltaic system.

Analyzing the feature importances of the MLP branch of the Combined Classifier for detecting electric vehicle charging from power consumption data, it is found that the model also relies on negative energy values. These correspond to household photovoltaic systems. One possible explanation is that the Combined Classifier assumes that an owner of a photovoltaic system also charges an electric car at their home. The feature importances of the MLP branch of the Combined Classifier for photovoltaic system detection show that the model focuses on features in the $-2$ kW and 1 kW range. The range at $-2$ kW corresponds to photovoltaic systems and thus represents a plausible feature. The MLP branch of the Combined Classifier for the detection of electric vehicle charging as well as for the detection of photovoltaic systems is based on features in the low positive kW range. This range corresponds to devices with low basic consumption and thus does not represent a plausible feature for classification. One approach to solving this problem is to expand the database so that there is a balanced ratio of customers with or without a photovoltaic system and an electric vehicle. This makes it easier for the Combined Classifier to learn both cases.

The highest classification accuracy for electric vehicle charging detection is 90.50%, and that for photovoltaic system detection from smart meter power consumption data is 96.37%. In order to increase the classification accuracies in the future, the hyperparameters of the Combined Classifier would need to be further optimized. One possibility here would be to use a dynamic learning rate. The learning rate when training the Combined Classifier could be reduced from epoch to epoch. Another possibility would be to change the network topology—for example, by adjusting the number of layers and the neurons per layer. A further approach to optimizing the classification accuracy for photovoltaic system detection would be to build an ensemble model that combines the monthly and weekly Combined Classifiers. The two Combined Classifiers with the highest classification accuracies are combined to take advantage of the strengths of each. However, when optimizing the hyperparameters, it must be noted that a classification accuracy of 100% is difficult to achieve due to the presence of perturbations in the dataset. For example, an electric vehicle may be charging from a household socket with low charging power or the photovoltaic system may be overlaid by other large consumers.

5. Conclusions

In this paper, an approach for the detection of the charging cycles of electric vehicles and the presence of a photovoltaic system in smart meter data is presented, which uses a combined neural network consisting of a CNN branch and an MLP branch. The Combined Classifier is trained and evaluated on freely available datasets (Pecan Street dataset and ACN dataset) and thus offers the option to comprehend the shown findings and adapt as well as transfer the used approach. Additionally, a method is presented to generate a
dataset that contains synthetic electric vehicle charging cycles. Due to changing boundary conditions, especially in terms of solar radiation, the data synthesis approach offers the option to extend an existing set of smart meter data for other locations so that it contains the charging cycles of electric vehicles, even though they are not contained in the original dataset. This offers the option to use the Combined Classifier also for datasets from other locations and to evaluate its accuracy even if there are no data currently available for this location that include electric vehicle charging cycles. The introduced Combined Classifier utilizes time series information, which is processed in the CNN branch, as well as load distribution information, which is processed in the MLP branch. Due to the combination of both branches, the Combined Classifier achieves accuracy of 90.50% for the detection of electric vehicles and 96.37% in the case of photovoltaic systems. These results align with the scores that are listed in Section 1. The achieved accuracy decreases when data of a shorter span of time are provided to the Combined Classifier. An exception is the case where the consumption data of one week are used to identify photovoltaic systems in the smart meter data. In this case, the Combined Classifier achieves 1% higher accuracy compared to the case where the data of one month are used. Additionally, the feature importance of the MLP branch is assessed. The calculation of the corresponding feature importances indicates that the classifier thereby focuses on meaningful information and therefore promises to be robust also when applied on real-world data. The conducted investigations further show that the classification accuracy rises with the increasing time duration of the input consumption data.

The presented approach offers the possibility to separately identify whether a household has an electric vehicle and a photovoltaic system. As described in Section 1, the distribution network operator can install flexible consumption control in case of high maximum loads based on this knowledge. Additionally, the electricity provider has the option to offer the corresponding flexible contracts to the customer. It would be conceivable for the owners of electric vehicles to specify to the electricity provider the time by which the vehicle should be fully charged. During this period, the electricity provider could dynamically adjust the charging time so that the electricity load is smoothed over the affected area. In the event of an unexpectedly high load on the power grid, electric vehicles could also feed in electricity to stabilize the distribution network before they are charged, possibly even in combination with the buffer storage of the photovoltaic systems. Furthermore, the approach presented in this paper offers the possibility for the distribution network operator to establish a control instance regarding whether the homeowners comply with their possible legal obligation to report electric vehicles and photovoltaic systems.

**Author Contributions:** Conceptualization, O.G., O.M. and A.S.; methodology, O.G. and O.M.; software, M.N.; validation, M.N.; formal analysis, M.N.; investigation, M.N.; data curation, M.N.; writing—original draft preparation, M.N. and O.G.; writing—review and editing, O.G., A.S.; visualization, M.N.; supervision, O.G. and O.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the German Federal Ministry of Education and Research, grant number 01QE1948C.

**Data Availability Statement:** Publicly available datasets were analyzed in this study. The following datasets were used: [Pecan Street Dataset] Pecan Street Inc. Dataport; https://www.pecanstreet.org/; (accessed on 4 April 2021). [ACN Dataset] Lee, Z.J.; Li, T.; Low, S.H. 2019. ACN-Data: Analysis and Applications of an Open EV Charging Dataset; https://ev.caltech.edu/dataset, accessed on 4 April 2021.

**Conflicts of Interest:** The authors declare no conflict of interest.
References

1. Pörtner, D.C.; Roberts, E.S.; Poloczanska, K.; Mintenbeck, M.; Tignor, A.; Alegria, M.; Craig, S.; Langsdorf, S.; Löschke, V.; Möller, A. IPCC, 2022: Summary for Policymakers. In Climate Change 2022: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change; Cambridge University Press: Cambridge, UK, 2022; in press.

2. Agora Verkehrswende, Agora Energiewende, Regulatory Assistance Project (RAP) (2019): Verteilnetzausbau für die Energie-wende—Elektronomobilität im Fokus. Available online: https://static.agora-energiewende.de/fileadmin/Projekte/2018/Netzausbau_Elektromobilitaet/Agora-Verkehrswende-Agora-Energiewende_EV-Grid_WEB.pdf (accessed on 13 May 2022).

3. European Environment Agency. New Registrations of Electric Vehicles in Europe; European Environment Agency: Copenhagen, Denmark, 2021. Available online: https://www.eea.europa.eu/ims/new-registrations-of-electric-vehicles#:~:text=The%20uptake%20of%20electric%20vehicles,registrations%20in%20just%201%20year (accessed on 18 November 2021).

4. Wirth, H. Aktuelle Fakten zur Photovoltaik in Deutschland; Fraunhofer ISE: Breisgau, Germany, 2022. Available online: http://www.pv-fakten.de/ (accessed on 1 May 2022).

5. Zhang, Z.; Son, J.H.; Li, Y.; Trayer, M.; Pi, Z.; Hwang, D.Y.; Moon, J.K. Training-free non-intrusive load monitoring of electric vehicle charging with low sampling rate. In Proceedings of the IECON 2014—40th Annual Conference of the IEEE Industrial Electronics Society, Dallas, TX, USA, 29 October 29–1 November 2014; pp. 5419–5425. [CrossRef]

6. Shaw, A.; Nayak, B.P. Electric vehicle charging load filtering by power signature analysis. In Proceedings of the 2017 IEEE International Conference on Data Management, Analytics and Innovation (ICDMAI), Pune, India, 24–26 February 2017; pp. 71–75. [CrossRef]

7. Hoffmann, V.; Fesche, B.I.; Ingebrigtsen, K.; Christie, I.N.; Punnerud Engelstad, M. Automated detection of electric vehicles in hourly smart meter data. In Proceedings of the 25th International Conference on Electricity Distribution (CIRED 2019), Madrid, Spain, 3–6 June 2019. [CrossRef]

8. Munshi, A.A.; Mohamed, Y.A.R.I. Unsupervised Nonintrusive Extraction of Electrical Vehicle Charging Load Patterns. IEEE Trans. Ind. Inform. 2019, 15, 266–279. [CrossRef]

9. Verma, A.; Asadi, A.; Yang, K.; Tyagi, S. A data-driven approach to identify households with plug-in electrical vehicles (PEVs). Appl. Energy 2015, 160, 71–79. [CrossRef]

10. Verma, A.; Asadi, A.; Yang, K.; Maitra, A.; Asgeirsson, H. Analyzing household charging patterns of Plug-in electric vehicles (PEVs): A data mining approach. Comput. Ind. Eng. 2019, 128, 964–973. [CrossRef]

11. Barkost, P.H. Detecting EV Charging From Hourly Smart Meter Data. Master’s Thesis, UiT Norges Arktiske Universitet, Tromsø, Norway, 2020.

12. Donaldson, D.L.; Jayaweera, D. Effective solar prosumer identification using net smart meter data. Int. J. Electr. Power Energy Syst. 2020, 118, 105823. [CrossRef]

13. Brown, J.; Abate, A.; Rogers, A. Disaggregation of household solar energy generation using censored smart meter data. Energy Build. 2021, 231, 110617. doi: 10.1016/j.enbuild.2020.110617. [CrossRef]

14. Ling, W.; Yu, X.; Wang, J.; Sokolowski, P. A Motif-based Classification Algorithm for Identifying Solar Panel Installations. In Proceedings of the 2020 IEEE International Conference on Industrial Technology (ICIT), Buenos Aires, Argentina, 26–28 February 2020; pp. 595–600. [CrossRef]

15. Supplementary Information: Source Code Documentation of This Paper at Github; Available online: https://github.com/m-neubert/mdpi-2022-paper (accessed on 1 June 2022).

16. Pecan Street Inc. Dataport. Available online: https://www.pecanstreet.org/ (accessed on 4 April 2021).

17. Anforderungen an die Interoperabilität der Kommunikationseinheit eines Intelligenten Messsystems. Available online: https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Publikationen/TechnischeRichtlinien/TR03109/TR03109-1.pdf?__blob=publicationFile&v=1 (accessed on 17 September 2021).

18. EPEX SPOT SE Trading at EPEX SPOT, Paris. 2022. Available online: https://www.epexspot.com/en/downloads#market-data (accessed on 20 June 2022)

19. Lee, Z.; Li, T.; Low, S.H. ACN-Data: Analysis and Applications of an Open EV Charging Dataset. In Proceedings of the Tenth International Conference on Future Energy Systems, e-Energy ’19, Phoenix, AZ, USA, 25–28 June 2019.

20. Directive 2014/94/EU of The European Parliament And of The Council of 22 October 2014 on the Deployment of Alternative Fuels Infrastructure. Available online: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32014L0094 (accessed on 30 May 2022).

21. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. 2011, 12, 2825–2830.

22. Mey, O.; Schneider, A.; Enge-Rosenblatt, O.; Mayer, D.; Schmidt, C.; Klein, S.; Herrmann, H.G. Condition Monitoring of Drive Trains by Data Fusion of Acoustic Emission and Vibration Sensors. Processes 2021, 9, 1108. [CrossRef]

23. Neubert, M. Erkennung von Ladevorgängen für elektrisch betriebene Fahrzeuge mit Hilfe Deep-Learning-basierter Analysen von Smart-Meter-Stromverbrauchsdaten. Master’s Thesis, Fraunhofer IIS, Dresden, Germany, 2022.

24. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.S.; Davis, A.; Dean, J.; Devin, M.; et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. 2015. Available online: tensorflow.org (accessed on 1 May 2021).

25. Breiman, L. Random forests. Mach. Learn. 2001, 45, 5–32. [CrossRef]