energies

Article

A Machine Learning Application for the Energy Flexibility Assessment of a Distribution Network for Consumers

Jaka Rober [1], Leon Maruša [2] and Miloš Beković [1,*]

[1] Faculty of Electrical Engineering and Computer Science, University of Maribor, 2000 Maribor, Slovenia; jaka.rober@student.um.si
[2] Elektro Celje d.d, Vrunčeva ulica 2a, 3000 Celje, Slovenia; leon.marusa@elektro-celje.si
* Correspondence: milos.bekovic@um.si; Tel.: +386-2-220-7172

Abstract: This paper presents a step-by-step approach to assess the energy flexibility potential of residential consumers to manage congestion in the distribution network. A case study is presented where a selected transformer station exhibits signs of overloading. An analysis has been performed to evaluate the magnitude of the overloading and the timing of the overload occurrence based on their historical load data. Based on the historical load data, the four most prominent consumers have been chosen for the flexibility assessment. Temperature load dependency has been evaluated for the selected consumers. The paper's novel approach focuses on selecting individual consumers with the highest energy flexibility potential, and analysing their load patterns to address transformer overloading. To achieve this, machine learning algorithms, specifically, multiple linear regression and support vector machines, were used for load profile forecasting during the overload occurrences. Based on the forecast and measured load patterns, flexibility scenarios were created for each consumer. The generated models were evaluated and compared with the forecasting based on the average load of the past days. In the results, three potential consumers were identified who could resolve the transformer overloading problem. The machine learning models outperformed the average-based forecasting method, providing more realistic estimates of flexibility potential. The proposed approach can be applied to other overloaded transformer stations, but with a limited number of consumers.

Keywords: flexibility; baseline; demand response; distribution transformer; congestion management; power flow control; peak shaving; load shifting; predictive models; machine learning

Citation: Rober, J.; Maruša, L.; Beković, M. A Machine Learning Application for the Energy Flexibility Assessment of a Distribution Network for Consumers. *Energies* **2023**, *16*, 6168. https://doi.org/10.3390/en16176168

Academic Editor: Abu-Siada Ahmed

Received: 1 August 2023
Revised: 14 August 2023
Accepted: 21 August 2023
Published: 24 August 2023

1. Introduction

The electric power system has undergone drastic changes in the last two decades. The distribution network (DN) has, specifically, been put under huge constraints because of the increasing number of distributed energy resources, such as solar and wind power plants [1,2]. This is part of the European Green Deal programme that aims to make Europe carbon neutral and sustainable by 2050 with carbon-neutral electric energy production. The same programme also aims to decarbonise the transportation and housing sector through electrification. In such a way, the introduction of electric vehicles has been proposed as an alternative to gas-powered vehicles. However, integrating such high-intensity loads into the DN causes many problems. As was pointed out in [2], the stochastic load pattern of electric cars and ageing or inadequate infrastructure limits the integration process of electric vehicles. In the housing sector, conventional heating resources such as natural gas and oil are being replaced with heat pumps. As was presented in [3], large communities could be heated by heat pumps, along with energy provided by solar power plants. Heat pumps offer notable advantages, as they are characterised by their environmental friendliness and efficiency. However, it is essential to consider certain limitations associated with these systems. Notably, heat pumps can exhibit a considerable load, and their performance is dependent on external factors such as air temperature and humidity. During colder winter months, when solar power plants generate less energy, heat pumps tend to consume more electricity, potentially leading to energy shortages. This situation may necessitate the implementation of battery storage solutions, particularly in micro-networks, to address the fluctuations in energy demand and supply effectively.

Apart from high energy consumption, the usage of newly penetrating loads is not distributed evenly across the day, which means that a high power demand usually coincides with the peak load time and therefore causes an overload of the infrastructure. One way of solving these issues is to reinforce the DN, which is expensive and, at the same time, presents the risk of its infrastructure not being utilised to its maximum potential. It is also important to acknowledge that grid reinforcement may not be able to keep up with the increasing demand of rising consumption. The energy flexibility of consumers and producers has been introduced to tackle this problem cost-effectively and in a fast manner. Energy flexibility services in DNs have been explored widely, and adjusted to residential buildings and industry [4,5]. In a DN, a demand response service is used to mitigate the overloading of the infrastructure by shifting the load [6] or shaving peaks [7] at times of peak demand hours. Consumers have now been put in an active role, where they adjust their energy consumption or production to provide the necessary ancillary services for the DN operator. Consumers with the ability to adjust production profiles or energy consumption profiles are called prosumers. In reference [8], an energy flexibility review for household consumers is given, along with thorough characterisation and quantification methods. The utilisation of prosumer flexibility services presents significant prospects for enhancing distribution grid planning and infrastructure adequacy. It can lead to a reduced reliance on expensive grid reinforcements, and facilitate the seamless integration of renewable energy sources [9]. Integrating the concept of prosumer flexibility into energy-based maintenance and sustainable predictive maintenance practices can revolutionise grid management and maintenance strategies. By analysing and predicting prosumers' energy consumption patterns, load-shifting capabilities, and distributed energy resources, these maintenance approaches can identify potential grid stress points and predict maintenance needs with greater precision. An overview of the demand response programme implementation has been made in [10] for Europe and the United States. It was concluded that the United States has the advantage, due to the better regulatory and policy environment in some regions of the United States.

To identify possible energy flexibility providers, an assessment must be made of whether the magnitude and pattern of consumption are adequate to engage load demand services. Transformers and cables are essential components that form the foundation of the grid. The temporary overloading of these elements is possible; however, it can lead to a shortened life expectancy, increased losses, and, in the worst-case, outages. In [11], a methodology was described, which correlates how much overloading is needed to justify the willingness of DN operators to conduct energy flexibility services. Another challenge in achieving energy flexibility realisation lies in the willingness of consumers to engage actively in demand response programmes, as discussed in reference [12]. This willingness to participate is contingent upon factors such as household type, appliance usage patterns, and the financial incentives offered for their involvement. Consumers who are willing to participate in demand response programmes sell their services through aggregators, who combine the services of multiple prosumers and sell those on the energy flexibility market [13]. Prosumers are then rewarded financially according to the type of demand response service and the load difference between the forecast load (baseline) and load at the time of the demand response event.

The same methods used for evaluating demanded response programmes can be used for load profile identification [14]. A load profile can also be deducted by segmenting the user's past load profile and appliance activity, as was proposed in [15,16], together with an optimal scheduling of appliances. Without information about appliances, a different method has been proposed in [17], which is based on clustering consumers' hourly load data. In the case study of Pakistan [18], the consumer flexibility potential was calculated for selected household appliances using Monte Carlo simulations. Appliances were segmented by their load patterns, and an economic analysis was conducted for each to provide savings for consumers. In cases where measurements are lacking, reference [19] describes an approach to assess the flexibility provision of consumers by utilising optimisation on a range of nodal power injections for each consumer.

This article proposes a novel method of identifying the energy flexibility potential of residential distribution network consumers to manage congestion at transformer stations, and, at the same time, enable the integration of a higher number of distributed energy resources and newly penetrating loads due to the green transition. Energy flexibility has been determined by using machine learning algorithms to forecast selected consumer load profiles and evaluate patterns of consumption relative to transformer times of loading. We identified temperature-dependent consumers and correlated their consumption with temperature. An energy flexibility assessment can be used by aggregators to approach certain consumers, and motivate them to enter demand response programmes [13]. The same energy flexibility assessments can also be used in planning DNs; hence, it can be determined where the necessity for infrastructure upgrading is needed most.

The article is structured into six sections. As described, Section 1 serves as an introduction, outlining the current issue of incorporating renewable sources and proposing flexibility as a solution. Various methods for predicting consumption and evaluating their ability to determine flexibility are described in Section 2. Section 3 presents a case study, to assess the potential for the flexibility of an overloaded transformer station. In Section 4, a comprehensive analysis of transformer station consumption is conducted to identify overload intervals and select potential consumers for flexibility services based on their historical load data. Section 5 illustrates a comparison between consumption forecasts and actual data on the day of the highest overload for the selected consumers, presenting different flexibility scenarios. Finally, Section 6 concludes the article by summarising the findings and highlighting the novelty of the new approach.

2. Proposed Methodology of Energy Flexibility Assessment

The proposed methodology for assessing energy flexibility potential is based on forecasting load consumption on the day an overload has occurred. A forecast is then evaluated using evaluation metrics. Based on the evaluation of each method and forecast pattern, an assessment of energy flexibility potential is made for each consumer. Forecast methods for small-scale residential buildings are described in [20]. Simple average models were proposed, along with machine learning models such as regression and artificial neural networks. In this article, we propose one model based on averaging and two machine learning models based on regression. Multiple linear regression (MLR) with custom features is proposed, along with support vector regression (SVR) that has been adopted widely as a benchmark model for short-term load forecasting in [21], where multiple regression models were compared and tested in the case of a university campus load forecast.

The employment of probabilistic methods was also studied in [22], where a Gaussian process was used, and [23] where Bayesian approach was utilised for load forecast. Probabilistic methods focus on modelling uncertainty and providing probability distributions, which can be valuable when dealing with uncertain data. Probabilistic models are more robust to noise and can handle both linear and non-linear dependencies. Additionally, they may require fewer training examples to achieve meaningful results. The downside of a probabilistic models is that they are computationally intensive and hard to interpret compared to SVM and MLR, where the focus is on understanding relationships between variables.

The use of artificial neural networks (ANNs) is also one very commonly used method as studied in [24], that offers the capability to capture intricate relationships and patterns within load data, making them effective for handling complex and non-linear load behaviour. They can incorporate multiple influencing factors simultaneously and adapt to changing conditions, improving forecasting accuracy. However, ANNs can be computationally demanding and require careful tuning to prevent overfitting. The complexity of their internal workings can make interpretation challenging, and their performance heavily relies on data quality and the selection of relevant features.

Compared to a classic regression model, polynomial regression can capture non-linear relationships between input features and load, making it suitable for situations where the load behaviour deviates from linearity. Additionally, polynomial regression models are relatively easy to implement and understand as authors described in [25]. But they may struggle to capture highly complex and intricate load patterns compared to more sophisticated methods like ANNs. It can also be sensitive to outliers and noise in the data. As the degree of the polynomial increases, overfitting becomes a concern, and selecting the appropriate degree requires careful consideration to avoid model complexity or underfitting.

Clustering is also one commonly used method that can uncover hidden patterns and group similar load behaviours, providing insights into different load profiles, as was studied in [26], where each consumer load profile was estimated. It is useful for segmenting data into distinct clusters, which can aid in developing tailored forecasting models for specific load categories. Clustering can also help identify anomalies or unusual load patterns. While it cannot directly provide load predictions, it does offer a data exploration and segmentation tool. Selecting the appropriate number of clusters and deciding on the clustering algorithm can be challenging. Additionally, clustering may not fully capture the underlying dynamics driving load variations, and it may not be suitable for forecasting subtle changes or fine-grained load patterns.

The use of hybrid models has become a novelty in load forecast. In [27], the authors made a hybrid model based on data decomposition considering periodicity, trend, and randomness of the original load time series data. A forecast of the short-term load was made through preprocessing and analysing the original time series and using a genetic algorithm to optimise a generalised regression neural network. The results showed good fitting ability along with the option to approximate the actual values when dealing with non-linear time series data with periodicity, trend, and randomness.

In comparison to other methods, such as XGBoost, AdaBoost, and Random Forest, it is noteworthy that these ensemble algorithms can yield greater efficiency, especially in capturing complex relationships and non-linear patterns in the data [28]. However, MLR and SVR offer distinct advantages in terms of input features. Specifically, MLR and SVR allow for a higher degree of customisation and simplicity in selecting input features, as they focus on linear relationships, and, in the case of SVM, also non-linear. This simplicity can enhance the interpretability and the ease of incorporating domain knowledge. On the other hand, ensemble methods, while offering enhanced predictive capabilities, might require more advanced techniques to fine-tune feature selections. Therefore, while XGBoost, AdaBoost, and Random Forest can provide greater efficiency, it is important to recognise that MLR and SVR excel in offering customisable and interpretable input feature options. The choice depends ultimately on the trade-off between model complexity, predictive power, and the need for feature customisation.

Out of all the regression models, MLR has been chosen because of its simplicity and ease of adding and optimising input parameters, which aligns well with our objective of exploring various input features for an enhanced model performance. MLR efficiency has been validated in [29]. As for SVM, we recognised its significance as a benchmark model in regression tasks [21], particularly when dealing with non-linearities. SVM's ability to capture complex patterns and their robustness to non-linear relationships made it a suitable candidate for comparison with our proposed models. In the scope of this study, the generation of a solar power plant has not been modelled because we focused on the demands of consumers. In [30], an approach is described where the load is predicted as the median of interpolating days, and solar power plant generation is predicted using the Random Forest method.

2.1. Forecast Models

Average of the last Y days: The last Y working days were selected from the last Z calendar days, and then samples of the same time were averaged. While it is a straightforward approach, it may not capture the complex patterns and dynamics present in the load profile data due to weather changes. Two models are needed to forecast each day of the week. For working days, only previous working days are considered (1). As for weekends and national holidays, the same days of the week are considered in calculating the average (2) because of different consumption patterns. The previous Sundays are selected if a national holiday coincides with a working day. The method is described in [20], and the novelty lies in our addition of separate considerations for working and non-working days:

$$\hat{P}_i = \frac{1}{k}\sum_{j=1}^{k} P_{i,j}, \ \ i = 1,2,3\ldots\ n \tag{1}$$

$$\hat{P}_i = \frac{1}{k}\sum_{j=1}^{k} P_{i,j\cdot 7}, \ \ i = 1,2,3\ldots\ n \tag{2}$$

In (1) and (2), $\hat{P}_i$ represents an element of the vector containing the predicted 15 min load samples for the observing day. Elements $P_{i,j}$ form a matrix that contains load samples for the previous k days of observation, where index j represents the day of observation, and index i represents each 15 min sample, resulting in the collection of 96 samples for one day of observation (n = 96). Another commonly utilised averaging technique is HighXofY, which involves selecting the X days with the highest average load from the last Y working days to construct a forecast model. In this study, the described model was not realised, because it was outperformed by the classic average of Y days. Averaging models are relatively easier to explain to consumers in the context of demand response programmes compared to machine learning models, which need additional explanation.

Multiple linear regression (MLR): represents the most basic form of regression models. The dependent (target) variable is formed as a combination of the independent (predictor) variables in a linear manner. The assumption within MLR is that a linear relationship exists between the dependent and independent variables. The general equation for MLR can be represented as (3):

$$y_i = \beta_0 + \beta_1 x_{\mathrm{i}1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + e_i, \ \ i = 1,2,3\ldots\ n \tag{3}$$

The dependent variable is marked with y_i; x_{ip} are independent variables; β_p are the estimated regression coefficients; β_0 is the value of y_i when all independent variables are equal to zero; and e_i is the model's error term, also known as the residual, and accounts for any unexplained variation in the dependent variable that is not captured by the linear relationship with the independent variables. The regression coefficients determine the impact and direction of each independent variable on the dependent variable. The coefficients β_p are estimated based on minimising the sum of squared residuals (SSR) to minimise the difference between the observed values and the predicted values from the model. This is achieved by taking partial derivatives of SSR with respect to each coefficient and setting them to zero. Solving the resulting system of equations yields the optimal values for the coefficients, which provide the best-fitting linear relationship between the dependent and independent variables. The formula for SSR is given below (4):

$$SSR = \sum_{i=1}^{n}\Big(z_i - \Big(\beta_0 + \beta_1 x_{\mathrm{i}1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + e_i\Big)\Big)^2 \tag{4}$$

In (4), first part (z_i) represents the measured value and the second part is the predicted value obtained from the MLR model for the i-th data point. The dependent variable in the case of load forecast is represented by electrical load, and the independent variables are given as input features to the model. We considered the model and features in [29] in addition to our proposed features to build the MLR model (5):

$$\hat{P} = \beta_0 + \beta_1 \cdot Day \cdot Minutes + \beta_2 \cdot Month + \beta_3 \cdot P + \beta_4 \cdot P_{\mathrm{RA}} + \beta_5 + P_{\mathrm{AVG}} + \beta_6 \cdot Holidays + \beta_7 \cdot T \cdot Month + \beta_8 \cdot T \cdot Minutes \quad (5)$$

$\hat{P}$ is the predicted load profile for the observing day. Class variables were considered of the 15 min samples, day of the week, month of the year, and national holidays. Numerical variables such as load samples of the previous three days and load of the same day of the week before are given as P; also, the average load of the previous day P_{AVG} is considered. The proposed model (5) can be assessed as a time series approach, and therefore, the mean rolling average was used as data preparation to smooth out short-term fluctuations and to capture the trend of the load change. A rolling moving average was used for each sample, with a three-hour window over the past three days, and is given as P_{RA}. To account for seasonal variations and temperature discrepancies between day and night, a linear interaction effect was incorporated between the temperature and classes of 15 min samples and months. The trend of increasing consumption was not considered due to the effect of COVID-19 and the inclusion of distributed energy resources, which caused disturbances in model prediction. Furthermore, in [29], the exploration of second- and third-order temperature interactions was undertaken to model temperature-dependent consumers for both lower and higher temperature ranges. However, for the scope of our analysis, it was found that consumers exhibited only lower-temperature dependency, thus rendering any additional interactions unnecessary. MLR requires certain conditions to ensure accurate predictions. In our study, we examined the linearity of temperature and the load of individual consumers thoroughly. We also conducted tests to check if the errors followed a normal pattern. To tackle multicollinearity, we took a proactive approach by removing one feature from the model, improving the reliability of our results. While we could not explore all conditions due to study limitations, these steps strengthened the reliability of our findings significantly.

Support vector regression (SVR): It is an extension of support vector machines (SVMs), which are used primarily for classification. SVR can exhibit both linear and non-linear behaviour, depending on the choice of the kernel function. This flexibility makes SVR a powerful regression technique, as it can handle a wide range of data patterns by using appropriate kernel functions, unlike MLR, which can only model linear interactions between variables. When a linear kernel is used, SVR behaves as a linear regression technique, aiming to find a hyperplane in the feature space to fit the training samples, while minimising errors and complexity. On the other hand, when a non-linear kernel is employed, such as the Radial Basis Function (RBF) or a polynomial kernel, SVR performs non-linear regression. These non-linear kernels enable SVR to map the input data into a higher-dimensional feature space, allowing for more complex relationships between the variables. A general formula for SVR can be written as function (6):

$$f(x) = \boldsymbol{w} \cdot \boldsymbol{\phi}(x_i) + b \quad (6)$$

The vector $\boldsymbol{w}$ contains the coefficients of weights that are minimised to fit the hyperplane to the data. Vector $\boldsymbol{\phi}$ is defined as a transformation function that transforms input variables x_i into a higher-dimensional space. The transformation is dependent on kernel type. When a b bias term is indicated, this represents the intercept of the hyperplane in the higher-dimensional feature space. We minimised function (7), intending to find a function of a hyperplane that minimises the errors of deviation (ζ) from the predefined margin of tolerance ε. For a general linear SVR model, certain conditions must be met (8), so that the SVR model can be built:

$$\min \frac{1}{2} \|\boldsymbol{w}\|^2 + C \sum_{i=1}^{n} (\zeta_i + \zeta_i^*) \quad (7)$$

$$\begin{gathered} y_i - \boldsymbol{w}\cdot\boldsymbol{\phi}(x_i) - b \leq \varepsilon + \zeta_i \\ \boldsymbol{w}\cdot\boldsymbol{\phi}(x_i) + b - y_i \leq \varepsilon + \zeta_i^* \\ \zeta_i,\ \zeta_i^* \geq 0 \end{gathered} \tag{8}$$

There are two deviations, $\zeta_i + \zeta_i^*$, one for the upper and one for the lower boundary that defines the distance between ε and the training points. Factor C is the regularisation parameter that controls the trade-off between maximising the margin and minimising the errors. Measured points are presented as y_i. Support vectors are defined as data points that lie on or within the margin boundaries, and have the most effect on the hyperplane function. With the first term of (7), we aimed to maximise the margin between the hyperplane and the support vectors. With the second term, we controlled how much penalisation was added to the deviations. An RBF kernel, also known as the Gaussian kernel, was used in our study. It was also used in [16], because it is infinitely divisible, and is therefore smooth. The general restrictions of SVR (8) do not apply in this case, as the RBF kernel defines $\boldsymbol{\phi}(x_i)$ implicitly, and therefore, no explicit transformation is required as in the linear kernel case. This type of kernel can also capture complex patterns and non-linearities that may be difficult to represent in the original input space. The RBF kernel is represented by (9):

$$K(x_i, x_j) = e^{-\gamma\|x_i - x_j\|^2} \tag{9}$$

In general, (9) represents the similarity between data points x_i and x_j. With the term $\|x_i - x_j\|$, we denote the Euclidian distance between those two data points. The hyperparameter γ in the RBF kernel determines the influence of nearby points on grouping. Higher γ values result in a more localised influence, where only data points very close to the reference point have a significant impact. Conversely, lower γ values lead to a broader influence, extending the influence of each training point to a larger region. In reference [16], a thorough comparison of various kernels was conducted, and as a result, no additional explanations will be provided in this context.

The input features of SVR were downsized compared to MLR because of the high computing time of SVR. The numerical features of historical load data stayed the same. Only categorical features were kept of the month and distinction between working and non-working days. There was no interaction of categorical features with temperature. Some features could be dropped because the SVR models non-linear interactions, which was not possible with MLR. In the proposed model, the following hyperparameters were used: $C = 1$, $\varepsilon = 0.3$, and γ was set to scale. The hyperparameters were localised by using the grid search method on a limited range of values due to computation time restrictions.

2.2. Performance Evaluation Metrics

Commonly used evaluation metrics were used to evaluate the forecast models. Each consumer has a unique pattern of consumption, and, for that purpose, different evaluation metrics were used. In the following equations, z_i represents a sample of the measured variable, while $\hat{z}_i$ represents the predicted variable, and n represents the number of observations. The mean absolute error (*MAE*) (10) measures the average magnitude of errors:

$$MAE = \frac{\sum_{i=1}^{n}|z_i - \hat{z}_i|}{n}\ [\mathrm{kW}] \tag{10}$$

The mean absolute percentage error (*MAPE*) (11) defines the relative overall fit, and is a method frequently used in predicting a forecast method in statistics. However, with small values of the measured points z_i, the *MAPE* can become relatively large and inaccurate:

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{z_i - \hat{z}_i}{z_i}\right|\ [\%] \tag{11}$$

Root-mean-squared error (*RMSE*) (12) measures the average magnitude of the residuals. It is similar to *MAE*, but it penalises large deviations; therefore, *RMSE* is more effective in the case of irregular consumption patterns where prediction errors are large:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(z_i - \hat{z}_i)^2}{n}} \text{ [kW]} \quad (12)$$

In the described metric, a smaller value indicates a better performance of the model. Another commonly used metric is R-squared (R^2), which measures how closely the fitted regression line aligns with the observed results. R-squared values range between 0 and 1, and a higher value indicates a better fit of the model.

3. Case Study Description

For this study, a set of data has been provided by the distribution network operator Elektro Celje for selected transformer stations, covering the period between 1 January 2020 and 1 June 2023. With the help of an advanced metering infrastructure, 15 min readings of a transformer's apparent power were collected, along with active power readings of 19 consumers out of the available 22. The remaining three consumers have old meters that only measure energy consumption at one-hour intervals or daily, and therefore could not be included in the analysis. The rated power of the transformer in the transformer station is 50 kVA. There are seven consumers who have their own source of energy production in solar power plants. Figure 1 illustrates the six key steps of our novel approach to determine consumer flexibility. At point 1, all measurements from consumers and the transformer are collected from the metering database. At point 2, the collected data are processed, cleaned, and prepared for further analysis. In step 3a, an analysis of the transformer station's overload is conducted, based on historical consumption data. Simultaneously, potential consumers for flexibility services are identified using past load information in point 3b. In point 4, consumption forecasts are made for the day of transformer station overload for each selected consumer. In step 5, based on the predicted consumption patterns and critical overload intervals of the transformer, flexibility assessments are developed for each consumer. Finally, at point 6, an evaluation of the prediction model is carried out to assess how accurately we can determine consumer flexibility based on the predicted and measured consumption data for each individual consumer.

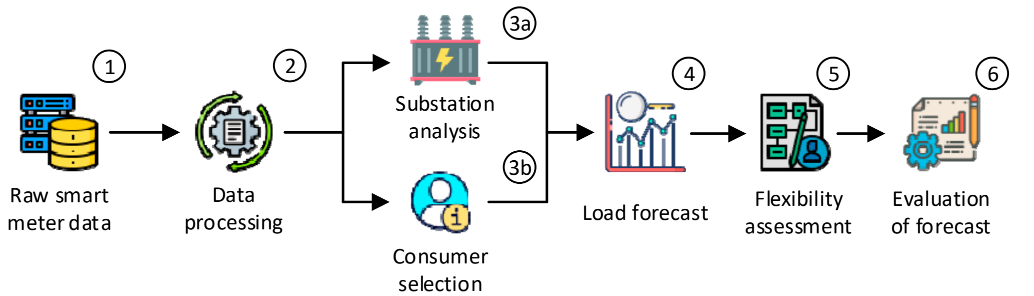

Figure 1. Process of consumer flexibility assessment.

The provided dataset underwent preprocessing to address missing values. Specifically, the missing data points were imputed using the load samples from the corresponding week before. The same approach has been used to replace extreme values caused by disturbances in the meter reading of data. Extreme values were defined as those exceeding 10 times the fuse nominal loading. The number of values that needed to be replaced varied among consumers. On average, approximately 567 samples had to be substituted for each consumer on a whole dataset, equivalent to 6 days' worth of values. The generated and consumed power was summarised for consumers with solar power plants. Following the data processing stage, it was observed that four consumers had invalid or incomplete readings for the entire observation period. Consequently, these consumers were excluded from further analysis. This was due to the machine learning's need for training models on valid historical load data over a longer period.

The first two years of the dataset were used for training the models, and a third year was used for the validation of the created model. The testing of the models was conducted on the remaining half a year of data. The split of the dataset and process of the training model are shown in Figure 2. In [16], the share of the training set was 80% and the testing set share was 20%. In our analysis, we considered validating models on a whole year of data, so the testing data share was 14% and the training set share was 86%.

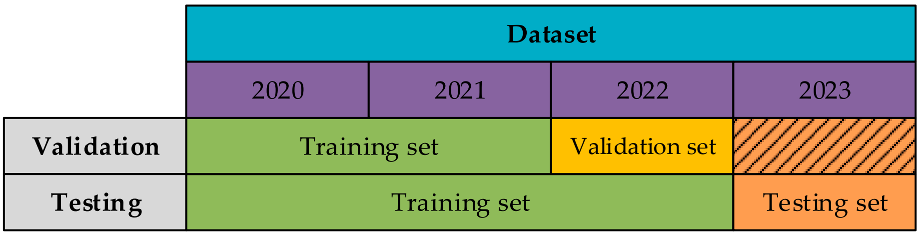

Figure 2. Split of the dataset for training, validation, and testing of the model.

In the process of machine learning, we first conducted the validation phase, where we trained the models using the training set's data, and then evaluated their performance on the validation set's data. During this iterative process, we experimented with different parameter configurations for the two models described in Section 2. Parameters leading to improved evaluation metrics were retained, while others were not included. Once satisfied with the model's performance, we proceeded to the testing phase, where we conducted the final evaluation on the testing set. In this step, the previous validation set was merged with the training set, and the model was trained further using this combined dataset. No further parameter modification was carried out at this stage, so the model parameters of validation and testing stayed the same. In the process of validation, we must avoid overfitting, so that the model has an equal performance on the training set as it does on the validation set. In our case, we considered overfitting by using the R^2 evaluation metrics, where values very close to 1 were considered as overfitting the model for the whole dataset evaluation. Addressing overfitting can also be made with other techniques, such as regularisation and cross-validation, where the train and test dataset are divided into multiple subsets.

4. Transformer Station Analysis

The selected transformer station was identified as a potential candidate for congestion management due to overloading that occurred at certain times of the year and specific hours. In Figure 3, the transformer measurement of apparent power for the whole period of observation is shown, along with the rated power of the transformer.

As Figure 3 shows, overloading occurred in the winter season between November and March. In the summer season, the power flow reversed; in May of the year 2023, the power of the solar power plant became so large that it overloaded the transformer. To gain comprehensive insights into the transformer load throughout the entire observation period, cumulative histograms were performed and are presented in Figure 4, showing the load distribution for each year of observation.

The loading of the transformer has varied over the years. The years 2020 and 2021 had the highest loading, with the year 2021 having a load of 50% rated power or higher for 37% of the time. For the years 2022 and 2023, this type of loading was present only 25% of the time. As for the overloading, the years 2020 and 2021 have had loads higher than the rated power for about 1% of the time, whereas the years 2022 and 2023 had it only for 0.5% of the time. Because of the increased number of connected solar power plants, the duration of transformer negative loading has become longer, with a higher magnitude of load. The variance of load can be attributed to the COVID-19 restrictions and penetration of solar power plants into the distribution grid. Figure 5 illustrates the minor proportion of overload duration observed in Figure 4, which is challenging to illustrate distinctly. Overloads are cumulated for each year, and categorised based on working and non-working days, represented using a bar diagram.

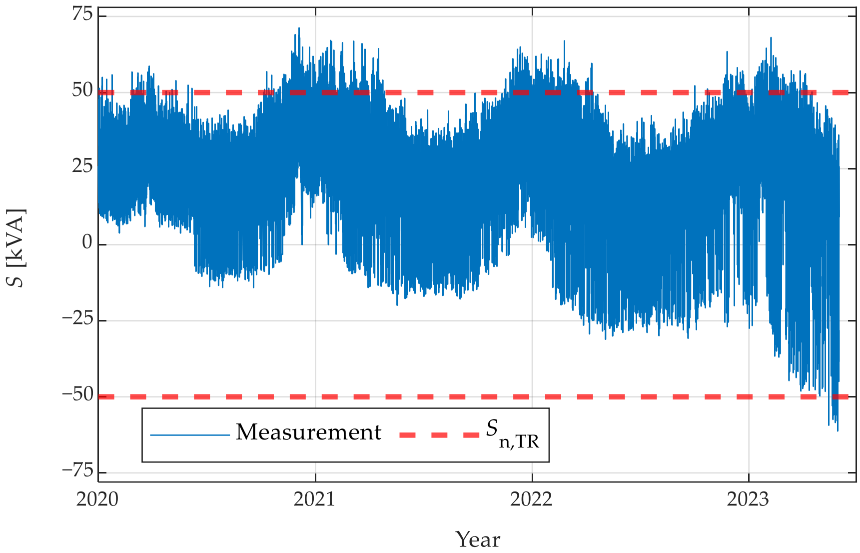

Figure 3. Apparent power measurement of the transformer at the observed transformer station.

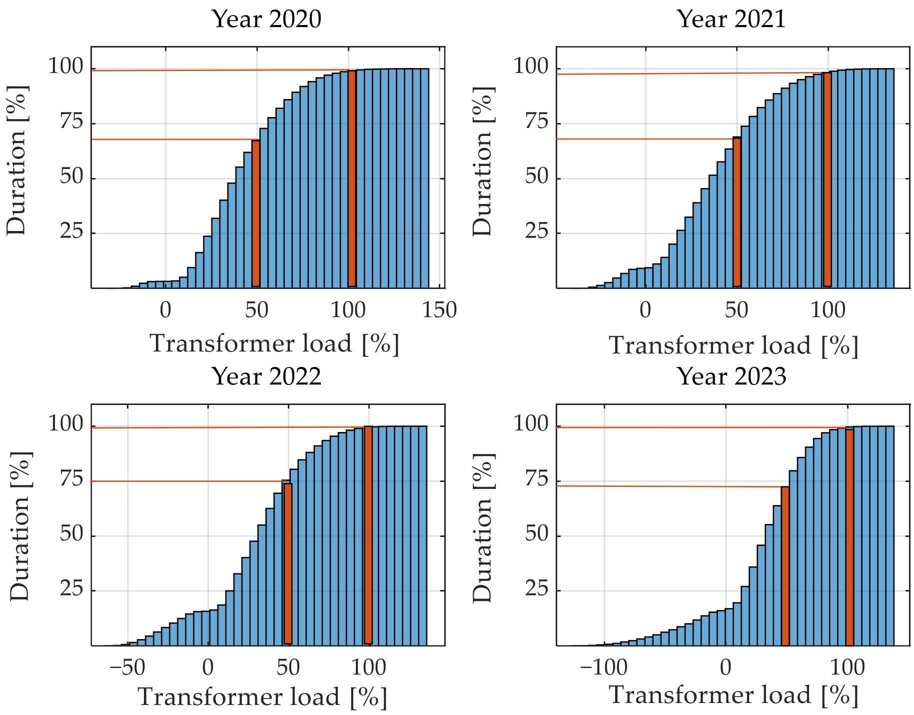

Figure 4. Cumulative histograms of transformer loading over the years.

As previously noted, the highest occurrences of overloading were observed in the years 2020 and 2021. Throughout all the years of observation, the proportion of overloading occurrences on weekdays versus weekends remained consistent, with the duration of weekday overloading being twice as high as that of weekends. In Figure 6, a bar diagram is shown, which shows which hours of the day had the most overloading over the past two years.

From Figure 6, it can be concluded that, in the year 2023, evening overloading had disappeared, with only the morning hours of 6 and 7 being the most critical. The drop in evening overloading can be contributed to the ending of the COVID-19 pandemic restrictions.

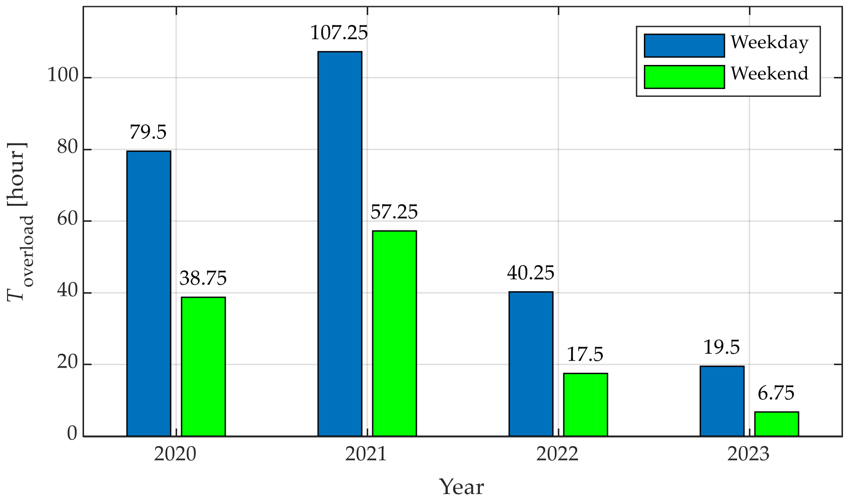

Figure 5. Overloading duration by the type of day for all the years of observation.

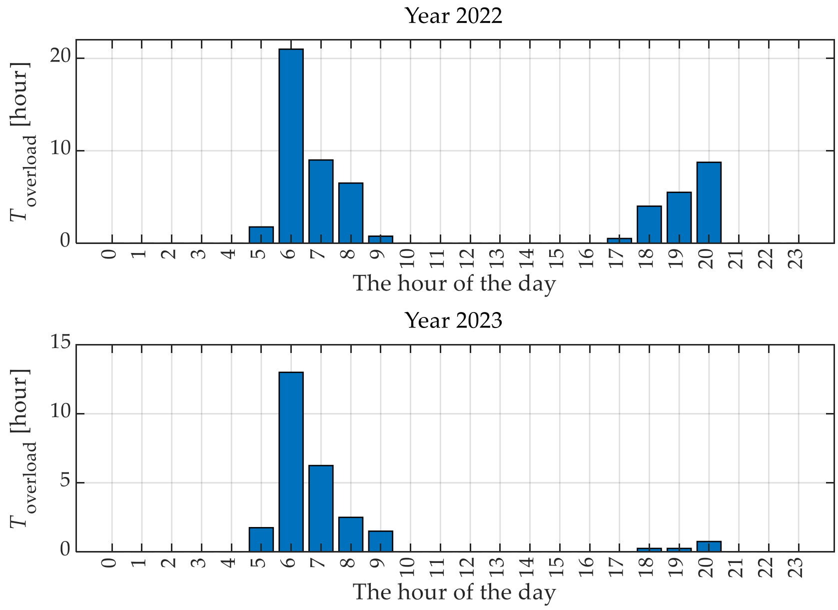

Figure 6. Duration of overloading by the hours for two years of observation.

Overloading can be conducted with the above analysis pattern of the transformer. The focus is on overloading between the hours of 6 and 7 with the emphasis on weekdays. To identify a set of potential consumers for assessing flexibility potential, the load of each consumer must undergo analysis. Figure 7 displays a box plot representing the consumers' load data for the year 2022, offering an efficient approach to examine the consumption patterns of multiple consumers simultaneously.

From the box plot, we can determine which consumer had their source of energy generation and what was their magnitude of consumption. Furthermore, we can detect unique patterns of consumption, as is the case of consumer 10, which had a small median and many outliers. This indicates that there was small consumption across the day, with exemptions where the load was very high. Consumers 5, 13, and 15 have a third quartile, with a wide area and a high magnitude of the maximum. Consumers 5, 10, 13, and 15 were therefore selected as the most prominent candidates for energy flexibility services. As potential candidates, consumers 3, 7, 8, and 9 were also considered, based on the number of outliers and the third quartile range. Analysis has shown that the load patterns did not coincide with overloading at critical hours. In Figure 8, a relationship between load and temperature is shown for the selected candidates 5, 10, 13, and 15. The cutoff temperature for the lower temperature range was established at 17 °C, while for the higher temperature range, a cutoff temperature of 25 °C was deemed appropriate.

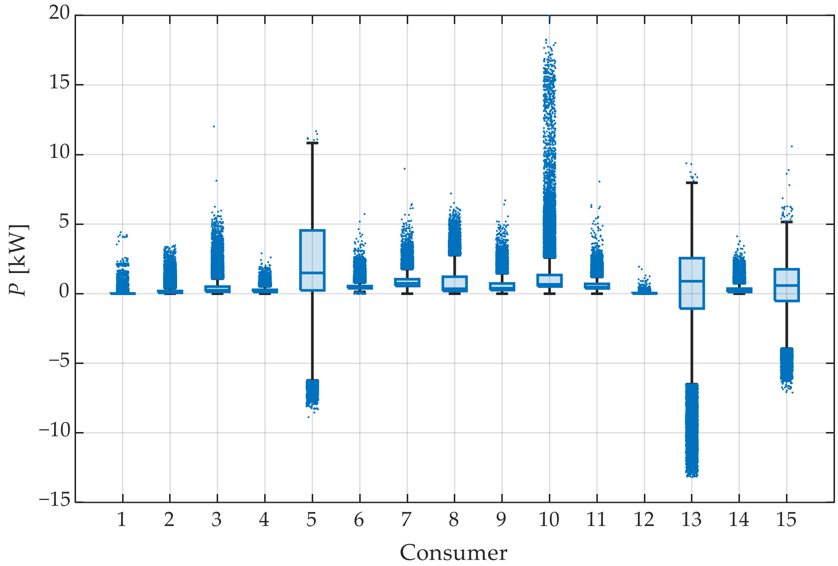

Figure 7. Box plot of consumers' consumption for the year 2022.

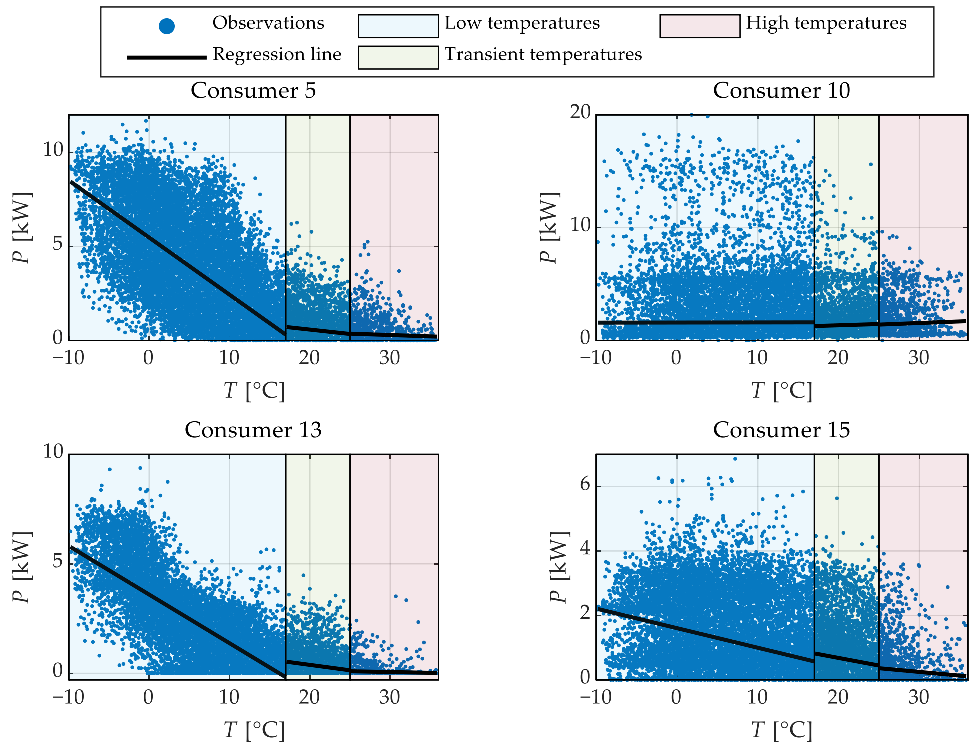

Figure 8. Relationship between the load and temperature for the selected consumers.

Based on the regression line, we can assume that consumers 5 and 13 were both temperature-dependent in the lower-temperature areas. All the observations are scattered evenly across the regression line compared to consumer 15, which has two constant loads split into two areas, therefore presenting an impression of temperature dependency. Consumer 10 has no temperature dependency. None of the selected consumers were temperature-dependent in the higher-temperature areas. The slope of the regression line indicates that consumer 5 was more temperature-dependent than consumer 13. The slope of the regression line for consumer 15 was significantly smaller in comparison to consumers 5 and 13, thereby supporting our indications of the existence of temperature dependency.

5. Results of the Energy Flexibility Assessment Study

The analysis of energy flexibility was conducted on the day of the testing set with the highest overload at the transformer station to establish the maximum amount of energy flexibility required. On 30 January, between 6:30 and 7:15 h, the transformer experienced its highest overload. The peak value during that period reached 64 kVA, which indicates that the transformer was operating at 128% of its nominal power. Theoretically, this suggests that a minimum power of 14 kVA is required in the worst-case scenario. The external temperature is also crucial for the actual overloading of the transformer; thus, local overloading is also allowed on colder days, which means that it would be necessary to upgrade the model. The load profile of the transformer for that specific day is depicted in Figure 9.

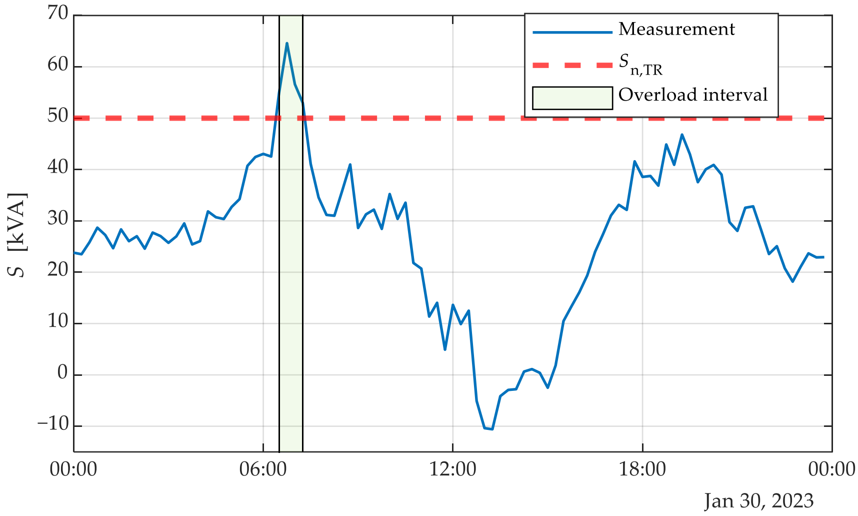

Figure 9. Transformer load profile on the day of the overload.

Following the morning overload event, there is evidence of another peak occurring later in the day. However, the power generated by the solar power plants mitigates this spike. In the evening, there was almost another overload event, with the load just under the nominal power of the transformer. Based on the observations in Figure 6, we can deduce that the overloads occurring after 7 h were due to the solar power plant not operating optimally, likely influenced by adverse weather conditions.

Figure 10 provides the daily measured load profile for the selected consumers, which will be used to assess the influence of consumers on the overloading event.

In the average model, the adoption of an 8-day averaging was assumed for optimal conditions, striking a balance between capturing specific load patterns and preventing excessive fluctuations, by avoiding both overly long and overly short averaging periods. Based on the presented load profiles in Figure 10, we can assume consumers' behaviour through the day. Consumer 5 exhibited a gradual power increase before the overload occurrence, and once again in the evening. On the other hand, consumer 10 experienced a rapid surge in power demand during the times of overload. Consumer 13 maintained a consistent high-power consumption throughout the day, disregarding the generation of solar plants. Predicting the load pattern for consumer 15 throughout the entire day poses a challenge; however, during the time of overload, there was a constant indication of an increase in power. As for consumers with solar power plants, their predicted power values were significantly off target, due to missing parameters in the prediction model, particularly the lack of accurate predicted solar observations.

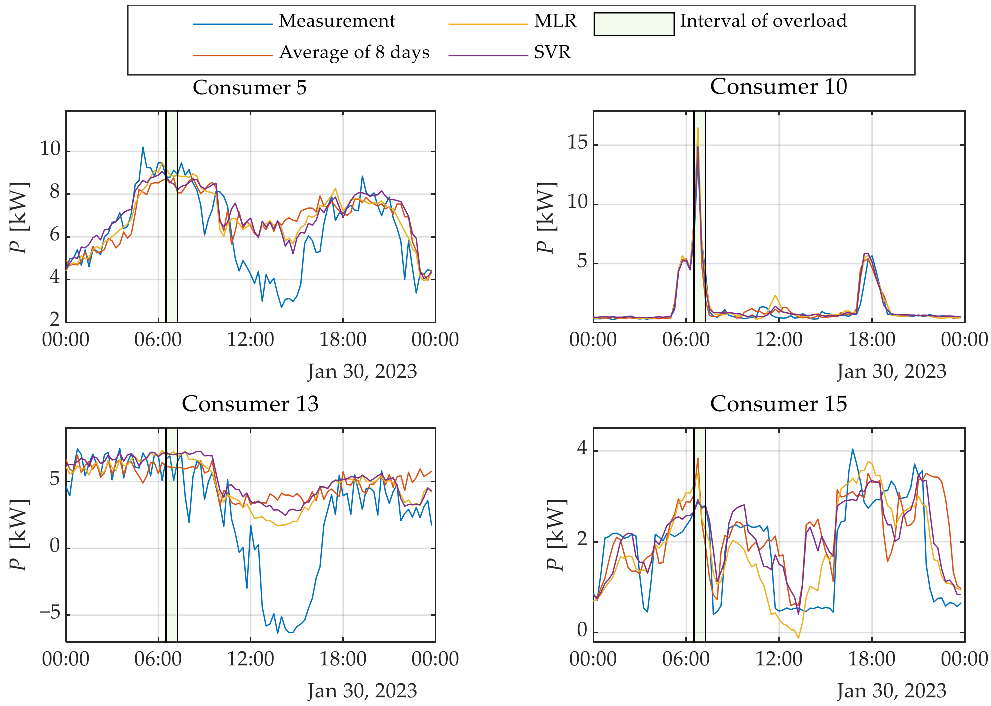

Figure 10. Load profiles of consumers 5, 10, 13, and 15 on the day with the highest overloading in the testing dataset

Figure 11 presents the measured and forecast loads for the time of the overload and before and after. Regression models generally provide a better match to the measured load profile. Notably, for consumer 13, the regression models outperformed the average model in capturing peak load behaviour. Furthermore, for consumer 15, the SVR model fitted the measured profile during the overload interval precisely.

Based on the forecast load, the consistency of the pattern can be determined, as behaviour throughout the day, which helps with the energy flexibility assessment. Consumer 15 was omitted from further analysis due to the inadequate magnitude of load at the time of the overloading event. As described, leveraging consumer energy flexibility allows us to handle transformer overloading adeptly by decreasing the consumer, thereby mitigating the risks of overloading. Demand response services such as peak shaving involves reducing consumer load partially, while load shifting entails optimising specific loads at alternative times. As mentioned in our case, to address the overloading issue between the hours of 6:30 and 7:15, an additional 15 kVA of power was required. Based on a comparison of the forecasted and measured load pattern of each consumer from Figure 11 and transformer overloading form Figure 9, the corelations can be made. Based on these corelations, the following scenarios are presented for decreasing transformer overloading:

- Scenario 1: Shifting the load of consumer 10 provides around 15 kW of energy flexibility.
- Scenario 2: By employing a combination of consumers 5 and 13, where each lowers their load by either 80% or 50%, there is sufficient flexibility to address the overloading problem effectively.
- Scenario 3: Any combination of load reduction and load shift of consumers 5 and 13, or load shift of consumer 10.

Consumer 5 exhibited a higher potential for load shifting in contrast to consumer 13, as the latter exhibited a constant load pattern with no room for shifting. Moreover, consumer 13 could provide flexibility throughout the entire day, as it maintained a constant load pattern. Above all, it is crucial to consider the rebound effect, which could result in overloading occurring at a different time, even after implementing load-shifting strategies.

Careful evaluation and planning are necessary to manage potential rebound effects and ensure a balanced and stable distribution network.

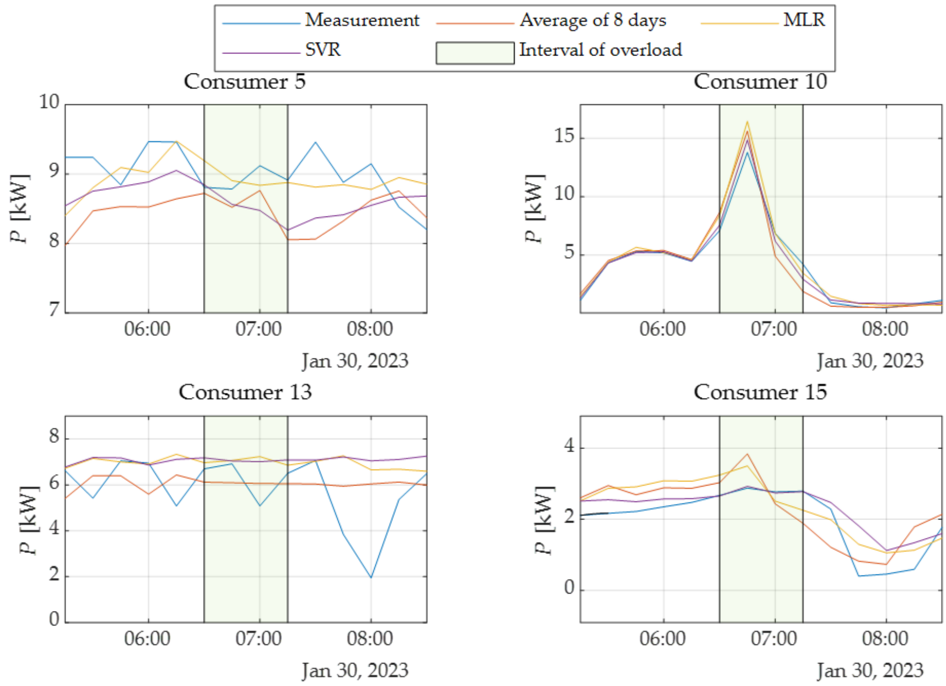

Figure 11. Load profile of consumers 5, 10, 13, and 15 in the interval of the highest overloading in the testing dataset.

SVR and MLR presented similar results, and to assess the performance of the utilised models, a straightforward method is to compare the predicted values with the measured values of the testing dataset on the same plot, as shown in Figure 12 for the case of SVR. This graphical representation enables a rapid evaluation of prediction accuracy. By comparing the predicted and measured data points visually, any discrepancies or deviations from the regression line can be observed, providing valuable insights into the accuracy and effectiveness of the models.

As can be seen from Figure 12, the observations are scattered evenly across the regression line for most consumers, except for consumer 10, which exhibited noticeable residuals and unevenly scattered observations at higher loads. Additionally, for consumers with solar power plants, the negative load observations are not scattered evenly due to the poor prediction model for solar generation. The irregular distribution of data points away from the regression line indicates discrepancies between the predicted values and the actual measurements, suggesting potential areas for model improvement and further analysis.

To provide a precise evaluation of each used model, evaluation metrics calculations were performed, as described in Section 2.2. The results of these evaluations are presented in Table 1 for the entire day of overload observations.

On the day of the overload event, the regression models outperformed the averaging model consistently in all cases. While some consumers may benefit from an SVR model, others may find better results with an MLR model. However, in general, both the SVR and MLR models produce similar outcomes, indicating their comparable performance in handling the overload event. The selection of the most suitable model may depend on the specific characteristics and behaviour of each consumer's load profile. The interval of the overload evaluation of models is given in Table 2.

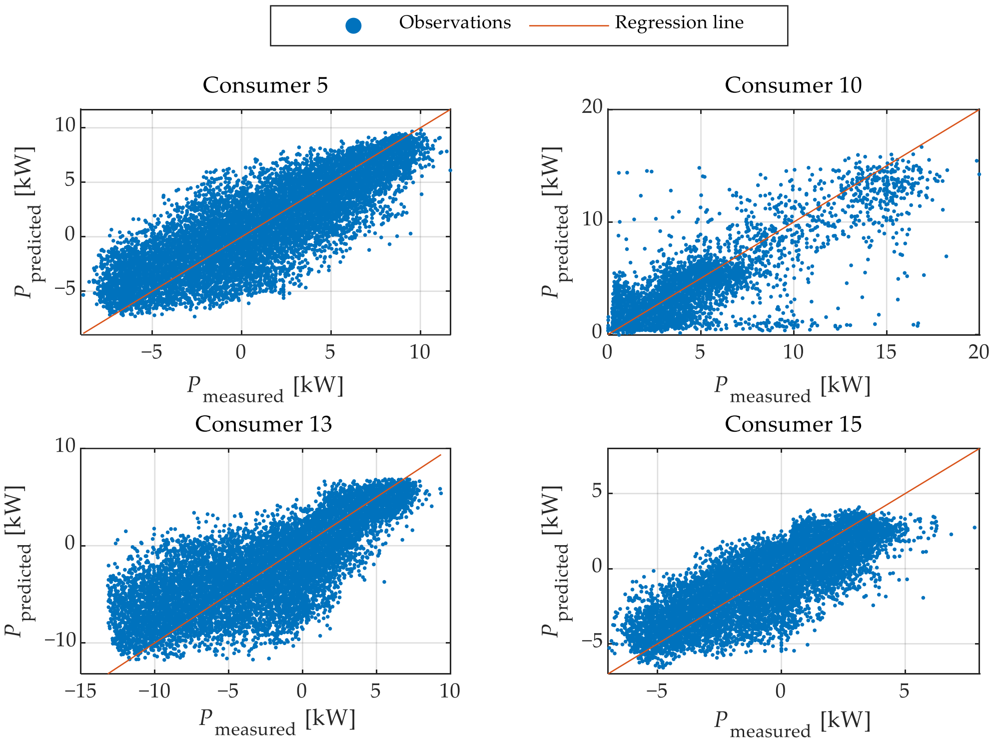

Figure 12. Prediction plot of SVR on the testing dataset for the selected consumers.

Table 1. Evaluation of models for a day of overload.

Consumer		**5**	**10**	**13**	**15**
MAE (kW)	Averaging	1.1224	0.3225	2.6475	0.7483
	MLR	0.9821	0.3053	2.2714	0.5810
	SVR	1.0297	0.2724	2.5141	0.5783
MAPE (%)	Averaging	24.39	33.97	187.05	81.89
	MLR	21.49	37.21	161.62	51.00
	SVR	22.48	33.48	202.19	65.71
RMSE (kW)	Averaging	1.5429	0.6104	4.0271	0.9948
	MLR	1.3636	0.5614	3.4253	0.6670
	SVR	1.4118	0.4771	3.7908	0.7413
R^2 (/)	Averaging	0.4263	0.9194	0.5300	0.2380
	MLR	0.5813	0.9483	0.7531	0.6146
	SVR	0.6164	0.9494	0.6961	0.5290

In the case of the overloading interval, the average model showed slightly better performance for consumer 13 compared to the other consumers. However, for all the other consumers, the SVR and MLR predictions outperformed the average model significantly. The evaluations for the entire testing dataset, providing a comprehensive assessment of the model's overall accuracy, can be found in Table 3.

Evaluation metrics play a crucial role in selecting the most appropriate model for each consumer in our novel approach to assess flexibility. For a general evaluation, the averaging model proved to be superior only in cases where the consumption pattern remained constant, as observed with consumer 10. However, for all the other cases, the regression models outperformed the averaging model. Both the MLR and SVR models yielded comparable results in terms of performance. It is worth noting again that the

MLR model was constructed with more features than the SVR. However, overall, the MLR model, with our customised parameters, was more efficient than the SVR model, as the computation time for the MLR model was significantly lower than SVR while yielding comparable results, which can make the process of novel energy flexibility assessment much faster. Therefore, we can consider the proposed MLR model as the best viable option for a quick and accurate assessment of consumer flexibility potential.

Table 2. Evaluation of models for an interval of overload.

Consumer		5	10	13	15
MAE (kW)	Averaging	0.3910	1.9206	0.7078	0.6404
	MLR	0.2055	1.1813	0.7298	0.4992
	SVR	0.4054	0.8699	0.7798	0.0285
MAPE (%)	Averaging	4.378	29.75	11.68	22.85
	MLR	2.305	13.93	13.46	17.96
	SVR	4.517	13.60	13.98	1.017
RMSE (kW)	Averaging	0.4851	1.9416	0.7371	0.7049
	MLR	0.2466	1.5248	1.0999	0.5188
	SVR	0.4957	0.9298	1.0378	0.0321
R^2 (/)	Averaging	0.0387	0.9371	0.3380	0.1015
	MLR	0.3294	0.9901	0.5171	0.0114
	SVR	0.1547	0.9894	0.3278	0.9498

Table 3. Evaluation of models for the whole testing dataset.

Consumer		5	10	13	15
MAE (kW)	Averaging	1.4918	0.4671	1.7	0.9734
	MLR	1.3267	0.4945	1.5103	0.9138
	SVR	1.3070	0.4673	1.4892	0.8974
MAPE (%)	Averaging	144.12	44.90	187.46	203.814
	MLR	126.63	52.83	226.90	177.682
	SVR	121.45	46.49	210.58	178.13
RMSE (kW)	Averaging	1.9297	0.8923	2.3994	1.2985
	MLR	1.76022	0.8866	2.0835	1.1798
	SVR	1.7326	0.8734	2.1286	1.2035
R^2 (/)	Averaging	0.6242	0.7885	0.6058	0.5848
	MLR	0.6823	0.7967	0.7185	0.6700
	SVR	0.6915	0.7931	0.7165	0.6486

The SVR model was optimised on a limited range, due to the lengthy computation time involved. For consumers with solar power plants (5, 10, and 15), the MAPE values were relatively high, attributed primarily to the poor forecasts of solar power plant generation and values close to zero during the transition into or from the generation operation. In this case, other validation metrics provide better accuracy over a longer observation period for an overall comparison.

During the evaluation of the test dataset, we examined the reaction of temperature-dependent consumer regression models to temperature changes, considering two scenarios: one without any input feature, and another with the input of temperature. Figure 13 presents a comparison of these two models for consumer 13, illustrating their performance with and without temperature incorporation into the model. With this approach, our goal was to achieve accurate assessments of energy flexibility for temperature-dependent consumers.

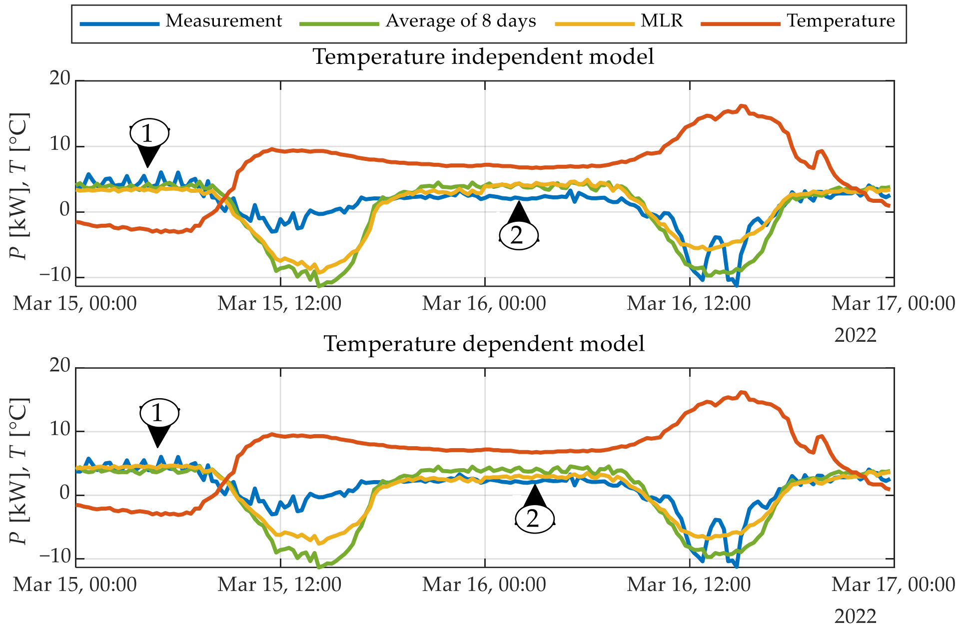

Figure 13. Comparison of the temperature-dependent and independent model in the case of the MLR load forecast for consumer 13.

As Figure 13 illustrates, the temperature decreased on the morning of March 15. In the temperature-independent model, at marked point (1), the load did not increase as it did in the temperature-dependent model, and therefore moved closer to the measured load. On the following day, at marked point (2), the temperature increased compared to the previous day, and in the temperature-independent model, the predicted load was high compared to the temperature-dependent model, where the load decreased.

This observation highlights the temperature-dependent model's ability to determine flexibility potential more accurately for temperature-dependent consumers, which helped us in our novel approach to assess energy flexibility. By capturing load changes more precisely based on temperature fluctuations, this model offers an advantage in managing and optimising energy consumption for such consumers. Figure 8 portrays visually how some consumers are responsive to temperature variations, and tend to increase their consumption during colder periods, leading to heightened energy demands and increased transformer loading. This elevated loading augments the potential for transformer overloading, especially when temperature fluctuations are in play.

Conversely, temperature-independent models exhibited similar behaviour to the averaging model, indicating their limitations in considering temperature variations for load prediction.

6. Conclusions

In response to the persistent challenges arising from ageing distribution grids and escalating congestion due to the integration of new loads, this paper introduces an innovative approach to evaluate the energy flexibility potential of DN consumers. Through a comprehensive case study conducted on a selected transformer station, which has been grappling with recurrent overloading issues, the research explores alternative methodologies for mitigating such problems. This paper presents a novel approach to assess flexibility, which prioritises the most prominent consumers for flexibility services and forecasts their load patterns to identify load profile matching with overloads. The approach suggests that machine learning models are the optimal methods for the quick and precise forecasting of consumer load profiles.

In conjunction with the conventional simple averaging model, two sophisticated regression models were employed to forecast consumption patterns during transformer station overloading. From the set of consumers connected to the transformer station, a careful selection process identified four candidates, with appropriate historical load data for in-depth analysis. Ultimately, three consumers emerged as viable candidates for providing energy flexibility services.

The precise evaluation of these models, utilising standard evaluation metrics, demonstrated the superior performance of the regression models compared to the simple averaging model. This showcases their efficacy in predicting consumption behaviour accurately during transformer station overloading events, thus enhancing the capabilities of our novel approach to assess flexibility. However, it is essential to acknowledge that the proposed approach applies to a specific subset of consumers and transformer stations characterised by high loads.

Using the case study data and our proposed novel approach for flexibility assessment based on consumer load patterns, we offer a quick and efficient method to assess flexibility potential for DN consumers. This unique approach provides DN operators with a data-driven foundation to make well-informed decisions concerning energy flexibility services, and optimise the integration of renewable energy sources into the grid. Consequently, consumers can be integrated into the flexibility market more easily, while simultaneously defining optimal flexibility scenarios tailored to each consumer's unique demands. The outcomes provide a basis for future studies and implementations aimed at enhancing the overall efficiency and resilience of distribution networks in the face of increasing demands and ageing infrastructure. In future work, it is possible to quantify consumer flexibility potential by analysing their appliance activities and utilising the flexibility indicators. Additionally, to make our assessment of flexibility potential more accurate, we can try using different models for load forecasting, as explained in Section 2. However, we need to keep in mind that calculating potential for each consumer takes time, which is a limitation here. While some methods are more accurate, we might not have full control over how exactly these models work, and they can be hard to tune.

Author Contributions: Conceptualization, M.B.; Methodology, J.R. and L.M.; Software, J.R.; Validation, J.R.; Data curation, L.M.; Writing—original draft, J.R.; Writing—review & editing, J.R.; Visualization, M.B.; Supervision, L.M. and M.B.; Project administration, M.B.; Funding acquisition, M.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

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

References

1. Majeed, I.B.; Nwulu, N.I. Reverse Power Flow Due to Solar Photovoltaic in the Low Voltage Network. *IEEE Access* **2023**, *11*, 44741–44758. [CrossRef]
2. Tavakoli, A.; Saha, S.; Arif, M.T.; Haque, M.E.; Mendis, N.; Oo, A.M.T. Impacts of grid integration of solar PV and electric vehicle on grid stability, power quality and energy economics: A review. *IET Energy Syst. Integr.* **2020**, *2*, 243–260. [CrossRef]
3. Pater, S. Increasing Energy Self-Consumption in Residential Photovoltaic Systems with Heat Pumps in Poland. *Energies* **2023**, *16*, 4003–4017. [CrossRef]
4. Li, P.; Wang, Y.; Ji, H.; Zhao, J.; Song, G.; Wu, J.; Wang, C. Operational flexibility of active distribution networks: Definition, quantified calculation and application. *Int. J. Electr. Power Energy Syst.* **2020**, *119*, 105872. [CrossRef]
5. Luo, Z.; Peng, J.; Cao, J.; Yin, R.; Zou, B.; Tan, Y.; Yan, J. Demand Flexibility of Residential Buildings: Definitions, Flexible Loads, and Quantification Methods. *Engineering* **2022**, *16*, 123–140. [CrossRef]
6. Binyet, E.; Chiu, M.-C.; Hsu, H.-W.; Lee, M.-Y.; Wen, C.-Y. Potential of Demand Response for Power Reallocation, a Literature Review. *Energies* **2022**, *15*, 863. [CrossRef]

7. Chen, L.; Xu, Q.; Yang, Y.; Song, J. Optimal Energy Management of Smart Building for Peak Shaving Considering Multi-Energy Flexibility Measures. *Energy Build.* **2021**, *241*, 110932. [CrossRef]
8. Li, H.; Wang, Z.; Hong, T.; Piette, M.A. Energy flexibility of residential buildings: A systematic review of characterization and quantification methods and applications. *Adv. Appl. Energy* **2021**, *3*, 100054. [CrossRef]
9. Klyapovskiy, S.; You, S.; Michiorri, A.; Kariniotakis, G.; Bindner, H.W. Incorporating flexibility options into distribution grid reinforcement planning: A techno-economic framework approach. *Appl. Energy* **2019**, *254*, 113662. [CrossRef]
10. Silva, C.; Faria, P.; Vale, Z. Demand Response Implementation: Overview of Europe and United States Status. *Energies* **2023**, *16*, 4043. [CrossRef]
11. Thorvaldsen, K.E.; Nessa, S.; Sandell, S.; Sæle, H. Spatial dependency on flexibility value considering stacked transformer overload cost. *TechRxiv* **2023**, preprint. [CrossRef]
12. Sridhar, A.; Honkapuro, S.; Ruiz, F.; Stoklasa, J.; Annala, S.; Wolff, A.; Rautiainen, A. Toward residential flexibility—Consumer willingness to enroll household loads in demand response. *Appl. Energy* **2023**, *342*, 121204. [CrossRef]
13. Mousavi, M.; Wu, M. A DSO Framework for Market Participation of DER Aggregators in Unbalanced Distribution Networks. *IEEE Trans. Power Syst.* **2021**, *14*, 2247–2258. [CrossRef]
14. Valentini, O.; Andreadou, N.; Bertoldi, P.; Lucas, A.; Saviuc, I.; Kotsakis, E. Demand Response Impact Evaluation: A Review of Methods for Estimating the Customer Baseline Load. *Energies* **2022**, *15*, 5259. [CrossRef]
15. Afzalan, M.; Jazizadeh, F. Residential loads flexibility potential for demand response using energy consumption patterns and user segments. *Appl. Energy* **2019**, *254*, 113693. [CrossRef]
16. Luo, X.; Yang, Y.; Liu, Y.; Zhao, T. Classification of energy use patterns and multi-objective optimal scheduling of flexible loads in rural households. *Energy Build.* **2023**, *283*, 112811. [CrossRef]
17. Kwac, J.; Flora, J.; Rajagopal, R. Household Energy Consumption Segmentation Using Hourly Data. *IEEE Trans. Smart Grid* **2014**, *5*, 420–430. [CrossRef]
18. Rehman, W.U.; Sajjad, I.A.; Khan, M.F.N. Assessment of Demand Flexibility Potential of Residential Consumers in Pakistan. In Proceedings of the 2018 2nd International Conference on Energy Conservation and Efficiency (ICECE), Lahore, Pakistan, 16–17 October 2018; pp. 77–85.
19. Bolfek, M.; Capuder, T. A practical approach to flexibility provision assessment in an unobservable distribution network. *Electr. Power Syst. Res.* **2022**, *212*, 108262. [CrossRef]
20. Jazaeri, J.; Alpcan, T.; Gordon, R.; Brandao, M.; Hoban, T.; Seeling, C. Baseline methodologies for small scale residential demand response. In Proceedings of the 2016 IEEE Innovative Smart Grid Technologies—Asia (ISGT-Asia), Melbourne, Australia, 28 November–1 December 2016; pp. 747–752.
21. Madhukumar, M.; Sebastian, A.; Liang, X.; Jamil, M.; Shabbir, M.N.S.K. Regression Model-Based Short-Term Load Forecasting for University Campus Load. *IEEE Access* **2022**, *10*, 8891–8905. [CrossRef]
22. Weng, Y.; Rajagopal, R. Probabilistic baseline estimation via Gaussian process. In Proceedings of the 2015 IEEE Power Energy Society General Meeting, Denver, CO, USA, 26–30 July 2015; pp. 1–5.
23. Tehrani, N.H.; Khan, U.T.; Crawford, C. Baseline load forecasting using a Bayesian approach. In Proceedings of the 2016 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), Vancouver, BC, Canada, 15–18 May 2016; pp. 1–4.
24. Oyedokun, J.; Bu, S.; Han, Z.; Liu, X. Customer Baseline Load Estimation for Incentive-Based Demand Response Using Long Short-Term Memory Recurrent Neural Network. In Proceedings of the 2019 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe), Bucharest, Romania, 29 September–2 October 2019; pp. 1–5.
25. Arunaun, A.; Pora, W. Baseline Calculation of Industrial Factories for Demand Response Application. In Proceedings of the 2018 IEEE International Conference on Consumer Electronics—Asia (ICCE-Asia), JeJu, Republic of Korea, 24–26 June 2018; pp. 206–212.
26. Li, K.; Wang, B.; Wang, Z.; Wang, F.; Mi, Z.; Zhen, Z. A baseline load estimation approach for residential customer based on load pattern clustering. *Energy Procedia* **2017**, *142*, 2042–2049. [CrossRef]
27. Dong, Y.; Ma, X.; Ma, C.; Wang, J. Research and Application of a Hybrid Forecasting Model Based on Data Decomposition for Electrical Load Forecasting. *Energies* **2016**, *9*, 1050. [CrossRef]
28. Ruiz-Abellón, M.D.C.; Gabaldón, A.; Guillamón, A. Load Forecasting for a Campus University Using Ensemble Methods Based on Regression Trees. *Energies* **2018**, *11*, 2038. [CrossRef]
29. Hong, T.; Wang, P.; Willis, H.L. A Naïve multiple linear regression benchmark for short term load forecasting. In Proceedings of the 2011 IEEE Power and Energy Society General Meeting, Detroit, MI, USA, 24–28 July 2011; pp. 1–6.
30. Limmer, S.; Einecke, N. An Efficient Approach for Peak-Load-Aware Scheduling of Energy-Intensive Tasks in the Context of a Public IEEE Challenge. *Energies* **2022**, *15*, 3718. [CrossRef]