

# Clustering Informed MLP Models for Fast and Accurate Short-Term Load Forecasting

Athanasios Ioannis Arvanitidis <sup>1,\*</sup>, Dimitrios Bargiotas <sup>1</sup>, Aspasia Daskalopulu <sup>1</sup>,  
Dimitrios Kontogiannis <sup>1</sup>, Ioannis P. Panapakidis <sup>1</sup> and Lefteri H. Tsoukalas <sup>2</sup>

<sup>1</sup> Department of Electrical and Computer Engineering, University of Thessaly, 38334 Volos, Greece; bargiotas@uth.gr (D.B.); aspasia@uth.gr (A.D.); dimkonto@uth.gr (D.K.); panapakidis@uth.gr (I.P.P.)

<sup>2</sup> Center for Intelligent Energy Systems (CiENS), School of Nuclear Engineering, Purdue University, West Lafayette, IN 47907, USA; tsoukala@ecn.purdue.edu

\* Correspondence: atarvanitidis@uth.gr

**Abstract:** The stable and efficient operation of power systems requires them to be optimized, which, given the growing availability of load data, relies on load forecasting methods. Fast and highly accurate Short-Term Load Forecasting (STLF) is critical for the daily operation of power plants, and state-of-the-art approaches for it involve hybrid models that deploy regressive deep learning algorithms, such as neural networks, in conjunction with clustering techniques for the pre-processing of load data before they are fed to the neural network. This paper develops and evaluates four robust STLF models based on Multi-Layer Perceptrons (MLPs) coupled with the K-Means and Fuzzy C-Means clustering algorithms. The first set of two models cluster the data before feeding it to the MLPs, and are directly comparable to similar existing approaches, yielding, however, better forecasting accuracy. They also serve as a common reference point for the evaluation of the second set of two models, which further enhance the input to the MLP by informing it explicitly with clustering information, which is a novel feature. All four models are designed, tested and evaluated using data from the Greek power system, although their development is generic and they could, in principle, be applied to any power system. The results obtained by the four models are compared to those of other STLF methods, using objective metrics, and the accuracy obtained, as well as convergence time, is in most cases improved.

**Keywords:** short-term load forecasting; multi-layer perceptrons; K-Means; Fuzzy C-Means



**Citation:** Arvanitidis, A.I.; Bargiotas, D.; Daskalopulu, A.; Kontogiannis, D.; Panapakidis, I.P.; Tsoukalas, L.H. Clustering Informed MLP Models for Fast and Accurate Short-Term Load Forecasting. *Energies* **2022**, *15*, 1295. <https://doi.org/10.3390/en15041295>

Academic Editor: Andrzej Bielecki

Received: 16 January 2022

Accepted: 8 February 2022

Published: 10 February 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Two requirements of the Net Zero by 2050 initiative for “greener” power grids, such as the integration of renewable energy sources, and the connection of volatile loads, such as electric vehicles, affect the stable and efficient operation of power systems dramatically. Satisfying load needs instantly and at all times becomes a major challenge, and all aspects of the operation and management of power plants, such as economic dispatch [1], demand side management [2], price forecasting [3,4], maintenance scheduling and the formulation of an effective bidding strategy in power system markets [5], along with the financial viability of electrical companies themselves, are increasingly relying on accurate load predictions [6].

Electric load forecasting has, justifiably, been the focus of much research, and work in the area is classified into three categories based on the time horizon and the operational choice that must be made, namely short-term, medium-term, and long-term forecasting. Long-term load forecasting generally spans 20 years and is required for planning purposes, such as the construction of new power plants and the upgrade of transmission system capacity. Medium term load forecasting ranges from a few weeks to a year and is mostly used for scheduling maintenance and fuel supply [7]. The day-to-day functioning of the power system necessitates Short-Term Load Forecasting (STLF), which is primarily

influenced by temporal factors (for example, weekly periodicity and seasonal fluctuations) and weather conditions (for example, humidity, temperature, wind speed, and cloud coverage) [8]. STLF is considered essential for the smooth and uninterrupted operation of a power system, because it enables load flow studies and contingency analysis, on issues such as bus voltages, line currents, power generation, and line flows. Therefore, in order to achieve high accuracy in forecasting results, various load forecasting models have been developed and investigated [9].

Traditional methodologies for STLF include time series models, regression models, and Kalman filtering-based procedures [10,11]. Artificial intelligence and deep learning approaches, on the other hand, are considered state-of-the-art and include Artificial Neural Networks (ANNs) [12], such as Multi-Layer Perceptrons (MLPs) [13], Radial Basis Function Neural Networks (RBFNNs) [14], Convolutional Neural Networks (CNNs) [15], Recurrent Neural Networks (RNNs) [16], Support Vector Machines (SVMs) [17], Decision Trees (DT) or Random Forests (RF) [18], and Fuzzy-Neural models, offering high accuracy and effective convergence time for the STLF problem.

In the context of these deep learning approaches, the demand for even more accurate predictions has led researchers to develop hybrid forecasting models, which integrate a clustering algorithm for the pre-processing of data before it is used to train the neural network. Typically, clustering methods are implemented in order to create clusters of the load data, which is first pre-processed using an enhanced min-max scaling method [13]. The subject of short-term load forecasting coupled with a clustering strategy, has been extensively studied using methods based on RNNs, Long Short-Term Memory (LSTM), CNNs, SVMs [19], ANNs, Simple Exponential Smoothing (SES) and Group Method of Data Handling (GMDH) algorithms [20].

Hernández et al. [21], use a hybrid clustering approach to evaluate short-term load forecasts on the Soria microgrid. First, a Self-Organizing Map (SOM) is used to categorize historical data and the K-Means clustering method is then applied to group the data of each category. To achieve appropriate forecasts of the load curve, a separate MLP for each data cluster is trained. Despite its complexity, this method achieves a Mean Absolute Percentage Error (MAPE) near 2%. In a similar attempt, Farfar et al. [22] utilize a hybrid forecasting model based on a clustering approach of load profiles alongside a daily temperature estimator. Artificial neural networks for the daily load forecast for each cluster are used in the regression phase, with initial weights computed through stacked denoising autoencoders. Each cluster's MAPE does not decrease below 1.9%.

In [23], K-Means is applied to cluster load data and CNNs are utilized to estimate the following day's load in conjunction with meteorological and consumer categorization data. The researchers recorded for winter days forecast values with MAPE equal to 7.41%, while for summer days the results showed MAPE close to 3.06%. Unlike prior studies, where the K-Means application was solely used to load data, a novel clustering technique is introduced in [24]. The authors suggest using the clustering technique to normalized input variables, which include weather and label data that reflect seasonal features in addition to load data. The MAPE of the predicted load values obtained using the suggested technique is about 2%.

In addition to K-Means, there is a plethora of papers about the Fuzzy C-Means (FCM) clustering method in the literature for both short-term load forecasting and power generation forecasting by solar systems [25] and wind turbines [26].

Bian et al. [27], propose data grouping in clusters based on the strong or weak correlation at adjacent moments and not on similar load profile. They apply FCM to further cluster the data to display similar values. Finally, the sets are fed into the NNs, which is used to STLF with MAPE at about 2.01%. In [28], the FCM clustering algorithm based on Principal Component Analysis (PCA) is applied to cluster real-time load data of power systems in NSW State, Australia, at half hourly intervals. The centers of an RBF neural network are determined using PCA. Although the forecasted values obtained through the proposed technique have a fairly high MAPE (specifically 5.1%), it is more accurate

than other simpler techniques such as an approach based on a RBF neural network and an approach based on a RBF neural network in conjunction with the FCM algorithm. A similar attempt to utilize the FCM algorithm for load prediction with the Self-Normalizing Gated Recurrent Units (GRU) application is described in [29]. FCM is applied to normalized data in order to create clusters of data that belong to similar days. In this scenario, the MAPE does not drop below 2.6%.

This paper presents four generic robust hybrid STLF models, which use MLPs neural networks and the K-Means and Fuzzy C-Means clustering techniques. The models are designed, tested, and evaluated using data from the Greek power system, however they are generic, in the sense that they can be applied to the specific load data of any power system. The first set of two models are developed by initially applying the K-Means and Fuzzy C-Means clustering techniques to the load data, in order to generate optimal clusters, and then feeding each cluster to a MLP to produce short-term load predictions. These two models are similar to existing methods, and they were developed in order to serve as a common reference point for comparison with the second set of two models. However, these first two models contribute some novelty, though similar in spirit to existing methods, because they achieve MAPE well below 2% (around 1.70, a 25% improvement), which is the current best, as indicated by the preceding discussion of related work. The second set of two models were developed in order to improve the first set further, by using a single MLP per clustering method, which is fed with the original load data set and an additional input variable containing the cluster label of each point in the load data set. Hence, one can think of these two models as improved versions of the first set. The labeling information that is used to extend the input to the MLP is produced by the K-Means and Fuzzy C-Means clustering techniques and the use of the elbow optimization method. The forecasting results obtained by the second set of models are also better than those of other approaches with MAPE well below the current best of 2%. Moreover, all four models are compared to other existing load forecasting approaches, and to each other and exhibit shorter convergence time compared to classical data pre-processing approaches [21–24,27].

In the remainder of this paper we explain the clustering algorithms that were employed, as well as the performance measures, before proceeding with the details of the four models that were developed. The results are shown and discussed for each model, in relation to the other models and related work.

## 2. Materials and Methods

### 2.1. Clustering Methods for Short-Term Load Forecasting

Clustering is an unsupervised machine learning approach that partitions a dataset into groups (clusters) so that data in the same cluster are close to one another and hence very similar. K-Means [30] and Fuzzy C-Means [31] are two of the most prevalent clustering algorithms used in STLF.

#### 2.1.1. K-Means Clustering Algorithm

K-Means clustering begins with the selection of  $K$  representative points among the dataset as the initial centroids. Based on the Euclidean distance metric, each point in the dataset is subsequently assigned to the nearest centroid. The centroids for each cluster are updated after the clusters are generated. The algorithm then iteratively executes these two steps until the centroids do not change any further. The selection of the optimum number of clusters, indicated by the parameter  $K$ , is derived by proper objective functions, the most important of which is the Sum of Squared Errors (SSE), which is defined mathematically by Equation (1), and must be minimized:

$$SSE(C) = \sum_{k=1}^K \sum_{x_i \in C_k} \|x_i - c_k\|^2 \quad (1)$$

where  $C$  indicates a cluster,  $x_i$  is an instance of the given dataset that consists of  $N$  points and  $c_k$  is the centroid of cluster  $C_k$ . The centroid of each cluster is updated iteratively through Equation (2):

$$c_k = \frac{\sum_{x_i \in C_k} x_i}{|C_k|} \quad (2)$$

where  $|C_k|$  is the total number of points in cluster  $k$ .

### 2.1.2. Fuzzy C-Means Clustering Algorithm

Strict assignment of points to clusters is not possible in incomplete datasets with overlapping clusters (i.e., where the original dataset cannot be partitioned). As a result, K-Means would produce an inappropriate segmentation of data into clusters. A fuzzy clustering approach (often called soft K-Means clustering) may be used to retrieve such overlapping structures. Each data point in the FCM technique is assigned a probability score that reflects its membership to a given cluster, therefore point membership in various clusters might range from 0 to 1, with 0 denoting no membership, 1 denoting total membership, and intermediate values denoting varying degrees of membership. The sum of memberships of a given point to various clusters must be 1.

The purpose of FCM, as in the case of K-Means, is the reduction of SSE. The membership weight of point  $x_i$  belonging to cluster  $C_k$  is represented by  $w_{xik}$  and is utilized as an FCM update step. The calculation of  $w_{xik}$  is derived from Equation (3):

$$w_{xik} = \frac{1}{\sum_{j=1}^K \left( \frac{x_i - c_k}{x_i - c_j} \right)^{\frac{2}{\beta-1}}} \quad (3)$$

where  $x_i$  is an instance of the given dataset that consists of  $N$  points,  $c_k$  is the centroid of cluster  $C_k$ , and  $\beta$  is a parameter that determines the fuzziness of the cluster. Equation (4) calculates the weighted centroid for  $C_k$  based on the fuzzy weights, and Equation (5) provides the SSE function for each cluster  $C$  defined by the FCM:

$$c_k = \frac{\sum_{x_i \in C_k} w_{xik}^{\beta} \cdot x_i}{\sum_{x_i \in C_k} w_{xik}} \quad (4)$$

$$SSE(C) = \sum_{k=1}^K \sum_{x_i \in C_k} w_{xik}^{\beta} \cdot \|x_i - c_k\|^2 \quad (5)$$

## 2.2. Elbow Optimization Method

The elbow method is a heuristic method used in cluster analysis to determine the optimal number of clusters into which a given dataset may be segmented [32]. The elbow technique depicts the value of the cost function, generally the Sum of Squared Errors (SSE), produced by a certain number of clusters and then determines the optimal number of clusters ( $K$ ) by picking the value of  $K$  for which the change in SSE first appears to reduce, thus forming an elbow in the curve, i.e., the point after which the distortion starts decreasing in a linear fashion. As  $K$  increases, the SSE decreases because each cluster has fewer data points that are closer to their respective centroids. The value of  $K$  at which the improvement in distortion decreases the most is known as the elbow of the curve, and it is at this point that splitting the dataset into additional clusters should cease.

## 2.3. Performance Metrics

Certain objective measures must be used to assess the predictive accuracy of a forecasting model, such as MLPs. Mean Absolute Percentage Error (MAPE) and coefficient of determination ( $R^2$  score) are the two most commonly used metrics in the application of neural networks to various regression problems, such as STLF.

In statistics, MAPE is a measure of the predictive accuracy afforded by a forecasting method. Because of its rather obvious definition in terms of relative error, it is often employed as a loss function for regression tasks and model evaluation. MAPE is defined by Equation (6) as follows:

$$\text{MAPE} = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{A_i - F_i}{A_i} \right| \quad (6)$$

where  $n$  is the number of data points,  $A_i$  is the actual value and  $F_i$  is the forecasted value of each data point.

The  $R^2$  score is an important metric for evaluating the performance of a regression-based machine learning model. It is the amount of the variation in the output dependent attribute, which is predictable from the input independent variable. It provides a measure of how well observed outcomes are replicated by the model, based on the proportion of total variation of outcomes explained by the model. An  $R^2$  value of 1 means that the model fits the data perfectly, and a value of 0 means that the model will perform badly on an unseen dataset, i.e., it has a very poor predictive power. This implies that the closer the value of the  $R^2$  score is to 1, the better the model is trained. The  $R^2$  score is calculated from Equation (7) as follows:

$$R^2 = 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \bar{y})^2} \quad (7)$$

where  $y_i$  is the actual output value that is associated with each input instance  $x_i$ ,  $f_i$  is the forecasted value for input instance  $x_i$ , and  $\bar{y}$  is the mean value of the dataset.

#### 2.4. Problem Formulation

This paper presents the development and evaluation of four hybrid STLF models that lead to high accuracy with fast convergence time. The models employ MLPs neural networks and optimized clustering methods, and they were developed and tested using historical hourly load data of the Greek power system from the period 2013–2017, obtained through ENTSO-E platform [33]. Air temperature was included in the input data of the MLPs, in addition to load data, to increase the precision of the prediction. The data of the period 2013–2016 are used for training purposes (a total of 35,040 data points), and the data of the year 2017 are used as the test set to assess the accuracy of the predictions (8,760 entries). The development and implementation of the K-Means and Fuzzy C-Means algorithms for generating the optimum number of clusters with the use of the elbow optimization method is discussed in what follows. In a nutshell, two experiments were conducted: First, each of the two clustering methods, namely K-Means and FCM, are applied to the input data, thus producing a set of clusters. For each cluster produced, a separate MLP is trained to produce STLF predictions, thus yielding two models. In a second experiment the two clustering methods are applied to the input data, thus producing a set of clusters each. Then the resulting labeling values of the data that are generated by the clustering method are fed, along with load and temperature data, back to one MLP for each clustering method, thus resulting in two STLF models with faster convergence and improved accuracy compared to existing approaches. The MLP neural networks, the clustering algorithms, and the elbow optimization method were developed via Python's library scikit-learn [34] and implemented in a computer used with Intel Core i7-4510U CPU and 8 GB installed RAM.

##### 2.4.1. Calculation of Optimum Number of Clusters

Since STLF is inextricably linked to data concerning temperature, humidity, and historical load values, the algorithms are applied to datasets that include weather and load data. The load data are processed in order to form clusters based on the load profile of each data point using the K-Means and FCM techniques. The variables considered for the application of the clustering methods are the load value at the same time on the same day

of the previous week (D-7 Load), the load value of the previous day at the same time (D-1 Load), and the load value at the previous hour load (H-1 Load). The load data used for clustering is pre-processed through the enhanced min-max scaling method, which leads to improved forecasting results, compared to the simple min-max scaling technique [13]. Figure 1 illustrates the process by which clusters are formed for each of the two clustering algorithms used.

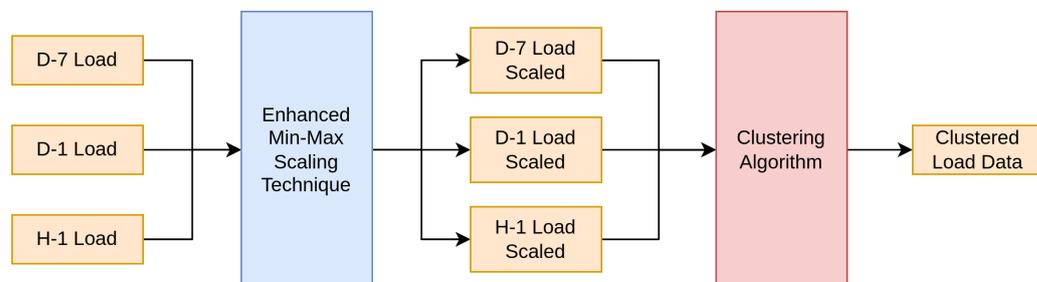


Figure 1. Classification of data, based on their load profile, using clustering techniques.

The optimal number of clusters for the data of the Greek power system is derived by using the Elbow optimization heuristic, which plots the explained variation as a function of the number of clusters by calculating the SSE and picking the elbow of the curve as the number of clusters to use. In order to generate the optimal number of clusters, the SSE is computed for a range of clusters in the interval [1, 10], which is a common procedure. Figure 2 shows the SSE for values of the variable  $K$  in the interval [1, 10]. After extensive experimentation and comparison of forecasting results, using MAPE as a metric, the optimal separation of data into clusters, based on the load profile for the specific dataset, occurs for  $K = 4$ .

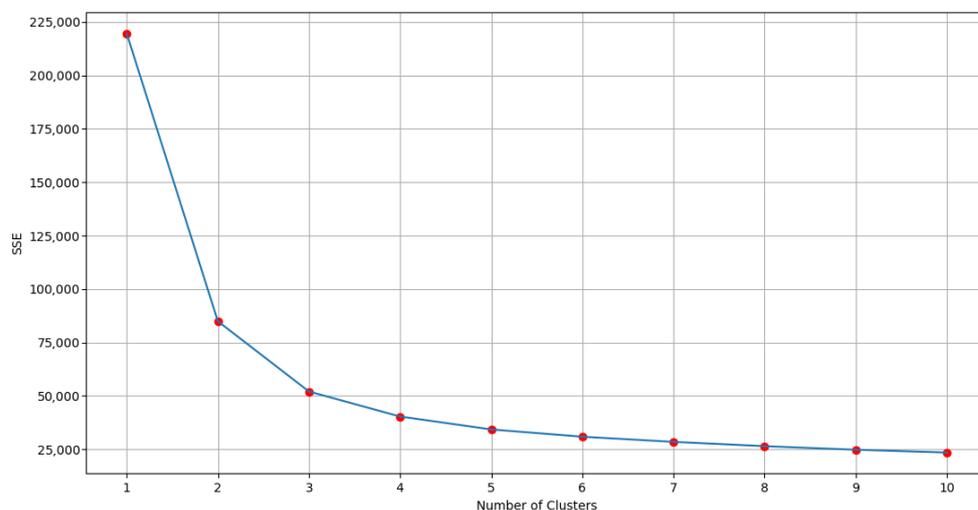
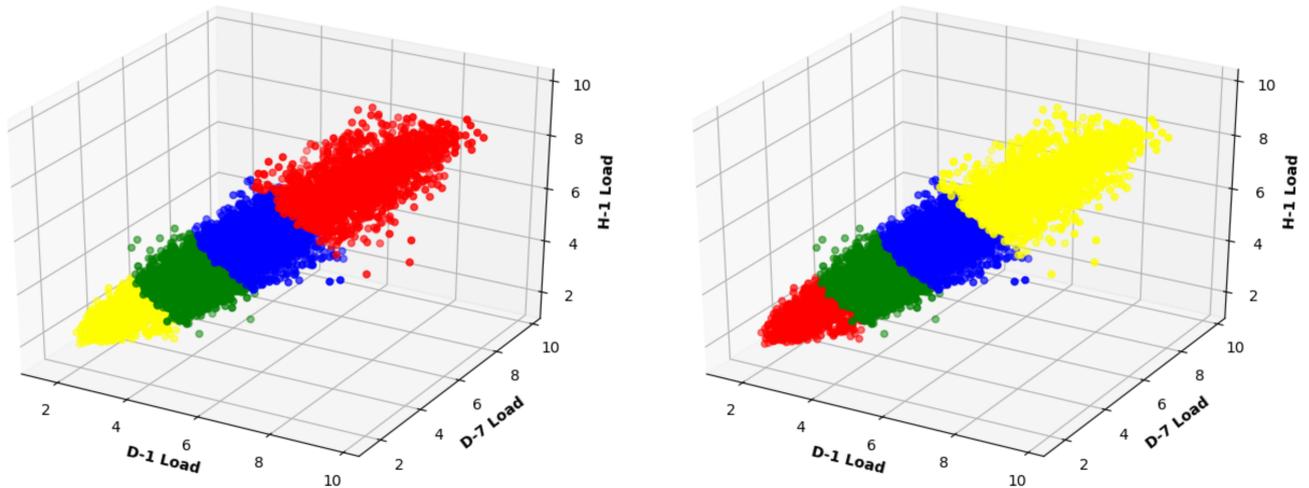


Figure 2. Variation of sum of squared errors for determining the optimum value of parameter  $K$ .

Since clustering is based solely on the three variables, D-7 Load, D-1 Load, and H-1 Load, each data point may be represented as a point in a three-dimensional space with the values of these variables as coordinates. Figure 3 depicts the clusters produced by the K-Means (on the left) and FCM (on the right) clustering algorithms, respectively, of the dataset used.

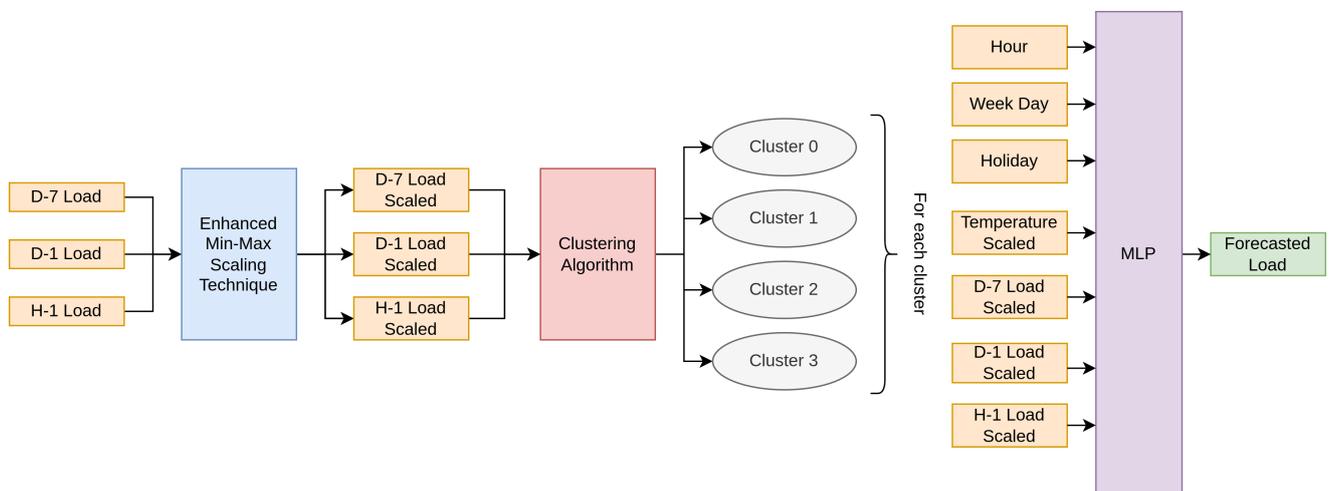


**Figure 3.** Clustered load data in space as a result of the implementation of K-Means (left) and FCM (right) clustering techniques.

2.4.2. Short-Term Load Forecasting Approaches in Conjunction with Clustering Techniques

Clustering algorithms have been used to generate input for MLPs in classical approaches to STLF, where a separate MLP neural network is used for each cluster. In the first experiment, the hybrid models created in this way are schematically shown in Figure 4. The neural network input variables are:

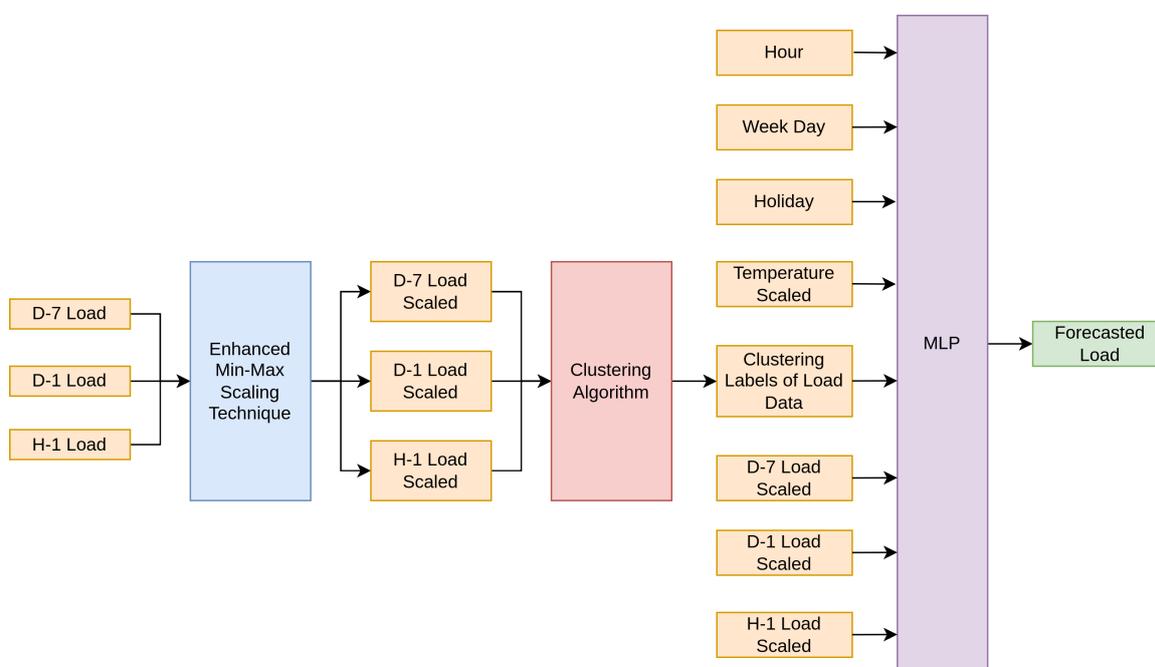
- Hour: Input variable within the range [0, 23] indicating the load forecast’s time of day;
- Week Day: Input variable denoting the day of the week, within the range [1, 7] (1 corresponding to Sunday, and so on);
- Holiday: Binary values are used to indicate whether a day is a holiday (1), which includes Greek state holidays, religious holidays and the weekends, or a normal working day (0);
- Temperature: Input variable indicating the temperature of the hour (in Celsius) for which the load is predicted, scaled by min-max technique;
- D-7 Load: Input variable denoting the corresponding load at the same hour on the same day in the prior week;
- D-1 Load: Input variable denoting the corresponding load at the same hour in the prior day;
- H-1 Load: Input variable denoting the corresponding load in the prior hour.



**Figure 4.** Short-term load forecasting using a MLP for each optimally generated cluster of the dataset.

In the second experiment, a single MLP, fed with an input variable containing the clustering label, produced by either K-Means or Fuzzy C-Means optimized via the elbow method clustering technique, is used for STLF.

The STLF model for this approach is shown in Figure 5. The input variables of the neural network include, in addition to the ones of the first experiment, the labeling of each data point of the load data produced and optimized by each clustering algorithm. The variable *Label for Cluster* receives integer values from 0 to 3, (since the separation of the dataset into four clusters was determined to be optimum) and indicates in which cluster each data point belongs. In total, four STLF models emerged, two from the first experiment which follows classical hybrid approaches, and two from the second, which augments/informs the neural network input with clustering information. The four models were applied to a dataset of the Greek power system and the forecasting results were compared to each other and to other existing load forecasting approaches. Moreover, this comparison indicates which of the clustering algorithms is more appropriate for partitioning a dataset into clusters. Extensive testing and experimentation show that the conjunction of MLPs neural networks with optimized dataset clustering, leads to improvement of the accuracy and the convergence time of the forecasting model.



**Figure 5.** Short-term load forecasting using a single MLP with the optimized clustering labels fed as input.

### 3. Results

This section presents the results obtained from the four STLF models that emerged from the two experiments. MAPE and  $R^2$  score are used as metrics, in order to evaluate the accuracy of the prediction of each model.

Table 1 shows the total MAPE and  $R^2$  score for the predictive method, and for each cluster individually, for the first model, where K-Means clustering followed by separate MLPs for each cluster was used. Table 2 provides the same information for the second model, where Fuzzy C-Means clustering followed by separate MLPs for each cluster was used. Figure 6 provides a graphical comparison of the actual load values and the prediction results obtained from these two models. Both approaches performed well, in terms of MAPE and  $R^2$  score, compared with existing methods.

**Table 1.** MAPE and  $R^2$  score of the MLPs' forecasted values via K-Means implementation.

Cluster	Number of Data	MAPE (%)	$R^2$ Score
Cluster 0	2052	1.66	0.95132
Cluster 1	2330	1.76	0.89331
Cluster 2	2955	1.67	0.88591
Cluster 3	1423	1.65	0.88591
Total	8760	1.69	0.98643

**Table 2.** MAPE and  $R^2$  score of the MLPs' forecasted values via Fuzzy C-Means implementation.

Cluster	Number of Data	MAPE (%)	$R^2$ Score
Cluster 0	1423	1.74	0.87951
Cluster 1	2128	1.66	0.95198
Cluster 2	2308	1.75	0.89383
Cluster 3	2901	1.70	0.88200
Total	8760	1.71	0.98632

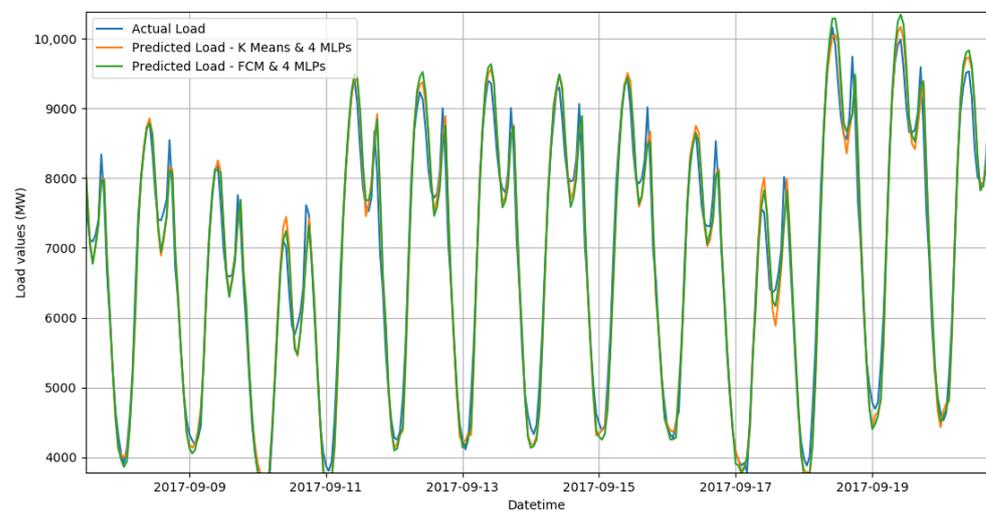
**Figure 6.** Actual and predicted load curves resulting from the method using a distinct MLP for each cluster.

Table 3 presents the MAPE and the  $R^2$  score of the third and fourth model, where the MLP input is informed with labeling information acquired from the application of K-Means and FCM clustering algorithms, respectively. Figure 7 provides a graphical comparison between real load values and the forecasted load values, indicatively for some days in February 2017, calculated using the third and fourth model.

**Table 3.** MAPE and  $R^2$  score of the forecasted values using clustering labels as input variables to MLP.

Approach	MAPE (%)	$R^2$ Score
MLPs and K-Means—Labels are fed as input to MLP	1.77	0.98583
MLPs and Fuzzy C-Means—Labels are fed as input to MLP	1.70	0.98678

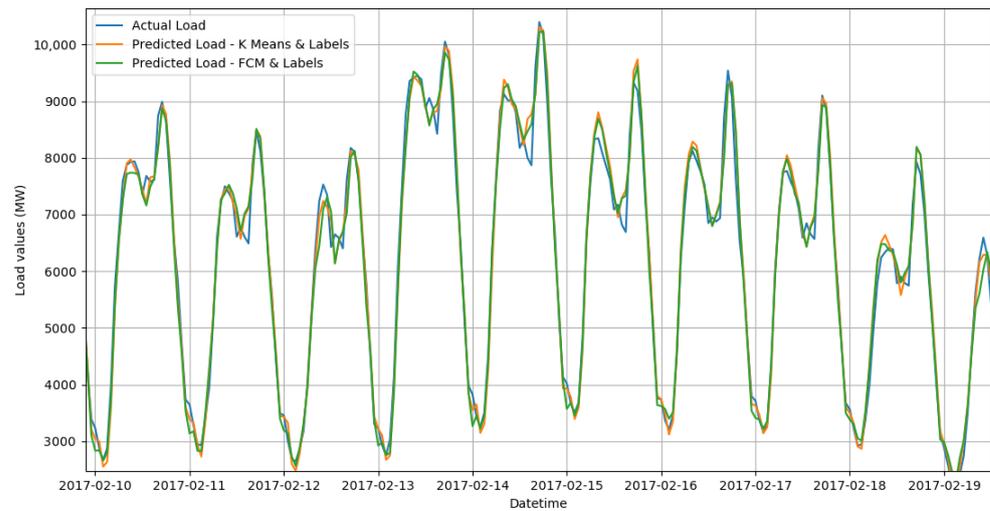


Figure 7. Actual and predicted load curves by MLPs with the use of K-Means and FCM.

Figure 8 focuses on the K-Means clustering method and graphically compares the results obtained from the first and third model. A similar graphical comparison of the results obtained with the use of Fuzzy C-Means clustering method, which is from the second and fourth models, is presented in Figure 9.

Apart from the MAPE and  $R^2$  score, the performance for STLF using MLPs in conjunction with K-Means and Fuzzy C-Means, is also evaluated by measuring the execution time required for each approach. Table 4 provides the time (in seconds) needed for the load forecasting of the year 2017 in all four models.

Table 4. Execution time for the four proposed forecasting models.

Approach	Time (s)
MLPs and K-Means—Individual MLP for each cluster	1690.27835
MLPs and Fuzzy C-Means—Individual MLP for each cluster	1353.51278
MLPs and K-Means—Labels are fed as input to MLP	1223.64124
MLPs and Fuzzy C-Means—Labels are fed as input to MLP	808.320161

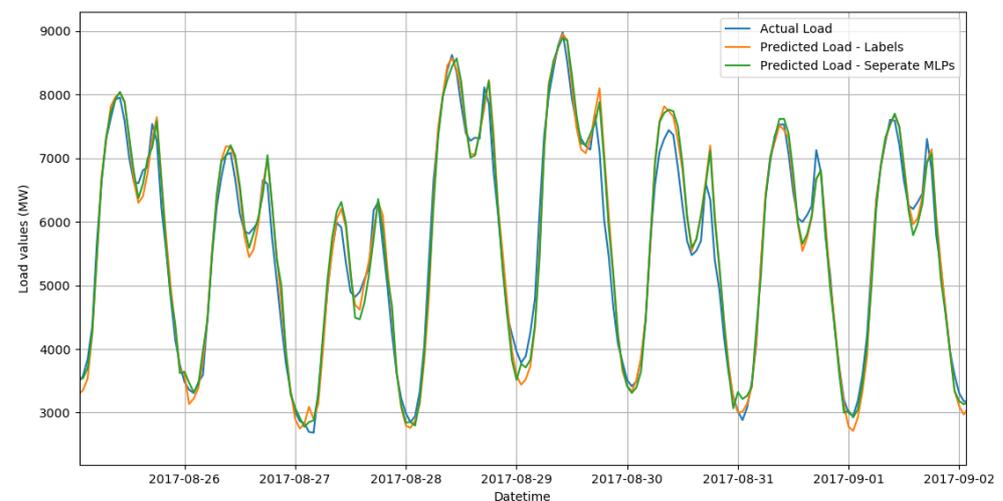
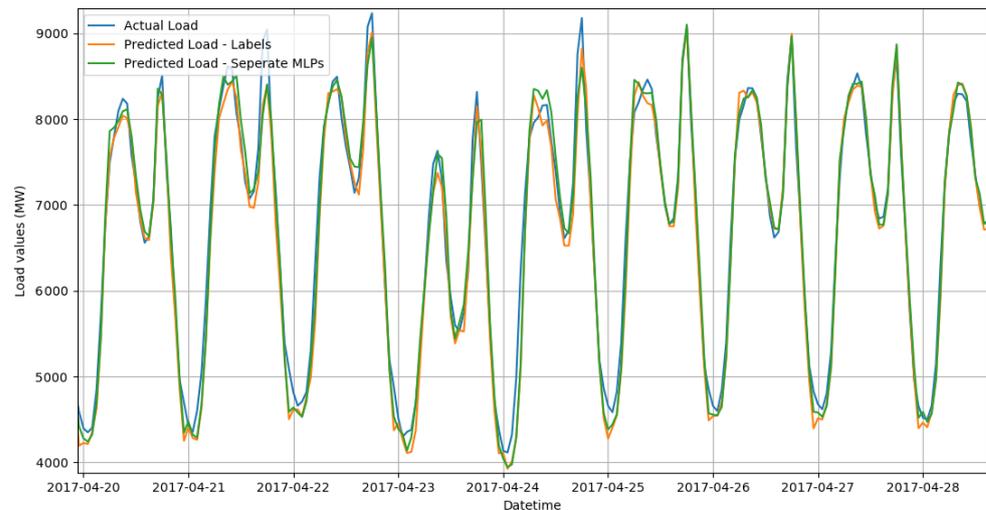


Figure 8. Actual and predicted load curves resulting from the two methods using K-Means.



**Figure 9.** Actual and predicted load curves resulting from the two methods using FCM.

#### 4. Discussion

The use of a clustering algorithm, which properly groups the data based on their load profile, clearly improves the accuracy of STLF results, as acknowledged by several related works in this area. The current best MAPE obtained is around 2%, although it should be noted that different datasets from different power systems are used. MAPE is equal to 1.80% in [13], which uses the same load data from the Greek power system as this work, and as in Tables 1 and 2, which demonstrate that the first set of models that we developed are more accurate with a better MAPE value, for both clustering methods employed.

The results presented in Table 3 demonstrate that for the second set of models, where a single MLP is employed, informed explicitly with the clustering labels of the input data points, both K-Means and FCM improve the load prediction compared to [13], and the FCM specifically has the best overall accuracy. However, both models yield slightly lower accuracy than their counterparts from the first set, but converge faster than them. In fact, Table 4 demonstrates that the fourth model using FCM performs remarkably better than the others.

A comparison of the results obtained from all four proposed models with similar STLF methods, which use neural network prediction techniques in conjunction with the application of a clustering algorithm, reveals that the methods described in this paper perform in most of the cases similarly or better. However, note that an exact comparison requires comparison on exactly the same dataset. In [21–24,27], who use similar techniques for short-term load forecasting, the MAPE gets values close to 2%, while in the present work the lowest MAPE is equal to 1.69%. Table 5 presents the results in terms of the achieved MAPE of various techniques suggested by other researchers considered in the related literature review. It is obvious that the models proposed here lead to improved MAPE and therefore greater prediction accuracy for STLF.

**Table 5.** MAPE for various forecasting techniques examined in the literature.

Approach	Proposed by	MAPE (%)
SOM—K-Means—MLP	Hernandez et al. [21]	3.18
K-Means—Stacked Denoising Autoencoders - ANNs	Farfar et al. [22]	1.85
Sparsified K-Means—ANN	Seon-Ju Ahn et al. [24]	2.06
K-Means—SVM	Xishuang Dong et al. [23]	2.92
K-Means—MLP	Xishuang Dong et al. [23]	3.12
K-Means—CNN	Xishuang Dong et al. [23]	3.06
K-Means—FCM—MLP	Bian Haihong et al. [27]	2.15
Enhanced STLF via MLPs	Arvanitidis et al. [13]	1.80
MLPs and K-Means—Individual MLP for each cluster	Proposed algorithm	1.69
MLPs and Fuzzy C-Means—Individual MLP for each cluster	Proposed algorithm	1.71
MLP and K-Means—Labels are fed as input to MLP	Proposed algorithm	1.77
MLP and Fuzzy C-Means—Labels are fed as input to MLP	Proposed algorithm	1.70

## 5. Conclusions

This paper examines the integration of clustering algorithms with neural networks for the purposes of developing fast and accurate STLF models. Two ways in which such integration can be implemented were considered, and as a result two sets of models were designed, tested, and evaluated on the same dataset. The first set of models followed the standard for hybrid STLF model development, in which first the dataset is clustered and then each cluster is used to train a MLP. Since we experimented with two clustering algorithms, namely K-Means and Fuzzy C-Means, this first set produced two models, which were used as a reference point. These first two models do present an improvement on the current best score in the relevant literature, because the dataset is initially subjected to enhanced scaling, which has been evaluated in a separate paper [13].

The second way in which clustering algorithms can be integrated with neural networks is explored in the second set of models that were developed. In this case, first the dataset is clustered (using K-Means and Fuzzy C-Means, again), and then a single MLP is trained, whose input variables are augmented with the inclusion of the labeling information produced by the clustering.

All four models were evaluated using load data of the Greek power system as a common reference point. All models yielded better accuracy than other methods (as reflected by MAPE values below 2%). Moreover, the models of the second set, where the MLP is informed by clustering, converged significantly faster. The experiments suggest that the FCM informed MLP is the fastest model, however, to be precise, it needs to be evaluated on other datasets as well, and this is one direction for future work. A second direction for future work involves experimenting with other clustering algorithms to establish whether they might offer even better accuracy and convergence time.

**Author Contributions:** Conceptualization, A.I.A.; methodology, A.I.A.; software, A.I.A.; validation, A.I.A., D.B. and A.D.; formal analysis, A.I.A.; investigation, A.I.A.; resources, A.I.A., D.K. and I.P.P.; data curation, A.I.A., D.K. and I.P.P.; writing—original draft preparation, A.I.A.; writing—review and editing, A.I.A., D.B., A.D. and L.H.T.; visualization, A.I.A., D.K. and I.P.P.; supervision, D.B., A.D. and L.H.T.; project administration, D.B. and A.D. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data are available in a publicly accessible repository. The data used in this study are openly available from the ENTSO-E portal in <https://open-power-system-data.org/data-sources> (accessed on 18 November 2021). The dataset was processed as the input for the

design and performance assessment of the clustering algorithms and the multi-layer perceptron neural network described in this article.

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

## References

1. Kalakova, A.; Nunna, H.S.S.K.; Jamwal, P.K.; Doolla, S. Genetic Algorithm for Dynamic Economic Dispatch with Short-Term Load Forecasting. In Proceedings of the 2019 IEEE Industry Applications Society Annual Meeting, Baltimore, MD, USA, 29 September–3 October 2019; pp. 1–6. [\[CrossRef\]](#)
2. Laitsos, V.M.; Bargiotas, D.; Daskalopulu, A.; Arvanitidis, A.I.; Tsoukalas, L.H. An Incentive-Based Implementation of Demand Side Management in Power Systems. *Energies* **2021**, *14*, 7994. [\[CrossRef\]](#)
3. Alamaniotis, M.; Ikononopoulos, A.; Alamaniotis, A.; Bargiotas, D.; Tsoukalas, L.H. Day-ahead electricity price forecasting using optimized multiple-regression of relevance vector machines. In Proceedings of the 8th Mediterranean Conference on Power Generation, Transmission, Distribution and Energy Conversion (MEDPOWER 2012), Cagliari, Italy, 1–3 October 2012; pp. 1–5. [\[CrossRef\]](#)
4. Alamaniotis, M.; Bargiotas, D.; Bourbakis, N.G.; Tsoukalas, L.H. Genetic Optimal Regression of Relevance Vector Machines for Electricity Pricing Signal Forecasting in Smart Grids. *IEEE Trans. Smart Grid* **2015**, *6*, 2997–3005. [\[CrossRef\]](#)
5. Alamaniotis, M.; Tsoukalas, L.H. Implementing smart energy systems: Integrating load and price forecasting for single parameter based demand response. In Proceedings of the 2016 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe), Ljubljana, Slovenia, 9–12 October 2016; pp. 1–6. [\[CrossRef\]](#)
6. Kyriakides, E.; Polycarpou, M. Short Term Electric Load Forecasting: A Tutorial. In *Trends in Neural Computation*; Chen, K., Wang, L., Eds.; Springer: Berlin/Heidelberg, Germany, 2007; pp. 391–418. [\[CrossRef\]](#)
7. Alamaniotis, M.; Bargiotas, D.; Tsoukalas, L.H. Towards Smart Energy Systems: Application of Kernel Machine Regression for Medium Term Electricity Load Forecasting. *SpringerPlus* **2016**, *5*, 58. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Kontogiannis, D.; Bargiotas, D.; Daskalopulu, A. Fuzzy Control System for Smart Energy Management in Residential Buildings Based on Environmental Data. *Energies* **2021**, *14*, 752. [\[CrossRef\]](#)
9. Kontogiannis, D.; Bargiotas, D.; Daskalopulu, A.; Tsoukalas, L.H. A Meta-Modeling Power Consumption Forecasting Approach Combining Client Similarity and Causality. *Energies* **2021**, *14*, 88. [\[CrossRef\]](#)
10. Hua, Y.; Wang, N.; Zhao, K. Simultaneous Unknown Input and State Estimation for the Linear System with a Rank-Deficient Distribution Matrix. *Math. Probl. Eng.* **2021**, *2021*, 1–11. [\[CrossRef\]](#)
11. Liu, C.; Li, Q.; Wang, K. State-of-charge estimation and remaining useful life prediction of supercapacitors. *Renew. Sustain. Energy Rev.* **2021**, *150*, 111408. [\[CrossRef\]](#)
12. Kontogiannis, D.; Bargiotas, D.; Daskalopulu, A. Minutely Active Power Forecasting Models Using Neural Networks. *Sustainability* **2020**, *12*, 3177. [\[CrossRef\]](#)
13. Arvanitidis, A.I.; Bargiotas, D.; Daskalopulu, A.; Laitsos, V.M.; Tsoukalas, L.H. Enhanced Short-Term Load Forecasting Using Artificial Neural Networks. *Energies* **2021**, *14*, 7788. [\[CrossRef\]](#)
14. Zhang, T.; Liu, D.; Yue, D. Rough Neuron Based RBF Neural Networks for Short-Term Load Forecasting. In Proceedings of the 2017 IEEE International Conference on Energy Internet (ICEI), Beijing, China, 17–21 April 2017; pp. 291–295. [\[CrossRef\]](#)
15. Sadaei, H.J.; de Lima e Silva, P.C.; Guimarães, F.G.; Lee, M.H. Short-term load forecasting by using a combined method of convolutional neural networks and fuzzy time series. *Energy* **2019**, *175*, 365–377. [\[CrossRef\]](#)
16. Liu, C.; Zhang, Y.; Sun, J.; Cui, Z.; Wang, K. Stacked bidirectional LSTM RNN to evaluate the remaining useful life of supercapacitor. *Int. J. Energy Res.* **2021**, 1–10. [\[CrossRef\]](#)
17. Yang, A.; Li, W.; Yang, X. Short-term electricity load forecasting based on feature selection and Least Squares Support Vector Machines. *Knowl.-Based Syst.* **2019**, *163*, 159–173. [\[CrossRef\]](#)
18. Moon, J.; Kim, Y.; Son, M.; Hwang, E. Hybrid Short-Term Load Forecasting Scheme Using Random Forest and Multilayer Perceptron. *Energies* **2018**, *11*, 3283. [\[CrossRef\]](#)
19. Almalaq, A.; Edwards, G. A Review of Deep Learning Methods Applied on Load Forecasting. In Proceedings of the 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA), Cancun, Mexico, 18–21 December 2017; pp. 511–516. [\[CrossRef\]](#)
20. Koo, B.g.; Lee, S.W.; Kim, W.; Park, J.H. Comparative Study of Short-Term Electric Load Forecasting. In Proceedings of the 2014 5th International Conference on Intelligent Systems, Modelling and Simulation, Langkawi, Malaysia, 27–29 January 2014; pp. 463–467. [\[CrossRef\]](#)
21. Hernández, L.; Baladrón, C.; Aguiar, J.M.; Carro, B.; Sánchez-Esguevillas, A.; Lloret, J. Artificial neural networks for short-term load forecasting in microgrids environment. *Energy* **2014**, *75*, 252–264. [\[CrossRef\]](#)
22. Farfar, K.E.; Khadir, M. A two-stage short-term load forecasting approach using temperature daily profiles estimation. *Neural Comput. Appl.* **2019**, *31*. [\[CrossRef\]](#)
23. Dong, X.; Qian, L.; Huang, L. Short-term load forecasting in smart grid: A combined CNN and K-means clustering approach. In Proceedings of the 2017 IEEE International Conference on Big Data and Smart Computing (BigComp), Jeju Island, Korea, 13–16 February 2017; pp. 119–125. [\[CrossRef\]](#)

24. Ngo, M.D.; Yun, S.Y.; Choi, J.H.; Ahn, S.J. Short-Term Load Forecasting of Buildings Based on Artificial Neural Network and Clustering Technique. *J. IKEEE* **2018**, *22*, 672–679.
25. Li, Z.; Bao, S.; Gao, Z. Short Term Prediction of Photovoltaic Power Based on FCM and CG-DBN Combination. *J. Electr. Eng. Technol.* **2019**, *15*, 333–341. [[CrossRef](#)]
26. Yang, M.; Shi, C.; Liu, H. Day-ahead wind power forecasting based on the clustering of equivalent power curves. *Energy* **2021**, *218*, 119515. [[CrossRef](#)]
27. Bian, H.; Zhong, Y.; Sun, J.; Shi, F. Study on power consumption load forecast based on K-means clustering and FCM-BP model. *Energy Rep.* **2020**, *6*, 693–700. [[CrossRef](#)]
28. Lu, Y.; Zhang, T.; Zeng, Z.; Loo, J. An improved RBF neural network for short-term load forecast in smart grids. In Proceedings of the 2016 IEEE International Conference on Communication Systems (ICCS), Shenzhen, China, 14–16 December 2016; pp. 1–6. [[CrossRef](#)]
29. Kuan, L.; Yan, Z.; Xin, W.; Yan, C.; Xiangkun, P.; Wenxue, S.; Zhe, J.; Yong, Z.; Nan, X.; Xin, Z. Short-term electricity load forecasting method based on multilayered self-normalizing GRU network. In Proceedings of the 2017 IEEE Conference on Energy Internet and Energy System Integration (EI2), Beijing, China, 26–28 November 2017; pp. 1–5. [[CrossRef](#)]
30. Jin, X.; Han, J., K-Means Clustering. In *Encyclopedia of Machine Learning*; Sammut, C., Webb, G.I., Eds.; Springer: Boston, MA, USA, 2010; pp. 563–564. [[CrossRef](#)]
31. Bezdek, J.C.; Ehrlich, R.; Full, W. FCM: The fuzzy c-means clustering algorithm. *Comput. Geosci.* **1984**, *10*, 191–203. [[CrossRef](#)]
32. Bholowalia, P.; Kumar, A. EBK-means: A clustering technique based on elbow method and k-means in WSN. *Int. J. Comput. Appl.* **2014**, *105*, 17–24.
33. Hirth, L.; Mühlenpfordt, J.; Bulkeley, M. The ENTSO-E Transparency Platform—A review of Europe’s most ambitious electricity data platform. *Appl. Energy* **2018**, *225*, 1054–1067. [[CrossRef](#)]
34. 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.