

Article

Energy Use Forecasting with the Use of a Nested Structure Based on Fuzzy Cognitive Maps and Artificial Neural Networks

Katarzyna Poczeta ^{1,*}  and Elpiniki I. Papageorgiou ² 

¹ Department of Applied Computer Science, Kielce University of Technology, 25314 Kielce, Poland

² Department of Energy Systems, Faculty of Technology, University of Thessaly, Geopolis, 41500 Larisa, Greece

* Correspondence: k.piotrowska@tu.kielce.pl

Abstract: The aim of this paper is to present a novel approach to energy use forecasting. We propose a nested fuzzy cognitive map in which each concept at a higher level can be decomposed into another fuzzy cognitive map, multilayer perceptron artificial neural network or long short-term memory network. Historical data related to energy consumption are used to construct a nested fuzzy cognitive map in order to better understand energy use behavior. Through the experiments, the usefulness of the nested structure in energy demand prediction is demonstrated, by calculating three popular metrics: Mean Square Error, Mean Absolute Error and the correlation coefficient. A comparative analysis is performed, applying classic multilayer perceptron artificial neural networks, long short-term memory networks and fuzzy cognitive maps. The results confirmed that the proposed approach outperforms the classic methods in terms of prediction accuracy. Moreover, the advantage of the proposed approach is the ability to present complex time series in the form of a clear nested structure presenting the main concepts influencing energy consumption on the first level. The second level allows for more detailed problem analysis and lower forecast errors.

Keywords: nested structure; energy use forecasting; fuzzy cognitive maps; artificial neural networks; long short-term memory networks



Citation: Poczeta, K.; Papageorgiou, E.I. Energy Use Forecasting with the Use of a Nested Structure Based on Fuzzy Cognitive Maps and Artificial Neural Networks. *Energies* **2022**, *15*, 7542. <https://doi.org/10.3390/en15207542>

Academic Editors: David Borge-Diez and Seung-Hoon Yoo

Received: 31 August 2022

Accepted: 10 October 2022

Published: 13 October 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

An important element of policy around the world and strategy formulation is energy use analysis. The aim of this paper is to introduce a novel approach for constructing a nested structure in which each concept at a higher level can be decomposed into another fuzzy cognitive map, multilayer perceptron artificial neural network or long short-term memory network. The proposed nested FCM allows for very good forecasting accuracy thanks to the use of effective models on the second level and at the same time can facilitate the understanding of the energy use behavior thanks to the general fuzzy cognitive map on the first level.

In recent years, many different techniques of machine learning and deep learning have been explored in the literature to solve the problem of energy use forecasting [1–3]. In [4], a forecasting system based on the support vector regression model and Markov Chain was developed in order to discover energy consumption patterns in China. The presented results show that the proposed approach is more accurate than the moving average model and the grey model. In [5], an Autoregressive Integrated Moving Average model was used to analyze the demand forecasting electricity based on real data from the World Bank on energy consumption in Morocco. The results indicated an upward trend in electrical energy consumption by the end of 2030. Forecasting energy use provides important information for energy suppliers and customers to improve energy management and load control.

In [6], the management of energy consumption in warehouse buildings was analyzed. Support Vector Regression, Random Forest, Extreme Gradient Boosting, Recurrent Neural Networks, Long Short-Term Memory network (LSTM), Gated Recurrent Unit, and Autoregressive Integrated Moving Average were compared in the problem of predicting daily

energy consumption. The results show that the Extreme Gradient Boosting models outperform all other machine learning and deep learning models for short-term load forecasting. Forecasting electricity consumption in Thailand with the use of multiple linear regression, artificial neural network (ANN), support vector machine, hybrid models, and ensemble models were implemented in [7]. The results show that a hybrid model of a multiple regression model and artificial neural networks provided the best forecasting accuracy.

In [8], the Long Short-Term Memory, the Gated Recurrent Unit and the Drop-Gated Recurrent Unit were used to predict power consumption in some French cities. In [9], a deep learning approach based on recurrent neural networks and Long Short-Term Memory networks was developed to forecast the photovoltaic output power. The proposed technique was compared with well-known regression, hybrid ANFIS and machine learning methods achieving higher forecasting accuracy. In [10], artificial neural networks were applied to short-term load forecasting. The enhanced min–max scaling procedure was proposed where the importance of certain input variables on the total outcome of the artificial neural network was taken into consideration. The results show that using some data preprocessing techniques can lead to improved prediction accuracy.

Various factors can affect energy consumption around the world. Most of the models used in forecasting energy consumption focus on the accuracy of the prediction. In black-box models, such artificial neural networks are less interpretable in terms of the form of the relationship between inputs and output variables. They can outperform other techniques, e.g., the decision trees, predicting energy demand with lower error, but do not provide feature importance insight [11]. Recently, there has been a growing interest in explainable machine learning. In [12], explainable machine learning was applied in the transportation energy field. The methodology was implemented based on the Household Travel Survey data and artificial neural network. The importance and effect of the available inputs on transportation energy consumption were analyzed.

Understanding and changing energy use behavior can protect environmental resources. This raises the need to develop new techniques that will allow for understanding the problem and facilitate decision-making based on the analysis of available data. A fuzzy cognitive map (FCM) is a soft computing technique that allows mimicking the analyzed problem in the form of important concepts and causal connections between them [13]. In [14], an FCM model was used to analyze the dynamic behavior of concepts and the relations among them in renewable energy systems. In [15], a novel technique based on fuzzy cognitive maps was applied for solar energy forecasting. Table 1 shows the relevant literature summary.

However, a too large amount of data attributes can complicate analysis. In the case of a large number of concepts, the problem can be described in the form of a nested fuzzy cognitive map that allows a more readable representation of knowledge than classic fuzzy cognitive maps. In [16], an FCM model containing a large number of concepts was simplified by merging related or similar concepts into the same cluster allowing a more understandable view of the problem. In [17], we have analyzed a nested fuzzy cognitive map consisting in decomposing each concept at a higher map level into another FCM.

Artificial neural networks can achieve higher accuracy in predicting data than classic fuzzy cognitive maps [18]. The motivation of this paper is to extend the approach for constructing a nested structure presented in [17] to include the possibility of using artificial neural networks at the second level for more accurate forecasting. Each concept of the nested fuzzy cognitive map can be decomposed into another fuzzy cognitive map, multi-layer perceptron artificial neural network or long short-term memory network. Historical data related to hourly energy demand, generation, prices and weather are used to evaluate the effectiveness of the proposed approach [19]. A comparative analysis is performed, applying classic artificial neural networks, long short-term memory networks and fuzzy cognitive maps.

Table 1. Selected methods used for energy use forecasting.

Analyzed Approach	Problem	Results	Reference
Support vector regression model and Markov Chain	Forecasting energy consumption	Outperforms moving average model and the grey model	[4]
Extreme Gradient Boosting models	Predicting daily energy consumption in warehouse buildings	Outperform Recurrent Neural Networks and Autoregressive Integrated Moving Average	[6]
Long Short-Term Memory networks	Predicting the photovoltaic output power	Outperform other well-known regression, hybrid ANFIS and machine learning methods	[9]
Artificial neural networks	Predicting energy demand	Outperform other techniques but do not provide feature importance insight	[11]
Randomized-based fuzzy cognitive map	Solar energy forecasting-univariate time series	Outperforms Probabilistic Weighted Fuzzy Time Series and Seasonal Autoregressive Integrated Moving Average models	[15]

The objectives of the paper are briefly summarized as follows:

- To develop a methodology for the construction of nested structures based on fuzzy cognitive maps and artificial neural networks in order to better understand energy use behavior and extract useful knowledge.
- To apply the developed approach for energy use forecasting.
- To evaluate the performance of the nested structure on the first and the second level and finally choose the most accurate model.
- To compare the proposed approach with classic artificial neural networks, long short-term memory networks and fuzzy cognitive maps.

The paper is organized as follows. Section 2 describes the proposed nested structure and the used dataset and methodologies: fuzzy cognitive maps, multilayer perceptrons and long short-term memory networks. The results are presented in Section 3. Section 4 contains the discussion of results and further research directions.

2. Materials and Methods

This section details the proposed nested structure as well as the materials and methods used in this research study.

2.1. Nested Structure

A nested fuzzy cognitive map in which each concept of a higher level can be decomposed into another fuzzy cognitive map that has been proposed in [17]. In the current research, the approach has been extended to include the possibility of using neural networks at the second level for more accurate forecasting. Each concept of the nested fuzzy cognitive map can be decomposed into another fuzzy cognitive map, multilayer perceptron artificial neural network or long short-term memory network. Figure 1 visualizes a sample nested structure.

The nested structure may have two or more levels depending on the complexity of the analyzed problem. In this paper, a two-level nested structure is used. The construction of a nested structure consists of three main steps described below: data preprocessing, constructing the first level in the form of a general fuzzy cognitive map, and constructing the second level based on fuzzy cognitive maps, multilayer perceptrons and long short-term memory networks. Figure 2 shows the architecture of the proposed algorithm for constructing a nested structure.

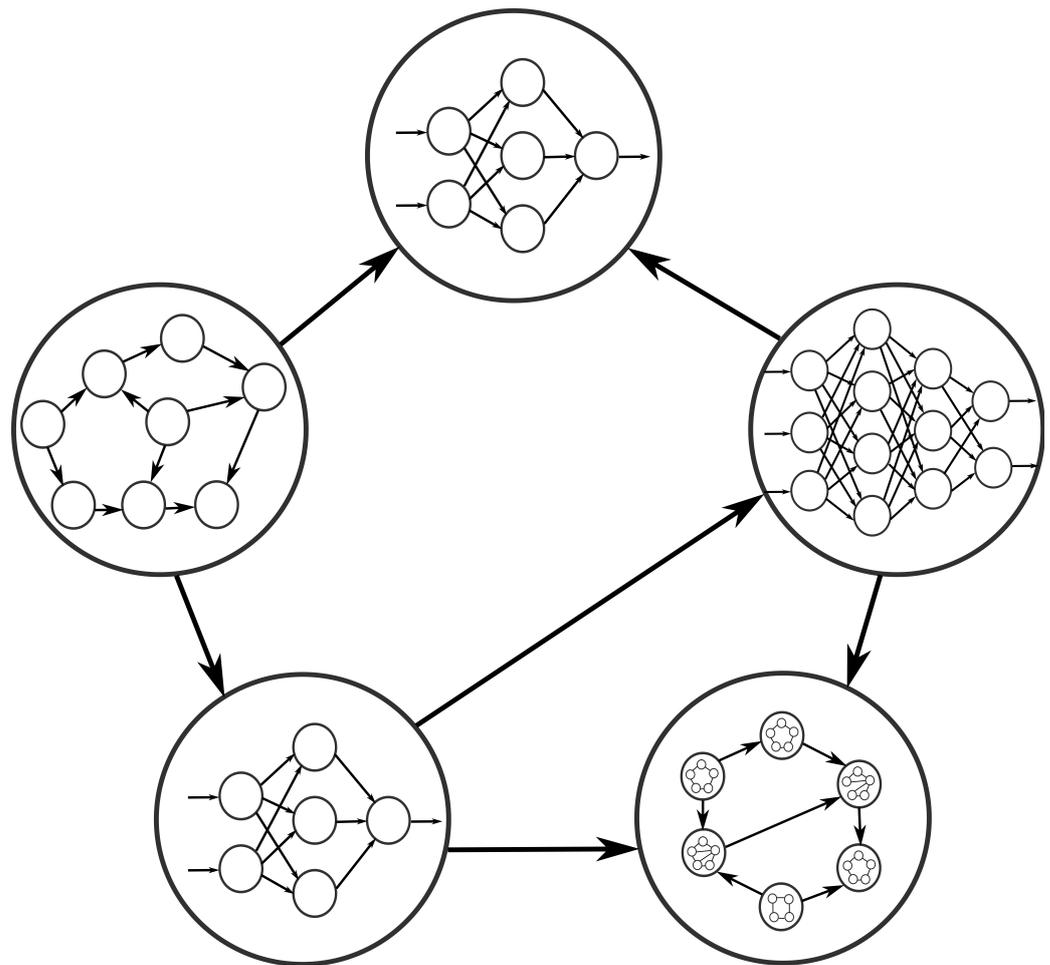


Figure 1. Example of a nested structure.

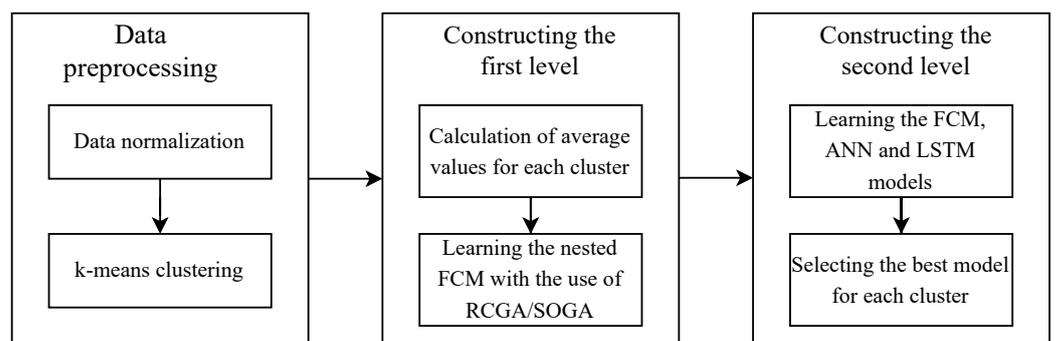


Figure 2. Architecture of the proposed approach.

Data preprocessing.

The first step of the proposed approach is complex data preprocessing. The available data is normalized in the range [0,1] with the standard min-max normalization, described by the formula [20]:

$$x_{new} = \frac{x_{old} - \min_A}{\max_A - \min_A} \tag{1}$$

where \min_A is the minimum value of an attribute A, \max_A is the maximum value of an attribute A, x_{old} is the value of an attribute A, and x_{new} is the normalized value. In our work, we performed data normalization in python based on MinMaxScaler from the sklearn.preprocessing package [21,22].

Then, the normalized data is clustered using the k-means method [23]. The k-means process involves partitioning an N -dimensional population into K disjoint groups on the basis of a sample. This method starts with K clusters, each containing a random point (centroid). Each new point (describing data attribute) is added to the cluster whose centroid is closest to the new point. The K-means algorithm aims to choose centroids that minimize the potential function [24]:

$$\Phi = \sum_{x \in X} \min_{c \in C} \|x - c\|^2, \quad (2)$$

where X is a set of data points, x is a single point, C is a set of centroids, and c is a single centroid.

When the points are added, the centroids are calculated based on the average of all points assigned to each cluster. The algorithm repeats until the centroids stop changing. The k-means method is a popular and fast clustering technique [24] and allows us to group similar data attributes. All data attributes are grouped apart from decision-making elements (the output data attributes that we want to forecast). We implemented data clustering in python with the use of K-means from the sklearn.cluster module [21,25].

Constructing the first level

The next step is the construction of the first level of the nested structure. This is the most general level in the form of a fuzzy cognitive map in which concepts are determined based on the clusters and the output (decision) data attributes. The input data for this FCM model are calculated as the average values for data within each cluster:

$$X_k(t) = \frac{1}{n_k} \sum_{x \in C_k} x(t), \quad (3)$$

where $t = 1, \dots, T$, T is the number of records, C_k is a set of data points belonging to the k -th cluster, x is a single point, $k = 1, \dots, K$, K is the number of clusters, n_k is the number of points in the k -th cluster.

The connections between the concepts at the first level of the nested structure are determined with the use of the learning process and the input data. If we have some expert knowledge about possible connections between concepts or about the weights of the connections, we can also include it in the learning process. In our analysis, we have used two genetic algorithms: RCGA and SOGA to determine the connections between the concepts.

In this paper, we focus on a case in which we do not have expert knowledge but have a large amount of data about the problem. Therefore, the calculation of the input data in each cluster is based on the mean value. Where expert knowledge is available, the input data could be calculated on the basis of a weighted average, where the weights for individual data attributes would be determined by experts. Another approach that could be explored in future research is the use of more advanced data processing techniques, for example, grey numbers.

Constructing the second level

In the last step, the more detailed level of the nested structure is determined. Each concept of a higher level, if it is necessary, can be decomposed into a fuzzy cognitive map, multilayer perceptron or long short-term memory network. All models are assessed and, finally, the most accurate model is chosen. A separate model is constructed for each cluster and can be used to predict energy use. Each model of the second level makes a prediction on the basis of all data attributes belonging to its cluster and the average data from other clusters.

2.2. Fuzzy Cognitive Maps

A fuzzy cognitive map is a graph structure for representing causal reasoning. Nodes are concepts important for the analyzed problem. Edges are causal relationships [13].

Values of the concepts can be calculated on the basis of the selected dynamic model [26]. In the analysis, a nonlinear model is used:

$$X_i(t+1) = f(X_i(t) + \sum_{j=1, j \neq i}^n w_{j,i} X_j(t)), \quad (4)$$

where $X_i(t)$ is the value of the i -th concept, t is the number of iteration, $w_{j,i}$ is the weight of the relationships between the j -th concept and the i -th concept, $i, j = 1, \dots, n$, n is the number of concepts, f is a normalization function. There are four functions that can be used to normalize the values of the concepts [26]:

- sigmoid function;
- hyperbolic tangent function;
- step function;
- threshold linear function.

The selection of the function depends on the problem. In this work, the most popular sigmoid function is used and described as follows:

$$f(x) = \frac{1}{1 + e^{-\lambda x}} \quad (5)$$

where λ is a constant value that indicates the function slope.

Fuzzy cognitive maps can be learned with the use of evolutionary algorithms: the Real-Coded Genetic Algorithm (RCGA) [27] and the Structure Optimization Genetic Algorithm (SOGA) [28]. The learning process enables finding important relationships between concepts at every nesting level and determining the weights of these relationships on the basis of the available data.

2.3. Artificial Neural Networks

The structure of an artificial neural network that combines single neurons into a network and is proposed [29]. One of the most well-known artificial neural networks is a multilayer perceptron. It is composed of multiple layers of perceptrons. A multilayer perceptron consists of an input layer, hidden layers, and an output layer [30]. The neurons in the layers are linked by synaptic weights. These weights can be determined with the use of the learning process.

In our analysis, multilayer perceptrons are learned based on the backpropagation method with momentum [30]. The sigmoid function (5) is used as the activation function. The process of learning fuzzy cognitive maps and multilayer perceptrons is carried out using the previously developed ISEMK application [31].

2.4. Long Short-Term Memory Networks

A Long Short-Term Memory network is a recurrent network with memory cell structures developed to overcome the error-back flow problems [32]. Like a multilayer perceptron, an LSTM model consists of an input layer, hidden layers, and an output layer.

LSTM models can give more accurate time series forecasts compared to traditional recurrent neural networks or multilayer perceptrons. In [33], Long Short-Term Memory networks were analyzed based on environmental time-series problems: forecasting water pollution, air pollution, and ozone alarm. The presented results confirm that LSTM shows the most accurate results compared to multilayer perceptron and traditional recurrent neural networks. The LSTM shows the best accuracy in almost all experiments.

The LSTM network is implemented with the use of Keras API and python language [34]. Two effective models are used in the analysis: a Vanilla LSTM and a Bidirectional LSTM. A Vanilla LSTM is a network with a single hidden layer and an output layer used to make a prediction. Bidirectional LSTMs are an extension of traditional LSTMs that allow learning the input sequence both forward and backwards. The Adam algorithm

is used to learn the LSTM models. It is a stochastic gradient descent method that is based on adaptive estimation of first-order and second-order moments [35].

2.5. Dataset

The analyzed dataset contains hourly energy demand, generation, prices and weather [19]. The data was initially collected on an hourly basis and contains about 0.14% of missing values. The missing values are replaced with zero. The following 40 data attributes are used in the analysis:

- X_1 —year
- X_2 —month
- X_3 —day
- X_4 —hour
- X_5 —generation biomass
- X_6 —generation fossil brown coal/lignite
- X_7 —generation fossil coal-derived gas
- X_8 —generation fossil gas
- X_9 —generation fossil hard coal
- X_{10} —generation fossil oil
- X_{11} —generation fossil oil shale
- X_{12} —generation fossil peat
- X_{13} —generation geothermal
- X_{14} —generation hydro pumped storage aggregated
- X_{15} —generation hydro pumped storage consumption
- X_{16} —generation hydro run-of-river and poundage
- X_{17} —generation hydro water reservoir
- X_{18} —generation marine
- X_{19} —generation nuclear
- X_{20} —generation other
- X_{21} —generation other renewable
- X_{22} —generation solar
- X_{23} —generation waste
- X_{24} —generation wind offshore
- X_{25} —generation wind onshore
- X_{26} —**total load actual**
- X_{27} —price day ahead
- X_{28} —price actual
- X_{29} —temp
- X_{30} —temp min
- X_{31} —temp max
- X_{32} —pressure
- X_{33} —humidity
- X_{34} —wind speed
- X_{35} —wind degree
- X_{36} —rain 1 h
- X_{37} —rain 3 h
- X_{38} —snow 3 h
- X_{39} —clouds all
- X_{40} —weather id

Figure 3 visualizes the selected data attributes.

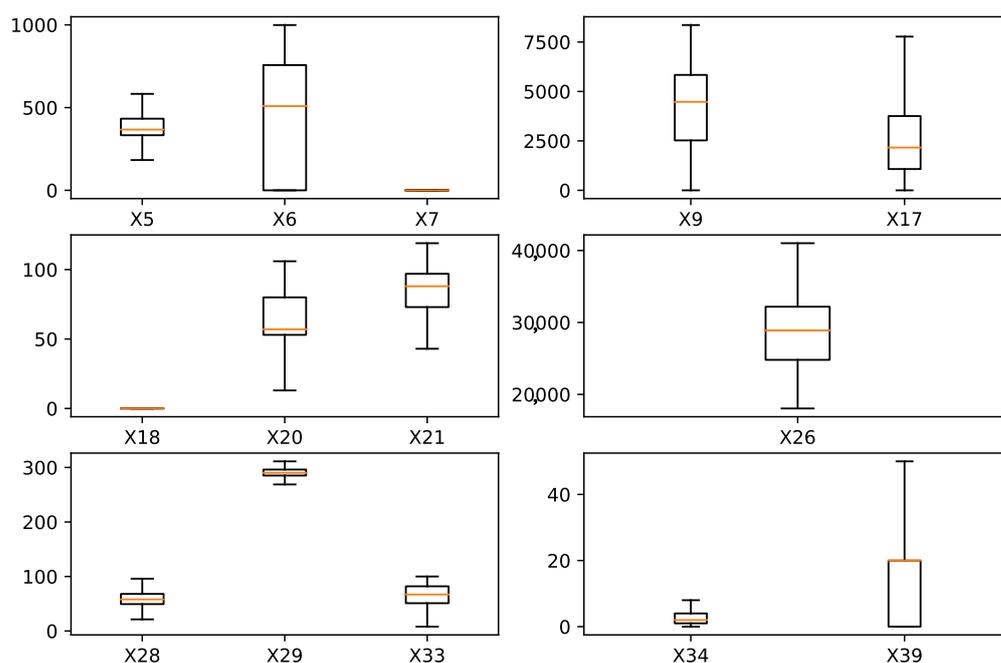


Figure 3. Box plots of selected data attributes.

3. Results

The aim of simulations is a one-step-ahead prediction of the energy use (X_{26} total load actual) based on available data. Other data attributes are grouped into eight clusters with the use of k-means clustering:

- **cluster 1:** X_7 (generation fossil coal-derived gas), X_{11} (generation fossil oil shale), X_{12} (generation fossil peat), X_{13} (generation geothermal), X_{14} (generation hydro pumped storage aggregated), X_{15} (generation hydro pumped storage consumption), X_{18} (generation marine), X_{24} (generation wind offshore), X_{34} (wind speed), X_{36} (rain 1h), X_{37} (rain 3h), X_{38} (snow 3h)
- **cluster 2:** X_5 (generation biomass), X_{10} (generation fossil oil), X_{19} (generation nuclear), X_{21} (generation other renewable), X_{23} (generation waste), X_{33} (humidity), X_{40} (weather id)
- **cluster 3:** X_8 (generation fossil gas), X_{17} (generation hydro water reservoir), X_{22} (generation solar), X_{25} (generation wind onshore), X_{39} (clouds all)
- **cluster 4:** X_6 (generation fossil brown coal/lignite), X_9 (generation fossil hard coal)
- **cluster 5:** X_2 (month), X_3 (day), X_4 (hour), X_{16} (generation hydro run-of-river and poundage), X_{20} (generation other), X_{27} (price day ahead), X_{28} (price actual), X_{32} (pressure)
- **cluster 6:** X_{35} (wind degree)
- **cluster 7:** X_1 (year)
- **cluster 8:** X_{29} (temp), X_{30} (temp min), X_{31} (temp max)
- **output:** X_{26} (total load actual)

The first level of the nested structure is a fuzzy cognitive map constructed based on the grouped data. It contains nine concepts: cluster 1, cluster 2, cluster 3, cluster 4, cluster 5, cluster 6, cluster 7, cluster 8 and the output concept. The learning process is realized with the use of genetic algorithms (RCGA, SOGA).

Each concept of the nested fuzzy cognitive map can be decomposed into another fuzzy cognitive map, multilayer perceptron artificial neural network or long short-term memory network for more detailed energy use forecasting (the second level of the nested structure).

The dataset is divided into training data and testing data. The training dataset contains 1000 records and is used during the learning process. The testing dataset contains the next 1000 records and is used to evaluate the nested structure.

Three popular metrics are calculated based on testing data: Mean Square Error (MSE), Mean Absolute Error (MAE) and the correlation coefficient (R):

$$MSE = \frac{1}{T} \sum_{t=1}^T (Z(t) - Y(t))^2 \quad (6)$$

$$MAE = \frac{1}{T} \sum_{t=1}^T |Z(t) - Y(t)| \quad (7)$$

$$R = \frac{\sum_{t=1}^T (Z(t) - \bar{Z})(Y(t) - \bar{Y})}{\sqrt{\sum_{t=1}^T (Z(t) - \bar{Z})^2 \sum_{t=1}^T (Y(t) - \bar{Y})^2}} \quad (8)$$

where $t = 1, \dots, T$, T is the number of testing records, $Z(t)$ is the true normalized value of the energy use, $Y(t)$ is the predicted value of the energy use, \bar{Z} the mean value of the true normalized values, \bar{Y} the mean value of the predicted values.

The ANN models are learned based on the backpropagation method with momentum. The LSTM models are learned based on the Adam algorithm. Both methods are stochastic gradient descent algorithms consisting of the modification of weights in the neural network. The amount that the weights are updated is determined by the learning rate (Lr). It controls the speed at which the ANN or LSTM model learns. The momentum (M) allows the oscillation to be dampened while the gradient is decreasing. The number of epochs (E) controls the number of complete passes through the learning dataset.

We choose the optimal values of the parameters by the grid search method. For ANN, the number of epochs E is searched from {100, 200, 300, 400, 500, 1000}. Learning rate Lr is chosen from {0.001, 0.005, 0.01, 0.02}. The parameter momentum M is searched from {0, 0.5, 0.9}. The parameter λ of the sigmoid activation function is chosen from {1, 2, 3}. The number of hidden layers is chosen from {1, 2, 3}. The number of neurons is searched from {5, 10, 20, 40}. For LSTM, the number of epochs E is searched from {100, 200, 300, 400, 500, 1000}. Learning rate Lr is chosen from {0.001, 0.01}. The number of hidden layers is chosen from {1 (Vanilla LSTM), 2 (Bidirectional LSTM)}. The number of neurons is searched from {4, 5, 10}.

The FCM models are learned with the use of genetic algorithms (RCGA, SOGA). The population size (P) denotes the number of individuals in a population. A mutation is a random change of a single individual with some probability (Pm). Crossover is a genetic operator that enables combining the genetic information of two parents to generate new individuals with some probability (Pc). For FCM, the population size P is chosen from {100, 200}. The maximum number of generation G is selected from {100, 200, 500}. The mutation is searched from {random, Mühlensbeins}. The mutation probability Pm is chosen from {0.2, 0.5}. A uniform crossover is used. The crossover probability Pc is chosen from {0.5, 0.8}. The parameter combination yielding the best MAE and MSE metrics is selected as the final model. Table 2 presents the sample results obtained for the nested fuzzy cognitive map.

The results show that the best forecasting accuracy is obtained for LSTM models with two hidden layers (Bidirectional LSTM) in the second level. Figure 4 illustrates the optimum nested structure based on the fuzzy cognitive map in the first level and the LSTM models in the second level.

Table 2. Results for the nested fuzzy cognitive map.

Approach	Model	MSE	MAE	R
First level of the nested structure				
Nested FCM	FCM RCGA	0.0025	0.0393	0.9023
	FCM SOGA	0.0026	0.0400	0.9025
Second level of the nested structure				
Cluster 1	FCM RCGA	0.0032	0.0452	0.8751
	FCM SOGA	0.0034	0.0476	0.9019
	ANN	0.0017	0.0315	0.9376
	LSTM	0.0013	0.0277	0.9480
Cluster 2	FCM RCGA	0.0034	0.0464	0.8953
	FCM SOGA	0.0035	0.0455	0.8622
	ANN	0.0019	0.0343	0.9310
	LSTM	0.0014	0.0289	0.9441
Cluster 3	FCM RCGA	0.0031	0.0466	0.8945
	FCM SOGA	0.0029	0.0431	0.8990
	ANN	0.0018	0.0332	0.9294
	LSTM	0.0015	0.0295	0.9433
Cluster 4	FCM RCGA	0.0028	0.0432	0.8921
	FCM SOGA	0.0027	0.0415	0.8949
	ANN	0.0017	0.0315	0.9369
	LSTM	0.0015	0.0291	0.9434
Cluster 5	FCM RCGA	0.0044	0.0542	0.8635
	FCM SOGA	0.0033	0.0479	0.9088
	ANN	0.0017	0.0315	0.9413
	LSTM	0.0015	0.0297	0.9433
Cluster 6	FCM RCGA	0.0025	0.0393	0.9023
	FCM SOGA	0.0027	0.0416	0.9025
	ANN	0.0017	0.0312	0.9377
	LSTM	0.0014	0.0291	0.9442
Cluster 7	FCM RCGA	0.0025	0.0393	0.9023
	FCM SOGA	0.0027	0.0416	0.9025
	ANN	0.0017	0.0312	0.9377
	LSTM	0.0014	0.0291	0.9442
Cluster 8	FCM RCGA	0.0025	0.0397	0.9097
	FCM SOGA	0.0025	0.0404	0.9034
	ANN	0.0017	0.0316	0.9378
	LSTM	0.0015	0.0290	0.9439

To evaluate the performance of the proposed approach, a comparative analysis between the nested structure (for the first and the second level), the standard FCM, the multilayer perceptron artificial neural network and the Long Short-Term Memory network is performed. We choose the optimal values of the parameters by the grid search method. For ANN, the number of epochs E is searched from {100, 200, 300, 400, 500, 1000}. Learning rate Lr is chosen from {0.001, 0.005, 0.01, 0.02}. The parameter momentum M is searched from {0, 0.5, 0.9}. The parameter λ of the sigmoid function is chosen from {1, 2, 3}. The number of hidden layers is chosen from {1, 2, 3}. The number of neurons is searched from {5, 10, 20, 40}. For LSTM, the number of epochs E is searched from {100, 200, 300, 400, 500, 1000}. Learning rate Lr is chosen from {0.001, 0.01}. The number of hidden layers is chosen from {1 (Vanilla LSTM), 2 (Bidirectional LSTM)}. The number of neurons is searched from {4, 5, 10}. For FCM, the population size P is chosen from {100, 200}. The number of generation G is selected from {100, 200, 500}. The mutation is searched from {random, Mühlensbeins}. The mutation probability Pm is chosen from {0.2, 0.5}. A uniform crossover is used. The crossover probability Pc is chosen from {0.5, 0.8}.

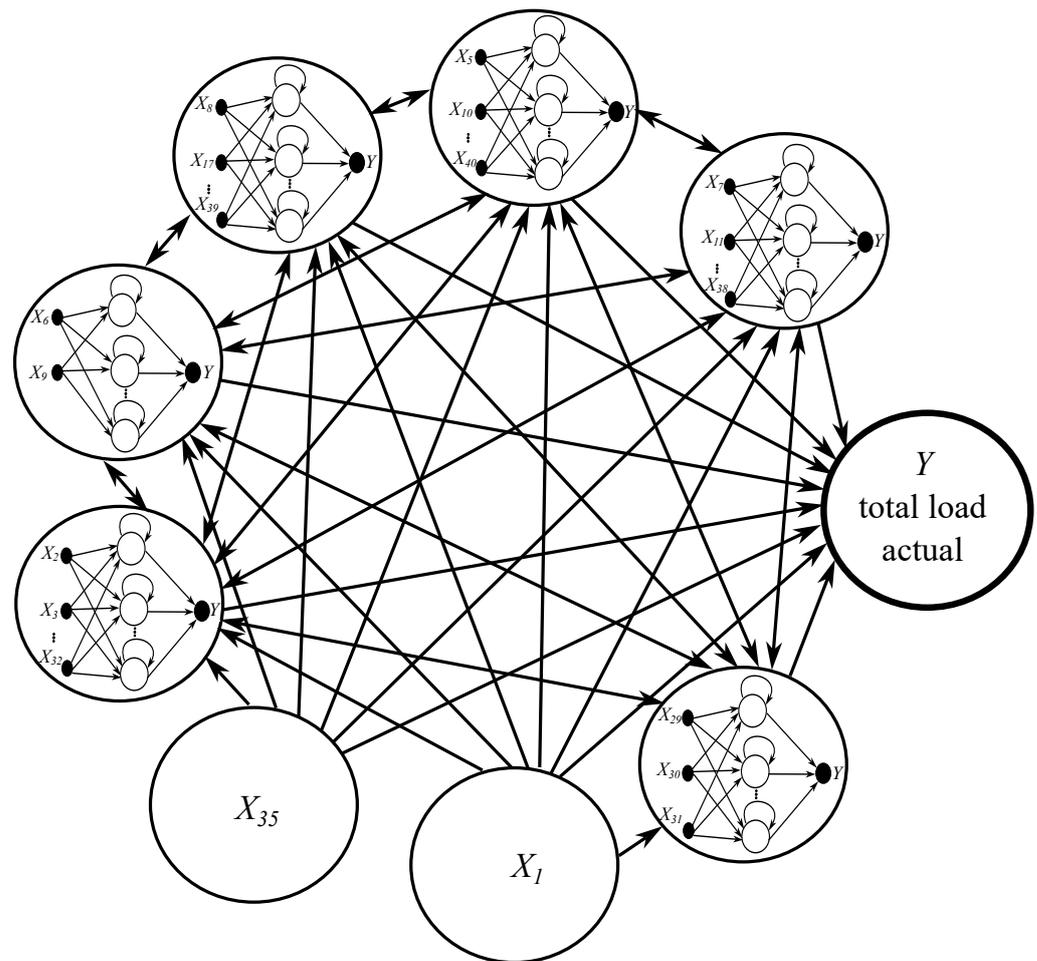


Figure 4. Sample nested structure.

Table 3 shows the results of a comparison. These results correspond to models that are trained using three best-obtained parameters.

Both the general fuzzy cognitive map (first structure level) and the best LSTM model of the second level obtained lower MAE and MSE error values and a higher value of the correlation coefficient than classic fuzzy cognitive maps and artificial neural networks. Moreover, the second level of the proposed structure outperforms the classic Vanilla LSTM models and Bidirectional LSTM models.

Additionally, the Wilcoxon Signed-Rank [36] and Friedman [37] tests for the forecasting error $FE = |Z(t) - Y(t)|$ generated by the nested structure (first level and second level) against the compared models are analyzed. The test is realized with the use of scipy.stats python module [38]. With a 0.05 significance level and a 2-tailed hypothesis, significant results are shown in Table 4. The Wilcoxon Signed-Rank test labels all the models as statistically different to the second-level model and FCM models and Bidirectional LSTM is statistically different from the first-level model (p -value < 0.05). The Friedman test comparing the first-level model with the remaining models (FCM, ANN, LSTM) gives a p -value of $1.28 \cdot 10^{-28} < 0.05$. The Friedman test comparing the second-level model with the remaining models (FCM, ANN, LSTM) gives a p -value of $1.01 \cdot 10^{-73} < 0.05$. It means that the difference among these models is statistically significant.

Table 3. Results of comparative analysis.

Model	Parameters	MSE	MAE	R
FCM	SOGA: $P = 100, G = 100$, Mühlensbeins mutation, $P_m = 0.2, P_c = 0.8$	0.0035	0.0480	0.8635
	RCGA: $P = 100, G = 100$, Mühlensbeins mutation, $P_m = 0.2, P_c = 0.8$	0.0037	0.0471	0.8573
	RCGA: $P = 100, G = 100$, Random mutation, $P_m = 0.5, P_c = 0.5$	0.0038	0.0498	0.8618
ANN	1 hidden layer (20), $Lr=0.01, M = 0.9, E = 300, \lambda = 2$	0.0026	0.0402	0.8976
	1 hidden layer (10), $Lr=0.01, M = 0.9, E = 400, \lambda = 2$	0.0028	0.0402	0.8948
	2 hidden layers (10;5), $Lr=0.01, M = 0.9, E = 1000, \lambda = 2$	0.0028	0.0414	0.8954
LSTM	Vanilla LSTM, 1 hidden layer (10), $Lr = 0.001, E = 1000$	0.0021	0.0345	0.9196
	Vanilla LSTM, 1 hidden layer (20), $Lr = 0.001, E = 1000$	0.0028	0.0399	0.9075
	Bidirectional LSTM, two hidden layers (10;5), $Lr = 0.01, E = 500$	0.0039	0.0422	0.8754
First level	SOGA: $P = 200, G = 200$, Mühlensbeins mutation, $P_m = 0.5, P_c = 0.5$	0.0026	0.0400	0.9025
	RCGA: $P = 100, G = 200$, Mühlensbeins mutation, $P_m = 0.2, P_c = 0.8$	0.0025	0.0393	0.9023
	SOGA: $P = 100, G = 100$, Random mutation, $P_m = 0.2, P_c = 0.8$	0.0025	0.0401	0.9068
Second level	Cluster 1, Bidirectional LSTM, two hidden layers (10;5), $Lr = 0.01, E = 400$	0.0013	0.0277	0.9480
	Cluster 2, Bidirectional LSTM, two hidden layers (4;4), $Lr = 0.001, E = 1000$	0.0014	0.0289	0.9441
	Cluster 8, Bidirectional LSTM, two hidden layers (4;4), $Lr = 0.01, E = 500$	0.0015	0.0290	0.9439

The training time of all models was also compared. The average training time for the ANN models is 80 s for the number of epochs $E = 500$. The average training time for the LSTM models is 30 s for the number of epochs $E = 500$. The difference is due to different implementations of both models (C# and Python). The average training time for the general FCM (first level of the nested structure) is 150 s for the number of generations $G = 100$ and the population size $P = 100$. The average training time for the classic FCM models is the longest and equals 1800 s for the number of generations $G = 100$ and the population size $P = 100$.

Table 4. p -values achieved by the paired-sample Wilcoxon signed-rank test comparing the forecasting error FE of the nested structure (first level, second level) and the remaining models.

Approach	First Level	Second Level
FCM SOGA	$1.73 \cdot 10^{-14}$	$3.73 \cdot 10^{-58}$
FCM RCGA	$1.35 \cdot 10^{-11}$	$6.95 \cdot 10^{-49}$
ANN	0.36	$2.47 \cdot 10^{-19}$
LSTM	0.21	$1.39 \cdot 10^{-17}$

The prediction performance of the analyzed models is presented with the use of bar charts, scatter plots and line charts. Figure 5 shows forecasting results for the best-analyzed models in the form of line charts. Figure 6 shows forecasting results for the analyzed models in the form of scatter plots. Figure 7 visualizes the resulting MSE, MAE and R metrics for the compared models in the form of bar charts.

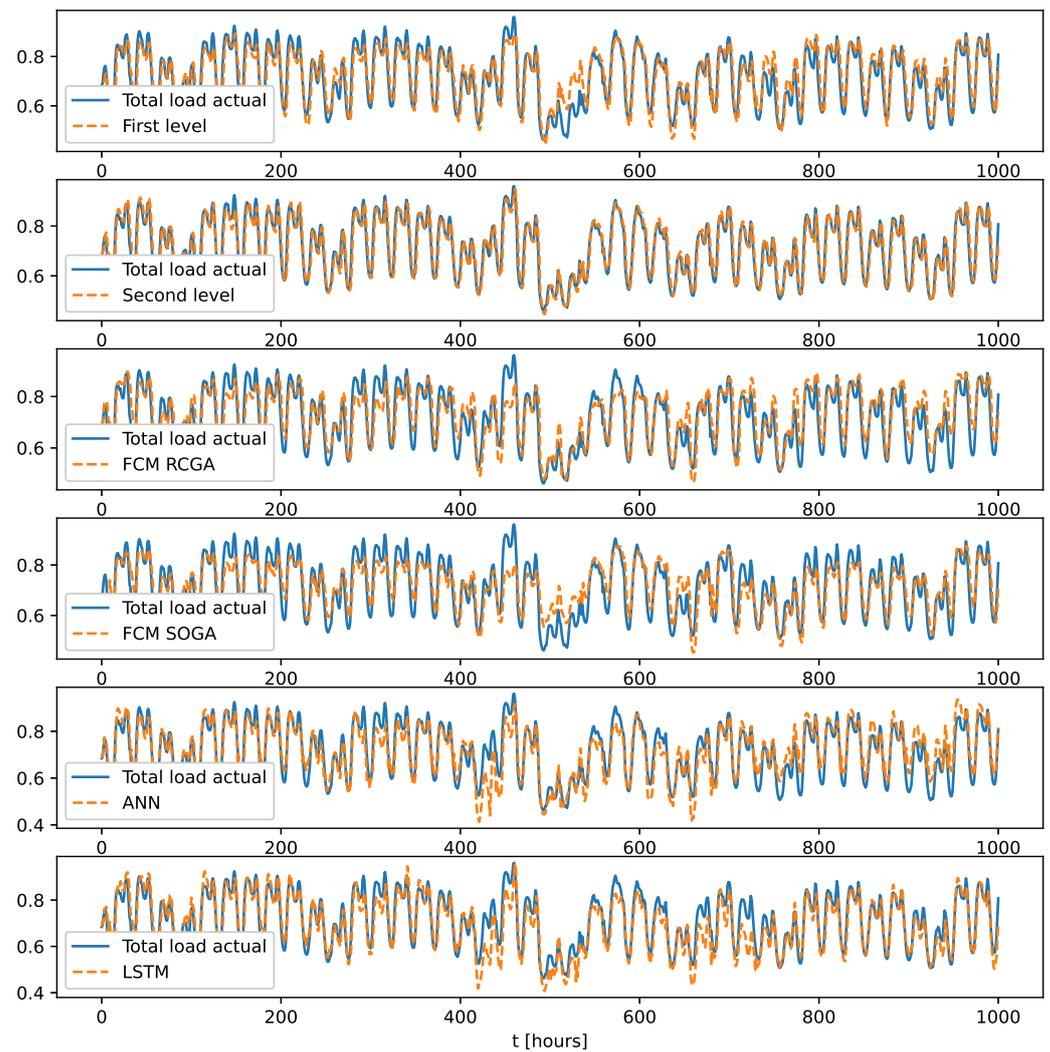


Figure 5. Forecasting results for the best model obtained with the use of the first and the second level of the nested structure, the FCM RCGA approach, the FCM SOGA approach, the ANN model and the LSTM model.

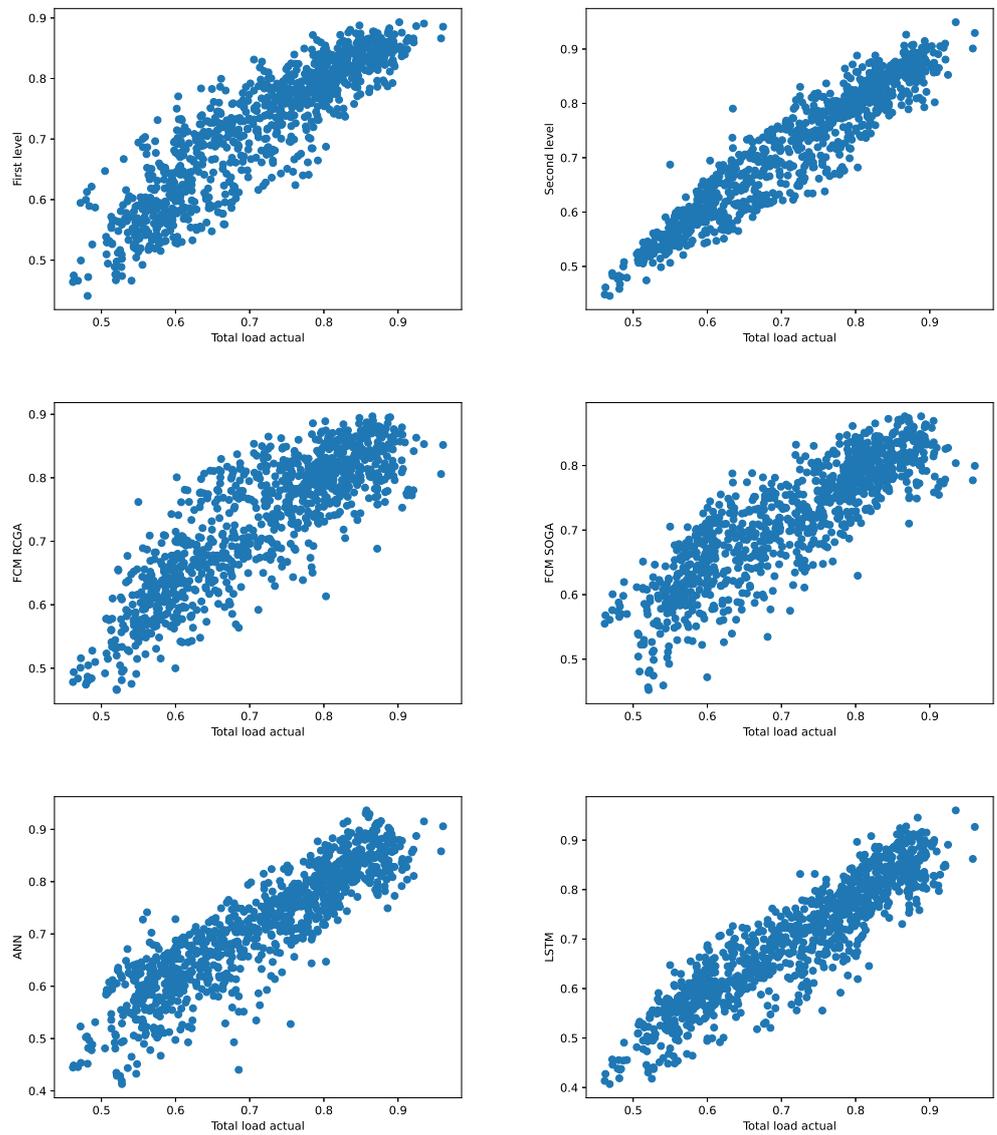


Figure 6. Scatter plots for the best models obtained with the use of the nested structure (first level, second level), FCM approaches (RCGA, SOGA), the LSTM model and the ANN model.

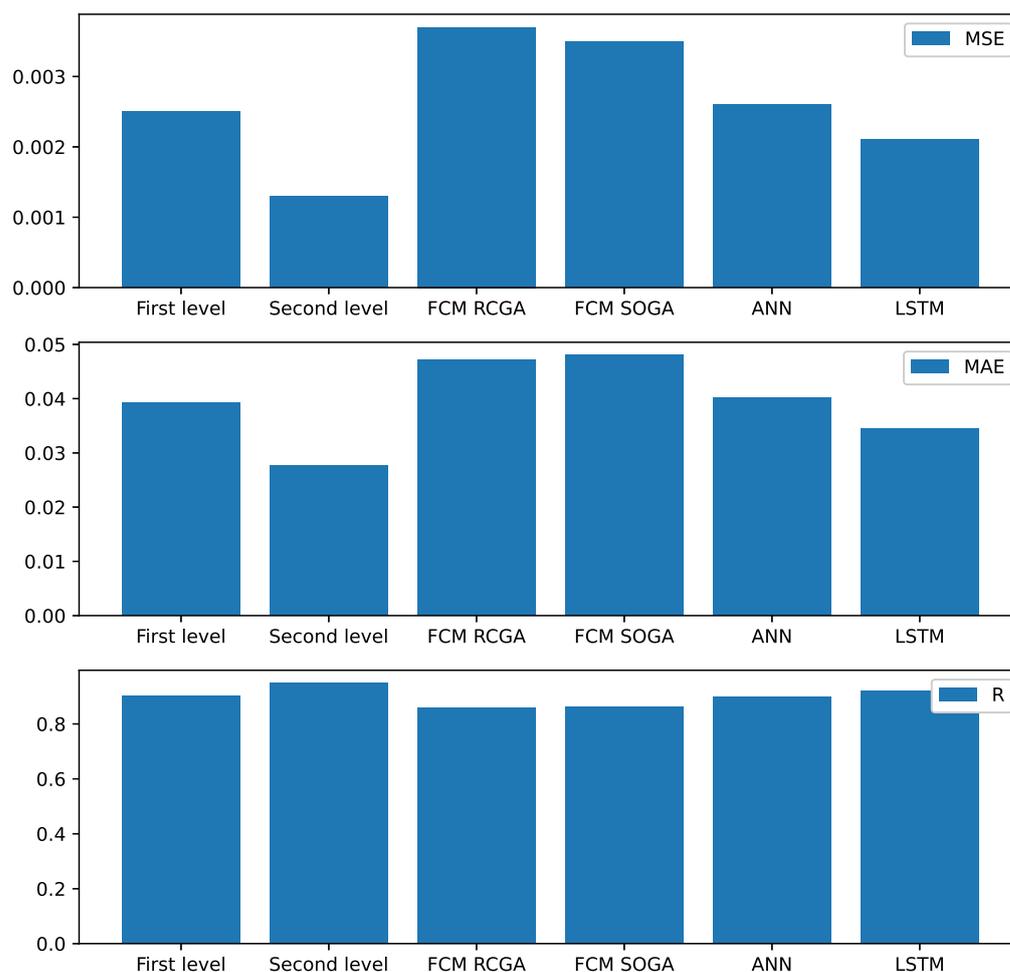


Figure 7. Forecasting metrics: MSE, MAE and R obtained for the compared models.

4. Discussion and Conclusions

This paper presents a novel nested structure based on fuzzy cognitive maps and artificial neural networks. The construction of a nested structure consists of three main steps described below: data preprocessing, constructing the first level in the form of a general fuzzy cognitive map, and constructing the second level based on fuzzy cognitive maps, multilayer perceptrons and long short-term memory networks. Historical data related to hourly energy demand, generation, prices and weather are used to evaluate the nested structure.

The proposed approach allows us to obtain a readable structure with nine concepts constructed based on the clusters and the output concept in the first level. The concepts of the first level can be decomposed, providing a more detailed representation of energy use time series. As shown in Table 2, a nested structure based on LSTM models in the second level provides the lowest forecasting errors MAE, MSE and the highest value of the correlation coefficient R. The results presented in Table 3 confirm that the proposed nested structure outperforms the standard well-known fuzzy cognitive maps, classic multilayer perceptron artificial neural networks and long short-term memory networks.

The main disadvantage of the proposed approach is the need to build a separate model for each cluster, which may be time-consuming in the case of large data sets. However, we can limit this approach to only the most general level of the nested structure.

The advantages of the proposed nested structure can be summarized as follows:

- The presented approach can be used to analyze and facilitate the understanding of complex multivariate time series, especially in the absence of expert knowledge.

- The nested structure allows you to analyze the problem of energy consumption at various levels of detail - from the most general (first level) to more detailed (second level).
- The proposed approach presents complex data with 40 various data attributes in the form of a clear nested structure presenting on the first level the main concepts influencing energy consumption. Each first-level concept contains data attributes with similar behavior.
- Such presentation of the problem can provide valuable information about energy use and facilitate decision-making.
- The second level allows for more detailed problem analysis. The use of artificial neural networks at the second level, especially LSTM models, made it possible to obtain very accurate forecasts related to energy load.
- The proposed nested structure outperforms the classic fuzzy cognitive maps and artificial neural networks in terms of prediction accuracy.
- The approach is flexible and can also be used to analyze multivariate time series in various fields, e.g., in economics, management or medicine.

Future work is oriented toward using fuzzy clustering techniques to construct the nested structure. We plan to analyze the application of the developed approach to construct fuzzy cognitive maps with three or more levels based on more complex time series. Where expert knowledge is available, the input data could be calculated on the basis of a weighted average, where the weights for individual data attributes would be determined by experts. Another approach that could be explored as the next step is the use of more advanced data processing techniques, for example, grey numbers to construct the first level of the nested structure.

Author Contributions: Conceptualization, K.P.; methodology, K.P.; software, K.P.; validation, K.P. and E.I.P.; formal analysis, K.P. and E.I.P.; investigation, K.P.; resources, K.P. and E.I.P.; data curation, K.P.; writing—original draft preparation, K.P. and E.I.P.; writing—review and editing, K.P. and E.I.P.; visualization, K.P.; supervision, K.P. and E.I.P. 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: Not applicable.

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

Abbreviations

The following abbreviations are used in this manuscript:

FCM	Fuzzy cognitive map
LSTM	Long Short-Term Memory
ANN	Multilayer perceptron artificial neural network
SOGA	Structure Optimization Genetic Algorithm
RCGA	Real-Coded Genetic Algorithm
ANFIS	adaptive neuro-fuzzy inference system
MSE	Mean Square Error
MAE	Mean Absolute Error
R	Correlation coefficient
ISEMK	Intelligent expert system based on cognitive maps

References

1. Runge, J.; Zmeureanu, R. A Review of Deep Learning Techniques for Forecasting Energy Use in Buildings. *Energies* **2021**, *14*, 608. [CrossRef]
2. Patsakos, I.; Vrochidou, E.; Papakostas, G.A. A Survey on Deep Learning for Building Load Forecasting. *Math. Probl. Eng.* **2022**, *2022*, 1–25. [CrossRef]
3. Zhang, L.; Wen, J.; Li, Y.; Chen, J.; Ye, Y.; Fu, Y.; Livingood, W. A review of machine learning in building load prediction. *Appl. Energy* **2021**, *285*, 116452. [CrossRef]
4. Meng, Z.; Sun, H.; Wang, X. Forecasting Energy Consumption Based on SVR and Markov Model: A Case Study of China. *Front. Environ. Sci.* **2022**, *10*, 1–15. [CrossRef]
5. Jamii, M.; Maaroufi, M. The Forecasting of Electrical Energy Consumption in Morocco with an Autoregressive Integrated Moving Average Approach. *Math. Probl. Eng.* **2021**, *2021*, 1–9. [CrossRef]
6. Ribeiro, A.M.N.C.; do Carmo, P.R.X.; Endo, P.T.; Rosati, P.; Lynn, T. Short- and Very Short-Term Firm-Level Load Forecasting for Warehouses: A Comparison of Machine Learning and Deep Learning Models. *Energies* **2022**, *15*, 750. [CrossRef]
7. Pannakkong, W.; Harncharnchai, T.; Buddhakulsomsiri, J. Forecasting Daily Electricity Consumption in Thailand Using Regression, Artificial Neural Network, Support Vector Machine, and Hybrid Models. *Energies* **2022**, *15*, 3105. [CrossRef]
8. Mahjoub, S.; Chrifi-Alaoui, L.; Marhic, B.; Delahoche, L. Predicting Energy Consumption Using LSTM, Multi-Layer GRU and Drop-GRU Neural Networks. *Sensors* **2022**, *22*, 4062. [CrossRef] [PubMed]
9. Akhter, M.N.; Mekhilef, S.; Mokhlis, H.; Almohaimeed, Z.M.; Muhammad, M.A.; Khairuddin, A.S.M.; Akram, R.; Hussain, M.M. An Hour-Ahead PV Power Forecasting Method Based on an RNN-LSTM Model for Three Different PV Plants. *Energies* **2022**, *15*, 2243. [CrossRef]
10. Arvanitidis, A.I.; Bargiotas, D.; Daskalopulu, A.; Laitos, V.M.; Tsoukalas, L.H. Enhanced Short-Term Load Forecasting Using Artificial Neural Networks. *Energies* **2021**, *14*, 7788. [CrossRef]
11. Shams Amiri, S.; Mostafavi, N.; Lee, E.R.; Hoque, S. Machine learning approaches for predicting household transportation energy use. *City Environ. Interact.* **2020**, *7*, 100044. [CrossRef]
12. Amiri, S.S.; Mottahedi, S.; Lee, E.; Hoque, S. Peeking inside the black-box: Explainable machine learning applied to household transportation energy consumption. *Comput. Environ. Urban Syst.* **2021**, *88*, 101647. [CrossRef]
13. Kosko, B. Fuzzy cognitive maps. *Int. J. -Man-Mach. Stud.* **1986**, *24*, 65–75. [CrossRef]
14. Çoban, V.; Onar, S. Modeling renewable energy usage with hesitant Fuzzy cognitive map. *Complex Intell. Syst.* **2017**, *3*, 155–166. [CrossRef]
15. Orang, O.; Silva, P.C.d.L.; Guimarães, F.G. Introducing Randomized High Order Fuzzy Cognitive Maps as Reservoir Computing Models: A Case Study in Solar Energy and Load Forecasting. *arXiv* **2022**, arXiv:2201.02158.
16. Hatwagner, M.F.; Koczy, L.T. Parameterization and concept optimization of FCM models. In Proceedings of the 2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Istanbul, Turkey, 2–5 August 2015. [CrossRef]
17. Poczeta, K.; Papageorgiou, E.I.; Gerogiannis, V.C. Fuzzy Cognitive Maps Optimization for Decision Making and Prediction. *Mathematics* **2020**, *8*, 2059. [CrossRef]
18. Papageorgiou, K.I.; Poczeta, K.; Papageorgiou, E.; Gerogiannis, V.C.; Stamoulis, G. Exploring an Ensemble of Methods that Combines Fuzzy Cognitive Maps and Neural Networks in Solving the Time Series Prediction Problem of Gas Consumption in Greece. *Algorithms* **2019**, *12*, 235. [CrossRef]
19. Kazuki, H. Electric Price EDA & Prediction. Available online: <https://www.kaggle.com/code/kazukihirahara/electric-price-eda-prediction-lightgbm/data> (accessed on 21 May 2022).
20. Han, J.; Kamber, M.; Pei, J. 3-Data Preprocessing. In *Data Mining (Third Edition)*, 3rd ed.; Han, J., Kamber, M., Pei, J., Eds.; The Morgan Kaufmann Series in Data Management Systems; Morgan Kaufmann: Boston, MA, USA, 2012; pp. 83–124. [CrossRef]
21. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
22. sklearn.preprocessing.MinMaxScaler. Available online: <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html> (accessed on 11 July 2022).
23. MacQueen, J.B. Some Methods for Classification and Analysis of MultiVariate Observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*; Cam, L.M.L., Neyman, J., Eds.; University of California Press: Berkeley, CA, USA, 1967; Volume 1, pp. 281–297.
24. Arthur, D.; Vassilvitskii, S. K-Means++: The Advantages of Careful Seeding. In Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms SODA '07, New Orleans, LA, USA, 7–9 January 2007; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 2007; pp. 1027–1035.
25. sklearn.cluster.KMeans. Available online: <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html#sklearn.cluster.KMeans> (accessed on 11 July 2022).
26. Bueno, S.; Salmeron, J.L. Benchmarking main activation functions in fuzzy cognitive maps. *Expert Syst. Appl.* **2009**, *36*, 5221–5229. [CrossRef]
27. Stach, W.; Kurgan, L.; Pedrycz, W.; Reformat, M. Genetic learning of fuzzy cognitive maps. *Fuzzy Sets Syst.* **2005**, *153*, 371–401. [CrossRef]

28. Poczeta, K.; Yastrebov, A.; Papageorgiou, E.I. Learning fuzzy cognitive maps using Structure Optimization Genetic Algorithm. In Proceedings of the 2015 Federated Conference on Computer Science and Information Systems (FedCSIS), Lodz, Poland, 13–16 September 2015; pp. 547–554. [[CrossRef](#)]
29. McCulloch, W.; Pitts, W. A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.* **1943**, *5*, 115–133. [[CrossRef](#)]
30. Haykin, S. *Neural Networks: A Comprehensive Foundation*; Macmillan Publishing: New York, NY, USA, 1994; ISBN 0-02-352781-7.
31. Papageorgiou, E.I.; Poczeta, K. A two-stage model for time series prediction based on fuzzy cognitive maps and neural networks. *Neurocomputing* **2017**, *232*, 113–121. [[CrossRef](#)]
32. Hochreiter, S.; Schmidhuber, J. Long Short-term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)] [[PubMed](#)]
33. Kim, K.; Kim, D.K.; Noh, J.; Kim, M. Stable Forecasting of Environmental Time Series via Long Short Term Memory Recurrent Neural Network. *IEEE Access* **2018**, *6*, 75216–75228. [[CrossRef](#)]
34. Chollet, F. Keras. GitHub. 2015. Available online: <https://github.com/fchollet/keras> (accessed on 16 July 2022).
35. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980v9.
36. Wilcoxon, F. Individual Comparisons by Ranking Methods. *Biom. Bull.* **1945**, *1*, 80–83. [[CrossRef](#)]
37. Daniel, W.W. *Applied Nonparametric Statistics*, 2nd ed.; Duxbury Advanced Series in Statistics and Decision Sciences; PWS- Kent Publishing Company: Boston, MA, USA, 1990.
38. Statistical Functions (scipy.stats). Available online: <https://docs.scipy.org/doc/scipy/reference/stats.html> (accessed on 16 September 2022).