energies

MDPI

Article

An Ensemble Deep-Learning-Based Model for Hour-Ahead Load Forecasting with a Feature Selection Approach: A Comparative Study with State-of-the-Art Methods

Fatma Yaprakdal

Engineering Faculty, Electrical and Electronics Engineering Department, Kirklareli University, 39100 Kirklareli, Turkey; fatmayaprakdal@klu.edu.tr

Abstract: The realization of load forecasting studies within the scope of forecasting periods varies depending on the application areas and estimation purposes. It is mainly carried out at three intervals: short-term, medium-term, and long-term. Short-term load forecasting (STLF) incorporates hour-ahead load forecasting, which is critical for dynamic data-driven smart power system applications. Nevertheless, based on our knowledge, there are not enough academic studies prepared with particular emphasis on this sub-topic, and none of the related studies evaluate STLF forecasting methods in this regard. As such, machine learning (ML) and deep learning (DL) architectures and forecasters have recently been successfully applied to STLF, and are state-of-the-art techniques in the energy forecasting area. Here, hour-ahead load forecasting methods, the majority of which are frequently preferred high-performing up-to-date methods in the literature, were first examined based on different forecasting techniques using two different aggregated-level datasets and observing the effects of these methods on both. Case and comparison studies have been conducted on these high-performing methods before, but there are not many examples studied using data from two different structures. Although the data used in this study were different from each other in terms of the time step, they also had very different and varied features. In addition, feature selection was studied on both datasets and a backward-eliminated exhaustive approach based on the performance of the artificial neural network (ANN) on the validation set was proposed for the development study of the forecasting models. A new DL-based ensemble approach was proposed after examining the results obtained on two separate datasets by applying the feature selection approach to the working forecasting methods, and the numerical results illustrate that it can significantly improve the forecasting performance compared with these up-to-date methods.

Keywords: hour-ahead load forecasting; aggregated-level; DL; ML; forecaster ensemble; feature selection

check for updates

Citation: Yaprakdal, F. An Ensemble Deep Learning-Based Model for Hour-Ahead Load Forecasting with a Feature Selection Approach: A Comparative Study with State-of-the-Art Methods. *Energies* **2023**, *16*, 57. https://doi.org/10.3390/en16010057

Academic Editor: Jaroslaw Krzywanski

Received: 16 November 2022
Revised: 14 December 2022
Accepted: 16 December 2022
Published: 21 December 2022

Copyright: © 2022 by the author. 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

The electrical power system is a dynamic system that is known as one of the most complex systems designed by humans. This system, consisting of power system generators, transmission and distribution lines, transformers, loads, and protective devices, must always maintain its stability, including its conditions of failure [1]. Moreover, this system is equipped with smart technologies and keeps up with developments in communication and intelligent power control technologies while preserving its smart grid feature [2]. One of the most important cases in the operation of power systems is to maintain a supply–demand balance to instantly meet the electrical load demand without any interruption. This demand-dependent case requires a sufficient energy production scheme at all times [3]. Therefore, having prior knowledge of electrical energy consumption will enable power utility companies to correctly manage each distributed energy source in a user-friendly manner to advance the future power market. In addition, this situation is extremely important in preventing sudden power system outages and losses caused by supply–demand imbalances. In this regard, accurate load estimation is of great interest, especially

in power system management [4]. The accuracy and reliability of the estimations directly affect the operators of power systems in the deregulated power market environment, resulting in substantial savings in operating and maintenance costs. Moreover, this will indirectly affect the end-user, and electricity consumption costs will decrease. Furthermore, the majority of the operationally decisive processes, such as operational energy planning, dynamic electrical energy consumption pricing in deregulated power market environments, generation and maintenance scheduling, energy transactions, optimal unit commitment, and economic power dispatch, need accurate load forecasting [5].

Electrical load forecasting (ELF) is studied in three different time intervals, namely short-term, medium-term, and long-term, depending on these power system application areas. Long-term forecasts have periods ranging from months to years, medium-term forecasts have periods ranging from one week to several months, and short-term forecasts have periods ranging from an hour to a week. On the other hand, power system applications that generally require long-term forecasts are power system planning, those that require mid-term applications are maintenance and fuel supply planning, and hourly and daily power system operations, management, and planning require short-term applications. ELF can generally be performed in two levels: aggregated and building (individual) levels. With regard to the building level, load estimation is a priority in energy management systems. Aggregated-level forecasting gives an estimate of the total load demand for a group of users at such particular levels as system level [6,7], regional level [8], and feeder level [9]. As data-driven methods, ML-based methods rely on historical load data aggregated over previous periods. These methods are of great interest to most researchers, but their accuracy can decrease when utilized on new electrical load data [10].

Instantaneous forecasts concerning hour-ahead electrical demand are utilized by energy providers in the realization of fluctuations and power outages that are unpredictable by day-ahead and week-ahead electrical demand prediction. Although operationally decisive processes such as optimal unit commitment, economic power dispatch, and maintenance scheduling need hour-ahead demand prediction, energy sector stakeholders have not gained much from ML techniques in hour-ahead demand prediction in comparison with the commissioning of month-ahead and day-ahead demand prediction models [11]. This subject has not been extensively studied in the literature since it has not been utilized frequently in practice until today. Related articles mostly focus on day-ahead load forecasting or next-day peak load forecasting.

It is possible to classify methods of hour-ahead ELF as a form of STLF into three categories: traditional, artificial intelligence (AI), and hybrid. A sufficient accuracy may not be achieved when using time-series approaches as a traditional method since they only consider the electrical load, whereas multiple linear regression (MLR) models have been commonly utilized for STLF studies even in recent decades because of their satisfactory accuracy [12]. Furthermore, they perform poorly in terms of accuracy, performance, and speed compared to AI-based methods, especially ML methods used in power system applications. The most popular ML-based methods in STLF are ANNs [13–15], support vector machines (SVMs) [16] , regression tree (RTs) [17], MLR [12,18], k-nearest neighbor (KNN) algorithms [19], recurrent neural networks (RNNs) [20,21], convolutional neural networks (CNNs) [22,23], and different ML ensemble models [4,24,25] and hybrid models, which come through deficits of conventional and AI forecasting models [26–29].

On the other hand, short-term load sequences have a highly fluctuating and stochastic nature, which makes it difficult to obtain accurate results in STLF. Factors affecting this character of load sequences are varied and complex, such as climatic, geographical, energy prices, and holidays. The original load sequence has a large amount of non-periodic information due to its complexity. If raw data are directly inserted into the deep learning model, the complexity of the hidden layer neuron coefficients will rise and the fitting performance will also decrease [30]. At this point, feature selection plays an important role in removing information complexity. Advanced ML methods such as mutual information

(MI), ReliefF (RF), and correlation-based selection (CFS) can be used to interpret the grade of candidate features and give weights to features [31].

In the remainder of this article, the literature review and problem statement are presented in Section 1. Section 2 gives datasets in detail and information on the methodologies used in the experimental phase of the study, and then outlines the development processes of proposed methods. In Section 3, experimental results and a benchmarking study are presented with the support of related tables. Finally, Section 4 summarizes the findings of the suggested model, emphasizing its originality, and offers recommendations for further research.

1.1. Literature Review

There are a variety of research studies regarding STLF, but most of them are related to day-ahead, week-ahead, and longer periods, rather than for hour-ahead horizon targets. This section reviews selected literature associated with hour-ahead LF.

Chitalia et al. applied DL algorithms to hour-ahead and day-ahead building-level ELF, which displayed a superior performance, with an improvement of 20–45% compared to the latest results available in the literature. Moreover, a 30% improvement was also observed in hour-ahead results by the authors with available 15 min resolution data [21].

A multi-scaled RNN-based DL technique was proposed for one-hour-ahead ELF by Bui et al. Regarding the approach, observation values were classified into two scales—short (6-hour duration) and long-term (1-week duration)—and RNNs were fed with these [32].

Rafati et al. proposed LTF-RF-MLP, a new hybrid method, and applied it on New England ISO data for the study of one-hour-ahead forecasting. This hybrid method extracts innovative features from load data by using load tracing features, and then uses RF for selecting the most suitable features and MLP as the forecaster. Unlike others, this method only uses load data, but is more accurate than the benchmarking models by 0.42% in terms of the yearly mean absolute percentage error (MAPE) [27].

Day and hour-ahead forecasting models were developed for several types of buses (urban, suburban, and industrial) based on an ANN by Panapakidis et al. In addition to the ANN model, a time-series clustering model was used to improve the accuracy of the ANN model. According to the overall test results, hour-ahead forecasting displayed a better accuracy compared with day-ahead forecasting due to the operation of the ANN itself. When the ANN model was taken as reference, a 5.23% average improvement in the MAPE value was achieved with the developed model [14].

Dagdougui et al. carried out day and hour-ahead ELF on real-world data from a campus in downtown Montreal. An ANN model was utilized for the aforementioned study to analyze the ANN performance in many aspects, such as back-propagation algorithm, hidden layer, and input number. It was observed that increasing the number of neuron layers has a significant contribution on the learning process while making no significant difference in MAPE. The Bayesian regularized learning algorithms and hour-ahead predictions displayed a higher accuracy and performance compared with the Levenberg–Marquardt learning algorithm and day-ahead prediction, respectively [15].

The ensemble re-forecast method was proposed for the prediction of one-hour and day-ahead electrical loads over a state-wide load dataset by Kaur et al. With the proposed method, the hour-ahead and day-ahead forecasts displayed a 47% improvement in MAPE, whereas the re-forecasts displayed a considerable improvement during off-peak hours and a small improvement during peak hours. The ensemble re-forecast method was proposed for the prediction of one-hour and day-ahead electrical loads over a state-wide load dataset. The hour-ahead and day-ahead forecast results obtained with the proposed method displayed an improvement in MAPE of up to 47%, whereas the re-forecasts showed a significant improvement in off-peak hours and a minor improvement in on-peak hours [24].

One-hour-ahead forecasting was realized by a new data-driven approach in terms of the energy Internet by Du et al. In this context, the spatiotemporal correlation-based multidimensional input automatically adds miscellaneous information concerning load variations followed by data preprocessing based on low-rank decomposition and load

gradients, thus clarifying the features to learn more targeted features. Finally, the developed 3D CNN-GRU model was implemented, and a good accuracy was achieved, with a mean absolute error (MAE) of 2.14% [28].

Laouafi et al. developed three electrical-demand-forecasting models by performing the adaptive neuro-fuzzy inference system approach in parallel load series. For forecasting the one-hour-ahead load, the real-time quart-hourly metropolitan France electricity load was used by implementing these three models. One of the models showed the best performance, with a nearly 0.5 absolute percentage error for 56% of the forecasted loads. The authors have concluded that a highly accurate model can be developed with fewer data through a combined model of a neural network and fuzzy logic [33].

Fallah et al. carried out a study on computational intelligence techniques used for load forecasting. They concluded that the performances of the methods first depend on datasets, and that hybrid methods outperform single methods in terms of accuracy [34].

1.2. Problem Statement

This paper provides research results regarding hour-ahead ELF, a subfield of STLF, while investigating other state-of-the-art papers in the field. The estimation performance of ELF methods is firstly dependent on the datasets used in the estimation procedure [34]. Hence, datasets with different structures from two countries, the Czech Republic and Australia, were used in the present study. The hour-ahead electrical consumption is a forecast using fine-tuned time-series prediction models that make use of supervised and ensemble learning methods. Hour-ahead MLR, CNN, and RNN forecasters were built for the first model, and hour-ahead SVR, KNN, and tree forecasters were built for the second model, with the goal of shortening the training period. Additionally, feature selection was carried out independently for each analyzed method on two datasets in order to observe the impact of each feature on each dataset. For feature selection, a backward-eliminated exhaustive method based on the ANN performance on the validation set was suggested. The proposed ensemble models and state-of-the-art techniques were then evaluated for their performance using the following accuracy metrics using normalized data and the z-score method [35]: MAPE [36] and MAE [37]. Since these metrics have been commonly used in other similar studies, they were preferred for comparison of results. Experiments were conducted using Matlab r2021a, providing rich support for ML.

2. Methodology

Selecting a proper forecaster structure is the first important step when designing a forecasting system. Many different forecaster architectures are presented here for this study on one-hour-ahead load forecasting. These forecasters used the last six days' one-hour resolution data as input. The number of previous days (i.e., six days) taken to form the input sequence was coarsely determined by trying the performance of a few possible values on the validation set. Here, the sequence length 's' was 144 for hourly resolution data (Czech data) and 288 for 30 min resolution data (Australian data). Adjustable parameters that belong to forecasters are stated with respect to the protocol represented in Section 4 (cross-validation) by using another subset of the data.

2.1. Input Datasets

The Czech data form a one-hour resolution dataset that includes information about year, day, month, hour, wind power plant (WPP) energy generation, photovoltaic power plant (PVPP) energy generation, load consumption including pumping load, temperature, direct horizontal radiation, and diffuse horizontal radiation features from 1 January 2012 to 31 December 2016 in the Czech Republic. All of the information in the dataset was considered for predicting the next hour's electrical energy demand value. Thus, 11 separate features were used as input.

Ausdata form a 30 min resolution (48 measurements per day) dataset with data on day, month, year, hour, dry bulb temperature, dew point, electricity consumption price,

electrical energy consumption, wet bulb temperature, and humidity from 1 January 2006 to 31 December 2010 in Australia. Electrical load consumption and price data were provided by the Australian Energy Market Operator, and meteorological data were obtained from the Bureau of Meteorology for Sydney and New South Wales. Henceforth, the next hour's electrical energy demand value was predicted by taking into consideration all of the information in the dataset. As such, nine separate features were used as input.

For data preprocessing, month, day, and hour values were divided by corresponding maximum values and one was added to the resulting values. As for the other values, preprocessing was handled by subtracting the mean and then dividing to standard deviation, ending up with zero mean and unit variance.

2.2. Backward-Eliminated Exhaustive Feature Selection Approach

A backward-eliminated exhaustive approach based on the performance of ANN on the validation set was proposed for feature selection. The algorithm was initialized with the complete set of features. A reasonable amount of best features (i.e., six features) was first selected as an initial elimination using the backwards feature elimination. Backward-elimination works by eliminating the worst performing feature by trying to leave out each feature one by one and finding the best performing subset of one less feature. This process was repeated until the remaining number of features was 6.

In the second stage, best feature subset was determined utilizing exhaustive feature selection, which only works in $2^6 - 1 = 63$ (except the empty subset, thus 63 trials in total) iterations. Performance of candidate features $\hat{k}$ was decided using the MAE metric. The proposed algorithm is explained in Algorithm 1.

Algorithm 1 Feature selection

Initialize the feature set R as the complete set of features
while $|R| > 6$ **do**
$k \leftarrow argmin_{\hat{k}} MAE(F_{ANN}(X_{R \setminus \hat{k}}))$ where the function MAE gives the MAE of the given model and X is the observations matrix in the validation set with the denoted feature subset
$R \leftarrow R \setminus k$
end while
$r \leftarrow argmin_{\hat{r}} MAE(F_{ANN}(X_{\hat{r} \in \mathcal{P}(R \setminus \emptyset)}))$ where $\mathcal{P}(R \setminus \emptyset)$ is the power set of $R \setminus \emptyset$, $\hat{r}$ thus denotes any subset of R except $\emptyset$
Output r

2.3. Feedforward ANN

Neural networks are appropriate structures for the development of forecasting methods, owing to their nonlinear architecture and capability to learn from data. They were early attempts toward the development of load forecasting methods [38].

This work utilized two different ANNs. First ANN had 1 hidden layer of 12 hidden units. Second ANN had five hidden layers. Hyperbolic tangent sigmoid transfer function was used in hidden layers and pure linear transfer function was used in output layers. Figure 1 demonstrates the ANNs. The trained ANN function was $F_{ANN_c}(\hat{x})$, where c was either 1 or 2 for some test sample $\hat{x}$.

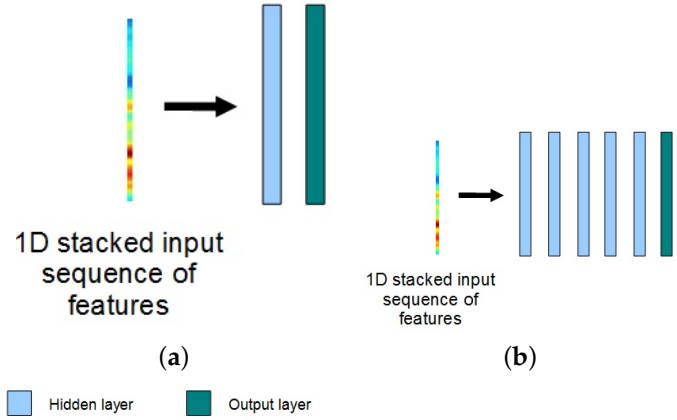

Figure 1. Proposed ANN architectures. (**a**) ANN_1: Hidden layer #units: 12.; (**b**) ANN_2: Hidden layer #units (from left to right): 50, 40, 30, 20, 10.

2.4. CNN

CNN is a feed-forward neural network and the neurons in this neural network structure are created by imitating human neurons. CNN, which is a deep learning method, has been widely applied in areas such as visual and audio processing and natural language processing. It is employed to process the power system data topology within the scope of the ELF [39].

CNN was utilized in the present study to learn continuous load values for forecasting task. This study utilized two independent CNNs (CNN_1 and CNN_2) and an ensemble of these two forecasters. CNN input size was $f \times s$ (number of features times sequence length). Figure 2 demonstrates the proposed CNN architectures. The trained CNN function was $F_{CNN_c}(\hat{x})$, where c is one of 1, 2, 3, or 4 for some test sample $\hat{x}$.

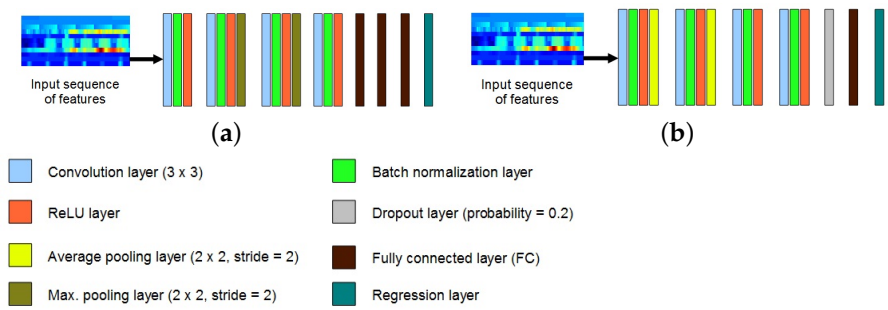

Figure 2. Proposed CNN architectures. (**a**) CNN_1: Convolution layer #filters (from left to right): 8, 16, 32, 64. FC sizes (from left to right): 5096, 512, s; (**b**) CNN_2: Convolution layer #filters (from left to right): 8, 16, 32, 32. FC size: s.

2.5. MLR

The goal of regression analysis is to specify a function that defines the correlation among the variables in such a way that the value of the dependent variables can be forecasted using a set of independent variable values [12]. MLR describes a linear relationship between a dependent variable y and one or more independent variables (features) matrix X of observations:

$$y_i = \beta_0 + \sum_{k=1}^{s \times f} beta_k X_{ik} + \epsilon_i, i = 1, ..., n, \tag{1}$$

where y_i is the ith response, β_k is the kth coefficient, X_{ik} is the ith observation on the kth element of input sequence, s is the sequence length, f is the number of features, n is the number of observations, and ϵ_i denotes the ith noise term. The fitted linear function can be expressed as:

$$\hat{y}_i = b_0 + \sum_{k=1}^{s \times f} b_k X_{ik} + \epsilon_i, i = 1, ..., n, \tag{2}$$

where $\hat{y}_i$ is the estimated response and b_ks are the coefficients, which are learned by minimizing the mean squared difference between y and $\hat{y}$. The multiple linear forecaster function F_{MLR} for a test sample $\hat{x}$ of length $s \times f$ is thus given as:

$$F_{LR}(\hat{x}) = b_0 + \sum_{k=1}^{s \times f} b_k \hat{x}_k + \epsilon. \tag{3}$$

2.6. SVM

SVMs were presented by Vapnik in 1995 for analyzing machine learning problems such as regression, pattern recognition, and density estimation. SVM as a learning mechanism applies a high-dimensional feature space and yields functions that are defined in terms of a subset of support vectors. SVM can model complex structures with a few support vectors. SVMs comprise support vector classification (SVC), which defines classification with SVMs, and support vector regression (SVR), which defines regression with SVMs [40].

In this study, cost was chosen as 0.08 and ν was chosen as 0.5 without a detailed grid search for hyper-parameter fine-tuning. The trained SVR function was $F_{SVR}(\hat{x})$ for some test sample $\hat{x}$. Hyper-parameters were coarsely optimized using the validation set by evaluating a few values instead of fine-tuning. Aim was to prevent overfitting to the validation set and to be more general and adaptive to the possible best feature subset found by the proposed feature selection algorithm.

2.7. RT

Because of the division circumstances, the decision tree clarifies the correlation between input and output variables and provides a visual representation of the problem structure to be worked with a binary tree; that is, the decision tree can find out the rules embedded in the input and output data, and, therefore, this method can be included in data mining techniques. Decision trees can be classified into two groups: classification and regression trees. In classification, the classification phase is handled with a discrete number, whereas regression trees care for one with an uninterrupted output variable [41].

Hyper-parameters of the tree (such as minimum sample leaf size, maximum number of splits, and number of variables to sample) were automatically optimized from each training set in the current study by way of Bayesian optimization using an inner cross-validation within the current training set, where the current training set was also determined from an outer cross-validation as explained in Section 3.1. The trained tree function was $F_{tree}(\hat{x})$ for some test sample $\hat{x}$.

2.8. KNN

In the KNN method, the k-nearest neighbor of an unknown pattern is found and classified by considering the labels of these neighbors. For regression problems, KNNs are utilized by taking the average of the ground-truths of k nearest neighbors [7].

It was determined in the present study, by applying a few tests on validation set (first partition of Czech data-Section 3.1), that the simplest KNN architecture of $k = 1$ and the Euclidean distance performs well as the metric; hence, this simple architecture was preferred for KNN tests. The trained KNN function was denoted as $F_{KNN}(\hat{x})$ for some test sample $\hat{x}$.

2.9. RNN

An RNN is a type of ANN that uses sequential information as a result of directed connections between units in each layer. RNNs are known as 'recurrent' since they perform the same duty for each member [42]. They are able to identify patterns in time series data.

These models process an input sequence at a time and keep hidden units in their state vector, which comprise knowledge about the history of all of the former components of the series. Nevertheless, RNNs have a gradient vanishing problem and hence cannot maintain long-range dependences [43].

Here, fully connected layer gave one output of sequence length s (e.g., 144 for Czech data) as only one feature (i.e., load) was forecasted. Figure 3 demonstrates the proposed RNN. The trained RNN function was $F_{RNN}(\hat{x})$ for some test sample $\hat{x}$.

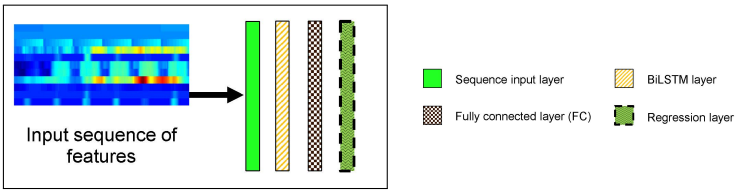

Figure 3. Proposed RNN architecture.

2.10. Proposed Forecaster Models

Depending on the characteristics of the forecasters, two different models were utilized. Input was either represented as image of size (sequence length × feature size for CNN and RNN) or one-dimensional vector (others).

$Model_1$ forecasted the next hour's single value. The models were trained with all training data, with inputs of past 6 days' feature values to predict just the next hour, obviously including all in-between hours (e.g., using previous 144 h's measurements, forecast the value at 14:00). One forecaster instance of a kind suffices to generate the $Model_1$. Hour-ahead ANN, MLR, CNN, and RNN forecasters were built following this approach. $Model_1$ is shown in Figure 4.

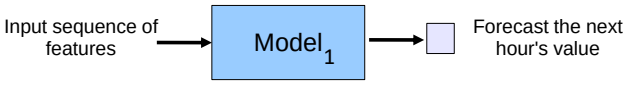

Figure 4. Forecaster $Model_1$.

Similar to $Model_1$, $Model_2$ forecasted the next hour's single value and all training data were utilized during training. Different from $Model_1$, a specialized forecaster was trained depending on the hour to be forecasted. In total, there were H different forecasters $\{Model_{2_1}, Model_{2_2}, ..., Model_{2_H}\}$ that built up $Model_2$. Hour-ahead SVR, KNN, and tree forecasters were built following this approach. The main motivation was to reduce the training time dramatically while preserving performance. $Model_2$ is illustrated in Figure 5.

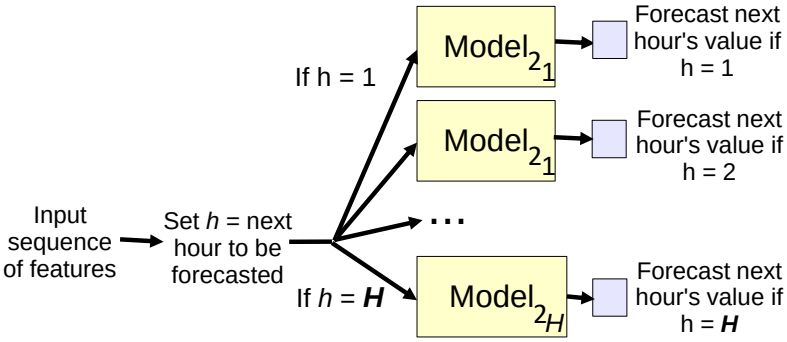

Figure 5. Forecaster $Model_2$.

Hourly Ensemble

Best ensemble for each hour was determined independently by finding the weight set W^h for each hour h to apply a hourly weighted sum of forecasts by different forecasters:

$$F^h(\hat{x}) = \sum_{k=1}^{N} W_k^h F_k^h(\hat{x}), \tag{4}$$

where F_k^h is the output of forecaster model k (e.g., CNN, RNN...) for hour h, W_k^h is the corresponding weight ($0 \leq W_k^h \leq 1$ and $\sum_{k=1}^{N} W_k^h = 1$), and N represents the number of different forecasters. W_k^hs were determined using the validation set. Forecaster ensemble can be helpful for improving the same kind of single forecasters with different hyper-parameters (e.g., ANNs, RNNs, CNNs) by taking the average of forecasts.

3. Experimental Results and Benchmarking

3.1. Cross-Validation

The days were split into five equal folds in size for cross-validation. One of these folds was reserved for testing, whereas the remaining four folds were employed for training. The first fold was used for ensemble weight optimization, selective ensemble hours, model hyper-parameter optimization, and feature selection, which are development goals here. However, that fold was utilized for training the models in the remaining tests, whereas the other folds took part in training. The results eventually consist of average and standard deviation values for four cross-validation tests.

3.2. Feature Selection

The initially selected features for the Czech dataset were day, hour, PVPP, load, wind speed (10 m), and radiation (diffuse horizontal). The best subsets of these features were exhaustively found as day, hour, load, and radiation (diffuse horizontal). The initially selected features for Ausdata were day, month, hour, dew point, load, and humidity. The best subsets of these features were exhaustively identified as hour and load.

3.3. Proposed Models' Results with Full Set of Features and Selected Features

The results with the full set of features and selected features are presented in Tables 1 and 2 for Czech data and Ausdata, respectively.

It was observed that, when the hour-ahead load forecasting results obtained with the Czech Republic dataset were examined, although the highest MAPE and MAE values were obtained with the KNN method, both when all features were used and when a feature selection was performed, the lowest values for these metrics were obtained with the suggested hourly ensemble method. Whereas the MAPE and MAE values closest to the suggested method were obtained using the SVR method when all features were used, the closest MAPE and MAE values were obtained with the CNN_1 and CNN_2 (13) ensemble method when feature selection was made.

The results obtained with the Australia dataset are slightly better compared with those obtained with the Czech dataset. The primary reason for this is considered to be the fact that this dataset has a 30 min resolution (48 measurements per day). Here, the highest MAPE and MAE values were again obtained with the KNN method, both when all features were used and for feature selection, whereas the lowest values for these metrics were obtained using the suggested hourly ensemble method. The MAPE and MAE values closest to the suggested method were obtained using the ANN_1 and ANN_2 (11) ensemble method, and the closest MAPE and MAE values obtained with feature selection were also obtained here with the CNN_1 and CNN_2 (13) ensemble method.

Table 1. Czech dataset hour-ahead load forecasting results.

	Whole Set of Features		Reduced Feature Subset	
Model	**MAE**	**MAPE**	**MAE**	**MAPE**
Mean	87.15 ± 0.00%	31.21 ± 0.00%	87.15 ± 0.00%	31.21 ± 0.00%
Random	117.38 ± 0.29%	42.45 ± 0.17%	117.51 ± 0.25%	42.40 ± 0.25%
ANN_1 (1)	5.84 ± 0.93%	2.71 ± 0.63%	4.50 ± 0.11%	2.03 ± 0.26%
ANN_2 (2)	6.32 ± 1.08%	2.99 ± 0.80%	5.05 ± 0.80%	2.22 ± 0.18%
SVR (3)	5.13 ± 0.38%	2.37 ± 0.27%	5.40 ± 0.31%	2.45 ± 0.27%
KNN (4)	36.72 ± 2.41%	16.67 ± 1.95%	23.16 ± 1.12%	10.46 ± 1.05%
MLR (5)	5.30 ± 0.45%	2.52 ± 0.21%	5.49 ± 0.52%	2.63 ± 0.19%
Tree (6)	6.73 ± 0.58%	2.94 ± 0.28%	6.27 ± 0.55%	2.73 ± 0.24%
RNN_1 (7)	14.88 ± 3.24%	6.43 ± 0.66%	6.06 ± 0.80%	2.74 ± 0.66%
RNN_2 (8)	8.93 ± 0.97%	4.00 ± 0.76%	4.49 ± 0.16%	1.99 ± 0.24%
CNN_1 (9)	7.25 ± 0.81%	3.31 ± 0.75%	3.88 ± 0.40%	1.71 ± 0.18%
CNN_2 (10)	8.34 ± 0.25%	3.84 ± 0.47%	4.83 ± 0.53%	2.16 ± 0.21%
1&2 (ANNs) (11)	5.53 ± 0.80%	2.57 ± 0.56%	4.44 ± 0.42%	1.96 ± 0.16%
7&8 (RNNs) (12)	8.93 ± 0.93%	3.98 ± 0.74%	4.48 ± 0.12%	1.98 ± 0.26%
9&10 (CNNs) (13)	6.83 ± 0.62%	3.11 ± 0.57%	3.78 ± 0.41%	1.66 ± 0.17%
Hourly ensemble	4.05 ± 0.37%	1.84 ± 0.24%	3.50 ± 0.38%	1.53 ± 0.17%

Table 2. Ausdata hour-ahead load forecasting results.

	Whole Set of Features		Reduced Feature Subset	
Model	**MAE**	**MAPE**	**MAE**	**MAPE**
Mean	79.07 ± 0.00%	45.89 ± 0.00%	79.07 ± 0.00%	45.89 ± 0.00%
Random	111.68 ± 0.25%	59.19 ± 0.24%	111.45 ± 0.79%	59.07 ± 0.54%
ANN_1 (1)	4.15 ± 0.26%	1.80 ± 0.12%	3.26 ± 0.09%	1.42 ± 0.09%
ANN_2 (2)	4.09 ± 0.40%	1.79 ± 0.24%	3.32 ± 0.05%	1.42 ± 0.06%
SVR (3)	7.13 ± 1.15%	3.13 ± 0.38%	5.49 ± 0.64%	2.42 ± 0.27%
KNN (4)	39.96 ± 2.63%	16.71 ± 0.75%	21.14 ± 1.64%	8.32 ± 0.31%
MLR (5)	3.75 ± 0.06%	1.67 ± 0.09%	3.72 ± 0.08%	1.65 ± 0.09%
Tree (6)	6.66 ± 1.54%	2.72 ± 0.56%	6.09 ± 1.07%	2.49 ± 0.37%
RNN_1 (7)	9.66 ± 1.34%	4.31 ± 0.49%	4.83 ± 0.40%	2.05 ± 0.22%
RNN_2 (8)	7.16 ± 0.39%	3.06 ± 0.15%	3.94 ± 0.31%	1.72 ± 0.09%
CNN_1 (9)	12.06 ± 3.33%	6.27 ± 2.60%	3.07 ± 0.10%	1.33 ± 0.09%
CNN_2 (10)	6.47 ± 0.28%	2.78 ± 0.23%	3.85 ± 0.11%	1.66 ± 0.11%
1&2 (ANNs) (11)	3.78 ± 0.22%	1.65 ± 0.16%	3.08 ± 0.04%	1.33 ± 0.07%
7&8 (RNNs) (12)	6.83 ± 0.41%	2.94 ± 0.13%	3.86 ± 0.29%	1.67 ± 0.10%
9&10 (CNNs) (13)	6.64 ± 0.60%	3.05 ± 0.56%	3.05 ± 0.08%	1.31 ± 0.09%
Hourly ensemble	3.37 ± 0.13%	1.48 ± 0.12%	2.96 ± 0.05%	1.28 ± 0.08%

3.4. Benchmarking

As can be seen from Table 3, a comparison of the present study with some related studies in the literature reveals that the lowest values for MAPE were observed in our study. Although the studies performed with data at other levels were also included in the comparison table, the results differ slightly due to the internal dynamics of the data at each level. According to the table, MAPE results are in a wide range, as the dynamics of energy consumption vary significantly inside the building in studies using loads at the building level [15,21]. The reason for this is that the consumer behavior, which we refer to as the internal dynamics of the data, is different at each data level. Whereas the data at the aggregated level follow a smoother load consumption pattern, the data at the building level have a higher variability and fluctuations due to the consumer load profile in the residence [44]. Among the other studies, ref. [14] bus-load-level data are the closest to the aggregated-level data used in this study in terms of the data structure. However, the MAPE value is 5.23% here as well, which is considerably higher than that of the present study. Although the lowest MAPE value of 1.47% was reached in [24] by utilizing data with a

characteristic comparable to ours, it is considerably higher than the value determined in the present study.

Table 3. Benchmarking study.

Reference	Data Level	Feature Selection Method or Feature Types	Method	MAPE (%)
Present study	Aggregated	Backward-eliminated exhaustive approach	DL-based ensemble	1.28
[14]	Bus load	Selected lagged load hours, temperature, day type, holiday type	ANNs and time-series clustering	5.23 (avr.)
[15]	Building	Previous correlated load data of the same week, day, hour, and quarter-hour load, dew temperature, humidity, and dry-bulb temperature	ANN	1.70–2.5
[21]	Building	Feature set selection and Pearson correlation	DL ensemble	1.87–10.97
[24]	Aggregated	No exogenous features	Ensemble re-forecast	1.47%

4. Conclusions and Future Work

This article presents a proper approach utilizing ensemble and supervised learning techniques in hour-ahead load forecasting that tries to improve dynamic-data-driven smart power system applications. The proposed ensemble approach was thoroughly tested and evaluated by comparing it with the state-of-the-art STLF methods.

For hour-ahead load forecasting with Czech data, SVR performs the best, followed by MLR and ANNs. Deep models perform worse, indicating that simpler models are appropriate as the problem gets simpler. However, the CNN performs better than the RNN in DL methods. This phenomenon is more extreme in Ausdata, where MLR performs the best, followed by ANNs. However, the proposed ensemble method may still provide the best results for both data sets. The difference between mean as well as random guesses becomes more than 20-fold compared to the hourly ensemble, showing the significance of the proposed forecasters and ensembles.

On the other hand, feature selection was realized for each data. Although many feature selection methods have been used in previous ELF studies so far, the proposed backward-eliminated exhaustive approach has not been used on ELF before. It was observed that the process significantly improves the hour-ahead forecasting results by almost 1%. Although the data sets used in this study were different from each other in terms of the time step, they also had very different and various features. It should be noted that the effect of distinctive features on the estimation results is different for each data set due to feature selection. Therefore, making the feature selection special for each dataset while studying ELF will produce more effective results.

A more comprehensive comparison can be made for the proposed ensemble approach with other state-of-the-art methods in future studies. Moreover, implementing or modifying the proposed method for more diverse energy forecasting studies, such as solar energy forecasting, electricity price forecasting, and other load forecasting horizons such as very-short-term and day-ahead, is planned.

Funding: This research received no external funding.

Data Availability Statement: Publicly available datasets were analyzed in this study. These datas can be found here: [https://www.mathworks.com/matlabcentral/fileexchange/31877-electricity-load-forecasting-for-the-australian-market-case-study, https://power.larc.nasa.gov/data-access-viewer/, https://www.ceps.cz/en/all-data].

Conflicts of Interest: The author declares no conflict of interest.

References

1. 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. ._16. [CrossRef]
2. Tiwari, A.; Pindoriya, N.M. Automated Demand Response in Smart Distribution Grid: A Review on Metering Infrastructure, Communication Technology and Optimization Models. *Electr. Power Syst. Res.* **2022**, *206*, 107835. [CrossRef]
3. Godinho, G.C.; Lima, D.A. Security of power supply in hydrothermal systems: Assessing minimum storage requisites for hydroelectric plants. *ELectric Power Syst. Res.* **2020**, *188*, 106523. [CrossRef]
4. Saviozzi, M.; Massucco, S.; Silvestro, F. Implementation of advanced functionalities for Distribution Management Systems: Load forecasting and modeling through Artificial Neural Networks ensembles. *Electr. Power Syst. Res.* **2019**, *167*, 230–239. [CrossRef]
5. Hou, H.; Liu, C.; Wang, Q.; Wu, X.; Tang, J.; Shi, Y.; Xie, C. Review of load forecasting based on artificial intelligence methodologies, models, and challenges. *Electr. Power Syst. Res.* **2022**, *210*, 108067. [CrossRef]
6. Quilumba, F.L.; Lee, W.; Huang, H.; Wang, D.Y.; Szabados, R.L. Using Smart Meter Data to Improve the Accuracy of Intraday Load Forecasting Considering Customer Behavior Similarities. *IEEE Trans. Smart Grid* **2015**, *6*, 911–918. [CrossRef]
7. Zhang, P.; Wu, X.; Wang, X.; Bi, S. Short-term load forecasting based on big data technologies. *CSEE J. Power Energy Syst.* **2015**, *1*, 59–67. [CrossRef]
8. Guo, Z.; Zhou, K.; Zhang, X.; Yang, S. A deep learning model for short-term power load and probability density forecasting. *Energy* **2018**, *160*, 1186–1200. [CrossRef]
9. Kim, Y.; gu Son, H.; Kim, S. Short term electricity load forecasting for institutional buildings. *Energy Rep.* **2019**, *5*, 1270–1280. [CrossRef]
10. Liu, N.; Tang, Q.; Zhang, J.; Fan, W.; Liu, J. A hybrid forecasting model with parameter optimization for short-term load forecasting of micro-grids. *Appl. Energy* **2014**, *129*, 336–345. [CrossRef]
11. Velasco, L.C.P.; Arnejo, K.A.S.; Macarat, J.S.S. Performance analysis of artificial neural network models for hour-ahead electric load forecasting. *Procedia Comput. Sci.* **2022**, *197*, 16–24. [CrossRef]
12. Amral, N.; Ozveren, C.S.; King, D. Short term load forecasting using Multiple Linear Regression. In Proceedings of the 42nd International Universities Power Engineering Conference, Brighton, UK, 4–6 September 2007; pp. 1192–1198. [CrossRef]
13. Satish, B.; Swarup, K.; Srinivas, S.; Rao, A.H. Effect of temperature on short term load forecasting using an integrated ANN. *Electr. Power Syst. Res.* **2004**, *72*, 95–101. [CrossRef]
14. Panapakidis, I.P. Clustering based day-ahead and hour-ahead bus load forecasting models. *Int. J. Electr. Power Energy Syst.* **2016**, *80*, 171–178. [CrossRef]
15. Dagdougui, H.; Bagheri, F.; Le, H.; Dessaint, L. Neural network model for short-term and very-short-term load forecasting in district buildings. *Energy Build.* **2019**, *203*, 109408. [CrossRef]
16. Li, Y.; Che, J.; Yang, Y. Subsampled support vector regression ensemble for short term electric load forecasting. *Energy* **2018**, *164*, 160–170. [CrossRef]
17. Dong, H.; Gao, Y.; Fang, Y.; Liu, M.; Kong, Y. The short-term load forecasting for special days based on bagged regression trees in qingdao, China. *Comput. Intell. Neurosci.* **2021**, *2021*, 3693294. [CrossRef] [PubMed]
18. Dhaval, B.; Deshpande, A. Short-term load forecasting with using multiple linear regression. *Int. J. Electr. Comput. Eng.* **2020**, *10*, 3911. [CrossRef]
19. Fan, G.F.; Guo, Y.H.; Zheng, J.M.; Hong, W.C. Application of the Weighted K-Nearest Neighbor Algorithm for Short-Term Load Forecasting. *Energies* **2019**, *12*, 916. [CrossRef]
20. Jian Zheng.; Cencen Xu.; Ziang Zhang.; Xiaohua Li. Electric load forecasting in smart grids using Long-Short-Term-Memory based Recurrent Neural Network. In Proceedings of the 2017 51st Annual Conference on Information Sciences and Systems (CISS), Baltimore, MD, USA, 22–24 March 2017; pp. 1–6. [CrossRef]
21. Chitalia, G.; Pipattanasomporn, M.; Garg, V.; Rahman, S. Robust short-term electrical load forecasting framework for commercial buildings using deep recurrent neural networks. *Appl. Energy* **2020**, *278*, 115410. [CrossRef]
22. Bendaoud, N.; Farah, N.; Ahmed, S.B. Applying load profiles propagation to machine learning based electrical energy forecasting. *Electr. Power Syst. Res.* **2022**, *203*, 107635. [CrossRef]
23. Li, L.; Ota, K.; Dong, M. Everything is Image: CNN-based Short-Term Electrical Load Forecasting for Smart Grid. In Proceedings of the 14th International Symposium on Pervasive Systems, Algorithms and Networks 2017 11th International Conference on Frontier of Computer Science and Technology 2017 Third International Symposium of Creative Computing (ISPAN-FCST-ISCC), Exeter, UK, 21–23 June 2017; pp. 344–351. [CrossRef]
24. Kaur, A.; Pedro, H.T.; Coimbra, C.F. Ensemble re-forecasting methods for enhanced power load prediction. *Energy Convers. Manag.* **2014**, *80*, 582–590. [CrossRef]
25. Feng, C.; Zhang, J. Assessment of aggregation strategies for machine-learning based short-term load forecasting. *Electr. Power Syst. Res.* **2020**, *184*, 106304. [CrossRef]
26. Eskandari, H.; Imani, M.; Moghaddam, M.P. Convolutional and recurrent neural network based model for short-term load forecasting. *Electr. Power Syst. Res.* **2021**, *195*, 107173. [CrossRef]
27. Rafati, A.; Joorabian, M.; Mashhour, E. An efficient hour-ahead electrical load forecasting method based on innovative features. *Energy* **2020**, *201*, 117511. [CrossRef]

28. Du, L.; Zhang, L.; Wang, X. Spatiotemporal Feature Learning Based Hour-Ahead Load Forecasting for Energy Internet. *Electronics* **2020**, *9*, 196. [CrossRef]
29. Aly, H.H. A proposed intelligent short-term load forecasting hybrid models of ANN, WNN and KF based on clustering techniques for smart grid. *Electr. Power Syst. Res.* **2020**, *182*, 106191. [CrossRef]
30. Yu, F.; Wang, L.; Jiang, Q.; Yan, Q.; Qiao, S. Self-Attention-Based Short-Term Load Forecasting Considering Demand-Side Management. *Energies* **2022**, *15*. 4198. [CrossRef]
31. Chen, J.; Li, T.; Zou, Y.; Wang, G.; Ye, H.; Lv, F. An Ensemble Feature Selection Method for Short-Term Electrical Load Forecasting. In Proceedings of the 2019 IEEE 3rd Conference on Energy Internet and Energy System Integration, Changsha, China, 8–10 November 2019; pp. 1429–1432. [CrossRef]
32. Bui, V.; Pham, T.L.; Kim, J.; Jang, Y.M.; et al. RNN-based deep learning for one-hour ahead load forecasting. In Proceedings of the 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), Fukuoka, Japan, 19–21 February 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 587–589.
33. Laouafi, A.; Mordjaoui, M.; Dib, D. One-hour ahead electric load forecasting using neuro-fuzzy system in a parallel approach. In *Computational Intelligence Applications in Modeling and Control*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 95–121.
34. Fallah, S.N.; Deo, R.C.; Shojafar, M.; Conti, M.; Shamshirband, S. Computational intelligence approaches for energy load forecasting in smart energy management grids: State of the art, future challenges, and research directions. *Energies* **2018**, *11*, 596. [CrossRef]
35. Henderi, H.; Wahyuningsih, T.; Rahwanto, E. Comparison of Min-Max normalization and Z-Score Normalization in the K-nearest neighbor (kNN) Algorithm to Test the Accuracy of Types of Breast Cancer. *Int. J. Inform. Inf. Syst.* **2021**, *4*, 13–20. [CrossRef]
36. De Myttenaere, A.; Golden, B.; Le Grand, B.; Rossi, F. Mean Absolute Percentage Error for regression models. *Neurocomputing* **2016**, *192*, 38–48. [CrossRef]
37. Qi, J.; Du, J.; Siniscalchi, S.M.; Ma, X.; Lee, C.H. On mean absolute error for deep neural network based vector-to-vector regression. *IEEE Signal Process. Lett.* **2020**, *27*, 1485–1489. [CrossRef]
38. Malki, H.A.; Karayiannis, N.B.; Balasubramanian, M. Short-term electric power load forecasting using feedforward neural networks. *Expert Syst.* **2004**, *21*, 157–167. [CrossRef]
39. 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, Cancun, Mexico, 18–21 December 2017; pp. 511–516. [CrossRef]
40. Mitchell, G.; Bahadoorsingh, S.; Ramsamooj, N.; Sharma, C. A comparison of artificial neural networks and support vector machines for short-term load forecasting using various load types. In Proceedings of the 2017 IEEE Manchester PowerTech, Manchester, UK, 18–22 June 2017; pp. 1–4. [CrossRef]
41. Mori, H.; Kosemura, N. Optimal regression tree based rule discovery for short-term load forecasting. In Proceedings of the 2001 IEEE Power Engineering Society Winter Meeting. Conference Proceedings, Columbus, OH, USA, 28 January–1 February 2001; Volume 2, pp. 421–426. [CrossRef]
42. Bouktif, S.; Fiaz, A.; Ouni, A.; Serhani, M.A. Optimal deep learning lstm model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches. *Energies* **2018**, *11*, 1636. [CrossRef]
43. 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]
44. Khan, A.N.; Iqbal, N.; Rizwan, A.; Ahmad, R.; Kim, D.H. An Ensemble Energy Consumption Forecasting Model Based on Spatial-Temporal Clustering Analysis in Residential Buildings. *Energies* **2021**, *14*, 3020. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.