Peer-reviewed version available at *Energies* **2020**, *13*, 2242; doi:10.3390/en13092242

Article

Energy Demand Forecasting Using Deep Learning: Application to the French Grid

Alejandro .J. del Real [1,*], Fernando Dorado [2] and Jaime Durán [3]

[1] University of Seville; adelreal@us.es
[2] IDENER; fernando.dorado@idener.es
[3] IDENER; jaime.duran@idener.es
* Correspondence: adelreal@us.es;

Abstract: This paper investigates the use of deep learning techniques to perform energy demand forecasting. Specifically, the authors have adapted a deep neural network originally thought for image classification and composed of a convolutional neural network (CNN) followed by a multilayered fully connected artificial neural network (ANN). The convolutional part of the network was fed with a grid of temperature forecasting data distributed in the area of interest in order to extract a featured temperature. The subsequent ANN is then fed with this calculated temperature along with other data related to the timing of the forecast. The proposed structure was first trained and then used in a real setting aimed to provide the French energy demand forecast using ARPEGE forecasting weather data. The results show that the performance of this approach is in the line of the performance provided by the reference RTE subscription-based service, which opens the possibility to obtain high accuracy forecasting using widely accessible deep learning techniques through open-source machine learning platforms.

Keywords: energy; demand; forecasting; deep; learning; machine; convolutional; artificial; neural; networks.

1. Introduction

The forecasting of demand plays an essential role in the electric power industry. Thus, there is a great variety of methods raging from short- (minute order) to long-term (week order) electricity demand prediction, while considering microscopic (individual consumer) to macroscopic (country level) aggregation level. This paper is framed in the mid-term (hour order) macroscopic power forecast.

So far, the researchers agree in that the electrical demand arises from complex interactions between multiple personal, corporative and socio-economic factors [1]. All these sources make power demand forecasting difficult. Indeed, an ideal model able to forecast the power demand with the highest possible level of accuracy would require access to virtually infinite data sources in order to feed such a model with all the relevant information. Unfortunately, both the availability of the data and the associated computational burden make researchers to investigate approximate models fed with partial input information.

Within this framework, the prediction of the power consumption has been tackled from different perspectives and using different forecasting methodologies so far. Indeed, there is a rich state of the art of methods which, according to the classification made by the authors in [1], can be divided into the next main types:

- Statistical models: Purely empirical models where inputs and outputs are correlated using statistical inference methods, such as:
    - Cointegration analysis and ARIMA;
    - Log-linear regression models;

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 March 2020 doi:10.20944/preprints202003.0158.v1

Peer-reviewed version available at *Energies* **2020**, *13*, 2242; doi:10.3390/en13092242

    - Combined bootstrap aggregation (bagging) ARIMA and exponential smoothing.

  Despite some authors [2] report substantial improvements of the forecast accuracy of the demand for energy end-use services in both developed and developing countries, the related models require to excel on sophisticated statistical methods which are case-dependent on the application. These facts hinder statistical models to form an affordable and consistent basis for a general power demand approach.

- Grey models: They combine a partial theoretical structure with empirical data to complete the structure. When compared to purely statistical models, grey models would require only a limited amount of data to infer the behavior of the electrical system. Therefore, grey models can deal with partially known information through generating, excavating and extracting useful information from what is available. In return, the construction of the partial theoretical structure required by grey models is resource-demanding in terms of modelling and thus, the cost of an accurate grey model for a particular application is usually high.
- Artificial Intelligence models: Traditional machine learning are data-driven techniques used to model complex relationships between input and output. Despite the basis of machine learning is mostly statistical, the current availability of open platforms to easily design and train models greatly contribute to the access of this technology. This fact, along with the high performance achieved by well designed and trained machine learning models, provides an affordable tool for power demand forecasting.

  However, traditional machine learning models require a pre-processing step to perform the extraction of the main features of the input data, which generally demands a manual construction of such module. To overcome this requirement (as depicted in next Figure 1) modern deep learning techniques [3] may integrate feature learning and model construction in just one model. According to this approach, each layer's representation is transferred from the original input sets into more abstracted representations in order to allow the subsequent model layers to find the inherent structures.

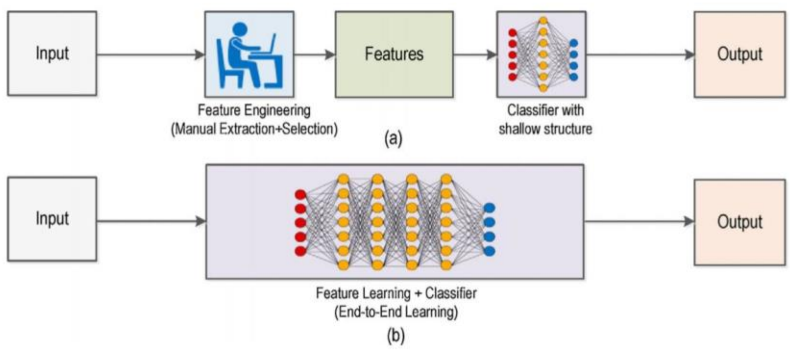


Figure 1. Comparison between traditional machine learning models (a) requiring manual feature extraction and modern deep learning structures (b) which can automate all the feature and training process in an end-to-end learning structure.

Despite modern deep learning techniques have attracted the attention of many researches in a myriad of areas, the relative high number of publications related to power demand forecasting mostly use traditional machine learning approaches such as artificial neural networks. Thus, the authors of this paper found a promising topic related to the application of modern deep learning structures to the problem of power demand forecasting.

More specifically, this paper describes the novel use of a particular deep neural network structure composed of a convolutional neural network followed by an artificial neural network (widely used in image classification) for the forecasting of power demand with a limited number of information sources available. As described in the next sections, this deep learning structure is an effective approach to deal with the power demand time series forecasting problem with multiples inputs variables, complex nonlinear relationships, and missing data.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 March 2020 doi:10.20944/preprints202003.0158.v1

Peer-reviewed version available at Energies 2020, 13, 2242; doi:10.3390/en13092242

Furthermore, the proposed deep learning structure has been applied to the particular problem of French power demand in a real setting approach. Next section comprehensively describes the materials and data sources used to reach such an end, so other researchers can replicate and adapt the work of this paper to other power demand forecasting applications. As shown in the later section related to the results, the performance of this approach equals (if not outperforms) the reference RTE (Réseau de Transport d'Electricité) French power demand forecast subscription-based service.

2. Materials and Methods

As described in next sections 2.2.1 and 2.2.2, the main source of input information for predicting the power demand is the weather forecast. Given that weather forecast providers release their predictions only at certain times, a real problem setting must be capable of predicting the power demand (in this case, the French power demand) a day ahead (D+1) based on the weather forecasts published at day D. In the setting done in this paper and as depicted in the next Figure 2, the power demand forecast model is run at 08.00 every day D with the most recent weather forecast information available (released at 00.00) so it provides a prediction of the energy demand during day D+1.

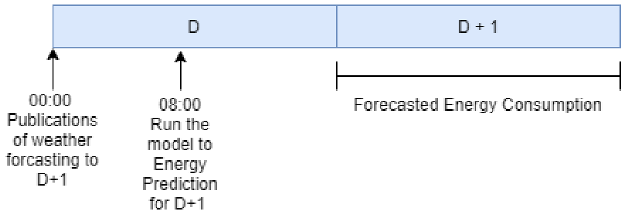


Figure 2. Real setting of the energy demand forecasting problem

2.1. Data Analysis

2.2.1. Power demand data

For this paper's application, the historical data of French energy consumption was downloaded from the official RTE website [4], which provides data from 2012. A first analysis of this data results in the next Figure 3, which shows a clear seasonal pattern on the energy demand.

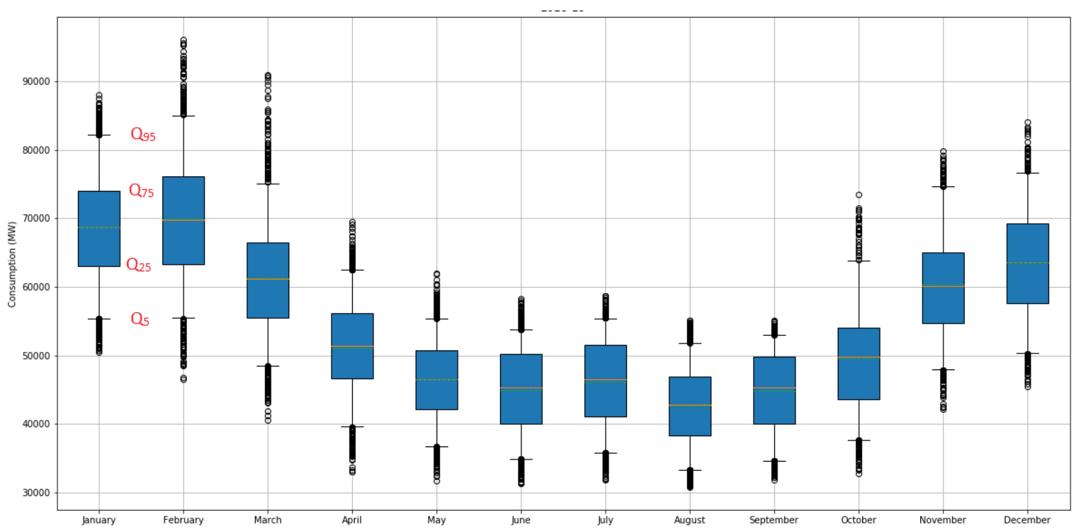


Figure 3. Monthly French energy demand during the period 2018-2019. Q_x indicates the X data percentiles. The colored lines within the Q_{25} and Q_{75} quartiles box represents the median (orange line) and the mean (dashed green line). Points below and above Q_{25} and Q_{75} quartiles are shown as well.

The strong seasonal pattern on the energy demand is further backed by next Figure 4 provided by RTE, which depicts the correlation between energy consumption and temperature. As shown, an average variation of 1º C during winter on the entire territory can lead to a variation of around 2,500 MW on the peak consumption (equivalent to the average winter consumption of about 2 million homes) [4]. In summer, the temperature gradient related to air conditioning is about 400 MW per °C.

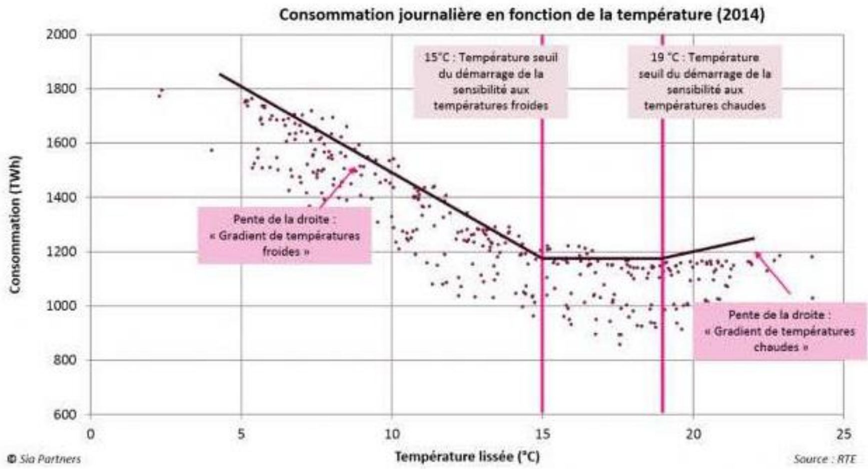


Figure 4. Correlation between energy consumption and temperature as provided by RTE.

2.2.2. Weather forecast data

For this paper's application the weather forecast historical data was collected from ARPEGE (Action de Recherche Petite Echelle Grande Echelle), which is a main numerical weather prediction provider over the Europe-Atlantic domain.

As indicated in the ARPEGE documentation [5], the initial conditions of this model are based on a 4-dimensional variational assimilation (4D-Var) that incorporates a very large and varied amount of conventional observations (radio sounding, airplane measurements, ground stations, ships, buoys, etc) and also from remote sensing (ATOVS, SSMI/S, AIRS, IASI, CRIS, ATMs, SEVIRI, GPS sol, GPS satellite, SATOB, etc...).

Despite ARPEGE provides multiple forecast data related to weather (pressure, wind, temperature, humidity) with a full resolution of 0.1º, the authors of this paper found that just the temperature data with a 1º resolution (as shown in Figure 5) was sufficient for the purpose of energy demand forecasting while maintaining easy-to-handle data sets. This finding is also aligned with the strong correlation of energy demand and temperature described in the preceding section 2.2.1.

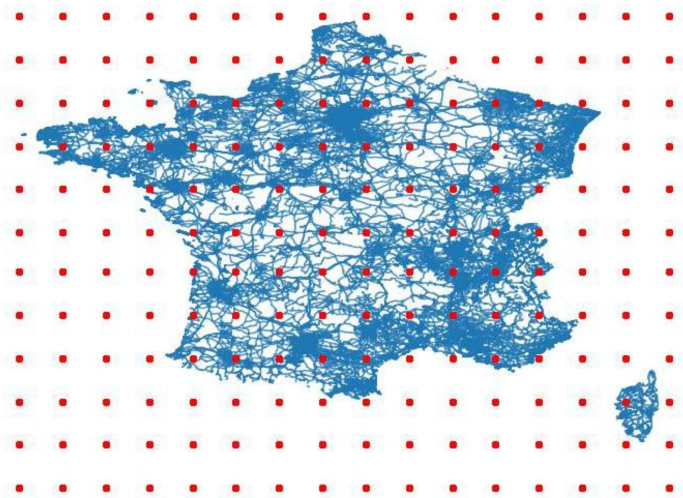


Figure 5. Locations of temperature forecasting with a resolution of 1º along France .

Peer-reviewed version available at Energies 2020, 13, 2242; doi:10.3390/en13092242

2.2. Data preparation

First, the historical datasets related to the French energy consumption (section 2.2.1) and forecasted temperature (section 2.2.2) were pre-processed to eliminate outliers, clean unwanted characters and to filter null data. Then, and as usual practice when training machine learning models, the resulting data was divided into three datasets: training, validation and test.

Despite the historical French consumption data provided by RTE dates back to 2012, the authors of this paper had only access to the ARPEGE historical weather forecast data in the period spanning from 2018-10-01 to 2019-30-09. Although a wider availability of historical weather forecast data would had benefitted the generalisation capability of the resulting machine learning model, the data set available still covered a whole year, so the seasonal influence was fully captured. Additioanlly, the authors randomly extracted eight full days from the original data set in order to further test the generalisation performance of the model (as depicted in Figure 6 and further discussed in section 3). This way, the remaining data sets were randomly divided as follows:

- Training Dataset (80% of the data): The sample of data used to fit the model.
- Validation Dataset (10% of the data): The sample of data used to provide an unbiased evaluation of a model fit on the training dataset while tuning model hyperparameters.
- Test Dataset (10% of the data): The sample of data used to provide an unbiased evaluation of a final model fit on the training dataset.

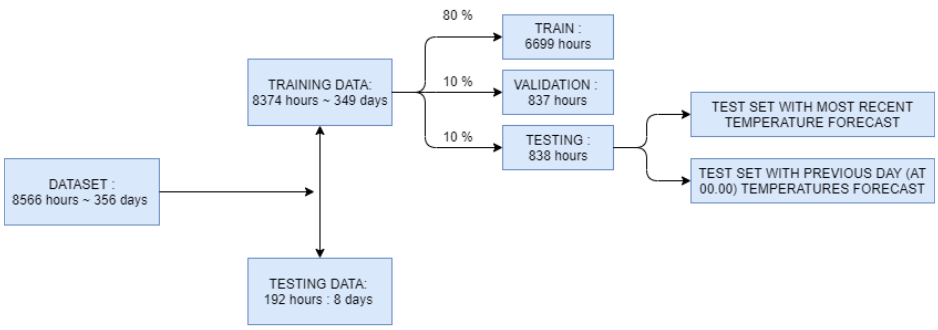


Figure 6. Division of the original dataset (365 days) into testing and training data. The testing data was used as a complementary means to further analyse the generalisation performance of the resulting model. The remaining training data was divided as usual: 80% train, 10% validation and 10% testing.

2.3. Deep learning architecture

The deep learning architecture used in this paper (as shown in Figure 7) resembles those structures widely used in image classification: a convolutional neural network followed by an artificial neural network. The novelty of this paper is not the deep neural network itself but its application to the macroscopic forecast of energy demand. In fact, the aforementioned deep learning architecture was originally thought to automatically infer features from an input image (made of pixels) in order to subsequently classify such image in a certain category attending to the inferred specific features.

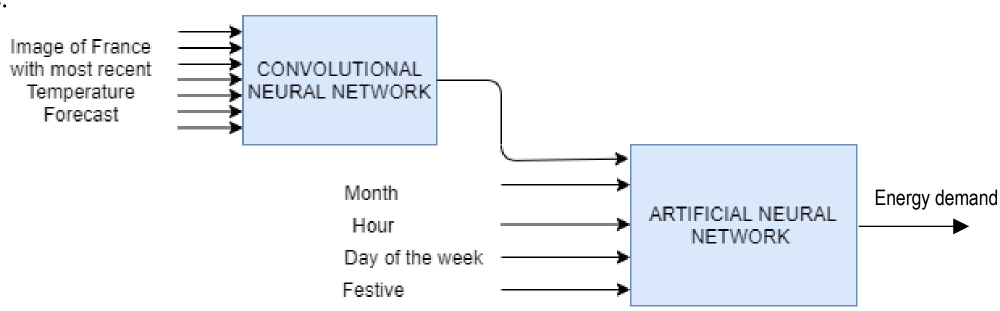


Figure 7. Deep learning structure composed by a convolutional neural network followed by and artificial neural network and adapted to the energy demand forecasting problem.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 March 2020 doi:10.20944/preprints202003.0158.v1

Peer-reviewed version available at Energies 2020, 13, 2242; doi:10.3390/en13092242

As adapted to this paper's application, the convolutional network receives the temperature forecasts of multiple locations within the area of interest (in this case, France) instead of an image. Still, the convolutional network extracts a "feature" of such input, which may be understood as a representative temperature of France automatically calculated attending to the individual contribution of each location to the aggregated energy consumption. For instance, the temperature locations close to large consumption sites (such as highly populated areas) would be automatically assigned a larger weight when compared to other less populated areas.

As also discussed in the introductory section 1, the advantage of the proposed deep learning structure with respect to traditional (and less sophisticated) machine learning structures is that this feature extraction is implicit to the model and thus, there is no need to manually design the feature extraction step.

As also shown in Figure 7, the artificial neural network receiving the "featured" temperature from the convolutional network is also fed with additional information found to highly influence the energy demand as well, namely:

- Month: a number from 1 to 12.
- Hour: a number from 1 to 24.
- Day of the week: a number from 1 to 7.
- Holiday: True (1) or false (0).

2.3.1. Convolutional neural network

As described by the authors in [6], Convolutional Neural Networks or CNNs, are a specialized kind of neural network for processing data that has a known, grid-like topology. The name of "convolutional neural network" indicates that the network employs a mathematical operation called convolution, which is a specialised kind of linear operation. Traditional neural network layers use matrix multiplication to model the interaction between each input unit and each output unit. This means every output unit interacts with every input unit. Convolutional networks, however, typically have sparse interactions (also referred to as sparse connectivity or sparse weights). This characteristic provides the CNNs a series of benefits, namely:

- They use a fewer number of parameters (weights) with respect to fully connected networks.
- They are designed to be invariant in object position and distortion of the scene when used to process images, which is a property shared when they are fed with other kind of inputs as well.
- They can automatically learn and generalise features from the input domain.

Attending to these benefits, this paper uses a CNN to extract a representative temperature of the area of interest (France) from the historical temperature forecast data as explained before. For the sake of providing an easy replication of the results by other researcher, next bullet points lists the main characteristics of the CNN designed for this paper:

- 2D convolutional layer. This layer creates a convolution kernel that is convolved with the layer into to produce a tensor of outputs. It was set with the next parameters:
    - o Filters: 8 Integers, the dimensionality of the output space.
    - o Kernel Size: (2,2). A list of 2 integers, specifying the height and width of the 2D convolution window.
    - o Strides: (1,1). A list of 2 integers, specifying the stride of the convolution along with the height and width.
    - o Activation Function: Rectified Linear Unit (ReLU)
    - o Padding: Apply padding to input (if needed) so that input image gets fully covered by filter and stride specified.
- AveragePooling2D: Average Pooling operation for special data. It was set with the next parameters:
    - o Pool Size: (2,2). Factors by which to downscale.
    - o Strides: (1,1)
- Flatten: Flattens the input.
- Fully Connected Network providing the featured output temperature:

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 March 2020 doi:10.20944/preprints202003.0158.v1

Peer-reviewed version available at *Energies* **2020**, *13*, 2242; doi:10.3390/en13092242

- Layer 1: 64 neurons, activation function: ReLU.
- Layer 2: 24 neurons, activation function: ReLU.
- Layer 3: 1 neuron, activation function: ReLU.

2.3.4. Artificial neural network.

As shown in Figure 7, the CNN output was connected to a fully connected multilayer Artificial Neural Network (ANN) with the following structure.

- Layer 1: 256 neurons, activation function: ReLU.
- Layer 2: 128 neurons, activation function: ReLU.
- Layer 3: 64 neurons, activation function: ReLU.
- Layer 4: 32 neurons, activation function: ReLU.
- Layer 5: 16 neurons, activation function: ReLU.
- Layer 6: 1 neuron, activation function: ReLU.

The weights of all layers were initialised following a normal distribution with mean 0.1 and standard destination 0.05.

2.4. Training

The training process of the proposed deep neural network was aimed to adjust its internal parameters (resembling a mathematical regression) so the structure is able to correlate its output (the French energy demand forecast) with respect to its inputs.

What separates deep learning from a traditional regression problem is the handling of the generalisation error, also called validation error. Here, the generalisation error is defined as the expected value of the error when the deep learning structure is shown new input. Typically, the usual approach is to estimate the generalisation error by measuring its performance on the validation set of examples that were collected separately from the training set. The factors determining how well a deep learning algorithm performs are its ability to:

- Make the training error as low as possible while
- Making the gap between the training and validation errors as low as possible as well.

The tradeoff of these factors results in a deep neural network structure either underfitted or overfitted. In order to prevent overfitting, the usual approach is to update the learning algorithm to encourage the network to keep the weights small. This is called weight regularisation and it can be used as a general technique to reduce overfitting of the training dataset and improve the generalization of the model.

In the model used in this paper the authors used the so-called *L2 regularisation* in order to reduce the validation error. This regularization strategy drives the weights closer to the origin by adding a regularization term to the objective function. L2 regularisation adds the sum of the square of the weights to the loss function [7].

The rest of the training parameters were selected as follows:

- Batch Size: 100. The number of training examples in one forward/backward pass. The higher the batch size, the more memory space needed.
- Epochs: 30.000. One forward pass and one backward pass of all the training examples .
- Learning Rate: 0.001. Determines the step size at each iteration while moving toward a minimum of a loss function.
- β1 parameter: 0.9. The exponential decay rate for the first moment estimates (momentum).
- β2 parameter: 0.99. The exponential decay rate for the first moment estimates (RMSprop).
- Loss function: Mean Absolute Percentage Error .

3. Results

The error metric used in this paper is the Mean Absolute Percentage Error (MAPE):

$$\text{MAPE}\ (\%) = \frac{100}{n}\sum\left|\frac{y-\hat{y}}{y}\right|, \quad (1)$$

 doi:10.20944/preprints202003.0158.v1

Peer-reviewed version available at Energies 2020, 13, 2242; doi:10.3390/en13092242

where y is the reference measure (in our case, the real energy demand) and $\hat{y}$ is the estimated measure (in our case, the forecasted energy demand).

Once the deep learning structure proposed in this paper was trained with the training data set shown in Figure 6 and its output was tested against the real French energy demand, the authors calculated the MAPE shown in Table 1 below. As an additional performance metric, the authors calculated the MAPE of the reference RAE energy demand forecast, which has been also included in Table 1 for comparison.

Table 1. Performance comparison metrics

Model	MAPE (%)
Deep learning network	1.4933
RTE forecast service	1.4940

These outcomes are also presented in the graphical form provided in next Figure 8.

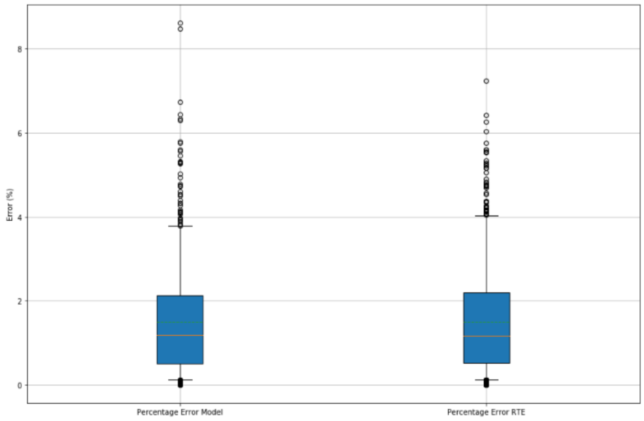


Figure 8. MAPE distribution provided by the deep learning structure proposed in this paper and the RTE subscription-based service.

An also-interesting measure of the performance of the proposed structure is the MAPE monthly distribution along a full year as show in next Figure 9 and Table 2.

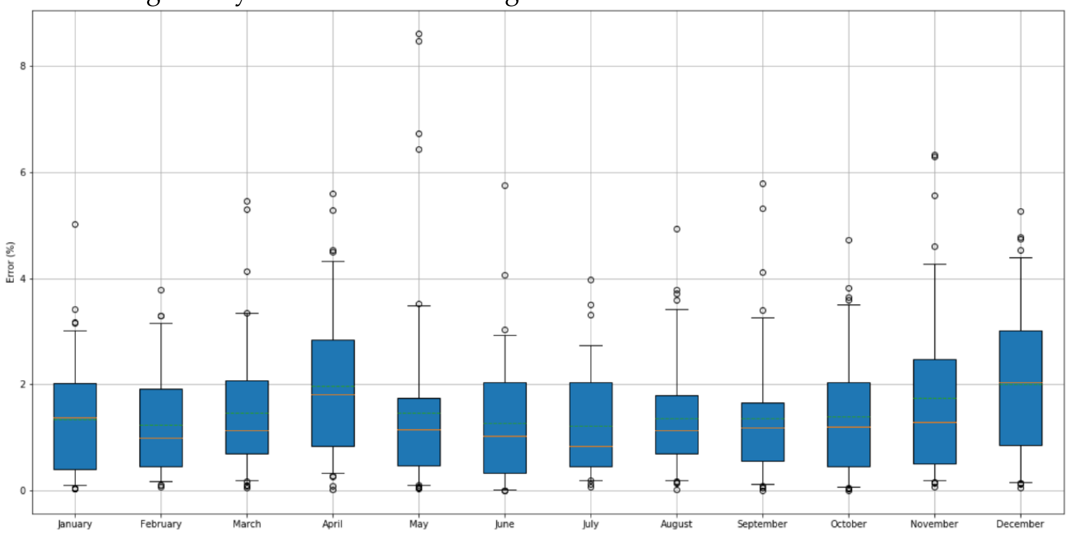


Figure 9. MAPE specific monthly metrics over a full year as provided by the proposed deep neural network.

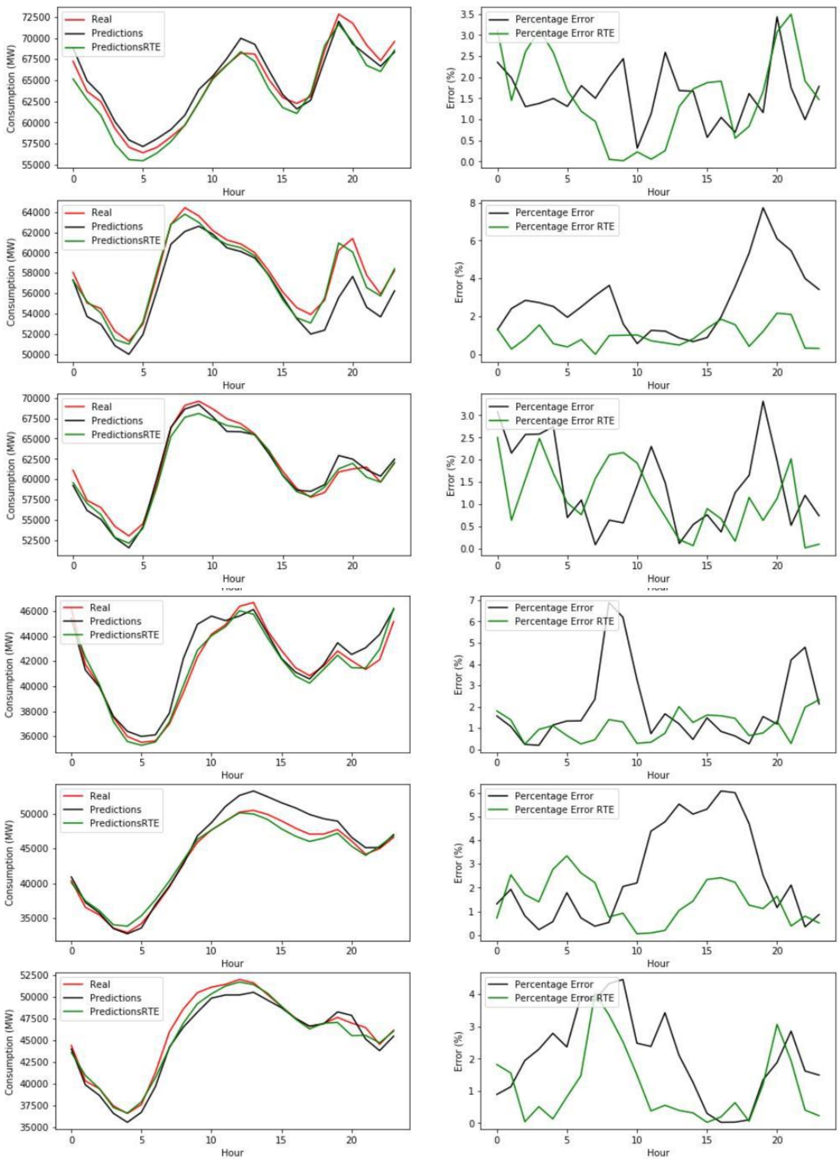
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 March 2020 doi:10.20944/preprints202003.0158.v1

Peer-reviewed version available at Energies 2020, 13, 2242; doi:10.3390/en13092242

Table 2. Performance comparison of monthly MAPE metrics.

Month	MAPE (%) provided by the proposed deep learning structure	MAPE(%) provided by the RTE subscription-based service
January	1.3542	1.5249
February	1.2424	1.5078
March	1.4667	1.8752
April	1.9774	1.4599
May	1.4718	1.5619
June	1.2693	1.3278
July	1.2318	1.3025
August	1.3592	1.4180
September	1.3575	1.2333
October	1.3964	1.4759
November	1.7511	1.5558
December	2.0131	1.6306

Lastly, next Figure 10 depicts the results for the testing of the performance of the forecast provided by the proposed deep neural network on the eight full days extracted from the original data.



doi:10.20944/preprints202003.0158.v1

Peer-reviewed version available at Energies 2020, 13, 2242; doi:10.3390/en13092242

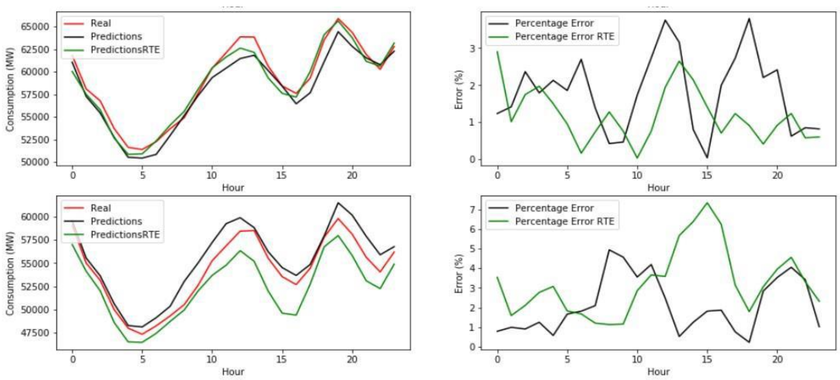


Figure 10. Performance of the forecast provided by the proposed deep neural network on the eight full days extracted from the original data.

4. Discussion

In this paper, the authors present the adaptation of a deep neural network structure commonly used for image classification applied to the forecast of energy demand. In particular, the structure was trained for the French energy grid.

The results show that the performance of the proposed structure competes with the results provided by the RTE subscription-based reference service. Specifically, the overall MAPE metric of the proposed approach delivers an error of 1.4933%, which is slightly better to the 1.4940% metric obtained with the RTE forecast data.

When analysed in a monthly basis, the errors are uniformly distributed through the year despite there are noticeable increments during late autumn and winter season. This fact is also in accordance with the reference RTE forecast data and may be due to the intermittency of the energy consumption profile observed when French temperatures are low.

Also, the proposed deep neural network was tested against eight full days randomly selected from the original dataset in order to provide an additional measure of generalisation performance. On the one hand and as shown in Figure 10, the errors are uniformly distributed along the selected days. On the other hand, the predictions provided by this paper are quite similar to those predictions provided by the reference RTE subscription service and also are aligned with the overall MAPE metrics. These results indicate that the proposed neural network structure is well designed and trained and that it generalises as expected.

The performance achieve in this paper is a promising result for those researchers within the electrical energy industry requiring accurate energy demand forecasting at multiple levels (both temporal and geographical). Indeed, and despite that the application of this paper is dedicated to the French macroscopic energy demand problem, the flexibility of the proposed deep neural network and the wide availability of open platforms for its design and training make the proposed approach an accessible an easy-to-implement project. To further facilitate the replication of this paper by other researchers in this area, the authors have included detailed information about the topology and design of the proposed structure.

Author Contributions: "Conceptualization, A.J. del Real; software, F. Dorado.; validation, A.J. del Real and J. Durán; writing—original draft preparation, A.J. del Real; writing—review and editing, A.J. del Real, F. Dorado and J. Durán; supervision, A.J. del Real. All authors have read and agreed to the published version of the manuscript."

References

1. H. Hu, L. Wang, L. Peng and Y.-R. Zeng., Effective energy consumption forecasting using enhanced bagged echo state network. *Energy* 2020, *Volume 193*, pp 1167-1178.
2. E. M. Oliveira, F. Luiz and C. Oliveira, Forecasting mid-long term electric energy consumption through bagging ARIMA and exponential smoothing methods. *Energy* 2018, *Volume 144*, pp 776-778.

3. J. Wang, Y. Ma, L. Zhang, R. X. Gao and D. Wu, Deep learning for smart manufacturing: Methods and applications. *Journal of Manufacturing Systems* 2018, *Volume* 48, pp 144-156.
4. RTE web site. Available online: https://www.rte-france.com/
5. ARPEGE web site. Available online: https://www.weatheronline.co.uk
6. I. Goodfellow, Y. Bengio and A. Courville, Convolutional Neural Networks. *Deep Learning* 2017, pp 800
7. J. Browlee, Machine learning mastery. Available online: https://machinelearningmastery.com/weight-regularization-to-reduce-overfitting-of-deep-learning-models/ (Accessed on 03 February 2020)
8. Title of Site. Available online: URL (accessed on Day Month Year).