Energy 243 (2022) 123060

Contents lists available at ScienceDirect

Energy

journal homepage: www.elsevier.com/locate/energy

Electricity consumption forecasting based on ensemble deep learning with application to the Algerian market

D. Hadjout [a, 1], J.F. Torres [a, 1], A. Troncoso [a], A. Sebaa [b, c], F. Martínez-Álvarez [a, *]

[a] Data Science and Big Data Lab, Universidad Pablo de Olavide, ES-41013, Seville, Spain
[b] LIMED Laboratory, Faculty of Exact Sciences, University of Bejaia, 06000, Bejaia, Algeria
[c] Higher School of Sciences and Technologies of Computing and Digital, Amizour, 06000, Bejaia, Algeria

ARTICLE INFO

Article history:
Received 21 May 2021
Received in revised form
22 December 2021
Accepted 28 December 2021
Available online 31 December 2021

Keywords:
Time series
Forecasting
Deep learning
Ensemble learning

ABSTRACT

The economic sector is one of the most important pillars of countries. Economic activities of industry are intimately linked with the ability to meet their needs for electricity. Therefore, electricity forecasting is a very important task. It allows for better planning and management of energy resources. Several methods have been proposed to forecast energy consumption. In this work, to predict monthly electricity consumption for the economic sector, we develop a novel approach based on ensemble learning. Our approach combines three models that proved successful in the field, namely: Long Short Term Memory and Gated Recurrent Unit neural networks, and Temporal Convolutional Networks. The experiments have been conducted with almost 2000 clients and 14 years of monthly electricity consumption from Bejaia, Algeria. The results show that the proposed ensemble models achieve better performance than both the company's requirements and the prediction of the traditional individual models. Finally, statistical tests have been carried out to prove that significance of the ensemble models developed.

1. Introduction

It is well known that electricity cannot be stored efficiently, especially in large quantities. Thus, it must be consumed in real-time as it is generated. Therefore, it is important to avoid significantly exceeding the required amount of electricity (the load plus the losses) [1].

The distribution subsidiary orders electricity from the generation subsidiary and then the energy produced is distributed to the customers. The energy produced but not distributed constitutes an overproduction which is considered a dead loss for the company. Thus, the over-production or production losses can be defined as the difference between the electricity produced and the actual electricity distributed and, consequently, a better prediction of customers' consumption allows to reduce considerably the errors of production orders which minimizes the losses due to the overproduction.

In addition, at the level of the distribution subsidiary, there are also electricity losses that are so important that the company is putting in place a whole strategy to control its impact. The perfect aim is that the electricity purchased should be equal to the billed energy. However, in the real situation, there are losses that constitute a loss of income for the electricity suppliers [2]. The electrical losses are of two types, technical and non-technical. Technical losses are losses due to the electrical networks and equipment. Non-technical losses are also called management losses and concern the unbilled energy due to meter failures, fraudulent behavior of some customers, and management failures [3]. Therefore, the electricity losses can be defined as the addition of both technical and non-technical losses.

Given the impact of losses especially the management failures type, the suppliers set up a system of annual piloting with a monthly point of verification which aims to follow losses and to make sure to set up the operational actions to maintain a rate of normative losses. In this context, the accurate monthly prediction of customers plays a central role in preventing any deviation of results from what is expected. Hence the objective of our work is to propose a new method of predictions with a high accuracy that improves the piloting system of electricity suppliers such as Sonelgaz.

* Corresponding author.
E-mail addresses: dhad@alu.upo.es (D. Hadjout), jftormal@upo.es (J.F. Torres), atrolor@upo.es (A. Troncoso), abderrazak.sebaa@univ-bejaia.dz (A. Sebaa), fmaralv@upo.es (F. Martínez-Álvarez).
[1] Equally contributing authors.

https://doi.org/10.1016/j.energy.2021.123060

In this context, predictive models for accurately forecasting electricity consumption are important solutions that guide the electricity planning and monitoring for the economic sector. Among the possible direct applications, some of them highlight:

1. Evaluation of loss rate simulations with a high degree of accuracy, which will enhance the operational management
2. Evaluation of energy forecasts with a high degree of accuracy, which will reduce electricity losses due to overproduction.
3. Analysis of the sales of large customers which will reduce the losses of management.

Moreover, analysis and forecasting of electricity consumption play an important and strategic role in the local and national levels whatever the forecasting granularity. Indeed, Mocanu et al. [4] grouped electricity demand forecasting compared to the time horizon of prediction into three categories: first, a period between an hour to a week represents the short-term forecasting, second, the period between a week to a year which represents the medium-term forecasting, third, a time period of more than one year represents long-term forecasting. Thus, at long-term time horizons and at a regional level, electricity consumption forecasting is useful for planning and trading on electricity markets [5].

Electricity consumption forecasting has shown to be a complex problem, it can be considered as a non-linear time series problem [6] since the future values cannot be explained as linear combinations of past values [7]. Thus, to address this problem, several studies used statistical or machine learning-based time series forecasting models to predict the near future electricity load. However, applying such predictive models to large regions that have millions of consumers faces two challenges [8]. First, training a prediction model per-customer is computationally costly. Second, the variability in the energy usage per consumer causes the prediction models to have high errors.

In this paper, we focus on addressing the problem of electricity consumption forecasting of the economic sector for Algeria. We investigate this problem considering medium-long term time horizons and individual consumption forecasting. At the Sonelgaz company level, the management time cycle is one month. Thus, each month, a set of decisions are made such as:

1. The customers are invoiced monthly.
2. The performances energy balances are done monthly. Especially, the performance parameters evaluation such as the loss rate, which depends on monthly purchases and sales.
3. Energy forecasts in terms of purchases and sales are made monthly.
4. Simulations of loss rates are done monthly.

Moreover, the control of monthly sales forecasting improves considerably all the management decisions which strongly depend on the accuracy of the prediction. Indeed, Sonelgaz is a national company that ensures all of the production and management of electrical energy. Since 2011, it has been set up to 39 subsidiaries and 5 joint ventures as an industrial group. These various subsidiaries are responsible for the production, transport, and distribution of electricity [9]. Mainly by gas thermal, combined cycle, and steam thermal. Sonelgaz offers three categories of tension in the electricity distribution system for its customers [10]:

1. Low voltage customers (LV). This first category represented by customers supplied with low voltage at 220 V or 380 V. They are usually from the household sector connected to the distribution electric network.
2. High voltage type A (HVA) customers. This second category represented by customers supplied with a voltage less or equal to 30 kV and a maximum power of 15 MW.
3. High voltage type B (HVB) customers. This category represented by customers supplied with connection voltage greater or equal to 60 kV and subscribed power greater than 15 MW. They are supplied from the high voltage network for electricity.

Generally, the two latter categories are directly connected to the electric power transmission network. This mainly involves big and industrial companies, medium and heavy industries, such as refineries, steel industry, and cement factories.

The most important challenges addressed in this work are summarized below:

1. Consumers of the Algerian economic sector are characterized by their high variability and variety of consumption over time. For instance, during one month the consumption of one consumer can be equal to the consumption of more than 50 consumers.
2. Electricity demand is growing very quickly. For instance, in Algeria electricity production was around 64 662 MW in 2015 [10]. However, in 2019, this production has increased to 72 395 MW.
3. According to Sonelgaz statistics [10], the share of the high voltage electricity consumers is more than 45% of the total electricity consumption. However, they represent only less than 1% of the total number of consumers. Thus, in our case study, we focus on this consumer's category.
4. For years, consumption of this type of customer is collected manually by agents, but just recently in 2017, consumption is measured automatically every first day of the month via a remote reading from an indexing platform, resulting in non-accurate data and requiring much preprocessing.

The main goal of this paper is the development of ensemble models by combining deep learning approaches to address the problem of electricity consumption forecasting. Thus, three deep learning models are studied, Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Temporal Convolutional Networks (TCN) and two ensemble learning approaches for electricity consumption prediction are proposed, one static and another dynamic. The results and developed models of this work fall within the needs of the SONELGAZ company' planning strategy. These models will be therefore used by the planning department of the company to plan the sales and purchase of electricity.

The rest of this paper is organized as follows: Section 2 reviews related work about electricity consumption forecasting. Section 3 describes our methodology and the proposed ensemble learning approaches. It describes how the three models are used to build the ensemble models. Section 4 illustrates the most important results of electricity consumption forecasting in the Bejaia department (Algeria) obtained using our ensemble learning approaches, including statistical tests, followed by a meaningful discussion about the results. Section 5 concludes the paper.

2. Related works

In this section, we review some recently published approaches for electricity consumption forecasting. For more information about more work, readers are advised to consult review [11] which has widely reviewed the existing machine learning techniques for energy consumption forecasting. Thus, several studies have addressed energy consumption forecasting problem considering a well-defined geographic area, a country, or only a specific building. According to the used methodology, we describe and classify some recent works in non-deep learning and deep learning approaches.

2.1. Non-deep learning approaches

Early approaches for electricity consumption prediction rely on statistical methods [12]. However, recent research uses this type of method for comparison purposes. Thus, to achieve better load management of electricity trough different regions in India, a Multiplicative Seasonal Autoregressive Integrated Moving Average (MSARIMA) to forecast monthly electricity demand peaks from five different regions of India was used in Ref. [13]. To make effective managerial decisions and plans, the authors in Ref. [14] proposed to forecast energy price and consumption in the Iranian industrial sector using an Adaptive Neuro-Fuzzy Inference System (ANFIS) and an Artificial Neural Networks. In Ref. [15] daily electricity consumption of an administration building located at the Southwark campus of London South Bank University was predicted. The authors applied multiple linear regression models and a genetic programming model. This technique eliminates unnecessary variables to improve predictive performance. However, it is hard to obtain explanatory variables using linear regression. To make better decisions in reducing total primary energy consumption using next week's energy consumption prediction, authors in Ref. [16] used a random forest regression technique. They exploited human dynamics analysis derived and processed from GSM network call data records in the prediction process. Even if most of the above methods have proven their effectiveness by considering some useful features from the time series. However, they can't handle multiple times series and lack to get a more effective mechanism to find time dependence information between short term and long term. Moreover, the performance of machine learning methods is weak in relation to high-dimensional data.

2.2. Deep learning approaches

There has been considerable work predicting electricity consumption based on deep learning. For instance, in Ref. [17] a solution based on a deep feed-forward neural network for electricity consumption forecasting was proposed. The authors used H2O big data analysis framework with the Apache Spark platform for distributed computing. The authors in Ref. [18] built an LSTM-based univariate model for electricity demand forecasting over both short-term (few days to 2 weeks) and medium-term (few weeks to few months) time horizons. They used features selection and a genetic algorithm to optimize a LSTM model for France metropolitan's electricity consumption data. A similar work was conducted in Ref. [19], where the authors proposed a method for short-term residential load forecasting based on LSTM. The accuracy of the approach was assessed on several sets of residential smart meter data. To predict future energy consumption, the authors in Ref. [20] implemented two approaches, a neural network based genetic algorithm and a neural network based particle swarm optimization. These approaches give better results than the Convolutional Neural Network (CNN) models.

To predict the housing energy consumption, the study in Ref. [21] combined CNN and LSTM that can extract spatial and temporal features. However, the above presented deep learning methods have rarely considered multiple time series. The models proposed in this work use LSTM cells and GRU, which are efficient types of recurrent neural network (RNN) cells. Our proposal differs from existing methods in consumption prediction by considering multiple sources, while allowing the deep neural network to be efficiently trained. Only a few works [14,22] have addressed the electricity consumption forecasting of the economic sector.

3. Methodology

This section is composed of two parts. In the first subsection, we introduce the theoretical background including some related concepts, deep learning models, and architectures used in this work. Then, we describe in the second subsection the new proposal in detail.

3.1. Theoretical background

3.1.1. Times series

All the proposed models in this study fall into the time-series model. Indeed, a time series $y(t)$ is a set of measurements of a variable y, recorded as time progresses. Usually, an assumption is that these points are obtained at equally spaced points of time t, where t is the time index and the resulting samples $y(1)$, $y(2)$, …, $y(n)$ are the discrete-time series [23]. Time-series methods avoid potential issues brought about by inappropriate time-index labeling and they are able to detect the time dependency inherently embedded in the input data [24]. Deep learning is more suitable for dealing with data indexed over time, specifically, for time series forecasting. More information on deep learning for time series forecasting can be found in Refs. [25,26].

3.1.2. Deep learning models

Different deep learning models and architectures are proposed in the literature. An ensemble deep learning method is the process of skillfully combining multiple learning algorithms to improve prediction performance. In this work, we used an exhaustive strategy-based ensemble algorithm by combining different deep learning models.

Recurrent Neural Networks (RNNs). Among deep learning models, RNNs can accommodate dependencies between consecutive time steps so, they are able to store past information in time series and they are generally used to predict sequential tasks. However, they suffer from the problem of vanishing/exploding gradient [27]. Thus, different models were proposed in the literature such as LSTM [28] and GRU neural networks [29] to overcome the RNN vanishing gradient problem. These models and others are described below.

Long Short Term Memory (LSTM). LSTM networks are a special architecture of the RNNs widely used to deal with linear and nonlinear time series problems. LSTM networks are neural networks where the problems of vanishing/exploding gradient, which makes learning long-term dependencies difficult, are addressed [28]. For this purpose, LSTM uses three gates to keep longstanding relevant information and discard irrelevant information, namely update gate (it decides what information is used to update the memory state), output gate (output value that will be the input of the next hidden unit) and forget gate (information thrown away or saved).

Gated Recurrent Unit (GRU). GRUs are also long-term memory networks but they emerged as a simplification of LSTMs due to the high computational cost that such networks have. GRU is one of the most commonly used versions that researchers have converged on and found to be robust and useful for many different problems. The use of gates in RNNs has made it possible to improve the capture of very long-range dependencies, making RNNs much more effective. The LSTM is more powerful and more effective since it has three gates instead of two, but the GRU is a simpler model and it is computationally faster as it only has two gates (update and relevance gates).

Temporal Convolutional Networks (TCN). TCNs are a variant of CNN. They were conceived for data sequence and they compete directly with Deep Recurrent Neural Networks (DRNNs) in terms of execution times and memory requirements. They have two mains restrictions: first, the input and output of the network should have the same length; second, TCNs only use the information from past time steps.

3.1.3. Ensemble learning

Generally, the ensemble learning is the process of combining multiple learning algorithms to improve prediction accuracy obtained by each individual model by reducing their bias and variance In the literature there are few works based on ensemble learning as an electricity consumption forecasting method. Some ensemble learning methods, including bagging [30], are employed only to reduce the variance, while other ensemble methods, including boosting [31] and stacking [32] reduce both the bias and variance. Besides, The work [33] proposes a stacking ensemble scheme. The authors employed a scheme formed by three base learning methods and a top method. The basic learning methods are regression trees based on evolutionary algorithms, artificial neural networks and random forests. At the top level, they have used the generalized boosted regression model in order to combine the predictions produced by the bottom level. In paper [34], two models are used, namely the extreme gradient boosting forest and feedforward deep networks. These models are combined by ridge regression. In Ref. [35], authors developed a cooperative ensemble framework which divides the forecasting problem into several subtasks based on peak and off-peak conditions. Each subtask is then solved using multiple forecasting models that include classification and regression. In the paper [36], three ensembles learning models are developed and the respective results compared: gradient boosted regression trees, random forests and an adaptation of Adaboost. Our paper extends the research by proposing a new method based on deep learning ensemble by training three different deep learning models namely LSTM, GRU and TCN with the electric consumptions of the customers of the economic sector. From the predictions of the trained models we apply two types of methods static and dynamic ensemble that will allow us to implement the combination based on the weights between the different models and thus subtract a better configuration of the ensemble.

3.2. Methodology description

3.2.1. Problem formulation

Our problem is a one-step multivariate time series forecasting, which means our objective is to predict the next value (prediction horizon equal to 1, $h = 1$) using a window of W previous values. This is formulated in Eq. (1):

$$y(t+1) = f(y(t), y(t-1), \ldots, y(t-(W-1))), \tag{1}$$

where the goal is to find the model f.

In this particular case, the consumption of N clients for the next month is wanted to be forecast. Let $C_i(t+1)$ be the consumption of the i-th customer at time $t = 1$ and $\hat{C}_i(t+1)$ the associated forecast for such consumption. Therefore, the goal is to accurately predict the value at $t = 1$ for each of the N clients, as shown in Eq. (2):

$$\hat{C}(t+1) = \sum_{i=1}^{N} \hat{C}_i(t+1) \tag{2}$$

Where $\hat{C}(t+1)$ denotes the prediction of the consumption for the next value of all N customers.

The problem formulation is illustrated in Fig. 1, where the consumption of N customers over time is depicted. The goal is to forecast N consumptions of N customers at time $t = 1$, that is, the set of values $\{C_1(1), C_2(1), C_3(1), \ldots, C_N(1)\}$.

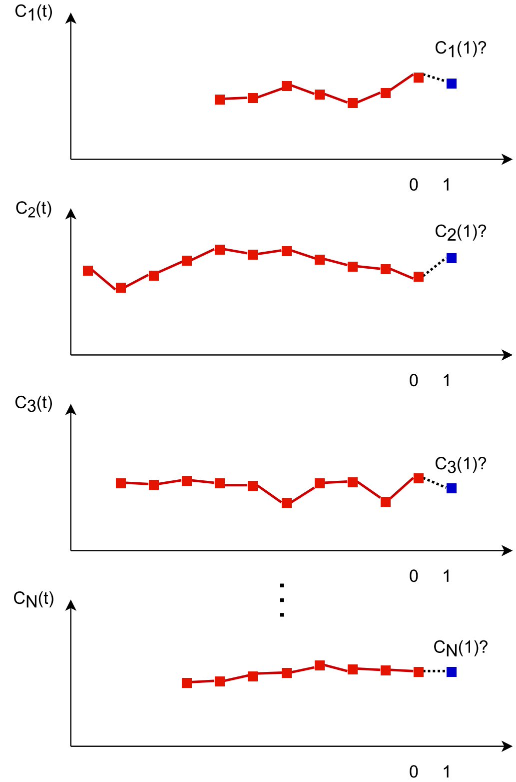


Fig. 1. Graphical formulation of the problem.

3.2.2. Proposed approach

The overall process for forecasting electricity consumption is formulated following the four steps depicted in Fig. 2, which stands for a standard data mining procedure. The steps involved in this process are explained below.

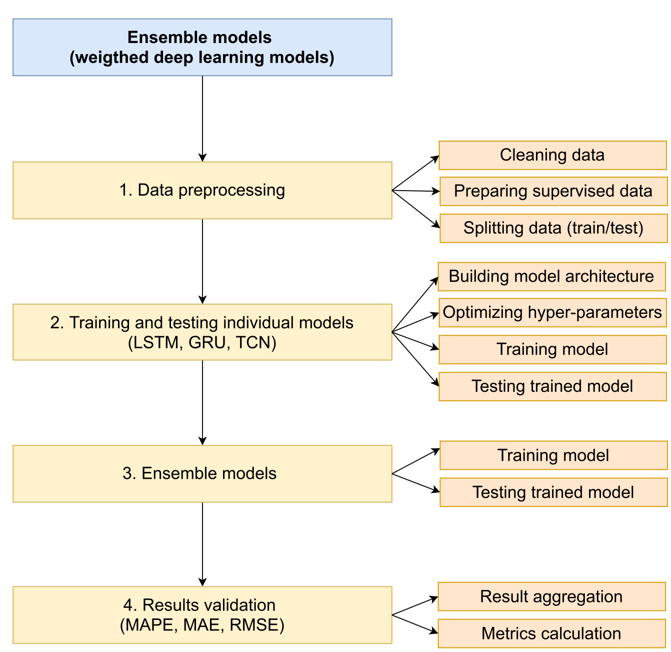


Fig. 2. Methodology steps.

Data pre-processing. Most analysis methods require complete time-series datasets, without missing values. However, in the real and industrial world, data from different applications often have missing or incomplete data ranges. Thus, in this step, missing values are removed. For this purpose, we keep for each client the starting consumption from which no more missing values are present, thus avoiding the application of imputation methods [37].

Note that clients starting at different timestamps are valid, that is, every client may have a different number of measures (different length of the time series). Therefore, every valid client has associated a time-series of different length but all ending at the present ($t = 0$).

Data preparation. At this stage, we transform the time series problem into a supervised learning problem and we normalize the input variables. Thus, for each client of length L, with $L > 12$, we generate $L - 12$ samples. Each sample consists of $W = 12$ inputs (the previous 12 months) and $h = 1$ output (the test value, or target current month). That is, we use a sliding window to generate the $L - 12$ samples. This choice has been made because the electricity consumption is considered a cyclical phenomenon that depends on climate, seasons, holidays, the customer's sector of activity and other factors. Thus, this choice is also motivated by the desire to capture the cyclical characteristics and the maximum of factors that influence the consumption of customers. Moreover, many authors have successfully used this window length [23].

For example, for a client with two years of consumption ($L = 24$), we would have 12 different samples ($L - 12 = 12$):

- Sample 1. Inputs: January (year 1) to December (year 1). Output: January (year 2).
- Sample 2. Inputs: February (year 1) to January (year 2). Output: February (year 2).
- …
- Sample 12. Inputs: December (year 1) to November (year2). Output: December (year 2).

Finally, data are normalized by scaling it to a range of [0,1] using the method called Min-Max Normalization [38]. This method convert the input value of the attribute to the range [min,max], as expressed in Eq. (3):

$$x' = \frac{x - min(\overline{x})}{max(\overline{x}) - min(\overline{x})} \tag{3}$$

Where x' is the normalized value for x, $\overline{x}$ is the set of values, and $max(\overline{x})$ and $min(\overline{x})$ are the maximum and minimum values in $\overline{x}$, respectively.

Data splitting. The dataset is split into 70% instances for training and 30% instances for test. Additionally, the training set is also split and a 30% is used as validation set in order to generate the model which will be applied to the test set.

Ensemble models. It is now introduced the ensemble learning forecasting models [33]. Firstly, the framework of ensemble learning is described. Then, we present how the optimal weight coefficients of integration are obtained using a grid search (GS).

The forecasting phase is composed of two main steps. First, LSTM, GRU and TCN deep learning models have been selected. This choice is motivated by their effectiveness in times series forecasting especially for energy forecasting [25]. Each of these models is constructed with different numbers of hidden layers, inputs, and outputs layers. This construction must be done following any smart strategy for hyperparameters configuration [17,39,40]. Fig. 3 shows the structure of the proposed ensemble learning. At the end of this step, we get three independent forecasting models.

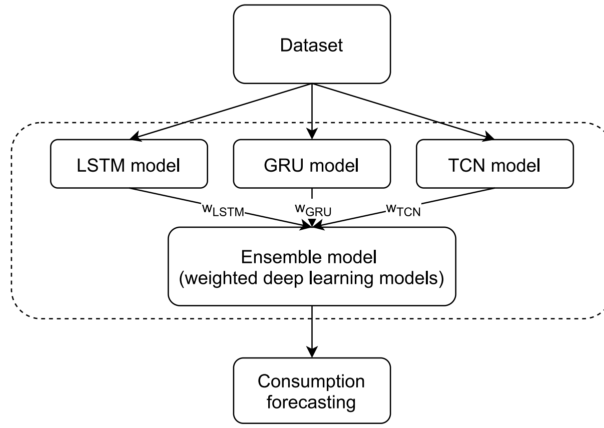


Fig. 3. Ensemble learning method.

After obtaining the three series of forecasting results based on different deep learning models, the GS is applied. Training datasets of all models are used to calculate the optimal weight coefficients of each model.

Let $\hat{P}_L(t)$, $\hat{P}_G(t)$ and $\hat{P}_T(t)$ be the results obtained during the training phase at time t for the LSTM, GRU and TCN models, respectively. Then we can represent the weight integration model as follows:

$$M(t) = w_L(t) \times \hat{P}_L(t) + w_G(t) \times \hat{P}_G(t) + w_T(t) \times \hat{P}_T(t) \tag{4}$$

where $w_L(t)$ denotes the weight for LSTM at time t, $w_G(t)$ denotes the weight for GRU at time t and $w_T(t)$ denotes the weight for TCN at time t.

An exhaustive GS is considered as the solver to optimize the coefficients of our final ensemble. Please note that every coefficient can vary from 0 to 1, with step 0.01, and that the sum of all of them must be equal to 1 as shown in Eq. (5), that is:

$$w_{LSTM}(t) + w_{GRU}(t) + w_{TCN}(t) = 1 \tag{5}$$

This is achieved by minimizing the root mean square error (RMSE) of the aggregated result. Hence, the optimal values of the weight coefficients for each model are used to construct the final model.

Finally, we explore two different strategies to build the ensembles:

1. Dynamic ensemble. Calculate the weights for each time t (in our case, each month). Therefore, to select the optimal weight for each time, we take the training results for each model and time. We denote this strategy by Ensemble-1.
2. Static ensemble. Calculate the weights in a global way, with the overall prediction of training, that is, without considering single time stamps by integrating all samples. Therefore, in this situation, the weight coefficients become independent of time and are calculated just once, regardless of the prediction horizon. We denote this strategy by Ensemble-2.

Evaluation metrics. To evaluate the performance of the proposed deep learning models, the following three performance metrics, namely Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Square Error (RMSE) were used. These selected metrics indicate the forecasting accuracy of the models. They are expressed in Equations (6)–(8), respectively.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|\hat{C}_i(t+1) - C_i(t+1)| \tag{6}$$

$$MAPE = \frac{100}{N}\sum_{i=1}^{N}\frac{|\hat{C}_i(t+1) - C_i(t+1)|}{C_i(t+1)} \tag{7}$$

$$RMSE = \sqrt{\left(\frac{1}{N}\sum_{i=1}^{N}(\hat{C}_i(t+1) - C_i(t+1))^2\right)} \tag{8}$$

where $C_i(t+1)$ denotes the actual consumption at time t+1 for the i-th client, $\hat{C}_i(t+1)$ denotes the estimated consumption for the i-th client at time t+1 and N is total number of clients considered.

4. Experiment and results

4.1. Dataset description

The experimental data that were collected from the Bejaia concession of Algerian Electricity and Gas Distribution Company (SADEG, one of Sonelgaz's subsidiary, https://www.sadeg.dz) related to the electrical energy consumption in Bejaia department from 2006 to 2019. This data includes 1699 economic sector consumers and 285 432 measures. Monthly consumption is measured and analyzed during this period.

The original dataset has been preprocessed in order to be used in different experiments. As shown in Table 1, from the original data only 59,01% of this data are used. Furthermore, it shows the results of the data cleaning stage of our methodology. The data distribution over time is depicted in Fig. 4. It can be seen that the number of measurements has been increasing since the initial 8376 values in 2016 until the final 16 104 values in 2019 (see Fig. 5).

In this study, the historical window is set to 12 ($W = 12$), thus identifying the 12 months of a year. The prediction horizon h has been set to 1, corresponding to the next month.

4.2. Experimental setup and models designing

To evaluate and validate the effectiveness of the proposed solution, we conducted several experiments to forecast electricity consumption. All deep learning algorithms and ensemble method were implemented using the Keras 2.3.1 API [41] with TensorFlow framework [42] version 2.1.0 under Python language version 3.7. Moreover, we implemented, LSTM, GRU, TCN, with different hyperparameters. After testing several different possible configurations, we selected parameters that provided the best accuracy for each model. Furthermore, the grid search is also used to optimize several

Table 1
Data cleaning results.

Measures	Collected	Missing	Removed	Used
Samples	285 432	91 994	117 000	168 432
Rate	100%	32.23%	40.99%	59.01%

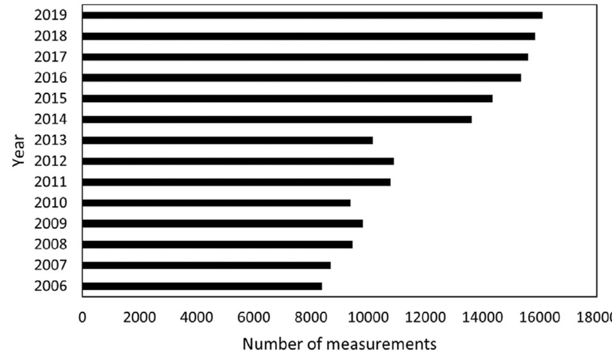
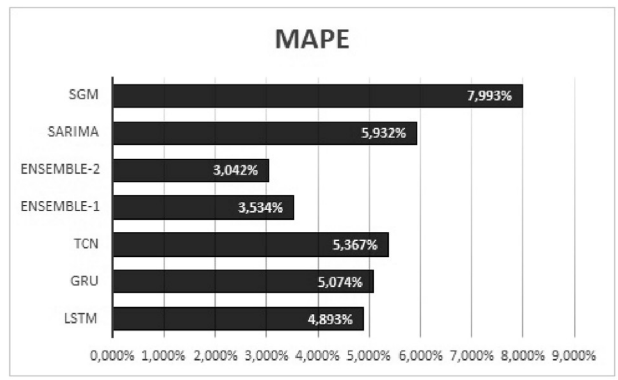
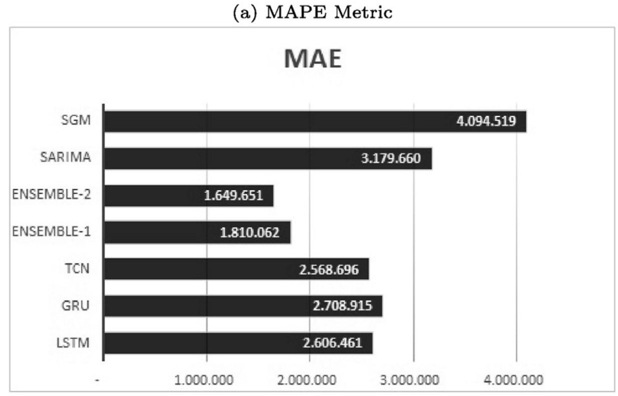


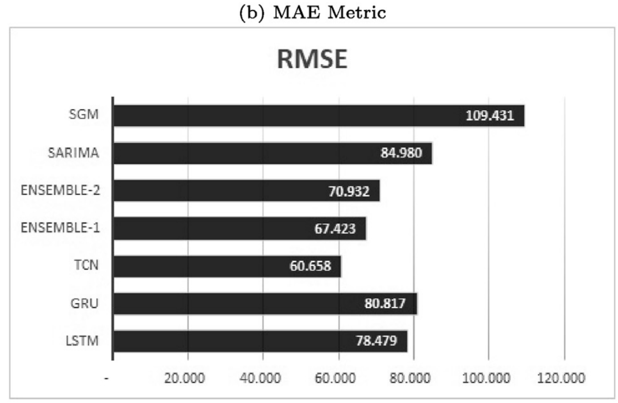
Fig. 4. Number of available measures over time.



(a) MAPE Metric



(b) MAE Metric



(c) RMSE Metric

Fig. 5. Comparison of three metrics MAPE, MAE and RMSE for proposed ensemble learning and other methods.

important hyper-parameters of different models [17]. Adjustment and evaluation on the main parameters in different models are made, including units in the last layer, dropout of the last layer, epochs and batch size. The resulting architectures and hyper-parameters are listed below:

1. LSTM. Layer 1: units = 200, dropout = 0.2; layer 2: units = 200, dropout = 0.2; layer 3: units = 100, dropout = 0.2; Layer 4: units = 1. Optimizer = Adam, loss = MSE, metrics = [accuracy, MSE].
2. GRU. Dense layer 1: units = 12; GRU layer 1: units = 64, dropout 0.2; GRU layer 2: units = 32, dropout 0.1; dense layer 2: units = 1. Optimizer = Adam, loss = MSE. metrics = [accuracy, MSE], epochs = 50, batch-size = 124.
3. TCN. Dense layer 1: units = 12; layer 2: units = 1. Optimizer = Adam, loss = MSE, metrics = [accuracy, MSE], epochs = 50, batch-size = 124.

4.3. Training and test phases

As explained in the previous section, a GS strategy is used to find the optimal weight coefficients for the ensemble model. We vary the weights with two decimal digits, i.e., 101 values for each weight, and their sum must be 1 according to Eq. (5). The metric selected to be minimized is the RMSE.

Tables 2 and 3 describe the optimal weights coefficients in the two proposed solutions. In Table 2, we present the optimal weight coefficients of each model for each predicted month, that is, for the first strategy described in the previous section. In Table 3, we show the optimal weight coefficients for the second strategy reached after the adjustment of the entire training set.

Finally, the values of the optimal weight coefficients for each model are used to construct the final result.

4.4. Results analysis and discussion

In this section the evaluation of the results of the proposed model are given. Individual results for each deep learning model is provided and compared to the aggregated result of the ensemble models. Referring to Fig. 6, the overall analysis of the results for the three evaluation metrics clearly shows that the proposed Ensemble-1 and Ensemble-2 methods obtain the best results for the MAPE, MAE metrics. The TCN model on the other hand shows a better result for the RMSE metric followed by Ensemble-1 and Ensemble-2.

Table 2
Coefficients for Ensemble-1.

t	$w_{GRU}(t)$	$w_{LSTM}(t)$	$w_{TCN}(t)$
January	0.22	0.50	0.28
February	0.27	0.00	0.73
March	0.00	0.82	0.18
April	0.00	0.57	0.43
May	0.00	0.00	1.00
June	0.25	0.17	0.58
July	0.09	0.18	0.73
August	0.00	0.54	0.46
September	0.33	0.07	0.60
October	0.14	0.29	0.57
November	0.29	0.18	0.53
December	0.27	0.00	0.73

Table 3
Coefficients for Ensemble-2.

t	$w_{GRU}(t)$	$w_{LSTM}(t)$	$w_{TCN}(t)$
Entire set	0.22	0.50	0.28

Table 4 shows the aggregated consumption for all the considered customers, along with the aggregated prediction made for the individual deep learning models and the ensemble models proposed. Note that all the values are expressed in megawatt-hour (MWh) for the sake of simplicity and comprehensibility.

Table 5 shows the MAE obtained for the individual deep learning, statistic and ensemble models. On the one hand, it is worth noting that all the individual models reach the best result in at least three months, which leads to the conclusion that each of them is suitable for different periods of time. One may wonder why each model performs better for a particular month. But this is, a priori, almost impossible to determine since each learning algorithm is based on a set of assumptions about the data, its inductive bias [43]. This means that it will only learn well if the bias matches the learning problem. A learning algorithm may perform very well in one domain, but not on the next. This poses strong restrictions on the use of machine learning or data mining techniques, since the relationship between the learning problem and the effectiveness of different learning algorithms is not yet understood. Anyway, this situation reinforces the idea of using ensemble learning to make the best of each individual model.

Thus, LSTM outperforms the other models in the coldest months (December, January, February) whereas TCN obtains quite good results for the warmest ones (July and August, as well as October). GRU emerges as a more suitable method for seasons with no extreme temperatures (March, April, May, June, September, November). On the other hand, the ensemble models reduce by almost a 36% the MAE obtained by the best individual model, from 2568.70 MWh obtained by TCN to 1649.65 MWh obtained by Ensemble-2. For these models, it seems that Ensemble-1 is highly influenced by TCN since it gets the better results for months April, July, August and October. Ensemble-2 gets better results for the remaining months which leads to the lowest average MAE (1649.65 MWh).

Table 6 shows the MAPE obtained for the individual deep, statistic and ensemble models. As expected, given the direct relation between MAPE and MAE, the lowest MAPE is achieved for Ensemble-2, with 3.04%, improving by almost a 38% the best individual deep learning model (LSTM, 4.89%). Ensemble-1 obtained an error slightly greater than Ensemble-2, with 3,53% but, again, better than all individual models. Overall, the errors are smoothed but the month of August remains a challenging month for both Ensemble-1 and Ensemble-2, with errors greater than 8% and 10%, respectively. Note that TCN reached a MAPE less than 2% for this month but the higher errors obtained by LSTM and GRU deeply influenced in the final error by the ensemble models.

Finally, Table 7 shows the RMSE obtained for the individual deep learning, statistic and ensemble models. This error shows a significant difference from those shown by MAE and MAPE: Ensemble-1 obtained the smaller error (67.42 MWh), while Ensemble-2 had reached the smaller ones for the latter. Additionally, TCN outperforms LSTM and GRU for all the months except for

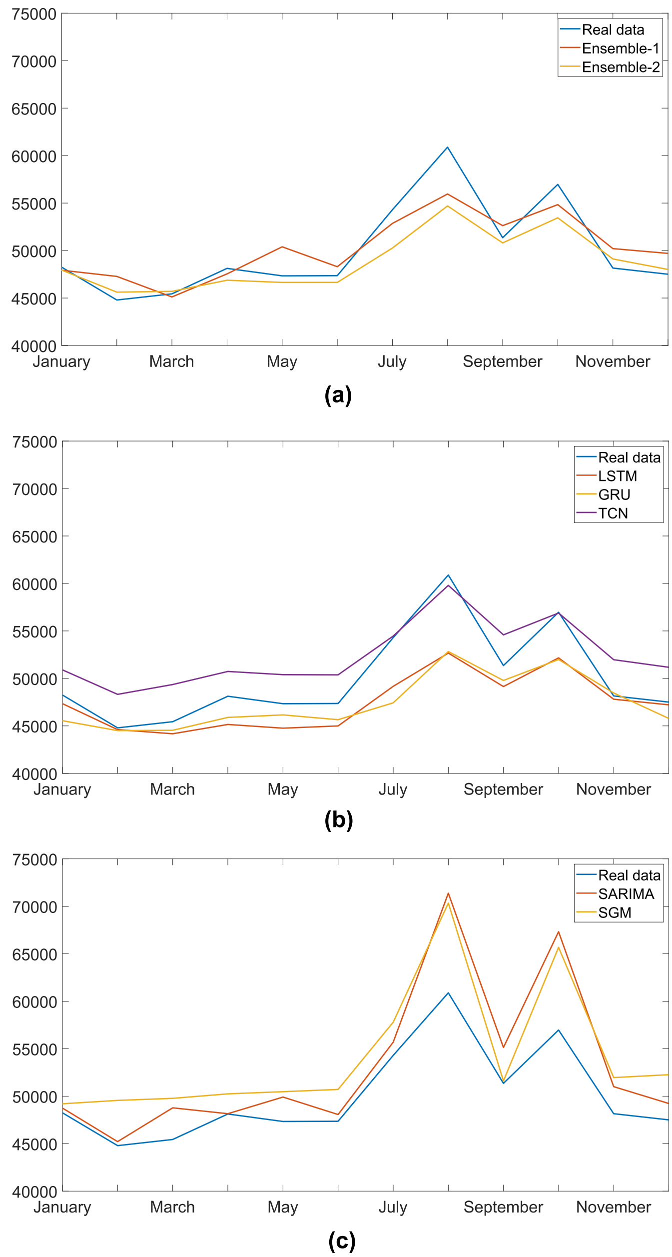


Fig. 6. The figures transpose the real values with the results of the prediction of the Ensemble 1 and Ensemble 2 methods (a), the deep leaning model (b) and the statistical method (c).

September. The RMSE by TCN (60.66 MWh) is even smaller than those by Ensemble-1 and Ensemble-2. At this point, it is important to highlight the differences between MAE (and MAPE) and RMSE, to understand why some methods are better in some metrics but worse in others. While both metrics express the average of the model prediction error, taking the square root of the average squared errors has some interesting implications for RMSE. Since the errors are squared before they are averaged, the RMSE gives a relatively high weight to large errors. This means the RMSE should be more useful when large errors are particularly undesirable. In other words, RMSE increases with the variance of the frequency distribution of error magnitudes. Therefore, Ensemble-2 obtained lower absolute and relative error values but higher variance. If no such variance is desired and the MAE (and MAPE) obtained is acceptable for the forecasting process, Ensemble-1 should be used instead.

4.5. Statistical significance

This section is devoted to evaluate the statistical significance of the ensemble models developed. The goal is to show that such models provide results significantly different from those of the individual models and from the other ensemble model.

The standard Wilcoxon-signed rank non-parametric test [44] has been chosen to perform the statistical comparison, as suggested in the literature. The metric selected has been RMSE.

This test evaluates paired sample tests and determine, through the p-value its statistical significance. When p-values are small enough (typically less than 0.05), then it can be assumed that the two distributions are independent.

Table 8 shows the p-values for all possible combinations for all the considered models. Small p-values are obtained for all the comparisons made (<0.05), leading to the conclusion that all the ensemble models proposed exhibit statistically different performance from the individual deep learning models and from the other ensemble. In other words, it is confirmed that both ensemble proposals (referred as Ens. 1 and Ens. 2, respectively, in Table 8) generate statistically independent outputs.

5. Conclusions and future work

In this paper, we have presented a novel model to achieve highly accurate long-medium-term electricity consumption forecasting for the Algerian economic sector, by using ensemble learning of diverse deep learning models. In the proposed ensemble learning, three different deep learning models, namely LSTM, GRU, and TCN models, are used. These three models separately learn the evolution of electricity consumption. Then, a weighting strategy is used to

Table 4
Actual predicted values for both the individual, statistic and ensemble models (in MWh).

Date	Actual	LSTM	GRU	TCN	Ensemble-1	Ensemble-2	SARIMA	SGM
January	48 249.24	47 342.19	45 545.62	50 897.86	47 930.64	47 942.53	48 748.66	49 200.34
February	44 792.72	44 605.71	44 503.30	48 320.30	47 279.30	45 623.26	45 213.66	49 558.25
March	45 441.18	44 171.16	44 537.75	49 356.48	45 113.95	45 703.70	48 773.7	49 774.14
April	48 123.62	45 151.03	45 891.67	50 732.71	47 543.18	46 876.84	48 156.99	50 245.55
May	47 337.01	44 760.07	46 157.14	50 394.91	50 394.91	46 645.18	49 909.91	50 475.11
June	47 358.06	44 995.29	45 653.49	50 374.80	48 297.89	46 646.36	48 068.61	50 713.36
July	54 294.37	49 165.44	47 429.25	54 459.22	52 857.62	50 265.73	55 686.26	57 792.3
August	60 887.86	52 668.30	52 833.62	59 797.83	55 958.85	54 700.94	71 379.98	70 339.92
September	51 353.44	49 140.60	49 785.21	54 586.79	52 623.18	50 807.35	55 128.29	51 604.47
October	56 967.99	52 164.75	51 992.70	56 885.81	54 837.92	53 448.80	67 322.05	65 673.68
November	48 155.82	47 809.49	48 464.87	51 969.48	50 204.60	49 118.47	51 005.18	51 961.24
December	47 507.31	47 217.07	45 785.13	51 172.37	49 703.12	48 009.53	49 231.25	52 264.48

Table 5
MAE for the individual deep learning, statistic and ensemble models (in MWh).

Date	LSTM	GRU	TCN	Ensemble-1	Ensemble-2	SARIMA	SGM
January	907.05	2703.62	2648.63	318.60	306.71	499.42	951.11
February	187.01	289.41	3527.58	2486.58	830.55	420.94	4765.53
March	1270.02	903.43	3915.30	327.23	262.52	3332.51	4332.96
April	2972.60	2231.95	2609.09	580.45	1246.78	33.38	2121.93
May	2576.94	1179.87	3057.90	3057.90	691.83	2572.9	3138.1
June	2362.77	1704.57	3016.74	939.83	711.70	710.55	3355.3
July	5128.94	6865.13	164.84	1436.75	4028.64	1391.88	3497.93
August	8219.56	8054.24	1090.03	4929.01	6186.92	10 492.13	9452.06
September	2212.84	1568.24	3233.35	1269.74	546.10	3774.84	251.03
October	4803.25	4975.29	82.18	2130.07	3519.20	10 354.06	8705.68
November	346.33	309.05	3813.66	2048.78	962.65	2849.36	3805.42
December	290.24	1722.18	3665.06	2195.81	502.22	1723.93	4757.17
Average	2606.46	2708.91	2568.70	1810.06	1649.65	3179.66	4094.52

Table 6
MAPE for the individual deep learning, statistic and ensemble models.

Date	LSTM	GRU	TCN	Ensemble-1	Ensemble-2	SARIMA	SGM
January	1.88%	5.60%	5.49%	0.66%	0.64%	1.04%	1.97%
February	0.42%	0.65%	7.88%	5.55%	1.85%	0.94%	10.64%
March	2.79%	1.99%	8.62%	0.72%	0.58%	7.33%	9.54%
April	6.18%	4.64%	5.42%	1.21%	2.59%	0.07%	4.41%
May	5.44%	2.49%	6.46%	1.46%	5.44%	5.44%	6.63%
June	4.99%	3.60%	6.37%	1.98%	1.50%	1.5%	7.08%
July	9.45%	12.64%	0.30%	2.65%	7.42%	2.56%	6.44%
August	13.50%	13.23%	1.79%	8.10%	10.16%	17.23%	15.52%
September	4.31%	3.05%	6.30%	2.47%	1.06%	7.35%	0.49%
October	8.43%	8.73%	0.14%	3.74%	6.18%	18.18%	15.28%
November	0.72%	0.64%	7.92%	4.25%	2.00%	5.92%	7.9%
December	0.61%	3.63%	7.71%	4.62%	1.06%	3.63%	10.01%
Average	4.89%	5.07%	5.37%	3.53%	3.04%	5.93%	7.99%

Table 7
RMSE for the individual deep learning, statistic and ensemble models (in MWh).

Date	LSTM	GRU	TCN	Ensemble-1	Ensemble-2	SARIMA	SGM
January	66.66	56.11	25.10	51.07	51.00	13.35	25.42
February	71.26	62.54	29.37	36.93	56.01	11.25	127.36
March	53.75	45.83	22.45	45.51	38.88	89.07	115.8
April	40.08	33.29	17.70	22.85	26.00	0.89	56.71
May	44.59	42.35	17.10	17.10	33.61	68.76	83.87
June	60.86	61.47	32.50	42.41	51.14	18.99	89.67
July	35.46	34.72	33.40	28.02	27.95	37.2	93.49
August	199.76	199.39	197.16	197.81	198.11	280.41	252.62
September	84.80	141.38	125.76	126.87	106.04	100.89	6.71
October	99.65	107.13	92.54	94.25	96.89	276.72	232.67
November	91.23	91.67	72.88	77.25	82.64	76.15	101.7
December	93.63	93.92	61.94	69.00	82.92	46.07	127.14
Average	78.48	80.82	60.66	67.42	70.93	84.98	109.43

Table 8
P-values for the Wilcoxon signed-rank statistical test.

	LSTM	GRU	TCN	Ens. 1	Ens. 2
Ens. 1	0.028	0.033	0.010	–	0.029
Ens. 2	0.024	0.016	0.007	0.029	–

obtain the final weight integration. To determine the optimal weight coefficients of each model, a GS strategy is used. Thus, the final electricity consumption forecasting results are achieved by the aggregation of three forecasting results of different deep learning models with the optimal weight coefficients obtained by the GS algorithm. To evaluate the proposed ensemble learning models in terms of MAE, MAPE and RMSE, we used the dataset of Bejaia HVA consumers consumption. Through comparison results, the proposed ensemble learning models improve significantly the forecasting performance compared to the different models separately in terms of MAE and MAPE. However, TCN obtained the best results in terms of RMSE due to the low variance exhibited in the errors. The statistical significance of the ensemble models developed has been assessed through the Wilcoxon signed-rank test, showing p-values smaller than 0,05 for all the paired combinations. For future work, it is an interesting subject to consider the categorization of economic sector consumers. Finally, it seems that August requires a specific model, given the high error rate achieved by all the methods.

Credit author statement

D. Hadjout: Data curation, Conceptualization, writing. J. F. Torres: Visualization, Investigation. A. Troncoso: Software, Validation. A. Sebaa: Conceptualization, reviewing, editing. F. Martínez-Álvarez: Conceptualization, reviewing, editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank the Spanish Ministry of Science, Innovation and Universities for the support under project PID2020-117954RB-C21, and to the Junta de Andalucía for projects PY20-00870 and UPO-138516. Also, this work has been partially supported by the General Directorate of Scientific Research and Technological Development (DGRSDT, Algeria), under the PRFU project (ref: C00L07UN060120200003). Funding for open access publishing: Universidad Pablo de Olavide/CBUA.

References

[1] Fiot JB, Dinuzzo F. Electricity demand forecasting by multi-task learning. IEEE Trans Smart Grid 2016;9(2):544–51.
[2] Nizar AH, Zhao JH, Dong ZY. Customer information system data pre-processing with feature selection techniques for non-technical losses prediction in an electricity market. In: Proceedings of the IEEE international conference on power system technology; 2006. p. 1–7.
[3] Viegas JL, Esteves PR, Melicio R, Mendes VMF, Vieria SM. Solutions for detection of non-technical losses in the electricity grid: a review. Renew Sustain Energy Rev 2017;80:1256–68.
[4] Mocanu E, Nguyen PH, Gibescu M, Kling WL. Deep learning for estimating building energy consumption. Sustainable Energy, Grids and Networks 2016;6:91–9.
[5] Goude Y, Nedellec R, Kong N. Local short and middle term electricity load forecasting with semi-parametric additive models. IEEE Trans Smart Grid 2013;5(1):440–6.
[6] Dedinec A, Filiposka S, Dedinec A, Kocarev L. Deep belief network based electricity load forecasting: an analysis of Macedonian case. Energy 2016;115:1688–700.
[7] Bradley E, Kantz H. Nonlinear time-series analysis revisited. Chaos 2015;25:097610.
[8] Simmhan Y, Noor Muhammad U. Scalable prediction of energy consumption using incremental time series clustering. In: Proceedings of the IEEE international conference on big data. IEEE; 2013. p. 29–36.
[9] Nezzar RM. *Approches multi-modèle pour la prévision de la charge électrique à moyen long terme*. PhD thesis. Université de Annaba; 2016.
[10] CREG. Electricity and gas regulation commission. 1999.
[11] Deb C, Zhang F, Yang J, Eang Lee S, Shah KW. A review on time series forecasting techniques for building energy consumption. Renew Sustain Energy Rev 2017;74:902–24.
[12] Trull Ó, García-Díaz JC, Troncoso A. Application of discrete-interval moving seasonalities to Spanish electricity demand forecasting during easter. Energies 2019;12(6):1083.
[13] Rallapalli SR, Ghosh S. Forecasting monthly peak demand of electricity in India—a critique. Energy Pol 2012;45:516–20.
[14] Mirsoltan S, Akhavan NST. Forecasting energy price and consumption for Iranian industrial sectors using ANN and ANFIS. Iranian Journal of Economic Studies; 2013.
[15] Amber KP, Aslam MW, Hussain SK. Electricity consumption forecasting models for administration buildings of the UK higher education sector. Energy Build 2015;90:127–36.
[16] Bogomolov A, Lepri B, Larcher R, Antonelli F, Pianesi F, Pentland A. Energy consumption prediction using people dynamics derived from cellular network data. EPJ Data Science 2016;5(1):13.
[17] Torres JF, Galicia A, Troncoso A, Martínez-Álvarez F. A scalable approach based on deep learning for big data time series forecasting. Integrated Comput Aided Eng 2018;25(4):335–48.
[18] Bouktif S, Fiaz A, Ouni A, Serhani MA. Optimal deep learning LSTM model for electric load forecasting using feature selection and genetic algorithm: comparison with machine learning approaches. Energies 2018;11(7):1636.
[19] Kong W, Dong ZY, Jia Y, Hill DJ, Xu Y, Zhang Y. Short-term residential load forecasting based on LSTM recurrent neural network. IEEE Trans Smart Grid 2019;10(1):841–51.
[20] Muralitharan K, Sakthivel R, Vishnuvarthan R. Neural network based optimization approach for energy demand prediction in smart grid. Neurocomputing 2018;273:199–208.
[21] Kim T-Y, Cho S-B. Predicting residential energy consumption using CNN-LSTM neural networks. Energy 2019;182:72–81.
[22] Wang Z-X, Li Q, Pei L-L. A seasonal GM (1, 1) model for forecasting the electricity consumption of the primary economic sectors. Energy 2018;154:522–34.
[23] Martínez-Álvarez F, Troncoso A, Asencio-Cortés G, Riquelme JC. A survey on data mining techniques applied to electricity-related time series forecasting. Energies 2015;8(11):13162–93.
[24] Pérez-Chacón R, Asencio-Cortés G, Troncoso A, Martínez-Álvarez F. Big data time series forecasting based on pattern sequence similarity and its application to the electricity demand. Inf Sci 2020;540:160–74.
[25] Torres JF, Hadjout D, Sebaa A, Martínez-Álvarez F, Troncoso A. Deep learning for time series forecasting: a survey. Big Data 2021;9(1):3–21.
[26] Lara-Benítez P, Carranza-García M, Riquelme JC. An experimental review on deep learning architectures for time series forecasting. Int J Neural Syst 2021;31(3):2130001.
[27] Rumelhart D, Hinton G, Williams R. Long short-term memory. Nature 1986;323:533–6.
[28] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput 1997;9:1735–80.
[29] Cho K, Merrienboer BV, Bahdanau D, et al. On the properties of neural machine translation: encoder-decoder approaches. In: Proceedings of SSST-8; 2014. p. 103–11.
[30] Breiman L. Bagging predictors. Mach Learn 1996;24:123–40.
[31] Schapire RE. The strength of weak learnability. Mach Learn 1990;5:197–227.
[32] Breiman L. Stacked regressions. Mach Learn 1996;24:49–64.
[33] Divina F, Gilson A, Gómez-Vela F, García-Torres M, Torres JF. Stacking ensemble learning for short-term electricity consumption forecasting. Energies 2018;11(4):949.
[34] Chen K, Jiang J, Zheng F, Chen K. A novel data-driven approach for residential electricity consumption prediction based on ensemble learning. Energy 2018;150:49–60.
[35] Weeraddana D, Khoa Nguyen-Lu D, Mahdavi N. Machine learning based novel ensemble learning framework for electricity operational forecasting. Elec Power Syst Res 2021;201:107477.
[36] Pinto T, Praça I, Vale Z, Silva J. Ensemble learning for electricity consumption forecasting in office buildings. Neurocomputing 2021;423:747–55.
[37] Bokde N, Beck MW, Martínez-Álvarez F, Kulat K. A novel imputation methodology for time series based on pattern sequence forecasting. Pattern Recogn Lett 2018;116:88–96.
[38] Bhanja S, Das A. Impact of data normalization on deep neural network for time series forecasting. 2018. ArXiv, abs/1812.05519.
[39] Torres JF, Gutiérrez-Avilés D, Troncoso A, Martínez-Álvarez F. Random hyper-parameter search-based deep neural network for power consumption forecasting. Lect Notes Comput Sci 2019;11506:259–69.
[40] Divina F, Torres JF, García-Torres M, Martínez-Álvarez F, Troncoso A. Hybridizing deep learning and neuroevolution: application to the Spanish short-term electric energy consumption forecasting. Appl Sci 2020;10(16):5487.
[41] Belcher J. An overview of time series analysis. Nurse Res 2001;9(1):26–40.
[42] Cai M, Pipattanasomporn M, Rahman S. Day-ahead building-level load forecasts using deep learning vs. traditional time-series techniques. Appl Energy 2019;236:1078–88.
[43] Gordon D, Desjardins M. Evaluation and selection of biases in machine learning. Mach Learn 1995;20:5–22.
[44] Wilcoxon F. Individual comparisons by ranking methods. Biometrics 1945;1:80–3.