

Article

# Application of Artificial Neural Networks for Natural Gas Consumption Forecasting

Athanasios Anagnostis <sup>1,2,\*</sup> , Elpiniki Papageorgiou <sup>1,3</sup>  and Dionysis Bochtis <sup>1</sup> 

<sup>1</sup> Institute for Bio-economy and Agri-technology (iBO), Center for Research and Technology—Hellas (CERTH), Thessaloniki GR57001, Greece; e.papageorgiou@certh.gr or elpinikipapageorgiou@uth.gr (E.P.); d.bochtis@certh.gr (D.B.)

<sup>2</sup> Department of Computer Science, University of Thessaly, Lamia GR35131, Greece

<sup>3</sup> Department of Energy Systems, Faculty of Technology, University of Thessaly, Geopolis Campus Ring Road of Larissa-Trikala, Larissa GR41500, Greece

\* Correspondence: a.anagnostis@certh.gr or athananagno@uth.gr; Tel.: +30-6947072707

Received: 3 July 2020; Accepted: 6 August 2020; Published: 9 August 2020



**Abstract:** The present research study explores three types of neural network approaches for forecasting natural gas consumption in fifteen cities throughout Greece; a simple perceptron artificial neural network (ANN), a state-of-the-art Long Short-Term Memory (LSTM), and the proposed deep neural network (DNN). In this research paper, a DNN implementation is proposed where variables related to social aspects are introduced as inputs. These qualitative factors along with a deeper, more complex architecture are utilized for improving the forecasting ability of the proposed approach. A comparative analysis is conducted between the proposed DNN, the simple ANN, and the advantageous LSTM, with the results offering a deeper understanding the characteristics of Greek cities and the habitual patterns of their residents. The proposed implementation shows efficacy on forecasting daily values of energy consumption for up to four years. For the evaluation of the proposed approach, a real-life dataset for natural gas prediction was used. A detailed discussion is provided on the performance of the implemented approaches, the ANN and the LSTM, that are characterized as particularly accurate and effective in the literature, and the proposed DNN with the inclusion of the qualitative variables that govern human behavior, which outperforms them.

**Keywords:** machine learning; artificial neural networks; natural gas; demand forecasting

## 1. Introduction

The consumption of natural gas has seen a substantial increase during recent years, as it presents a reliable and economical energy and heating solution for businesses as well as households. Its wide acceptance from large-scale infrastructures to small houses has created diverse consumption patterns, especially during high-demand occasions. Inevitably, this has perplexed any attempt of forecasting its demand, especially when one considers the diversity of the consumers and the finite restrictions of the natural gas infrastructure, i.e., low accumulation ability within the grid.

Analytical modelling of such complicated systems would require substantial effort in order to design the grid architecture and each of its consumers, apply correct heat losses throughout the pipes, and in general, include a variety of intricate parameters into the whole system before running the simulation computations. On the other hand, data-driven models are invariant of such parameter tuning and can properly model a system by learning valuable patterns from its collected data. Machine learning algorithms create models by recurrently learning from data, until they can model a process in the best way possible. Being dependent on data alone, alternative scenarios based on different

energy resources like fossil fuels, oil, or electricity may as well utilize these methods for their own forecasting purposes.

State-of-the-art published studies which focus on natural gas forecasting of production, consumption, demand, market volatility and fluctuation in prices, and income elasticity have been surveyed and are presented in [1]. Efficient energy supply planning is essential for any country's socio-economic state since it is crucial, especially for building successful development plans [2]. There is a large number of papers found in the relevant literature that tackle the problem of accurate forecasting of natural gas consumption, mostly focusing in hourly intervals [3]. Short-term forecasting is based on the pattern analysis of time series in order to predict accurate values of consumption or demand [4]. Artificial intelligence, machine learning, and other statistical methods are typically used in short-, medium-, and long-term forecasts of energy demand [5]. Based on research studies from the literature, there are notable findings that utilized artificial neural network (ANN) algorithms on forecasting natural gas demand, and whose day-ahead predictions had high accuracy [6–15]. Multiple variants of neural networks, especially deep neural networks, have been extensively used to tackle the problem of short-term demand forecasting of natural gas. Deep learning was firstly used by Merkel et al. for forecasting short-term load of natural gas [16,17], and then to be compared to traditional ANN and linear regression models on 62 different areas with at least 10 years of data [18].

Other data-driven approaches, such as neuro-fuzzy methods or genetic algorithms, have tackled the problem of natural gas demand [19–21]. Hybrid approaches including Wavelet Transform (WT), Genetic Algorithm (GA), Adaptive Neuro-Fuzzy Inference System (ANFIS), and Feed-Forward Neural Network (FFNN) have been used by Panapakidis and Dagoumas in order to forecast natural gas demand in the Greek natural gas grid [22]. Moreover, other soft computing techniques, like fuzzy cognitive maps (FCMs), enhanced by evolutionary algorithms, have been applied for modeling time series problems [23–28]. In [29], Poczeta and Papageorgiou conducted a preliminary study on implementing FCMs with ANNs for natural gas prediction, showing for first time the capabilities of evolutionary FCMs in this domain. Furthermore, the research team in [30] recently conducted a study for time series analysis devoted to natural gas demand prediction in three Greek cities, implementing an efficient ensemble forecasting approach through combining ANN, RCGAFCM, SOGA-FCM, and hybrid FCM-ANN. In this research study, the advantageous features of intelligent methods, through an ensemble to multivariate time series prediction, have emerged.

Many works can be found in the literature that address the accurate forecasting of natural gas demand with a methodology that was based solely on an artificial neural network, or was used in combination with other methods in hybrid forecasting systems; however, in the present work, an innovative approach that includes vital social factors in deep neural network (DNN) models was studied exclusively, contributing to the novelty of the current study. The main aim of this study is the development of a non-linear time series model that can accurately predict future energy demand and estimate how the introduction of important social factors can improve the accuracy of its predictions. As a case study for the demonstration of the approach's applicability, natural gas energy data from various cities in Greece, which present socio-economic aspects and thus different consumption attributes, have been implemented.

Contrary to most studies that focus on quantitative-only inputs, there are some studies that take into consideration the impact of social or socio-economic factors with machine learning based approaches [31]. The behavioral habits and characteristics of consumers are strong indicators in forecasting electricity load in households [32–34]. Social factors were taken into consideration in the prediction of total energy demand in several cases such as Spain [35], China [36], and Turkey [37,38]. The application of social components alongside meteorological and past consumption data was also studied in district heating networks [39,40]. In all relevant studies, the results showed that the inclusion of social parameters in the modelling can increase the model's overall accuracy [41,42].

Our effort focuses on investigating three types of approaches. The first approach relies on a simple Artificial Neural Network (ANN), namely a single-layer perceptron, that takes into account only

quantitative variables. The second approach is based on the state-of-the-art recurrent neural network (RNN), namely Long Short-Term Memory (LSTM) network, which uses single-variable time series and can predict a variable's value for the next point in time by "memorizing" past variations. The third approach is the proposed Deep Neural Network (DNN) that takes as input not only quantitative, but also qualitative variables. The DNN consists of more nodes and layers than the ANN since it needs to process a larger and more diverse amount of inputs, both numerical and categorical, in a more appropriate fashion. The qualitative variables that are being used in the proposed DNN approach are social factors that fit the characteristics of the country of Greece and will be described in detail in paragraph 2.2. For the case of natural gas demand forecasting, the consumption of energy is bound to the behavior of the human population, which is dependent on social habits, a factor whose impact is investigated extensively in this study.

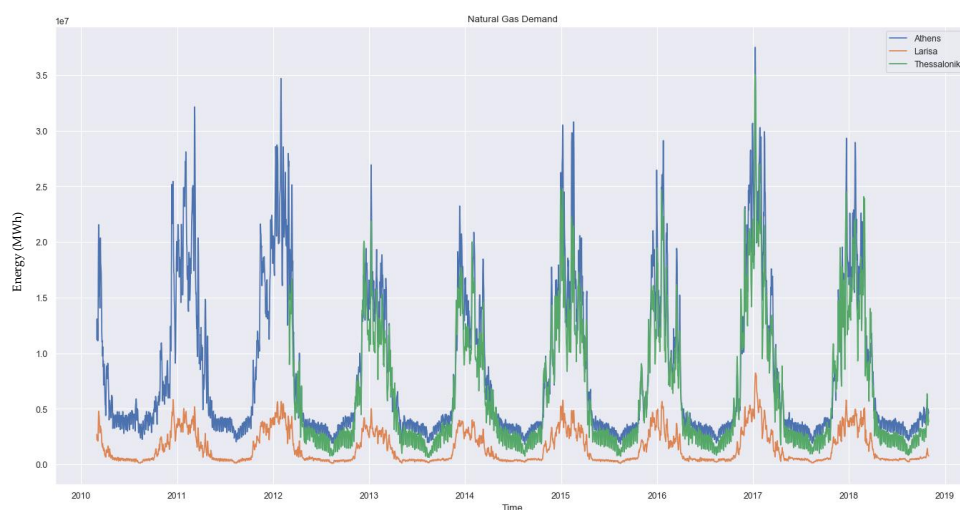
In the present study, the aim is to build a robust forecasting model based on a proposed deep neural network (DNN) and compare it with an artificial neural network (ANN) and a recurrent neural network (RNN), both of which are able to accurately forecast energy demand [43]. This comparative analysis aims to investigate whether the factors that dictate human behavior can offer crucial information and increase the accuracy of our forecasts. The results clearly demonstrate that the proposed DNN approach, with the inclusion of social factors, has attained better accuracies than other state of the art intelligent models for natural gas consumption forecasting.

## 2. Materials and Methods

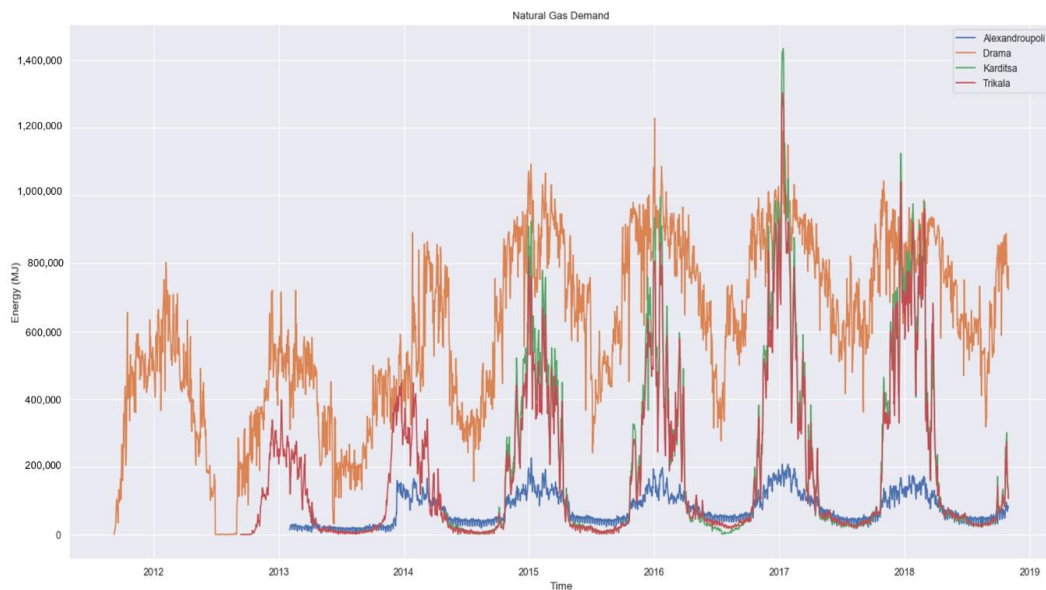
The Hellenic Gas Transmission System Operator S.A. (DESFA) ([www.desfa.gr](http://www.desfa.gr)) is the operator that manages and develops the Greek natural gas infrastructure. DESFA handles all natural gas off-takes, deliveries, and general distribution, as well as the collection of useful data. They have provided with the dataset that was used in this study. Details on the dataset and its features are provided below.

### 2.1. Dataset

The dataset contains historical data of time series from the natural gas consumption of multiple cities all over Greece, as well as the average daily temperature of each city's surrounding area. The data spans from 1 March 2010, or later on for some cities, until 31 October 2018. Specifically for some centralized larger cities of Greece such as Athens and Larissa, where the natural gas distribution system was installed early on, there is data since 1 March 2010, as seen in Figure 1, while in some other large cities, like Thessaloniki, also seen in Figure 1, or smaller ones, like Alexandroupoli, seen in Figure 2, data collection started later on. The exact starting dates of data collection for each city are given later on in Section 3.



**Figure 1.** Natural gas energy consumption [MWh] for Athens, Thessaloniki, and Larissa over the years.



**Figure 2.** Natural gas energy consumption [MWh] for Alexandroupoli, Drama, Karditsa, and Trikala over the years.

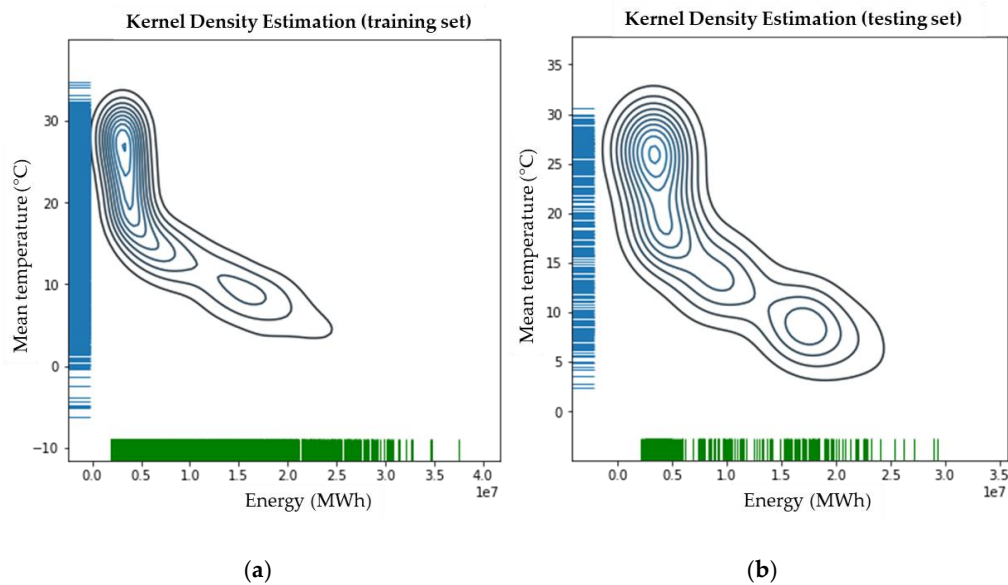
As provided, the dataset contained time and dates, natural gas consumption of each city's distribution point, and the daily average temperature of the area in Celsius degrees. On top of the existing data, social indicators have been added to the dataset such as a month indicator, a day indicator, a weekday/weekend indicator, and a bank holiday indicator. The proper addition of social variables is a key factor to the study since the aim is to see if qualitative social traits can improve the performance of a forecasting model, and by how much compared to other methods.

## 2.2. Feature Engineering

A certain amount of feature engineering is required for the qualitative data to take proper form, in order to be readable by the machine learning algorithms. This takes place during the preprocessing phase and is conducted in the following way: Months and days are described by a name, e.g., September, Tuesday, etc., and need to be transformed into categorical values, e.g., 9, 2, etc., in a serial way. Therefore, the following association is considered: January-1, February-2, etc., and Monday-1, Tuesday-2, etc. Each of these values are then transformed into vectors with the size of the value range of the variable. In detail, the "month" variable contains 12 different values, one for each month, therefore the size of the vector is  $1 \times 12$ . Respectively, the "day" variable contains 7 values, one for each day, therefore the size for this vector is  $1 \times 7$ . Consecutively, the "month" variable is transformed into 12 variables, one for each month, and the "day" variable is transformed into 7 variables, one for each day of the week. The "bank holiday" variable denotes a public or religious holiday that affects social behavior (businesses are closed, people are out celebrating, etc.) and is binary, therefore there is no need for any kind of further transformation.

Time and date data are transformed into a single timestamp variable which is then used as an index, thus leading to the total amount of 22 variables that are being taken as inputs for the modeling of the energy forecast of the proposed methodology. The desirable variable for the forecast is the natural gas energy consumption from the specific distribution point, which is used as output. The correlation of the consumption energy with the mean daily temperature for the city of Athens is shown in Figure 3.

The correlation plots of consumption energy and mean temperature for all the investigated cities are given in the Appendix A, where it is obvious that not all cities have the same pattern of correlation between the mean temperature and the consumption of natural gas. This variation in patterns is one of the reasons that the implemented models achieve different accuracies for each different city, as will be shown later.



**Figure 3.** Correlation of the energy demand and the mean temperature for the training set (a) and the test set (b) for the city of Athens.

### 3. Methodology

Three types of neural network variants were modeled and tested for accurately forecasting natural gas energy demand. Their implementations differ even though they all belong to the neural network family of algorithms. Each approach has different input variables, however, the general approach remains the same; historical data train a model that is able to produce accurate forecasts of natural gas energy consumption. The approaches, as well as the general process flow, are described in the following sections.

#### 3.1. Artificial Neural Networks (ANN)

Artificial Neural Networks (ANN) [44] were hypothesized as a method to imitate the human brain and its functions while it performs cognitive tasks, or when it learns. For the mathematical structure of such networks, nodes are used to represent the neurons, and layers are used to represent their interactive synapses in the same fashion as it is within the brain. They received wide acceptance ever since they were conceptualized, however, they started gaining more and more fame ever since the computational power, especially in graphic processing units, has become cheaper and thus easily attainable. Another important reason for their wide acceptance is the vast availability of immense amounts of data that have been collected throughout the past years. This adoption has enhanced the scientific progress of the ANN algorithms, which started from the single-layer perceptron [45], moved to multi-layer perceptron [46], introduced the back-propagation algorithm [47], and led to many new derivatives of ANNs, the most well-known being the deep neural networks that are described in the next paragraph.

#### 3.2. Long Short-Term Memory Networks (LSTM)

Long Short-Term Memory are also neural networks which are built upon a recurrent fashion by introducing memory cells and their in-between connections, in order to construct a graph directed over a sequence. In general, recurrent networks process sequences by using said memory cells in a fashion that is different than that of the simple ANNs, and even though they are well suited for problems with time dependency, they often face the problem of vanishing gradients, or not being able to “memorize” large portions of data. LSTMs [48] solved this problem because of the specific cell structure they have, which allows the network the ability to variate the amount of retained information. These cell

structures are called gates, and they control which information is stored in the long memory and which is discarded, thus optimizing the memorizing process. Dynamic temporal behavior problems, i.e., time sequences, were suited for such approaches.

### 3.3. Deep Neural Networks (DNN)

Deep neural networks [49] are the basis of deep learning, one of the most influential areas of the artificial intelligence for the past decade. Based on the ANN, the DNN is comprised by more layers and nodes in the same sequential fashion. For very deep implementations, the problem of “vanishing” or “exploding” gradients would not allow the network to learn properly, therefore new techniques were introduced, such as “identity shortcut connections” as seen in Resnet [50], as well as others, in order to solve these obstacles. The “deep” approach has been used in all derivatives of neural networks. Deep convolutional neural networks (CNN) have been used for image classification and object detection [51,52], and deep recurrent neural networks (RNN) have been used for word prediction [53] and time series forecasting [54].

### 3.4. Process Flow

In order to accurately model the natural gas energy forecast, a specific process flow has been designed. The steps that have been followed, are described below.

#### 3.4.1. Preprocessing

The initial preprocessing is focused on the organization of the original unstructured data. Same-variable columns have been aligned and set in correct time and date order, and columns that were empty or contained plenty of NaN (Not a Number) values were removed completely. The columns “date” and “time” were merged into a TIMEDATE column, which was consecutively used as index. The MONTH and DAY variables were manually added and later transformed into multiple categorical variables as described in Section 2.

For the ANN implementation, only the daily mean temperature is considered as input, for the LSTM, only the energy consumption is used as both input and output (past and present values respectively), and for the DNN implementation, all the aforementioned variables are used with the addition of vectorial representations of qualitative variables. The inputs that are used in each neural network variant, as well as the output variable, that are being used in this study, are shown in Table 1.

**Table 1.** Variables that are used as output and input for each implementation of the neural network variants. ANN: Artificial neural network; LSTM: Long Short-Term Memory; DNN: Deep neural network.

	Output		Inputs				
	Energy Current Day (MJ)	Daily Mean Temperature (°C)	Energy 1 Previous Day (MJ)	Energy 2 Previous Days (MJ)	Month (Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct, Nov, Dec)	Day of the Week (Mon, Tue, Wen, Thu, Fri, Sat, Sun)	Bank Holiday (Yes, No)
ANN	×	×					
LSTM	×						
DNN	×	×	×	×	×	×	×

#### 3.4.2. Data Split

The data was split in the following way: The last year, ranging from 1 November 2017 till 31 October 2018, was used as the testing period for all models, approaches, and for all cities. The starting date of data collection for each city, as well as the ratio of training/testing portions of each dataset, is shown in Table 2.

During the training phase, 20% of the training dataset is used as a validation set, in order to identify whether our model tends to under- or overfit, and to be able to measure its loss and accuracy.

**Table 2.** Starting dates and ratio of training/testing portions of the dataset per city.

City	Start Date	Ratio of Training/Testing
Agioi Theodoroi	07/06/2014	3.41
Alexandroupoli	02/02/2013	4.78
Athens	01/03/2010	7.68
Drama	07/09/2011	6.15
Karditsa	01/05/2014	3.51
Kilkis	01/03/2010	7.68
Lamia	01/02/2013	4.75
Larissa	01/03/2010	7.68
Laurio	01/03/2010	7.68
Markopoulo	01/03/2010	7.68
Serres	01/06/2013	4.42
Thessaloniki	01/03/2012	5.67
Trikala	12/09/2012	5.14
Volos	01/03/2010	7.68
Xanthi	01/03/2010	7.68

For the ANN implementation, only the numerical variables were used as input, i.e., the energy consumption of 2 prior days and the mean temperature. For the LSTM implementation, the natural gas energy consumption is used both as input and output, so the previous 365 values are used to find the future trend, i.e., the energy demand. For the DNN implementation, all the variables described in paragraph 2.2 are used as inputs, and the natural gas energy consumption is used as output, as it is with all implementations.

#### 3.4.3. Standardization

All data was normalized with Python’s SciKit Learn MinMaxScaler, between 0 and 1 values [55]. This way, the performance metrics are common for all cities, therefore direct comparisons can be conducted, but more importantly because during training it allows the non-convex cost function to converge to the global minimum faster and in a more appropriate fashion.

#### 3.4.4. Processing

In all ANN, LSTM, and DNN approaches, we used the “ReLU” activation function, the “adam” optimizer for the cost function, mean squared error for measuring the loss of training and validation, and an early stopping function (EarlyStopping callback in Keras [56] with 5 epochs patience) in order to avoid overfitting [57]. For the LSTM approach, the previous 365 energy values were used as input, and the consumption for the next 365 days was forecasted. After the model training, all models are being evaluated on the testing portion of the dataset. Their performance is measured based on certain metrics that are described in the following section.

### 4. Evaluation metrics

The performance and robustness of each studied natural gas forecasting model is based on four of the most common evaluation metrics. Mean square error (MSE), absolute error (MAE), mean absolute percentage error (MAPE), and coefficient of determination ( $R^2$ ) are all being used in order to determine the best performing model [5,7,43].

All the modelling, tests, and evaluations were performed with the use of Python 3.7 and the Tensorflow 1.14, Keras 2.3, SciKit Learn 0.21, Pandas 0.25, Numpy 1.17, Matplotlib 3.1, Seaborn 0.9 libraries. The mathematical equations of these evaluation metrics are described below:

Mean Squared Error:

$$\text{MSE} = \frac{1}{T} \sum_{t=1}^T (Z(t) - X(t))^2, \quad (1)$$

Mean Absolute Error:

$$\text{MAE} = \frac{1}{T} \sum_{t=1}^T |Z(t) - X(t)|, \quad (2)$$

Mean Absolute Percentage Error:

$$\text{MAPE} = \frac{1}{T} \sum_{t=1}^T \left| \frac{Z(t) - X(t)}{Z(t)} \right|, \quad (3)$$

Coefficient of Correlation:

$$R = \frac{T \sum_{t=1}^T Z(t) \cdot X(t) - (\sum_{t=1}^T Z(t)) (\sum_{t=1}^T X(t))}{\sqrt{T \sum_{t=1}^T (Z(t))^2 - (\sum_{t=1}^T Z(t))^2} \cdot \sqrt{T \sum_{t=1}^T (X(t))^2 - (\sum_{t=1}^T X(t))^2}}, \quad (4)$$

where  $X(t)$  is the predicted value,  $Z(t)$  is the real value,  $t$  is the iteration at each point ( $t = 1, \dots, T$ ), and  $T$  is the number of testing records.

Low MSE, MAE, and MAPE values signify small error, therefore higher accuracy. On the contrary,  $R^2$  value close to 1 is preferred, signifying better performance for the model and that the regression curve is well fit on the data. A coefficient of determination value of 1 would signify that the regression line fits the data perfectly; however, this could also denote overfitting on the data.

To summarize, the whole process so far can be visually represented into an algorithmic flowchart. Starting from the data preprocessing, to the training of the algorithm and the prediction of the results, all the consecutive steps are shown in Figure 4.

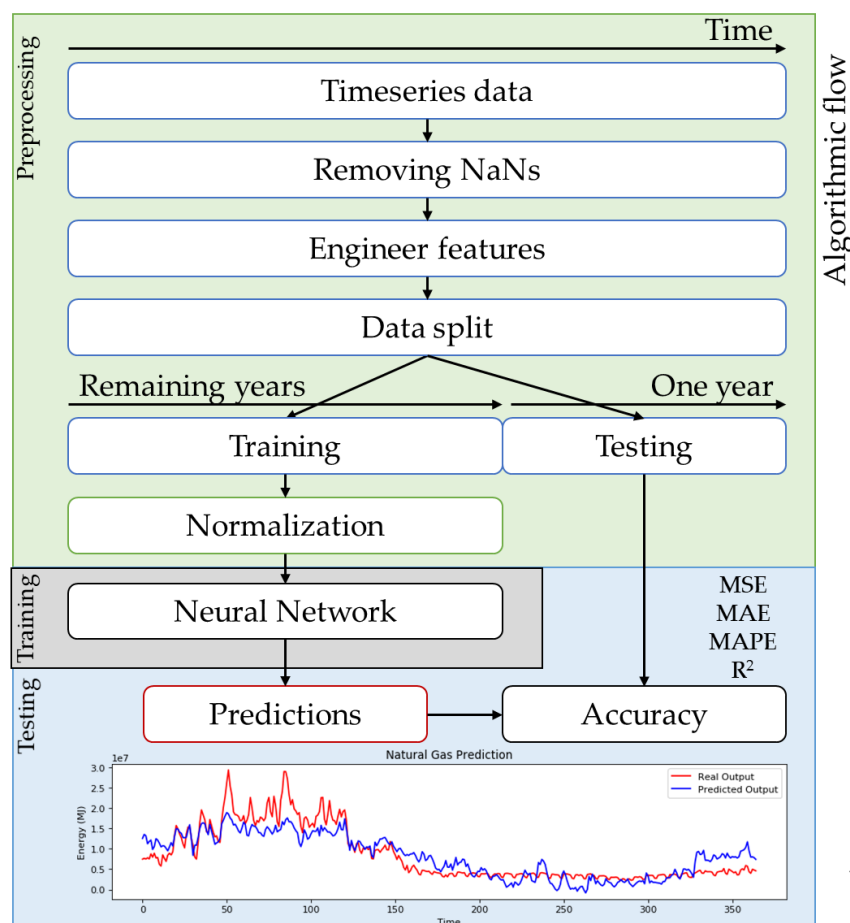


Figure 4. Process flow for the algorithm that has been used for the study.

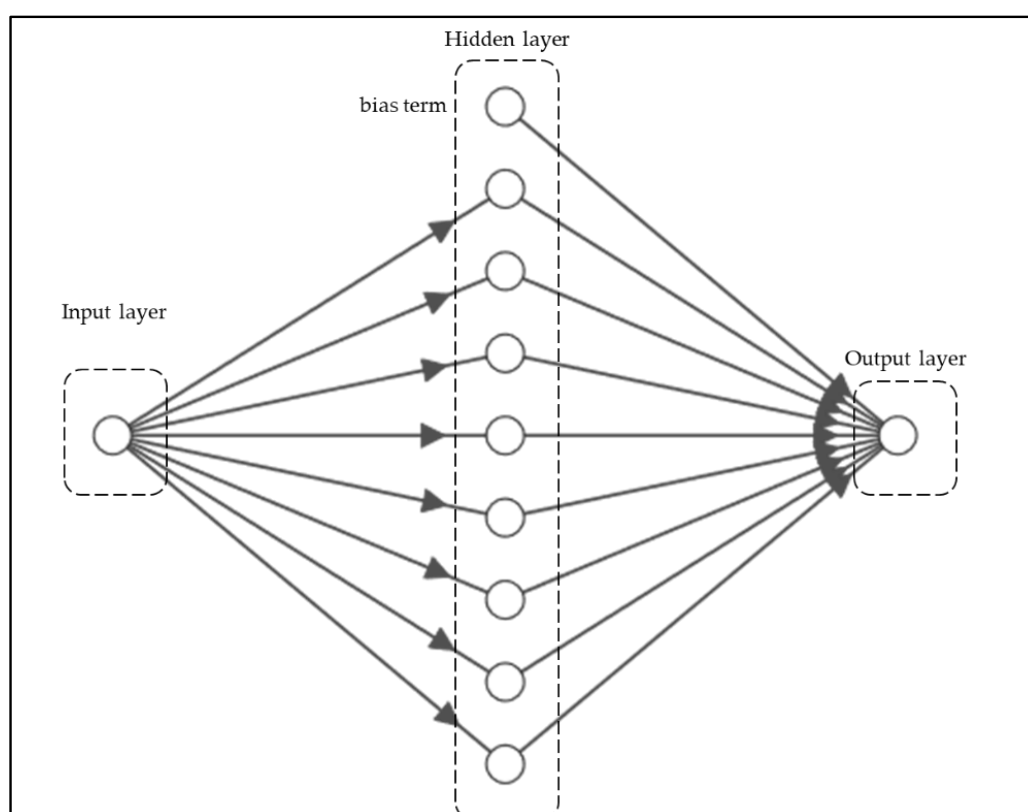


## 5. Results

Athens was chosen as a reference point for searching the best fitting parameters of each method. The outcomes of each parameter selection for the ANN, the LSTM, and the DNN architectures are presented in paragraphs 5.1, 5.2, and 5.3. The parameters that resulted to the best performing model for all 15 cities are presented in paragraph 5.4. For the evaluation of all tests, MSE, MAE, MAPE, and  $R^2$  were used.

### 5.1. Results from ANN

For the ANN implementation, an architecture of a single-layer perceptron with 8 nodes in the hidden layer was selected, after a concise exploratory analysis. The simple ANN was selected for benchmarking purposes. Having a simplistic model as a baseline, we can investigate the performance improvement of the other approaches. The initial architecture, seen in Figure 5, was tested and evaluated without any dropout function, and the performance metrics are shown in Table 3.



**Figure 5.** Architecture of the ANN approach.

**Table 3.** Performance metrics of the selected ANN architecture. MSE: Mean square error; MAE: Mean absolute error; MAPE: Mean absolute percentage error;  $R^2$ : Coefficient of determination.

ANN		Architecture Selection			
Layers	Nodes	MSE ( $MJ^2$ )	MAE ( $MJ^2$ )	MAPE (%)	$R^2$
1	8	$7.70 \times 10^{-3}$	$7.01 \times 10^{-2}$	16.41	0.87

Next, the effect of the dropout rate is investigated, and in order to understand how it affects the model's performance, four distinct percentages have been tested. The results are presented in Table 4.

**Table 4.** Comparison of dropout rate in ANN.

1 Layer/ 8 Nodes		Dropout Comparison		
Dropout	MSE (MJ <sup>2</sup> )	MAE (MJ <sup>2</sup> )	MAPE (%)	R <sup>2</sup>
0	$7.70 \times 10^{-3}$	$7.01 \times 10^{-2}$	16.41	0.87
0.25	$8.20 \times 10^{-3}$	$6.94 \times 10^{-2}$	30.30	0.86
0.50	$9.00 \times 10^{-3}$	$7.41 \times 10^{-2}$	52.74	0.84
0.75	$1.25 \times 10^{-2}$	$8.66 \times 10^{-2}$	73.02	0.78

The ability of forecasting the energy demand in yearly intervals was also investigated. Even though this study is focused in one-year ahead forecasts, a timeframe of yearly depths up to four years ahead was investigated. This investigation was conducted in order to see the magnitude of the forecasts' accuracy through time, and the results are presented in Table 5.

**Table 5.** Comparison of year-ahead forecasting in ANN.

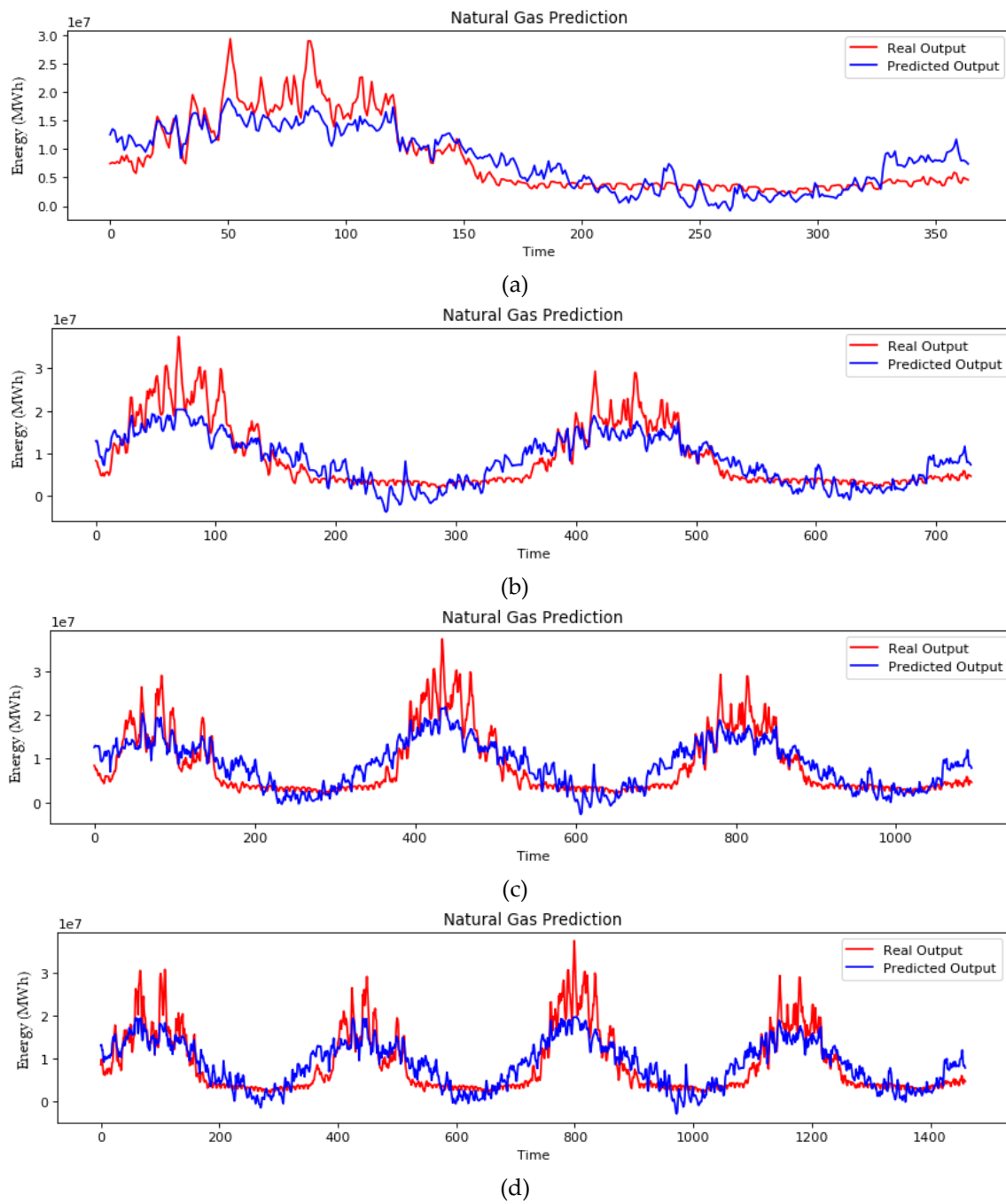
1 Layer/ 8 Nodes		Forecasting Comparison		
Years ahead	MSE (MJ <sup>2</sup> )	MAE (MJ <sup>2</sup> )	MAPE (%)	R <sup>2</sup>
1	$7.70 \times 10^{-3}$	$7.01 \times 10^{-2}$	16.41	0.87
2	$1.08 \times 10^{-1}$	$3.07 \times 10^{-1}$	48.61	-1.39
3	$9.28 \times 10^{-2}$	$2.83 \times 10^{-1}$	52.46	-1.31
4	$1.42 \times 10^{-1}$	$3.53 \times 10^{-1}$	49.73	-2.57

The ANN approach is able to capture the general trend, however, it deviates significantly from the real consumption values, something that could signify that the model cannot give better forecasts for longer time ahead. Figure 6 shows the plots of the ANN implementation for (a) one-, (b) two-, (c) three-, and (d) four-year ahead forecasting. The prediction of the energy demand in MWh is depicted in blue, and the real output is depicted in red.

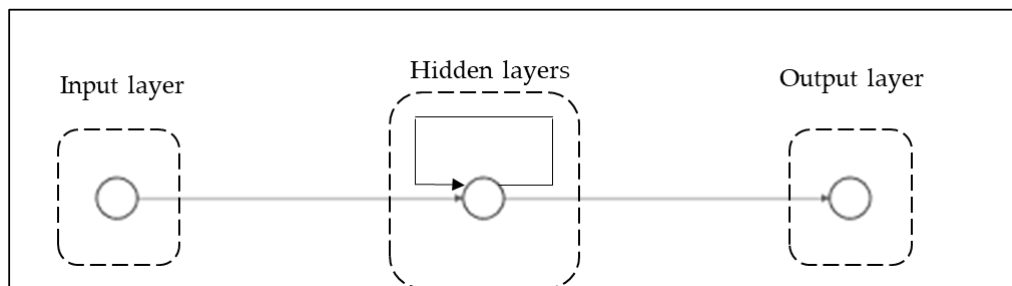
It is evident that the ANN model, even though it can follow the trend of the fluctuation, fails to forecast accurately the consumption of the natural gas. Furthermore, the more the forecasting time increases, the greater this deviation gets. Even though ANNs are powerful algorithms for forecasting, in this particular problem, the single-layer perceptron is not enough to model the problem accurately.

## 5.2. Results from LSTM

An investigatory analysis was conducted also for the LSTM implementation. The number of layers and memory units were explored in order to find the best combination, which was comprised of one LSTM layer with 200 memory units. The architecture of this implementation is seen in Figure 7, and its performance is shown in Table 6.



**Figure 6.** Forecasts of (a) one-, (b) two-, (c) three-, and (d) four-year ahead of natural gas demand with the use of ANN.



**Figure 7.** The architecture of the LSTM approach.

**Table 6.** Performance metrics of the selected LSTM architecture.

LSTM		Architecture Selection			
Layers	Units	MSE (MJ <sup>2</sup> )	MAE (MJ <sup>2</sup> )	MAPE (%)	R <sup>2</sup>
1	200	$2.20 \times 10^{-3}$	$2.95 \times 10^{-2}$	11.07	0.96

The dropout's effect on the LSTM implementation was also investigated, and its effect on the performance of the model is seen in Table 7.

**Table 7.** Comparison of dropout rate in LSTM.

1 Layer/ 200 Units		Dropout Comparison		
Dropout	MSE (MJ <sup>2</sup> )	MAE (MJ <sup>2</sup> )	MAPE (%)	R <sup>2</sup>
0	$2.20 \times 10^{-3}$	$2.95 \times 10^{-2}$	11.07	0.96
0.25	$2.40 \times 10^{-3}$	$3.31 \times 10^{-2}$	20.94	0.96
0.50	$2.20 \times 10^{-3}$	$3.10 \times 10^{-2}$	18.50	0.96
0.75	$2.10 \times 10^{-3}$	$2.92 \times 10^{-2}$	9.77	0.96

Dropout application seems to increase performance over a non-dropout approach, and in fact the highest rate selected has given the best results.

Again, forecasts of up to four years ahead were evaluated in order to investigate how the predictions are affected. The results are presented in Table 8.

**Table 8.** Comparison of year-ahead forecasting in LSTM.

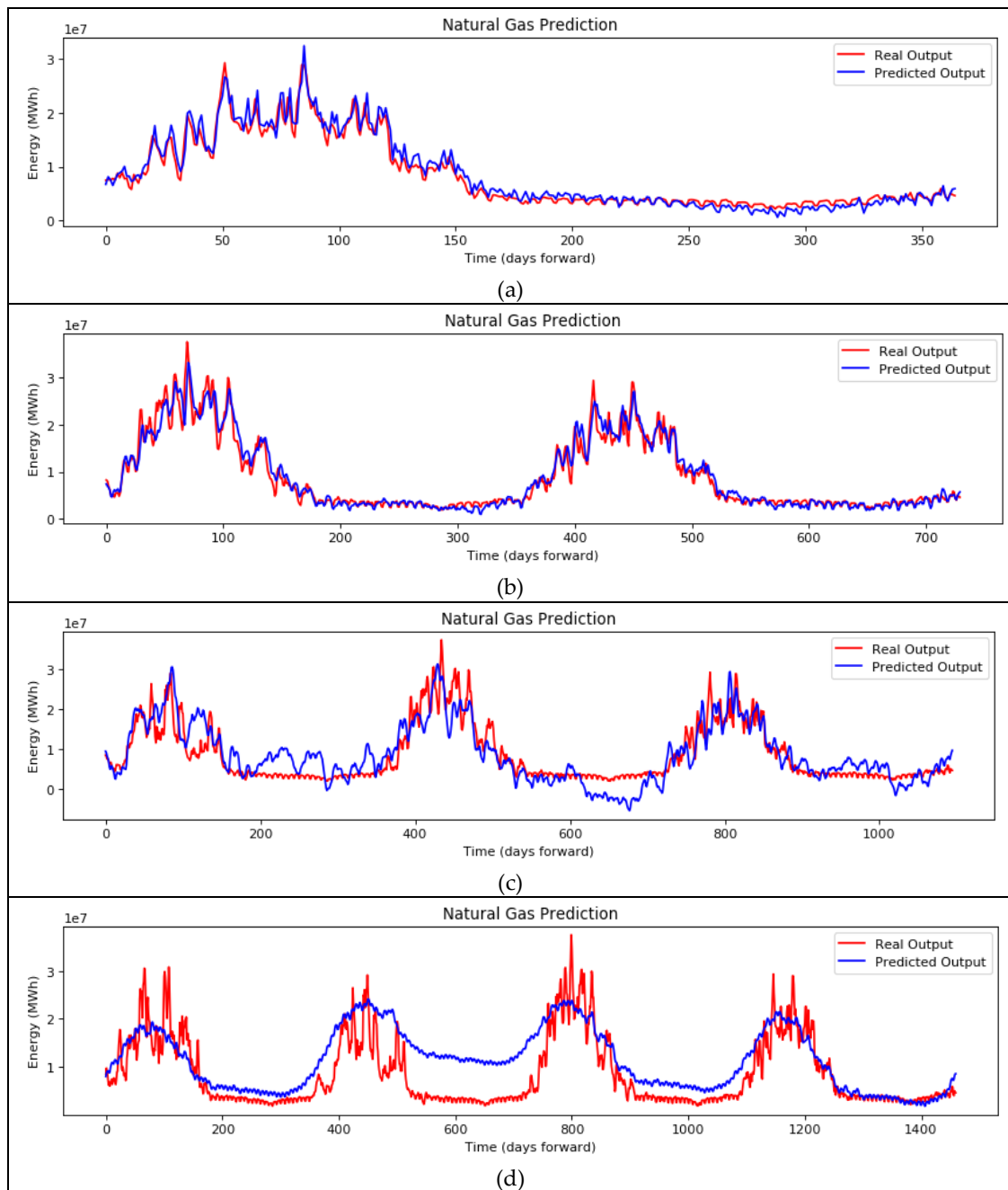
1 Layer/ 8 Nodes		Forecasting Comparison		
Years ahead	MSE (MJ <sup>2</sup> )	MAE (MJ <sup>2</sup> )	MAPE (%)	R <sup>2</sup>
1	$2.10 \times 10^{-3}$	$2.92 \times 10^{-2}$	9.77	0.96
2	$6.30 \times 10^{-3}$	$5.90 \times 10^{-2}$	15.00	0.86
3	$5.26 \times 10^{-2}$	$2.06 \times 10^{-1}$	59.06	-0.31
4	$8.73 \times 10^{-2}$	$2.44 \times 10^{-1}$	80.87	-1.20

The plots of the LSTM setup are shown in Figure 8 for (a) one-, (b) two-, (c) three-, and (d) four-year ahead forecasting. The prediction of the energy demand in MWh is depicted in blue, and the real output is depicted in red.

In the case of LSTM implementation, it is clear that the forecasts for the one- and two-year ahead demands are more accurate than that of the ANN implementation. However, it is evident that anything beyond the two-year ahead forecast is tremendously inaccurate, resulting even in negative R<sup>2</sup> values. LSTMs can offer excellent accuracy for single-variable time series; however, it is evident that they are highly susceptible to the depth of the forecasting period, as well as to the data that are required for proper training.

### 5.3. Results from DNN

For the DNN implementation, a deeper, more complex network was constructed that is comprised of 4 hidden layers with 32 nodes in each layer. The proposed architecture is structured in such way so that it can take as input the vectorial representations of categorical values, the ones mentioned above, the quantitative values from the current time (in each step), as well as the energy values from past inputs. The architecture of the ANN approach is seen in Figure 9. No dropout was initially set for the exploratory analysis, and the performance metrics for the selected setup is shown in Table 9.



**Figure 8.** One year ahead forecast of natural gas demand with the use of LSTM.

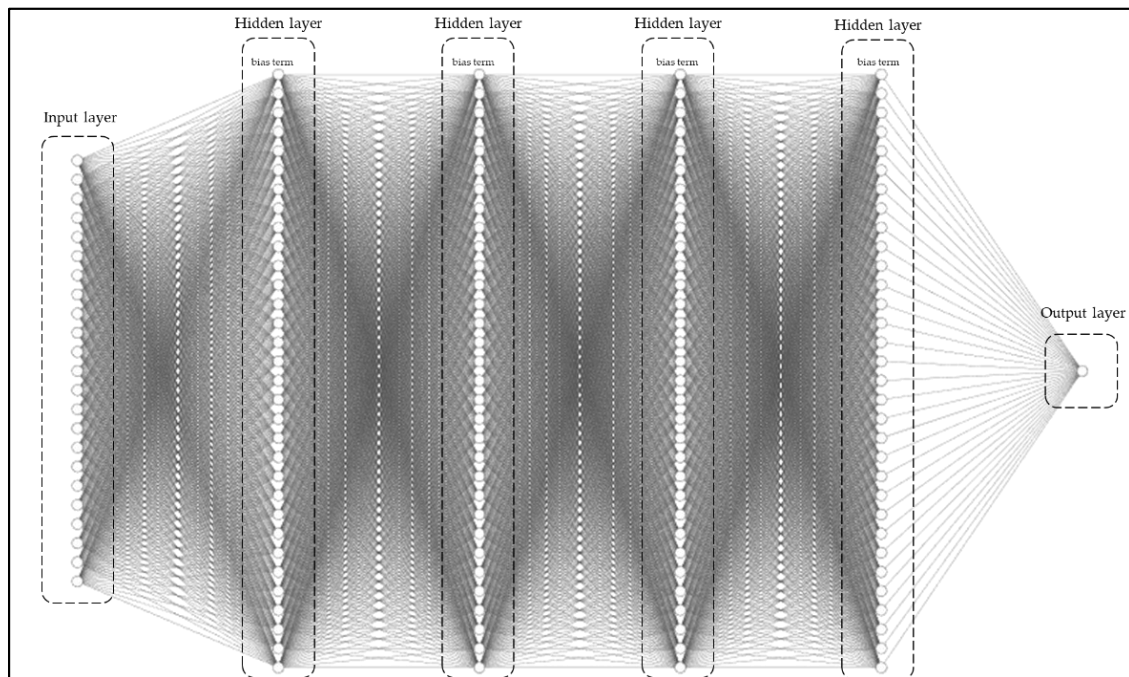


Figure 9. Architecture of the DNN approach.

Table 9. Performance metrics of the selected DNN architecture.

DNN		Architecture Selection			
Layers	Nodes	MSE ( $\text{MJ}^2$ )	MAE ( $\text{MJ}^2$ )	MAPE (%)	$R^2$
4	32	$1.70 \times 10^{-3}$	$2.75 \times 10^{-2}$	7.53	0.97

The effect of the dropout rate is investigated once again, and the same four percentages have been tested and evaluated, the results of which are presented in Table 10.

Table 10. Comparison of dropout rate in DNN.

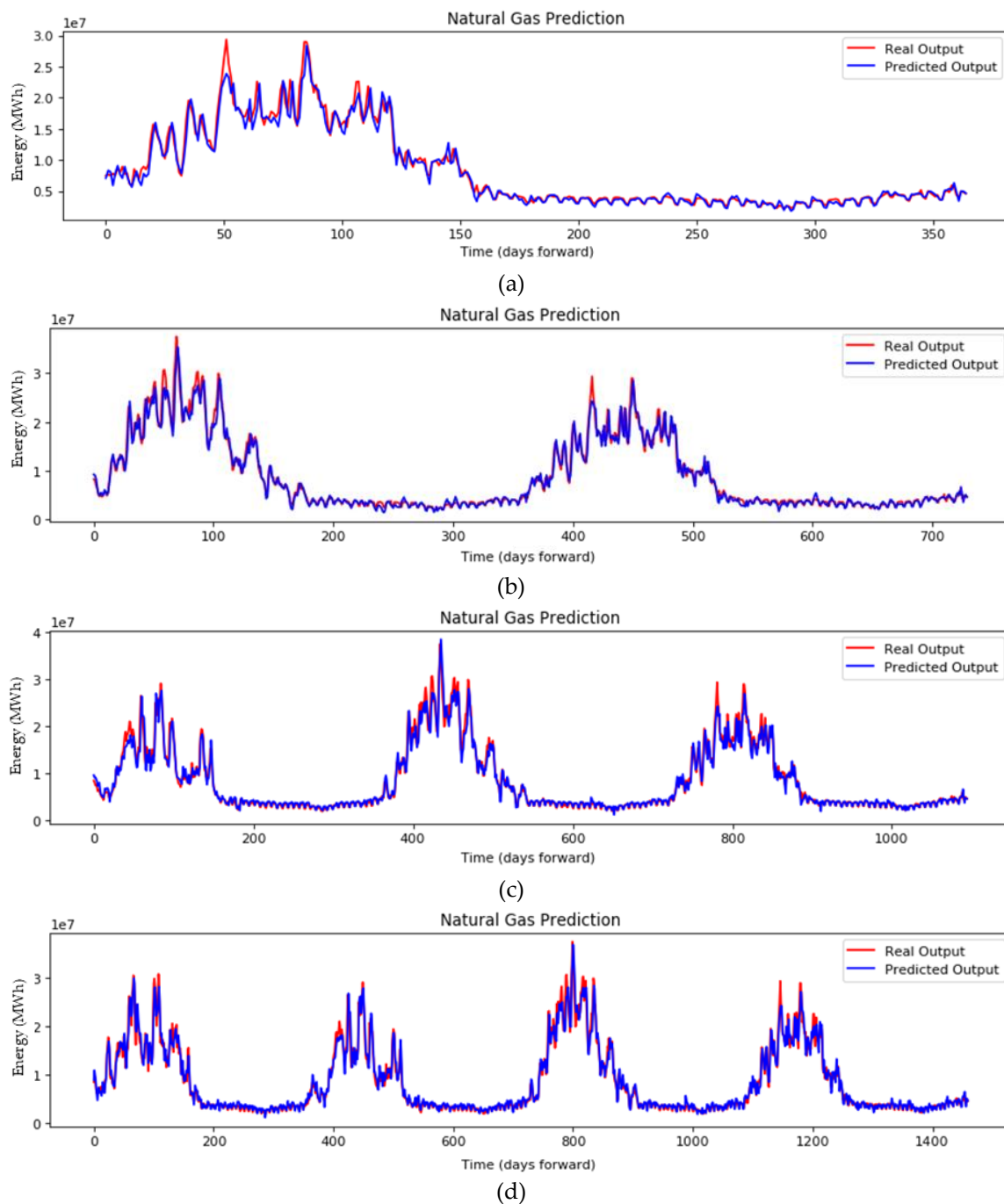
4 Layer/ 32 Nodes		Dropout Comparison			
Dropout	MSE ( $\text{MJ}^2$ )	MAE ( $\text{MJ}^2$ )	MAPE (%)	$R^2$	
0	$1.70 \times 10^{-3}$	$2.75 \times 10^{-2}$	7.53	0.97	
0.25	$2.20 \times 10^{-3}$	$3.11 \times 10^{-2}$	10.98	0.96	
0.50	$2.90 \times 10^{-3}$	$4.32 \times 10^{-2}$	26.45	0.95	
0.75	$3.00 \times 10^{-3}$	$3.82 \times 10^{-2}$	48.61	0.95	

It is evident that the proposed DNN model performs better than the ANN and the LSTM. Testing its forecasting capabilities for up to four years ahead will show its ability to generalize well and properly model the consumption pattern. The results are presented in Table 11.

Table 11. Comparison of year-ahead forecasting in DNN.

4 Layer/ 32 Nodes		Forecasting Comparison			
Years ahead	MSE ( $\text{MJ}^2$ )	MAE ( $\text{MJ}^2$ )	MAPE (%)	$R^2$	
1	$1.70 \times 10^{-3}$	$2.75 \times 10^{-2}$	7.53	0.97	
2	$1.50 \times 10^{-3}$	$2.76 \times 10^{-2}$	8.98	0.97	
3	$1.30 \times 10^{-3}$	$2.06 \times 10^{-2}$	8.68	0.97	
4	$1.70 \times 10^{-3}$	$3.22 \times 10^{-2}$	9.40	0.96	

At this point, it is interesting to mention that the accuracy of the predictions is hardly affected for up to four years ahead. This is due to the yearly periodicity of the energy demand that is caused not only by the general temperature trends, but also by the social aspects that govern human behavior in certain periods of time. Figure 10 shows the plots of the best performing setup for the DNN implementation for (a) one-, (b) two-, (c) three-, and (d) four-year ahead forecasting. The prediction of the energy demand in MWh is depicted in blue, and the real output is depicted in red.



**Figure 10.** Forecasts of (a) one-, (b) two-, (c) three-, and (d) four-year ahead of natural gas demand with the use of DNN.

For the proposed DNN, it is clear that its forecasting capabilities surpass by far the ANN and the state-of-the-art LSTM models. The inclusion of qualitative social variables alongside measurable

quantities has improved not only the accuracy of the forecasts, but also the depth of forecasting time into the future. This is an indicator that the deeper network, alongside with behavioral knowledge, has offered a generalized understanding of the energy consumption trend.

#### 5.4. Comparison (Cities)

The trained models of all approaches were applied on a range of fifteen cities all around Greece for the sake of comparison. Each city had different energy distributions during the documented years depending on its size, population, and specific natural gas network characteristics. Testing all implementations on different cities, offers insight on whether each model can provide accurate one-year ahead forecasts in cities that are in different geographical locations, but also with different behavioral patterns. The confidence interval (CI) [58] is also included in the following analysis, to demonstrate the range of energy values, in which 95% of the predictions fall within, for each city. The performance metrics for all cities are shown in Table 12 for the ANN, Table 13 for the LSTM, and Table 14 for the proposed DNN.

**Table 12.** Comparison of cities for the ANN implementation.

ANN		Cities Comparison			
Cities	MSE (MJ <sup>2</sup> )	MAE (MJ <sup>2</sup> )	MAPE (%)	R <sup>2</sup>	CI
Agioi Theodoroi	$4.96 \times 10^{-2}$	$1.70 \times 10^{-1}$	58.29	0.14	[92,246–101,257]
Alexandroupoli	$8.30 \times 10^{-3}$	$6.72 \times 10^{-2}$	13.43	0.89	[80,527–88,566]
Athens	$2.40 \times 10^{-3}$	$3.53 \times 10^{-2}$	10.94	0.96	[7,710,057–9,006,556]
Drama	$8.50 \times 10^{-3}$	$6.77 \times 10^{-2}$	6.42	0.78	[736,302–764,636]
Karditsa	$2.60 \times 10^{-3}$	$3.54 \times 10^{-2}$	28.66	0.97	[234,893–293,622]
Kilkis	$2.87 \times 10^{-2}$	$1.23 \times 10^{-1}$	20.16	0.43	[1,023,806–1,088,807]
Lamia	$3.96 \times 10^{-2}$	$1.62 \times 10^{-1}$	33.76	0.20	[125,129–135,568]
Larissa	$2.60 \times 10^{-3}$	$3.43 \times 10^{-2}$	14.18	0.96	[1,329,088–1,609,373]
Laurio	$1.32 \times 10^{-2}$	$7.59 \times 10^{-2}$	86,605.00	0.74	[6,004,693–7,432,331]
Markopoulo	$3.06 \times 10^{-2}$	$1.38 \times 10^{-1}$	19.38	0.27	[250,400–264,171]
Serres	$3.60 \times 10^{-3}$	$4.52 \times 10^{-2}$	11.69	0.96	[400,111–467,284]
Thessaloniki	$3.40 \times 10^{-3}$	$4.52 \times 10^{-2}$	16.80	0.95	[6,283,167–7,449,531]
Trikala	$2.40 \times 10^{-3}$	$2.76 \times 10^{-2}$	25.62	0.97	[213,058–266,020]
Volos	$3.90 \times 10^{-3}$	$4.92 \times 10^{-2}$	11.41	0.91	[1,662,841–1,845,579]
Xanthi	$2.94 \times 10^{-2}$	$1.33 \times 10^{-1}$	36.12	0.17	[153,935–168,218]

**Table 13.** Comparison of cities for LSTM implementation.

LSTM		Cities Comparison			
Cities	MSE (MJ <sup>2</sup> )	MAE (MJ <sup>2</sup> )	MAPE (%)	R <sup>2</sup>	CI
Agioi Theodoroi	$3.56 \times 10^{-2}$	$1.50 \times 10^{-1}$	59.68	0.38	[86,688–94,424]
Alexandroupoli	$4.20 \times 10^{-3}$	$4.60 \times 10^{-2}$	8.29	0.94	[86,103–95,178]
Athens	$2.10 \times 10^{-3}$	$2.92 \times 10^{-2}$	9.77	0.96	[7,634,555–8,986,641]
Drama	$1.06 \times 10^{-2}$	$7.82 \times 10^{-2}$	9.86	0.72	[765,187–790,777]
Karditsa	$8.40 \times 10^{-3}$	$7.61 \times 10^{-2}$	72.99	0.89	[290,871–357,268]
Kilkis	$1.15 \times 10^{-2}$	$8.17 \times 10^{-2}$	10.97	0.77	[1,008,757–1,071,517]
Lamia	$1.86 \times 10^{-2}$	$1.11 \times 10^{-1}$	32.19	0.62	[157,729–167,902]
Larissa	$3.30 \times 10^{-3}$	$4.43 \times 10^{-2}$	17.39	0.95	[1,377,457–1,661,699]
Laurio	$2.45 \times 10^{-2}$	$1.30 \times 10^{-1}$	78,544.00	0.53	[8,712,884–10,054,310]
Markopoulo	$8.00 \times 10^{-3}$	$6.79 \times 10^{-2}$	9.70	0.81	[258,990–272,945]
Serres	$4.00 \times 10^{-3}$	$4.85 \times 10^{-2}$	15.78	0.96	[375,137–443,421]
Thessaloniki	$3.90 \times 10^{-3}$	$4.79 \times 10^{-2}$	19.54	0.94	[6,268,458–7,571,520]
Trikala	$4.60 \times 10^{-3}$	$5.41 \times 10^{-2}$	32.77	0.94	[256,379–318,363]
Volos	$4.80 \times 10^{-3}$	$5.21 \times 10^{-2}$	13.18	0.89	[1,480,596–1,649,836]
Xanthi	$9.80 \times 10^{-3}$	$7.60 \times 10^{-2}$	24.43	0.72	[169,605–180,582]



**Table 14.** Comparison of cities for the proposed DNN approach.

DNN Cities	Cities Comparison				
	MSE (MJ <sup>2</sup> )	MAE (MJ <sup>2</sup> )	MAPE (%)	R <sup>2</sup>	CI
Agioi Theodoroi	$2.46 \times 10^{-2}$	$1.19 \times 10^{-1}$	38.70	0.57	[89,946–98,778]
Alexandroupoli	$3.00 \times 10^{-3}$	$4.01 \times 10^{-2}$	7.57	0.96	[80,268–88,430]
Athens	$1.70 \times 10^{-3}$	$2.75 \times 10^{-2}$	7.53	0.97	[7,571,538–8,863,422]
Drama	$1.03 \times 10^{-2}$	$7.69 \times 10^{-2}$	7.32	0.73	[729,057–757,898]
Karditsa	$3.50 \times 10^{-3}$	$4.84 \times 10^{-2}$	25.10	0.95	[230,724–290,272]
Kilkis	$8.20 \times 10^{-3}$	$6.73 \times 10^{-2}$	10.36	0.84	[984,497–1,049,268]
Lamia	$1.41 \times 10^{-2}$	$9.71 \times 10^{-2}$	21.23	0.71	[126,988–137,111]
Larissa	$1.10 \times 10^{-3}$	$2.40 \times 10^{-2}$	9.95	0.98	[1,413,437–1,698,290]
Laurio	$2.03 \times 10^{-2}$	$1.17 \times 10^{-1}$	78,518.00	0.61	[6,822,570–8,233,887]
Markopoulo	$6.80 \times 10^{-3}$	$6.29 \times 10^{-2}$	10.10	0.84	[245,857–259,800]
Serres	$4.50 \times 10^{-3}$	$5.21 \times 10^{-2}$	17.46	0.95	[396,910–466,370]
Thessaloniki	$1.60 \times 10^{-3}$	$2.72 \times 10^{-2}$	10.25	0.98	[6,289,857–7,456,553]
Trikala	$4.40 \times 10^{-3}$	$5.04 \times 10^{-2}$	22.98	0.95	[222,492–276,310]
Volos	$5.10 \times 10^{-3}$	$5.70 \times 10^{-2}$	11.73	0.89	[1,644,087–1,825,076]
Xanthi	$1.02 \times 10^{-2}$	$7.44 \times 10^{-2}$	26.04	0.71	[165,997–179,896]

For the ANN implementation, the performance of the model ranges from ~14% for Agioi Theodoroi till ~97% for Trikala considering R<sup>2</sup>. Seven out of fourteen cities achieved an accuracy of >90%, however, for the other seven cities, the performance of the model is disappointing.

For the LSTM implementation, the prediction accuracies are better than the ANN, ranging from ~39% for Agioi Theodoroi to ~96% for Athens, using R<sup>2</sup> as the primary metric. Here, six cities achieved an accuracy of >94%, with the rest achieving higher accuracy when compared to the ANN.

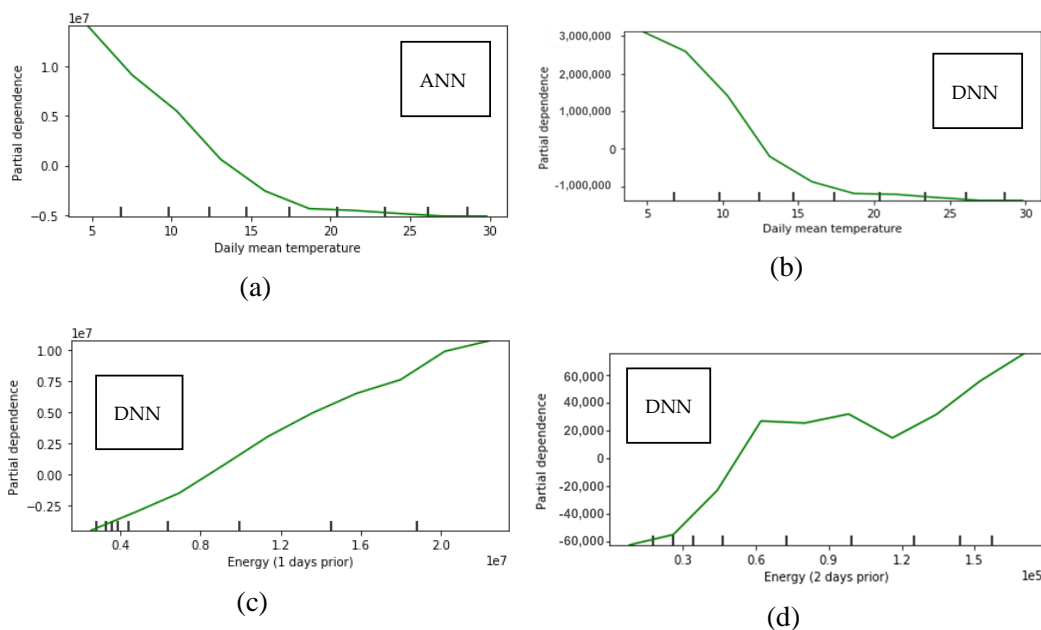
For the proposed DNN implementation, the performance of the model ranges from ~58% for Agioi Theodoroi till ~99% for Larissa considering R<sup>2</sup>. For seven out of fourteen cities, the proposed methodology achieved an accuracy of >94%, which is considered very satisfactory for prediction, considering that the MSE of these models is also very low.

### 5.5. Sensitivity Analysis

A sensitivity analysis is conducted on the dataset of Athens. The selected method for the sensitivity analysis is the Partial Dependence Plots (PDP) [59,60], where target variables (features) of the dataset are investigated through their range of values in order to visualize their dependence to the target outcome. The numerical variables used in the datasets, i.e., daily mean temperature, 1-day and 2-days prior consumption energy are used as the target features and the dependence of the target outcome, i.e., the present-day consumption energy, is shown in Figure 11.

Both for the ANN and the DNN, the mean daily temperature is inversely proportional to the daily energy consumption, which is expected since heating needs are lower when the external temperature is high. For the DNN, the 1-day prior consumption is directly proportional to the target outcome, the same applying for the 2-days prior consumption as well. We notice that for the 2-days prior, the scale is two orders of magnitude less than for the 1-day prior and one order less than the mean daily temperature. This signifies that the model interprets a weaker relationship with this variable and the outcome, meaning that the dependence of the target outcome from this feature is less significant than the others.

Qualitative values cannot be included in the sensitivity analysis because they don't span over a range of values. Also, for the LSTM model, there can be no sensitivity analysis because only one variable is used as time series, therefore only past values of energy consumption are used for future predictions.



**Figure 11.** Partial Dependence Plots for the dataset of Athens on the mean daily temperature for the ANN (a) and the DNN (b) models, and the energy consumption values of 1-day (c) and 2-days (d) prior.

## 6. Discussion

Three methods, focused on tackling the problem of accurate forecasting of natural gas energy consumption in fifteen cities all over Greece, were investigated and applied in this study. The first method is an ANN that takes into consideration quantitative variables only, like the energy consumption and the external daily temperature. The second method is a LSTM that takes into consideration 365 previous values of only energy consumption of each city. The third method is a DNN that takes into consideration not only the quantitative variables used in the ANN, but also qualitative variables that govern human behavior such as weekdays, weekends, and bank holidays. Comparison analyses were conducted for each method in order to find the optimal architecture for each one.

All models perform adequately in most cases. The value of artificial neural networks and their derivatives is well known, however, the purpose of this study is to increase the accuracy and the time-depth of the forecasting capabilities. For the larger cities, high accuracies in forecasting energy consumption is achieved. The proposed DNN implementation achieved the highest  $R^2$  for the city of Larissa (0.9846) while the LSTM implementation for the city of Athens (0.9644) and the ANN for the city of Trikala (0.699). For the worst-case scenario, the city of Agioi Theodoroi, has consistently obtained the worst accuracies, with the DNN (0.5748) achieving significant higher accuracy, even though still not so good, compared to the LSTM (0.3848) and ANN (0.1440) implementation. The dataset of Agioi Theodoroi is the smallest compared to the rest, being one reason for achieving these low accuracies. It can be argued that the size of the city (<5000 habitants) is another important reason, since the consumption trends are sparser due its low population.

For the city of Agioi Theodoroi, the DNN increased the accuracy of its forecasts by 49.362% compared to the LSTM, and 299.195% compared to the ANN. Regarding the city of Athens, the DNN increased the accuracy of its forecasts by 0.682% compared to the LSTM, and 1.292% compared to the ANN. The proposed DNN increased the accuracy of the forecasts in almost all cases, however, its main impact was on small-scale cities such as Kilkis (+94.698%), Lamia (+259.457%), Markopoulo (+207.667%), and Xanthi (+330.273%), which have small populations, energy consumption, and less amount of gathered data. In a previous study [43] where only the ANN and LSTM approaches were applied on quantitative-only datasets, the LSTM approach offered the best results. The particular problem of forecasting energy values, is time-dependent, thus allowing the LSTM approach to excel. However, since there are other factors that affect the behavior of the consumers, and consequently the

consumption of energy, the DNN was considered as an approach that could improve the accuracy of the forecasts.

For the implementation of the LSTM, the application of dropout has improved performance for the one-year-ahead forecast. By selecting 200 units in the layer of the selected implementation, the LSTM is able to capture a measurable part of the input (365 days); however, in order to generalize well, the model should randomly drop a percentage of the weights it has “learned”. This way it has the ability to “memorize” large inputs, however, these inputs are generalized and do not overfit on the past data. This conflicts with the other implementations of ANN and DNN; however, LSTM utilizes information in a different way than the ANN and DNN. The reason why the LSTM performs better with a high dropout rate is because it tends to overfit soon during training, and even if it could reach high training accuracy, its validation (and therefore testing) accuracy would be weak. In this study, there is a trend based on seasonality, and in order to have an LSTM model that is not overly simplistic (therefore needing at least 200 units), and to train as long as possible, generalization was achieved via high dropout [61].

For the long-term predictions, the ANN and LSTM models fail to produce accurate predictions, resulting in negative  $R^2$  values. This can be derived from several facts. The more important is that since the dataset is finite, the further ahead in time the prediction is, the less training data the model is left with to “learn” from. Machine learning models are highly dependent on data, and their performance is highly correlated to the data quality and quantity. Particularly for the ANN approach, its simplistic implementation cannot capture the complexity that is required for the long-term forecasting, even if in general ANNs are powerful. Another reason is that the scale of the energy prediction units is large (in absolute numbers), thus the worse the prediction is, the larger is the penalty for it. Additionally, since the forecasting timescale increases for additionally 1, 2, and 3 years, the ill-fitted models produce large errors in predictions which are accumulated, because the forecasting time is 1, 2, and 3 times larger, respectively. The  $R^2$  metric is based on the MSE, and is scale-dependent, while MAPE is not, therefore it is useful for understanding the performance of the models. It is considered that  $R^2$  is still probably the best metric for forecasts [62], however, MAPE can still be used because the percentage of error makes sense and there are no zero values in our dataset.

In our proposed architecture, social behavior variables were added as inputs and the number of layers and nodes in our neural network was increased, in order to investigate the effect of these additions on the forecasting accuracy. These variables are strong indicators of social behavior and habits of the majority of the Greek population, which can affect the energy consumption in specific days/occasions. Overfitting was avoided by monitoring loss and accuracy throughout the training phase.

Furthermore, in order to show the effectiveness of the proposed DNN forecasting methodology, a comparative analysis was conducted with a SOGA-FCM, which was applied in one year ahead of natural gas consumption predictions concerning the same dataset of the three Greek cities (Athens, Thessaloniki, and Larissa) in [29], and a recent soft computing technique for time series forecasting using evolutionary fuzzy cognitive maps and their ensemble combination [30]. This comparison is shown in Table 15, where the MSE and MAE are used as performance metrics.

**Table 15.** Comparison of results between machine learning and soft computing methods for three benchmark cities.

Methods	Cities					
	Athens		Thessaloniki		Larissa	
	MSE (MJ <sup>2</sup> )	MAE (MJ <sup>2</sup> )	MSE (MJ <sup>2</sup> )	MAE (MJ <sup>2</sup> )	MSE (MJ <sup>2</sup> )	MAE (MJ <sup>2</sup> )
ANN	$7.70 \times 10^{-3}$	$7.01 \times 10^{-2}$	$3.40 \times 10^{-3}$	$4.52 \times 10^{-2}$	$2.60 \times 10^{-3}$	$3.43 \times 10^{-2}$
LSTM	$2.10 \times 10^{-3}$	$2.92 \times 10^{-2}$	$3.90 \times 10^{-3}$	$4.79 \times 10^{-2}$	$3.30 \times 10^{-3}$	$4.43 \times 10^{-2}$
Hybrid FCM	$3.20 \times 10^{-3}$	$3.28 \times 10^{-2}$	$3.30 \times 10^{-3}$	$3.81 \times 10^{-2}$	$4.10 \times 10^{-3}$	$4.17 \times 10^{-2}$
Ensemble (EB)	$3.10 \times 10^{-3}$	$3.28 \times 10^{-2}$	$3.10 \times 10^{-3}$	$3.69 \times 10^{-2}$	$4.00 \times 10^{-3}$	$4.17 \times 10^{-2}$
<b>Proposed DNN</b>	<b><math>1.70 \times 10^{-3}</math></b>	<b><math>2.75 \times 10^{-2}</math></b>	<b><math>1.60 \times 10^{-3}</math></b>	<b><math>2.72 \times 10^{-2}</math></b>	<b><math>1.10 \times 10^{-3}</math></b>	<b><math>2.40 \times 10^{-2}</math></b>

It is evident that all methods achieve high accuracy in the predictions of the energy consumption patterns in their relative timescales. The ensemble and hybrid methods achieve the same accuracy as the ANN, with the LSTM performing slightly better. The proposed DNN, by utilizing inputs of social variables into its learning patterns and having a deeper architecture, outperforms all other methods with significant difference. The significance of this outcome lies on the fact that qualitative variables that dictate human behavior can be learned by computational algorithms and be utilized to improve forecasting accuracy furthermore.

The case of Greece has some sensitive aspects, since multiple dynamics in natural gas consumption were introduced due to the financial crisis of the previous years. This instability created an additional obstacle to the accurate forecasting of energy demand, thus increasing the need for efficient forecasting models that can accurately offer in-depth insight on the demand trends of each city and adapt to their different conditions. Since the proposed method offers high accuracy and long forecasting capabilities, it can be used by any utilities and distribution operators, as a solution to upgrade operational long-term planning, as well as to provide insight on policy making from the side of the state.

## 7. Conclusions

Summing up, three different forecasting approaches have been implemented in order to develop models for predicting energy demand of natural gas. Investigative analysis took place for an ANN, a LSTM, and the proposed DNN implementations in order to find a desired architecture for each method. Fifteen cities all around Greece were tested, each one with a dataset of measurements that spanned from 3 to 7 years. The investigated cities differ both in size as well as in geographical location, amplifying as much as possible the variability of each use case examined. Despite the fact that this study is focused on cities that are only in Greece, the proposed methodology is highly generalizable for any other city that can provide sufficient amount of data, both measurable and behavioral.

The goal of this study was to propose an efficient neural network implementation that utilizes a variety of quantitative and qualitative inputs, as well as a deep architecture with many layers and nodes, to demonstrate how social factors can improve the performance of the model and increase the accuracy of its forecasts. The proposed methodology has outperformed both the simple ANN approach as well as the state-of-the-art LSTM approach even though both still offer good accuracy in most cases. The inclusion of social factors in the proposed DNN approach offered consistently more generalized, high-accuracy results. This derives from the fact that by exploring longer forecasts, the four-year ahead forecast was achieved only with the proposed DNN implementation, while the LSTM could only provide accurate results up to two years ahead, and the ANN was deviating systematically.

Applying a combination of multi-parametric social factors, by also taking advantage of the memory cells structure of the LSTM implementation will be the base of the future work that will aim to outperform the DNN implementation. Additional Fuzzy Cognitive Maps structures will be also considered for increasing the interpretability of the models and how the inputs affect the performance.

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

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

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

### Appendix A Appendix

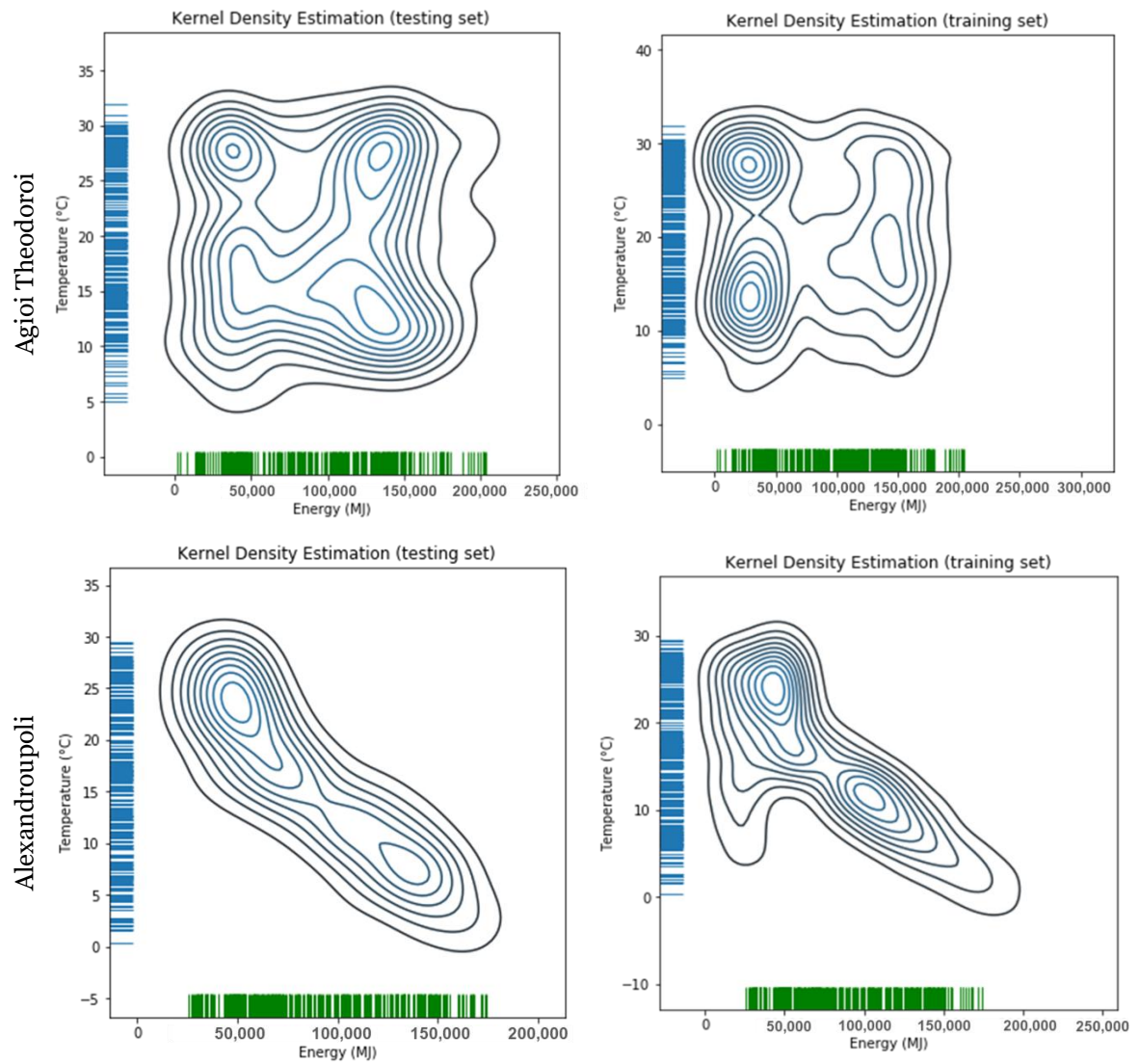


Figure A1. Cont.

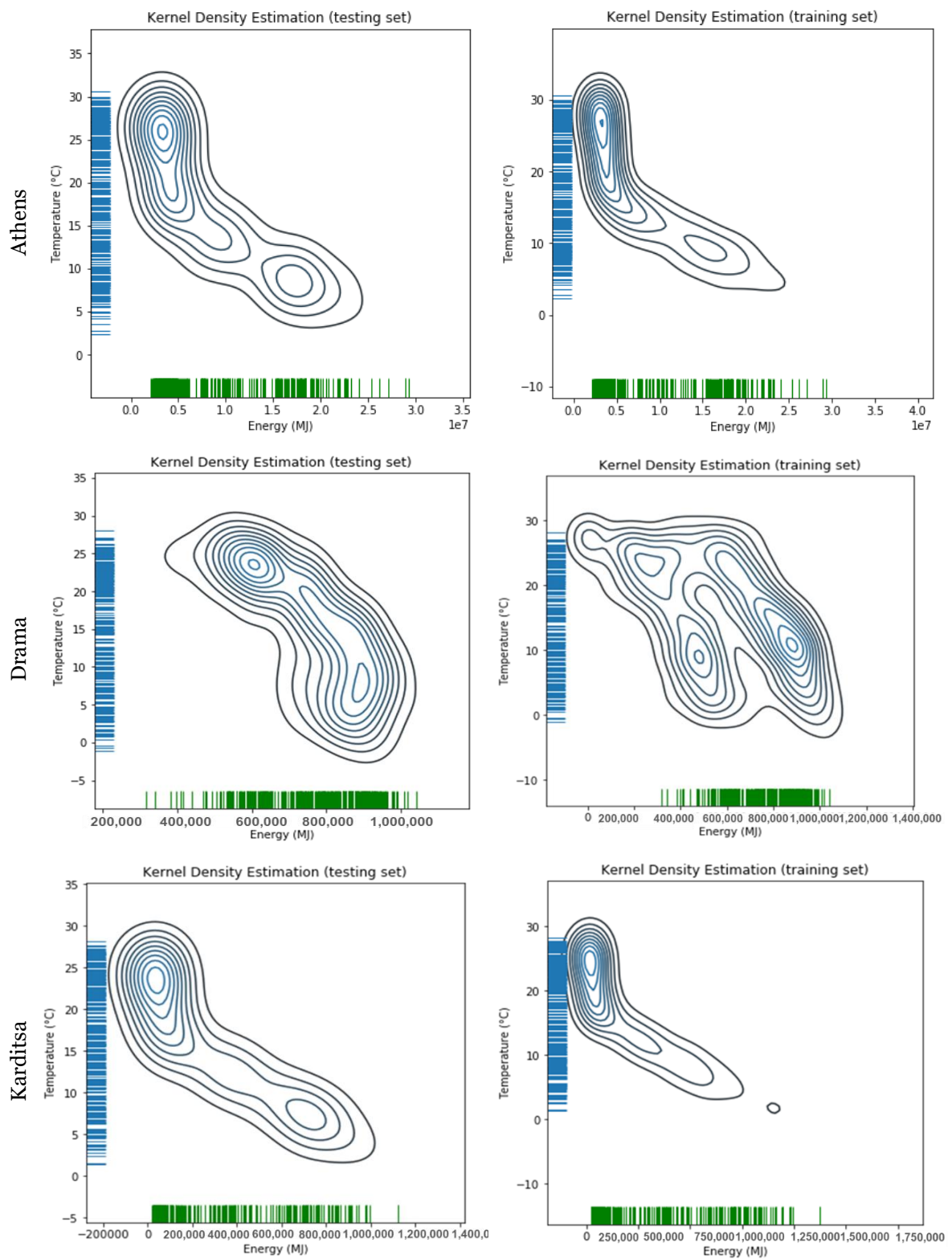


Figure A1. Cont.

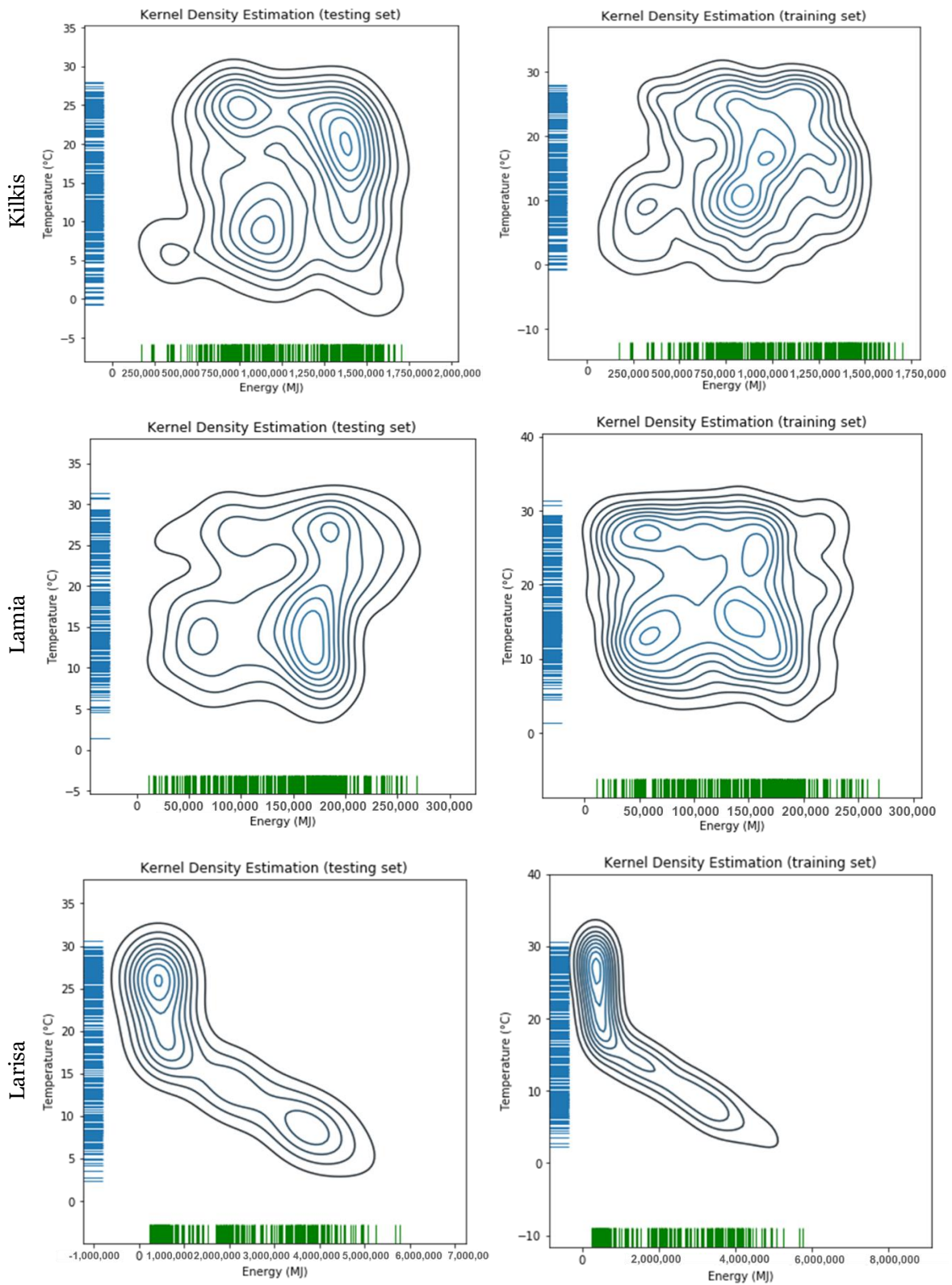


Figure A1. Cont.

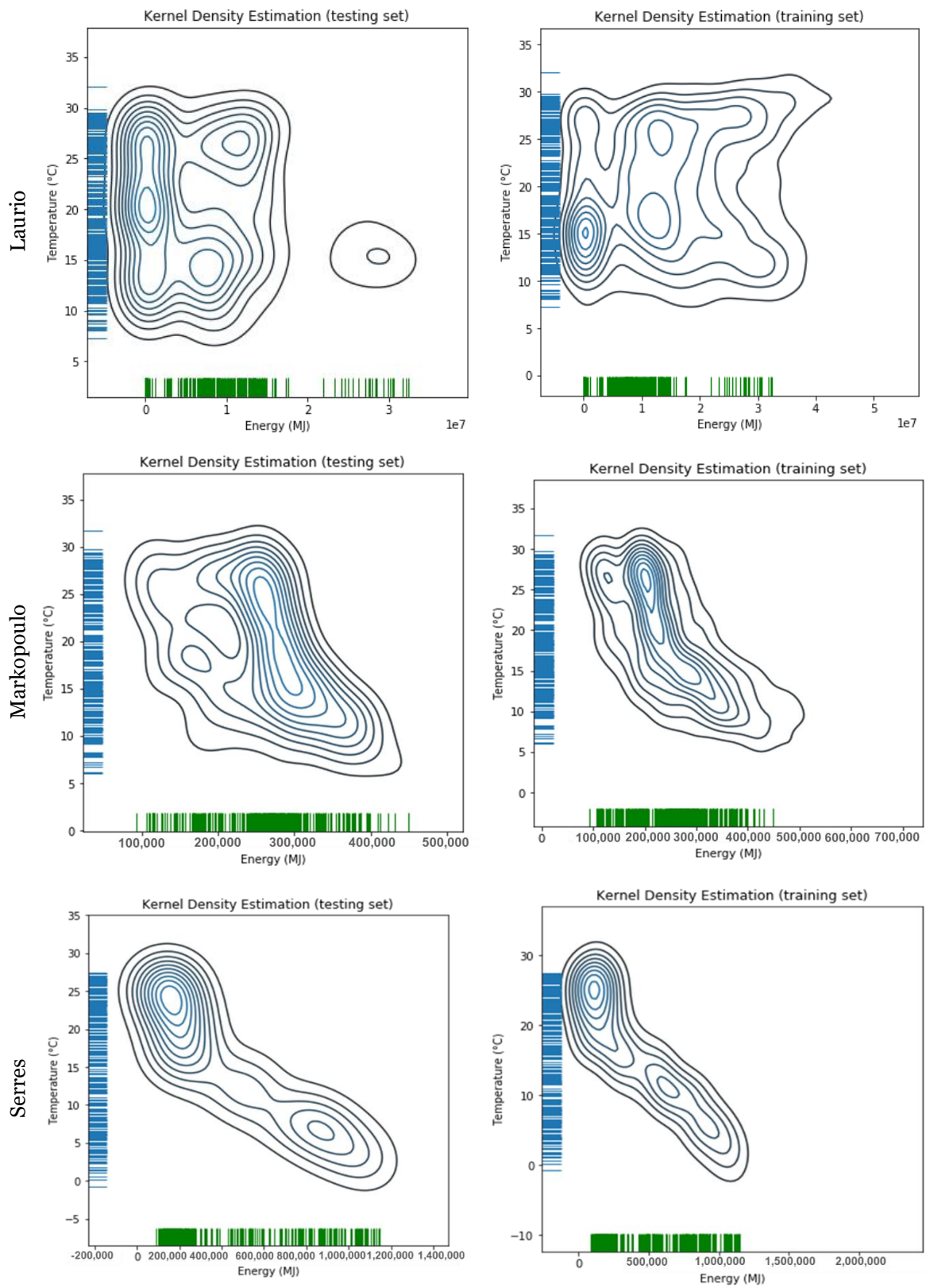


Figure A1. Cont.



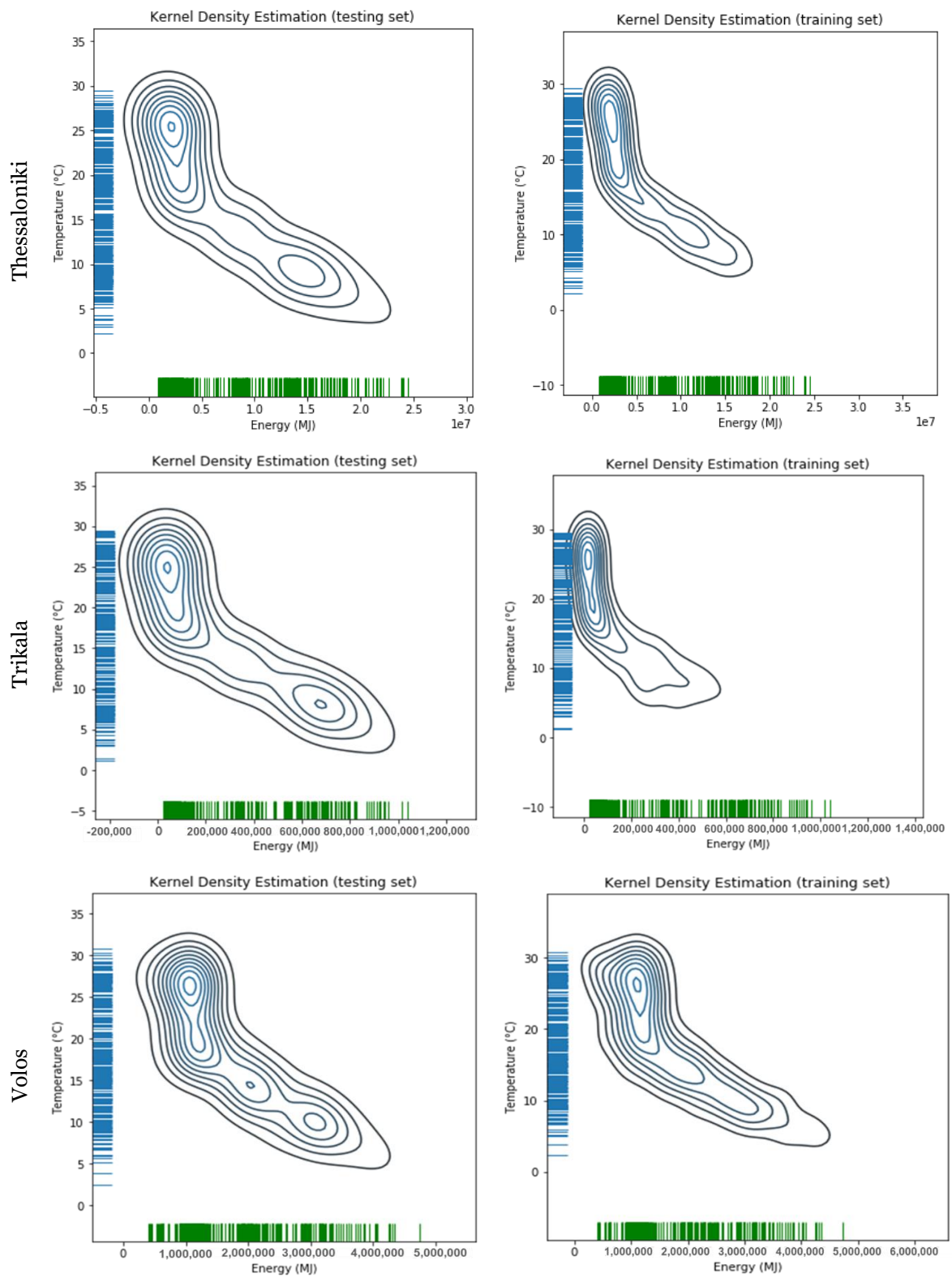
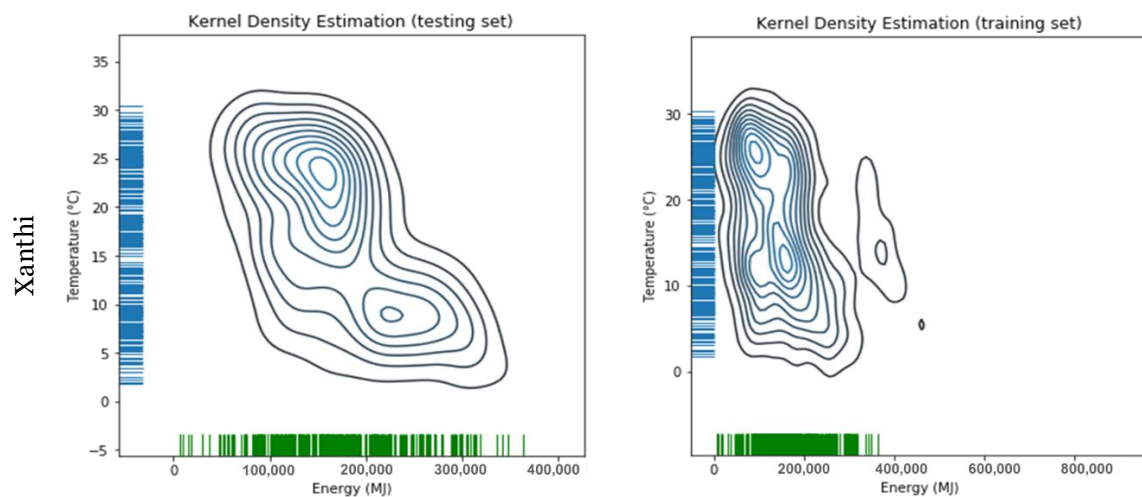


Figure A1. Cont.



**Figure A1.** Correlation of the energy demand and the mean temperature for the training set and the test set for all the examined cities.

## References

1. Tamba, J.G.; Essiane, S.N.; Sapnken, E.F.; Koffi, F.D.; Nsouandélé, J.L.; Soldo, B.; Njomo, D. Forecasting natural gas: A literature survey. *Int. J. Energy Econ. Policy* **2018**, *8*, 216–249.
2. Ivezić, D. Short-Term Natural Gas Consumption Forecast. *FME Trans.* **2006**, *34*, 165–169.
3. Ghalekhondabi, I.; Ardjmand, E.; Weckman, G.R.; Young, W.A. An overview of energy demand forecasting methods published in 2005–2015. *Energy Syst.* **2017**, *8*, 411–447. [[CrossRef](#)]
4. Motlagh, O.; Grozev, G.; Papageorgiou, E.I. A Neural Approach to Electricity Demand Forecasting. In *Artificial Neural Network Modelling*; Shanmuganathan, S., Samarasinghe, S., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 281–306.
5. Raza, M.Q.; Khosravi, A. A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings. *Renew. Sustain. Energy Rev.* **2015**, *50*, 1352–1372. [[CrossRef](#)]
6. Gorucu, F.B.; Geriş, P.U.; Gumrah, F. Artificial Neural Network Modeling for Forecasting Gas Consumption. *Energy Sources* **2004**, *26*, 299–307. [[CrossRef](#)]
7. Papageorgiou, E.I.; Poczęta, K. A two-stage model for time series prediction based on fuzzy cognitive maps and neural networks. *Neurocomputing* **2017**, *232*, 113–121. [[CrossRef](#)]
8. Karimi, H.; Dastranj, J. Artificial neural network-based genetic algorithm to predict natural gas consumption. *Energy Syst.* **2014**, *5*, 571–581. [[CrossRef](#)]
9. Khotanzad, A.; Elragal, H. Natural gas load forecasting with combination of adaptive neural networks. In Proceedings of the International Joint Conference on Neural Networks, Washington, DC, USA, 10–16 July 1999; IEEE: Piscataway Township, NJ, USA, 2002.
10. Khotanzad, A.; Elragal, H.; Lu, T.L. Combination of artificial neural-network forecasters for prediction of natural gas consumption. *IEEE Trans. Neural Netw.* **2000**, *11*, 464–473. [[CrossRef](#)]
11. Kizilaslan, R.; Karlik, B. Comparison neural networks models for short term forecasting of natural gas consumption in Istanbul. In Proceedings of the 1st International Conference on the Applications of Digital Information and Web Technologies, ICADIWT 2008, Ostrava, Czech Republic, 4–6 August 2008.
12. Kizilaslan, R.; Karlik, B. Combination of neural networks forecasters for monthly natural gas consumption prediction. *Neural Netw. World* **2009**, *19*, 191–199.
13. Musilek, P.; Pelikan, E.; Brabec, T.; Simunek, M. Recurrent Neural Network Based Gating for Natural Gas Load Prediction System. In Proceedings of the 2006 IEEE International Joint Conference on Neural Network Proceedings, Vancouver, BC, Canada, 16–21 July 2006.
14. Soldo, B. Forecasting natural gas consumption. *Appl. Energy* **2012**, *92*, 26–37. [[CrossRef](#)]
15. Szoplik, J. Forecasting of natural gas consumption with artificial neural networks. *Energy* **2015**, *85*, 208–220. [[CrossRef](#)]

16. Merkel, G.D.; Povinelli, R.J.; Brown, R.H. Deep neural network regression as a component of a forecast ensemble. In Proceedings of the International Symposium on Forecasting, Cairns, Australia, 25–28 June 2017.
17. Merkel, G.D. Deep Neural Networks as Time Series Forecasters of Energy Demand. Master's Thesis, Marquette University, Milwaukee, WI, US, 2017. Available online: [https://epublications.marquette.edu/theses\\_open/434/](https://epublications.marquette.edu/theses_open/434/) (accessed on 5 May 2020).
18. Merkel, G.D.; Povinelli, R.J.; Brown, R.H. Short-term load forecasting of natural gas with deep neural network regression. *Energies* **2018**, *11*, 2008. [[CrossRef](#)]
19. Azadeh, A.; Asadzadeh, S.M.; Ghanbari, A. An adaptive network-based fuzzy inference system for short-term natural gas demand estimation: Uncertain and complex environments. *Energy Policy* **2010**, *38*, 1529–1536. [[CrossRef](#)]
20. Behrouznia, A.; Saberi, M.; Azadeh, A.; Asadzadeh, S.M.; Pazhooheshfar, P. An adaptive network based fuzzy inference system-fuzzy data envelopment analysis for gas consumption forecasting and analysis: The case of South America. In Proceedings of the 2010 International Conference on Intelligent and Advanced Systems, ICIAS 2010, Kuala Lumpur, Malaysia, 15–17 June 2010.
21. Yu, F.; Xu, X. A short-term load forecasting model of natural gas based on optimized genetic algorithm and improved BP neural network. *Appl. Energy* **2014**, *134*, 102–113. [[CrossRef](#)]
22. Panapakidis, I.P.; Dagoumas, A.S. Day-ahead natural gas demand forecasting based on the combination of wavelet transform and ANFIS/genetic algorithm/neural network model. *Energy* **2017**, *118*, 231–245. [[CrossRef](#)]
23. Homenda, W.; Jastrzebska, A.; Pedrycz, W. Modeling time series with fuzzy cognitive maps. In Proceedings of the IEEE International Conference on Fuzzy Systems, Beijing, China, 6–11 July 2014.
24. Homenda, W.; Jastrzebska, A.; Pedrycz, W.; Pedrycz, W. Nodes selection criteria for fuzzy cognitive maps designed to model time series. *Adv. Intell. Syst. Comput.* **2015**, *323*, 859–870.
25. Salmeron, J.L.; Froelich, W. Dynamic optimization of fuzzy cognitive maps for time series forecasting. *Knowl. Based Syst.* **2016**, *105*, 29–37. [[CrossRef](#)]
26. Froelich, W.; Salmeron, J.L. Evolutionary learning of fuzzy grey cognitive maps for the forecasting of multivariate, interval-valued time series. *Int. J. Approx. Reason.* **2014**, *55*, 1319–1335. [[CrossRef](#)]
27. Papageorgiou, E.I.; Poczeta, K.; Lapidou, C. Application of Fuzzy Cognitive Maps to water demand prediction. In Proceedings of the IEEE International Conference on Fuzzy Systems, Istanbul, Turkey, 2–5 August 2015.
28. Poczeta, K.; Yastrebov, A.; Papageorgiou, E.I. Learning fuzzy cognitive maps using structure optimization genetic algorithm. In Proceedings of the 2015 Federated Conference on Computer Science and Information Systems, FedCSIS 2015, Lodz, Poland, 13–16 September 2015.
29. Poczeta, K.; Papageorgiou, E.I. Implementing fuzzy cognitive maps with neural networks for natural gas prediction. In Proceedings of the Proceedings—International Conference on Tools with Artificial Intelligence, ICTAI, Volos, Greece, 5–7 November 2018.
30. Papageorgiou, K.I.; Poczeta, K.; Papageorgiou, E.; Gerogiannis, V.C.; Stamoulis, G. Exploring an ensemble of methods that combines fuzzy cognitive maps and neural networks in solving the time series prediction problem of gas consumption in Greece. *Algorithms* **2019**, *12*, 235. [[CrossRef](#)]
31. Bhattacharyya, S.C.; Timilsina, G.R. Energy Demand Models for Policy Formulation: A Comparative Study of Energy Demand Models. 2009. Available online: <https://core.ac.uk/reader/6521667> (accessed on 24 November 2019).
32. Han, Y.; Sha, X.; Grover-Silva, E.; Michiardi, P. On the impact of socio-economic factors on power load forecasting. In Proceedings of the 2014 IEEE International Conference on Big Data, IEEE Big Data 2014, Washington, DC, USA, 27–30 October 2014; IEEE: Piscataway Township, NJ, USA, 2015.
33. McLoughlin, F.; Duffy, A.; Conlon, M. Characterising domestic electricity consumption patterns by dwelling and occupant socio-economic variables: An Irish case study. *Energy Build.* **2012**, *48*, 240–248. [[CrossRef](#)]
34. Beckel, C.; Sadamori, L.; Santini, S. Automatic socio-economic classification of households using electricity consumption data. In Proceedings of the e-Energy 2013—Proceedings of the 4th ACM International Conference on Future Energy Systems, Berkeley, CA, USA, 22–24 May 2013.
35. Sánchez-Oro, J.; Duarte, A.; Salcedo-Sanz, S. Robust total energy demand estimation with a hybrid Variable Neighborhood Search—Extreme Learning Machine algorithm. *Energy Convers. Manag.* **2016**, *123*, 445–452. [[CrossRef](#)]

36. Liu, B.; Fu, C.; Bielefield, A.; Liu, Y.Q. Forecasting of Chinese Primary Energy Consumption in 2021 with GRU artificial neural network. *Energies* **2017**, *10*, 1453. [CrossRef]
37. Oğcu, G.; Demirel, O.F.; Zaim, S. Forecasting Electricity Consumption with Neural Networks and Support Vector Regression. *Procedia Soc. Behav. Sci.* **2012**, *58*, 1576–1585. [CrossRef]
38. Günay, M.E. Forecasting annual gross electricity demand by artificial neural networks using predicted values of socio-economic indicators and climatic conditions: Case of Turkey. *Energy Policy* **2016**, *90*, 92–101. [CrossRef]
39. Parfenenko, Y.; Shendryk, V.; Vashchenko, S.; Fedotova, N. The forecasting of the daily heat demand of the public sector buildings with district heating. In *Information and Software Technologies*; Springer: Cham, Switzerland, 2015.
40. Benalcazar, P.; Kamiński, J. Short-term heat load forecasting in district heating systems using artificial neural networks. In Proceedings of the IOP Conference Series: Earth and Environmental Science, Krakow, Poland, 14–17 November 2017; IOP: London, UK, 2019.
41. Dolinay, V.; Vasek, L.; Novak, J.; Chalupa, P.; Kral, E. Heat demand model for district heating simulation. In Proceedings of the MATEC Web of Conferences, Majorca, Spain, 14–17 July 2018.
42. Powell, K.M.; Sriprasad, A.; Cole, W.J.; Edgar, T.F. Heating, cooling, and electrical load forecasting for a large-scale district energy system. *Energy* **2014**, *74*, 877–885. [CrossRef]
43. Anagnostis, A.; Papageorgiou, E.; Dafopoulos, V.; Bochtis, D. Applying Long Short-Term Memory Networks for natural gas demand prediction. In Proceedings of the 2019 10th International Conference on Information, Intelligence, Systems and Applications, Patras, Greece, 15–17 July 2019; pp. 1–7.
44. McCulloch, W.S.; Pitts, W. A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.* **1943**, *5*, 115–133. [CrossRef]
45. Rosenblatt, F. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychol. Rev.* **1958**, *65*, 386–408. [CrossRef]
46. Pal, S.K.; Mitra, S. Multilayer Perceptron, Fuzzy Sets, and Classification. *IEEE Trans. Neural Netw.* **1992**, *3*, 683–697. [CrossRef]
47. Linnainmaa, S. Taylor expansion of the accumulated rounding error. *BIT* **1976**, *16*, 146–160. [CrossRef]
48. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]
49. Lecun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef] [PubMed]
50. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016.
51. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, 21–26 July 2017.
52. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015.
53. Kenton, L.; Kristina, T.; Devlin, J.; Chang, M.-W. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv* **2017**, arXiv:1810.04805.
54. Chung, J.; Gulcehre, C.; Cho, K.; Bengio, Y. Gated feedback recurrent neural networks. In Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6–11 July 2015.
55. Anagnostis, A.; Asiminari, G.; Papageorgiou, E.; Bochtis, D. A Convolutional Neural Networks Based Method for Anthracnose Infected Walnut Tree Leaves Identification. *Appl. Sci.* **2020**, *10*, 469. [CrossRef]
56. Chollet, F.; Allaire, J.J. Keras 2015. Available online: [https://keras.io/getting\\_started/faq/#how-should-i-cite-keras](https://keras.io/getting_started/faq/#how-should-i-cite-keras) (accessed on 27 March 2015).
57. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
58. Hazra, A. Using the confidence interval confidently. *J. Thorac. Dis.* **2017**, *9*, 4125–4130. [CrossRef]
59. Ruppert, D. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. *J. Am. Stat. Assoc.* **2004**, *99*, 567. [CrossRef]
60. Molnar, C. *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*; Lulu: Morrisville, NC, USA, 2019.

61. Graham, B.; Reizenstein, J.; Robinson, L. Efficient batchwise dropout training using submatrices. *arXiv* **2015**, arXiv:1502.02478.
62. Cameron, A.C.; Windmeijer, F.A.G. An R-squared measure of goodness of fit for some common nonlinear regression models. *J. Econom.* **1997**, *77*, 329–342. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).