energies

Article

Comparison of Hospital Building's Energy Consumption Prediction Using Artificial Neural Networks, ANFIS, and LSTM Network

Dimitrios K. Panagiotou and Anastasios I. Dounis *

Department of Biomedical Engineering, University of West Attica, 12243 Athens, Greece
* Correspondence: aidounis@uniwa.gr

Abstract: Since accurate load forecasting plays an important role in the improvisation of buildings and as described in EU's "Green Deal", financial resources saved through improvisation of the efficiency of buildings with social importance such as hospitals, will be the funds to support their mission, the social impact of load forecasting is significant. In the present paper, eight different machine learning predictors will be examined for the short-term load forecasting of a hospital's facility building. The challenge is to qualify the most suitable predictors for the abovementioned task, which is beneficial for an in-depth study on accurate predictors' applications in Intelligent Energy Management Systems (IEMS). Three Artificial Neural Networks using a backpropagation algorithm, three Artificial Neural Networks using metaheuristic optimization algorithms for training, an Adaptive Neuro-Fuzzy Inference System (ANFIS), and a Long-Short Term Memory (LSTM) network were tested using timeseries generated from a simulated healthcare facility. ANFIS and backpropagation-based trained models outperformed all other models since they both deal well with complex nonlinear problems. LSTM also performed adequately. The models trained with metaheuristic algorithms demonstrated poor performance.

Keywords: artificial neural networks; adaptive neuro-fuzzy adaptive inference system; long short-term memory networks; backpropagation algorithms; metaheuristic algorithms; machine learning; load forecasting

check for updates

Citation: Panagiotou, D.K.; Dounis, A.I. Comparison of Hospital Building's Energy Consumption Prediction Using Artificial Neural Networks, ANFIS, and LSTM Network. *Energies* **2022**, *15*, 6453. https://doi.org/10.3390/en15176453

Academic Editors: Jaroslaw Krzywanski and F. Pacheco Torgal

Received: 26 July 2022
Accepted: 28 August 2022
Published: 3 September 2022

1. Introduction

1.1. General Context and Importance of the Present Study

Buildings play an essential part in human lives, and people spend large parts of their daily life in a building environment. The building sector is associated with nearly 40% of the EU's total energy consumption, in addition to 36% of greenhouse gas emissions in the EU [1].

Therefore, the EU has established a set of targets to be achieved concerning the improvement of energy efficiency in buildings, such as those described in the EU 31/2010 and 27/2012 directives with their revision in 2018. Among other measures, there are certain actions to be taken to reach near zero energy and positive energy buildings goal. Through the Horizon 2020 program, the EU invested in technologies capable of improving the energy efficiency of buildings.

Due to their social relevance, size, and operational characteristics, health facilities can be considered a special case [2,3]. They have the highest energy consumption per unit of floor area, as they contribute 10% of the total energy use and they are responsible for 5% of the CO_2 emissions [4]. In England, the carbon footprint of the *National Health System* (*NHS*) and relevant services represented 40% of the greenhouse gas emissions among the public sector buildings. Such a negative influence on public health is quite contradictory to the health sector [5]. High energy consumptions of the hospitals are due to extremely high demands of heating and cooling, ventilation, the substantial demand for hot water and electricity, and the fast advances in the diagnostic techniques that require continuous remodeling of the existing spaces in operative healthcare facilities, involving both architectural-distributive aspects and building plant systems [6].

Electricity load forecasting is an important aspect of the improvisation of buildings for several reasons; it is well known that accurate prediction (Albeit the terms "Forecast" and "Prediction" have conceptual differences, in the present study are used as interchangeable terms) of energy consumption has a positive impact on the operational costs [7]. Additionally, when referring to healthcare buildings, load forecasting is essential to an *Intelligent Energy Management System* (*IEMS*), which allows all the related subsystems to operate effectively and harmoniously while at the same time ensuring the operation of the hospital under extreme conditions while at the same time reduces costs [3], using the saving funds to improve the hospital's services, as described in EU's "Green Deal" [8]: "*Focus should also be put on renovating schools and hospitals, as the money saved through building efficiency will be money available to support education and public health*".

In the present paper, different machine learning predictors will be examined for the short-term load forecasting of a healthcare building in San Diego, California. *Short-Term Load Forecast* (*STLF*) usually predicts the electricity consumption from some minutes up to a week ahead and plays a vital role in the proper management of buildings and facilities. This is due to the fact that accurate forecasts can be beneficial in carrying out many decisions, such as peak load mitigation, optimal energy scheduling, and demand side management [9]. Moreover, the unbalance control of the network during the interchange between the site and the main grid is of great importance in microgrid configurations [10,11] and with an economic interest since energy management on buildings can be bounded with different pricing options [12], which is largely expected in power systems in the future [13]. Load forecasting is a rather challenging task since consumption depends on many stochastic and nonlinear factors that affect the system's behavior, such as weather conditions, social and economic changes, and many other parameters.

1.2. Research Papers: Contribution and Limitation to Data-Driven Methods

There are three main approaches developed in the last two decades: physical, data-driven, and hybrid methods [14].

The physical approach [15] employs thermodynamic rules to calculate the heat and mass balance inside a building. It is a "*white-box*" method since the governing physical laws of the system are known. A major limitation of this approach is that all building characteristics must be well known and that their physical mechanisms can be described with high accuracy [16].

Data-driven approaches provide the ability to learn from data [17], using machine learning algorithms to evaluate and estimate/predict the building's energy consumption [18]. Since the physical characteristics are ignored, these methods are called "*black-box*" methods. Black-box methods' major limitation is that they need large amounts of data. Moreover, the results obtained by these approaches are difficult to interpret in physical terms [16].

Hybrid approaches use combinations of the two abovementioned methods to overcome their limitations and exploit the advantages they have. These hybrid methods are called "*grey-box*" methods [19].

Data-driven models include *linear regression*, *Artificial Neural Networks* (*ANNs*), *Support Vector Machines* (*SVMs*), *Long-Short Term Memory* (*LSTM*) algorithm, and ensemble techniques [20]. ANNs are the most employed methods because of their straightforward implementation and their high performance [13,21,22]. In [1], the authors tested various ANN models for a day- and an hour-ahead prediction, selecting as inputs features climate data (dew and dry-bulb temperature, humidity), previous power demand data, and the calendar data (week pointer, day pointer, hour pointer, and quarter-hour pointer) and two training algorithms (*Levenberg-Marquardt* and *Bayesian Regularization*). For one-hour-ahead prediction, *MAPE* is in the range of 1.78–2.24%. The authors of [23] studied an LSTM-based model for one-hour resolution, comparing it with other models such as *Multi-Layer Perceptron* (*MLP*) neural networks, *Random Forest* (*RF*), and SVMs, showing that the LSTM model is the best, achieving a *MAPE* of 5.28% over a week's hourly predictions. The authors of [24] developed an *Adaptive Neuro-Fuzzy Inference System* (*ANFIS*) model using weather and load data as inputs and achieving an *R*-value of 0.98059 for the testing data for an hour ahead of energy forecasting. Classical algorithms such as backpropagation and Levenberg–Marquardt (LM) have various limitations: They easily get trapped in local optima, and they also are highly dependent on initialization, resulting in a lack of predicting better outcomes for new input patterns. Gradient-descent approaches also exhibit slow convergence because, for stable learning, small learning rates are desirable. Since stochastic approaches such as metaheuristic algorithms rely heavily on randomness for the results, they can adapt better to new input patterns. They can also avoid local optima. Of course, some drawbacks may arise, such as their slow convergence compared to the classical algorithms [25]. Hybrid models that combine evolutionary algorithms and ANNs were proposed by [26,27], who investigated methods that involve *Particle Swarm Optimization* (*PSO*) and *Genetic Algorithms* (*GA*) for the training of ANNs.

Although several methods of buildings' load forecasting have been reported in the literature over the last decade [15], only a small number of them refer to hospitals and healthcare facilities. The authors of [28] analyzed different machine learning techniques to predict the daily consumption of a hospital in Spain, achieving an average daily error of 3.8%. The authors of [2] analyzed the dynamics of short- and middle-term predictions by implementing unsupervised, supervised, and ensemble methods, and the proposed models are evaluated by means of a year's dataset with satisfactory results for the one of the two examined healthcare buildings. The building's results "*revealed an abrupt change in the average prediction error, which helped the authors to identify changes in the energy consumption patterns implemented by the managers of the building under study*" [3], using a multi-layer perceptron ANN based on a backpropagation algorithm for forecasting day ahead load consumption of a large hospital, with satisfactory results (*MAPE* close to 7%), using as inputs the previous day consumption, the previous week consumption, the mean consumption of the previous day, the weekday, a workday/holiday indicator, the timestamp, and the temperature [4], compared the precision of eight machine learning models trained with daily and weekly datasets from a general hospital in Shanghai. The results indicated that for daily load prediction, RF, XGBoost, and SVR are the most accurate and single learning models, with a *MAPE* on the test dataset of 9.64%, 9.81%, and 10.67%, respectively.

The present study focuses on electricity consumption short-term (one-hour-ahead) forecasting of a healthcare building with data-driven models such as Artificial Neural Networks (ANNs), Adaptive Neuro-Fuzzy Inference Systems (ANFIS), and Long Short-Term Memory (LSTM) networks. Since the existing literature is relatively small, and the models available are numerous, the primary target of this research paper is to propose a simple and accurate model for this critical infrastructure, contributing to ongoing global research.

2. Models and Methods

2.1. Forecasting Models

This study exhaustively tested 6 ANNs, an ANFIS, and an LSTM network, using timeseries generated from a simulated healthcare facility.

2.1.1. Artificial Neural Networks

Artificial Neural Networks can be considered as models of reasoning that imitate the function of the human brain. Like the brain is consisted of many biological neurons, an *Artificial Neural Network* (*ANN*) is consisted of several simple, highly interconnected processors, which are also called neurons. The neurons are connected by links, and every single link has an associated weight, which expresses the strength of each neuron input. The learning process of a neural network is carried through repeated adjustments of its weights until a goal is met; on most occasions, the goal is to achieve an *MSE* (*Mean Square Error*) lower than a predefined value within a certain number of iterations-called *epochs*.

All the neural networks that are investigated in the current paper have 24 inputs representing the consumption of the past 24 h (x_1 to x_{24}), one hidden layer with hyperbolic tangent transfer functions (f), and one linear transfer function output layer (y), which gives the hour-ahead prediction. W_{in} and W_{out} are the weights matrices of the input and output accordingly. According to the "rule of thumb" [29], the hidden layer has 6 neurons:

$$n_h = \sqrt{n_o + n_i} + l \tag{1}$$

where n_h represents the number of the hidden layer neurons, n_i is the number of inputs, and n_o is the number of outputs; l represents an integer between 1 and 10. Here, l equals one. The neural network's structure [30] is depicted in Figure 1. $Load_{t-1}$–$Load_{t-24}$ represent the inputs (the load of every hour of the past 24 h), and $Load_t$ represents the prediction for the next hour's consumption.

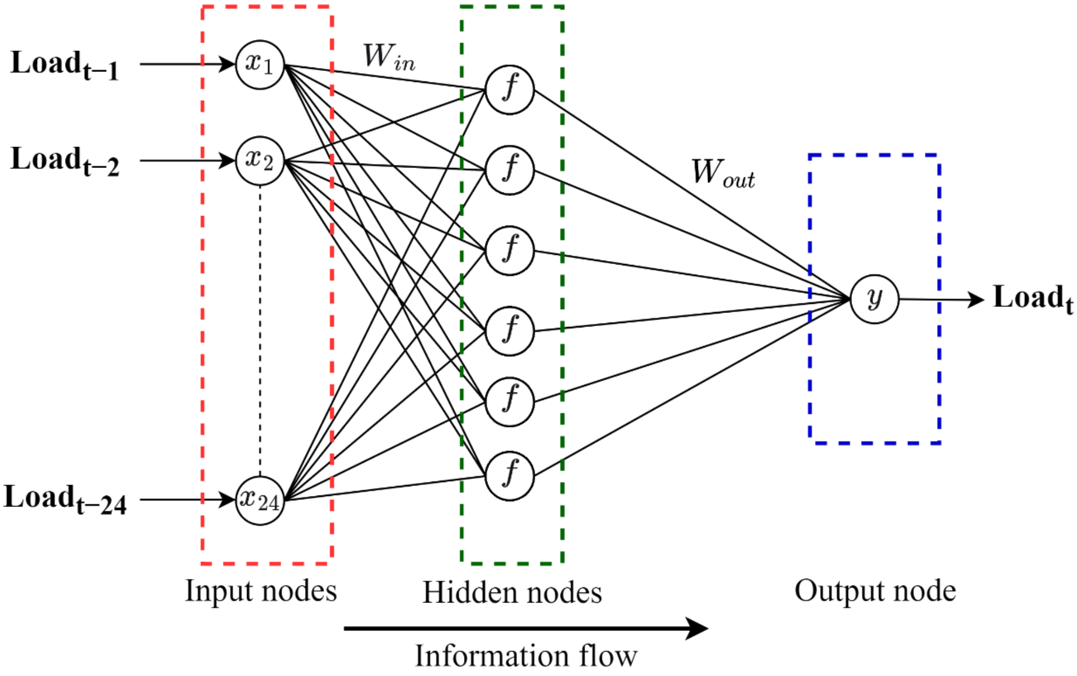

Figure 1. Structure of the feedforward ANN used in the study.

Among the studied neural networks, three use a backpropagation algorithm, such as *Levenberg Marquardt* (*LM*), *Scaled Conjugate Gradient* (*SCG*) backpropagation, and gradient descent with momentum and adaptive learning rate (*GDX*) backpropagation:

1. *Levenberg–Marquardt* algorithm is an iterative method that finds the minimum of a multivariate function expressed in terms of sum of squares of nonlinear real-valued functions and has become a standard technique for nonlinear least squares problems. It can be described as a combination of the steepest descent and the Gauss–Newton method [31]. LM algorithm is claimed to be the fastest method for training moderate-sized feedforward neural networks [32]. For the present study, the LM-ANN model has the following parameters:
    - Minimum performance gradient: 10^{-7};
    - Initial μ: 0.001;
    - μ decrease factor: 0.1;
    - μ increase factor: 10;
    - Maximum μ: 110;
    - Training goal: 0.01;

- Performance metric: *MSE.*

2. The *Scaled Conjugate Gradient* algorithm, introduced by Møller [33], is a combination of the LM algorithm and the *Conjugate Gradient* (*CG*) approach. *SCG* is considerably faster than standard backpropagation and other conjugate gradient methods. For the present study, the *SCG-ANN* model has the following parameters:
    - Minimum performance gradient: 10^{-6};
    - μ adjustment parameter: 0.005;
    - sigma (σ): 5×10^{-5};
    - lamda (λ): 5×10^{-7};
    - Training goal: 0.01;
    - Performance metric: *MSE*.
3. *Adaptive Learning Rate with Momentum*. In the standard steepest descent algorithm, the *Learning Rate* (*LR*) is fixed during training. For the present study, the *GDX-ANN* model has the following parameters:
    - Learning rate: 0.01;
    - LR increase rate: 1.05;
    - LR decrease rate: 0.7;
    - Momentum constant: 0.9;
    - Training goal: 0.01;
    - Performance metric: *MSE*.

Three neural networks have been trained using metaheuristic algorithms: *Genetic Algorithm* (*GA*), *Particle Swarm Optimization* (*PSO*), and *Big Bang-Big Crunch* (*BB-BC*):

1. *Genetic Algorithms* consist of a category of stochastic search algorithms inspired by Darwin's theory of natural evolution. The main idea is that a problem can be solved using a set of candidate solutions represented as binary strings by applying the following method:
    i. Creation of random initial population (chromosomes).
    ii. Creation of a series of new populations. At every iteration, the individuals in the current generation are the parents of the next population, called the offspring. The mating of the individuals is performed by the algorithm and includes the following steps:
        a. Every member of the existing population is evaluated as a solution for its fitness.
        b. The parents are selected based on their fitness.
        c. Some of the individuals who achieve the best fitness are passed on to the next generation as elite.
        d. The parents produce children either by mutation of a single parent or by mating a pair of parents, which is called crossover.
        e. The children and the elite replace the current population with the next generation.
    iii When one of the stopping criteria is met, the algorithm stops.

The concept behind the application of GAs in artificial neural network training is that a general-purpose optimization tool should be applicable to train a neural network for which an evaluation function can be derived [34]. This approach offers an alternative to backpropagation because it outperforms the latter in the following issues [35]: the scaling problem, where the performance of backpropagation appears to be poor when the problem complexity increases; and the inability to handle discontinuous optimality criteria or node transfer functions. GAs overcome these problems. For the present study, the GA-ANN model has the following parameters:

- Maximum generations: 1000;
- Population size: 100;
- Training goal: 0.01;

- Performance metric: *MSE*.

2. *Particle Swarm Optimization* is a metaheuristic algorithm first introduced by [36] to simulate social behavior. The PSO algorithm is known to outperform backpropagation training in terms of speed and is independent of the bias value [37] but seems to be inferior to the GA in terms of accuracy [38]. For the present study, the PSO-ANN model has the following parameters:
    - Maximum number of iterations: 1000;
    - Population size: 100;
    - Damping ratio: 0.99;
    - Personal learning coefficient c_1: 1.5;
    - Global learning coefficient c_2: 2;
    - Training goal: 0.01;
    - Performance metric: *MSE*.

The PSO algorithm used for the optimization of the neural network is presented in detail in [39].

3. The *Big Bang-Big Crunch* algorithm is a simple and novel metaheuristic algorithm developed by [40] and inspired by the Big Bang theory about the beginning of the universe. According to this theory, the particles-which represent the candidate solutions-are initially spread uniformly over the search space. In the Big Crunch phase of the algorithm, a convergence operator is determined, and the particles converge to a specific point. In the Big Bang phase that follows, the particles will be drawn in an orderly way. These two phases are repeated until a stopping criterion is met. The aim is to lead the particles to a global optimum. The convergence operator can be either the weighted average of the positions of the candidate solutions or the position of the best candidate solution. The convergence operator, called the *Center of Mass* (*CM*), is calculated by Equation (2):

$$CM(i) = \sum_{j=1}^{nP} \frac{P(j,i)/PFit(j)}{\sum_{j=1}^{nP}(1/PFit(j)),\ i = 1,2,\ldots,nv} \tag{2}$$

where *nP* represents the population of the particles, *P* is the matrix of the candidate solutions, *PFit* is the vector of the penalty function, *nv* the number of design variables and *bestP* is the particle with the best fit.

On the Big Bang phase, the particles are updated according to Equation (3):

$$newP = (CM\ or\ bestP) + \frac{rand \times (Ub - Lb)}{nIT} \tag{3}$$

where *rand* is a random number drawn from the uniform distribution in (0, 1), *nIT* represents the number of iterations, and *Ub* and *Lb* are the upper and lower bounds of the search space, respectively. In [41] BB-BC algorithm is introduced for the parameter optimization of the Type-2 fuzzy neural network. For the present study, the BB-BC-ANN model had the following parameters:

- Maximum number of iterations: 1000;
- Population size: 100;
- Training goal: 0.01;
- Performance metric: *MSE*.

The training of an ANN using metaheuristic algorithms is schematically depicted in Figure 2.

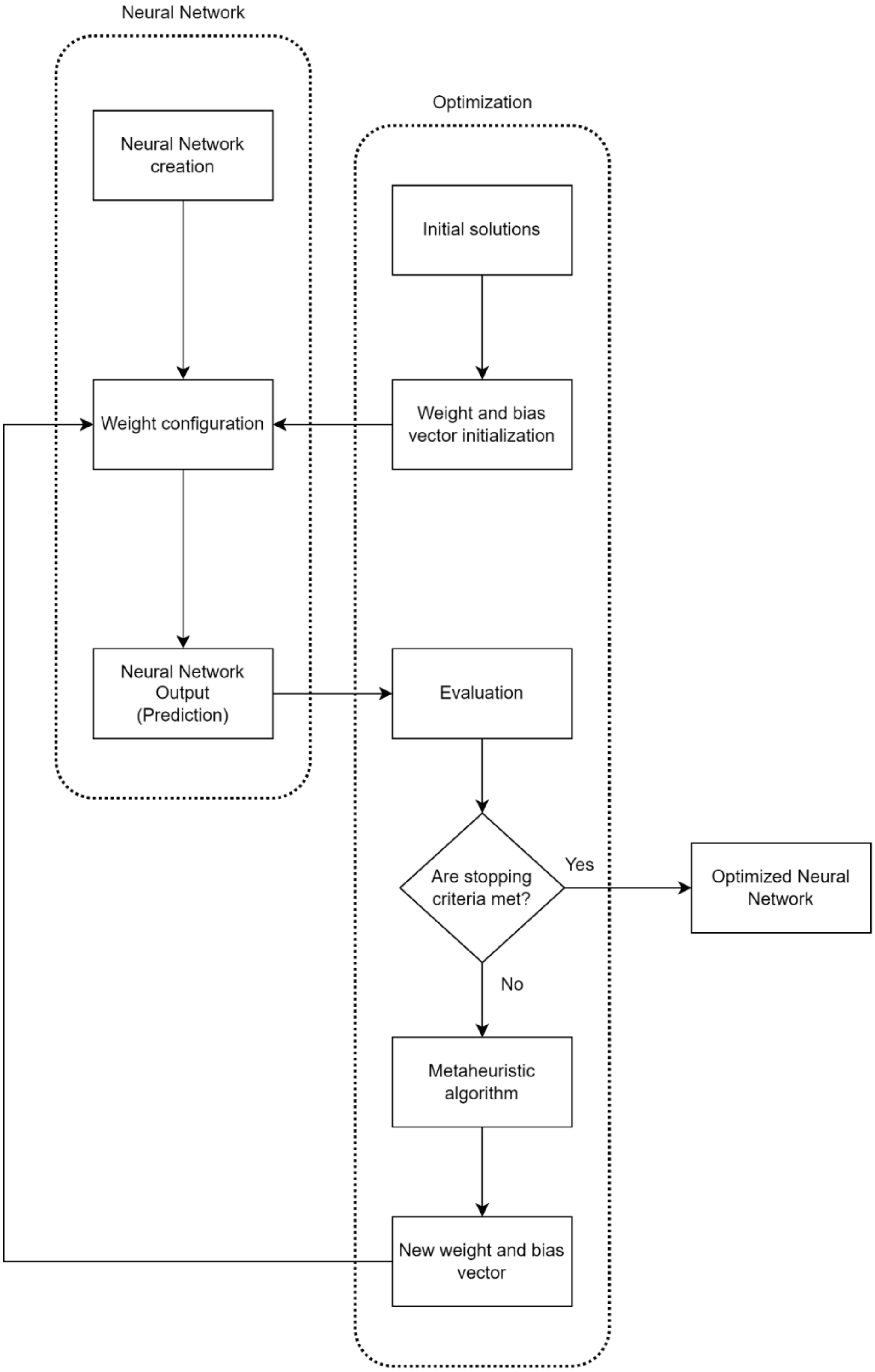

Figure 2. ANN's training using a metaheuristic algorithm.

2.1.2. Adaptive Neuro-Fuzzy Inference System

ANFIS proposed by Jang [42] has a similar to ANNs structure, with 24 inputs and one output (past 24-h load data as inputs and the next hour's prediction as output). Figure 3 illustrates the structure of the ANFIS used in the present study.

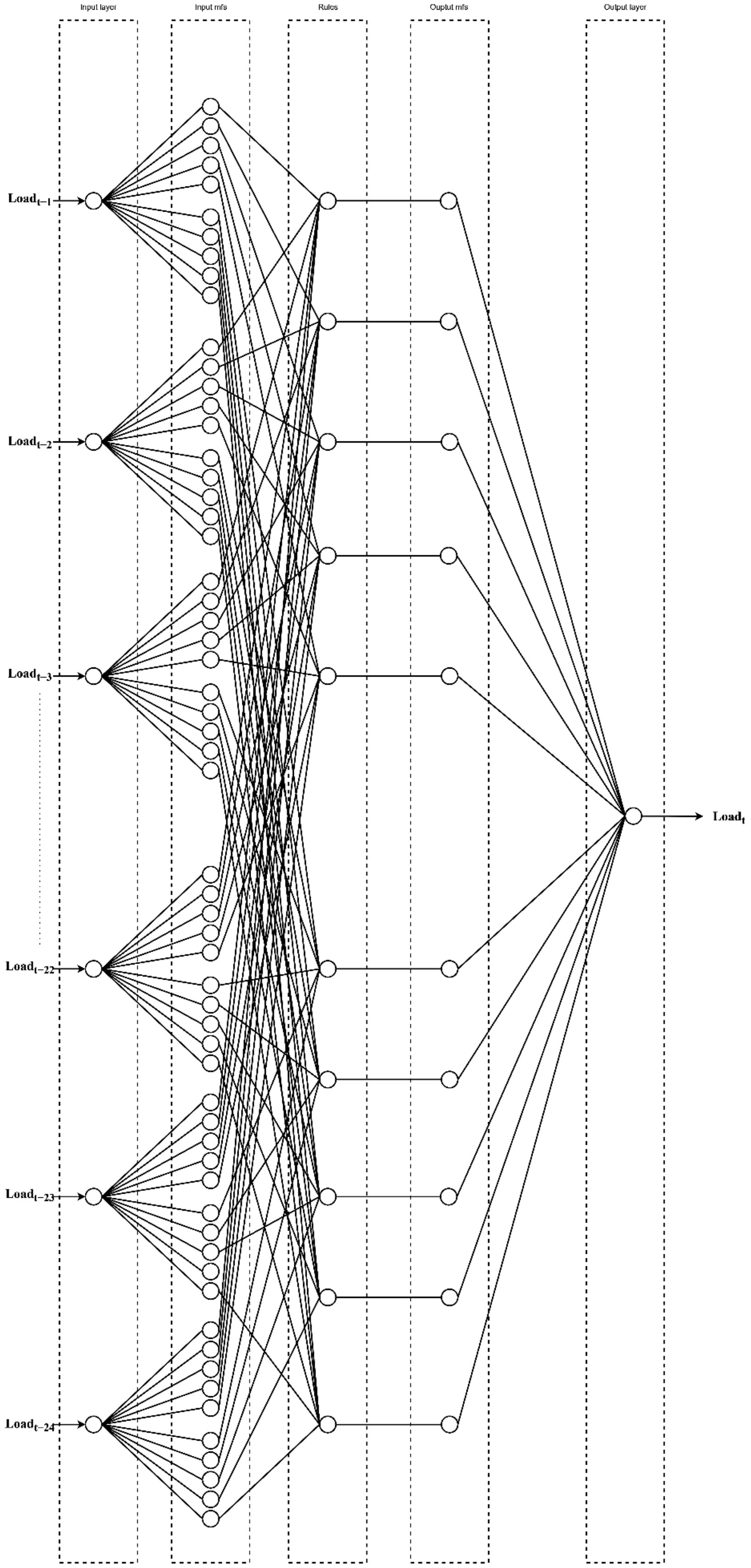

Figure 3. Structure of the TSK ANFIS used in the study.

ANFIS is a neural network functionally equivalent to a *Sugeno* fuzzy inference system. Sugeno's fuzzy model generates fuzzy rules from an input–output data set. The Sugeno fuzzy rule has the following form:

IF $Load_{t-1}$ is A_1
AND $Load_{t-2}$ is A_2
AND $Load_{t-24}$ is A_m
THEN $Load_t = f(Load_{t-1}, Load_{t-2}, \ldots, Load_{t-24})$

where $Load_{t-1}, Load_{t-2}, \ldots, Load_{t-24}$ are the input variables; $A_1, A_2, \ldots, A_m$ are the fuzzy sets; and $Load_t$ is either a constant or a linear function of the inputs. If $Load_t$ is a constant, the Sugeno fuzzy system is called a *zero-order Sugeno fuzzy model*, and the result of a rule is a singleton. When $Load_t$ is a first-order polynomial of the form:

$$Load_t = (k_0 + k_1 Load_{t-1} + k_2 Load_{t-2} + \ldots + k_{24} Load_{t-24}) \tag{4}$$

The model is called the first-order Sugeno fuzzy model.

Solving a problem using an ANFIS has the advantage that it is not necessary to have any prior knowledge of the rule parameters since the system learns these parameters by itself. Compared to ANNs, ANFIS seem to show a good performance on complex, nonlinear, and multivariate problems with very accurate results due to their remarkable generalization ability, while ANNs' performance is strongly dependent on their architecture-which is not a problem for ANFIS [43]. For the present study, the ANFIS had the following parameters:

- Clustering technique: Fuzzy C-Means clustering;
- Number of clusters: 10;
- Exponent: 2;
- Maximum number of epochs: 1000;
- Training goal: 0.01;
- Performance metric: *MSE*.

The ANFIS model used for the timeseries prediction is based on the model presented in [44].

2.1.3. Long Short-Term Memory Networks

LSTM networks belong to the category of recurrent neural networks specially designed to learn long-term dependencies between timesteps of sequences of data. They consist of a sequence input layer and an LSTM layer. The sequence input layer inputs sequential data into the network, while the LSTM layer learns the long-term dependencies between data timesteps. Figure 4 demonstrates the flow of an electrical Load time series of length S, which flows through an LSTM layer is given, where h_t and c_t denote the hidden state (the output) and the cell state at time step t, respectively.

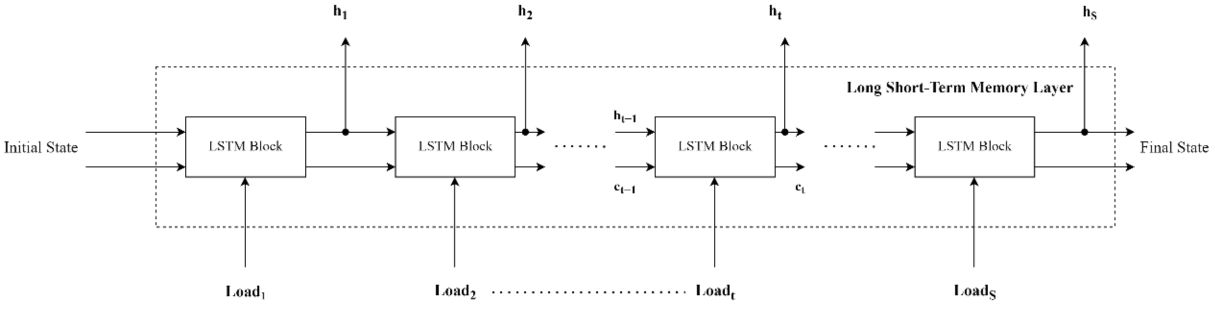

Figure 4. Time series flow through LSTM layer.

The initial state and the first timestep of the timeseries are used from the first LSTM block to calculate the first output and to update the cell's state. When in timestep t, the block computes the output and updates the cell state c_t by using the current state of the network, which is (c_{t-1}, h_{t-1}). The hidden state and the cell state form the layer's state. At timestep t, the hidden state contains this timestep's output of the LSTM layer. Information learned from previous timesteps is stored in the cell's state, and the layer adds or removes information at every timestep using gates to control these changes. Figure 5 shows the functions of a gate.

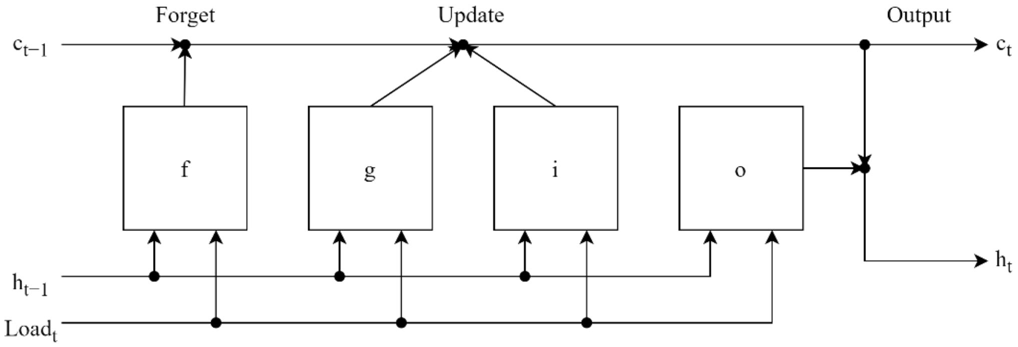

Figure 5. Functions of LSTM block's gate.

For the present study, the LSTM network had the following parameters:

- Gradient threshold: 1;
- Initial learning rate: 0.005;
- LR drop factor: 0.2;
- Performance metric: *MSE*.

The signal flow diagram of an LSTM network is shown below in Figure 6 [45].

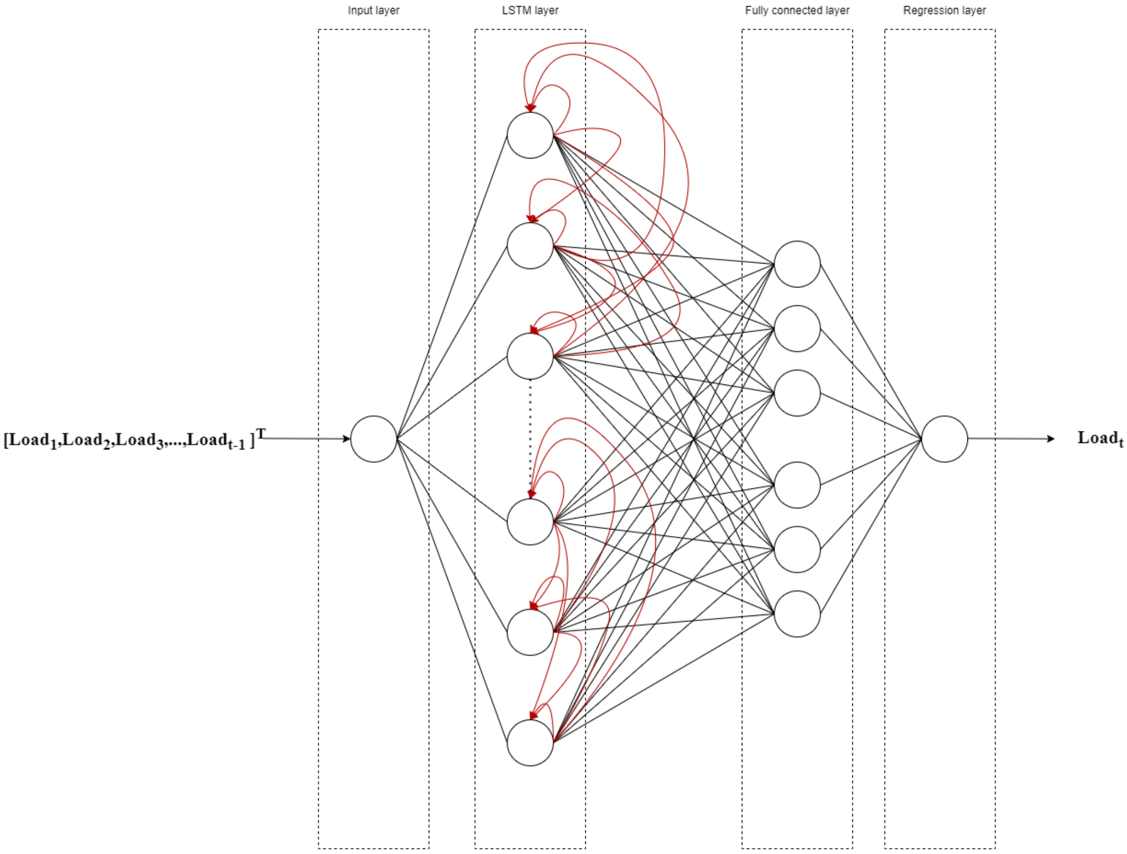

Figure 6. Signal flow diagram of an LSTM network.

2.2. Reference Building Model Description

The data are retrieved from the energy consumption simulation of a healthcare facility in San Diego, CA, USA, with **ENERGYPLUS** Version 9.0.1 software [39]. The building is a reference building developed by the *US Deparment of Energy*, and the general building's specifications are given in Table 1.

Table 1. Benchmark building's general specifications.

		Comments
Vintage	New Construction	
Location	Zone 3C: San Diego, CA, USA (warm, marine)	Selected climate based on ASHRAE Standard 169-2013
Available fuel types	Gas, electricity	
Building type (Principal building function)	Healthcare	
Building prototype	Hospital	
Total floor area including basement	22,436.18 m^2	241,410 sq.feet
Number of floors	5 plus basement	
Coordinates	Latitude [deg] 32.58, Longitude [deg]−117.0	
Elevation	159 m	

Figure 7 depicts the shape of the building under study.

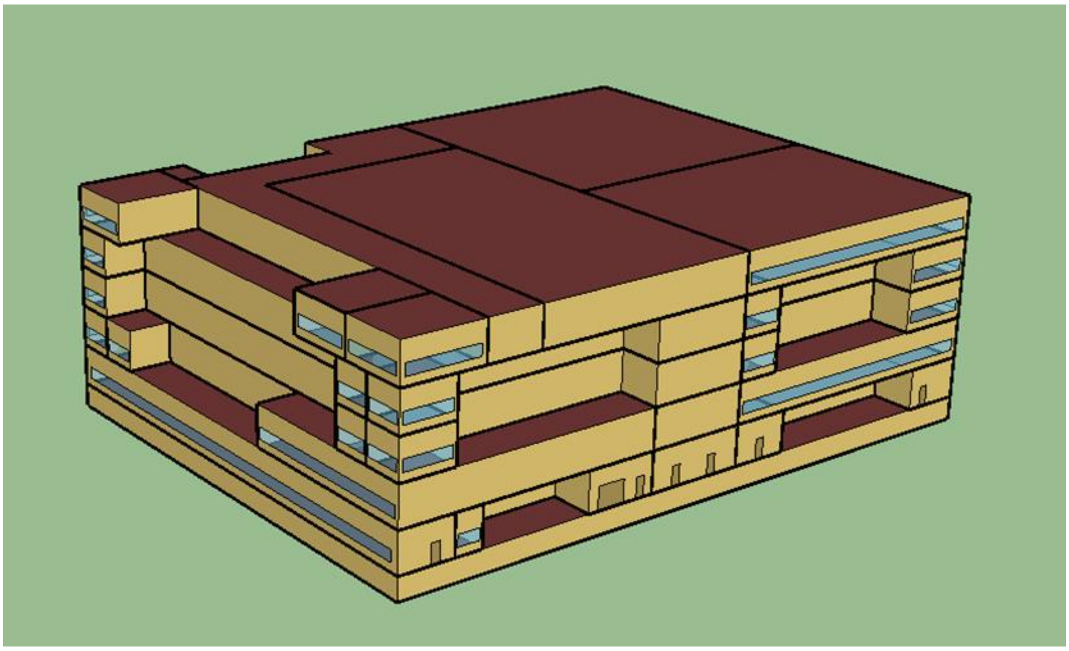

Figure 7. The shape of the reference building.

2.3. Obtained Data

The simulated period is from 1 January 2014, 01:00 to 1 January 2020, 01:00 (52,584 hourly timesteps). The sample has been divided into training and test sets, which are randomly selected, with 80% of the samples used for the training set (42,048 data points) and the rest 20% of the samples for the test set (10,512 data points). All neural networks (LM-ANN, SCG-ANN, GDX-ANN, GA-ANN, PSO-ANN, and BB-BC-ANN) and ANFIS had 24 inputs representing the hourly loads for the past 24 h, and one output, the load prediction for the next hour. The training set consists of 24 × 42,048 input samples and 42,048 targets, and the test set of 24 × 10,512 inputs and their associated targets. LSTM training set consists of 1 × 42,048 input samples and the according targets, and the test set consists of 10,512 data samples.

The models are created, trained, and tested using **MATLAB** R2020b software. Figure 8 demonstrates the training and test sets.

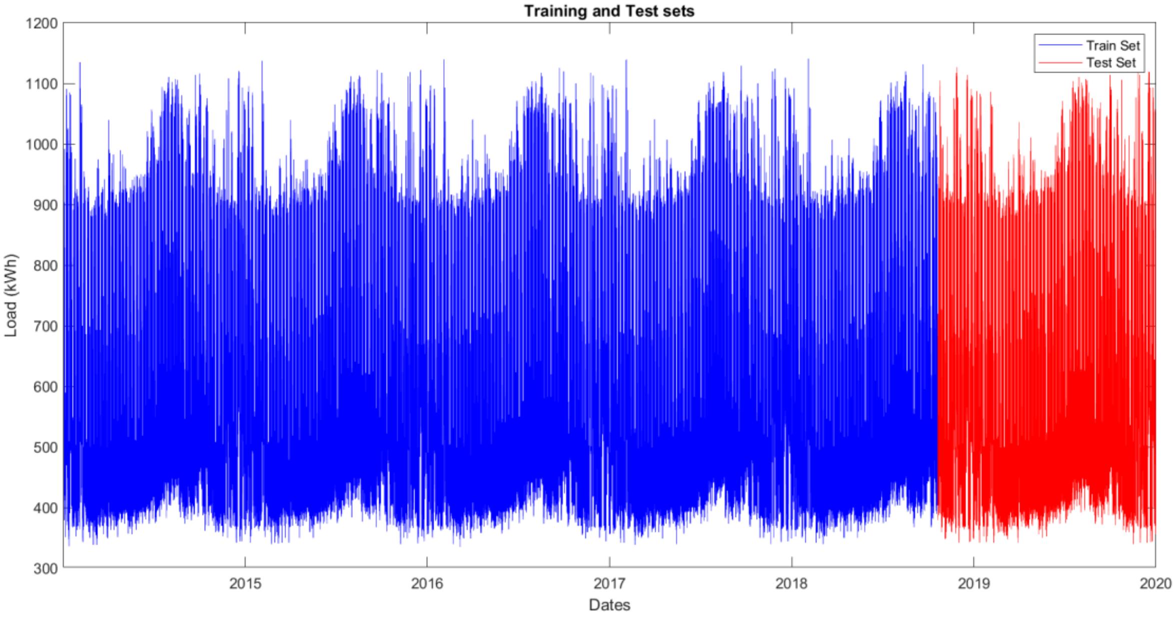

Figure 8. The train and test sets.

2.4. Evaluation Metrics

Mean Square Error used as the performance function during the training of the models:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{5}$$

where y and $\hat{y}$ denote the observed and the predicted values, respectively, and n the number of observations. A model with less *MSE* leads to more precise predictions.

The correlation between the predicted and the observed values is measured by the correlation coefficient R (also known as Pearson's correlation coefficient):

$$R = \frac{n\cdot\sum y_i\hat{y}_i - (\sum y_i)\cdot(\sum \hat{y}_i)}{\sqrt{\left[n\cdot(\sum {y_i}^2) - (\sum y_i)^2\right]\cdot\left[n\cdot(\sum {\hat{y}_i}^2) - (\sum \hat{y}_i)^2\right]}} \tag{6}$$

An R-value approximate to 1 indicates a strong correlation between the observed and the predicted variables, stating that the model is a good fit.

The coefficient of determination (R^2, indicates the extent to which the energy load is predictable, showing how well the model fits the observed data:

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \overline{y}_i)^2} \tag{7}$$

As a helpful metric to evaluate the models, *Mean Absolute Percentage Error* (*MAPE*) is chosen because of its ability to be scale independent, and thus, it can be used as a metric for comparison between different models:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{8}$$

Moreover, expressing the models' accuracy in percentages helps the reader to obtain a better insight into the models since it is easily interpretable.

Additionally, the models' precision can be estimated using the *confidence interval* (*CI*). Assuming that the number of samples is large enough, even if their distribution is not normal, according to the central limit theorem, the confidence interval is [46]:

$$CI = \left(MAPE - z^{*}\cdot\frac{s}{\sqrt{n}}\,,\ MAPE + z^{*}\cdot\frac{s}{\sqrt{n}}\right) \tag{9}$$

Confidence intervals allow us to make statements concerning the likely range that a population parameter (in a certain case, the *MAPE* value) lies within. To obtain the confidence interval, a confidence level must be defined [47]. A 95% confidence level means that there is a 95% possibility for the absolute percentage error of measurement to lie between the confidence interval bounds. z^{*} represents the number of standard deviations for a given confidence level. For 95% confidence level, the z^{*} (or *critical z-score*) is 1.96. Note that in Table 2, the *MAPE* confidence intervals were computed using the standard deviation of the absolute percentage errors for each prediction–observation pair.

Table 2. Results of the training process.

Predictor	*MSE*	*R*	R^2	*MAPE* %	*MAPE* with 95% *CI*	*MBE*
LM-ANN	539.90	0.99423	0.9885	2.62	2.59–2.65	-7.8828×10^{-10}
SCG-ANN	623.57	0.99333	0.9867	2.83	2.80–2.87	0.0098745
GDX-ANN	510.73	0.99454	0.9891	2.46	2.43–2.50	3.1315×10^{-5}
GA-ANN	2871.6	0.96895	0.9389	6.85	6.79–6.91	2.3268
PSO-ANN	1234.6	0.98675	0.9737	4.19	4.14–4.23	0.089787
BB-BC-ANN	1682.0	0.9819	0.9641	5.27	5.22–5.32	0.15657
ANFIS	442.98	0.99526	0.9905	2.10	2.07–2.13	−0.0011551
LSTM	680.84	0.99271	0.9855	3.07	3.04–3.11	0.020457

Finally, *MBE* (*Mean Bias Error*) is also employed as a metric of the forecasting algorithm's tendency to over- or underpredict:

$$MBE = \frac{1}{n}\sum_{i=1}^{n}\hat{y}_i - y_i \tag{10}$$

3. Results

3.1. Training Phase

The models are trained, and the results of the training process are shown in Table 2.

At a glance, ANFIS, GDX-ANN, LM-ANN, SCG-ANN, and LSTM predictors perform better than the ANNs trained with a metaheuristic algorithm, achieving significantly better results in all performance metrics during training. PSO-ANN and BB-BC-ANN metrics group these predictors as "*moderate*", and, finally, GA-ANN predictor performance can be labeled as "*poor*". ANFIS has the best performance in terms of *MSE*, *R*, and R^2. Confidence intervals are slightly narrower in ANFIS and LM-ANN models than in the CIs of the SCG-ANN, GDX-ANN, and LSTM. Metaheuristic-algorithm-trained models have broader *CIs*, showing that the models' precision is lower than the others. LM-ANN and ANFIS have small (especially LM-ANN's *MBE* can practically be considered zero) but negative *MBEs*, indicating the models' tendency to underpredict. GDX-ANN has also *MBE* close to zero—but positive—and all other predictors have various *MBE* positive values, showing their tendency to overpredict.

3.1.1. Performance during Training

Figure 9 shows the graphs of *MSE* versus Epochs, Generations, or Iterations for each predictor.

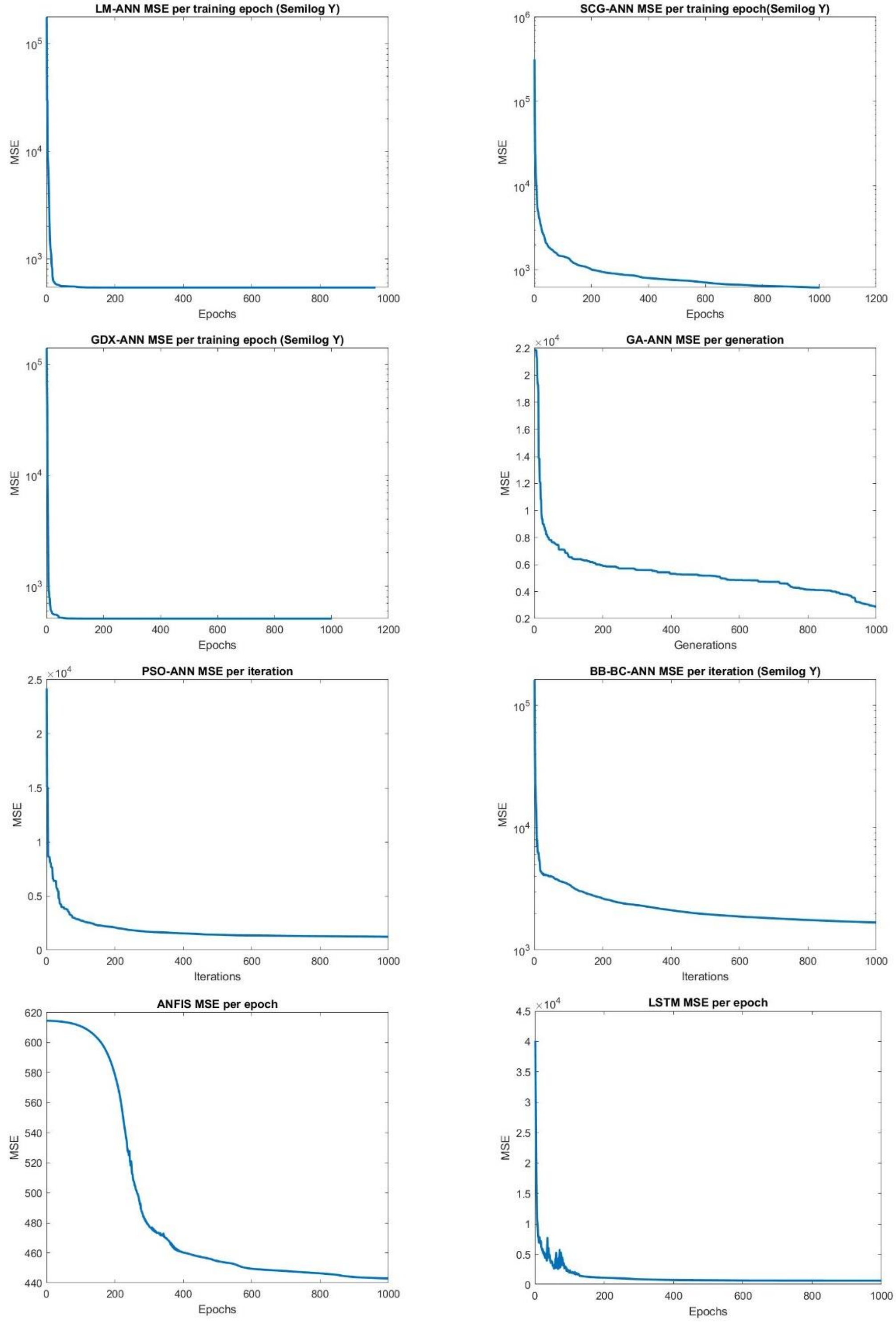

Figure 9. *MSE* versus number of iterations for the examined predictors.

All predictors reached their best performance at the maximum number of epochs/iterations except LM-ANN, which reached the best performance at 960 epochs, and ANFIS, which reached its best at 999 epochs. During training, the *MSE* of the LM-ANN predictor starts from very high values of 10^6~10^5 and converges to *MSE* values near to best before epoch 100. Practically, the *MSE* slightly changes after that point. SCG-ANN predictor's *MSE* also starts from values larger than 10^5 and declines exponentially to its best at epoch 1000, which is the stopping criterion. GDX-ANN predictor also starts training with a large *MSE* near 10^5 and then rapidly converges. However, *MSE* presents minor changes from iteration 70 and up. PSO-ANN starts from high *MSE* values and decreases exponentially at a very slow rate, while *MSE* from epoch 600 and up does not seem to change dramatically. BB-BC-ANN predictor falls rapidly from high *MSE* values and then converges slowly to its best performance until the stopping criterion of 1000 epochs is met. ANFIS *MSE* starts from 614, falls to 450 at the first 600 epochs, and declines at a very slow rate until the stopping criterion of 1000 epochs is met, forming an inversed S-shaped graph. LSTM starts with a large *MSE* which bounces heavily on the first 130 epochs while decreasing at the same time and decreases slowly to its best value afterward until the stopping criterion of maximum epochs is met.

3.1.2. Error Distribution for Training Phase

From the error distribution plots in Figure 10, the distribution of errors for LM-ANN seems to be strongly leptokurtic with a kurtosis of 11.64, exhibiting a "spike" close to zero. The skewness of -0.72 indicates a weak error distribution concentration to the right. SCG-ANN error distribution also has a similar form to LM-ANN error distribution with a kurtosis of 10.71 and a skewness of -0.63. GDX-ANN also demonstrates a large kurtosis of 14.56 and a positive skewness of -1.06. GA-ANN predictor has a kurtosis of 5.06 and skewness of 0.65. PSO-ANN error distribution demonstrates a kurtosis of 7.38 and a skewness of -0.2. BB-BC-ANN error distribution has a kurtosis of 5.6 and a skewness of 0.4. ANFIS error distribution is also strongly leptokurtic with a kurtosis of 17.04 and a skewness of -0.91.

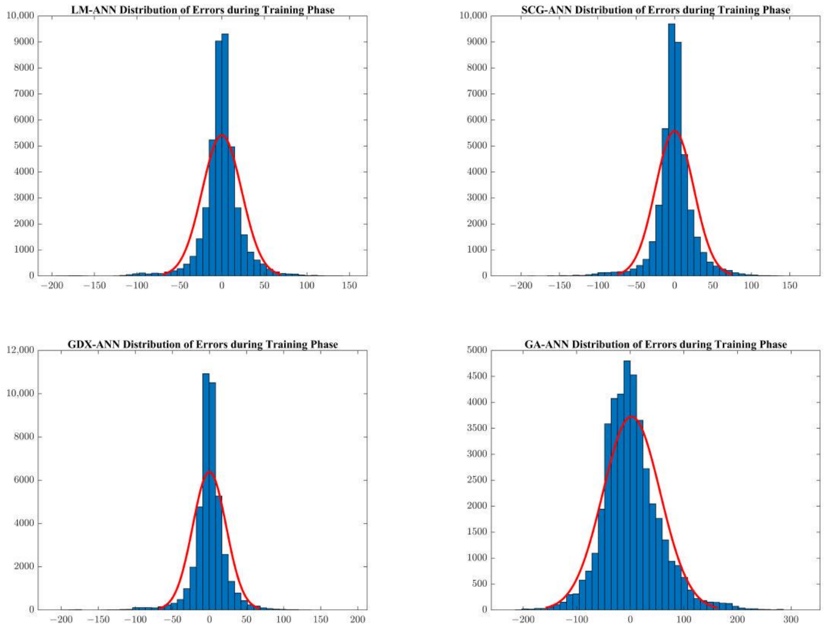

Figure 10. *Cont.*

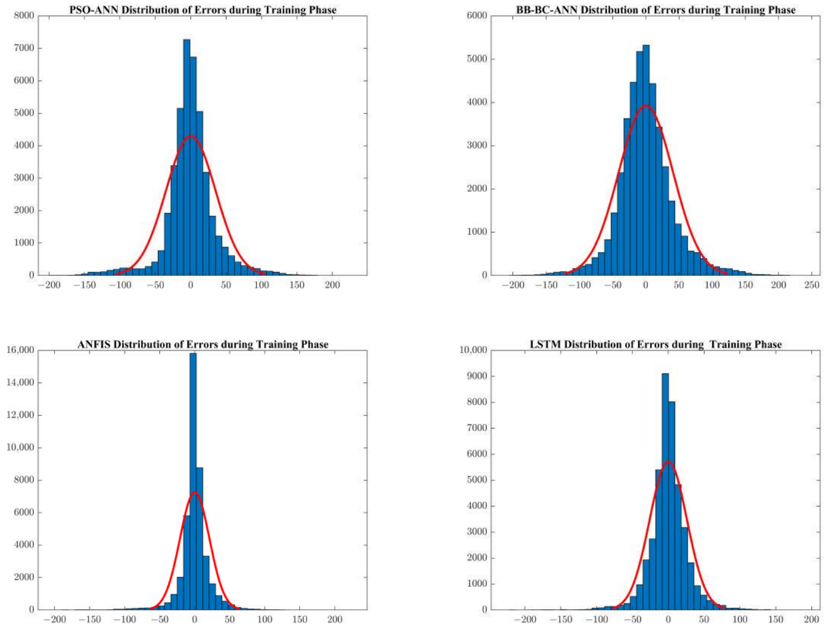

Figure 10. Error distribution for the training phase.

Lastly, LSTM network error distribution also demonstrates a kurtosis of 10.45 and a skewness of −0.45. One can observe that kurtosis and skewness of error distribution may be indicative of a predictor's accuracy, which is rational since kurtosis and skewness are dependent on the distribution of errors which in certain case is not normal.

3.1.3. Regression Plots for the Training Phase

From the plots in Figure 11, the predictions and their respective observations are almost perfectly matched in the cases of ANFIS, GDX-ANN, LM-ANN, SCG-ANN, and LSTM models, as expected from Table 2, and their *R*-values are close to 1. PSO-ANN and BB-BC-ANN present moderate fitness, while GA-ANN fitness can be described as "*poor*", with the lowest *R*-value of 0.96895. Examining Figures 10 and 11, it can be concluded that the more samples are close to zero, the more the *R*-value closes to one.

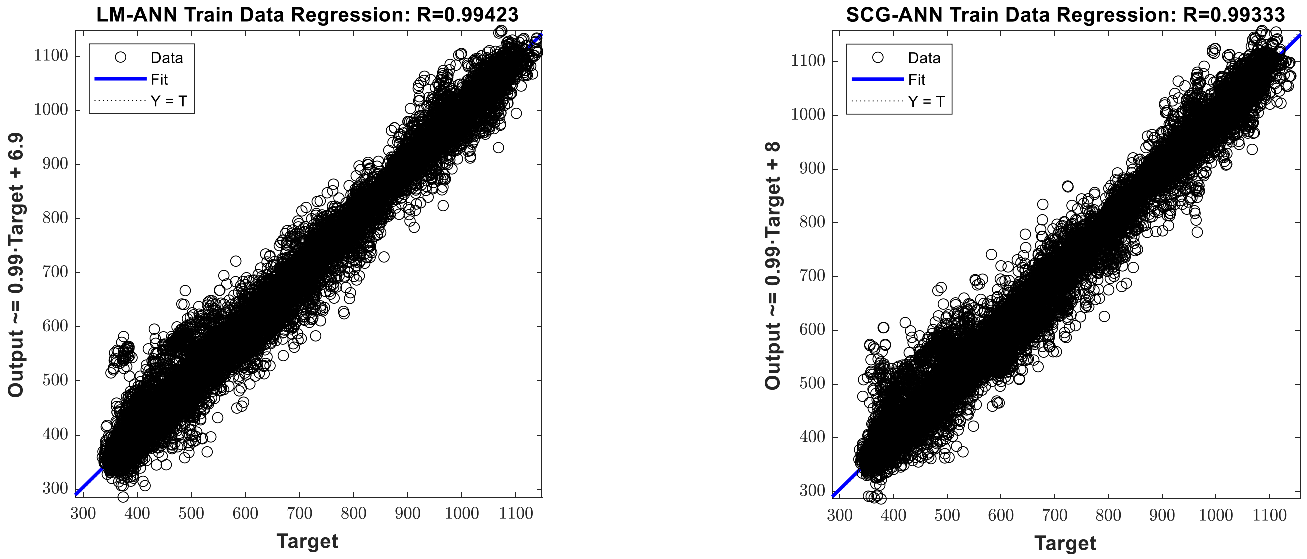

Figure 11. *Cont.*

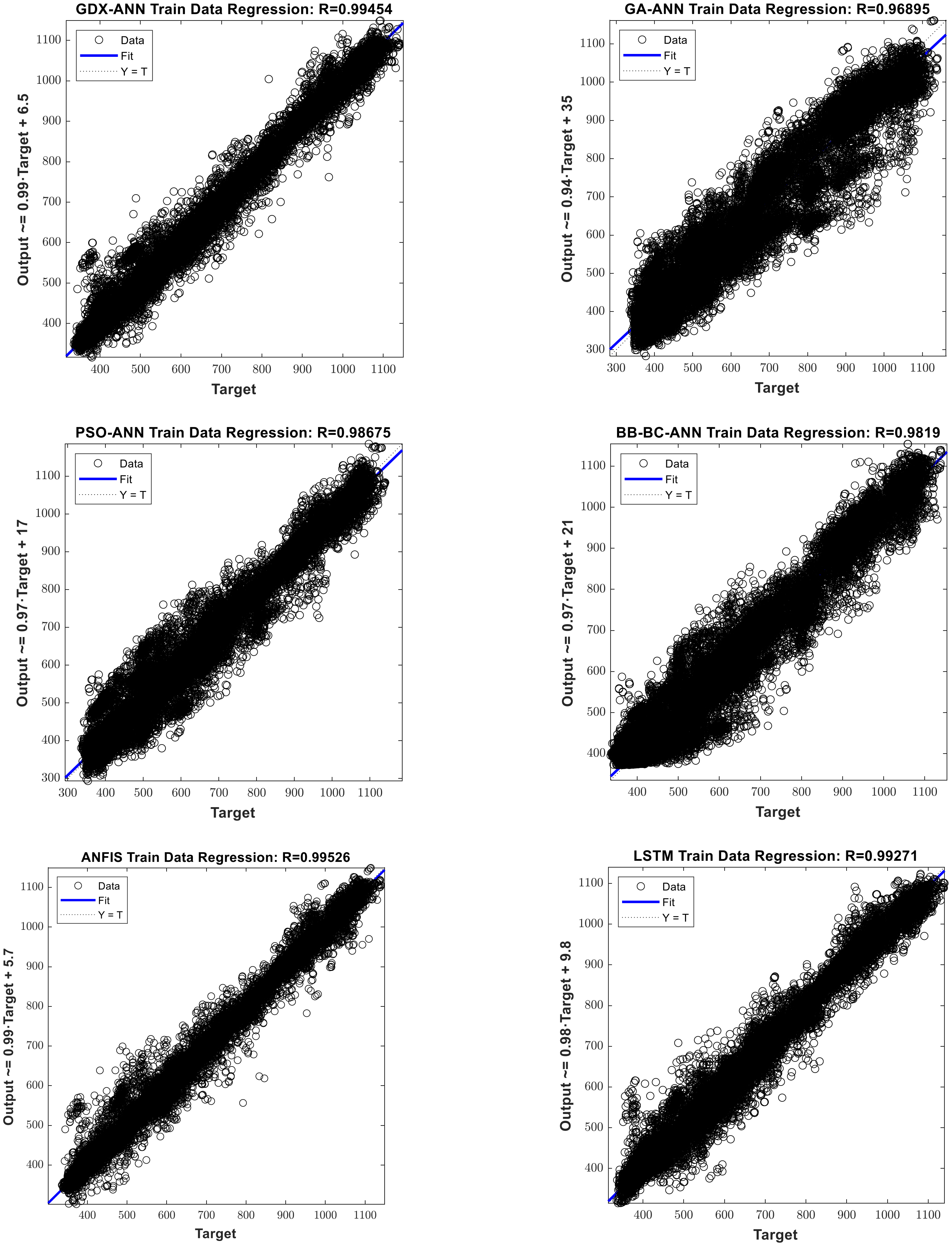

Figure 11. Regression plots for the training phase.

3.2. Testing Phase

The models are tested, and the results of the testing process are shown in Table 3.

Table 3. Testing results.

Predictor	*MSE*	*R*	R^2	*MAPE* %	*MAPE* with 95% *CI*	*MBE*
LM-ANN	613.95	0.99339	0.9868	2.78	2.71–2.85	−0.09
SCG-ANN	711.77	0.99234	0.9847	3.02	2.94–3.10	−0.35154
GDX-ANN	599.48	0.99355	0.9871	2.66	2.59–2.74	−0.27765
GA-ANN	3064.0	0.96664	0.9344	7.1	6.98–7.22	2.2024
PSO-ANN	1320.7	0.98581	0.9718	4.36	4.26–4.46	−0.035637
BB-BC-ANN	1825.0	0.98037	0.9611	5.52	5.42–5.62	0.41037
ANFIS	537.71	0.99422	0.9885	2.27	2.20–2.34	0.062046
LSTM	787.18	0.99152	0.9831	3.25	3.17–3.32	0.033282

Again, ANFIS, GDX-ANN, LM-ANN, SCG-ANN, and LSTM predictors perform better than the ANNs trained with a metaheuristic algorithm, achieving significantly better results in all performance metrics during testing. PSO-ANN and BB-BC-ANN metrics group these predictors as "*moderate*", and, finally, GA-ANN predictor performance can be labeled as "*poor*". ANFIS has the best performance in terms of *MSE*, *R*, R^2, and *MAPE*. ANFIS also has the best precision according to the confidence interval showing that the 95% of the predictions fall in the range of 2.2~2.34% of percentage error. LM-ANN demonstrates an *MBE* much bigger than the training phase's *MBE*. *MBE* also shows that in the case of GA-ANN, the model has a strong tendency to overpredict, while ANFIS has positive *MBE*, in contrast to the training phase. SCG-ANN, GDX-ANN, and PSO-ANN demonstrated negative *MBE*, also in contrast with the training phase.

3.2.1. Error Distribution for Testing Phase

From Figure 12, the distribution of errors for LM-ANN seems to be strongly leptokurtic with a kurtosis of 11.67, exhibiting a "spike" close to zero. The skewness of −0.89 indicates a weak error distribution concentration to the right. SCG-ANN error distribution also has a similar form to LM-ANN error distribution with a kurtosis of 11.25 and a skewness of −0.76. GDX-ANN also demonstrates a large kurtosis of 14.85 and a negative skewness of −1.19. GA-ANN predictor has a kurtosis of 4.73 and skewness of 0.57. PSO-ANN error distribution demonstrates a kurtosis of 7.01 and a skewness of −0.26. BB-BC-ANN error distribution has a kurtosis of 5.17 and a skewness of 0.35. ANFIS error distribution is also strongly leptokurtic, with a kurtosis of 17.91 and a skewness of −0.71.

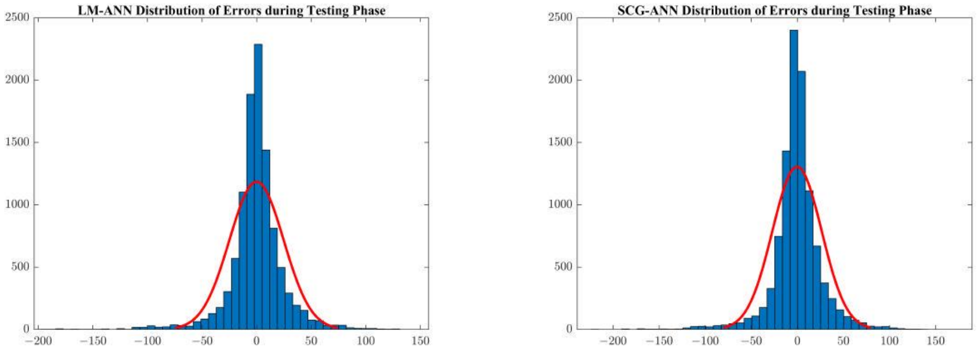

Figure 12. *Cont.*

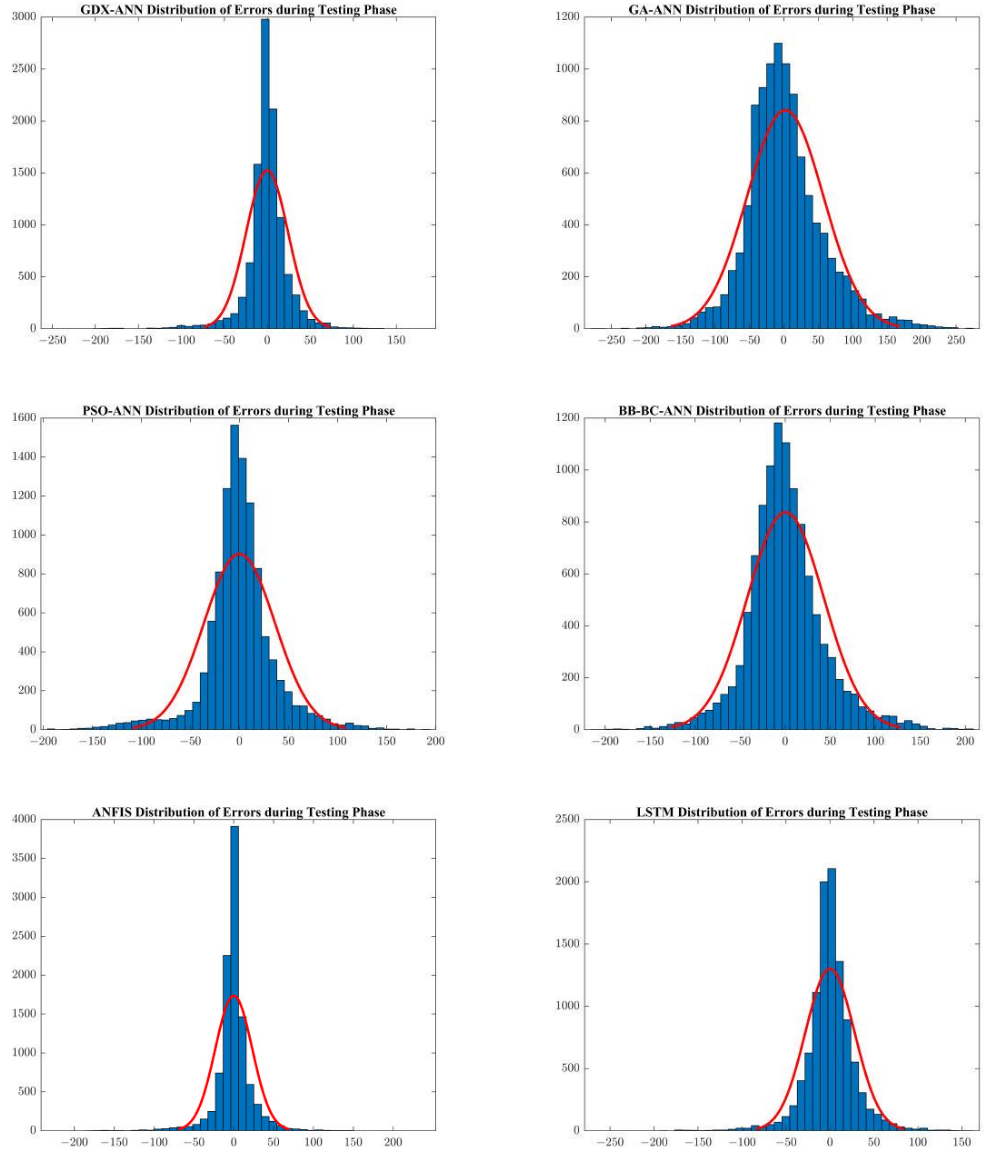

Figure 12. Error distribution for the testing phase.

Finally, LSTM network error distribution also demonstrates a kurtosis of 11.07 and a skewness of −0.64. Since the skewness and kurtosis are not zero for any of our models, the error distribution of the testing phase is not normal.

3.2.2. Regression Plots for the Testing Phase

From the plots in Figure 13, the predictions and their respective actual values are almost perfectly matched in the cases of ANFIS, SCG-ANN, LM-ANN, GDX-ANN, and LSTM models, as expected from the Table 3. *R*-values are close to 1 for these predictors. PSO-ANN and BB-BC-ANN present moderate fitness, while GA-ANN fitness can be described as "*poor*", with the lowest *R*-value of 0.96664.

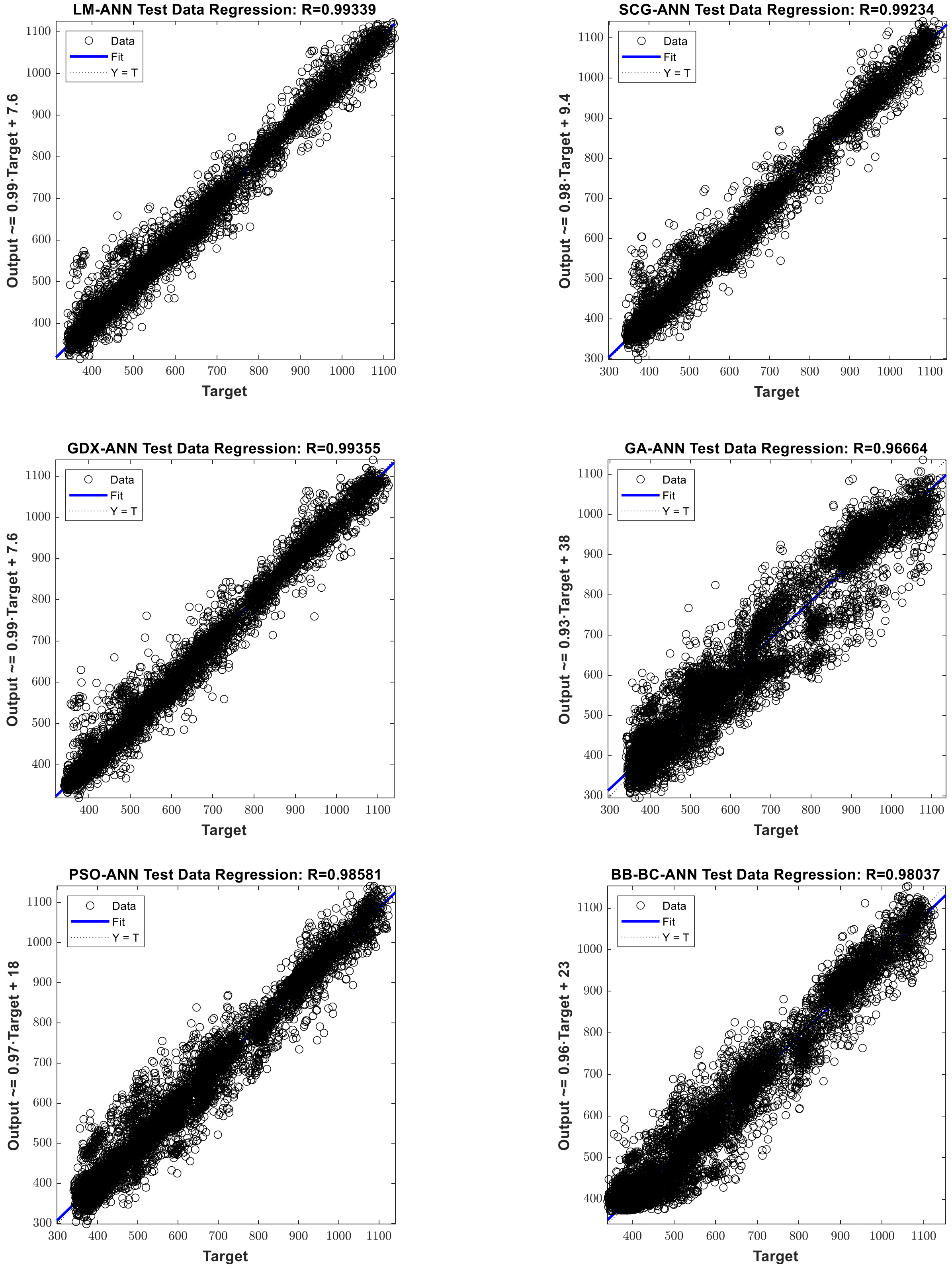

Figure 13. *Cont.*

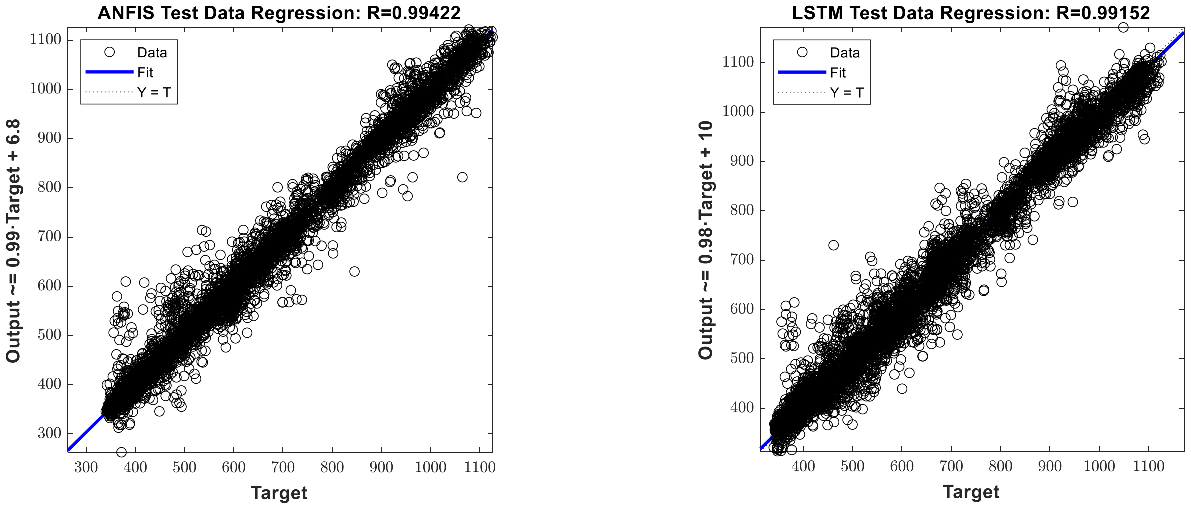

Figure 13. Regression plots for the testing phase.

3.2.3. *MAPE* Boxplot of the Predictors

In Figure 14, *MAPE* boxplots of the predictors are presented. Boxplots are a useful graphic tool for explanatory data analysis. Again, LM-ANN, SCG-ANN, GDX-ANN, and ANFIS have similar behavior, while GA-ANN has a bigger median value as well as upper and lower whiskers. LM-ANN, PSO-ANN, and BB-BC-ANN show a good performance concerning the outliers, which are significantly fewer than the other models' outliers.

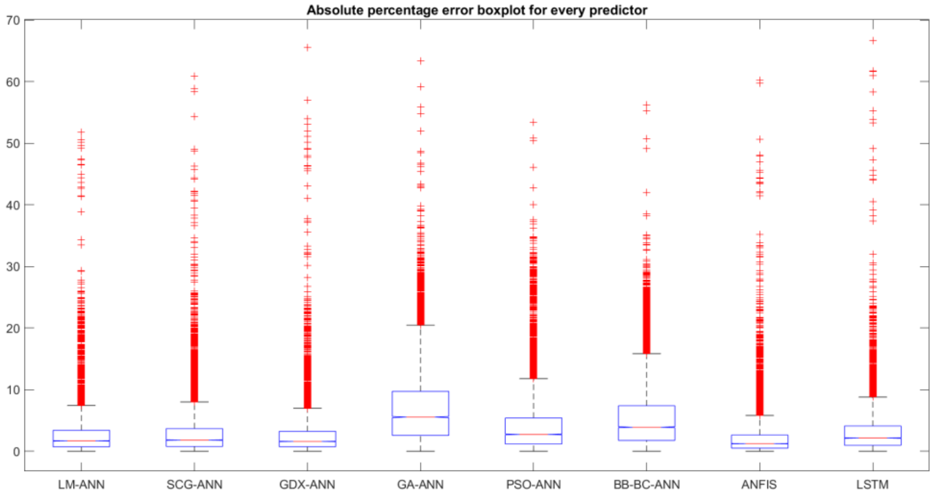

Figure 14. Boxplot for *MAPE* of the test sets.

In Figure 14, the model's accuracy is visually represented since the difference between the upper and the lower quartile, known as the interquartile range (IQR), indicates the distribution of absolute percentage errors for each predictor. As expected, LM-ANN, SCG-ANN, GDX-ANN, and ANFIS achieved low distribution of errors and fewer outliers than the other models. GA-ANN, BB-BC-ANN, and PSO-ANN, likewise, have a large IQR, indicating the spreading of the errors, so the models are not precise.

3.2.4. *MAPE* Results from a Randomly Selected Day

Table 4 shows the *MAPE* achieved from the predictors over a day's prediction from 08:00 of 15 April 2019 to 07:00 of 16 April 2019. Again, ANFIS has the best performance, while GA-ANN has the worst. In Figure 15, the predictions obtained from ANFIS and GA-ANN against the actual load are shown. ANFIS prediction almost fits perfectly to the actual load curve, while, on the contrary, GA-ANN's prediction is poor.

Table 4. *MAPE* results for a random day.

Predictor	*MAPE* %
LM-ANN	1.37
SCG-ANN	1.88
GDX-ANN	1.59
GA-ANN	5.53
PSO-ANN	3.06
BB-BC-ANN	4.54
ANFIS	1.23
LSTM	1.88

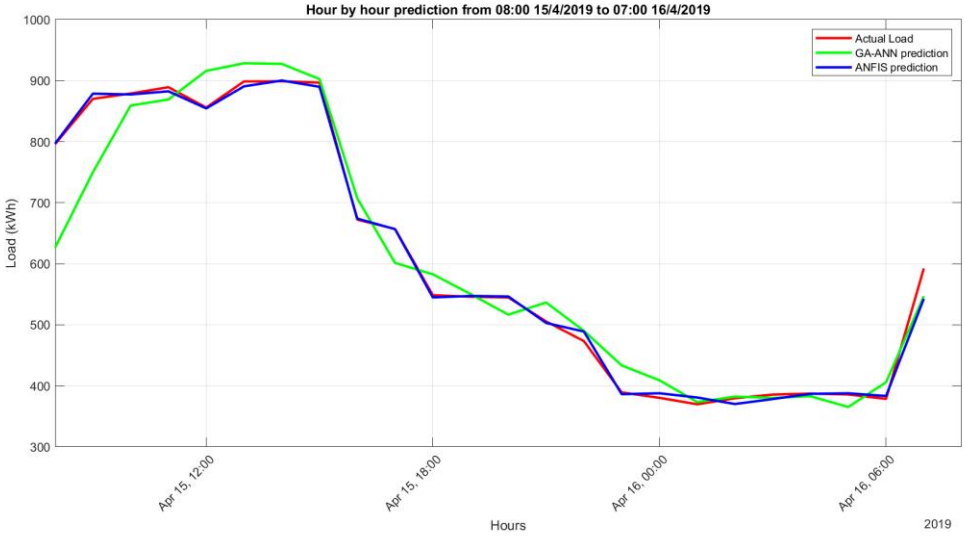

Figure 15. ANFIS and GA-ANN 24-h predictions and actual loads.

3.2.5. GA-ANN Optimization

Since the GA-ANN algorithm's performance is poor, a brief screening of certain parameters is necessary, and trials with combinations of different parameters took place to examine if there is room for improvement. The parameters' values and the according performance are given in Table 5.

Table 5. GA-ANN's parameters and the according performances during training.

Combination Number	Population	Crossover Rate	Mutation Rate	Generations	*MSE*	*MAPE* %
1	60	0.9	0.01	1000	2609	6.41
2	70	0.8	0.0034	1000	2785	6.69
3	90	0.7	0.0017	95	10,281	14.04
4	100	0.6	0.0001	81	13,688	15.92

From the results above, the combination which seems to perform better is combination number 1, which includes 60 individuals as a population, a Crossover rate of 0.9, a Mutation rate of 0.01 with an *MSE* of 2609, and a *MAPE* of 6.41%. Close enough is the GA-ANN model with combination number 2 parameters with an *MSE* of 2785 and a *MAPE* of 6.69. Combinations number 3 and 4 have large *MSE* values since they stall at 95 and 81 generations, respectively. Combinations 1 and 2 performances are close to the initial combination's performance.

3.2.6. Reproducibility of the Models

To study the reproducibility of the models, a test of the models is performed for a healthcare facility located in New York City. The simulated periods are January and June of 2021. *MAPEs* of the models are presented in Tables 6 and 7.

Table 6. *MAPE* results for January 2021.

Predictor	*MAPE* %
LM-ANN	17.81
SCG-ANN	17.8
GDX-ANN	16.31
GA-ANN	17.67
PSO-ANN	17.3
BB-BC-ANN	18.07
ANFIS	16.81
LSTM	3.69

Table 7. *MAPE* results for June 2021.

Predictor	*MAPE* %
LM-ANN	17.81
SCG-ANN	17.8
GDX-ANN	16.31
GA-ANN	17.67
PSO-ANN	17.3
BB-BC-ANN	18.07
ANFIS	16.81
LSTM	3.69

From Tables 6 and 7, it seems that GDX-ANN and ANFIS perform better than the other predictors. However, the accuracy of the predictors is not so good because the climate zone of the building is now different. LSTM network achieves a remarkable *MAPE*, probably because of its ability to learn long-term dependencies from timeseries data. Additionally, the predictors have the same *MAPE* for both January and June.

4. Discussion

In the present research paper, eight different predictors are tested on data obtained from a simulated healthcare building model. ANFIS outperformed all other predictors in terms of accuracy and had a better *R* score than in [6] and in any other metric as well. However, backpropagation algorithms such Levenberg–Marquardt, Scaled Conjugate Gradient, and Adaptive Learning Rate with Momentum performed well. Big Bang–Big Crunch and Particle Swarm Optimization performance can be labeled as "*moderate*". Genetic Algorithms performance is characterized as "*poor*" and not suitable for neural network training. LSTM also performed nearly as well as ANFIS and BP algorithms. Moreover, the LM algorithm converged faster than every other algorithm [7], and future research may pay attention to this, focusing on techniques that may help algorithms to converge faster without compromising in accuracy. The matter of time and computing resources is a subject that needs further investigation in future work.

Furthermore, metaheuristic algorithms strongly rely on their parameter sets, and optimization techniques for every single of them must be investigated in future work.

Another interesting aspect of the study is the replicability of the models. Since the resulting *MAPEs* are significantly higher than the values obtained from the benchmark building, which is used for the training of the models, future research must focus on refining them in order to give better results.

Since each predictor's capabilities in load forecasting are well known and accurate forecasting is feasible, future research will also focus on combinations of load and price forecasting techniques to develop a decision-making system that may automatically perform profitable energy transactions within a smart grid context while ensuring comfortable conditions for the residents but can additionally keep the facility resilient under extreme circumstances of load demand.

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

Funding: This research received no external funding.

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

References

1. European Commission. 2022. Available online: https://ec.europa.eu/info/news/focus-energy-efficiency-buildings-2020-lut-17_en (accessed on 2 May 2022).
2. Gordillo-Orquera, R.; Lopez-Ramos, L.M.; Muñoz-Romero, S.; Iglesias-Casarrubios, P.; Arcos-Avilés, D.; Marques, A.G.; Rojo-Álvarez, J.L. Analyzing and Forecasting Electrical Load Consumption in Healthcare Buildings. *Energies* **2018**, *11*, 493. [CrossRef]
3. Kyriakarakos, G.; Dounis, A. Intelligent Management of Distributed Energy Resources for Increased Resilience and Environmental Sustainability of Hospitals. *Sustainability* **2020**, *12*, 7379. [CrossRef]
4. World Bank Group. Climate-Smart Healthcare: Low-Carbon and Resilience Strategies for the Health Sector. 2017. Available online: https://documents1.worldbank.org/curated/en/322251495434571418/pdf/113572-WP-PUBLIC-FINAL-WBG-Climate-smart-Healthcare-002.pdf (accessed on 2 May 2022).
5. Health Care Without Harm. The Energy Efficiency Directive. 2017. Available online: https://noharm-europe.org/sites/default/files/documents-files/5047/2017-10-09HCWHEurope_Energy_Efficiency_Position_Paper.pdf (accessed on 23 July 2022).
6. Coccagna, M.; Cesari, S.; Valdiserri, P.; Romio, P.; Mazzacane, S. Energy Consumption in Hospital Buildings: Functional and Morphological Evaluations of Six Case Studies. *Int. J. Environ. Sci.* **2017**, *2*, 443–452. [CrossRef]
7. Soliman, S.A.-H.; Al-Kandari, A.M. *Electrical Load Forecasting: Modeling and Model Construction*; Elsevier: Amsterdam, The Netherlands, 2010.
8. European Commission. Delivering the European Green Deal. Available online: https://ec.europa.eu/info/strategy/priorities-2019-2024/european-green-deal/delivering-european-green-deal_en (accessed on 2 May 2022).
9. Dagdougui, H.; Bagheri, F.; Le, H.; Dessaint, L. Neural network model for short-term and very-short-term load forecasting in district buildings. *Energy Build.* **2019**, *203*, 109408. [CrossRef]
10. Manno, A.; Martelli, E.; Amaldi, E. A Shallow Neural Network Approach for the Short-Term Forecast of Hourly Energy Consumption. *Energies* **2022**, *15*, 958. [CrossRef]
11. Slowik, M.; Urban, W. Machine Learning Short-Term Energy Consumption Forecasting for Microgrids in a Manufacturing Plant. *Energies* **2022**, *15*, 3382. [CrossRef]
12. Kampelis, N.; Tsekeri, E.; Kolokotsa, D.; Kalaitzakis, K.; Isidori, D.; Cristalli, C. Development of Demand Response Energy Management Optimization at Building and District Levels Using Genetic Algorithm and Artificial Neural Network Modelling Power Predictions. *Energies* **2018**, *11*, 3012. [CrossRef]
13. Bagnasco, A.; Fresi, F.; Saviozzi, M.; Silvestro, F.; Vinci, A. Electrical consumption forecasting in hospital facilities: An application case. *Energy Build.* **2015**, *103*, 261–270. [CrossRef]
14. Mariano-Hernández, D.; Hernández-Callejo, L.; García, F.S.; Duque-Perez, O.; Zorita-Lamadrid, A.L. A Review of Energy Consumption Forecasting in Smart Buildings: Methods, Input Variables, Forecasting Horizon and Metrics. *Appl. Sci.* **2020**, *10*, 8323. [CrossRef]
15. Zhao, H.-X.; Magoulès, F. A review on the prediction of building energy consumption. *Renew. Sustain. Energy Rev.* **2012**, *16*, 3586–3592. [CrossRef]
16. Foucquier, A.; Robert, S.; Suard, F.; Stéphan, L.; Jay, A. State of the art in building modelling and energy performances prediction: A review. *Renew. Sustain. Energy Rev.* **2013**, *23*, 272–288. [CrossRef]
17. Seyedzadeh, S.; Rahimian, F.P.; Glesk, I.; Roper, M. Machine learning for estimation of building energy consumption and performance: A review. *Vis. Eng.* **2018**, *6*, 5. [CrossRef]
18. Sadeghi, A.; Sinaki, R.Y.; Young, W.A.; Weckman, G.R. An Intelligent Model to Predict Energy Performances of Residential Buildings Based on Deep Neural Networks. *Energies* **2020**, *13*, 571. [CrossRef]
19. Korkidis, P.; Dounis, A.; Kofinas, P. Computational Intelligence Technologies for Occupancy Estimation and Comfort Control in Buildings. *Energies* **2021**, *14*, 4971. [CrossRef]
20. Cao, L.; Li, Y.; Zhang, J.; Jiang, Y.; Han, Y.; Wei, J. Electrical load prediction of healthcare buildings through single and ensemble learning. *Energy Rep.* **2020**, *6*, 2751–2767. [CrossRef]

21. Mosavi, A.; Salimi, M.; Ardabili, S.F.; Rabczuk, T.; Shamshirband, S.; Varkonyi-Koczy, A.R. State of the Art of Machine Learning Models in Energy Systems, a Systematic Review. *Energies* **2019**, *12*, 1301. [CrossRef]
22. Hill, T.; Marquez, L.; O'Connor, M.; Remus, W. Artificial neural network models for forecasting and decision making. *Int. J. Forecast.* **1994**, *10*, 5–15. [CrossRef]
23. Wang, X.; Fang, F.; Zhang, X.; Liu, Y.; Wei, L.; Shi, Y. LSTM-based Short-term Load Forecasting for Building Electricity Consumption. In Proceedings of the 28th IEEE International Symposium on Industrial Electronics, ISIE 2019, Vancouver, BC, Canada, 12–14 June 2019. [CrossRef]
24. Ghenai, C.; Al-Mufti, O.A.A.; Al-Isawi, O.A.M.; Amirah, L.H.L.; Merabet, A. Short-term building electrical load forecasting using adaptive neuro-fuzzy inference system (ANFIS). *J. Build. Eng.* **2022**, *52*, 104323. [CrossRef]
25. Si, T.; Bagchi, J.; Miranda, P.B. Artificial Neural Network training using metaheuristics for medical data classification: An experimental study. *Expert Syst. Appl.* **2022**, *193*, 116423. [CrossRef]
26. Chenglei, H.; Kangji, L.; Guohai, L.; Lei, P. Forecasting building energy consumption based on hybrid PSO-ANN prediction model. In Proceedings of the 2015 34th Chinese Control Conference (CCC), Hangzhou, China, 28–30 July 2015. [CrossRef]
27. Wahid, F.; Fayaz, M.; AlJarbouh, A.; Mir, M.; Aamir, M. Imran Energy Consumption Optimization and User Comfort Maximization in Smart Buildings Using a Hybrid of the Firefly and Genetic Algorithms. *Energies* **2020**, *13*, 4363. [CrossRef]
28. Ruiz, E.; Pacheco-Torres, R.; Casillas, J. Energy consumption modeling by machine learning from daily activity metering in a hospital. In Proceedings of the 2017 22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), Limassol, Cyprus, 12–15 September 2017. [CrossRef]
29. Runge, J.; Zmeureanu, R. Forecasting Energy Use in Buildings Using Artificial Neural Networks: A Review. *Energies* **2019**, *12*, 3254. [CrossRef]
30. Muhammad, W.; Zhenzhi, L.; Shengyuan, L.; Zhang, J.; Mian, R.; Intisar, A.S. Optimal BRA based electric demand prediction strategy considering instance-based learning of the forecast factors. *Int. Trans. Electr. Energy Syst.* **2021**, *31*, e12967. [CrossRef]
31. Lourakis, M.I.A.; Argyros, A.A. Is Levenberg-Marquardt the Most Efficient Optimization Algorithm for Implementing Bundle Adjustment? In Proceedings of the 10th IEEE International Conference on Computer Vision (ICCV 2005). Beijing, China, 17–20 October 2005. [CrossRef]
32. Ruiz, L.G.B.; Cuéllar, M.P.; Calvo-Flores, M.D.; Jiménez, M.D.C.P. An Application of Non-Linear Autoregressive Neural Networks to Predict Energy Consumption in Public Buildings. *Energies* **2016**, *9*, 684. [CrossRef]
33. Møller, M.F. A scaled conjugate gradient algorithm for fast supervised learning. *Neural Netw.* **1993**, *6*, 525–533. [CrossRef]
34. Khosravani, H.R.; Castilla, M.D.M.; Berenguel, M.; Ruano, A.E.; Ferreira, P.M. A Comparison of Energy Consumption Prediction Models Based on Neural Networks of a Bioclimatic Building. *Energies* **2016**, *9*, 57. [CrossRef]
35. Montana, D.J.; Davis, L. Training feedforward neural networks using genetic algorithms. In Proceedings of the 11th International Joint Conference on Artificial Intelligence, IJCAI-89, Detroit, MI, USA, 20–25 August 1989; Volume 1.
36. Kennedy, J.; Eberhart, R. Particle swarm optimization. In Proceedings of the ICNN'95—International Conference on Neural Networks, Perth, Australia, 27 November 1995. [CrossRef]
37. Gudise, V.G.; Venayagamoorthy, G.K. Comparison of particle swarm optimization and backpropagation as training algorithms for neural networks. In Proceedings of the 2003 IEEE Swarm Intelligence Symposium, SIS'03, (Cat. No.03EX706). Indianapolis, IN, USA, 24–26 April 2003. [CrossRef]
38. Le, L.T.; Nguyen, H.; Dou, J.; Zhou, J. A Comparative Study of PSO-ANN, GA-ANN, ICA-ANN, and ABC-ANN in Estimating the Heating Load of Buildings' Energy Efficiency for Smart City Planning. *Appl. Sci.* **2019**, *9*, 2630. [CrossRef]
39. Heris, M.K. Particle Swarm Optimization in MATLAB. *Yarpiz*. 2015. Available online: https://yarpiz.com/50/ypea102-particle-swarm-optimization (accessed on 21 July 2022).
40. Erol, O.; Eksin, I. A new optimization method: Big Bang-Big Crunch. *Adv. Eng. Softw.* **2006**, *37*, 106–111. [CrossRef]
41. Wang, J.; Kumbasar, T. Parameter optimization of interval Type-2 fuzzy neural networks based on PSO and BBBC methods. *IEEE/CAA J. Autom. Sin.* **2019**, *6*, 247–257. [CrossRef]
42. Jang, J.S. ANFIS: Adaptive-Network-Based Fuzzy Inference System. *IEEE Trans. Syst. Man Cybern. Syst.* **1993**, *23*, 665–685. [CrossRef]
43. Kontogiannis, D.; Bargiotas, D.; Daskalopulu, A. Fuzzy Control System for Smart Energy Management in Residential Buildings Based on Environmental Data. *Energies* **2021**, *14*, 752. [CrossRef]
44. Heris, M.K. Time-Series Prediction using ANFIS in MATLAB. 2015. Available online: https://yarpiz.com/327/ypfz102-time-series-prediction-using-anfis (accessed on 21 July 2022).
45. Xie, C.; Wang, D.; Wu, H.; Gao, L. A long short-term memory neural network model for knee joint acceleration estimation using mechanomyography signals. *Int. J. Adv. Robot. Syst.* **2020**, 17. [CrossRef]
46. Pek, J.; Wong, A.C.M.; Wong, O.C.Y. Confidence Intervals for the Mean of Non-Normal Distribution: Transform or Not to Transform. *Open J. Stat.* **2017**, *07*, 405–421. [CrossRef]
47. Myatt, G.J. *Making Sense of Data-A Practical Guide to Exploratory Data Analysis and Data Mining*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2007.