Article

An Intelligent Data-Driven Approach for Electrical Energy Load Management Using Machine Learning Algorithms

Shamim Akhtar [1,*] , **Muhamad Zahim Bin Sujod [1]** and **Syed Sajjad Hussain Rizvi [2]**

[1] Department of Electrical Engineering, College of Engineering, Universiti Malaysia Pahang, Kuantan 26300, Pahang, Malaysia
[2] Shaheed Zulfikar Ali Bhutto Institute of Science and Technology, Karachi 75600, Pakistan
* Correspondence: shamimakhtar1@gmail.com

Abstract: Data-driven electrical energy efficiency management is the emerging trend in electrical energy forecasting and management. This fusion of data science, artificial intelligence, and electrical energy management has turned out to be the most precise and robust energy management solution. The Smart Energy Informatics Lab (SEIL) of the Indian Institute of Technology (IIT) conducted an experimental study in 2019 to collect massive data on university campus energy consumption. The comprehensive comparative study preparatory to the recommendation of the best candidate out of 24 machine learning algorithms on the SEIL dataset is presented in this work. In this research work, an exhaustive parametric and empirical comparative study is conducted on the SEIL dataset for the recommendation of the optimal machine learning algorithm. The simulation results established the findings that Bagged Trees, Fine Trees, and Medium Trees are, respectively, the best-, second-best-, and third-best-performing algorithms in terms of efficacy. On the contrary, a reverse ranking is observed in terms of efficiency. This is grounded in the fact that Bagged Trees is most effective algorithm for the said application and Medium Trees is the most efficient one. Likewise, Fine Trees has the optimum tradeoff between efficacy and efficiency.

Keywords: data driven energy efficiency management; machine learning; SEIL dataset; artificial intelligence; energy forecasting

1. Introduction

Electrical energy is an essential and likely increasingly scarce resource around the globe. Its scarcity could be handled in the following bifold manner. The first solution is to increase generation capacity, and the second is to improve load management and energy demand forecasting [1]. The scope of this study is mainly focused on electrical energy management and demand forecasting. This domain is well represented in the literature and a large number of researchers have submitted their contributions. The research has established that the major venues of energy consumption are buildings. This includes such things as residential buildings, commercial buildings, and industrial buildings [2]. It is also reported in the literature that building energy consumption accounts for 39% of global energy consumption and 38% of greenhouse gas emissions [3]. The SEIL published a study in 2019 endeavoring to collect massive amounts of data on the energy consumption of residential buildings and university campuses. Both datasets are reported as the most recent and benchmark dataset of data-driven energy forecasting systems considering residential buildings and university campuses [4]. In this research study, a university campus energy consumption dataset is under consideration.

It has been observed that large-scale university campuses demand a significant amount of electricity to fulfill their HVAC requirements. This includes, but is not limited to, the power requirements of lab equipment, machines, office equipment, classes, auditorium IT, and support equipment [5]. However, on the other hand, electrical energy is a constrained

resource. Electrical energy management systems (EEMS) are one of the most widely advocated solutions for achieving the goal of reconciling electricity demand with limited electricity resources. Furthermore, these systems contribute to a significant reduction in electricity consumption bills [6]. Electrical energy demand forecasting is the essential and integral part of the EEMS. It aims to manage, monitor, optimize, and analyze the day-to-day electricity demand of a specific area [7]. Therefore, the efficiency and efficacy of the EEMS is highly correlated with the performance of the electrical forecasting system [8]. Accurate, robust, adaptive, and efficient energy forecasting promises efficiency in the energy management system.

Moreover, an efficient electrical energy forecasting system complements other energy management policies, making for optimum energy consumption [9]. This eventually turns into a competitive advantage in particular and sustainable development in general. Researchers have recently strongly advocated for a data-driven approach for robust, adaptive, and efficient energy forecasting systems [10]. In the literature, this is technically termed the "black box approach". On the other hand, engineering or "white box" and statistical or "grey box" approaches are already in place [11]. In the data-driven, or "black box" approach, the artificial intelligence (machine learning or deep learning) model is trained and customized for the specific application. Since the training of the model is purely based on the data provided to it, it is logically termed a data-driven approach [12]. The recent literature is enriched by many applications of various machine learning algorithms for data-driven energy efficiency management. This includes, but is not limited to, probabilistic modeling [13], Artificial Neural Network, Random Forest [14], Regression [15] and many more machine learning algorithms that are trained for the respective application's dataset.

It has been observed that many datasets for building energy management are available. Likewise, different researchers have employed different machine learning algorithms for classification and prediction on their respective datasets. However, the need for a benchmark dataset was felt in the literature. Moreover, a unified application-oriented machine learning algorithm is also a pressing need. The SEIL then conducted a study in the year 2019 to collect a massive amount of data on the energy consumption of residential buildings and university campuses. In this research, a university campus energy consumption dataset is under consideration. This work also submits an exhaustive parametric and empirical study of machine learning algorithms on the relevant SEIL dataset (University Campus). Finally, the recommendation of the optimal machine learning algorithm for university campus energy demand forecasting is submitted.

The organization of the manuscript is as follows: Section 2 presents a Literature Review of the recent studies in the domain. The research gap is highlighted in Section 3. Sections 4 and 5, respectively, show the system setup and methodology of the proposed study. Sections 6 and 7 comprise the results and discussions. Finally, Section 8 presents the conclusion of the study.

2. Literature Review

This section will provide an in-depth review of the literature on the application of machine learning algorithms to energy predictions. Primarily, the scope of this literature review is bifold. First, the performance evaluation of various machine learning algorithms for electrical energy forecasting is considered. This establishes the logical rationale for the utility of machine learning algorithm energy forecasting. Secondly, the existing work on the benchmark dataset of SEIL University Campus is presented. This clarifies the research gap in the SEIL University Campus dataset.

In 2019, Johannesen et al. [16] presented a study to investigate the response of a regression model to a Sydney dataset. This dataset comprised weather, time stamps, and load demand. The dataset was locally gathered for four years. The research has diffused/mapped the time stamp, weather, and load demand data as common instances. Finally, the authors employed the Random Forest Regressor, k-Nearest Neighbor Regressor and Linear Regression for load forecasting. The authors presented a conclusion that for

the Sydney region dataset of electrical consumption, Random Forest is the best candidate for short-term (30 min) prediction. Likewise, the kNN Regression Tool is found to be relatively efficient for long term prediction (24 h) [17]. Another study was conducted on the power consumption of higher education university campuses in Korea. In this study, the authors also mapped the weather data and power consumption data. The authors in this study reduced the features dimension using principal component analysis (PCA) and then employed ANN and SVM for energy demand prediction. The authors then concluded that for the said dataset, the ANN is found to be the better candidate for energy demand prediction [18].

In parallel to the university campus, residential buildings are similar venues for energy consumption. Chou et al. [19] have presented a study on energy demand forecasting for residential buildings. In particular, the authors presented a hybrid model of prediction and optimization. They reported that the hybrid evolutionary-neuro system is found to be better as compared to the classical machine learning network for their respective datasets. The dataset was locally gathered from a residential building. In the same year, another group of authors also presented a study to rationalize the evolutionary-neuro system for energy forecasting. In this study, the authors have presented a hybrid model of evolutionary algorithms, i.e., teaching learning-based optimization (TLBO) and ANN. They also advocated the efficacy of the hybrid model for their respective datasets [20]. In another study, its author has presented a hybrid model for optimization and energy demand forecasting. This approach was tested on a dataset of South Korea's hourly energy consumption. The author claimed that the proposed model could be useful for other datasets. However, a limited rationale of this claim is set forth in their paper [21]. Ahmad et al. [22] have published a survey in *Sustainable Cities and Society*. In this survey, the authors have presented a comprehensive comparison through a literature survey about the efficacy of machine learning algorithms for energy demand forecasting. They have concluded that the Bayesian regularization back-propagation neural networks (BRB-NNs) and the Levenberg-Marquardt back-propagation neural networks (LMBNNs) are found to be relatively efficient predictors for electrical energy demand forecasting.

The second fold of this literature review is related to the existing investigation into the benchmark dataset of SEIL. In this section, the research contribution of the SEIL University Campus dataset is presented. In addition, the research gap is highlighted in the automatic learning application on electrical power prediction and existing work on the SEIL. A group of SEIL researchers used LSTM and improved the algorithm for optimizing the sinusoidal cosine. It translates into accurate and reliable power consumption predictions for short, medium and long-term forecasts. They argued that the hybridization of the enhanced Sine Cosine and LSTM algorithms has transformed into a robust power consumption model [4]. In a separate publication, SEIL researchers used kCNN-LSTM to provide accurate predictions of energy consumption in buildings. This experiment was based on real-time energy consumption data from the Kanwal Rekhi building, an academic building of the Indian Institute of Technology (IIT), Mumbai. The proposed approach uses k means clustering to conduct cluster analysis and understand the energy usage model. The proposed approaches were formed and tested using real-time energy consumption data from a four-floor building at IIT Bombay, India [4,23,24]. In-addition to the above recent work, Table 1 shows a literature matrix to highlight recent work of energy management using artificial intelligence with the title, proposed work, and the corresponding limitations of the proposed work.

Table 1. Literature map on energy management using artificial intelligence.

Ref	Year	Proposed Work	Limitations
[25]	2018	In this literature, the large data is called Big Data for energy management.	The availability of the referenced massive datasets is limited.
[26]	2019	Deep learning approaches have outdone themselves in dealing with big data.	There is the challenge of managing the large data package.
[27]	2019	The authors have strongly argued for the usefulness of Deep Learning frameworks to design the electrical energy efficiency management system.	Huge standard reference limits set of data for electric energy.
[28]	2016		
[29]	2016	This is a comparative consideration of two varieties of the Deep Learning network: (LSTM) and sequence architecture (S2 S).	In this study, coverage is limited to a single residential customer.
[30]	2016	They then compared the proposed model with the regression of existing support vector and Deep Learning frameworks. The result of the simulation shows that the local RVS surpassed the RVS and H2 O in-Deep Learning.	The authors stated that they used data on submissions; however, the details of the data were not included in this document.
[31]	2016	This paper presented the Factored Conditional Limited Boltzmann Machine (FCRBM) to forecast energy demands. The model has been tested on the EcoGrid EU data set.	The author of this paper needed to compare his research study with other variations of Deep Learning architecture and currently performed systems.
[32]	2017	The authors have compared the convolutional neural network (CNN/ConvNet) with the study presented in 2016.	No new CNN/ConvNet architecture was presented in this study, and neither was the pre-trained network described.
[33]	2017	Initially formed the Recurrent neural networks (RNN) using a data-driven approach.	This model-less, evidence-based approach has surpassed the approach of model-based studies in management of energy.
[34]	2017	This study provided a comprehensive comparison of conventional machine learning algorithms, including vector support machines, Gaussian processes, regression trees, overall amplification and linear regression, and the Deep Learning method.	Validation of the claim related to the energy management system is observed to be unclear.
[35]	2018	This method is optimized for building energy management, and explores two DL algorithms, namely, Deep Q-learning (DQN) and Deep Policy Gradient (DPG), at the same time.	The parametric fringe of the proposed technique proved insufficient. Moreover, the cognitive scope of the gadget seems to be very trendy.
[36]	2018	The authors have used Recurrent neural networks (RNN) to forecast time series data of energy consumption for a university campus.	The robustness of this work could be enhanced if the master data set had been chosen.
[37]	2019	The authors have suggested using the methods of alternating direction of multipliers (ADMM) and accelerated alternating direction of multipliers (AADM) to find the optimum value of operation of the micro network distribution.	This study did not include the data-based approach to energy forecasting. Moreover, it was felt that the parametric comparison was missing in this work.
[38]	2019	This work submitted a data-driven, Deep Learning approach to district-wide energy demand forecasting.	This study appears deficient because of the absence of a baseline data set and extensive comparison with pre-existing models.
[39]	2020	The FS-FCRBM-GWDO hybrid model is superior to the existing models presented in this study.	The gap between the existing real-world reference data set and the pre-established model is deficient.
[40]	2020	Major contributions include device-based real-time power management via a common cloud data monitoring server.	The actual application was outside the scope of study.

Table 1. *Cont*.

Ref	Year	Proposed Work	Limitations
[41]	2020	The authors have used a convolutional neural network (CNN) and a multilayer bi-directional synchronized recurrent unit (MB-GRU) to predict the consumption load of a residential building.	The degree of contribution to research could increase if the updated architecture of Deep Learning could be introduced. Moreover, the systematic evaluation of the novel model may be established in comparison to the existing pre-trained Deep Learning architectures.
[42]	2021	A machine (computer) based vision approach, You Only Look Once (YOLO v3), was utilized to calculate the number of individuals within the region. It is more in correlation with the temperature range of the air conditioning units.	The author's study would be made more effective by implementing new variants of Deep Learning.
[43]	2021	A new machine learning model for forecasting the energy usage on an hourly basis in a residential building is proposed.	The performance of the machine learning algorithm is compromised as a result of the performance plateau highlighted with big data.
[44]	2021	Deep Learning is the best candidate for power prediction based on time series. The concern in this study is increasingly associated with the fact that the performance of ML and DL is found.	Three small data sets were used to validate the study. the authors could not pursue the novel Deep Learning architecture of a machine learning model for data-driven energy efficiency forecasting.
[45]	2020	Comparison of Machine Learning and DL algorithms with the residential building dataset. The comparison was based on competition between ML and DL.	These works also highlight the urgent need for a master data set for domain-specific applications such as hospitals, schools, universities, residential buildings, etc.
[46]	2020		
[47]	2020		
[48]	2019	The work continued to use the RNN with the LSTM approach to forecast energy demand. This work also suggested a gap in the development of any new DL architecture and a pre-formed network of master datasets.	This work also suggests a gap in the development of new ML on master datasets.
[49]	2019	Smart Energy Informatics Lab (SEIL) offered data-driven reinforcement learning for predicting demand.	
[50]	2020	SEIL-IIT introduced the database medium for adaptive data visualization of large sensors.	
[51]	2020	Smart Energy Informatics Lab (SEIL), the same group, studied the hybrid model for predicting energy consumption in buildings through LSTM networks.	
[52]	2017	There has been a push for a data-driven approach to the intelligent energy management system.	
[53]	2019	In this work reference is made to the data set generated by their research. Another area that this group has targeted is solar photovoltaic optimisation and building thermal modelling [53–55]. This is beyond the scope of our study.	The Pre-Trained model is deficient.
[54]	2019		
[55]	2018		

3. Gap Analysis

After careful analysis of the existing work in the domain of data-driven energy management, it has been determined that the utilization of artificial intelligence is now inevitable for robust and precise energy management. In this connection, the benchmarking of the domain specific dataset is an essential need. Moreover, the development of robust machine learning algorithms will facilitate the objective. After detailed analysis, the SEIL dataset is found to be the most suitable dataset for energy forecasting for a university campus. However, the comprehensive empirical comparison of machine learning algorithms in the

literature is found to be deficient. In this study, the gap will be bridged by submitting an exhaustive evaluation of large number of machine learning algorithms on the SEIL dataset. The key deliverable of this study is the recommendation of the best machine learning algorithm. The recommendation will be based on empirical facts and figures.

4. Dataset Description and System Setup

The SEIL dataset was gathered from an IIT university building. The building has four floors and is divided into three wings. The dataset includes data from December 2016 through July 2018. All datasets are in CSV format. The datasets are all at one-minute granularity with current, voltage, and power as input attributes, and real energy consumption as an output attribute. The dataset is massive, with a volume of 20 GB. Data has been extracted from various units in the university building, such as building level, class level, auditorium level, lab level, office level, etc. In this study, the building-level data is taken into account in order to predict the total energy consumption of the building. Table 2 illustrates the list of attributes of the dataset. Since the dataset is labeled and continuous, the machine learning prediction algorithms have been selected for training and testing. A total of 24 machine learning prediction algorithms were tested to determine the best machine learning algorithm. The grounds for the decision are the functions of RMSE, R-Squared, MSE, MAE, prediction speed, and computation time. Please refer to Table 2. Attributes 1 to 23 served as the input attributes and attribute number 24 is taken as the output attribute. The power meter records the phase voltage (V1, V2, and V3) and the phase current (A1, A2, and A3). Attributes 8 to 10 describe the apparent power, whereas attributes 12 to 14 represent the real power consumed. The reactive power for each phase is shown in attributes 16 to 19, and the power factor for each phase is shown as PF1, PF2, and PF3 in attributes 20 to 22. The output real power in Wh is taken as the response variable. It is to be noted that in the dataset, the real power Wh is identical to the apparent and reactive power. Therefore, real power has been selected as the response variable.

Table 2. Attributes of the dataset.

Attribute Number	Attribute Symbol	Attribute Number	Attribute Symbol	Attribute Number	Attribute Symbol	Attribute Number	Attribute Symbol
1.	V1	8.	VA1	15.	W	22.	PF3
2.	V2	9.	VA2	16.	VAR1	23.	PF
3.	V3	10.	VA3	17.	VAR2	24.	FwdWh
4.	A1	11.	VA	18.	VAR3	25.	FwdVAh
5.	A2	12.	W1	19.	VAR	26.	FwdVARh, FwdVARhC
6.	A3	13.	W2	20.	PF1		
7.	A	14.	W3	21.	PF2		

5. Methodology

A quad-folded cascading methodology was used in this study. Figure 1 shows the pictorial illustration of the proposed methodology. In the first phase, the SEIL dataset is used, and the total energy consumption at the building level is considered. The building level energy consumption includes the auditorium, classroom, conference room, building floor, labs, offices, server room, and sub-server room. In the second phase, the building level dataset is first divided into 70% training samples and 30% testing samples with random permutation. The training set is used to train 24 machine learning algorithms. In the third phase, the parametric performance of each ML algorithm is evaluated as the function of training parameters like RMSE, R-squared, MSE, MAE, and Prediction Speed. Similarly, the same parameters for the testing phase are computed. Finally, the ranking of algorithms based on their efficacy and efficiency are established. Figure 2 illustrates the functional inside view of the training testing phase. The training and testing phase constitutes the inside workings of Phase 2 of the proposed methodology (Figure 1).
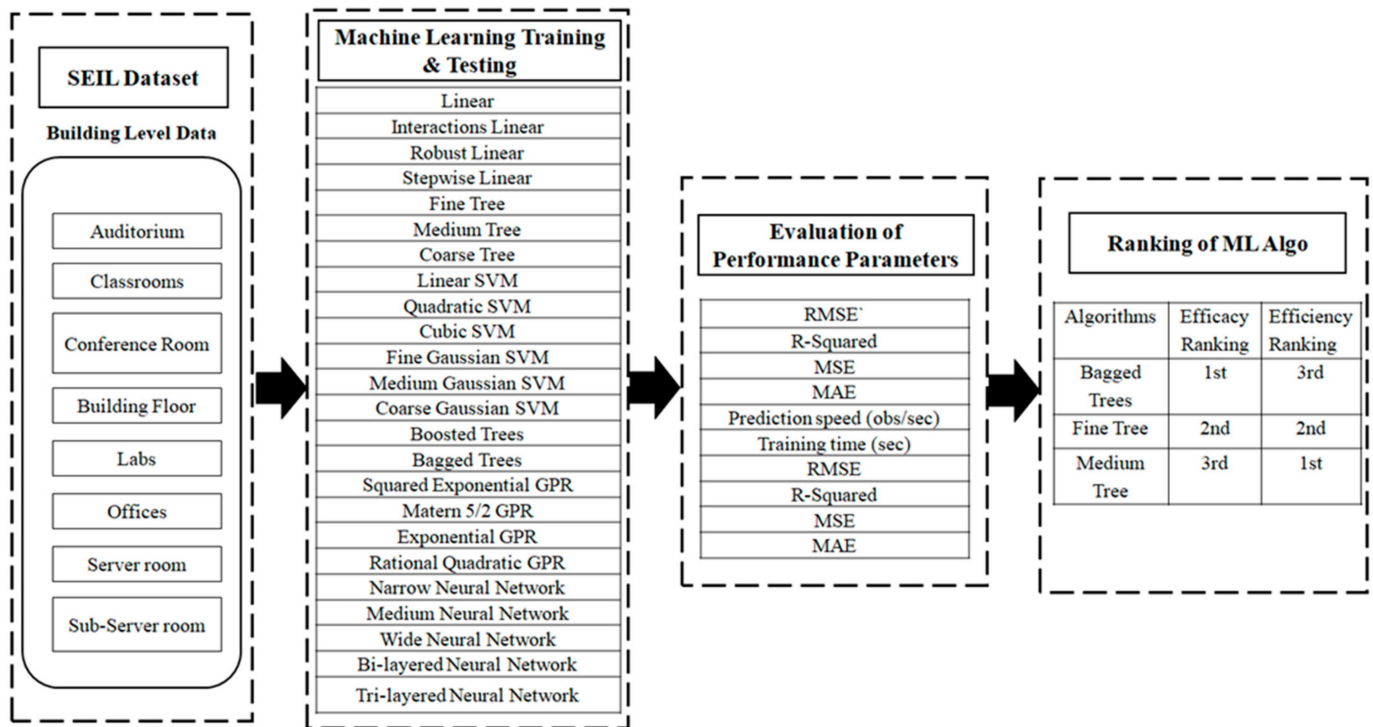
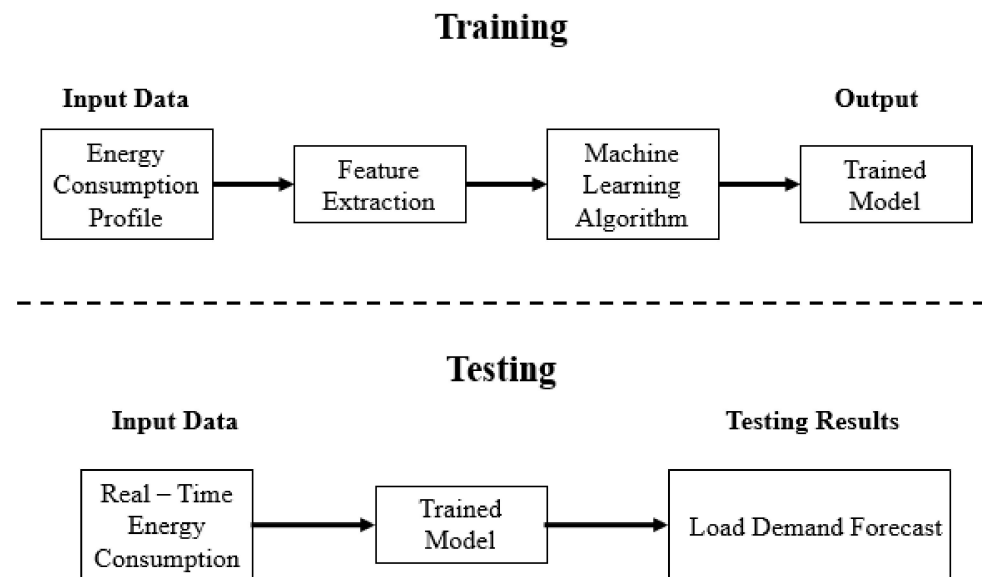**Figure 1.** Layout of proposed methodology.

Training

Testing

Figure 2. General framework of system training and testing [24].

6. Results

This section presents an exhaustive and comprehensive empirical evaluation of the best candidate of machine algorithm for energy demand prediction using the SEIL dataset. Figures 3–8 show the visual inferences of Table 3 for easy reference. Figures 9–20 illustrate the prediction vs. actual and residual of training and testing for each algorithm. In this study, 24 machine learning prediction algorithms were evaluated on benchmark performance parameters. In the predicted vs. actual graph, the x-axis shows the true response and the y-axis represents the predicted response. The approximate linearity of these curves is shown in black and the actual observation is depicted with blue dots. In the predicted vs. actual graph, the perceptual variance can be observed as the distance between the predicted value and the actual value. The larger variation corresponds to

the poor prediction performance of the respective algorithm. Moreover, the degree and measure of variation for the respective algorithm are shown in Table 3. The degree and measure of variation are the function of RMSE, R-Squared, MSE, and MAE for training and testing events. Likewise, prediction speed and computation time also establish grounds for the efficiency of the algorithm. The higher value of the error measure translates into the poor performance of the candidate algorithm. Figures 9–20 also include the Residual Error of training and testing for the top three performing algorithms. The x-axis of the residual curve refers to the predicted response, and the y-axis corresponds to the residual error. The close proximity of residual error to the predicted observation corresponds to the higher efficacy of the candidate algorithm. The inference is also correlated with the empirical and absolute values illustrated in Table 3. Foremost in this investigation, the *top three* performing machine learning algorithms for energy demand prediction at a university campus based on SEIL datasets are selected in general. Subsequently, an in-depth investigation of the performance parameter is performed. The graphical illustration and empirical findings have established that Bagged Trees (1), Fine Trees (2), and Medium Trees (3) are the *top three* performing algorithms in terms of efficacy. On the contrary, a reverse ranking is observed in terms of efficiency. This can also be inferred from Table 4. The performance measures such as RMSE, R-Squared, MSE, and MAE refer to the efficacy of the algorithm. Likewise, prediction speed and training time reflects the efficiency of the algorithm.
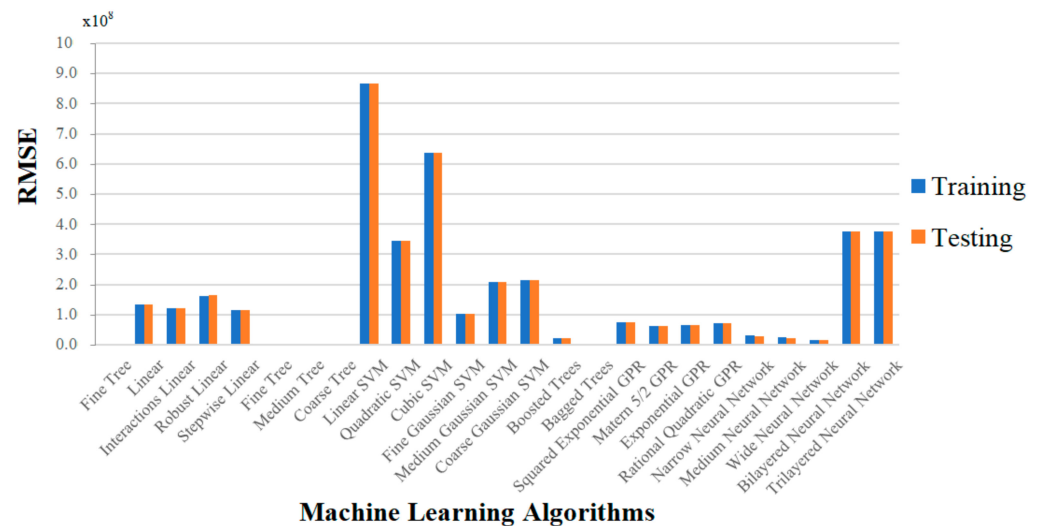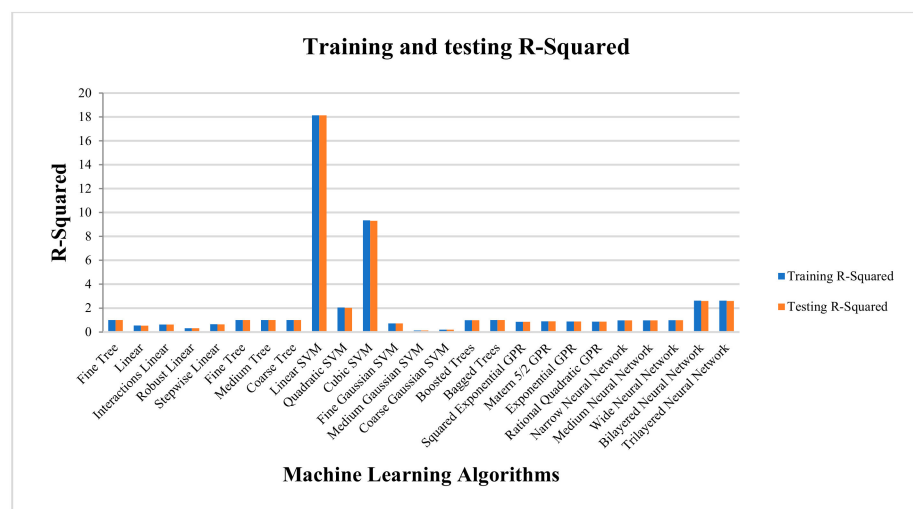

Figure 3. Training and testing RMSE.
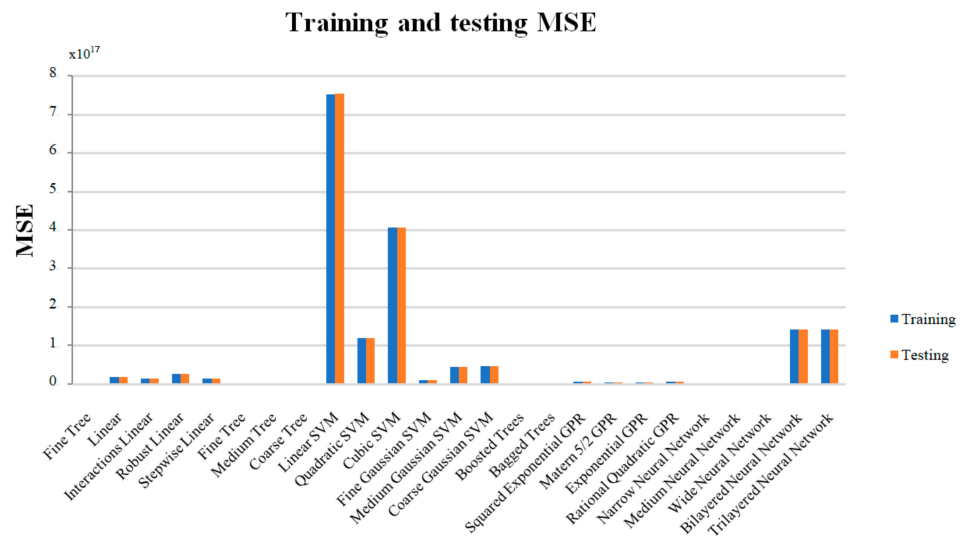

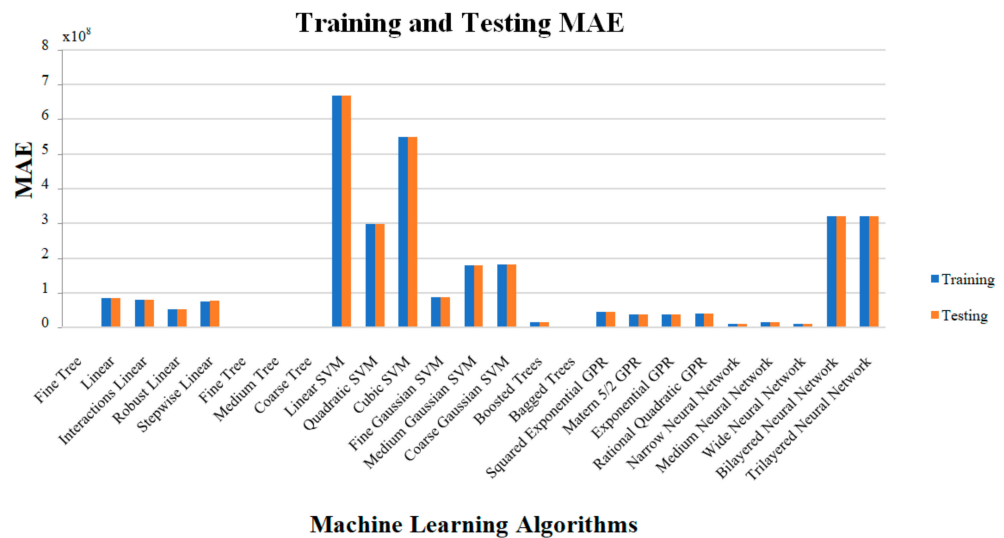
Figure 4. Training and testing R-Squared.

Training and testing MSE

Figure 5. Training and testing MSE.

Training and Testing MAE

Machine Learning Algorithms

Figure 6. Training and testing MAE.

Training Prediction speed (obs/sec)
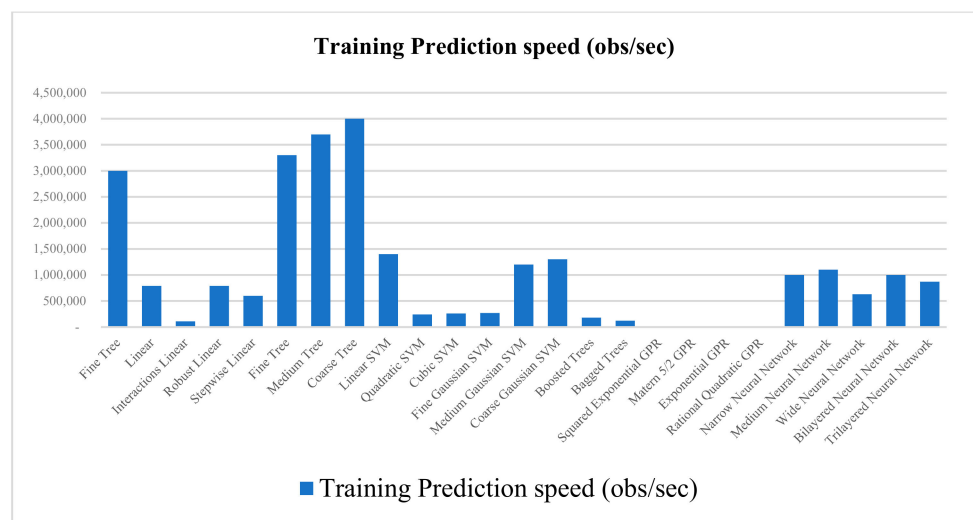

■ Training Prediction speed (obs/sec)
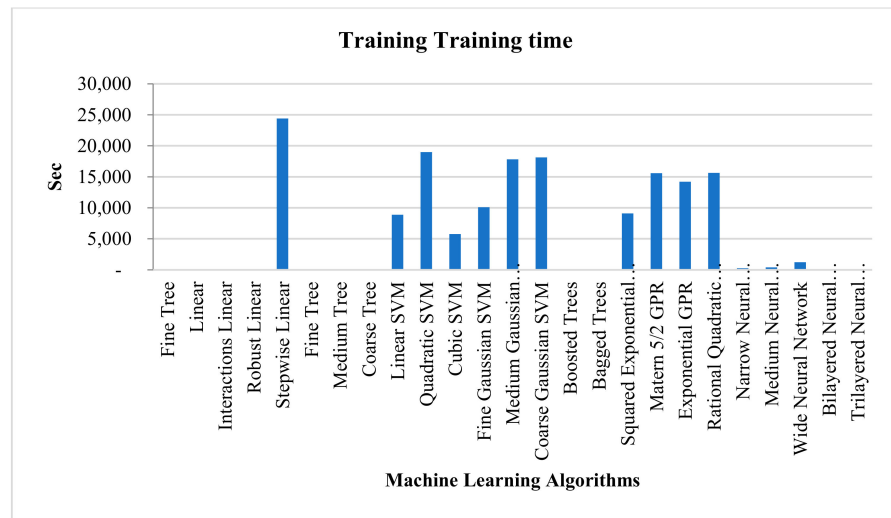
Figure 7. Prediction speed.
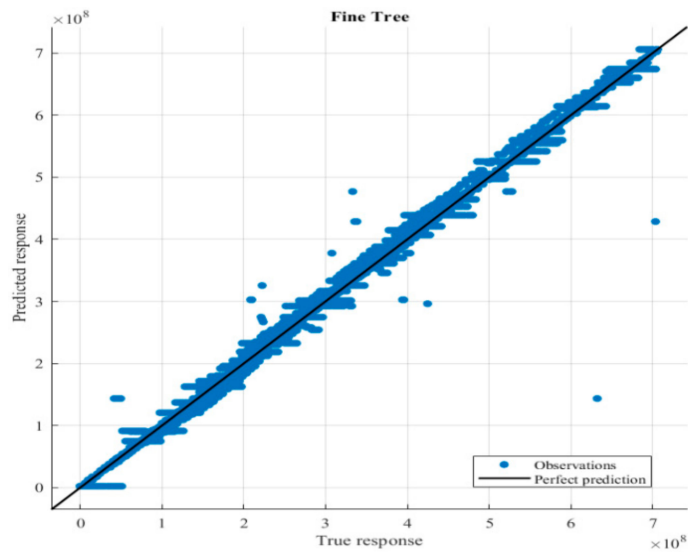
Figure 8. Training time.

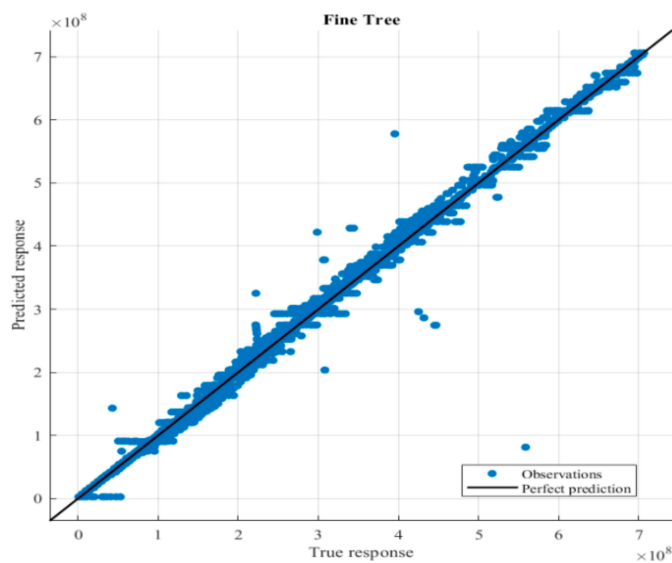**Figure 9.** Fine Trees Prediction vs. Actual training.

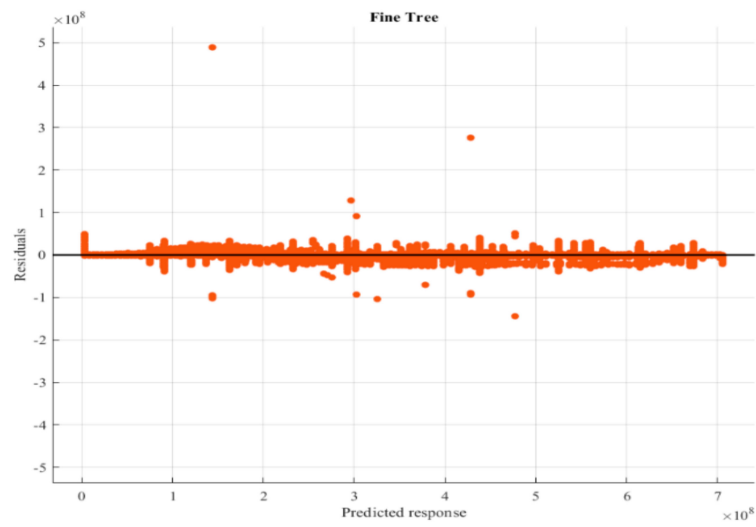**Figure 10.** Fine Trees Prediction vs. Actual testing.

Figure 11. Fine Trees Residual training.

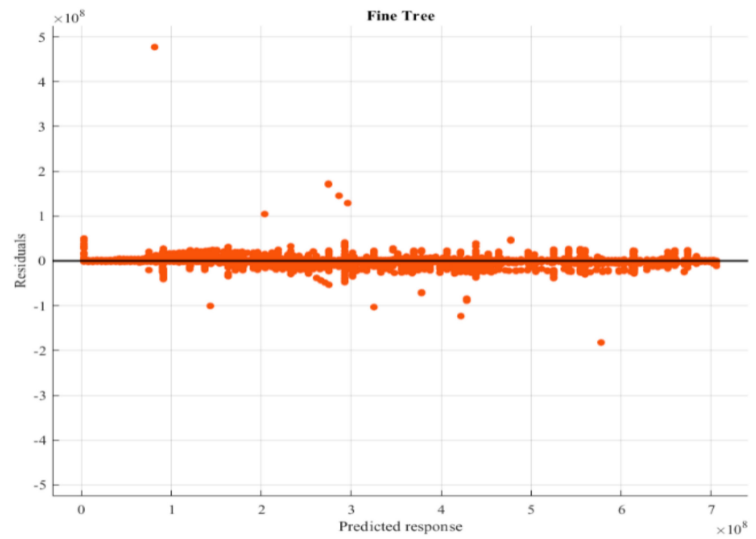**Figure 12.** Fine Trees Residual testing.

Figure 13. Medium Trees Prediction vs. Actual training.

Figure 14. Medium Trees Prediction vs. Actual testing.

Figure 15. Medium Trees Residual training.

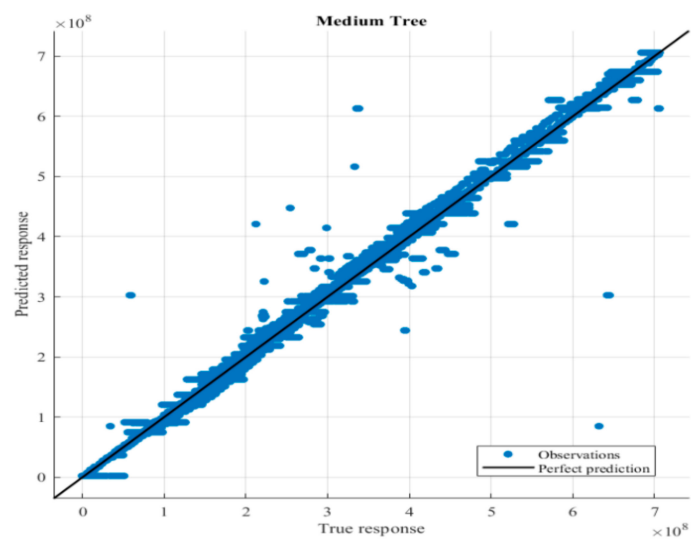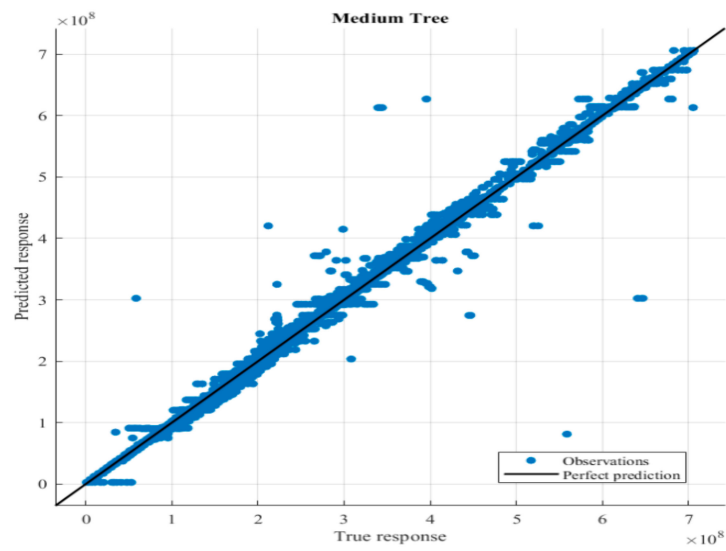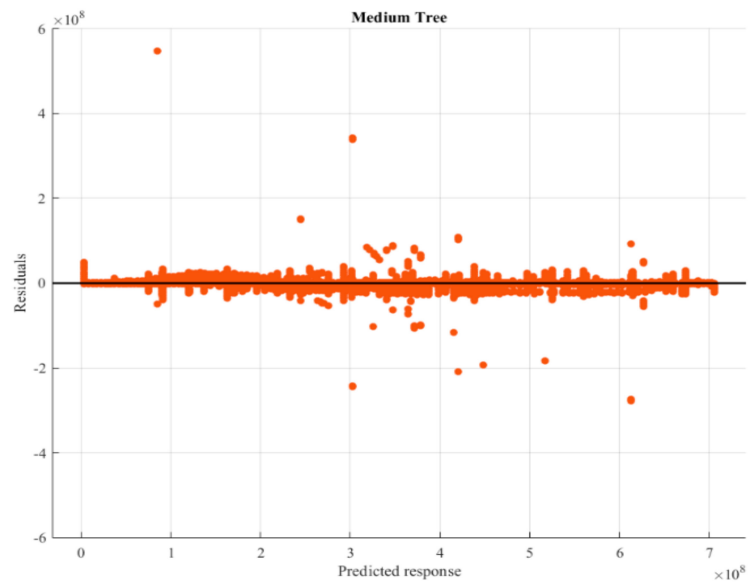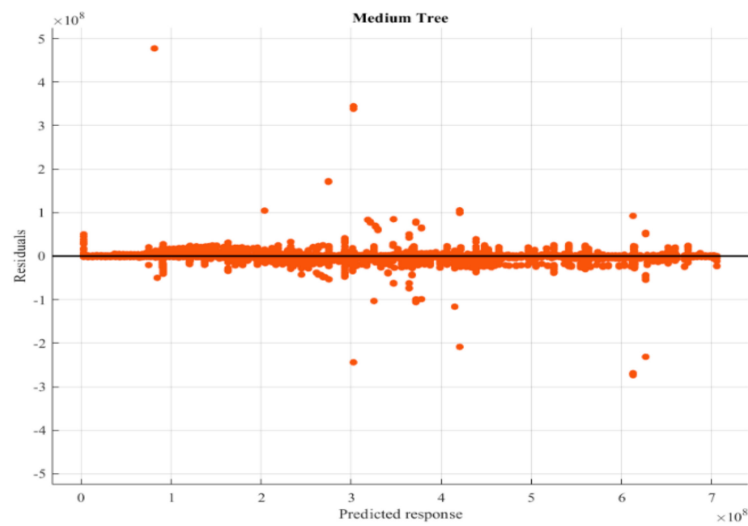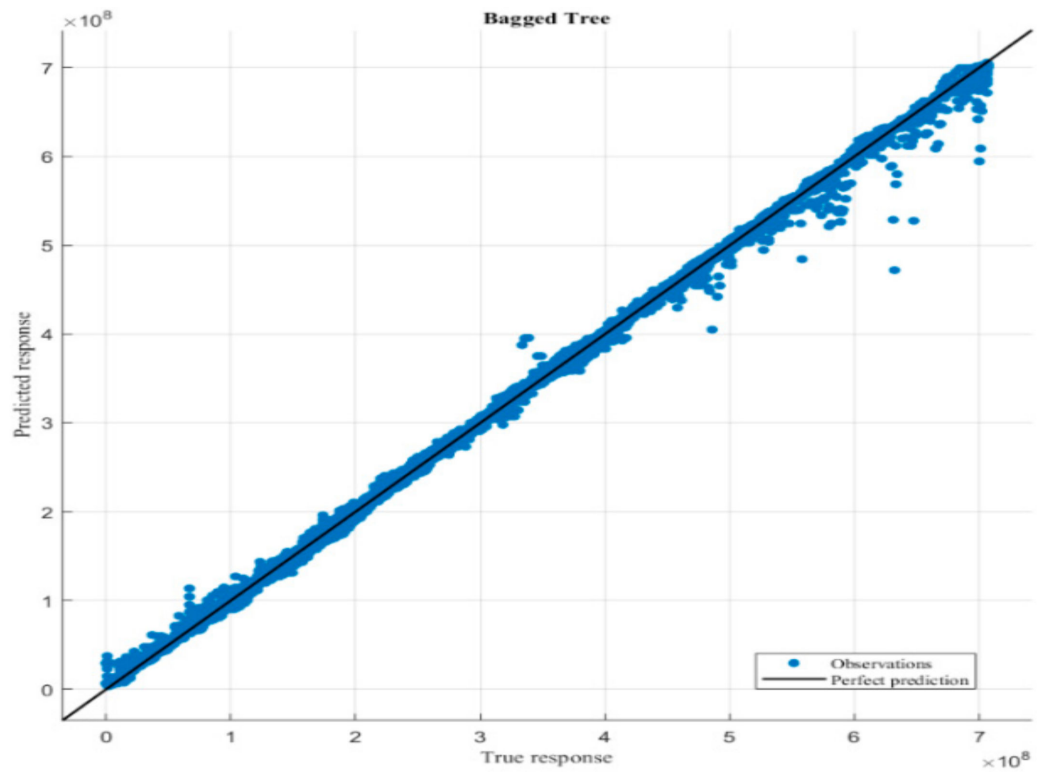**Figure 16.** Medium Trees Residual testing.

Figure 17. Bagged Trees Prediction vs. Actual training.
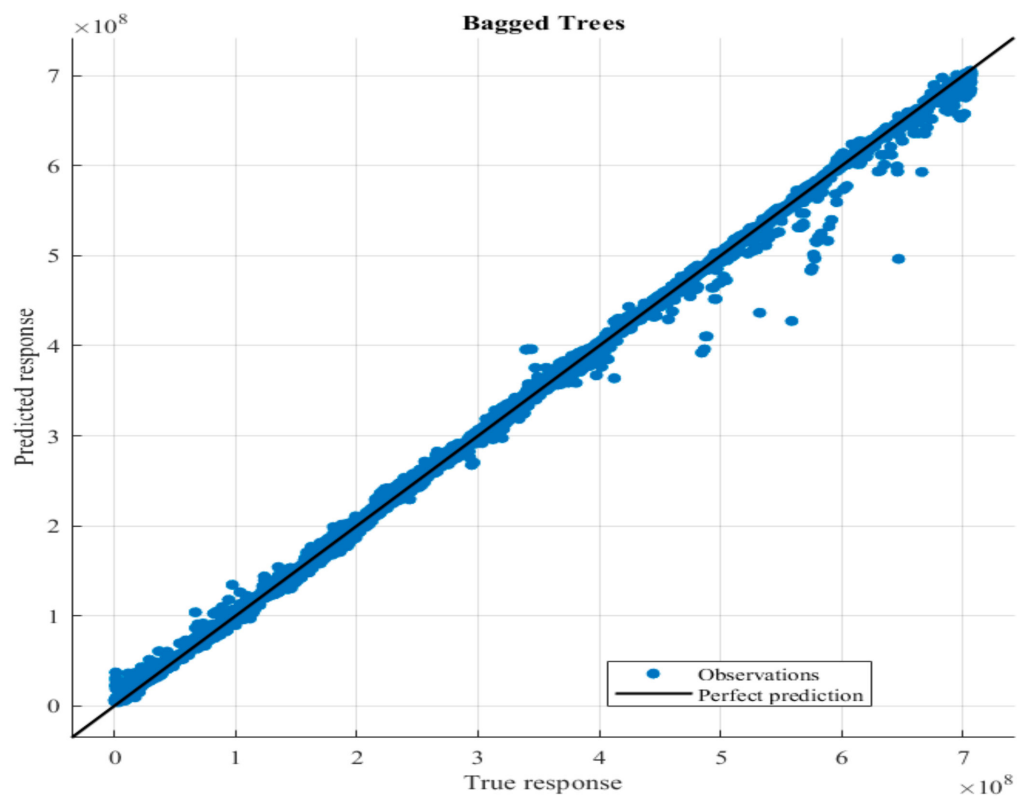

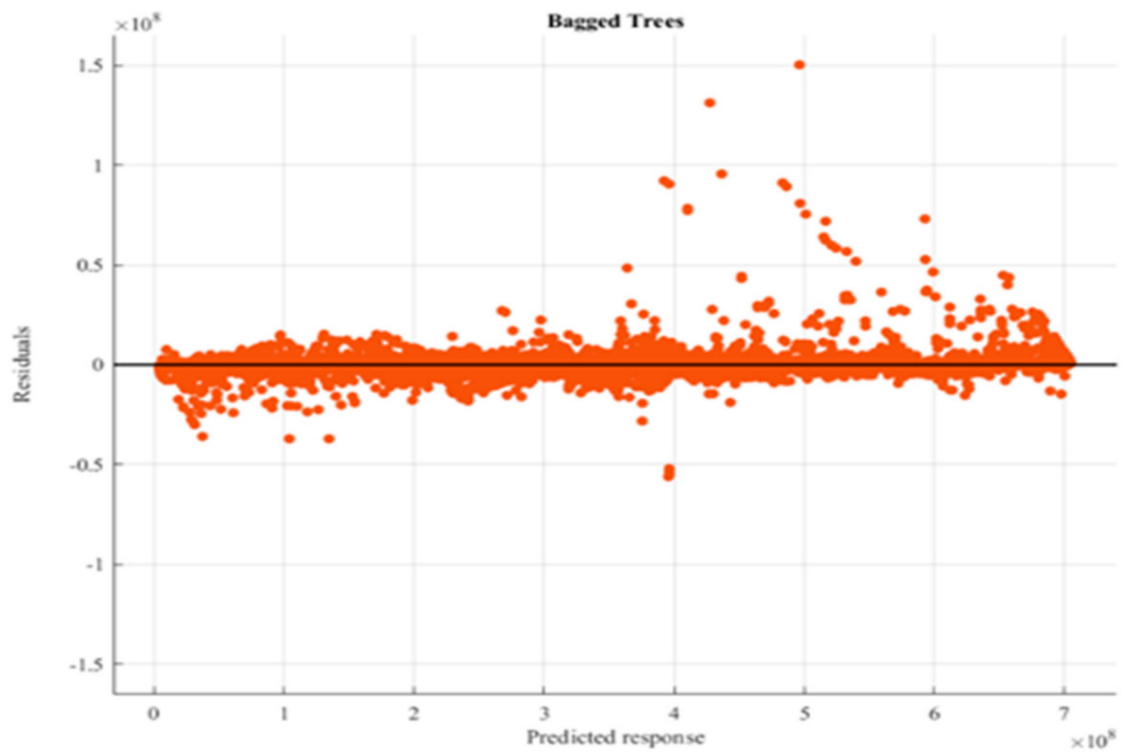**Figure 18.** Bagged Trees Prediction vs. Actual testing.
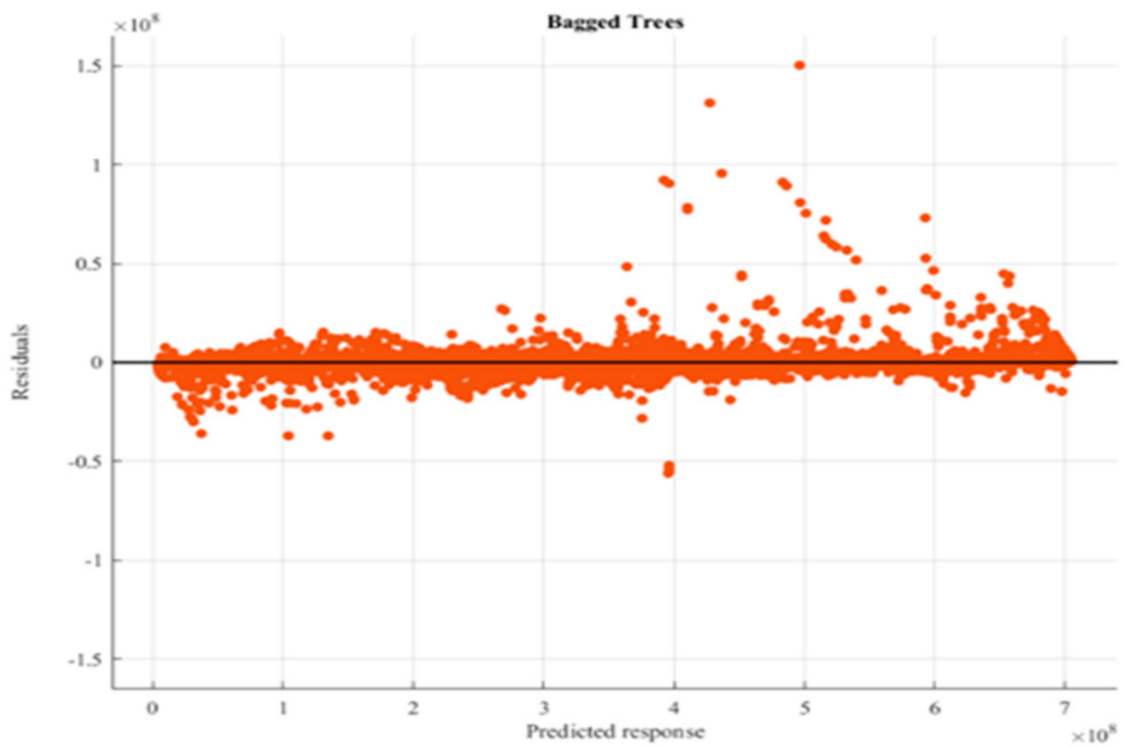
Figure 19. Bagged Trees Residual training.

Figure 20. Bagged Trees Residual testing.

Table 3. Performance evaluation of machine learning algorithm for energy prediction on SEIL dataset.

Algorithm	Training						Testing			
	RMSE	R-Squared	MSE	MAE	Prediction Speed (Obs/s)	Training Time (s)	RMSE	R-Squared	MSE	MAE
Linear	1.35×10^8	0.54	1.83×10^{16}	8.42×10^7	790,000	22.065	1.36×10^8	0.53	1.84×10^{16}	8.45×10^7
Interactions Linear	1.22×10^8	0.62	1.48×10^{16}	8.02×10^7	110,000	65.108	1.23×10^8	0.62	1.50×10^{16}	8.05×10^7
Robust Linear	1.64×10^8	0.32	2.68×10^{16}	5.25×10^7	790,000	19.44	1.65×10^8	0.31	2.71×10^{16}	5.30×10^7
Stepwise Linear	1.18×10^8	0.65	1.39×10^{16}	7.63×10^7	600,000	24,417	1.18×10^8	0.64	1.40×10^{16}	7.66×10^7
Fine Trees	2.11×10^6	1	4.44×10^{12}	1.40×10^6	3,300,000	8.3244	2.38×10^6	1	5.64×10^{12}	1.41×10^6
Medium Trees	2.81×10^6	1	7.87×10^{12}	1.42×10^6	3,700,000	8.0721	3.20×10^6	1	1.02×10^{13}	1.43×10^6
Coarse Tree	4.26×10^6	1	1.81×10^{13}	1.49×10^6	4,000,000	7.5477	4.63×10^6	1	2.15×10^{13}	1.50×10^6
Linear SVM	8.67×10^8	-18.13	7.52×10^{17}	6.69×10^8	1,400,000	8892.6	8.68×10^8	-18.13	7.54×10^{17}	6.69×10^8
Quadratic SVM	3.46×10^8	-2.05	1.20×10^{17}	2.98×10^8	240,000	18,985	3.45×10^8	-2.02	1.19×10^{17}	2.98×10^8
Cubic SVM	6.38×10^8	-9.35	4.07×10^{17}	5.50×10^8	260,000	5761.7	6.37×10^8	-9.31	4.06×10^{17}	5.50×10^8
Fine Gaussian SVM	1.04×10^8	0.72	1.08×10^{16}	8.83×10^7	270,000	10,099	1.04×10^8	0.72	1.08×10^{16}	8.85×10^7
Medium Gaussian SVM	2.10×10^8	-0.13	4.43×10^{16}	1.80×10^8	1,200,000	17,835	2.11×10^8	-0.13	4.44×10^{16}	1.80×10^8
Coarse Gaussian SVM	2.17×10^8	-0.19	4.69×10^{16}	1.82×10^8	1,300,000	18,127	2.17×10^8	-0.2	4.71×10^{16}	1.82×10^8
Boosted Trees	2.33×10^7	0.99	5.42×10^{14}	1.66×10^7	180,000	62.082	2.31×10^7	0.99	5.34×10^{14}	1.66×10^7
Bagged Trees	1.58×10^6	1	2.48×10^{12}	1.06×10^6	120,000	119.87	1.78×10^6	1	3.17×10^{12}	1.09×10^6
Squared E × ponential GPR	7.68×10^7	0.85	5.89×10^{15}	4.62×10^7	200	9102.8	7.66×10^7	0.85	5.87×10^{15}	4.62×10^7
Matern 5/2 GPR	6.40×10^7	0.9	4.10×10^{15}	3.87×10^7	110	15,589	6.40×10^7	0.9	4.10×10^{15}	3.88×10^7
E × ponential GPR	6.86×10^7	0.88	4.70×10^{15}	3.73×10^7	130	14,212	6.87×10^7	0.88	4.72×10^{15}	3.73×10^7
Rational Quadratic GPR	7.30×10^7	0.86	5.33×10^{15}	4.11×10^7	110	15,637	7.28×10^7	0.87	5.29×10^{15}	4.11×10^7
Narrow Neural Network	3.18×10^7	0.97	1.01×10^{15}	1.04×10^7	1,000,000	254.45	3.12×10^7	0.98	9.76×10^{14}	1.03×10^7
Medium Neural Network	2.54×10^7	0.98	6.46×10^{14}	1.46×10^7	1,100,000	396.17	2.51×10^7	0.98	6.32×10^{14}	1.46×10^7
Wide Neural Network	1.81×10^7	0.99	3.27×10^{14}	1.14×10^7	630,000	1238.8	1.79×10^7	0.99	3.19×10^{14}	1.14×10^7
Bi-layered Neural Network	3.77×10^8	-2.62	1.42×10^{17}	3.21×10^8	1,000,000	35.635	3.77×10^8	-2.6	1.42×10^{17}	3.20×10^8
Tri-layered Neural Network	3.77×10^8	-2.62	1.42×10^{17}	3.21×10^8	870,000	48.817	3.77×10^8	-2.6	1.42×10^{17}	3.20×10^8

Table 4. Efficacy vs. Efficiency ranking.

Algorithm	Efficacy Ranking	Efficiency Ranking
Bagged Trees	1st	3rd
Fine Trees	2nd	2nd
Medium Trees	3rd	1st

Table 4 reveals an intriguing fact: on the SEIL dataset, Bagged Trees is the most effective algorithm for said application, while Medium Trees is the most efficient algorithm for electrical energy demand prediction for university campuses. Likewise, Fine Trees have the optimum tradeoff between efficacy and efficiency. The Bagged Trees produces 75%, 56%, and 76% improvement in RMSE, MSE, and MAE, respectively, in training and testing, as compared to the Fine Trees. Similarly, the Bagged Trees produces 56%, 32%, and 75% improvement in RMSE, MSE, and MAE, respectively, in training and testing as compared to the Medium Trees algorithm. This metric shows the percentage of improvement in terms of efficacy between the algorithms. Similarly, in terms of efficiency improvement, the Medium Trees is 32 times more efficient in prediction speed and 14.8 times more efficient in training time, as compared to the Bagged Trees algorithm. Likewise, it is 1.3 times more efficient in prediction speed and 3 times more efficient in training time, as compared to the Fine Trees algorithm.

7. Discussion

Careful analysis of the simulation results presented in Section 6 establishes the rationale for the best candidate of machine learning algorithm for energy demand forecasting. To the best of the authors' knowledge, little work has been presented in the literature to date attempting to determine the best ML algorithm based on empirical facts. The systematic and empirical evaluation of a wide range of machine learning algorithms reveals that the Bagged Trees, Fine Trees, and Medium Trees algorithms are the top three ranked algorithms for energy demand forecasting using the SEIL dataset. This finding presents a knowledge add-on to the SEIL project consisting of the recommendation of the best machine learning algorithm for energy demand forecasting. Moreover, the new and customized algorithm is suggested to have further improvements in efficiency and efficacy. Based on this study, it is to be suggested that use of the novel and customized variant of the Medium Trees is strongly advised if efficiency is the primary goal. Similar to this, use of the innovative and customized Bagged Trees method is recommended if higher order efficacy is sought. The aforementioned conclusions are solely based on the empirical data and graphical facts presented in this work.

It has been inferred from the literature that the performance of a load management system is mainly the function of its efficiency and effectiveness, depending upon the application area. Most of the time, the knowledge of the optimum trade-off between efficiency and effectiveness is the dire need. Moreover, machine learning algorithms have been reported to be the best candidates for load management and demand forecasting. However, the selection of the relevant algorithm(s) for the specific application in order to attain higher performance is the pressing need. This study has contributed to a further extension of the research on the SEIL dataset by proposing the best candidate machine learning algorithm for the higher degree of performance. The proposed claim is also advocated by the empirical performance parameters of machine learning algorithms.

8. Conclusions

This study has presented a comprehensive and exhaustive empirical evaluation of machine learning algorithm for energy demand prediction on the SEIL dataset. The Smart Energy Informatics Lab (SEIL) of the Indian Institute of Technology (IIT) Bombay India, conducted an experimental study in 2019 to collect a massive dataset on university campus energy consumption. A comprehensive comparative study for the recommendation of the best candidate of machine learning algorithm on the SEIL dataset was the missing element

in the recent literature. This study has well filled that gap left for subsequent investigation. After the careful and detailed empirical assessment of performance parameters, it has been concluded that the Bagged Trees is the most effective algorithm for energy demand prediction applications and the Medium Trees is the most efficient algorithm for real-time systems. Moreover, Fine Trees has the optimum tradeoff between efficacy and efficiency.

References

1. Tian, J.; Li, K.; Xue, W. An adaptive ensemble predictive strategy for multiple scale electrical energy usages forecasting. *Sustain. Cities Soc.* **2021**, *66*, 102654. [CrossRef]
2. Sepehr, M.; Eghtedaei, R.; Toolabimoghadam, A.; Noorollahi, Y.; Mohammadi, M. Modeling the electrical energy consumption profile for residential buildings in Iran. *Sustain. Cities Soc.* **2018**, *41*, 481–489. [CrossRef]
3. Spandagos, C.; Ng, T.L. Equivalent full-load hours for assessing climate change impact on building cooling and heating energy consumption in large Asian cities. *Appl. Energy* **2017**, *189*, 352–368. [CrossRef]
4. Somu, N.; Gauthama, R.M.R.; Ramamritham, K. A hybrid model for building energy consumption forecasting using long short term memory networks. *Appl. Energy* **2020**, *261*, 114131. [CrossRef]
5. Batlle, E.A.O.; Palacio, J.C.E.; Lora, E.E.S.; Reyes, A.M.M.; Moreno, M.M.; Morejón, M.B. A methodology to estimate baseline energy use and quantify savings in electrical energy consumption in higher education institution buildings: Case study, Federal University of Itajubá (UNIFEI). *J. Clean. Prod.* **2020**, *244*, 118551. [CrossRef]
6. He, F.; Zhou, J.; Feng, Z.-k.; Liu, G.; Yang, Y. A hybrid short-term load forecasting model based on variational mode decomposition and long short-term memory networks considering relevant factors with Bayesian optimization algorithm. *Appl. Energy* **2019**, *237*, 103–116. [CrossRef]
7. Zhong, H.; Wang, J.; Jia, H.; Mu, Y.; Lv, S. Vector field-based support vector regression for building energy consumption prediction. *Appl. Energy* **2019**, *242*, 403–414. [CrossRef]
8. Wang, R.; Wang, J.; Xu, Y. A novel combined model based on hybrid optimization algorithm for electrical load forecasting. *Appl. Soft Comput.* **2019**, *82*, 105548. [CrossRef]
9. Li, C.; Tang, M.; Zhang, G.; Wang, R.; Tian, C. A hybrid short-term building electrical load forecasting model combining the periodic pattern, fuzzy system, and wavelet transform. *Int. J. Fuzzy Syst.* **2020**, *22*, 156–171. [CrossRef]
10. Ahmad, T.; Chen, H.; Guo, Y.; Wang, J. A comprehensive overview on the data driven and large scale based approaches for forecasting of building energy demand: A review. *Energy Build.* **2018**, *165*, 301–320. [CrossRef]
11. Wei, Y.; Zhang, X.; Shi, Y.; Xia, L.; Pan, S.; Wu, J.; Han, M.; Zhao, X. A review of data-driven approaches for prediction and classification of building energy consumption. *Renew. Sustain. Energy Rev.* **2018**, *82*, 1027–1047. [CrossRef]
12. Amasyali, K.; El-Gohary, N.M. A review of data-driven building energy consumption prediction studies. *Renew. Sustain. Energy Rev.* **2018**, *81*, 1192–1205. [CrossRef]
13. Wang, Y.; Yang, D.; Zhang, X.; Chen, Z. Probability based remaining capacity estimation using data-driven and neural network model. *J. Power Sources* **2016**, *315*, 199–208. [CrossRef]
14. Koschwitz, D.; Frisch, J.; van Treeck, C. Data-driven heating and cooling load predictions for non-residential buildings based on support vector machine regression and NARX Recurrent Neural Network: A comparative study on district scale. *Energy* **2018**, *165*, 134–142. [CrossRef]
15. Smarra, F.; Jain, A.; de Rubeis, T.; Ambrosini, D.; D'Innocenzo, A.; Mangharam, R. Data-driven model predictive control using random forests for building energy optimization and climate control. *Appl. Energy* **2018**, *226*, 1252–1272. [CrossRef]
16. Johannesen, N.J.; Kolhe, M.; Goodwin, M. Relative evaluation of regression tools for urban area electrical energy demand forecasting. *J. Clean. Prod.* **2019**, *218*, 555–564. [CrossRef]

17. Jain, R.K.; Smith, K.M.; Culligan, P.J.; Taylor, J.E. Forecasting energy consumption of multi-family residential buildings using support vector regression: Investigating the impact of temporal and spatial monitoring granularity on performance accuracy. *Appl. Energy* **2014**, *123*, 168–178. [CrossRef]

18. Moon, J.; Park, J.; Hwang, E.; Jun, S. Forecasting power consumption for higher educational institutions based on machine learning. *J. Supercomput.* **2018**, *74*, 3778–3800. [CrossRef]

19. Chou, J.-S.; Tran, D.-S. Forecasting energy consumption time series using machine learning techniques based on usage patterns of residential householders. *Energy* **2018**, *165*, 709–726. [CrossRef]

20. Li, K.; Xie, X.; Xue, W.; Dai, X.; Chen, X.; Yang, X. A hybrid teaching-learning artificial neural network for building electrical energy consumption prediction. *Energy Build.* **2018**, *174*, 323–334. [CrossRef]

21. Khan, P.W.; Byun, Y.-C.; Lee, S.-J.; Park, N. Machine learning based hybrid system for imputation and efficient energy demand forecasting. *Energies* **2020**, *13*, 2681.

22. Ahmad, T.; Chen, H. A review on machine learning forecasting growth trends and their real-time applications in different energy systems. *Sustain. Cities Soc.* **2020**, *54*, 102010. [CrossRef]

23. Akhtar, S.; Sujod, M.Z.B.; Rizvi, S.S.H. A Hybrid Soft Computing Framework for Electrical Energy Optimization. In Proceedings of the 2021 6th International Multi-Topic ICT Conference (IMTIC), Lahore, Pakistan, 14 November 2021.

24. Akhtar, S.; Sujod, M.Z.b.; Rizvi, S.S.H. A Novel Deep Learning Architecture for Data-Driven Energy Efficiency Management (D2 EEM)-Systematic Survey. In Proceedings of the 2021 International Conference on Engineering and Emerging Technologies (ICEET), Tianjin, China, 1–3 July 2021.

25. Zhang, Q.; Yang, L.T.; Chen, Z.; Li, P. A survey on deep learning for big data. *Inf. Fusion* **2018**, *42*, 146–157. [CrossRef]

26. Li, Y.; Hang, C.; Ding, L.; Pan, Y.; Gao, X. Deep learning in bioinformatics: Introduction, application, and perspective in the big data era. *Methods* **2019**, *166*, 4–21. [CrossRef]

27. Wang, H.; Lei, Z.; Zhang, X.; Zhou, B.; Peng, J. A review of deep learning for renewable energy forecasting. *Energy Convers. Manag.* **2019**, *198*, 111799. [CrossRef]

28. Manic, M.; Amarasinghe, K.; Rodriguez-Andina, J.J.; Rieger, C. Intelligent buildings of the future: Cyber aware, deep learning-powered, and human interacting. *IEEE Ind. Electron. Mag.* **2016**, *10*, 32–49. [CrossRef]

29. Marino, D.L.; Amarasinghe, K.; Manic, M. Building energy load forecasting using deep neural networks. In Proceedings of the IEEE IECON 2016—42nd Annual Conference of the IEEE Industrial Electronics Society, Florence, Italy, 23–26 October 2016.

30. Grolinger, K.; Capretz, M.A.M.; Seewald, L. Energy consumption prediction with big data: Balancing prediction accuracy and computational resources. In Proceedings of the 2016 IEEE International Congress on Big Data (BigData Congress), San Francisco, CA, USA, 27 June–2 July 2016.

31. Mocanu, E.; Nguyen, P.H.; Gibescu, M.; Larsen, E.M.; Pinson, P. Demand forecasting at low aggregation levels using factored conditionally restricted Boltzmann machine. In Proceedings of the IEEE 2016 Power Systems Computation Conference (PSCC), Genoa, Italy, 20–24 June 2016.

32. Amarasinghe, K.; Marino, D.L.; Manic, M. Deep neural networks for energy load forecasting. In Proceedings of the 2017 IEEE 26th International Symposium on Industrial Electronics (ISIE), Edinburgh, UK, 19–21 June 2017.

33. Paterakis, N.G.; Mocanu, E.; Gibescu, M.; Stappers, B.; van Last, W. Deep learning versus traditional machine learning methods for aggregated energy demand prediction. In Proceedings of the 2017 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe), Torino, Italy, 26–29 September 2017.

34. Chen, Y.; Shi, Y.; Zhang, B. Modeling and optimization of complex building energy systems with deepneural networks. In Proceedings of the IEEE 2017 51st Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, USA, 29 October–1 November 2017.

35. Mocanu, E.; Mocanu, D.C.; Nguyen, P.H.; Liotta, A.; Webber, M.E.; Gibescu, M.; Slootweg, J.G. On-line building energy optimization using deep reinforcement learning. *IEEE Trans. Smart Grid* **2018**, *10*, 3698–3708. [CrossRef]

36. Nichiforov, C.; Stamatescu, G.; Stamatescu, I.; Calofir, V.; Fagarasan, I.; Iliescu, S.S. Deep learning techniques for load forecasting in large commercial buildings. In Proceedings of the IEEE 2018 22nd International Conference on System Theory, Control and Computing (ICSTCC), Sinaia, Romania, 10–12 October 2018.

37. Afrasiabi, M.; Mohammadi, M.; Rastegar, M.; Kargarian, A. Multi-agent microgrid energy management based on deep learning forecaster. *Energy* **2019**, *186*, 115873. [CrossRef]

38. Ahmad, T.; Chen, H. Deep learning for multi-scale smart energy forecasting. *Energy* **2019**, *175*, 98–112. [CrossRef]

39. Hafeez, G.; Alimgeer, K.S.; Wadud, Z.; Shafiq, Z.; Khan, M.U.A.; Khan, I.; Khan, F.A.; Derhab, A. A Novel Accurate and Fast Converging Deep Learning-Based Model for Electrical Energy Consumption Forecasting in a Smart Grid. *Energies* **2020**, *13*, 2244. [CrossRef]

40. Han, T.; Muhammad, K.; Hussain, T.; Lloret, J.; Baik, S.W. An Efficient Deep Learning Framework for Intelligent Energy Management in IoT Networks. *IEEE Internet Things J.* **2020**, *8*, 3170–3179. [CrossRef]

41. Khan, Z.A.; Ullah, A.; Ullah, W.; Rho, S.; Lee, M.; Baik, S.W. Electrical Energy Prediction in Residential Buildings for Short-Term Horizons Using Hybrid Deep Learning Strategy. *Appl. Sci.* **2020**, *10*, 8634. [CrossRef]

42. Elsisi, M.; Tran, M.-Q.; Mahmoud, K.; Lehtonen, M.; Darwish, M.M.F. Deep learning-based industry 4.0 and internet of things towards effective energy management for smart buildings. *Sensors* **2021**, *21*, 1038. [CrossRef] [PubMed]

43. Truong, L.H.M.; Chow, K.H.K.; Luevisadpaibul, R.; Thirunavukkarasu, G.S.; Seyedmahmoudian, M.; Horan, B.; Mekhilef, S.; Stojcevski, A. Accurate Prediction of Hourly Energy Consumption in a Residential Building Based on the Occupancy Rate Using Machine Learning Approaches. *Appl. Sci.* **2021**, *11*, 2229. [CrossRef]

44. Hamdoun, H.; Sagheer, A.; Youness, H. Energy time series forecasting-analytical and empirical assessment of conventional and machine learning models. *J. Intell. Fuzzy Syst. Prepr.* **2021**, *40*, 12477–12502. [CrossRef]

45. Hafiz, F.; Awal, M.A.; de Queiroz, A.R.; Husain, I. Real-Time Stochastic Optimization of Energy Storage Management Using Deep Learning-Based Forecasts for Residential PV Applications. *IEEE Trans. Ind. Appl.* **2020**, *56*, 2216–2226. [CrossRef]

46. Wu, L.-X.; Lee, S.-J. A deep learning-based strategy to the energy management-advice for the time-of-use rate of household electricity consumption. *J. Internet Technol.* **2020**, *21*, 305–311.

47. Arienti, J.H.L. Time Series Forecasting Applied to an Energy Management System—A Comparison between Deep Learning Models and Other Machine Learning Models. Doctoral Dissertation, Universidade Nova de Lisboa, Lisbon, Portugal, 2020.

48. Aragón, G.; Puri, H.; Grass, A.; Chala, S.; Beecks, C. Incremental deep-learning for continuous load prediction in energy management systems. In Proceedings of the 2019 IEEE Milan PowerTech, Milan, Italy, 23–27 June 2019.

49. Kumar, H.; Priyanka, M.M.; Krithi, R. Explainable ai Deep reinforcement learning agents for residential demand-side cost savings in smart grids. *arXiv* **2019**, arXiv:1910.08719.

50. Tanted, S.; Agarwal, A.; Mitra, S.; Bahuman, C.; Ramamritham, K. Database and Caching Support for Adaptive Visualization of Large Sensor Data. In Proceedings of the ACM IKDD CoDS and 25th COMAD, Hyderabad, India, 5–7 January 2020; pp. 98–106.

51. Luo, X.J.; Oyedele, L.O. Forecasting building energy consumption: Adaptive long-short term memory neural networks driven by genetic algorithm. *Adv. Eng. Inform.* **2021**, *50*, 101357. [CrossRef]

52. Ramamritham, K.; Karmakar, G.; Shenoy, P.J. Smart Energy Management: A Computational Approach. In *International Conference on Big Data Analytics*; Springer: Cham, Switzerland, 2017.

53. Jois, S.; Ramamritham, K.; Agarwal, V. *Impact of Facade Based Building Integrated Photo Voltaicson the Indoor Thermal Comfort in Tropical Urban Areas—Mumbai as a Case Study*; AGU Fall Meeting 2019; American Geophysical Union: Washington, DC, USA, 2019.

54. Jois, S.; Ramamritham, K.; Agarwal, V. *Modeling Tools Development for Assessing Façade Based PV Feasibility in a Data-Scarce Developing Nation Using Open Source Technology*; AGU Fall Meeting 2019; American Geophysical Union: Washington, DC, USA, 2019.

55. Lee, S.; Shenoy, P.; Ramamritham, K.; Irwin, D. vSolar: Virtualizing Community Solar and Storage for Energy Sharing. In Proceedings of the Ninth International Conference on Future Energy Systems (e-Energy '18), Karlsruhe, Germany, 12–15 June 2018; ACM: New York, NY, USA, 2018.