



Alexandria University
Alexandria Engineering Journal

www.elsevier.com/locate/aej
www.sciencedirect.com



ORIGINAL ARTICLE

Prediction of energy consumption in campus buildings using long short-term memory



Muhammad Faiq^a, Kim Geok Tan^{a,*}, Chia Pao Liew^{b,*}, Ferdous Hossain^a,
 Chih-Ping Tso^a, Li Li Lim^c, Adam Yoon Khang Wong^d, Zulhilmi Mohd Shah^e

^a Faculty of Engineering and Technology, Multimedia University, Melaka 75450, Malaysia

^b School of Energy and Chemical Engineering, Xiamen University Malaysia, Selangor 43900, Malaysia

^c Faculty of Engineering and Technology, Tunku Abdul Rahman University of Management and Technology, Kuala Lumpur 53300, Malaysia

^d Faculty of Electrical and Electronic Engineering Technology, Universiti Teknikal Malaysia Melaka, Melaka 76100, Malaysia

^e Facilities Management Department (FMD), Multimedia University, Melaka 75450, Malaysia

Received 2 July 2022; revised 17 October 2022; accepted 6 December 2022

KEYWORDS

Energy consumption;
 Long short-term memory;
 Support vector regression;
 Gaussian process regression;
 Weather forecasting

Abstract In this paper, Long Short-Term Memory (LSTM) was proposed to predict the energy consumption of an institutional building. A novel energy usage prediction method was demonstrated for daily day-ahead energy consumption by using forecasted weather data. It used weather forecasting data from a local meteorological organization, the Malaysian Meteorological Department (MET). The predictive model was trained by considering the dependencies between energy usage and weather data. The performance of the model was compared with Support Vector Regression (SVR) and Gaussian Process Regression (GPR). The experimental results with a dataset obtained from a building in Multimedia University, Malacca Campus from January 2018 to July 2021 outperformed the SVR and GPR. The proposed model achieved the best RMSE scores (561.692–592.319) when compared to SVR (3135.590–3472.765) and GPR (1243.307–1334.919). Through experimentation and research, the dropout method reduced overfitting significantly. Furthermore, feature analysis was done with SHapley Additive exPlanation to identify the most important weather variables. The results showed that temperature, wind speed, rainfall duration and the amount had a positive effect on the model. Thus, the proposed approach could aid in the implementation of energy policies because accurate predictions of energy consumption could serve as system fault detection and diagnosis for buildings.

© 2022 THE AUTHORS. Published by Elsevier BV on behalf of Faculty of Engineering, Alexandria University. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

* Corresponding authors.

E-mail addresses: kgtan@mmu.edu.my (K. Geok Tan), chiapao.liew@xmu.edu.my (C. Pao Liew).

Peer review under responsibility of Faculty of Engineering, Alexandria University.

<https://doi.org/10.1016/j.aej.2022.12.015>

1110-0168 © 2022 THE AUTHORS. Published by Elsevier BV on behalf of Faculty of Engineering, Alexandria University. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

International Energy Agency (IEA) reported that global electricity demand will increase up to 3 % annually due to

continued economic growth in 2022 [1]. Carbon dioxide emissions increased from 1990 to 2020, and then declined by 5.8 % due to Covid-19 restrictions [2]. Despite the decrease in greenhouse gases emission in 2020, the concentration of carbon dioxide reached the highest average annual concentration. World Meteorological Organization (WMO) concluded that the global mean temperature of 2021 was 1.09 °C higher compared to the era 1850 to 1900 [3].

Building energy consumption accounts for a significant portion of Malaysia's total energy consumption. Residential and commercial buildings accounted for 13 % of total energy consumption in 2020, according to Malaysia's energy consumption trend [4]. Buildings consumed 54 % of the total electricity usage in Malaysia [5], which in turn was expected to rise 4.3 % annually on average from 2004 to 2030 [4]. The huge increase in energy demand was due to the growing population and economic development, which is also true globally. Hence there was motivation to improve the utilization of energy. Research showed the inefficiency of Malaysian office buildings and that optimizations could be done to reduce energy usage [6]. Furthermore, according to an energy benchmark conducted on higher education buildings in Australia, only 7 % of buildings were energy efficient [7]. Universities as a group are ideal for energy studies, as Malaysia has more than 100 institutions which consist of 20 public universities and private universities that contribute to energy consumption [8]. Furthermore, institutional buildings could be easily optimized by computer modelling [9]. Therefore, a predictive model was created and reported here to help foresee the energy consumption of Multimedia University Malacca.

The data-driven model created the most predictive models [20]. Which required complex and comprehensive information regarding buildings in creating a physical model [10]. Physical models are relatively complicated, yet they fail to accurately anticipate energy demand in buildings [43]. Data-driven models such as Artificial Neural Networks (ANNs) and SVR were preferred, due to their advantage in being able to exclude the complexity of structures, which was influenced by a variety of factors such as thicknesses of walls and types of building materials in developing a predictive model [11]. In Greece, the Multilayer Perceptron Model (MLP) is being used to predict long-term energy usage [12]. A feed-forward neural network was developed to forecast diurnal energy consumption for buildings had shown great prediction accuracy [13]. In Hong Kong, the decision tree model and neural network are preferred in predicting electricity energy consumption [16].

Environmental variables such as temperature could greatly affect the accuracy of the prediction model [14]. Table 1 showed the analysis of external variables that affected the performance of the predictive models. Analysis showed environmental variables could be useful indicators in creating a good predictive model [15]. In a study, 9 parameters including environmental variables were used in predicting solar irradiance prediction [19]. In another study, 11 external parameters were used in hourly energy prediction for two educational buildings with accurate results [15]. Therefore, weather data including temperature, wind velocity, humidity and air pressure were obtained from the Malaysian Meteorological Department (MET) in creating a predictive model here. The LSTM is used to predict the energy consumption of the institutional building due to its ability to solve long-term and short-term dependencies among the data [36]. LSTM was used

Table 1 Analysis of external variables affecting the performance of the predictive models.

Ref.	External variables used	Findings
[14]	<ul style="list-style-type: none"> • Principal building activity • Square footage • Number of floors • Heating degree days • Cooling degree days 	The square footage of a building was an important factor in affecting the performance of the model followed by the climate-related features (i.e., heating degree days and cooling degree days).
[15]	<ul style="list-style-type: none"> • Outdoor temperature • Dew point • Relative humidity • Barometric pressure • Precipitation • Wind speed • Solar radiation • Number of occupants • Time of day • Workday type • Day type 	Different buildings had its variable importance pattern shown for each predictive model. One of the buildings showed that energy usage depended on the occupancy variables (i.e., weekday, time of day and number of occupants) while the other building showed both occupancy data and environmental variables (i.e., dew point and pressure) highly impact the prediction performance of the model.
[30]	<ul style="list-style-type: none"> • Ambient temperature • Solar flux • Humidity • Hour of the day • Day of the week • Day of the year 	The prediction of the electric load in buildings showed that environmental variables such as ambient temperature and solar radiation highly impacted the results of the prediction while others such as wind velocity and humidity barely affected the outcome. The day and time variables and occupancy variable were significant in developing a good predictive model.

in short-term residential load forecasting that showed good prediction despite the high volatility and uncertainty of the data [17]. Genetic Algorithm (GA) was used to optimise the LSTM model for electric load forecasting using French metropolitan electricity data [18]. It showed that LSTM was able to show good prediction accuracy from meteorological observation in rainfall-runoff modelling [32].

The model created was used to predict the daily day-ahead for energy consumption for the institutional building. It was noted that forecast and prediction were defined as the estimation of the future magnitude of a variable [20]. These predictions can be benchmarked and used for system fault detection. Optimisations could be done by building retrofit to increase the efficiency of energy usage [21]. To achieve COP26's objective to reduce global warming by reducing greenhouse gases emission [22]. Furthermore, the amount of electricity consumption in Malaysia was correlated with its

Gross Domestic Product (GDP) which was a common metric to assess economic growth [5]. Therefore, the proposed work could be helpful in implementing energy policies as accurate energy predictions may have a significant impact on capital expenditure [12].

2. Literature review

The concept of energy prediction is to identify and predict the improvements that can be made to optimize energy usage in a building. There are many energy prediction techniques such as the physical method, statistical method and hybrid method [10], each having advantages and disadvantages.

A review of energy prediction techniques used in building energy performance forecasting has been presented [10,20,23]. Physical methods or white-box models such as the nodal method, zonal method and CFD method require detailed information and description of the building as the inputs. As a result, the energy prediction techniques become complex and take high computational time. However, these techniques do not require historical data of the buildings as it predicts the data of the building through simulation of the building properties. There is multiple software available for each method. There are EnergyPlus and TRNSYS for the nodal approach. The COMSOL and ANSYS are for CFD method, while SimSPARK is using the zonal approach. Wurtz et al. [24] predicted accurately the temperature field of a room using the zonal method. The study in [25] created a model of the building in EnergyPlus to analyze the energy performance of the building by changing the properties of the materials of the building.

The statistical method or black-box model is currently the most used in predicting building energy consumption. The method is comprised of GA, ANN, SVR, Auto-Regressive Integrated Moving Average (ARIMA), Multiple Linear Regression (MLR), GPR, Gradient Boosting and Principal Component Analysis [23]. González and Zamarreño [26] used feedback ANN for electric load forecasting. Ekonomou [12] created a predictive model using MLP for long-term energy consumption prediction in Greece. Deb et. al [13] showed that feedforward ANN can forecast the next day's diurnal cooling energy load with great accuracy for institutional buildings. The computational burden and complexity of physical methods cause the statistical method to be preferred. In [23], the developed ANN was able to significantly reduce the computational effort and time. As the statistical method requires historical data of the building, the technique in [21] used other institutional buildings to improve predictions made.

The hybrid method or grey-box model is the combination of physical and statistical techniques to produce a better model. In [27], a simplified tool for the thermal evaluation of a building in the Mediterranean coupled with GA was used to optimize the building parameter for optimum thermal evaluation. The ANN was used in the analysis for neural identification architecture to improve the HVAC system with a single-zone thermal system model [28]. In [29], DOE-2 physical software was coupled with GA by optimising values of architecture structure related to the envelope to increase energy efficiency for residential buildings. Table 2 summarises the features of the models.

Table 2 Advantages and disadvantages of several energy prediction techniques.

Model	Advantage	Disadvantage
Physical model	Great understanding of the building properties and exception of historical data as input.	High computation, is more complex and requires a detailed description of the building.
Statistical model	Less computational time and less complexity in the model.	Requires a huge amount of historical data to create an accurate model.
Hybrid model	Requires less historical data and retains physical interpretation of the building.	Requires less historical data compared to the statistical method, and medium computation time.

For the present research, a data-driven model which is a deep-learning model, called LSTM, will be used as the predictive model. Deep learning falls under the sub-category of machine learning and can store more information, and thus more data can be incorporated to make more accurate predictions [31]. The disadvantage of deep learning architectures is that they require a huge quantity of parameters, and are complex to train [20]. Examples of deep learning models include deep neural networks (DNN), recurrent neural networks (RNN) and autoencoders (AE). Table 3 shows some previous studies that contribute significantly to providing a platform for future studies, suggestions to improve certain aspects and their comparisons. Some studies reported on several iterations instead of batch size and number of epochs.

External variables such as temperature, wind velocity, pressure, humidity and dew point can be important factors in increasing prediction accuracy [15]. However, not all environmental variables are significant and some can be omitted, such as wind velocity or humidity [30]. In [25], a building's energy consumption is influenced by the insulation materials used. In [14], it was shown temperature was an important component and an indicator of energy used in either heating or cooling a building. These are important aspects that should be identified and analyzed in building retrofit and in energy consumption forecasting. By predicting building energy consumption accurately, a better energy benchmark and better energy efficiency can be obtained [34].

3. Energy prediction techniques

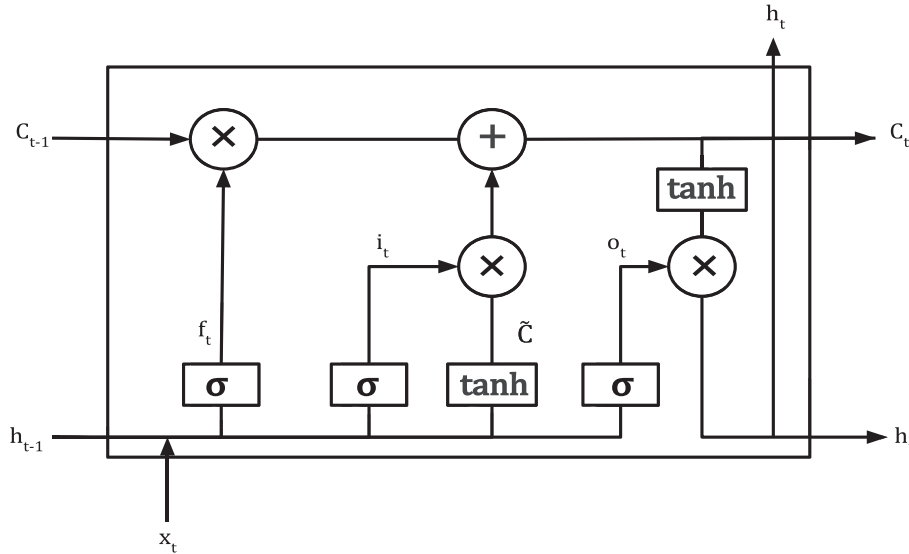
3.1. Proposed prediction algorithm, Long Short-Term Memory (LSTM)

LSTM is similar to RNN but with some distinguishing features. RNN architecture may consist of nodes, layers, and connections such as Feedforward Neural Networks (FNNs) [35]. An RNN can establish self-loop connections from one node to another through time step intervals [36].

Fig. 1 describes the structure of an LSTM unit. Instead of a simple RNN unit, LSTM consist of memory blocks with three

Table 3 Comparison of the proposed algorithm with related existing deep learning models.

Ref.	Model type	Number of epochs	Number of hidden neurons	Batch size	Is dropout used?	Dropout value
[17]	LSTM	150	20		No	
[18]	LSTM	150	100	125	No	
[19]	LSTM	50	30	50	No	
[31]	LSTM	50	20	1	Yes	0.1
[32]	CNN-LSTM		100		No	
[33]	DE-LSTM	3000	50	500	No	
[41]	LSTM		40		Yes	0.5

**Fig. 1** Structure of Long Short-term Memory (LSTM).

gates: input gate i_t , output gate o_t and forget gate f_t . The gates are important in avoiding the vanishing gradient problem as there will be a significant number of environmental variables [33]. Accurate predictive modelling as LSTM factors in the dependency between variables on the relevant timestamp [38]. In Reference [37], a detailed description of the LSTM model is given.

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (1)$$

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \quad (2)$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (3)$$

$$\tilde{C}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c) \quad (4)$$

$$C_t = f_t \otimes C_{t-1} + i_t \otimes \tilde{C}_t \quad (5)$$

$$h_t = o_t \otimes \tanh(C_t) \quad (6)$$

In the equations above, i_t , f_t and o_t are the three gates, input, output and forget gates, respectively at the time, t W_i and W_f . The W_o denotes the weight matrices from the input, forget and output gates to the input, respectively. The b_i , b_f and b_o are the bias of input, forget and output gate, respec-

tively. The U_i , U_f and U_o denote the weight matrices from the input, forget and output gates to the hidden, respectively. σ is a logistic sigmoid function and \otimes denotes the element-wise multiplication of two vectors. x_t is a vector that is located in the input layer of the LSTM. h_t is an output vector of the hidden layer and is located in the LSTM unit at the time, t .

C_t denotes the current cell state and \tilde{C}_t denotes the new candidate value for the next cell state. h_{t-1} denotes the previous state and is determined by the forget gate, f_t , by how much is passed to the next state. C_{t-1} denotes the update of the old cell state to the new cell state, C_t .

3.2. Dropout for proposed prediction algorithm

Deep neural network with many parameters can cause overfitting, especially when the datasets are small [42]. Big improvements in model performance can be observed when dropout is applied to the model since dropout encourages each hidden unit to identify useful features without relying on other hidden units to do its correction [44]. A brief description of dropout is when a neural network model is updating its hidden layer where the dropout is applied, it arbitrarily does not update neurons in the layer. Fig. 2. shows a neural network model when dropout is applied and dropped. The neurons

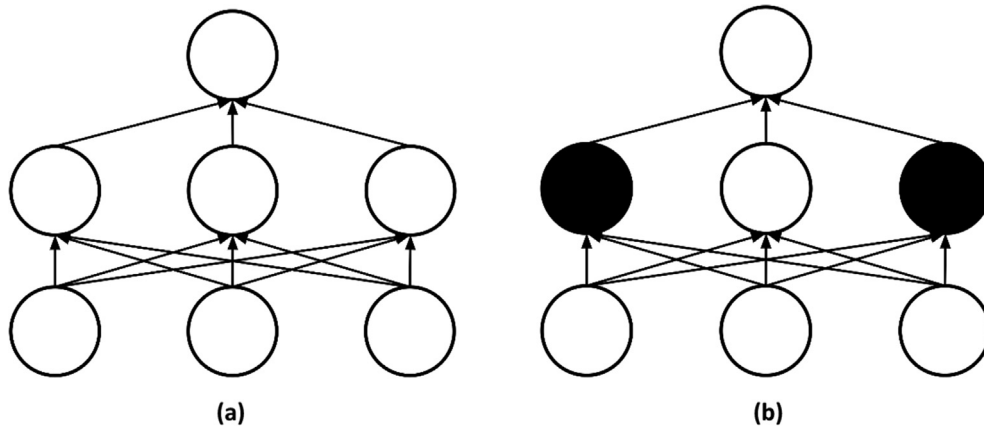


Fig. 2 (a) Standard neural network with a hidden layer; (b) Neural network with a hidden layer and dropout applied.

highlighted in black are dropped by the model when dropout is applied.

3.3. Importance of features

By using external environmental variables such as temperature, dew point, wind velocity, pressure and occupancy, we can increase the effectiveness of the model. A study in [14] showed increased accuracy when external variables are trained in the model. Another study [15] showed the significance of each external variable in creating a good prediction. The significance of a variable can be evaluated by comparing it with the prediction accuracy of other variables.

Each variable will be dropped and the model will be trained. The model's performance will be evaluated with and without the variable to determine the variable's significance. The model will be simulated 20 times to give a fair experiment among features. Feature analysis is important to understand the relation between the input and output. The significance of each variable can lead to the understanding and selection of highly influential variables in the model. Feature analysis is important to understand and correlate the highly influential variables that influence the changes in the trend of energy consumption of buildings.

3.4. Evaluation metrics

The performance of the prediction is determined by commonly used prediction accuracy evaluation which is the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE), defined below:

$$MAE = \frac{1}{N} \sum_{i=1}^N |(y_{predicted} - y_{actual})| \quad (7)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_{actual} - y_{predicted})^2} \quad (8)$$

Here, N is the sample size or the size of training sets, $y_{predicted}$ and y_{actual} are the predicted and actual values, respectively.

The MAE is a conventional statistical error indicator that represents the performance of the model by computing the

average of absolute errors between the predicted and actual values. The RMSE is a sample standard deviation of the predicted and the actual values. RMSE amplifies and contrasts the error severely due to its mathematical structure. Smaller values of MAE and RMSE indicate good predictive modelling.

4. Experimental configuration

4.1. Data

There are 8 variables used as input data and a variable will be obtained as output data. The input data are environmental variables (i.e., pressure (hPa), temperature ($^{\circ}$ C), relative humidity (%), wind velocity (m/s), rainfall duration (min), rainfall amount (mm)), and type of (occupancy related) data. Table 4 summarizes the input parameters and their ranges. As strings are hard to be analyzed by the deep learning model, weekday is set to 0 while weekend and holiday are set to 1. Meanwhile, the type of lockdown is set such that 0 is set when no movement control order (MCO) or total lockdown is applied, 1 when movement control order (MCO) or total lockdown is issued, and 2 for recovery movement control order (RMCO). Information on the model with total sets and the period of datasets taken are shown in Table 5.

The environmental variables are obtained from the Malaysian Meteorological Department (MET). The latitude and longitude of the weather station are $2^{\circ} 16' N$ and $102^{\circ} 15' E$, respectively. The displacement between the weather station and the institutional building is 3.47 km.

The output data is the usage peak of the electricity usage used by the appliances, or heating, ventilation and air conditioning (HVAC) system. Initially, the data for electricity usage were obtained monthly from Tenaga Nasional Berhad (TNB). The data is then extracted and divided into daily data depending on the number of days in the month. The value for each daily data is changed approximately depending on the type of day. According to research [39], energy consumption during weekends and holidays is lower by 15%. Fig. 3. depict the daily energy usage of our institutional building from 2018 to 2021. Instead of the classic four seasons in the western country, Malaysia has no seasonal changes in climate. Therefore, the temperature of Malaysia ranges from $22^{\circ} C$ to $33^{\circ} C$ including an average of sunshine between 7 and 12 h a day.

Table 4 Weather variables and values range at the MET Melaka Weather Station.

Variable	Abbreviation	Type	Measurement	Range of values
Pressure	Press	Continuous	hPa	1006 – 1014
Environmental temperature	Temp	Continuous	Deg. C	22 – 31
Relative humidity	Hum	Continuous	%	58 – 94
Wind velocity	Wind	Continuous	m/s	0.5 – 5.3
Rainfall duration	RainDur	Continuous	Minutes	0 – 60
Rainfall amount	RainAm	Continuous	mm	–14 – 4
Type of day	Day	Categorical	Weekday, weekend and holiday	0 for weekday and 1 for weekend and holiday
Type of lockdown	Lock	Categorical	No MCO, MCO and RMCO	0 for no MCO, 1 for MCO and 2 for RMCO

Table 5 Summary of algorithms and total datasets used.

Algorithm	Data	Period	Total datasets
Long Short-term Memory (LSTM)	Daily data	01/01/2018 to 31/7/2021	1308
Regression Tree (RT)	Daily data	01/01/2018 to 31/7/2021	1308
Support Vector Regression (SVR)	Daily data	01/01/2018 to 31/7/2021	1308

4.2. Experimental procedure

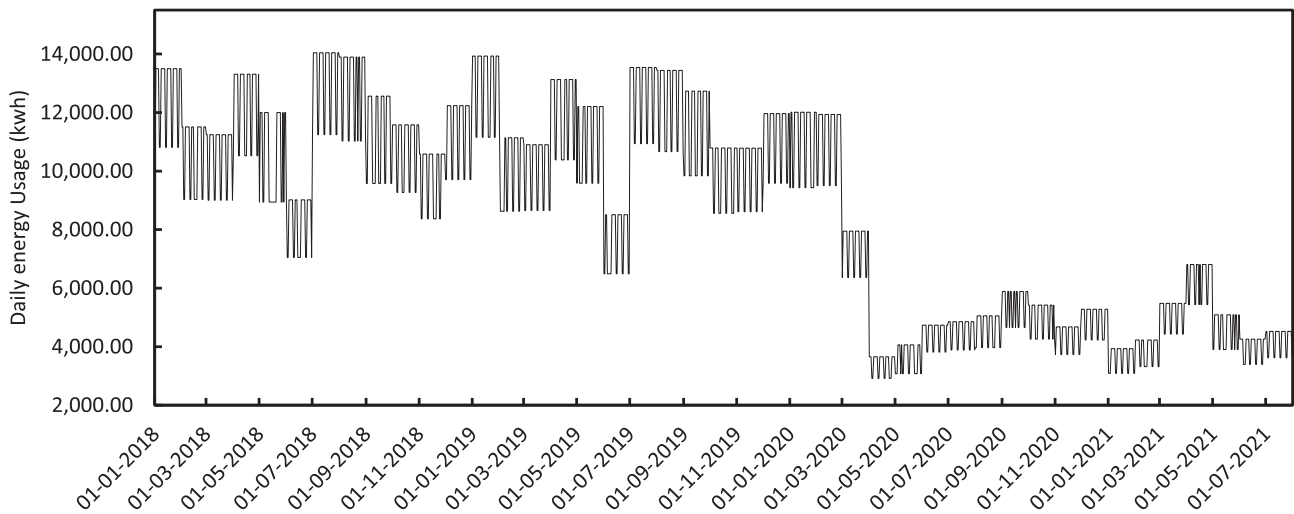
Fig. 4 shows the experimental approach in which three prediction models, LSTM, SVR, and GPR, were developed and deployed for the institutional building to compare their performance. Each input variable will be tested to understand the relation between the variable and the accuracy of the model.

There were several series of steps taken to train the three predictive models. Initially, the training and validating datasets were divided into two parts. 85 % of the total datasets are used to train the model and 15 % to validate the model. Thus, the training and testing datasets are composed of energy

usage and environmental variables, respectively. The performance of the model was compared by using statistical tools stated in the evaluation metrics.

The predictive model architectures are described as follows:

- The SVR model is created and implemented using the Scikit toolbox, using Gaussian or Radial Basis Function (RBF) kernel. Among other kernel functions such as linear function, Sigmoid function, and polynomial function, the RBF is best to represent complex dependencies between input and output variables [40]. An SVR model is composed of penalty parameter C and the radius ϵ . A grid search of hyperparameters with 5-fold cross-validations is done to improve the parameters. The optimized parameters are $C = 5$ and $\epsilon = 1$.
- The GPR model is created and implemented using the Scikit toolbox. Using common kernel models Radial-basis Function (RBF), Matern 3/2 (Ma3), Matern 5/2 (Ma5) and rational quadratic are assessed and compared. 20-fold cross-validations are done on kernel simulations to find the best performance evaluation. The Radial-basis Function (RBF) performs the best based on MAE and RMSE findings.
- The LSTM model is created and implemented using the deep learning tool in Keras. The LSTM parameters are provided in Table 6. The LSTM model is trained with

**Fig. 3** Daily usage peak of the main building of Multimedia University Malacca from 01/01/2018 to 31/08/2021.

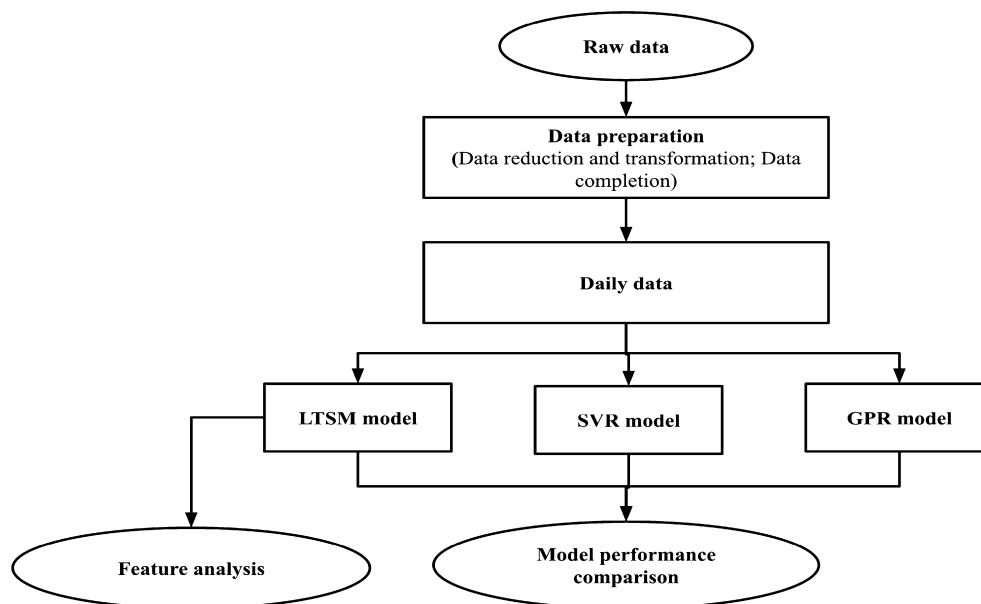


Fig. 4 Experimental procedure in the study.

7 features consisting of environmental variables and an external variable. An output layer is provided for the output variable, energy usage. Hyperparameters are optimised by tuning and interpreting the results of the number of epochs, the batch size and the number of hidden neurons. The optimised hyperparameters for the model is 2500 epochs, 100 hidden neurons and 70 batch size.

The data are normalized to [0,1] using the normalization technique from the pre-processing library from Scikit and from Section I in A, 0.5 dropouts are used. The models are simulated 20 times to have a fair comparison between the models. Table 6 shows the summary of the performance evaluation of MAE and RMSE of SVR, GPR, and LSTM.

5. Results and analysis

5.1. Performance evaluation

We can observe from Table 7 that LSTM has the best performance compared to others. The RMSEs of the predicted energy usage value using SVR, GPR and LSTM models are

Table 6 Predicting energy consumption using LSTM architecture based on Keras.

```

model = Sequential()
model.add(LSTM(100, input_shape = (1,15))
model.add(Dropout(0.5))
model.add(Dense(1))
model.compile(loss = 'mean_squared_error', optimize = 'adam',
metrics = ['accuracy'])
history = model.fit(train_X, train_Y, epochs = 2500,
batch_size = 70, validation_data = (test_X, test_Y))
  
```

3270.836 (kWh), 1310.105 (kWh), and 572.545 (kWh). The RMSE is relatively large due to a big drop in energy consumption during Movement Control Order (MCO) started on 18 March 2020 to 3 May 2020 and again on 11 January 2021 to 31 May 2021. Furthermore, energy consumption is still relatively low during the recovery movement control order (RMCO) due to the absence of students on campus and the restriction on the number of people in the office due to government policies. These results in huge prediction errors for SVR and GPR. The LSTM model can predict accurately even with a huge change in value. We can observe that the LSTM model has relatively lower MAE and RMSE compared to SVR and GPR for 20 simulations in Fig. 5 and Fig. 6.

We can observe from Table 8 the performance of the proposed algorithm with other predictive models. The proposed algorithm has the best MAE, 165.20, and the second-highest RMSE, 572.55, among the predictive models. However, each predictive model is created with different input variables and hyperparameters and thus, the comparison may be inaccurate. With respect to our data, the lockdown period was found to be incredibly significant as input to the proposed algorithm. In addition, the proposed algorithm takes advantage of rainfall duration and rainfall amount to further increase the prediction accuracy. The predictive model in [18] used weather variables (i.e., temperature, humidity, and wind speed) and schedule-related variables (i.e., weekday and month) as input variables.

Table 7 Summary of performance evaluation on the models.

Model	MAE (kWh)	RMSE (kWh)
LSTM	165.203	572.545
SVR	2851.339	3270.836
GPR	999.880	1310.105

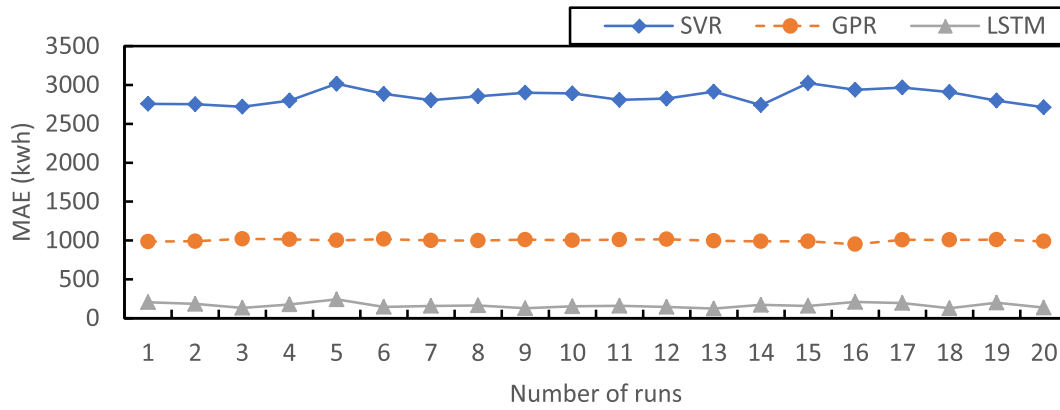


Fig. 5 MAE of 20 simulations using SVR, GPR and LSTM.

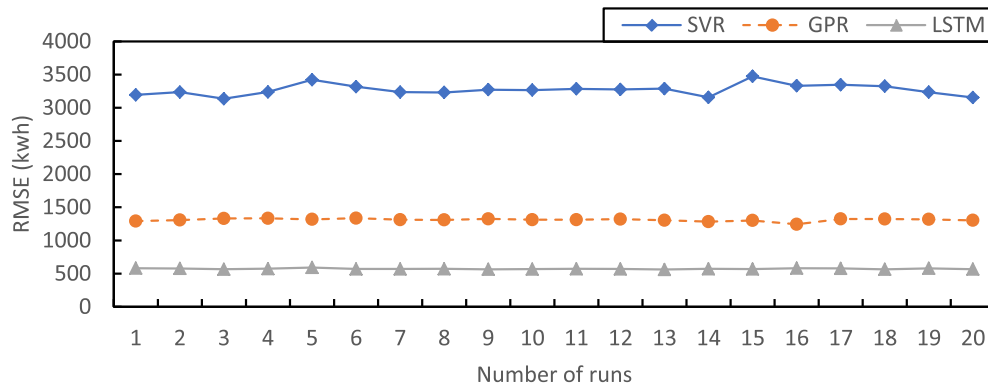


Fig. 6 RMSE of 20 simulations using SVR, GPR and LSTM.

Table 8 Comparison of the proposed algorithm with related existing deep learning models.

Algorithm	MAE	RMSE
Proposed	165.203	572.545
LSTM-RNN [18]	263.14	353.38
CNN-LSTM [32]	692.14	1134.18

The deep learning model in [32] used energy usage only as the input variable.

Fig. 7 shows an example of daily day-ahead energy consumption for the institutional building using LSTM for August 1, 2021, to August 7, 2021. August 1, 2021, and August 7, 2021, are weekends, and hence observation of a drop in energy consumption. The forecast done by the LSTM model is fairly accurate as the historical data has a huge drop due to the Covid-19 pandemic.

5.2. Features importance analysis

Feature analysis is done by removing a weather variable for every 20 simulations. Table 9 shows the summary of the performance evaluation for each simulation run. There are 6 input parameters used to train the LSTM model. We can observe the

lowest in RMSE, 568.868 (kWh) and 568.941 (kWh) when average pressure and humidity are not used as input parameters for the model. High RMSE, 586.694 (kWh) and 577.557 (kWh) are obtained when average temperature and rainfall amount are not used to train the predictive model. Without average temperature and average rainfall amount, the MAE is the highest at 212.792 (kWh) and 181.361 (kWh) respectively. The lowest MAE, 163.780 (kWh) and 146.238 (kWh) when average pressure and humidity are not included in the simulation of the model. In summary, the amount of temperature and rainfall are important parameters to be included in training the model. Research in [45] indicates rainfall may decrease energy consumption for cooling in hot climates especially wind-driven rain (WDR).

By using the SHapley Additive exPlanation (SHAP), we are able to evaluate the significance of input variables within the model. The focus of analysis is to figure out the most influential features of the LSTM model. Fig. 8 and Fig. 9 show the SHAP summary plot that arranges the features-based level of dependencies to the value of energy usage. Specifically, the type of lockdown has the greatest influence on the performance of the model followed by the type of day whether weekday, weekend, or holiday. This is due to movement restrictions order as a result of government policy in reducing the case of COVID-19. The author in [47] stated COVID-19 lockdown had reduced energy demand in the industrial sector while an increase in energy consumption at the household level. Furthermore, the observation is likely due to a decrease of 15 %

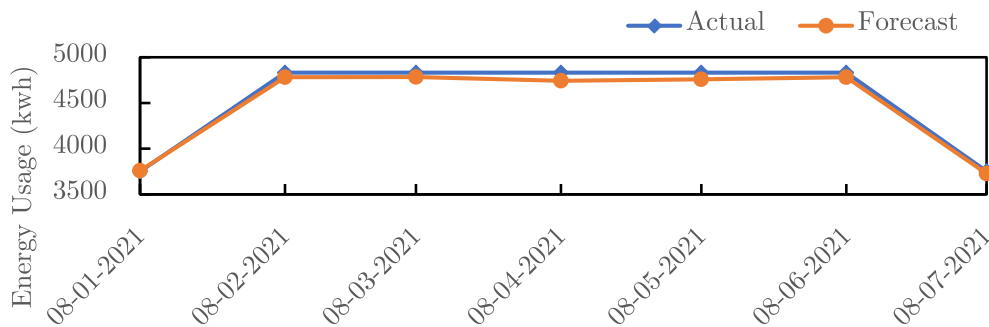


Fig. 7 Prediction of energy consumption from 1/8/2021 to 7/8/2021.

Table 9 Summary of performance evaluation for environmental variables analysis.

Parameter	MAE (kWh)	RMSE (kWh)
Without average pressure	163.780	568.868
Without average temperature	212.792	586.694
Without average humidity	146.238	568.941
Without average wind speed	181.206	575.423
Without average rainfall duration	166.868	573.403
Without average rainfall amount	181.361	577.557

to 20 % in energy consumption on weekends and holidays. The building is mainly open on weekdays, Monday to Friday, from 9:00 a.m. to 5:00p.m. During the holiday and weekdays, the building is closed, and thus, the occupancy is low to none.

A variable can have a positive and negative impact on the model which leads to an increase or decrease in SHAP values. Subsequently, the humidity in a day may have a significant impact on the changes in the energy consumption of the building. Although humidity has a significantly lower SHAP value and it has a negative impact on energy usage based on SHAP analysis, research in [46] shows lower thermal comfort and higher temperature at high humidity. Temperature, wind speed, rainfall amount and rainfall duration are important

variables in increasing the performance of the model shown in Fig. 9. High outdoor temperature affects the thermal comfort of people in a building and thus increases energy consumption in controlling indoor temperature [48]. Furthermore, the neutral thermal comfort for people in Malaysia is 25.6 °C for mechanical cooling (CL) mode where HVAC was turned on for cooling purposes and 26.8 °C for free-running (FR) mode, where HVAC was turned off [49]. Wind speed may reduce the energy consumption of the building due to its influence in increasing the ventilation rate around the building or causing wind-speed rain [45]. Temperature is lower during the rainy season and thus offer better thermal comfort condition [50].

6. Conclusion

The LSTM model is developed to forecast the energy consumption of an institutional building. The model uses the previous year’s energy data and forecasted weather as the input parameter to forecast the next day. The energy consumption data will be changed depending on whether the next day is a weekday, a weekend or a holiday or the type of lockdown is no MCO, MCO or RMCO. In comparison to SVR and GPR, the LSTM model is able to forecast better. LSTM has better MAE and RMSE than other algorithms. Even though there is a huge change in energy consumption in 2020 and onwards due to the pandemic, LSTM achieves more accurate

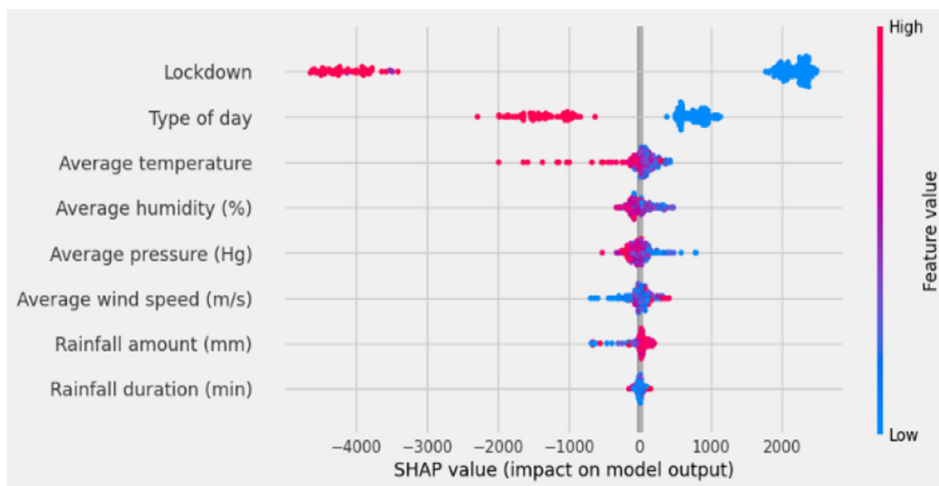


Fig. 8 Feature analysis using SHAP summary plot.

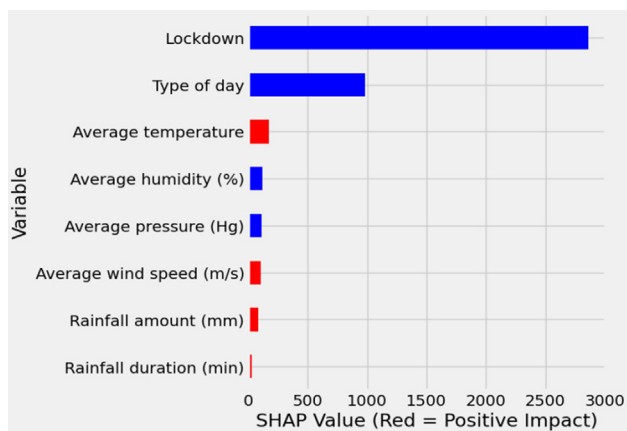


Fig. 9 Feature analysis using SHAP bar graph summary plot.

results. From the feature analysis, we can conclude that temperature and rainfall amount are important parameters in training the predictive model based on the MAE and RMSE findings. Using SHAP tools, we understood that type of lockdown (i.e., no MCO, MCO or RMCO) and type of day (i.e., weekday, weekend or holiday) are the most influential parameters in either increasing or decreasing the accuracy of the model. However, the precision of the day-ahead weather forecasting parameters is crucial to the prediction accuracy. However, LSTM requires huge historical data to predict accurately. In addition, LSTM requires external variables such as environmental variables (i.e. temperature, and wind speed), and schedule-related variables to further improve the accuracy of the predictive model. Future work could be able to improve the accuracy of the model and add more features to the model such as the occupancy of data (i.e., number of occupants).

CRediT authorship contribution statement

Muhammad Faiq: Conceptualization, Methodology, Software, Validation, Investigation, Formal analysis, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Kim Geok Tan:** Conceptualization, Methodology, Validation, Formal analysis, Writing – review & editing, Supervision. **Chia Pao Liew:** Writing – review & editing. **Ferdous Hossain:** Conceptualization, Methodology, Software, Validation, Formal analysis, Writing – review & editing. **Chih-Ping Tso:** Writing – review & editing. **Li Li Lim:** Writing – review & editing. **Adam Yoon Khang Wong:** Writing – review & editing. **Zulhilm Mohd Shah:** Writing – review & editing.

Declaration of Competing Interest

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

Acknowledgment

The research was supported by the Malaysian Meteorological Department (MET). Thank you to MET for providing the weather variables that were used. Gratitude to Multimedia University, Malacca Campus, for providing the data on energy

consumption. It was supported in part by the Mobile IoT: Location Aware project (Grant no. MMUE/180025) and Indoor Internet of Things (IoT) Tracking Algorithm Development based on Radio Signal Characterization project (Grant no. FRGS/1/2018/TK08/MMU/02/1). A special thanks also go to Telekom Malaysia for funding the research (under the project title, “Transparent Filtering Antenna”, Project ID MMUE 2200012).

References

- [1] I. I. E. Agency, *Electricity Market Report - Jan 2022*.
- [2] I. E. A. International Energy Agency, *Global Energy Review 2021*.
- [3] W. M. O. World Meteorological Organization, *WMO Provisional Report on the State of the Global Climate 2021*, 2021.
- [4] S. T. Suruhanjaya Tenaga, “Malaysia Energy Statistic Handbook 2020,” 2020.
- [5] J.S. Hassan, R.M. Zin, M.Z. Abd Majid, S. Balubaid, M.R. Hainin, *Building Energy Consumption in Malaysia: An Overview*, *Jurnal Teknologi* 70 (October 2014).
- [6] R. Saidur, *Energy consumption, energy savings, and emission analysis in Malaysian office buildings*, *Energy Policy* 37 (2009) 4104–4113.
- [7] M. Khoshbakht, Z. Gou, K. Dupre, *Energy use characteristics and benchmarking for higher education buildings*, *Eng. Build.* 164 (2018) 61–76.
- [8] K. P. M. Kementerian Pelajaran Malaysia, “Malaysia Educational Statistic 2020,” 2020.
- [9] C. Scheuer, G.A. Keoleian, P. Reppe, *Life cycle energy and environmental performance of a new university building: modeling challenges and design implications*, *Eng. Buildings* 35 (2003) 1049–1064.
- [10] A. Fouquier, S. Robert, F. Suard, L. Stéphan, A. Jay, *State of the art in building modelling and energy performances prediction: A review*, *Renew. Sustain. Energy Rev.* 23 (2013) 272–288.
- [11] A.S. Ahmad, M.Y. Hassan, M.P. Abdullah, H.A. Rahman, F. Hussin, H. Abdullah, R. Saidur, *A review on applications of ANN and SVM for building electrical energy consumption forecasting*, *Renew. Sustain. Energy Rev.* 33 (2014) 102–109.
- [12] L. Ekonomou, *Greek long-term energy consumption prediction using artificial neural networks*, *Energy* 35 (2010) 512–517.
- [13] C. Deb, L.S. Eang, J. Yang, M. Santamouris, *Forecasting diurnal cooling energy load for institutional buildings using Artificial Neural Networks*, *Eng. Buildings* 121 (2016) 284–297.
- [14] C. Robinson, B. Dilkina, J. Hubbs, W. Zhang, S. Guhathakurta, M.A. Brown, R.M. Pendyala, *Machine learning approaches for estimating commercial building energy consumption*, *Appl. Energy* 208 (2017) 889–904.
- [15] Z. Wang, Y. Wang, R. Zeng, R.S. Srinivasan, S. Ahrentzen, *Random Forest based hourly building energy prediction*, *Eng. Buildings* 171 (2018) 11–25.
- [16] G.K.F. Tso, K.K.W. Yau, *Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks*, *Energy* 32 (2007) 1761–1768.
- [17] W. Kong, Z.Y. Dong, Y. Jia, D.J. Hill, Y. Xu, Y. Zhang, *Short-Term Residential Load Forecasting Based on LSTM Recurrent Neural Network*, *IEEE Trans. Smart Grid* 10 (2019) 841–851.
- [18] S. Bouktif, A. Fiaz, A. Ouni, M.A. Serhani, *Optimal Deep Learning LSTM Model for Electric Load Forecasting using Feature Selection and Genetic Algorithm: Comparison with Machine Learning Approaches*, *Energies*, vol. 11, 2018.
- [19] X. Qing, Y. Niu, *Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM*, *Energy* 148 (2018) 461–468.

- [20] J. Runge, R. Zmeureanu, A Review of Deep Learning Techniques for Forecasting Energy Use in Buildings, *Energies*, vol. 14, 2021.
- [21] F. Ascione, N. Bianco, C.D. Stasio, G.M. Mauro, G.P. Vanoli, Artificial neural networks to predict energy performance and retrofit scenarios for any member of a building category: A novel approach, *Energy* 118 (2017) 999–1017.
- [22] U. N. United Nation, “COP26 The Glasgow Climate Pact,” 2021.
- [23] S. Fathi, R. Srinivasan, A. Fenner, S. Fathi, Machine learning applications in urban building energy performance forecasting: A systematic review, *Renew. Sustain. Energy Rev.* 133 (2020) 110287.
- [24] E. Wurtz, L. Mora, C. Inard, An equation-based simulation environment to investigate fast building simulation, *Build. Environ.* 41 (2006) 1571–1583.
- [25] S. Naji, A. Keivani, S. Shamshirband, U.J. Alengaram, M.Z. Jumaat, Z. Mansor, M. Lee, Estimating building energy consumption using extreme learning machine method, *Energy* 97 (2016) 506–516.
- [26] P.A. González, J.M. Zamarreño, Prediction of hourly energy consumption in buildings based on a feedback artificial neural network, *Energy. Build.* 37 (2005) 595–601.
- [27] E. Znouda, N. Ghrab-Morcous, A. Hadj-Alouane, Optimization of Mediterranean building design using genetic algorithms, *Energy. Build.* 39 (2007) 148–153.
- [28] J. Teeter, M.-Y. Chow, Application of functional link neural network to HVAC thermal dynamic system identification, *IEEE Trans. Ind. Electron.* 45 (1998) 170–176.
- [29] D. Tuhus-Dubrow, M. Krarti, Genetic-algorithm based approach to optimize building envelope design for residential buildings, *Build. Environ.* 45 (2010) 1574–1581.
- [30] S. Karatasou, M. Santamouris, V. Geros, Modeling and predicting building’s energy use with artificial neural networks: Methods and results, *Energy. Build.* 38 (2006) 949–958.
- [31] F. Kratzert, D. Klotz, C. Brenner, K. Schulz, M. Herrnegger, Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks, *Hydrol. Earth Syst. Sci.* 22 (2018) 6005–6022.
- [32] C. Tian, J. Ma, C. Zhang, P. Zhan, A Deep Neural Network Model for Short-Term Load Forecast Based on Long Short-Term Memory Network and Convolutional Neural Network, *Energies*, 11, 2018.
- [33] L. Peng, S. Liu, R. Liu, L. Wang, Effective long short-term memory with differential evolution algorithm for electricity price prediction, *Energy* 162 (2018) 1301–1314.
- [34] A.C. Tsoi, A. Back, Discrete time recurrent neural network architectures: A unifying review, *Neurocomputing* 15 (1997) 183–223.
- [35] N. Somu, R. Gauthama, M. Raman, K. Ramaritham, A hybrid model for building energy consumption forecasting using long short term memory networks, *Appl. Energy* 261 (2020) 114131.
- [36] Z. Cen, J. Wang, Crude oil price prediction model with long short term memory deep learning based on prior knowledge data transfer, *Energy* 169 (2019) 160–171.
- [37] S. Ghimire, R.C. Deo, N. Raj, J. Mi, Deep solar radiation forecasting with convolutional neural network and long short-term memory network algorithms, *Appl. Energy* 253 (2019) 113541.
- [38] J. Pan, R. Jain, S. Paul, T. Vu, A. Saifullah, M. Sha, An Internet of Things Framework for Smart Energy in Buildings: Designs, Prototype, and Experiments, *IEEE Internet Things J.* 2 (2015) 527–537.
- [39] Q. Li, Q. Meng, J. Cai, H. Yoshino, A. Mochida, Applying support vector machine to predict hourly cooling load in the building, *Appl. Energy* 86 (2009) 2249–2256.
- [40] M.R. Braun, H. Altan, S.B.M. Beck, Using regression analysis to predict the future energy consumption of a supermarket in the UK, *Appl. Energy* 130 (2014) 305–313.
- [41] J. Zhang, Y. Zhu, X. Zhang, M. Ye, J. Yang, Developing a Long Short-Term Memory (LSTM) based model for predicting water table depth in agricultural areas, *J. Hydrol.* 561 (2018) 918–929.
- [42] X. Lü, T. Lu, C.J. Kibert, M. Viljanen, Modeling and forecasting energy consumption for heterogeneous buildings using a physical–statistical approach, *Appl. Energy* 144 (2015) 261–275.
- [43] L. Li, L. Qin, Z. Xu, Y. Yin, X. Wang, B. Kong, J. Bai, Y. Lu, Z. Fang, Q., Song and others, “Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT,” *Radiology* 296 (2020) E65–E71.
- [44] G.E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, R.R. Salakhutdinov, Improving neural networks by preventing co-adaptation of feature detectors, *CoRR*, vol. abs/107.0580, 2012.
- [45] M. Abuku, H. Janssen, S. Roels, Impact of wind-driven rain on historic brick wall buildings in a moderately cold and humid climate: Numerical analyses of mould growth risk, indoor climate and energy consumption, *Energy. Build.* 41 (2009) 101–110.
- [46] H. Djamila, C.-M. Chu, S. Kumaresan, et al, Effect of humidity on thermal comfort in the humid tropics, *J. Build. Constr. Planning Res.* 2 (2014) 109.
- [47] S.M. Praveena, A.Z. Aris, The impacts of COVID-19 on the environmental sustainability: a perspective from the Southeast Asian region, *Environ. Sci. Pollut. Res.* 28 (2021) 63829–63836.
- [48] J.F. Nicol, M.A. Humphreys, Adaptive thermal comfort and sustainable thermal standards for buildings, *Energy. Build.* 34 (2002) 563–572.
- [49] S.A. Zaki, S.A. Damiaty, H.B. Rijal, A. Hagishima, A.A. Razak, Adaptive thermal comfort in university classrooms in Malaysia and Japan, *Build. Environ.* 122 (2017) 294–306.
- [50] W. Liping, W.N. Hien, The impacts of ventilation strategies and facade on indoor thermal environment for naturally ventilated residential buildings in Singapore, *Build. Environ.* 42 (2007) 4006–4015.

MUHAMMAD FAIQ BIN MUHAMAD FISOL is currently pursuing the Bachelor’s degree in Electronics Engineering majoring in telecommunications at the Faculty of Engineering and Technology (FET), Multimedia University (MMU) Melaka, Malaysia. He is in the final year of his degree and is expected to graduate in March 2022. His research interests include Energy Efficiency Prediction, Deep learning, and Machine learning.

TAN KIM GEOK received the B.E., M.E., and PhD. degrees from Universiti Teknologi Malaysia, in 1995, 1997, and 2000, respectively, all in electrical engineering. He has been a Senior R&D Engineer with EPCOS, Singapore, in 2000. From 2001 to 2003, he joined DoCoMo Euro-Labs, Munich, Germany. He is currently an Academic Staff with Multimedia University. His research interests include radio propagation for outdoor and indoor, RFID, multiuser detection technique for multicarrier technologies, and A-GPS.

LIEW CHIA PAO received his PhD in Engineering Education from Universiti Teknologi Malaysia, and a Master’s Degree in Electronics (Communication Systems) from the University of Hertfordshire, United Kingdom. He is currently the Assistant Dean and the Assistant Professor at the School of Energy and Chemical Engineering at Xiamen University Malaysia. His publications include refereed journals and conference papers in the domains of engineering education and wireless communication.

FERDOUS HOSSAIN received the B.Sc. degree in computer science and engineering and the M.Sc. degree in information and communi-

cation technology from Mawlana Bhashani Science and Technology University, Bangladesh, in 2012 and 2015, respectively, and the PhD degree in engineering from the Faculty of Engineering and Technology, Multimedia University, Malaysia, in 2019. He was a recipient of the Best PhD thesis award for the Doctor of Philosophy (Engineering) Award in 2020. He has authored and co-authored several international journals and conference papers. His research areas include radio propagation, next-generation wireless communication, image processing, machine learning, deep learning and big data analysis.

TSO CHIH-PING, a Malaysian engineering educator, and researcher received his Bachelor's degree from Loughborough in 1968, through a National Electricity Board scholarship, Master's from Massachusetts Institute of Technology through research assistantship, and Ph.D. from the University of California, Berkeley, in 1979, through a Fulbright-Hays scholarship. He has worked at the University of Malaya, Nanyang Technological University, and Multimedia University.

LIM LI LI received her Advanced Diploma in Technology (Electronics Engineering) from Tunku Abdul Rahman College (TARC) and subsequently obtained a M. Sc. in Electronics from the Queen's University of Belfast. In 2019, she obtained her PhD in Electrical and Electronics Engineering from the University of Nottingham Malaysia in the area

of MIMO channel coding and equalisation. Her current research interest is in statistical signal and image processing, machine learning, and next-generation wireless communication signal processing.

ADAM WONG YOON KHANG received his Ph.D. Degree from Universiti Teknologi Malaysia in 2018. He is currently a Senior Lecturer in the Fakulti Teknologi Kejuruteraan Elektrik dan Elektronik, Universiti Teknikal Malaysia Melaka (UTeM). He is also the Professional Technologist for the Malaysia Board of Technologists (MBOT). His current research interests are the Internet of Things, Hybrid Optical Wireless, simulation optimization, ad hoc network, and passive optical network but not limited to the mentioned topic here. He actively publishes research articles and received several research grants from the government and private sectors, universities and international collaboration.

ZULHILMI MOHD SHAH received his Degree in Electrical Engineering from Universiti Teknologi Malaysia in 2005. He is currently an Assistant Manager in the Facilities Management Department (FMD), Multimedia University, Melaka. He is also the Registered Electrical Energy Manager with Suruhanjaya Tenaga, Certified Energy Manager from AEMAS and Energy Management System Expert from UNIDO. His scope of work in FMD is Energy, Maintenance and Development.