TIME-SERIES BASED HOUSEHOLD ELECTRICITY CONSUMPTION FORECASTING

ABSTRACT

Data analytics using machine learning technologies when applied to the energy consumption data can provide valuable inputs for maintaining the perfect supply demand balance in a smart electrical grid system. In particular, the accurate predictions of energy consumption for future periods of time aids significantly in cost-cutting and energy saving for utility companies. Making use of the popular method of time-series forecasting and the Artificial Neural Networks (ANN) models, here in this paper, one of the variants of the Recurrent Neural Networks (RNN) model, the Long Short Term Memory (LSTM) model is applied for household electricity consumption forecasting. Real datasets from consumption building are used for experimenting the model and applied through Tensorflow platform with the keras functions in Python. The results obtained show significantly accurate values in predicting future consumption derived from models training with actual values of current consumption. Hence, this work provides yet another proof that the LSTM machine learning forecasting methods can be efficiently applied for household electricity forecasting.

1. INTRODUCTION

As an essential part of the SG systems and with the successful operations of the smart metering infrastructure in the household communities, there arises a scenario where different types of huge volumes of high velocity data is generated. The traditional ways of analysing these data does not prove to be useful for these enormous volumes of data which arrives in high frequencies from various heterogeneous sources. Hence applying the methods of big data analytics which is enriched with techniques like data mining, predictive and statistical analysis, artificial intelligence and machine learning approaches achieve the productive decision-making results in SGs. Big data analytical solutions can thus be applied to these enormous data from the disparate sources of the SG network and meaningful, real-time insights can be derived for the energy service utilities and control systems. Therefore, in general, the continuous data generated through smart consumer devices can be transformed into meaningful insights thereby discovering the necessary knowledge for optimizing business operations which can be illustrated as in Fig 1.

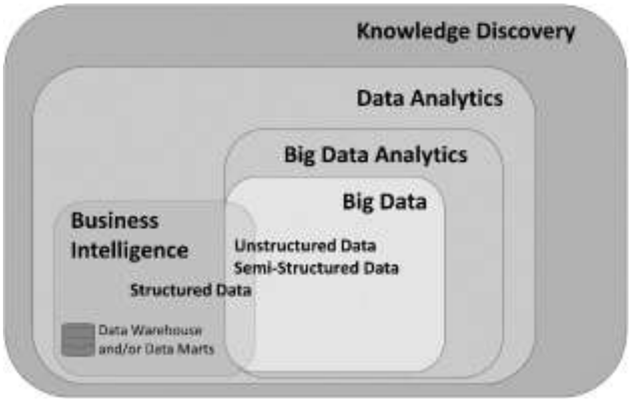

Fig 1 Relationship between data and business operations

In the geographically distributed systems like SGs with the capability of producing tremendous amount of data, the resulting data analytics insights greatly rely on the data processing speed, efficiency and accuracy. Hence, the timely computing and processing of data in such systems aid in delivering precise decision-making solutions. Considering the huge volume and velocity of big data generated from smart device, the data processing is generally carried out as

- Batch processing – A method of transmitting several sets of large volumes of stored data from different sources as a batch for processing, thus enabling more detailed analytics results.
- Stream processing – The data that is transmitted for processing are the information that arrives in a short period of time almost resembles that of real time data, providing immediate analytical results like fraud detection.

In the context of smart meters of the SG network, near real-time high-volume data is generated by the metering sources, like every few minutes.

The crucial data for electricity consumption analytics emerge from the consumer side smart devices. In the case of Smart electricity Grid systems, the Advanced Metering Infrastructure (AMI) plays a major role in generation of smart meter data. Intermediate nodes or data concentrators collect the smart meter data from individual households and further transmit that to meter data management system (MDMS) which in turn forwards the structured data to the analytical servers of the utility centre. Fig. 2 illustrates the various analytical applications executed in the utility end of the smart grids that influences business decisions. Forecasting of electricity consumption contributes majorly to the reliability and sustainability of smart grid systems and aids for an almost perfect balance in energy demand and supply.

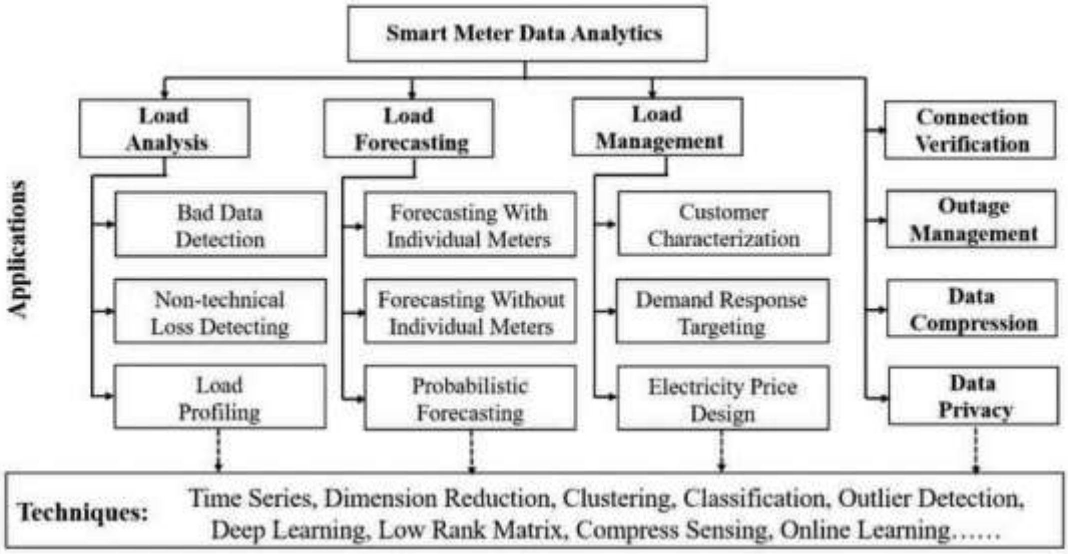

Fig. 2 Smart Meter Data Analytics [1]

The formal statistical forecast methods depend on the different types of data collected such as

a) time series data that is collected over constant time intervals,

b) cross-sectional data that records various entities at a given point of time and

c) longitudinal data that involves repeated observations of the same entity at different points of time.

Here, in this research work, a forecasting method using time series data is proposed for predicting short-term household electricity consumption. The proposed model shows accurate prediction results and is executed using Tensorflow / Keras functional API classes. The forecasting methods are classified according to the time-period over which the forecasts are predicted. This finite time-period is called as the forecasting horizon and can be broadly listed as very short term forecasting (VSTF), short term forecasting (STF), medium term forecasting (MTF) and long term forecasting (LTF) [2]. Forecasting solutions for time horizons below three hours form VSTF models. The typical applications for VSTF include forecasting renewable energy resources (RER) like solar and wind power generation [3]. STF is applied for applications with time horizon of few days, like grid reliability analysis and dispatch scheduling [4]. Solutions like system maintenance, financial risk management using consumption and price forecasting uses MTF that has time horizons up to a few months. LTF involves horizons up to a few years and applied for solutions like energy generation planning, trend analysis for long term, load growth analysis and is illustrated in Fig 3.

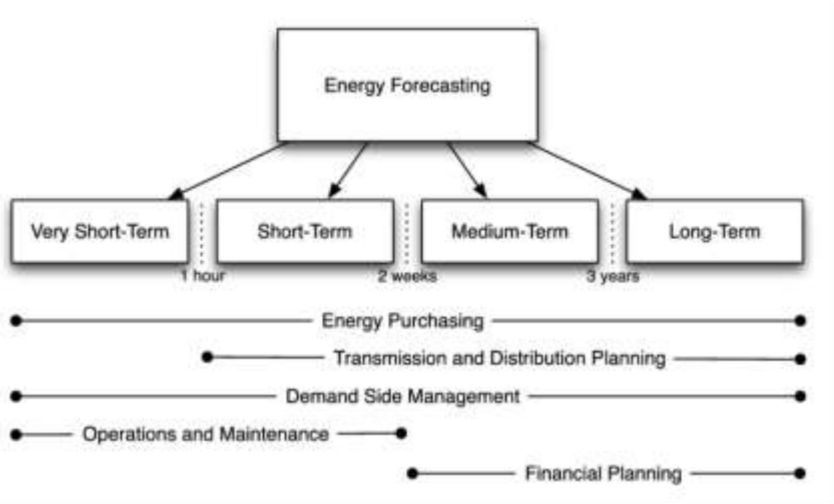

Fig 3 Classification of forecasting techniques based on time horizons [5]

Accordingly, Table 1 explains the forecasting types based on the time horizons and the applications along with the common forecasting technique in each category.

Table 1 Types of forecasting based on time horizons

Type of forecasting	**Application**	**Forecasting Techniques used**
Very short-term load forecasting (VSTF)	Renewable energy resources forecasting	Multiple regression Exponential smoothing Iteratively reweighted least-squares Adaptive load prediction Stochastic time series
Short term forecasting (STF)	Grid Operations	ARMA models Fuzzy logic ANN Expert systems
Medium term forecasting (MTF)	Planning for grid operations	ANN Particle swarm optimization Fuzzy logic Hybrid models
Long Term Forecasting (LTF)	Planning for system expansion	ANN Support vector machine Fuzzy logic Hybrid models

Energy forecasting techniques can be largely classified into Statistical methods, artificial intelligence (AI) methods and other innovative hybrid methods. The traditional statistical methods are simple and use the regression model on historical data collected on a time series fashion. Some techniques applied here include autoregressive moving average (ARMA), autoregressive integrated moving average (ARIMA) and vector auto regression (VAR). Machine learning (ML) methods use either supervised or unsupervised learning patterns from existing data and improve on prediction accuracy. The commonly applied ML techniques used in demand and generation forecasting include support vector machine (SVM), support vector regression (SVR) and Random Forest (RF) models. Deep learning (DL) methods use ANN that is defined as a collection of connected nodes in a sequence of layers. The variations in ANN methods like Feed forward neural networks (FFNN), back propagation neural networks (BPNN), recurrent neural networks (RNN), long short-term memory (LSTM) and convolutional neural networks (CNN) are some of commonly used techniques in DL forecasting. The ever-changing demand and growth of energy infrastructure, probabilistic forecasting methods are also existing to use particularly for RER generation and planning. To overcome the limitations of AI based methods, it is a common practice to use hybrid techniques with the ML and DL methods in combination with other methods to obtain optimization, however with an increase of computational complexity. Fig. 4 shows a brief classification of different forecasting techniques that can be applied for energy forecasting.

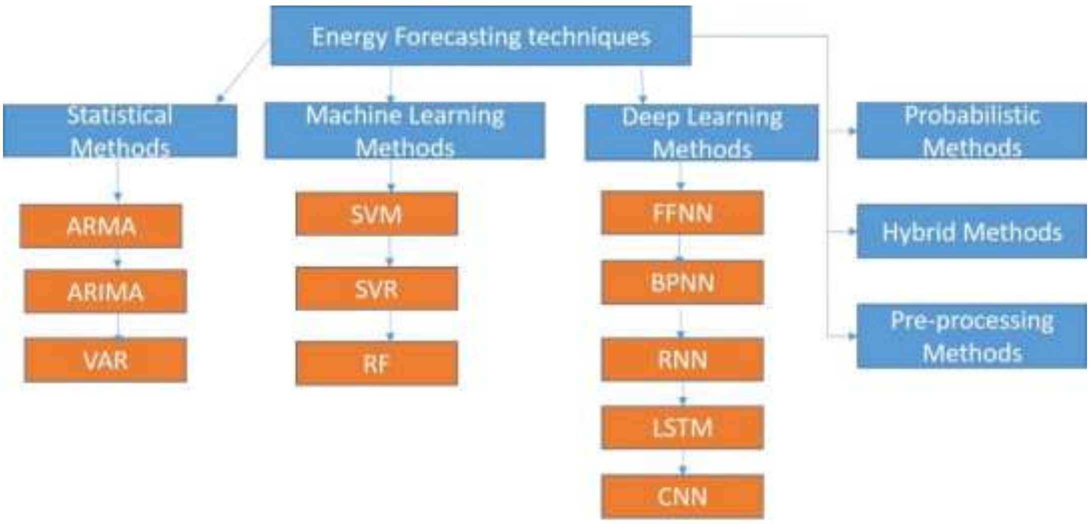

Fig. 4 Energy forecasting techniques

2. RELATED WORKS

Many forecasting methods have been proposed in literature for prediction of electricity consumption in household and industry usages. However, in general, machine learning models are usually applied for the purpose of load forecasting and electricity demand predictions. In [6], the authors have employed four machine learning methods namely Support Vector Regression (SVR), Gradient Boosting Regression Trees (GBRT), feedforward neural networks (FFNN) and Long Short-Term Memory (LSTM) for short term load forecasting of energy consumption in domestic buildings. These are data driven models and were trained merely on historical data recorded without external information and the performance measures are evaluated. Similarly, the authors in [7] have proposed four machine learning approaches namely nearest neighbor model, fuzzy neighborhood model, kernel regression and general regression neural network for mid-term load forecasting. These models are based on pattern similarity of time series pattern representations, that simplifies the forecasting complexity and have been analysed and compared with the classical forecasting models for performance metrics of accuracy and ease of optimization. The Prophet model for long term peak load forecasting has been investigated in the paper [8] and compared with the Holt–Winters model for feasibility, robustness and accuracy using real time data from Kuwait power plants. A bottom up approach using stock-and-flow model has been presented in [9] wherein long term forecasting of energy consumption of household appliances is proposed. Here, the prediction analysis of peak load results in the recommendation of usage of energy efficient equipment and standards. In the paper [10], a hybrid deep learning forecast model has been proposed that comprises of moving average, long short term memory (LSTM) and K-fold-correlation to provide peak load cost savings. This is achieved by forecasting the maximum load duration (MLD) based on time of use (TOU). Similarly using LSTM as a variant of recurrent neural network (RNN), a time series forecasting approach is presented in [11], which can be used for accurate short term prediction of individual energy consumption of households. Another work proposed in [12] also uses LSTM based RNN for accurate short term prediction of electricity load forecasting. A three layer LSTM forecasting model has also been proposed in [13] for short term electricity usage prediction in individual homes for hourly and daily prediction horizons.

A recurrent LSTM neural networks model is proposed in [14] to forecast short time energy load in a 24 hour prediction horizon. Here, the results prove to be successful in forecasting the irregular trends in time series along with high insensitivity to external elements. A hybrid model using auto-regressive integrated moving average (ARIMA) and RNN has been experimented in [15] wherein the seasonal component modelling technique and trend forecasting are used in combination to produce accurate results. A multi step time series load forecasting model using LSTM based RNN has been proposed in [16] and results show better performance when compared with ARIMA model of forecasting. In [17], an electricity demand estimation and long term forecast model has been studied and developed with real time data of Turkey using co-integration analysis and ARIMA modelling and proves to be more accurate than the current overestimated projections. Another long term electricity load prediction model has been successfully experimented in [18] using K-Means clustering and Artificial Neural Network (ANN) approach. A hybrid approach in [19] uses the ARIMA model for forecasting the linear load followed by the Support Vector Machines (SVM) model for correcting the deviation of the former forecast to achieve accurate predictions. A time series approach, singular spectral analysis (SSA) has been proposed in [20] for data analysis and forecasting of short term load and proves to be efficient in distinguishing and forecasting desired load time series.

3. NEURAL NETWORKS BASED FORECASTING

Forecasting using ANN models are widespread in many applications, owing to the reasons of non-linear data driven self-adaptive capabilities, learning from existing patterns and forming functional relationship between data. The working of the ANN approach for forecasting can be summarized as; the sequence of layers contain a collection of interconnected neurons which receive the input from other nodes or external stimuli, processes it and produces a transformed output. For a time-series forecasting solution, the ANN function mapping can be defined as shown in Equation 1

$$Y_{t+1} = f(Y_t,\ Y_{t-1}, Y_{t-2} \dots \dots \dots Y_{t-p}) \qquad [1]$$

whereY_t is the observation value at time t, p is the independent variable. Here, the input data is the historical information and the output is the prediction of future values. The training from historical data is performed by dividing the data into training set used for estimation and test set used for generalization. In a feed forward manner, the input values are transmitted through the intermediate hidden layers that is accumulated and fed as input to subsequent layers until the output is reached. Equation 2 shows the sum of Squared Errors (SSE) based cost function output E from N no. of training set observations Y_1, Y_2, ……Y_N with *n* input nodes resulting in N-n training patterns.

$$E = \frac{1}{2}\sum_{i=n+1}^{N}(y_i - a_i)^2 \quad [2]$$

where a_i is the actual output and the results are simplified the training algorithm computation [21]. The major advantages of using ANN for forecasting include data storage throughout the network, performance ability with lack of sufficient in formation, fault tolerance, distributed memory, uses unsupervised learning, efficient training patterns and parallel processing. ANNs are powerful models for forecasting to produce robust alternative for traditional forecasting methods especially in complex non-linear relationships. RNN is a special type of ANN model where the state of previous information is stored in memory, and there are inter dependencies between time series data points. For a neural network with input layers *x* and hidden layers *h*, the weight *w* is passed on with the inputs from the neurons of the input and hidden layers along with the bias *b*, then the first computed output *Y* is shown in Equation 3 and Equation 4.

$$h_{t+1} = f(w_x x_t + w_h h_t + b_h) \quad [3]$$

$$y_t = f(w_y . h_t + b_y) \quad [4]$$

with activation function *f*, x_t is the input at time step *t*, h_t is the hidden state at time *t*, y_t is the output of the network at time step *t*. w_x is the weight associated with the inputs in recurrent layer, w_h is the weight associated with the hidden units in recurrent layer, w_y is the weight associated with hidden to output layer and b_h is the bias associated with the recurrent layer [22]. The training of the weights of the network is carried out by using the Back Propagation algorithm of ANN to include the unfolding in time. The benefits of using RNN for forecasting include memory to store historical data, handle inputs of varying lengths and handle sequence data. The shortcoming of RNN, vanishing gradient problem occurs where the gradient that measures the change in all weights with regard to change error becomes too low and the learning process stops. Another similar issue in the RNN is exploding gradient where the gradient becomes too large. This is overcome by LSTM by enabling the network to remember the inputs for long time periods resulting in short trainings and more accuracy. The LSTM uses different activation function layers called input, output and forget gates to decide which information to be retained. The working of the LSTM model forecasting is explained in Equation 5, Equation 6 and Equation 7.

$$y^{out_j}(t) = f_{out_j}\left(net_{out_j}(t)\right) \quad [5]$$

$$y^{in_j}(t) = f_{in_j}\left(net_{in_j}(t)\right) \quad [6]$$

$$net_{out_j}(t) = \sum_u w_{out_j u}\, y^u(t-1),\ net_{in_j}(t) = \sum_u w_{in_j u}\, y^u(t-1) \qquad [7]$$

where out_j is the output gate, in_j is the input gate at j[th] memory cell, $y^{out_j}(t)$ is the activation of out_j at time t, $y^{in_j}(t)$ is the activation of in_j at time t, u may denote input, gate memory or even hidden cells [23]. The advantages of LSTM method include learning from previous time stamp information, greater accuracy, overcomes gradient problems of RNN, and more useful for datasets using long sequences.

4. PROPOSED LSTM MODEL BASED PREDICTION

4.1 Methodology

The dataset used has been extracted from Kaggle community service; the particular dataset selected for experimentation contains electricity consumption information of a residential building in Thailand. The consumption data collected are in 1-minute intervals from all the individual units and plug loads present in various floors of the building. The method applied for collecting data is through the Building Energy Management System (BEMS) comprising of smart maters, energy monitoring units, sensors and the central server [24]. The initial step is the visualization of raw data that helps in showing the presence of anomalies and identifying the distinct patterns. Then the correlation between the different features are mapped using heat-map. During the parameter selection, the redundant parameters are eliminated. The step that follows is the normalization and pre-processing of data which is essential for defining training and validation datasets and for obtaining accuracy. During the pre-processing, the split fraction is set to 0.915 to enable majority of data to be used. Since the consumption data is recorded every 1 minute, the step size is set to 60 for an hour, with a tracking data for 7200 timestamps and the prediction is evaluated for 720 timestamps. Then the training and validation datasets are defined. Subsequently the keras model is compiled for training with saving the intermediate results in checkpoints. The accuracy of training is measured with the validation and training loss graph. Finally, four sets of prediction values are plotted. The sequence of the proposed method for energy consumption forecasting using time-series LSTM model is illustrated in the form of a flow chart as in Fig. 5.

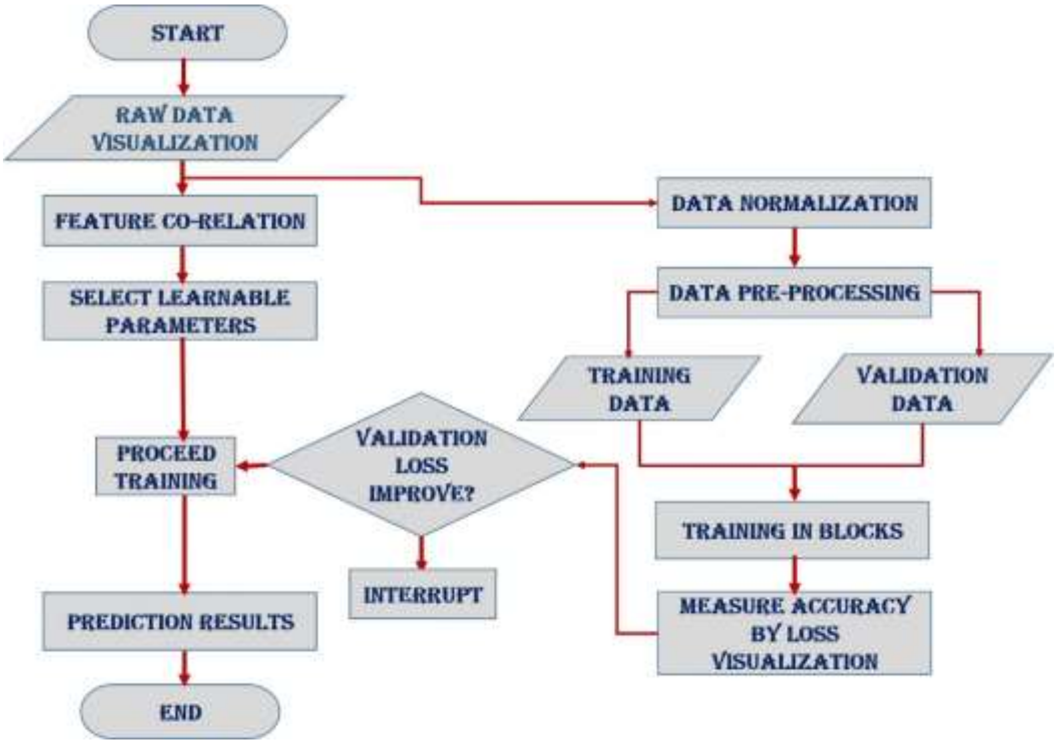

Fig 5 Flow chart of proposed LSTM forecasting method

The data analytical functions when applied to the electricity consumption data provides efficient input for energy supply decisions. In this research work, a time-series forecasting model using a variant of RNN is applied with LSTM training methodology. The steps followed in the method are, after the initial data normalization and pre-processing, the training and validation datasets are defined, followed by model training with the selected parameters.

5. RESULTS AND DISCUSSION

The plot of the Training and Validation loss as obtained from the execution of the code of the proposed method is shown in Fig 6. Accordingly, the below graph shows the relativity of fitting the model to training data. As seen from the plot, the training loss achieved stability and the validation loss decreases initially and after certain epochs reaches stability having lesser gap with the training loss. Hence, the model is fitting almost optimally, with the slightest chance of overfitting.

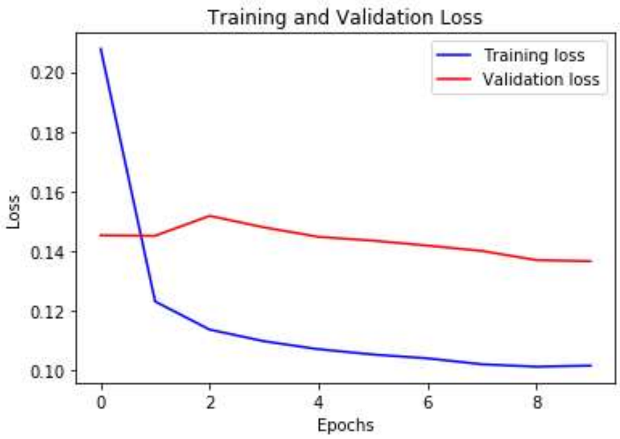

Fig. 6 Training and Validation Loss

A good fit model has relatively closer loss curves, with the validation loss higher than the training loss and after some epochs, both losses should ideally reach flat values [25]. Accordingly, considering the practical aspects of energy consumption and the inconsistency in customer behaviors, the available data and the proposed model provides a reasonably good fit for the training dataset. Further, the single step predictions are plotted for four days which is illustrated in Fig.7, Fig.8, Fig.9, Fig.10. As seen from the prediction plots, the actual future values are closely matching with the model predictions.

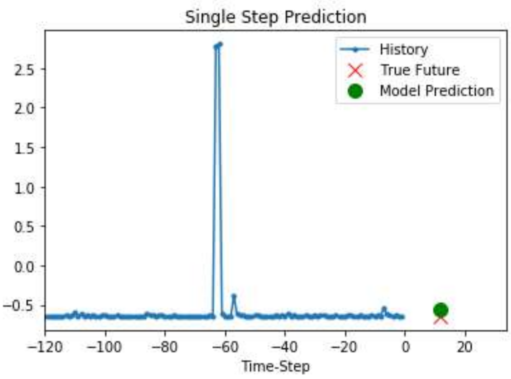

Fig. 7 Single step prediction Result Day 1

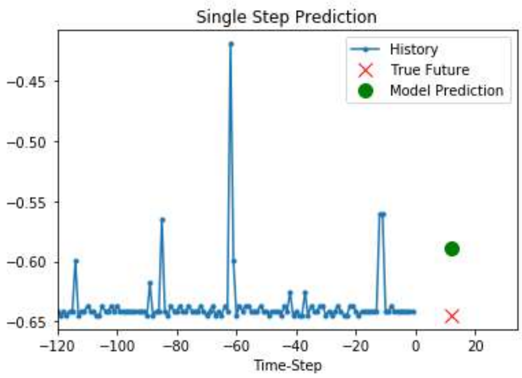

Fig. 8 Single step prediction Result Day 2

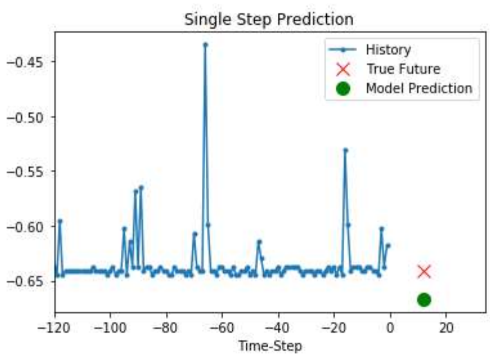

Fig. 9 Single step prediction Result Day 3

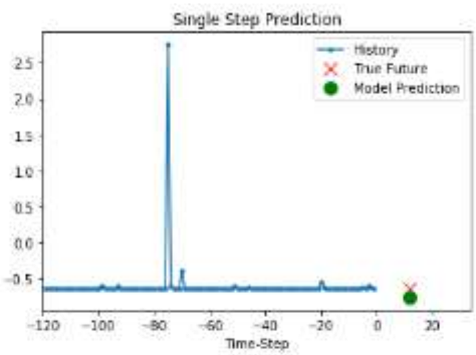

Fig. 10 Single step prediction Result Day 4

The line plots showing the predictions illustrated the comparison of actual true values with the prediction values obtained from past observations. For day one and four, the model predictions almost exactly match the actuals, implying an efficient training of parameters. Thus, we are able to prove from the proposed methodology of LSTM forecasting for energy consumption that the application of keras APIs enable faster and accurate prediction results. Finally, the validation loss is evaluated against the training loss, and a set of single step predictions are plotted. The execution of the steps are carried out through Tensorflow platform with keras functions and the anomaly detection results are shown in Fig 11

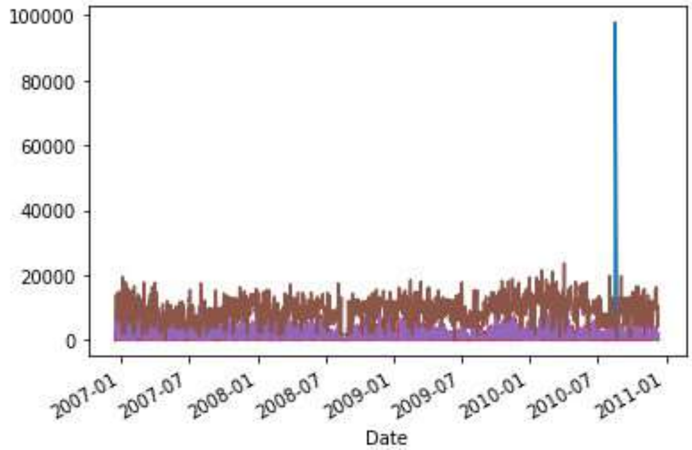

Fig 11 Anomaly detection results

The data analytical functions when applied to the electricity consumption data provides efficient input for energy supply decisions. In this paper, a time-series forecasting model using a variant of RNN is applied with LSTM training methodology. The steps followed in the method are, after the initial data normalization and pre-processing, the training and validation datasets are defined, followed by model training with the selected parameters. Finally, the validation loss is evaluated against the training loss, and a set of single step predictions are plotted. The execution of the steps is carried out through Tensorflow platform with keras functions. From the results obtained after experimentation with real datasets, it is found that the prediction of future values closely matches with true values and the accuracy is significant. Hence, we propose that the LSTM model of forecasting can be successfully used for energy consumption prediction. In the future, the model can be expanded to multi step predictions for large datasets, which can be used for planning renewable energy resources management. From the results obtained after experimentation with real datasets, it is found that the prediction of future values closely matches with true values and the accuracy is significant. Hence, we propose that the LSTM model of forecasting with appropriate pre-processing can be successfully used for energy consumption prediction. In the future, the model can be expanded to multi step predictions for large datasets, which can be used for planning renewable energy resources management.

6. CONCLUSION

In this paper, a time-series forecasting model using a variant of RNN were applied with LSTM training methodology. After the initial data normalization and pre-processing, the training and validation datasets are defined, followed by model training with the selected parameters. Finally, the validation loss was evaluated against the training loss, and a set of single step predictions are plotted. The execution of the steps was carried out through Tensorflow platform with keras functions. From the results obtained after experimentation with real datasets, it is found that the prediction of future values closely matches with true values and the accuracy is significant. Hence, the proposed LSTM model of forecasting can be successfully used for energy consumption prediction.

REFERENCES

[1] Y. Wang, Q. Chen, T. Hong, and C. Kang, "Review of Smart Meter Data Analytics: Applications, Methodologies, and Challenges," Feb. 2018, doi: 10.1109/TSG.2018.2818167.

[2] D. Kaur, S. N. Islam, M. A. Mahmud, and Z. Dong, "Energy Forecasting in Smart Grid Systems: A Review of the State-of-the-art Techniques," pp. 1–28, 2020, [Online]. Available: http://arxiv.org/abs/2011.12598.

[3] C. Guan, P. B. Luh, L. D. Michel, and Z. Chi, "Hybrid Kalman filters for very short-term load forecasting and prediction interval estimation," *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 3806–3817, 2013, doi: 10.1109/TPWRS.2013.2264488.

[4] J. C. Lopez, M. J. Rider, and Q. Wu, "Parsimonious Short-Term Load Forecasting for Optimal Operation Planning of Electrical Distribution Systems," *IEEE Trans. Power Syst.*, vol. 34, no. 2, pp. 1427–1437, 2019, doi: 10.1109/TPWRS.2018.2872388.

[5] O. Timur and A. Science, "Interpretation of Error Calculation Methods in the Context of Energy Forecastin," *12th Conf. Sustain. Dev. Energy, Water Environ. Syst.*, no. October, 2017, [Online]. Available: https://www.researchgate.net/publication/320310871_Interpretation_of_Error_Calculat

ion_Methods_in_the_Context_of_Energy_Forecasting.

[6] A. M. Pirbazari, M. Farmanbar, A. Chakravorty, and C. Rong, "Short-term load forecasting using smart meter data: A generalization analysis," *Processes*, vol. 8, no. 4, 2020, doi: 10.3390/PR8040484.

[7] G. Dudek and P. Pełka, "Pattern similarity-based machine learning methods for mid-term load forecasting: A comparative study," *Appl. Soft Comput.*, vol. 104, 2021, doi: 10.1016/j.asoc.2021.107223.

[8] A. I. Almazrouee, A. M. Almeshal, and A. S. Almutairi, "applied sciences Long-Term Forecasting of Electrical Loads in Kuwait," *Appliend Sci.*, vol. 5627, no. 10, pp. 2–17, 2020.

[9] T. Alajmi and P. Phelan, "Modeling and forecasting end-use energy consumption for residential buildings in Kuwait using a bottom-up approach," *Energies*, vol. 13, no. 8, 2020, doi: 10.3390/en13081981.

[10] J. Kim, B. Shah, and K. Il Kim, "Hybrid Deep Learning Architecture to Forecast Maximum Load Duration Using Time-of-Use Pricing Plans," *Comput. Mater. Contin.*, vol. 68, no. 1, pp. 283–301, 2021, doi: 10.32604/cmc.2021.016042.

[11] A. M. Alonso, F. J. Nogales, and C. Ruiz, "A single scalable LSTM model for short-term forecasting of massive electricity time series," *Energies*, vol. 13, no. 20, pp. 1–19, 2020, doi: 10.3390/en13205328.

[12] J. Zheng, C. Xu, Z. Zhang, and X. Li, "Electric load forecasting in smart grids using Long-Short-Term-Memory based Recurrent Neural Network," *2017 51st Annu. Conf. Inf. Sci. Syst. CISS 2017*, pp. 1–6, 2017, doi: 10.1109/CISS.2017.7926112.

[13] R. E. Alden, H. Gong, C. Ababei, and D. M. Ionel, "LSTM Forecasts for Smart Home Electricity Usage," *9th Int. Conf. Renew. Energy Res. Appl. ICRERA 2020*, no. Icrera, pp. 434–438, 2020, doi: 10.1109/ICRERA49962.2020.9242804.

[14] T. Ciechulski and S. Osowski, "High precision lstm model for short-time load forecasting in power systems," *Energies*, vol. 14, no. 11, pp. 1–15, 2021, doi: 10.3390/en14112983.

[15] P. G. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, pp. 159–175, 2003, doi: 10.1016/S0925-2312(01)00702-0.

[16] S. Masum, Y. Liu, and J. Chiverton, "Multi-step time series forecasting of electric load using machine learning models," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10841 LNAI, pp. 148–159, 2018,

doi: 10.1007/978-3-319-91253-0_15.

[17] E. Erdogdu, "Munich Personal RePEc Archive Electricity Demand Analysis Using Cointegration and ARIMA Modelling : A case study of Turkey," *Energy Mark. Regul. Auth.*, no. 19099, 2007.

[18] B. Shelly and M. Shivani, "Long term load forecasting using K-mean clustering and ANN approach," *Int. J. Adv. Res.*, vol. 4, no. 3, 2018, [Online]. Available: www.IJARIIT.com.

[19] H. Nie, G. Liu, X. Liu, and Y. Wang, "Hybrid of ARIMA and SVMs for short-term load forecasting," *Energy Procedia*, vol. 16, no. PART C, pp. 1455–1460, 2012, doi: 10.1016/j.egypro.2012.01.229.

[20] K. Afshar and N. Bigdeli, "Data analysis and short term load forecasting in Iran electricity market using singular spectral analysis (SSA)," *Energy*, vol. 36, no. 5, pp. 2620–2627, 2011, doi: 10.1016/j.energy.2011.02.003.

[21] G. Zhang, B. Eddy Patuwo, and M. Y. Hu, "Forecasting with artificial neural networks: The state of the art," *Int. J. Forecast.*, vol. 14, no. 1, pp. 35–62, 1998, doi: 10.1016/S0169-2070(97)00044-7.

[22] "An Introduction To Recurrent Neural Networks And The Math That Powers Them." https://machinelearningmastery.com/an-introduction-to-recurrent-neural-networks-and-the-math-that-powers-them/ (accessed Feb. 21, 2022).

[23] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.1735.

[24] "CU-BEMS, smart building energy and IAQ data | Kaggle." https://www.kaggle.com/claytonmiller/cubems-smart-building-energy-and-iaq-data/version/1 (accessed Nov. 23, 2021).

[25] "Learning Curve to identify Overfitting and Underfitting in Machine Learning | by KSV Muralidhar | Towards Data Science." https://towardsdatascience.com/learning-curve-to-identify-overfitting-underfitting-problems-133177f38df5 (accessed Nov. 24, 2021).