Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 7.056

IJCSMC, Vol. 14, Issue. 11, November 2025, pg.118 – 125

Mongolian Energy Consumption Prediction using Long and Short-Term Memory Model

Otgonzul Zorigt[1]; Seungjae Lee[2*]

[1, 2] Computer Engineering Department, Sunmoon University, Korea
[1] otgonzul.udachi2020@gmail.com; [2*](corresponding author) leeko@sunmoon.ac.kr

DOI: https://doi.org/10.47760/ijcsmc.2025.v14i11.008

Abstract: This study aims to forecast Mongolia's energy consumption using data from the national energy system. The goal is to determine whether a machine learning model can produce sufficiently accurate predictions for a complex, seasonal forecasting task, by exploring various ML techniques and building a data-driven forecasting model. The dataset comprises 54 years of monthly Mongolia energy consumption and is a univariate time series. An LSTM model was trained on the data, and model performance was assessed using RMSE to facilitate direct comparison with observed energy values.

Keywords: time series forecasting, recurrent neural network, energy consumption, deep learning

I. INTRODUCTION

Mongolian electric power (EP) consumption is continuously increasing year by year [1], which corresponds to GDP and population growth [2]. It is well known EP consumption depends on not only economic conditions but also social activity and weather conditions [3]. Since there is a fact that EP demand is increasing and most of EP demand is supplied by coal-fired thermal power plants (TPP) in Mongolia, introducing renewable energy sources and expanding the EP system with higher reliability are expected [4]. Furthermore, the Mongolian EP system is vulnerable due to the slow response of TPP [5]. It is necessary to improve a sufficient, manoeuvrable EP generation for the EP system. Even though, Mongolian EP demand was performed with hourly resolution [6].

The predicting the energy consumption and performance has recently been discussed using Artificial Intelligence (AI) in general and machine learning (ML) approaches in particular. Modern ML techniques are used to analyse, estimate, and benchmark energy use in buildings, industries, and transportation [7]. In Mongolia, there are not enough studies on applying machine learning models to the country's energy sector.

The purpose of this study is to create a machine learning model that predicts Mongolia's monthly energy consumption using the previous 54 years of monthly energy consumption data.

We used a long short-term memory (LSTM), a powerful Recurrent neural network (RNN) capable of learning long-term dependencies. Recurrent Neural Network is a class of Artificial neural networks where the connections between nodes can create a cycle, where the output from some nodes affects subsequent input to the same nodes. The RNN is different from other neural networks since it has a loop that allows information to be passed from one step of the network to the next. One of the challenges with Recurrent Neural Networks is vanishing an exploding gradient problem which is a common problem in all types of neural networks. Recurrent Neural Networks have also a cyclic connection making them powerful for modelling sequences. capable of learning order dependence in sequence problems such as time series forecasting [8].

The LSTM model has achieved the best results for many problems on sequential data by remembering in formation for long periods. The main idea behind LSTM cells is to learn the important parts of the sequence and forget the less important ones by the so-called gates. An LSTM module has a cell state and three gates (input, output, forget) which provides them with the power to selectively learn, unlearn or retain information from each of the units. The cell state helps the information flow through the units without being altered by allowing only a few linear interactions. Each unit has an input, output, and forget gate which can add or remove the information to the cell state. The forget gate decides which information from the previous cell state should be forgotten for which it uses a sigmoid function. The input gate controls the information flow to the current cell state using a point-wise multiplication operation of 'sigmoid' and 'tanh' respectively. Finally, the output gate decides which information should be passed on to the next hidden state [9]. One of the purposes to design LSTMs was to address the vanishing and exploding gradient problems of conventional RNNs.

II. Literature Review

A. *Energy consumption prediction*

Energy consumption prediction Energy consumption is on the rise as a result of economic development, technical advancement, and population growth. Additionally, given the current state of energy exploration and use, it is expected that all known energy sources will run out within a few decades. Our lives have been said to depend heavily on energy [10, 11]. When compared to global energy, Mongolia's situation in terms of primary energy resources is relatively poor. The majority of Mongolia's energy is derived from coal, with a smaller portion coming from crude oil exports, mainly to China. According to the World Energy 2020 statistics, 84% of global energy still relies on fossil fuels, while renewable energy accounts for only 11% of global primary energy consumption. The continued reliance on fossil fuels significantly contributes to greenhouse gas (GHG) emissions, exacerbating climate change and environmental pollution [12].

Electricity generation remains one of the largest sources of CO_2 emissions in Mongolia. In 2024, total electricity production reached 8,754.7 million kWh, reflecting a 2.7% increase from the previous year [13].

Knowing how much energy will be used for a task or process allows them to come up with modifications to reduce the amount of energy used. Making both short- and long-term predictions for future energy consumption can help us identify which type of energy is used most, and we can then work to defy the trend, as was done recently with the demise of fossil fuels and the rise of renewable energy. The amount of energy used varies by location depending on factors including water, wind, and temperature. Because there are so many variables involved, predicting energy use is a challenging job [14].

ML models are employed in many fields because they are useful and behave like a function that optimally maps the input data to output. Machine learning models are quite accurate at predicting energy use. To adopt energy- saving measures, governments can use them. ML models, for instance, can predict how much energy would be used during construction [15]. They can also be used to predict how various energy sources, such electricity or natural gas, would be consumed in the upcoming years [15].

B. *Application of Machine learning methods*

Machine learning in industry

Energy efficiency (EE) is rarely listed as one of an organization's top priority because industrial priorities are frequently centered on more urgent objectives. With the rise of Industry 4.0 and AI in particular, there is a strong desire across all industries to investigate the possibilities of ML technologies. ML includes a wide variety of effective techniques that make it simpler to extract crucial insights from unstructured data (such class prediction and pattern recognition), which can then be used to assist businesses in improving their tactical and operational decisions [16]. In Mongolia, there are not enough studies on applying machine learning models to the country's energy sector.

Machine learning in Electric Vehicles

The use of ML in effectively regulating Electric Vehicles Energy Consumption (ECEV), one of the issues being researched, is becoming essential for the creation of really intelligent EVs that make quick, unprogrammed decisions based on observed data and ensure minimal electricity use [24]. Traditional statistical methods like basic linear regression usually struggle to predict correct results since vast amounts of data may be scattered. Contrarily, ML algorithms that combine supervised and unsupervised learning techniques improve prediction accuracy by reducing the difference between the true and predicted data. Simple Linear Regression (SLR) was used to predict the remaining charge and driving range of the EVs [17]. Two of the most recent ML models that have gained popularity are Support Vector Regression (SVR) and Extreme Gradient Boosting (XGBoost) since they can predict events as accurately and quickly as possible [18, 19].

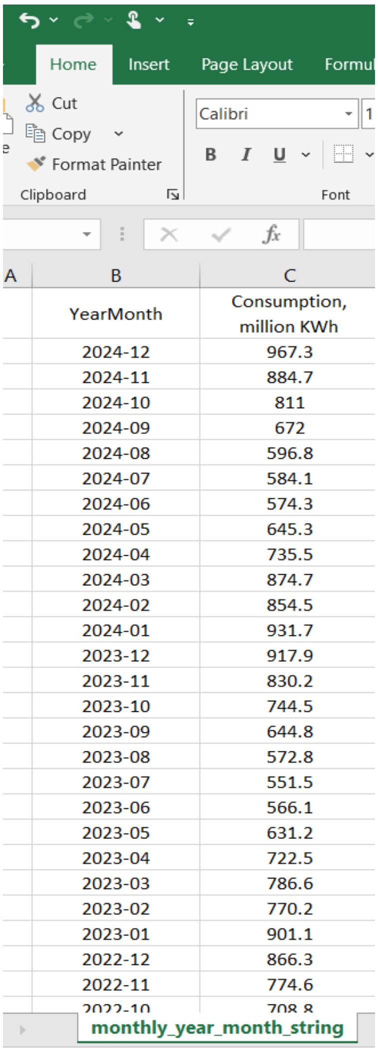
Machine learning in Buildings

The demand for energy in buildings accounts for a significant share of global energy consumption [20-22]. About 40% of the energy used in the US is used by commercial and residential buildings [20-23]. In addition, structures need energy for at least 50 years while they are in use. Buildings must therefore be efficiently constructed and operated in order to promote sustainable development in terms of energy usage. An essential part of energy management, planning, and conservation is predicting energy use in buildings [24]. Artificial neural networks (ANNs), SVR, XGBoost, and Extreme Learning Machines (ELM) are a few ML models that are effective at predicting building energy patterns [20]. With the use of ML models like SVR [24], ELM [20], ANNs [25], Deep Recurrent Neural Networks (DRNN) model [26], and Generative Adversarial Nets (GAN) [27], energy consumption in buildings has been forecasted.

III. DATA COLLECTION AND VISUALIZATION

A. Data collection

The data was imported from Mongolia's energy system as a CSV file and then exported to a Python (Fig.1). There was a total number of 660 observations and 2 variables in this dataset and no missing values were found. The minimum load volume is 26.1 million KWh, and the maximum load volume is 967.3 million KWh along with an average volume of 250.06 million KWh. There is no duplicate row in the dataset.



YearMonth	Consumption, million KWh
2024-12	967.3
2024-11	884.7
2024-10	811
2024-09	672
2024-08	596.8
2024-07	584.1
2024-06	574.3
2024-05	645.3
2024-04	735.5
2024-03	874.7
2024-02	854.5
2024-01	931.7
2023-12	917.9
2023-11	830.2
2023-10	744.5
2023-09	644.8
2023-08	572.8
2023-07	551.5
2023-06	566.1
2023-05	631.2
2023-04	722.5
2023-03	786.6
2023-02	770.2
2023-01	901.1
2022-12	866.3
2022-11	774.6
2022-10	708.8

Fig. 1 Electricity consumption in Mongolia

The first column presents the year and month, and the last column contains electricity consumption in Mongolia.

B. Data exploration

Data exploration considers an important process to understand data with statistical and visualization methods. One of the advantages of this process is to identify patterns and problems in this dataset. It seems by viewing the data frame that some columns are not needed where it presents different times records.

Outliers that have large or small values compared to most observations have a large impact on the model results. The max and min values looked like reasonable values compared to the mean values, so outliers were not found in the data (Fig.2).

	Consumption
count	660.000000
mean	250.069848
std	189.786728
min	26.100000
25%	133.750000
50%	193.100000
75%	292.725000
max	967.300000

Fig. 2 Descriptive statistics

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 660 entries, 0 to 659
Data columns (total 2 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   YearMonth    660 non-null    object
 1   Consumption  660 non-null    float64
dtypes: float64(1), object(1)
memory usage: 10.4+ KB
```

Fig. 3 Data information

The data which contain null values will make the analysis more complicated and affect the result of the model. Null values were not found in this data (Fig.3) and the total Number of years in this dataset is 54 years between 1970 - 2024.

C. Data Visualization

By looking for consumption according to years, it's seemed that there is seasonality in the data set, where the consumption increases in winter and decreases in summer. Since 2013, there has been a large increase in energy consumption, which could be related to the rapid development of industry [Fig.4].

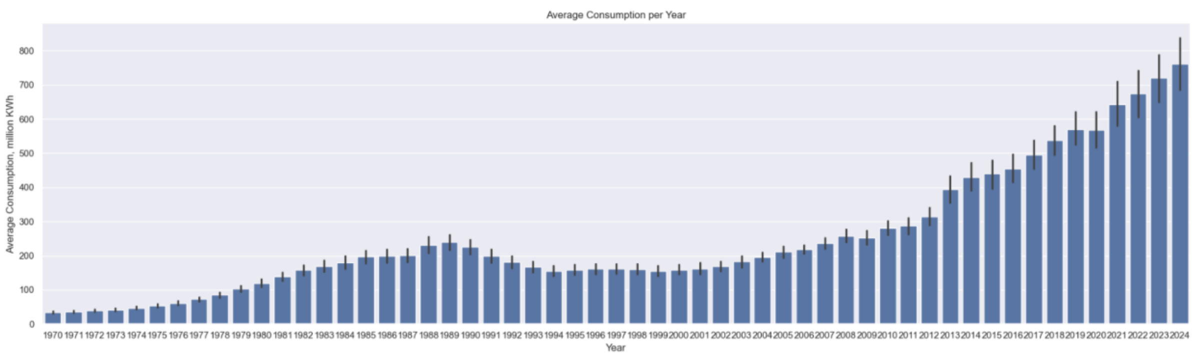


Fig. 4 Energy consumption by the year

The distribution plot shows that the data has unimodal with right skew distribution and the observations are concentrated beside no outliers were found (Fig.5).

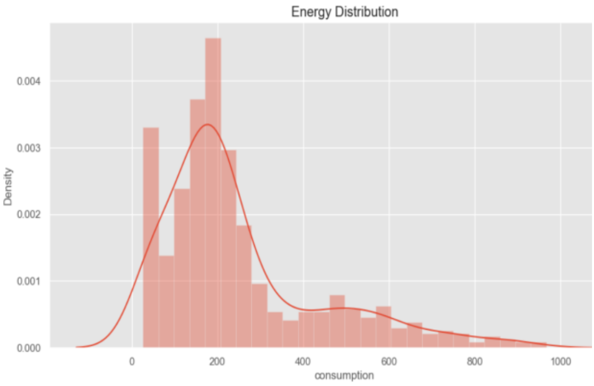


Fig. 5 Distribution of energy consumption

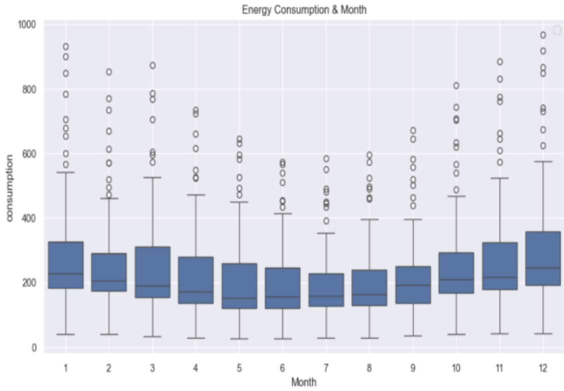


Fig. 6 Energy consumption & Month

IV. LONG-AND SHORT-TERM MEMORY MODEL

A. *Train, Validation, and Test Dataset*

The LSTM model is sensitive to the scale of the input data, so it can be a good practice to rescale the data to the range of 0-to-1. This is called data normalization. In this case, the data was normalized using the MinMaxScaler function. MinMaxScaler scaled the output and the input in the range between 0 - 1 to match the scale of the LSTM layer. MinMaxScaler will subtract the minimum value of Consumption and then divide it by the range.

To train the model we split the data into a training set (80%) and a testing set (20%). Also, we split training data into 20% for the validation set and 80% for training. The validation set is used to validate the model performance during training.

The input layer of the LSTM model requires 3D input, Where the first dimension represents the sample size, and the second represents the number of time steps. And the third dimension represents the number of features used to train the model [number of samples, time steps, and number of features]. The dataset was reshaped to look like a 3dimensional array, where our time steps were 2 and our feature is the Consumption (Fig.7).

```
X_train shape:  (420, 2, 1)
X_test shape:  (129, 2, 1)
X_val shape:  (102, 2, 1)
```

Fig. 7 Reshape Inputs to fit LSTM layers

B. *Model Structure*

Stacked LSTM will be built in this project, which contains four hidden LSTM layers one on top of another. The model building starts by creating a sequential model using Keras. Then we added the hidden LSTM layers to the model. Each layer had 50 units. A dropout layer was added to help in preventing overfitting. The last layer was the dense layer, which does the below operation on the input and returns the output (Fig. 8).

To update network weights iterative based on training data we used the Adam optimization algorithm which was used famously instead of the classical gradient descent procedure. Also, we used root mean squared error (RMSE) to test the model performance. A smaller RMSE means that our model is performing better.

Model: "sequential"

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 2, 50)	10,400
dropout (Dropout)	(None, 2, 50)	0
lstm_1 (LSTM)	(None, 2, 50)	20,200
lstm_2 (LSTM)	(None, 2, 50)	20,200
lstm_3 (LSTM)	(None, 50)	20,200
dense (Dense)	(None, 1)	51

Total params: 71,051 (277.54 KB)

Trainable params: 71,051 (277.54 KB)

Non-trainable params: 0 (0.00 B)

Fig. 8 LSTM Model Summary

C. *Model Training*

Training the LSTM model was done using the training set and the validation dataset for testing the results through the training process. The learning algorithm worked through the entire training dataset 70 times (Epoch), and the model weights were updated after each batch where the batch size is 20 (Fig. 9).

```
42/42 ━━━━━━━━━━━━━━━━━━━━ 0s 6ms/step - loss: 0.0025 - val_loss: 8.1035e-04
Epoch 54/70
42/42 ━━━━━━━━━━━━━━━━━━━━ 0s 6ms/step - loss: 0.0026 - val_loss: 6.8554e-04
Epoch 55/70
42/42 ━━━━━━━━━━━━━━━━━━━━ 0s 6ms/step - loss: 0.0025 - val_loss: 9.4907e-04
Epoch 56/70
42/42 ━━━━━━━━━━━━━━━━━━━━ 0s 6ms/step - loss: 0.0027 - val_loss: 8.2047e-04
Epoch 57/70
42/42 ━━━━━━━━━━━━━━━━━━━━ 0s 6ms/step - loss: 0.0024 - val_loss: 6.5388e-04
Epoch 58/70
42/42 ━━━━━━━━━━━━━━━━━━━━ 0s 6ms/step - loss: 0.0024 - val_loss: 6.9041e-04
Epoch 59/70
42/42 ━━━━━━━━━━━━━━━━━━━━ 0s 6ms/step - loss: 0.0027 - val_loss: 6.8429e-04
Epoch 60/70
42/42 ━━━━━━━━━━━━━━━━━━━━ 0s 6ms/step - loss: 0.0025 - val_loss: 7.5474e-04
Epoch 61/70
42/42 ━━━━━━━━━━━━━━━━━━━━ 0s 6ms/step - loss: 0.0027 - val_loss: 0.0016
Epoch 62/70
42/42 ━━━━━━━━━━━━━━━━━━━━ 0s 6ms/step - loss: 0.0025 - val_loss: 7.2417e-04
Epoch 63/70
42/42 ━━━━━━━━━━━━━━━━━━━━ 0s 6ms/step - loss: 0.0023 - val_loss: 6.7114e-04
Epoch 64/70
42/42 ━━━━━━━━━━━━━━━━━━━━ 0s 6ms/step - loss: 0.0024 - val_loss: 6.2646e-04
Epoch 65/70
42/42 ━━━━━━━━━━━━━━━━━━━━ 0s 6ms/step - loss: 0.0026 - val_loss: 7.9434e-04
Epoch 66/70
42/42 ━━━━━━━━━━━━━━━━━━━━ 0s 6ms/step - loss: 0.0028 - val_loss: 7.1120e-04
Epoch 67/70
42/42 ━━━━━━━━━━━━━━━━━━━━ 0s 6ms/step - loss: 0.0024 - val_loss: 5.8733e-04
Epoch 68/70
42/42 ━━━━━━━━━━━━━━━━━━━━ 0s 6ms/step - loss: 0.0024 - val_loss: 7.7128e-04
Epoch 69/70
42/42 ━━━━━━━━━━━━━━━━━━━━ 0s 6ms/step - loss: 0.0023 - val_loss: 7.3825e-04
Epoch 70/70
42/42 ━━━━━━━━━━━━━━━━━━━━ 0s 5ms/step - loss: 0.0025 - val_loss: 5.9596e-04
```

Fig. 9 Training LSTM Model

The check after training was to compare the training loss against the validation loss. The result shows that the two values were low, and no overfitting was detected (Fig. 10).

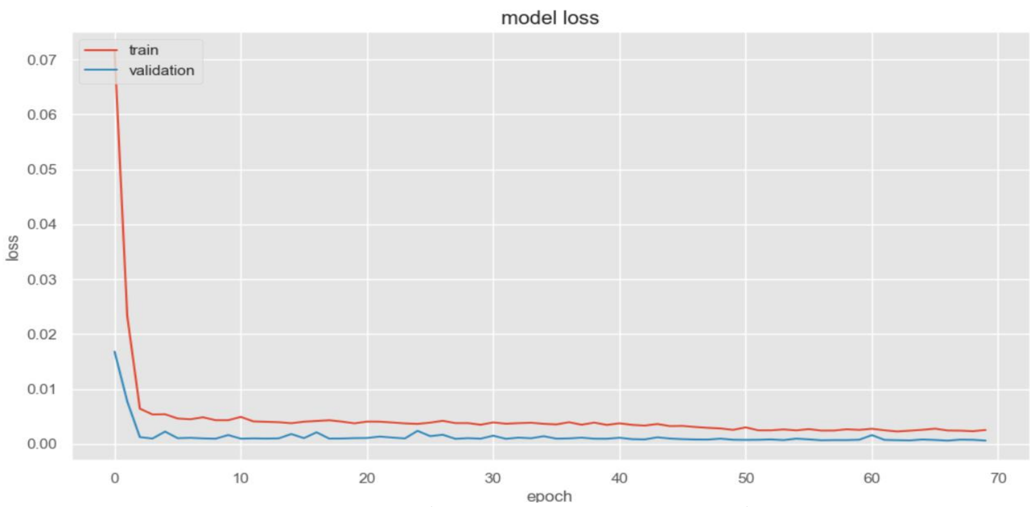


Fig. 10 Comparison Curve of Training and validation Accuracy of 60 epochs

V. RESEARCH RESULT

The results of the LSTM model predicting energy consumption using the training dataset (Fig. 11), validation dataset (Fig. 12), and test dataset (Fig. 13) show good predictive performance, with an MSE of 0.04.

	Train Predictions	Actuals
0	[727.653076171875]	[811.0]
1	[694.3245849609375]	[672.0]
2	[582.5758056640625]	[596.8]
3	[551.2528076171875]	[584.1]
4	[551.7942504882812]	[574.3]
...	...	...
415	[172.98416137695312]	[220.2]
416	[197.3787384033203]	[255.6]
417	[229.8621368408203]	[263.7]
418	[245.3514404296875]	[299.5]
419	[274.4111633300781]	[288.0]

420 rows × 2 columns

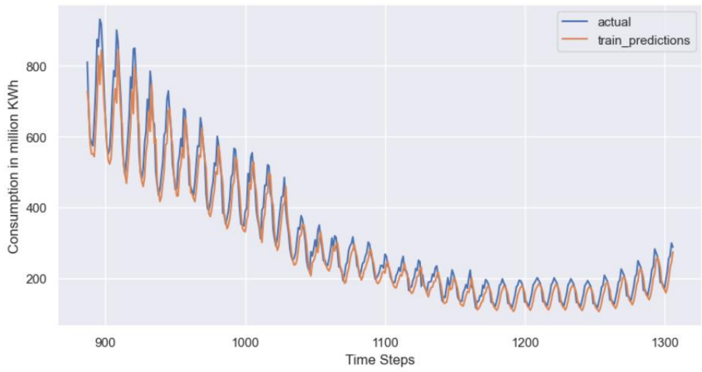


Fig. 11 Predict consumption using training data

	Val Predictions	Actuals_val
0	[198.2513427734375]	[201.6]
1	[185.2861785888672]	[190.1]
2	[178.25009155273438]	[181.4]
3	[168.96286010742188]	[207.4]
4	[184.09059143066406]	[233.3]
...	...	...
97	[102.47207641601562]	[109.4]
98	[98.90711975097656]	[104.5]
99	[93.62354278564453]	[119.4]
100	[101.1259994506836]	[134.3]
101	[115.57604217529297]	[155.9]

102 rows × 2 columns

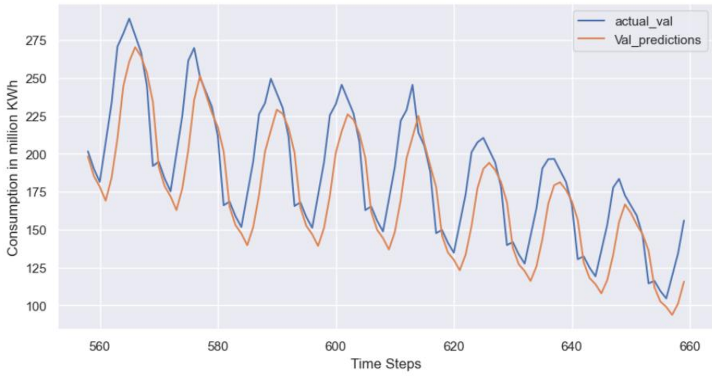


Fig.12 Predict consumption using validation data

	test Predictions	Actuals_test
0	[131.4945068359375]	[137.9]
1	[125.9340591430664]	[126.4]
2	[116.74889373779297]	[99.1]
3	[96.03764343261719]	[100.6]
4	[87.7029800415039]	[94.8]
...	...	...
124	[20.816097259521484]	[27.4]
125	[20.081451416015625]	[26.1]
126	[18.77117347717285]	[29.8]
127	[20.328025817871094]	[33.6]
128	[23.80988883972168]	[39.0]

129 rows × 2 columns

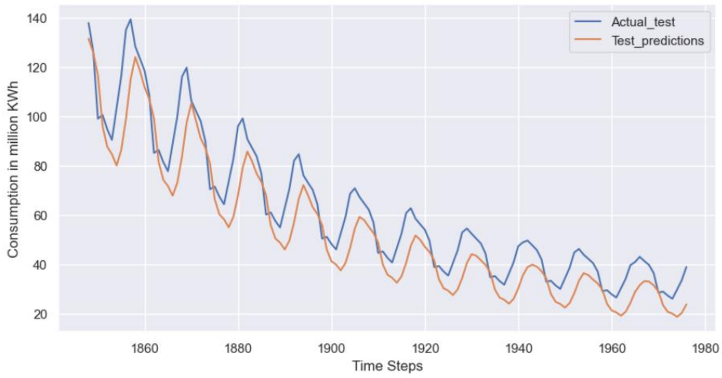


Fig. 13 Predict consumption using test data

VI. CONCLUSIONS

The study sought to forecast Mongolia's energy use by training a long short-term memory (LSTM) network on a 54-year dataset of electricity consumption. The hourly data were aggregated to daily figures and divided into samples of time steps. The model was trained on the initial 420 observations, with predictions beginning at observation 102. The architecture included four LSTM layers, a Dropout layer, and a final output layer, optimized with Adam. Performance was evaluated using RMSE. Findings indicate strong short-term predictive performance of the stacked LSTM, with assessments conducted across training, validation, and test sets.

REFERENCES

[1]. "Statistics on Energy Performance 2024," Mongolian Energy Regulatory Commission, Ulaanbaatar, Mongolia, 2024.

[2]. D. Sanchez, U. Batmunkh, D. Enkhtuya, A. Twum, "Mongolia Economic Update - Robust but Unbalanced Growth," World Bank Group, Washington, D.C., USA, Rep. 186065, Nov. 2023.

[3]. Ching-Lai Hor, S. J. Watson and S. Majithia, "Analyzing the impact of weather variables on monthly electricity demand," in IEEE Trans. on Power Syst.., vol. 20, no. 4, pp. 2078-2085, Nov. 2005, doi: 10.1109/TPWRS.2005.857397.

[4]. Z. Gansukh, "Mongol Dream Beyond Fossil Fuel: Prosperity of Greenification," Elsevier, Renewable Energy, Vol. 171, 2021, pp. 95-102. https://doi.org/10.1016/j.renene.2021.02.079.

[5]. C. Adiyabazar, F. Gonzalez-Longatt, M. N. Acosta, J. L. Rueda and P. Palensky, "Assessment of Under-Frequency Load Shedding in Mongolia Considering Inertia Scenarios," 2020 IEEE PES Innovative Smart Grid Technol. Europe (ISGT Europe), The Hague, Netherlands, 2020, pp. 1201-1205, doi: 10.1109/ISGT-Europe47291.2020.9248837.

[6]. T. Osgonbaatar, et al., "A Rank Analysis and Ensemble Machine Learning Model for Load Forecasting in the Nodes of the Central Mongolian Power System," MDPI Inventions, Vol. 8, 114. 2023, p. 1-20, https://doi.org/10.3390/inventions8050114

[7]. Zahra Eddaoudia, Zineb Aaraba, Khadija Boudmena, Asmae Elghazia,c,d, Moulay Driss Rahmania, "A Brief Review of energy consumption forecasting machine learning models", 2023, pp. 2, doi:10.1016/j.procs.2024.05.001

[8]. Sak, H. (2014, Feb 5). Long Short-Term Memory Based Recurrent Neural Network Architectures for Large Vocabulary Speech Recognition. Retrieved from arxiv.org: https://arxiv.org/abs/1402.1128

[9]. Time Series-LSTM Model. (2022, 10 16). Retrieved from tutorialspoint.com: https://www.tutorialspoint.com/time_series/time_series_lstm_model.htm

[10]. Groh, S., & Taylor, H. (2015). The role of microfinance in energy access–changing roles, changing paradigms and future potential. Micro Perspectives For Decentralized Energy Supply, 110.

[11]. Bazilian, M., Nakhooda, S., & Van de Graaf, T. (2014). Energy governance and poverty. Energy Research & Social Science, 1, 217 225.

[12]. Mubaarak.S.; Zhang.D.; Liu.J.; Chen.Y.; Wang.L.; Zaki.S.A.; Yuan.R.; Wu.J.; Zhang.Y.; Li.M et al. Potential Techno-Economic Feasibility of Hybrid Energy Systems for Electrifying Various Consumers in Yemen. Sustainability 2021, 13,228. https:// doi.org/10.3390/su13010228

[13]. Energy regulatory commission of Mongolia 2024, Energy Regulatory Commission

[14]. Zhao, H. X., & Magoulès, F. (2012). A review on the prediction of building energy consumption. Renewable and Sustainable Energy Reviews, 16(6), 3586-3592.

[15]. Ekici, B. B., & Aksoy, U. T. (2009). Prediction of building energy consumption by using artificial neural networks. Advances in Engineering Software, 40(5), 356-362.

[16]. Narciso, D. A., & Martins, F. G. (2020). Application of machine learning tools for energy efficiency in industry: A review. Energy Reports, 6, 1181-1199.

[17]. Amirkhani, A., Haghanifar, A., & Mosavi, M. R. (2019, December). Electric vehicles driving range and energy consumption investigation: A comparative study of machine learning techniques. In 2019 5th Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS) (pp. 1-6). IEEE.

[18]. Ma, F., & Yan, X. (2019, October). Research on the energy consumption estimation method of pure electric vehicle based on XGBoost. In 2019 3rd International Conference on Electronic Information Technology and Computer Engineering (EITCE) (pp. 1021-1026). IEEE.

[19]. Torlay, L., Perrone-Bertolotti, M., Thomas, E., & Baciu, M. (2017). Machine learning–XGBoost analysis of language networks to classify patients with epilepsy. Brain informatics, 4(3), 159-169.

[20]. Naji, S., Keivani, A., Shamshirband, S., Alengaram, U. J., Jumaat, M. Z., Mansor, Z., & Lee, M. (2016). Estimating building energy consumption using extreme learning machine method. Energy, 97, 506-516.

[21]. Chou, J. S., & Tran, D. S. (2018). Forecasting energy consumption time series using machine learning techniques based on usage patterns of residential householders. Energy, 165, 709-726.

[22]. Chou, J. S., Ngo, N. T., Chong, W. K., & Gibson Jr, G. E. (2016). Big data analytics and cloud computing for sustainable building energy efficiency. In Start-up creation (pp. 397-412). Woodhead Publishing.

[23]. Block, M., & Bokalders, V. (2010). The Whole Building Handbook:" How to Design Healthy, Efficient and Sustainable Buildings". Routledge.

[24]. Zhong, H., Wang, J., Jia, H., Mu, Y., & Lv, S. (2019). Vector field-based support vector regression for building energy consumption prediction. Applied Energy, 242, 403-414.

[25]. Chaudhuri, T., Soh, Y. C., Li, H., & Xie, L. (2019). A feedforward neural network based indoor-climate control framework for thermal comfort and energy saving in buildings. Applied energy, 248, 44-53.

[26]. Rahman, A., Srikumar, V., & Smith, A. D. (2018). Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks. Applied energy, 212, 372-385.

[27]. Tian, C., Li, C., Zhang, G., & Lv, Y. (2019). Data driven parallel prediction of building energy consumption using generative adversarial nets. Energy and Buildings, 186, 230-243.