Journal of Computer Science and Computational Intelligence
November 2025 | Vol. 1 | No. 2 | Page 53-60
E-ISSN: 3109-7383 P-ISSN: 3110-8393
DOI: https://doi.org/10.30598/algorithm.v1i2.53-60

Soil Moisture Prediction Model on Peatlands using Long Short-Term Memory

Helda Yunita Taihuttu [1*] Jemsri Stenli Batlajery[2]

[1] Department of Computer Science, Faculty of Science and Technology, Universitas Pattimura
Jl. Ir. M. Putuhena, Ambon, 97233, Indonesia
[2] Department of Computer Science, Faculty of Mathematics and Natural Sciences, IPB University
Jl. Raya Dramaga, Bogor 16680, Indonesia
[*]Corresponding author's e-mail: ***yunitahelda24@gmail.com**

Manuscript submitted: ***11th November 2025*** | ***Manuscript revision:*** ***20th November 2025*** | ***Accepted for publication:*** ***22nd November 2025***

Abstract

Peatlands play an important role in maintaining global ecosystem balance, but are highly susceptible to fires due to decreased soil moisture. This study aims to predict soil moisture on peatlands using the Long Short-Term Memory algorithm as a time series learning model. The data used includes variables of soil moisture (GWETPROF), rainfall (PRECTOTCORR), and temperature (T2M) obtained from NASA Langley Research Center's Prediction of Worldwide Energy Resources (POWER) for the period from August 1, 2019, to December 31, 2023. The preprocessing involved identifying and handling missing values using the mean imputation method and normalization with the Min-Max Scaling technique. Correlation analysis showed a weak relationship between variables, so all of them were used as independent features. The LSTM model was built with parameters of 50 neurons, ReLU activation function, Adam optimizer, and a dropout rate of 0.2. The test results showed that the model was able to accurately predict water content with a MAE value of 0.005, MSE ≈ 0.000, RMSE of 0.014, and R^2 of 0.97 on the test data. These results indicate that LSTM is effective in capturing temporal patterns and fluctuations in soil moisture, making it a potential tool for more adaptive and data-driven peatland fire mitigation.

Keywords: *Long Short-Term Memory, Peatland, Soil Moisture, Prediction Model.*

How to cite this article:
H. Y. Taihuttu, and J. S. Batlajery, "Soil Moisture Prediction Model on Peatlands using Long Short-Term Memory", *algorithm*, vol. 1, no. 2, pp. 53-60, November 2025.

Journal homepage: https://ojs3.unpatti.ac.id/index.php/algorithm
Research Article Open Access

1. INTRODUCTION

Indonesia has 24.67 million hectares of peatland that is prone to forest and land fires, causing greenhouse gas emissions and biodiversity loss [1]. Ogan Komering Ilir (OKI) Regency in South Sumatra is an area with a high frequency of fires and thousands of hotspots every [1], causing economic losses and threatening the ecological function of peat as a carbon sink[2].

Various studies show that soil moisture is a key indicator of fire [3], and [4] found a strong relationship between low soil moisture and increased fire occurrence in various countries. Meanwhile, Schaefer[5] showed that the relationship between soil moisture and fires depends on the type of land cover. However, previous studies have not focused on peatlands and have not been able to effectively handle real-time data fluctuations.

To address these limitations, this study proposes the use of Long Short-Term Memory (LSTM), a deep learning model capable of handling long- and short-term dependencies in sequential data. This model will integrate soil moisture, rainfall, and temperature data to predict fire risk more accurately and adaptively to peatland conditions with varying degrees of maturity and thickness. The developed LSTM model is expected to improve real-time peatland fire prediction, support data-driven decision-making for fire mitigation, and contribute to global climate change mitigation efforts through emission reduction.

2. RESEARCH METHODS

2.1. Study Area

The data in this study covers the Ogan Komering Ilir Regency in South Sumatra Province, an area of 19,023.47 km^2 that often experiences forest and land fires. Geographically, it is located between 104°20' East Longitude and 106°00' East Longitude and 2°30' South Latitude and 4°15' South Latitude.

2.2. Research Data

Soil Moisture, Rainfall, and Temperature Data: This data was collected from August 1, 2019, to December 31, 2023, using Prediction of Worldwide Energy Resources (POWER) provided by NASA Langley Research Center (LaRC) through their portal (https://power.larc.nasa.gov/data-access-viewer). POWER has been available since 2003. This data has a resolution of ½° latitude × ⅝° longitude for meteorological data sets, and the grid reference system is WGS84. Meteorological parameters are based on the MERRA-2 assimilation model. Details of the attributes of this data are presented in Table 1.

Table 1. Dataset attributes for prediction

Attributes	Description
LATITUDE	Latitude coordinates of station location (°)
LONGITUDE	Longitude coordinates of station location (°)
Date	Date of soil moisture measurement
GWETPROF	Soil moisture (%)
T2M	Temperature at 2 Meters (C)
PRECTOTCORR	Rainfall (mm/day)

Soil Moisture Active Passive (SMAP) is a NASA satellite project launched in January 2015 with the aim of measuring and mapping the moisture content of several inches of soil surface globally. This satellite has a temporal resolution that covers scanning every 2-3 days and a spatial resolution of 10x10 km. In this study, we used level 3 data from SMAP.

2.3. Research Stages

This research was conducted in several stages, namely data collection, data preprocessing, data partition, modeling, and evaluation. The overall research process can be seen in Figure 1.

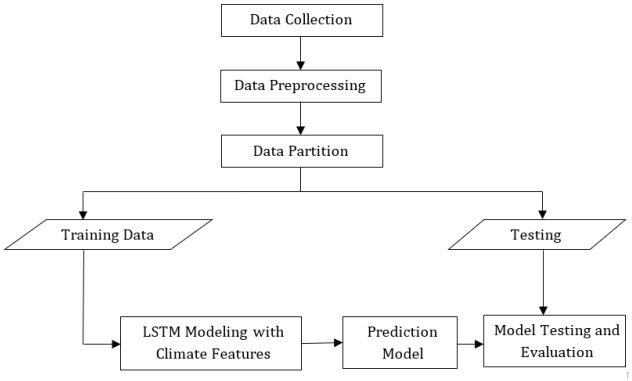

Figure 1. Research Stages

2.4. Data Preprocessing

The data used for modeling, namely rainfall, temperature, and soil moisture time series data, were checked and handled for missing values and normalization, and the model was designed to predict conditions one day ahead, as short-term prediction provides higher accuracy and is more relevant for timely decision-making in environmental monitoring and early warning systems. Missing values were handled using mean imputation, and normalization was performed using min-max normalization. Min-max normalization was calculated using Equation 1 [6].

$$x' = \frac{x - x(\min)}{x(\max) - x(\min)} \quad (1)$$

2.5. Long Short-Term memory (LSTM)

The LSTM architecture was introduced by Hochreiter and Schmidhuber in 1997 as an improvement on the Recurrent Neural Network (RNN) to overcome the vanishing gradient problem through the addition of a memory cell component. This structure is designed so that the network is able to store and utilize long-term information, making it very suitable for modeling and predicting time series data. An illustration of the LSTM architecture is shown in Figure 2.

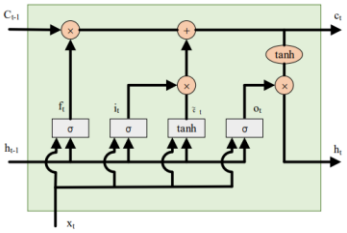

Figure 2. LSTM Architecture [7]

Based on Figure 2, LSTM has three main gates, namely the forget gate, input gate, and output gate. LSTM also has three inputs, namely x_t, C_{t-1}, and h_{t-1}, two outputs, namely c_t and h_t, and a bias value b. The following are the stages in the LSTM architecture [8] :

1. Forgetting Layer (f_t)

This layer uses the sigmoid activation function (σ) which provides an output value between

0 and 1. This layer is tasked with determining whether the information from the previous output (h_{t-1}) and the current input (x_t) is relevant or not. Information that is considered relevant will be passed on, while irrelevant information will be forgotten. Equation 2 for the forgetting layer can be written as.

$$f_t = \sigma(W_f . [h_{t-1}, x_t] + b_f) \quad (2)$$

2. Input gate (i_t)

The input gate also uses the sigmoid activation function to determine which part of the input information will be updated. The input gate can be seen in Equation 3.

$$i_t = \sigma(W_i . [\,[h_{t-1}, x_t] + b_i) \quad (3)$$

3. Candidate Layer $(\tilde{C}_t)$

This layer produces candidate values that will be used to update cell memory, using the tanh activation function. The candidate layer is calculated using Equation 4.

$$\tilde{C}_t = tanh(W_C . [\,[h_{t-1}, x_t] + b_C \quad (4)$$

4. Update Cell Memory (C_t)

The cell memory is updated by combining information from the forgetting layer and input gate, and integrating the new candidate from the candidate layer. Equation 5 is used to calculate the cell memory update.

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

5. Output gate (o_t)

The output gate determines the portion of the cell memory that will be sent to the output. It also uses the sigmoid activation function. The output gate calculation uses Equation 6.

$$o_t = (W_o . [\,[h_{t-1}, x_t] + b_o) \quad (6)$$

6. Output (h_t)

Finally, the output of the LSTM is calculated by multiplying the output from the output gate by the hyperbolic tangent value of the updated cell memory. The output value is calculated using Equation 7.

$$h_t = o_t \times \tanh(C_t) \quad (7)$$

2.6. Model Evaluation

After the prediction model is generated, it will be evaluated. This study uses several evaluation metrics, namely R-square (R^2), Mean Absolute Error (MAE), Mean Square Error (MSE), and Root Mean Square Error (RMSE). R^2 measures how well the model explains data variability. A value close to 1 indicates a better prediction. MAE calculates the average absolute error between the predicted and actual values. MSE calculates the average square error. RMSE is a technique often used to evaluate the difference between the actual value and the predicted result. MAE, MSE, and RMSE values closer to 0 indicate that the predictions are closer to the actual data. The R^2, MAE, MSE, and RMSE values can be calculated using Equation 8 [9], Equation 9[10] , Equation 10[10], and Equation 11[10].

$$R^2 = \frac{\sum_{t=1}^{n}(y_t-\bar{y})^2 - \sum_{t=1}^{n}(y_t-t)^2}{\sum_{t=1}^{n}(y_t-\bar{y})^2} \quad (8)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n} y_t - \hat{y}_t \quad (9)$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_t - \hat{y}_t)^2 \quad (10)$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}{n}} \quad (11)$$

Where:

y_t = Observed value for observation t
$\hat{y}_t$ = Predicted value for observation t
$\bar{y}$ = Average of all observed values
n = Number of data points

3. RESULTS AND DISCUSSION

3.1. Data Collection

The data in this study was obtained through a download process from the Prediction of Worldwide Energy Resources (POWER) system developed and managed by NASA Langley Research Center (LaRC). The dataset includes three variables, namely soil moisture (GWETPROF), rainfall (PRECTOTCORR), and temperature (T2M). The data was collected in daily resolution for the period from July 1, 2019, to December 31, 2023, and has 1614 rows × 3 columns.

3.2. Data Preprocessing

Data preprocessing began with the identification and handling of missing values to ensure data completeness. Missing values were handled using the mean ach falls under the category of central tendency imputation, which is simple yet effective in maintaining a stable data distribution without changing the number of observations. An example of the results of missing value handling can be seen in Table 2.

Table 2. Results of missing value handling

Date	Data Before Handling GWETPROF	Data After Handling GWETPROF
01/08/2019	0,86	0,86
02/08/2019	0,86	0,86
03/08/2019	0,85	0,85
04/08/2019	0,85	0,85
05/08/2019	0,84	0,84
…	…	…
27/12/2023	*NaN*	0.90509
28/12/2023	*NaN*	0.90509
29/12/2023	*NaN*	0.90509
30/12/2023	*NaN*	0.90509
31/12/2023	*NaN*	0.90509

Correlation analysis was performed on data that had undergone missing value handling to determine the relationship between the variables used in this study. The results of the correlation analysis between soil moisture (GWETPROF), temperature (T2M), and rainfall (PRECTOTCORR) are presented in Figure 3.

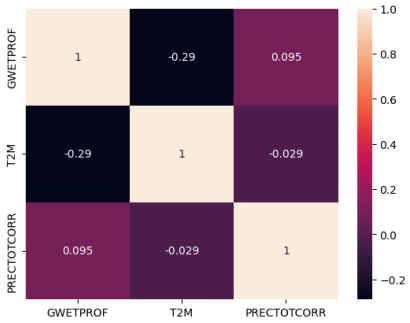

Figure 3. Correlation between variables

Figure 3 shows the results of the correlation analysis between climate variables, namely GWETPROF (soil moisture), T2M (temperature), and PRECTOTCORR (rainfall). The correlation value between GWETPROF and T2M of –0.29 indicates a weak negative relationship, suggesting that an increase in temperature tends to decrease soil moisture. The correlation between GWETPROF and PRECTOTCORR is 0.095, which is very weakly positive, while that between T2M and PRECTOTCORR is –0.029, which is almost unrelated. Overall, the three variables have weak relationships, so there is no indication of multicollinearity, and each can be used as an independent feature in the prediction model.

3.3. Data Partition

The data was divided into two parts, namely 80% as training data and 20% as test data, to ensure that the model could learn optimally and be evaluated objectively on data that was not used during training.

3.4. Modeling with LSTM

Soil moisture prediction modeling was carried out using the LSTM algorithms implemented with the Keras library based on TensorFlow in Python. The modeling process was carried out using the Long Short-Term Memory (LSTM) architecture with a series of optimal parameters. The parameters used were 50 neurons in the hidden layer to determine the network's capacity to capture temporal patterns, the ReLU activation function to regulate the model's non-linearity, the Adam optimizer to optimize weight updates during the training process, and a dropout rate of 0.2 as a regularization technique to reduce the risk of overfitting, as well as using 50 epochs.

After the model was obtained, it was then tested with test data to see the performance of the prediction model on data that had never been trained before. The visualization of the results (Figure 4) shows that the prediction curve (red) almost overlaps with the actual curve (blue) throughout the observation period, indicating that the LSTM model is able to recognize and predict soil moisture dynamics accurately, both when soil moisture is high at the beginning of the period and when it decreases sharply towards the end of the observation period.

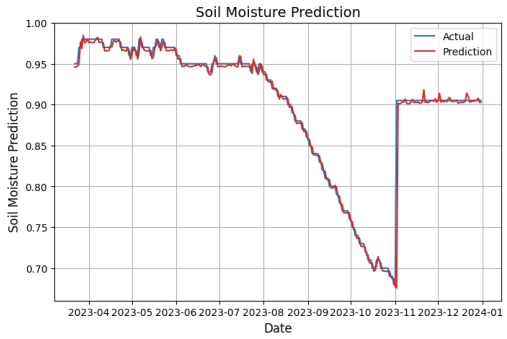

Figure 4. Comparison of actual and predicted soil moisture values

3.5. Model Evaluation

The performance of the LSTM model was evaluated using R-square (R^2), Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). The evaluation results can be seen in Figure 5 and Figure 6.

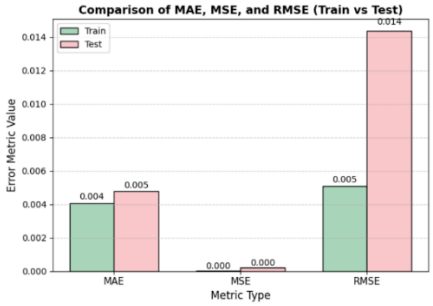

Figure 5. Comparison of MAE, MSE, and RMSE values for training and test data

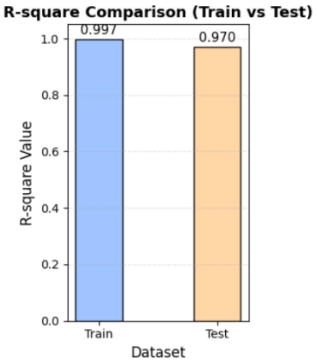

Figure 6. Comparison of R^2 values for training and test data

In Figure 5, it can be seen that the MAE (Mean Absolute Error) value for the training data is 0.004 and for the test data is 0.005, while the MSE (Mean Squared Error) value is close to 0.000 in both datasets. The Root Mean Squared Error (RMSE) values are also relatively small, 0.005 for the training data and 0.014 for the test data. These low error values indicate that the difference between the predicted values and the actual values is very small, indicating that the model has high prediction accuracy and is stable in the learning process.

Figure 6 shows a coefficient of determination (R^2) value of 0.997 for the training data and 0.970 for the test data. An R^2 value close to 1 indicates that the model is able to explain more than 97% of the data variation in the test dataset, with a slight decrease compared to the training data.

This indicates that the model has excellent generalization capabilities and shows no signs of significant overfitting.

Overall, these evaluation results reinforce that the LSTM model used has optimal performance, with a very low prediction error rate and high data pattern representation capabilities in both training and test data.

4. CONCLUSION

This study successfully developed a Long Short-Term Memory (LSTM)-based soil moisture prediction model for peatlands that shows excellent prediction performance with low error rates and high determination values. The model with a configuration of 50 neurons, ReLU activation, Adam optimizer, and 0.2 dropout was able to accurately recognize soil moisture dynamics patterns in both the increase and decrease phases. The MAE (0.005), RMSE (0.014), and R^2 (0.97) values indicate that the model is able to represent the complex relationship between climate factors and soil moisture very well. These findings confirm that the LSTM approach has great potential for application in early warning systems for peatland fires, particularly in vulnerable areas such as Ogan Komering Ilir Regency, and supports data-driven decision-making in climate change mitigation efforts.

5. REFERENCE

[1] [KLHK] Kementerian Lingkungan Hidup dan Kehutanan RI, *Status Hutan dan Kehutanan Indonesia*. 2018.

[2] P. Crippa *et al.*, "Population exposure to hazardous air quality due to the 2015 fires in Equatorial Asia," *Sci. Rep.*, vol. 6, no. August, pp. 1–9, 2016, doi: 10.1038/srep37074.

[3] D. Jensen, J. T. Reager, B. Zajic, N. Rousseau, M. Rodell, and E. Hinkley, "The sensitivity of US wildfire occurrence to pre-season soil moisture conditions across ecosystems," *Environ. Res. Lett.*, vol. 13, no. 1, 2018, doi: 10.1088/1748-9326/aa9853.

[4] N. Sazib, J. D. Bolten, and I. E. Mladenova, "Leveraging NASA Soil Moisture Active Passive for Assessing Fire Susceptibility and Potential Impacts over Australia and California," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 15, no. June, pp. 779–787, 2022, doi: 10.1109/JSTARS.2021.3136756.

[5] A. J. Schaefer and B. I. Magi, "Land-cover dependent relationships between fire and soil moisture," *Fire*, vol. 2, no. 4, pp. 1–15, 2019, doi: 10.3390/fire2040055.

[6] T. Wahyuningsih and E. Rahwanto, "Comparison of Min-Max normalization and Z-Score Normalization in the K-nearest neighbor (kNN) Algorithm to Test the Accuracy of Types of Breast Cancer," vol. 4, no. 1, pp. 13–20, 2021.

[7] J. Luo, L. Zhu, K. Zhang, C. Zhao, and Z. Liu, "Forecasting the 10.7-cm Solar Radio Flux Using Deep CNN-LSTM Neural Networks," pp. 1–11, 2022.

[8] S. and Hochreiter, "LONG SHORT-TERM MEMORY," vol. 9, no. 8, pp. 1–32, 1997.

[9] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)? -Arguments against avoiding RMSE in the literature," *Geosci. Model Dev.*, vol. 7, no. 3, pp. 1247–1250, 2014, doi: 10.5194/gmd-7-1247-2014.

[10] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *PeerJ Comput. Sci.*, vol. 7, pp. 1–24, 2021, doi: 10.7717/PEERJ-CS.623.