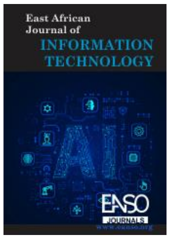


East African Journal of Information Technology

eajit.eanso.org

Volume 8, Issue 1, 2025

Print ISSN: 2707-5346 | Online ISSN: 2707-5354

Title DOI: https://doi.org/10.37284/2707-5354

EAST AFRICAN NATURE & SCIENCE ORGANIZATION

Original Article

Prediction of Forest Fire Danger by Using BiGRU Deep Learning Model and Comparable Data Scaling Methods: A Case of SAO Hill Forest Plantation, Tanzania

Isakwisa Gaddy Tende[1*]

[1] Dar es Salaam Institute of Technology, P. O. Box 2958, Dar es Salaam.
[*] Author for Correspondence ORCID ID; https://orcid.org/0009-0000-2264-7815; Email: isakwisa.tende@dit.ac.tz

Article DOI: https://doi.org/10.37284/eajit.8.1.3135

Date Published:

16 June 2025

Keywords:

Forest Fire Prediction, Fire Weather Index (FWI), Data Scaling Methods, SAO Hill Forest Plantation, Tanzania, BiGRU, Deep Learning.

ABSTRACT

Forest plantations are crucial in the daily lives of humans, playing an important role in producing raw materials for the wood industry, generating personal incomes, contributing to economies, attracting tourists, conserving biodiversity, and regulating the climate. Failure to accurately and timely predict forest fires can have devastating effects due to the destruction of forests by fire, resulting in loss of businesses and incomes, destruction of biodiversity, loss of tourist attractions, and shortage of wood raw materials. Fire Weather Index (FWI) is commonly used to indicate fire danger as it gives useful information on the impact of wind and fuel moisture on the behaviour and spread of fire. This study utilizes FWI by developing a Bidirectional Gated Recurrent Unit (BiGRU) Deep Learning model, which uses the previous 5 days FWI values as input to predict the output (next day FWI) at SAO Hill Forest Plantation located in Iringa region, Tanzania, using three commonly used data scaling methods: Min-Max, Standard, and Robust scalers. The 13-year SAO Hill Forest Plantation daily FWI dataset was pre-processed using a scaling (normalization) approach and split into training, validation and test sets before being used for training and testing the developed BiGRU Deep Learning model. The trained BiGRU Deep Learning model was then saved into .h5 format and integrated with a Gradio-based Web App to provide a user interface for officials at SAO Hill Forest Plantation to predict daily FWI. The evaluation findings reveal that the choice of data scaler has an impact on the daily FWI prediction performance of the developed BiGRU model, and Min-Max is the best performing and optimal data scaler with a Root Mean Squared Error (RMSE) score of 0.065 on test data, followed by Standard scaler with a test RMSE score of 0.157, followed by Robust scaler with a test RMSE score of 0.311. Major contributions of this study include a pre-processed 13-year FWI dataset for SAO Hill Forest Plantation ready for Artificial Intelligence (AI) research and development, a novel BiGRU model for predicting daily FWI at SAO Hill Forest Plantation, and a Web App integrated with the developed BiGRU model and Min-Max data scaler to help officials at SAO Hill Forest Plantation predict daily fire

This work is licensed under a Creative Commons Attribution 4.0 International License.

Article DOI: https://doi.org/10.37284/eajit.8.1.3135

danger and take precautionary measures to prevent forest fire ignition, respond to forest fire if it happens, and contain its spread.

APA CITATION

Tende, I. G. (2025). Prediction of Forest Fire Danger by Using BiGRU Deep Learning Model and Comparable Data Scaling Methods: A Case of SAO Hill Forest Plantation, Tanzania. *East African Journal of Information Technology*, *8*(1), 196-211. https://doi.org/10.37284/eajit.8.1.3135.

CHICAGO CITATION

Tende, Isakwisa Gaddy. "Prediction of Forest Fire Danger by Using BiGRU Deep Learning Model and Comparable Data Scaling Methods: A Case of SAO Hill Forest Plantation, Tanzania". *East African Journal of Information Technology* 8 (1), 196-211. https://doi.org/10.37284/eajit.8.1.3135.

HARVARD CITATION

Tende, I. G. (2025) "Prediction of Forest Fire Danger by Using BiGRU Deep Learning Model and Comparable Data Scaling Methods: A Case of SAO Hill Forest Plantation, Tanzania", *East African Journal of Information Technology*, 8(1), pp. 196-211. doi: 10.37284/eajit.8.1.3135.

IEEE CITATION

I. G. Tende "Prediction of Forest Fire Danger by Using BiGRU Deep Learning Model and Comparable Data Scaling Methods: A Case of SAO Hill Forest Plantation, Tanzania", *EAJIT*, vol. 8, no. 1, pp. 196-211, Jun. 2025.

MLA CITATION

Tende, Isakwisa Gaddy. "Prediction of Forest Fire Danger by Using BiGRU Deep Learning Model and Comparable Data Scaling Methods: A Case of SAO Hill Forest Plantation, Tanzania". *East African Journal of Information Technology*, Vol. 8, no. 1, Jun. 2025, pp. 196-211, doi:10.37284/eajit.8.1.3135.

INTRODUCTION

Forest plantations are crucial to humans in many ways, ranging from being the main sources of raw materials for the wood industry, generating personal incomes and contributing to the economies along the way to being tourist attractions and conserving the biodiversity and the ecosystem (Tian *et al*., 2017; Mgina *et al*., 2021, Zhang *et al*., 2022). In Tanzania, SAO Hill Forest Plantation (SAO Hill FP) is by far the biggest forest plantation (Kangalawe, 2021) and the biggest source of raw materials for the wood industry (Mgina *et al.,* 2021) which are consumed by the wood industry inside and outside the country, playing along the way a crucial role of generating personal incomes to many Tanzanians and contributing to the economy. Forest fires can be catastrophic, negatively impacting the availability of wood raw materials, personal incomes of people, and the economy, and can destroy biodiversity, ecosystems, and even nearby properties (Mgina *et al*., 2021; Shin *et al*., 2019). SAO Hill FP has been experiencing several forest fires over many years. For instance, between 2000 and 2011, SAO Hill FP experienced 143 fire incidents (Mgina *et al*., 2021). Failure to accurately predict forest fire danger in the SAO Hill FP and take precautionary measures can have devastating negative impacts. To address this issue, it is important to have in place effective Artificial Intelligence (AI) models that can accurately predict forest fire danger at SAO Hill FP and take precautionary measures to prevent forest fire from happening or to suppress and contain it in case it happens.

Fire Weather Index (Bouramdane, 2024; Copernicus, 2025) is a numerical index used to indicate how likely the meteorological conditions (fuel moisture and wind) can trigger the forest fire intensity and spread. If the FWI value is high, there is a greater likelihood that the meteorological conditions will trigger a forest fire. This study utilizes FWI by developing a Bidirectional Gated Recurrent Unit (BiGRU) Deep Learning model, which uses SAO Hill FP FWI data to predict daily fire danger, information that will be useful to officials at SAO Hill FP in preventing, suppressing, and containing forest fires.

Data scaling (normalization) is an important step when pre-processing data before feeding it into

This work is licensed under a Creative Commons Attribution 4.0 International License.

Article DOI: https://doi.org/10.37284/eajit.8.1.3135

Deep Learning models. The data scaling process usually involves using data scalers to scale down the data into smaller values, which helps the Deep Learning models to converge faster during training (Huang *et al.,* 2023). Several studies have used different data scalers to scale down data for training and testing Deep Learning models.

Liang *et al.* (2019) predicted forest fires in Canada by using a Min-Max scaler for normalizing input features and a Long Short-Term Memory (LSTM) Deep Learning model, achieving good results with 90.9% accuracy. Shahriar *et al*. (2025) predicted FWI in the Continental United States (CONUS) by using a Min-Max scaler for normalizing input features and Graph Neural Network-Long Short-Term Memory (GNN-LSTM) Deep Learning model, achieving good results with Root Mean Squared Error (RMSE) score of 1.590. Chhetri *et al.* (2020) predicted monthly rainfall in Bhutan by using a Min-Max scaler for normalizing input features and a Bidirectional Long Short-Term Memory-Gated Recurrent Unit (BiLSTM-GRU) Deep Learning model, achieving good results with a low MSE (Mean Squared Error) score of 0.0075. Baljon *et al.* (2023) used a Min-Max scaler for normalizing input features and a Function Fitting Artificial Neural Network classifier to predict rainfall rate in Saudi Arabia, achieving good results with 96.1% accuracy. Liu *et al*. (2024) predicted short-term traffic by using a Standard scaler and Convolutional Neural Network-Bidirectional Gated Recurrent Neural Network with Additive Attention Mechanism (CNN-BiGRU-AAM) Deep Learning model, achieving good results with a coefficient of determination (R^2) score of 0.97. Amiri *et al.* (2024) proposed a system for fault detection and photovoltaic diagnosis by using Standard Scaler and a combined CNN and BiGRU Deep Learning model, achieving good results in detecting and classifying various faults. Villegas-Ch *et al*. (2023) used a Standard scaler for the normalization of input features and an LSTM Deep Learning model for predicting drought in Ecuador, achieving good results with 98.5% accuracy. Yhdego *et al*. (2023) used a Robust scaler for input data normalization and a GRU Deep Learning model to forecast flight delays in America, achieving good results, with the results revealing most of the predicted delays were within 95% of the predefined confidence level. Pandit *et al*. (2022) used a Robust scaler for input data normalization and a GRU Deep Learning model to forecast long-term weather in Germany, achieving good results, with a Mean Absolute Error (MAE) score of 1.06 in forecasting wind speed. Tsokov *et al.* (2022) predicted air pollution in China by using a Robust scaler for data normalization and a hybrid spatiotemporal model based on CNN and LSTM, achieving good results, with an R^2 score of 0.908.

Although existing studies reveal good performances of different data scalers, there is still a wide research gap as to which is the best-performing data scaler, especially in the context of forecasting fire danger at SAO Hill FP, which is located in Iringa, Tanzania, with a unique ecosystem and climatic conditions. The reviewed literature shows that different studies have used different data scalers in achieving their optimal results, suggesting the performance of the data scaler depends on the nature of the problem and the type of Deep Learning model used. Due to this fact, the choice of data scaler to use in the BiGRU Deep Learning model for forecasting fire danger at SAO Hill FP cannot be generalized but rather needs to be studied.

To address this research gap, this study aims to first develop three instances of BiGRU Deep Learning model (with each instance implemented with one of the three commonly used data scalers (Min-Max, Standard and Robust)), second, to comparatively evaluate the impact of the three data scalers on the performance of the developed BiGRU Deep Learning model in forecasting daily FWI (fire danger) at SAO Hill FP and third, to develop a Web App integrated with best performing BiGRU instance to allow SAO Hill FP officials to predict daily FWI. Based on these objectives, this study

Article DOI: https://doi.org/10.37284/eajit.8.1.3135

intends to answer one key research question: What is the impact of Min-Max, Standard, and Robust data scalers on the performance of the BiGRU Deep Learning model in forecasting daily FWI at SAO Hill FP?

MATERIALS AND METHODS

Research Design

This study used experimental research design by developing the BiGRU Deep Learning model, training it using training and validation data and testing its performance to predict daily FWI at SAO Hill FP using test data (unseen data).

Research Approach

This study used a quantitative research approach by utilizing quantitative data (13-year daily FWI data at SAO Hill FP) to train and evaluate the performance of the developed BiGRU Deep Learning model in predicting daily FWI at SAO Hill FP.

Data Collection and Analysis Methods

This study utilized secondary data by downloading SAO Hill FP daily FWI data from the CEMS-Fire-800m-Daily (Copernicus Emergency Management Service-Fire) dataset available in the Google Earth Engine (GEE) cloud platform. This study used a timeseries analysis of 13-year daily FWI data at SAO Hill FP to train and test the BiGRU Deep Learning model to predict daily FWI. Also, descriptive statistics was used to analyse the pattern of the 13-year daily FWI data at SAO Hill FP.

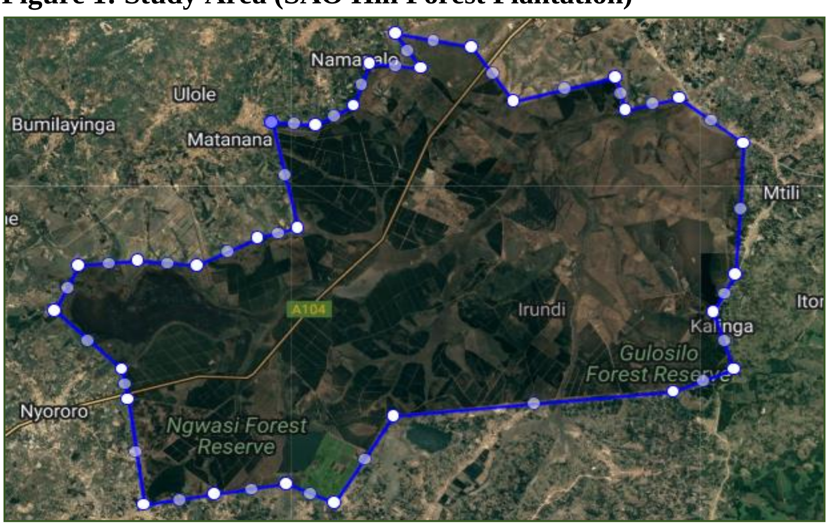
Sampling Technique and Sample Size

This study utilized purposive sampling, a non-probability sampling approach to select a single study area, SAO Hill FP. The study area was chosen because of the reliable availability of its historical daily FWI data, its susceptibility to FWI variability as well as its significance in the production of raw materials for the wood industry in Tanzania and outside Tanzania. A large sample size of temporal data points (daily FWI data for a total of 13 years) ensured sufficient data is available for training the developed BiGRU model and testing its prediction performance. Also, choosing only a single study area enhanced focus of the developed BiGRU model and reduced variability which could be caused by choosing many locations as they would have different climates.

Study Area

The area under this study is SAO Hill FP (enclosed by a blue-coloured polygon in *Figure 1*), located in the Mufindi district, Iringa region, Tanzania.

Figure 1: Study Area (SAO Hill Forest Plantation)



 This work is licensed under a Creative Commons Attribution 4.0 International License.

Article DOI: https://doi.org/10.37284/eajit.8.1.3135

Data Collection

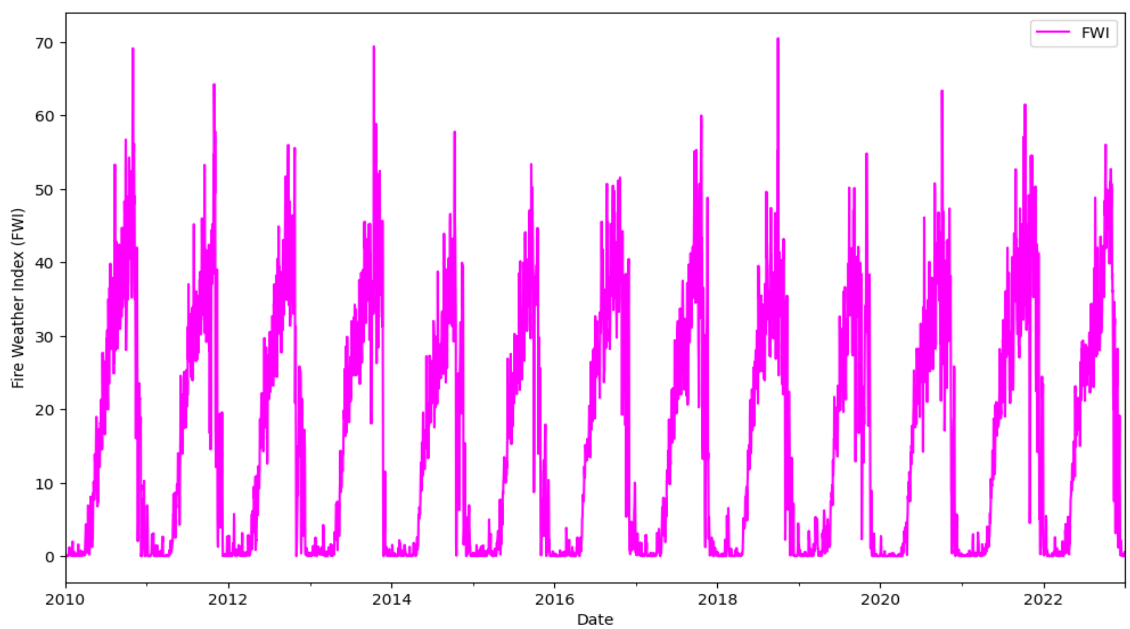
Daily average FWI timeseries data for SAO Hill FP for a period of 13 years (from January 1, 2010, to December 31, 2022) was downloaded from the CEMS-Fire-800m-Daily (Copernicus Emergency Management Service-Fire) dataset (CEMS, 2025). The dataset is hosted by the Google Earth Engine (GEE) cloud platform (Tamiminia *et al*., 2020) and was downloaded in CSV format from the GEE with the help of a Climate Engine application (Huntington, 2017). The downloaded daily FWI data were limited to an area enclosed by the blue-coloured polygon in Figure 1.

Data Pre-processing

Before feeding data into the Deep Learning models, the data needs to be pre-processed to be in a numerical format which is suitable for training and testing the Deep Learning models. This section describes the steps which were involved in pre-processing the data.

- Data Analysis**:** The 13-year daily FWI dataset for SAO Hill FP was visualized (refer to *Figure* 2) and afterwards analysed by using descriptive statistics, yielding the following results: count of 4748 records indicating a total number of daily FWI observations, a mean of 16.2742 indicating an average FWI value, a standard deviation of 16.0319 FWI value, a minimum FWI value of 0.0000, and a maximum FWI value of 70.5494.

Figure 2: Daily FWI Values for SAO Hill FP for 13 years



- Data Scaling: Three data scalers were used to normalize the 13-year daily FWI data for SAO Hill FP alternatively:
    - Min-Max Scaler: The FWI data was normalized (scaled down) to fit in the range of between 0 and 1 by using the Min-Max scaler (refer to equation (i)),

$$X_S = \frac{X - X_{Min}}{X_{Max} - X_{Min}} \qquad (i)$$

where X, X_{Max}, X_{Min} and X_S represent actual, maximum, minimum, and scaled FWI values, respectively.

 This work is licensed under a Creative Commons Attribution 4.0 International License.

Article DOI: https://doi.org/10.37284/eajit.8.1.3135

- Standard Scaler: The FWI data was normalized by using the Standard Scaler (refer to equation (ii)), where X, μ, σ and X_S represent actual, mean, standard deviation, and scaled FWI values, respectively.

$$X_S = \frac{X - \mu}{\sigma} \quad (ii)$$

- Robust Scaler: The FWI data was normalized by using the Robust Scaler (refer to equation (iii)), where X, X_{Median}, X_{75th}, X_{25th} and X_S represent actual, median, 75th percentile, 25th percentile, and scaled FWI values, respectively.

$$X_S = \frac{X - X_{Median}}{X_{75th} - X_{25th}} \quad (iii)$$

- Dataset Split: When developing a Deep Learning model, it is important to split the dataset into three sets (training, validation, and test sets). Training and validation sets are usually used during training of the model, with the role of a validation set being to evaluate the model's performance during training and tune (change) hyperparameters to attain best best-performing model. The role of a test set is to evaluate the prediction performance of the model on unseen data (never seen before) and measure its generalization capability. After scaling the 13-years FWI data for SAO Hill FP, the scaled FWI data was then split by using the following approach: training set (the first 60% of the data), validation set (the next 20% of the data) and test set (the last 20% of the data). There were three rounds of the data splitting process, each round corresponding to a particular data scaler.

- Input Features and Labels: The BiGRU model needs to learn how to predict the output (next day's FWI) by looking at the pattern of input (previous days' FWIs). A time-lag of 5 days was used, meaning a single pair of model input-output consisted of the previous 5 days' FWI values as input and the next day's FWI value as an output. The creation of input-output features was repeated for each record present in the training, validation, and test sets.

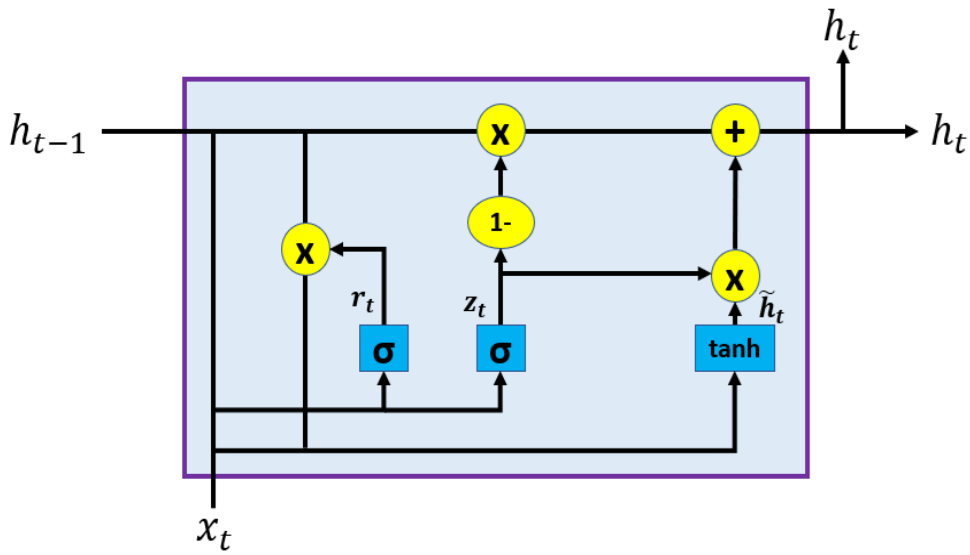
Architecture of the GRU Unit

GRU (refer to *Figure 3)* is a type of Recurrent Neural Network (RNN), and is usually used to process sequential data. The classical RNN model fails to remember useful information from past timesteps, which could become useful in future timesteps, a problem known as vanishing gradients. GRU addresses this problem by including in its architecture several gates to retain useful information from previous timesteps.

 This work is licensed under a Creative Commons Attribution 4.0 International License.

Article DOI: https://doi.org/10.37284/eajit.8.1.3135

Figure 3: GRU Architecture



At each timestep t The GRU unit receives the current input vector x_t and the previous hidden state h_{t-1} and provides an output which is a new hidden state h_t as follows.

- Reset Gate: The reset gate r_t is for deciding the extent of information to be forgotten from the previous hidden state before the new candidate hidden state is computed, as shown in equation (vi), where σ is the Sigmoid activation function (refer to equation (iv)), W_r is the weight matrix of the input, U_r is the weight matrix of the previous hidden state and b_r is the bias term. Because the Sigmoid activation function produces values ranging between 0 and 1, when r_t is close to 0, most of the previously hidden state information is forgotten, and when r_t It is close to 1, and most of the previously hidden state information is retained.

- Update Gate: The update gate z_t is for controlling the extent of information to be carried over to the next timestep from the previous hidden state, as shown in equation (vii), where W_z is the weight matrix of the input, U_z is the weight matrix of the previous hidden state and b_z is the bias term. When z_t is close to 0, most of the previous hidden state information is discarded, and when z_t It is close to 1, and most of the previously hidden state information is retained.

- Candidate Hidden State: The candidate hidden state $\tilde{h}_t$ controls the extent of the new information to be injected into the current hidden state, by combining the current input and past information from the reset gate, as shown in equation (viii), where $tanh$ (refer to equation (v) Is the activation function with e being the Euler's number, W_h is the weight matrix of the input, U_h is the weight matrix of the previous hidden state, b_h is the bias term and $\odot$ It is the element-wise multiplication. The past hidden state is retained if r_t is close to 1, enabling the candidate's hidden state to remember past information; otherwise, the previous hidden state is ignored if r_t is close to 0, forcing the candidate's hidden state to depend only on the new input.

- Final Hidden State: Computation of the final hidden state h_t is done by using the update gate, which makes decisions if the previous hidden state is kept or not kept (replaced by candidate hidden state) as shown in equation (ix). The hidden state remains the same if z_t is close to 1,

 This work is licensed under a Creative Commons Attribution 4.0 International License.

Article DOI: https://doi.org/10.37284/eajit.8.1.3135

while the hidden state is replaced by the candidate hidden state if z_t is close to 0.

$$\sigma(x) = \frac{1}{1+e^{-x}} \qquad (iv)$$

$$tanh(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}} \qquad (v)$$

$$r_t = \sigma(W_r x_t + U_r h_{t-1} + b_r) \qquad (vi)$$

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z) \qquad (vii)$$

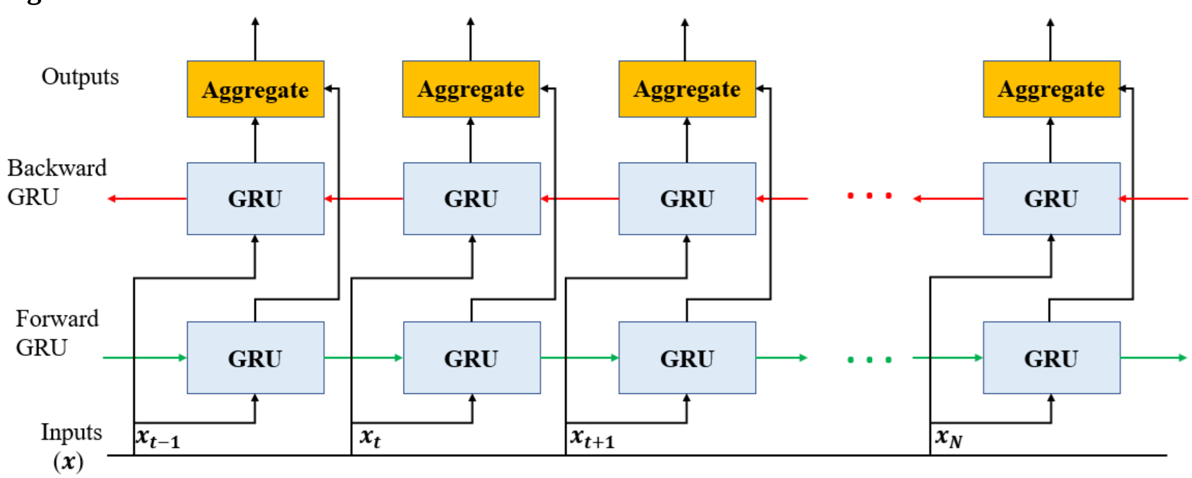
$$\tilde{h}_t = tanh(W_h x_t + U_h(r_t \odot h_{t-1}) + b_h) \qquad (viii)$$

$$h_t = (1 - z_t) \odot \tilde{h}_t + z_t \odot h_{t-1} \qquad (ix)$$

The architecture of BiGRU Unit

The architecture of the BiGRU unit shown in *Figure 4* is based on the GRU unit. However, unlike the GRU unit, which usually processes sequential information in only one direction (forward direction), the BiGRU unit processes information in two directions (forward and backward directions). The BiGRU unit contains two GRU units: the Forward GRU unit, which processes information from the first timestep to the last timestep, and the Backward GRU unit, which processes information from the last timestep to the first timestep. The output of the Forward GRU unit and the output of the Backward GRU unit are then aggregated to produce a single output of the BiGRU unit.

Figure 4: BiGRU Architecture



Proposed BiGRU Model

The architecture of the proposed BiGRU model is shown in *Figure 5* and consists of two BiGRU layers and one Dense layer. The role of the GRU layer is to learn patterns of input (previous days' FWI values) and how to map the input with the output (next day's FWI value). The role of a Dense layer is to produce a single output value as the predicted next day FWI.

This work is licensed under a Creative Commons Attribution 4.0 International License.

Article DOI: https://doi.org/10.37284/eajit.8.1.3135

Figure 4: Proposed BiGRU Model

Loss Function and Performance Evaluation Metrics

The loss Function plays a crucial role when training Deep Learning models because it computes the error between the true (actual) value y and the predicted value $\hat{y}$. Loss computation helps the BiGRU model to update its weights correctly to reduce FWI prediction error. By updating the model weights (parameters) correctly, the BiGRU model learns how to predict FWI values that are as close as possible to the actual FWI values and hence reduce the loss. In this study, Mean Squared Error (MSE) (refer to equation (x)) was used as a Loss Function.

On the other hand, the prediction performance of the developed BiGRU model on test data (unseen data) needs to be evaluated to measure its capability to generalize to new information that it has never seen before. This is done by measuring the error between the true (actual) value y and the predicted value $\hat{y}$. This study utilized the Root Mean Squared Error (RMSE) metric shown in equation (xi) to evaluate the performance of the developed BiGRU model.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \qquad (x)$$

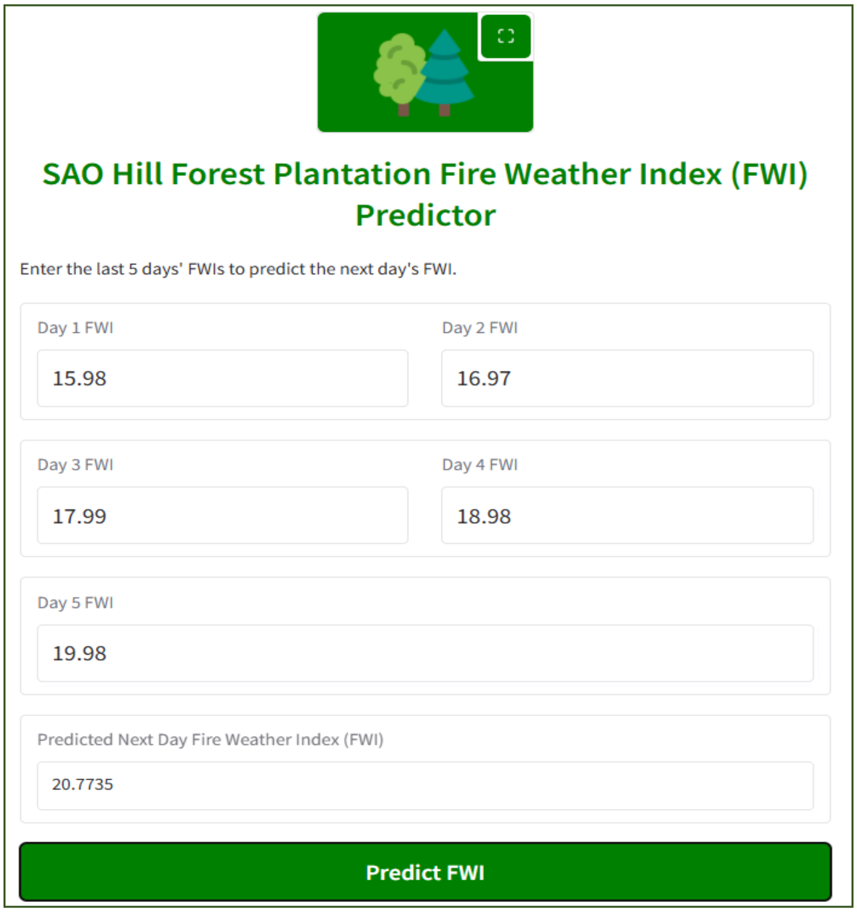
$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \qquad (xi)$$

WEB APP

This study developed a Web App shown in *Figure 5* using the Gradio framework (Ferreira *et al.*, 2024). SAO Hill FP official can access the Web App via a Web browser, use a CSS-styled Web page to enter the 5 previous days' FWI values, and click the ‘Predict’ button to predict the next day FWI. The Web App then takes the entered parameters (previous 5 days FWI values) and sends them to the imported and pretrained (already trained) BiGRU model in (.h5) format, which in turn takes the parameters as input, predicts the next day FWI, and returns the predicted next day FWI back to the Web App, which displays it back to the SAO Hill FP official.

 This work is licensed under a Creative Commons Attribution 4.0 International License.

Figure 5: Web App for Predicting Daily FWI at SAO Hill FP



RESULTS

Computation Environment

All of the training and testing experiments for the BiGRU model (developed in IPython notebook) were conducted on the Google Colab (Bisong, 2019) cloud platform using several software libraries including TensorFlow, Keras, Scikit-learn, Numpy, Pandas, and Matplotlib, and hardware run-time environment resource allocation of hard disk space of 107.7 GB and RAM of 12.7 GB.

Hyperparameters Tuning Experiments

Finetuning hyperparameters of the developed BiGRU model is a critical step during the training process, as hyperparameters have a direct impact on the performance of the Deep Learning model. After several hyperparameter tuning experiments for the BiGRU model, the following hyperparameters were chosen: 2 BiGRU layers, 1 Dense layer, output-dimensionality of 100 for the first BiGRU layer, output-dimensionality of 200 for the second GRU layer, batch-size of 16, learning rate of 0.001, Adam as an optimizer, and 100 training epochs.

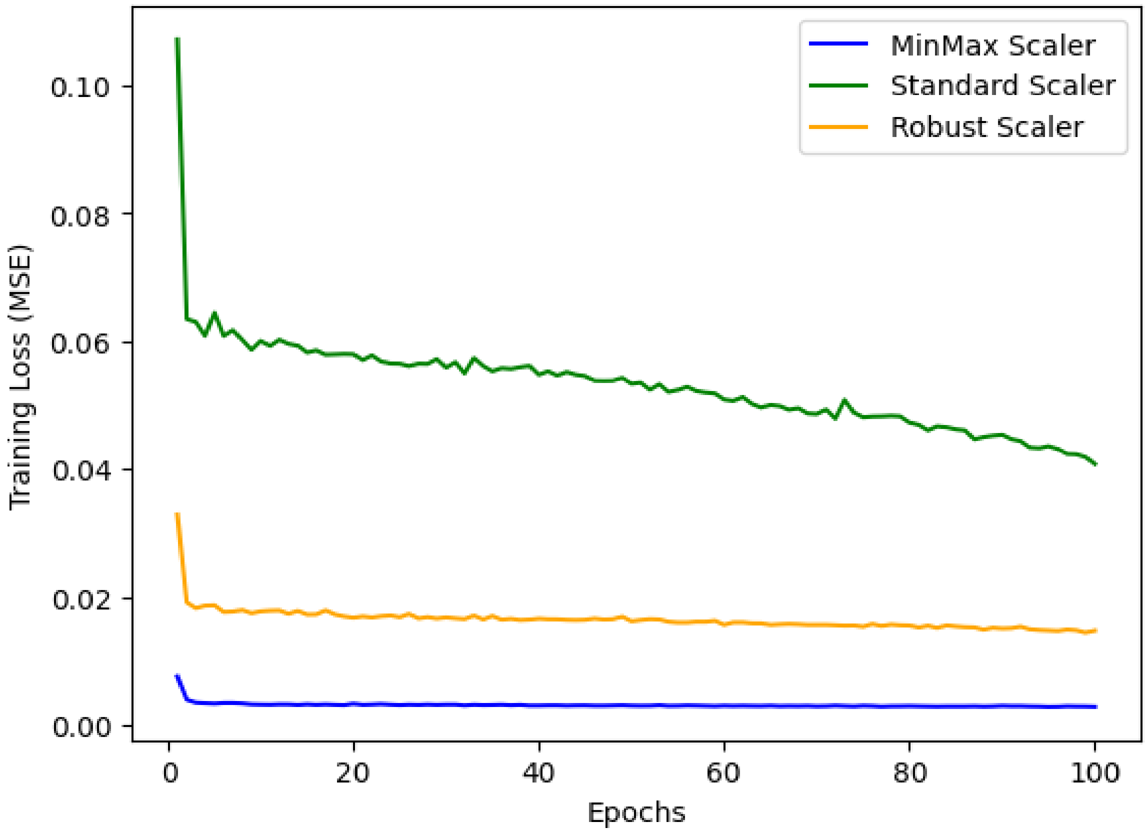
BiGRU Model Training Experiments

A total of three training experiments were conducted for the developed BiGRU model, with each training experiment using identical hyperparameters, the same BiGRU model, but a different data scaler. Three data scalers (Min-Max, Standard, and Robust) were used in training experiments 1, 2 and 3, respectively. In each experiment, the input was the 5 previous days' FWI values, and the output was the next day’s FWI value. After each training experiment, an instance of the trained BiGRU model (which corresponds to a particular data scaler) was saved in (.h5) format to

Article DOI: https://doi.org/10.37284/eajit.8.1.3135

enable future inference in the Web App by SAO Hill FP officials. *Figure 7* shows the training loss (MSE) of the BiGRU model when implemented with three different data scalers (Min-Max, Standard, and Robust). Since low training MSE scores indicate effective training and good training performance of the BiGRU model, these training results imply that the BiGRU model achieved best training performance when implemented with Min-Max data scaler, followed by Standard data scaler, and last, by Robust data scaler.

Figure 6: Training Loss of BiGRU Model Implemented with Three Different Data Scalers



BiGRU Model Performance Evaluation Results

After training experiments, each instance of the trained BiGRU model (the 3 instances correspond to 3 different data scalers used in the BiGRU model) was used to evaluate BiGRU's model performance on the test set (unseen data). *Figure 7* shows the plot of true (actual) FWI values against predicted FWI values by different instances of the BiGRU model.

On the other hand, Table 1 shows the prediction RMSE scores of the three instances of the BiGRU model on the test set. These results reveal that Min-Max is the best-performing data scaler in the BiGRU model for predicting daily FWI at SAO Hill FP, achieving a Test RMSE score of 0.065, followed by Standard scaler, which achieved a Test RMSE score of 0.157, followed by Robust scaler which achieved a Test RMSE score of 0.311.

This work is licensed under a Creative Commons Attribution 4.0 International License.

Article DOI: https://doi.org/10.37284/eajit.8.1.3135

Figure 7: True (Actual) vs Predicted Daily FWIs by BiGRU Model Implemented with Different Data Scalers at SAO Hill FP

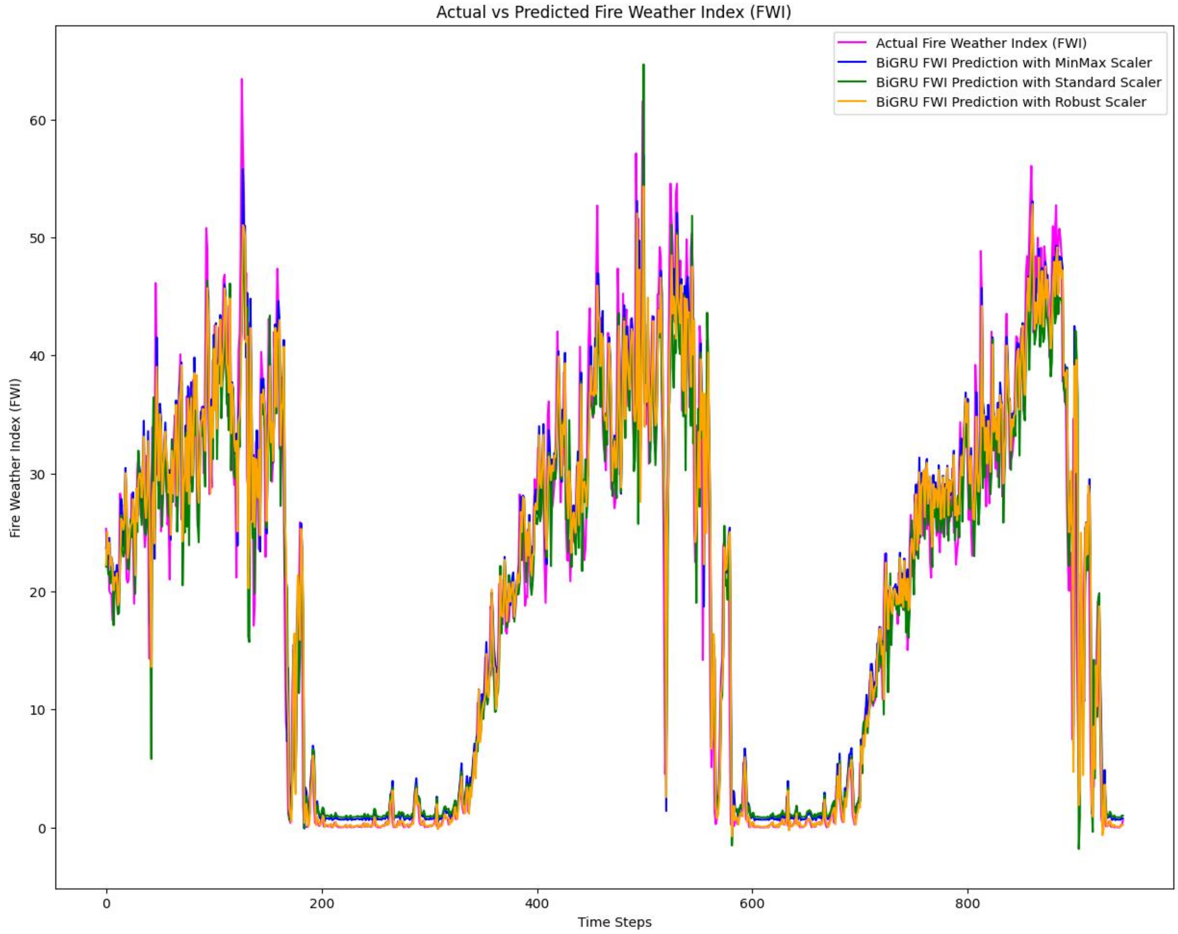


Table 1: Test RMSE Scores of BiGRU Model Implemented with Different Data Scalers at SAO Hill FP

Data Scaler	Test RMSE Score
Min-Max	0.065
Standard	0.157
Robust	0.311

DISCUSSION

Performance Differences of Data Scalers

The findings reveal that the Min-Max scaler achieved the best performance (lowest prediction RMSE score) on the test set, which demonstrates its superior performance in learning and preserving FWI data patterns in the BiGRU model. The next best-performing data scaler was the Standard scaler with a moderately higher prediction RMSE score on the test set, while the worst-performing data scaler was the Robust scaler with the highest prediction RMSE score on the test set. The difference in performance of the three data scalers may be attributed to their different characteristics. For instance, a Min-Max scaler scales the data to a fixed range, usually between 0 and 1, which helps Deep Learning models to converge faster and more accurately, especially in cases where the magnitude of data values differs but there are no extreme outliers. On the other hand, Standard Scaler scales the data to have a mean of 0 and a standard deviation

Article DOI: https://doi.org/10.37284/eajit.8.1.3135

of 1 and performs well in most use cases, although its performance might degrade when data is skewed or consists of outliers. Lastly, the Robust scaler is intentionally designed to handle outliers by using the median and interquartile range, hence, it might underperform when most of the data is clean and nicely distributed, as is the case in the SAO Hill FP FWI dataset.

Comparison with Findings from Literature

These findings suggest that the choice of data scaler has an impact on the performance of BiGRU models in forecasting (prediction) tasks of time series variables, and Min-Max is the best performing and optimal data scaler to use in BiGRU models when predicting daily FWI at SAO Hill FP, followed by Standard scaler, followed by Robust scaler. These findings align with the findings from the literature, which also suggest that a data scaler has an impact on the prediction performance of the Deep Learning model, but the best-performing data scaler depends on the type of Deep Learning model being used and the nature of the problem being addressed. This is evident in a study by Switrayana *et al*. (2025), which used a GRU model to predict stock prices with results revealing that the Min-Max scaler had 22.57% better RMSE performance than the Standard data scaler. Also, in a study by Christofi *et al*. (2024), which used an LSTM model to predict Arctic Sea ice melting, the results revealed that the Standard scaler was the best-performing data scaler with an RMSE score of 0.484, followed by the Robust scaler with an RMSE score of 0.574, followed by Min-Max scaler with an RMSE score of 0.758.

Major Contributions

The following are the major contributions of this study:

- Novel BiGRU Model: A Novel BiGRU model has been developed, which uses the optimal Min-Max data scaler to predict daily FWI at SAO Hill FP. The BiGRU model was trained and saved in (.h5) format to facilitate future inference by the Web App.
- Web App: A Gradio-based Web App has been developed to help officials at SAO Hill FP predict daily FWI and take appropriate measures in case the prediction shows high risk and danger of forest fires.
- Preprocessed Dataset: The 13-year SAO Hill FP daily FWI dataset has been preprocessed using several methods, including data scaling, creation of input and output features, and data splitting into training, validation, and test sets. The pre-processed dataset was then saved in (.pkl) format, ready to be imported and used by Deep Learning models. The preprocessed dataset will later be shared on the GitHub cloud platform to be freely accessed by the general public interested in AI research and development.
- Research Gap Filling: This study's findings help to fill the existing research gap on the impact of data scalers on the prediction performance of BiGRU Deep Learning models, especially in the context of predicting daily FWI at SAO Hill FP and environments resembling that of SAO Hill FP.

Practical Applications

The developed BiGRU model integrated with the Web App can be very useful in preventing forest fire ignition, responding to forest fire if it does happen, and mitigating its spread by helping the officials at SAO Hill FP take precautionary measures. After predicting FWI, which is deemed dangerous in the Web App, officials at SAO Hill FP can take the following precautionary measures:

- Forest Fire Prevention Measures: Officials can restrict all fire-related activities such as open burning, smoking, usage of sparks-generating chainsaws, patrol high-risk areas, and close access roads to prevent possible forest fires.

 This work is licensed under a Creative Commons Attribution 4.0 International License.

- Forest Fire Preparedness Measures: Officials can correctly position firefighting equipment such as fire trucks, water tankers, and portable pumps in high-risk areas, ensure water availability, remove dry grass and flammable debris around the forest, and create fire breaks.

- Forest Fire Emergence Response Readiness: Officials can alert emergency firefighters and conduct firefighting drills.

Study Limitations

Although this study reveals effective results in predicting forest fire danger at SAO Hill FP, it is worth mentioning the following limitations of this study:

- Limited Geographic Area: This study is based on SAO Hill FP and uses a single specific dataset of SAO Hill FP FWI to train the BiGRU model to predict forest fire. This limits its ability to capture variable environmental and climatic conditions across broader geographic locations.

- Operational Constraints Consideration: This study did not account for practical challenges and requirements of implementing the fire danger prediction system at SAO Hill FP, such as training of local staff, or integration with existing fire management workflows and protocols. This might impact the actual deployment of the developed Web App.

CONCLUSION

This study has developed the BiGRU Deep Learning model for predicting daily fire danger (FWI) at SAO Hill FP and evaluated its prediction performance when implemented with three different data scalers, with the results revealing Min-Max is the best-performing data scaler, achieving the lowest test RMSE score, followed by Standard scaler and Robust scaler respectively. This suggests the choice of data scaler has a direct impact on the performance of the BiGRU Deep Learning model in predicting FWI values. Also, the Min-Max is the optimal data scaler to use in the BiGRU model when predicting daily FWIs at SAO Hill FP and similar environments.

Recommendations

This study recommends the Min-Max data scaler as the optimal and practical data scaler to use in BiGRU-based Deep Learning models for predicting daily FWI at SAO Hill FP and environments with similar climatic conditions.

Acknowledgements

The author would like to give thanks to the Dar es Salaam Institute of Technology for supporting this research.

REFERENCES

Amiri, A. F., Kichou, S., Oudira, H., Chouder, A., & Silvestre, S. (2024). Fault Detection and Diagnosis of a Photovoltaic System Based on Deep Learning Using the Combination of a Convolutional Neural Network (CNN) and Bidirectional Gated Recurrent Unit (Bi-GRU). *Sustainability*, 16(3), 1012. DOI: https://doi.org/10.3390/su16031012

Baljon, M., & Sharma, S. K. (2023). Rainfall Prediction Rate in Saudi Arabia Using Improved Machine Learning Techniques. *Water*, 15(4), 826. DOI: https://doi.org/10.3390/w15040826

Bisong, E. (2019). Google colaboratory. In *Building machine learning and deep learning models on Google Cloud platform: a comprehensive guide for beginners*. Berkeley, CA: Apress.

Bouramdane, A. A. (2024). Assessing Global Wildfire Dynamics and Climate Resilience: A Focus on European Regions Using the Fire Weather Index. *Engineering Proceedings*, 68(1), 51. DOI: https://doi.org/10.3390/engproc2024068051

Article DOI: https://doi.org/10.37284/eajit.8.1.3135

Chhetri, M., Kumar, S., Pratim Roy, P., & Kim, B. G. (2020). Deep BLSTM-GRU Model for Monthly Rainfall Prediction: A Case Study of Simtokha, Bhutan. *Remote Sensing*, 12(19), 3174. DOI: https://doi.org/10.3390/rs12193174

Copernicus. (2025). Fire Weather Index. Retrieved April 8, 2025, from https://climate.copernicus.eu/fire-weather-index

Copernicus Emergency Management Service [CEMS]. (2025). CEMS-Fire. Retrieved April 19, 2025, from https://confluence.ecmwf.int/display/CEMS/CEMS-Fire

Ferreira, R., Canesche, M., Jamieson, P., Neto, O., & Nacif, J. A. (2024). Examples and tutorials on using Google Colab and Gradio to create online interactive student-learning modules. *Computer Applications in Engineering Education*, 32(4), e22729. DOI: https://doi.org/10.1002/cae.22729

Huang, L., Qin, J., Zhou, Y., Zhu, F., Liu, L., & Shao, L. (2023). Normalization techniques in training dnns: Methodology, analysis and application. *IEEE transactions on pattern analysis and machine intelligence*, 45(8), 10173-10196.

Huntington, J., Hegewisch, K., Daudert, B., Morton, C., Abatzoglou, J., McEvoy, D., & Erickson, T. (2017). Climate engine: Cloud computing and visualization of climate and remote sensing data for advanced natural resource monitoring and process understanding. *Bulletin of the American Meteorological Society*, 98(11), 2397-2410

Kangalawe, H. (2021). The Role of the World Bank in the Forestry Sector in Tanzania: A case of Sao Hill Plantation Forest, 1960-2010. *Tanzania Journal of Development Studies*, 19(1).

Liang, H., Zhang, M., & Wang, H. (2019). A neural network model for wildfire scale prediction using meteorological factors. *IEEE Access*, 7, 176746- 176755. DOI: https://doi.org/10.1109/ACCESS.2019.2957837

Liu, S., Lin, W., Wang, Y., Yu, D. Z., Peng, Y., & Ma, X. (2024). Convolutional Neural Network-Based Bidirectional Gated Recurrent Unit–Additive Attention Mechanism Hybrid Deep Neural Networks for Short-Term Traffic Flow Prediction. *Sustainability*, 16(5), 1986. DOI: https://doi.org/10.3390/su16051986

Mgina, M., & Wawa, A. (2021). Assessment of wildfires in Tanzania forest plantations: A case of Sao Hill in Mufindi District. *Huria: Journal of the Open University of Tanzania*, 27(2).

Pandit, R., Astolfi, D., Tang, A. M., & Infield, D. (2022). Sequential Data-Driven Long-Term Weather Forecasting Models' Performance Comparison for Improving Offshore Operation and Maintenance Operations. *Energies*, 15(19), 7233. DOI: https://doi.org/10.3390/en15197233

Shahriar, S. A., Choi, Y., & Islam, R. (2025). Advanced Deep Learning Approaches for Forecasting High-Resolution Fire Weather Index (FWI) over CONUS: Integration of GNN-LSTM, GNN-TCNN, and GNN-DeepAR. *Remote Sensing*, 17(3), 515. https://doi.org/10.3390/rs17030515

Shin, J. I., Seo, W. W., Kim, T., Park, J., & Woo, C. S. (2019). Using UAV multispectral images for classification of forest burn severity—A case study of the 2019 Gangneung forest fire. *Forests*, *10*(11), 1025.

Switrayana, I. N., Hammad, R., Irfan, P., Sujaka, T., & Nasri, M. (2025). Comparative Analysis of Stock Price Prediction Using Deep Learning with Data Scaling Method. *JTIM: Jurnal Teknologi Informasi dan Multimedia*, *7*(1), 78-90.

This work is licensed under a Creative Commons Attribution 4.0 International License.

Article DOI: https://doi.org/10.37284/eajit.8.1.3135

Tamiminia, H., Salehi, B., Mahdianpari, M., Quackenbush, L., Adeli, S., & Brisco, B. (2020). Google Earth Engine for geo-big data applications: A meta-analysis and systematic review. *ISPRS journal of photogrammetry and remote sensing*, 164, 152-170

Tian, M., Li, L., Wan, L., Liu, J., & de Jong, W. (2017). Forest product trade, wood consumption, and forest conservation—The case of 61 countries. *Journal of Sustainable Forestry*, 36(7), 717- 728. DOI: https://doi.org/10.1080/10549811.2017.1356736

Tsokov, S., Lazarova, M., & Aleksieva-Petrova, A. (2022). A Hybrid Spatiotemporal Deep Model Based on CNN and LSTM for Air Pollution Prediction. *Sustainability*, 14(9), 5104. DOI: https://doi.org/10.3390/su14095104

Villegas-Ch, W., & García-Ortiz, J. (2023). A Long Short-Term Memory-Based Prototype Model for Drought Prediction. *Electronics*, 12(18), 3956. DOI: https://doi.org/10.3390/electronics12183956

Yhdego, T. O., Wei, A.-T., Erlebacher, G., Wang, H., & Tejada, M. G. (2023). Analyzing the Impacts of Inbound Flight Delay Trends on Departure Delays Due to Connection Passengers Using a Hybrid RNN Model. *Mathematics*, 11(11), 2427. DOI: https://doi.org/10.3390/math11112427

Zhang, L., Wu, C., & Hao, Y. (2022). Effect of the Development Level of Facilities for Forest Tourism on Tourists' Willingness to Visit Urban Forest Parks. *Forests*, 13(7), 1005. DOI: https://doi.org/10.3390/f13071005

 This work is licensed under a Creative Commons Attribution 4.0 International License.