Detailed Report on Evapotranspiration (ET) and Rainfall Prediction

1. Introduction

This report presents an advanced forecasting approach for evapotranspiration (ET) and rainfall utilizing meteorological data acquired from the Hawaii Climate Data Portal (HCDP). Our study encompasses three major components: data acquisition and preprocessing, ET prediction using LSTM neural networks, and rainfall prediction employing a similar LSTM-based methodology. Additionally, future enhancements involving a novel rainfall prediction strategy, including an initial classification step followed by regression, are outlined.

1. Data Acquisition and Preprocessing

Data collection involved querying daily weather parameters—rainfall, maximum temperature, and minimum temperature—from the HCDP API across multiple farm locations from the years 2000 to 2025. The data obtained enabled computation of ET using the widely recognized Hargreaves–Samani equation, specifically designed to provide accurate estimations of ET based on temperature differentials and extraterrestrial radiation (Ra).

The data was systematically organized by farm and year into CSV files, facilitating streamlined model training and evaluation.

1. Locations Used in Model Creation

The models utilized data from various farm locations with distinct geographical coordinates as listed below:

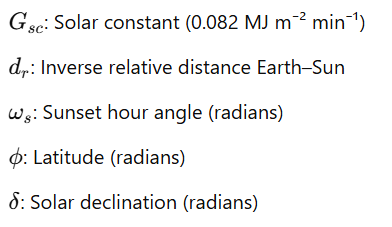
* Kahuku Farm: Latitude 21.6832, Longitude -157.9604
* Nozawa’ Farms: Latitude 21.6880, Longitude -157.9648
* Kuilima Farms: Latitude 21.6958, Longitude -158.0053
* Cabaero Farms: Latitude 20.8425, Longitude -156.3471
* Kupa'a Farms: Latitude 20.7658, Longitude -156.3513
* MAʻO Organic Farms (original site): Latitude 21.4645, Longitude -158.1132
* MAʻO Organic Farms (new site): Latitude 21.41505, Longitude -158.13707
* 2K Farm LLC: Latitude 21.445354, Longitude -158.181649
* Wong Hon Hin Inc: Latitude 21.466595, Longitude -158.164714
* Hawaii Taro Farm, LLC: Latitude 20.839723, Longitude -156.510438
* Hawaii Seed Pro LLC Farm: Latitude 20.796725, Longitude -156.359714
* Cabaero Farm: Latitude 20.791703, Longitude -156.358194
* Kupaa Farms: Latitude 20.765515, Longitude -156.35185
* Hirako Farm: Latitude 20.018748, Longitude -155.692546
* Hirako Farm: Latitude 20.002619, Longitude -155.694092
* Anoano Farms: Latitude 20.020913, Longitude -155.693966

These locations provide diverse climatic conditions that enhance the robustness of the predictive models.

1. Calculating Extraterrestrial Radiation (Ra)

Extraterrestrial radiation (Ra) refers to the solar radiation received at the Earth's atmosphere boundary, crucial for calculating evapotranspiration. The Ra is calculated using the following formula:



Where:

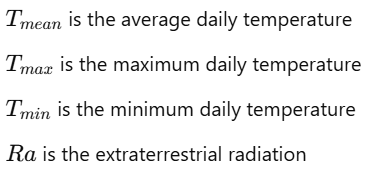
These parameters collectively determine the potential solar energy available at any specific location and date, accounting for variations due to Earth's elliptical orbit and axial tilt.

1. Computing Evapotranspiration (ET)

ET was calculated using the Hargreaves–Samani method, which leverages daily maximum and minimum temperature data alongside extraterrestrial radiation. The formula for ET calculation is:

Where:





This method effectively estimates ET based on temperature fluctuations and available solar radiation, making it a robust choice, especially in locations lacking extensive meteorological data. The calculated ET is an essential indicator for assessing water requirements in agricultural planning.

1. ET Prediction Using Multi-Step LSTM Model

The LSTM (Long Short-Term Memory) neural network was selected due to its proficiency in modeling sequential data, particularly effective in forecasting scenarios involving complex temporal dependencies.

6.1 Data Preparation

Historical data from multiple stations was meticulously prepared, segmenting datasets into training and testing subsets. Specifically, data from the Kupaa Farms station, spanning January 2021 to December 2022, served as the test dataset, while the remaining data contributed to model training.

Feature engineering techniques, such as extracting day and month from timestamps, were applied to enhance model effectiveness.

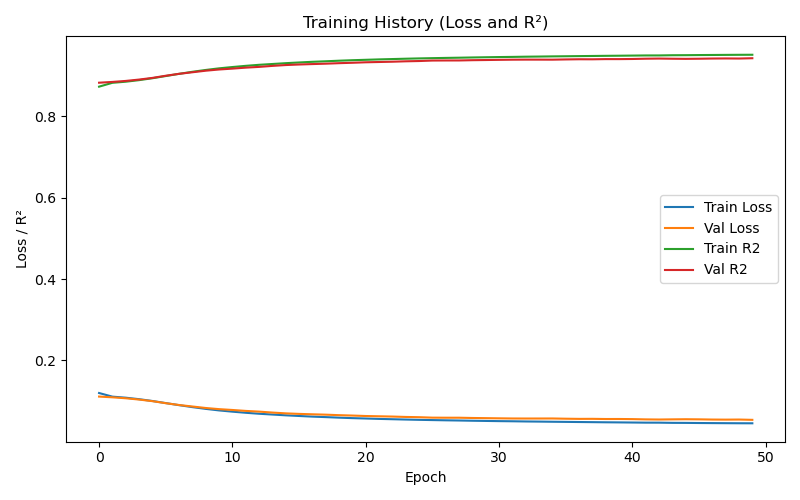
6.2 Modeling Approach

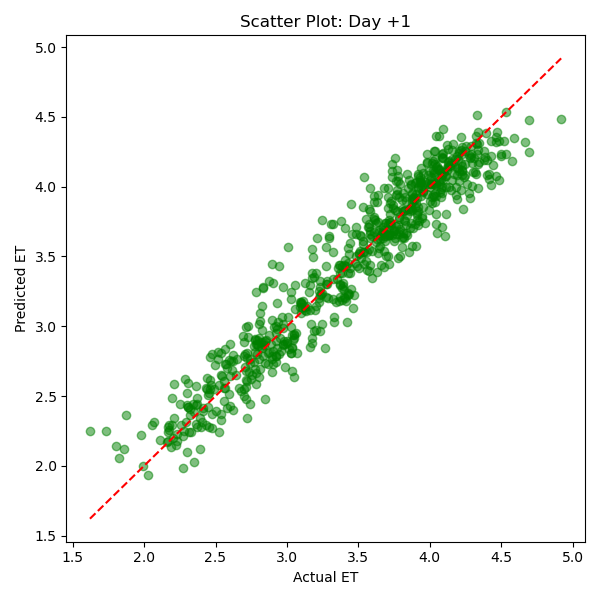
The ET model utilized sequences of 24 consecutive days as inputs to predict the subsequent three days' ET values. Standard scaling was applied to normalize data, ensuring uniform contribution of all features.

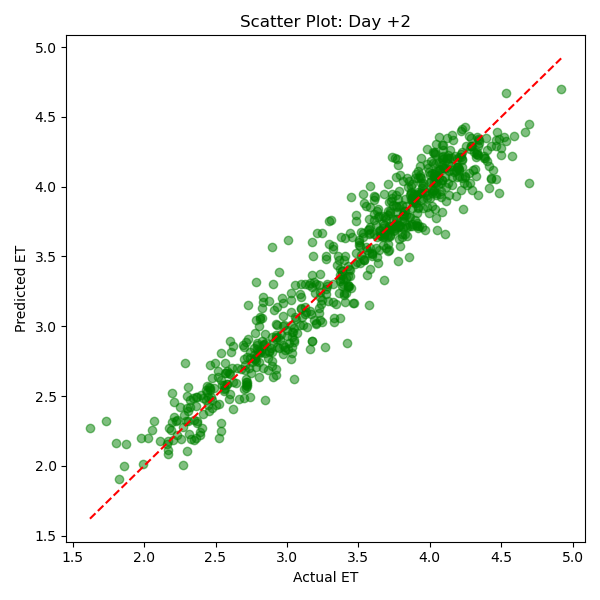
The LSTM model consisted of one recurrent layer optimized using the Adam optimizer with mean squared error (MSE) loss, alongside a custom-defined R² metric to evaluate predictive accuracy.

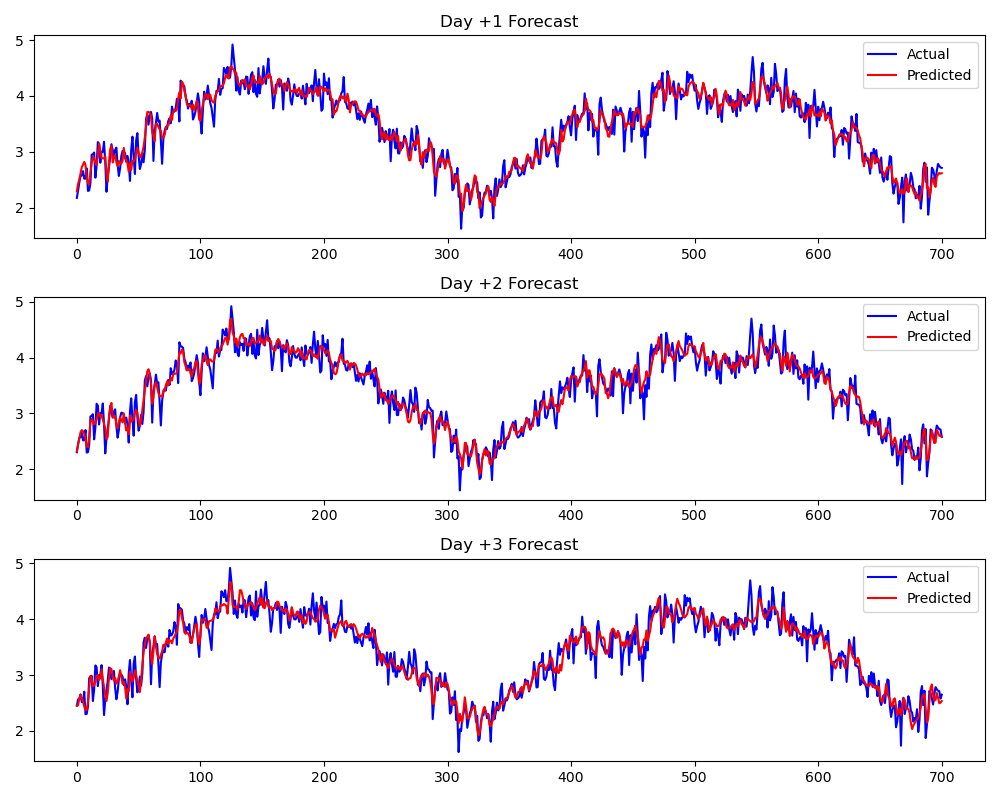
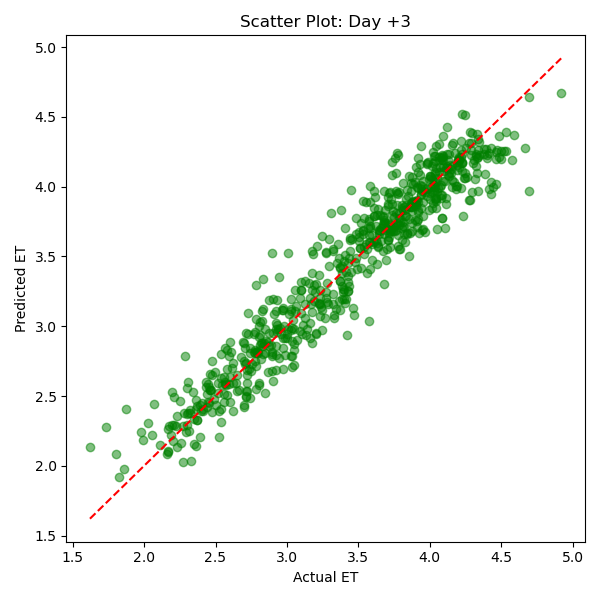
6.3 Model Training and Evaluation

The model training employed an 80/20 training-validation split, supported by early stopping to mitigate overfitting risks. Post-training evaluations utilized Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R² scores to assess model performance comprehensively.









7.Rainfall Prediction Using Multi-Step LSTM Model

The rainfall prediction followed an analogous methodology to the ET model, with adjustments primarily in the target variable. This uniformity in approach underscores the model's flexibility and general applicability to various meteorological predictions.

7.1 Data and Model Adjustment

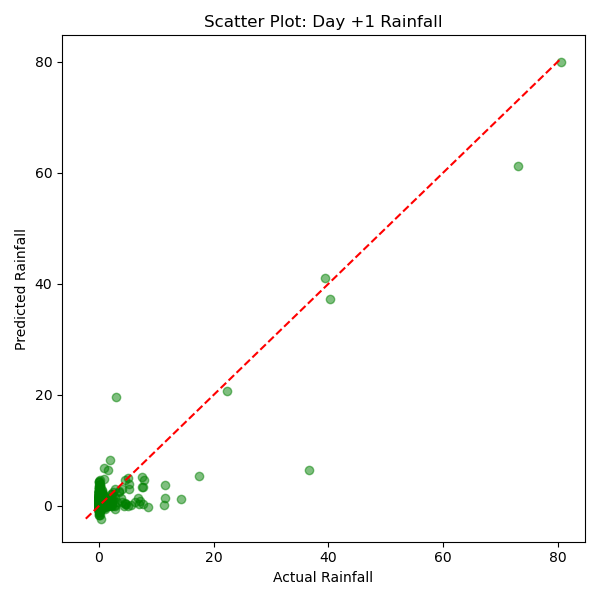
Rainfall data served as the primary target variable, following identical data preparation, scaling, and sequence creation protocols established in the ET model.

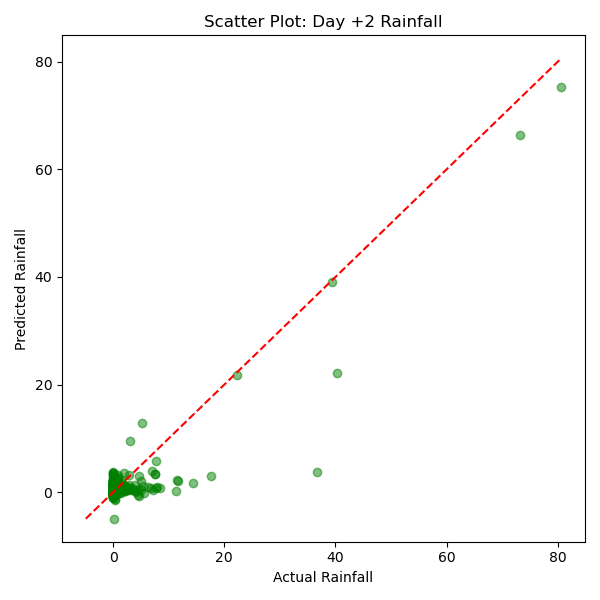
The architecture and training regimen remained consistent, ensuring comparability of results and ease of interpretation.

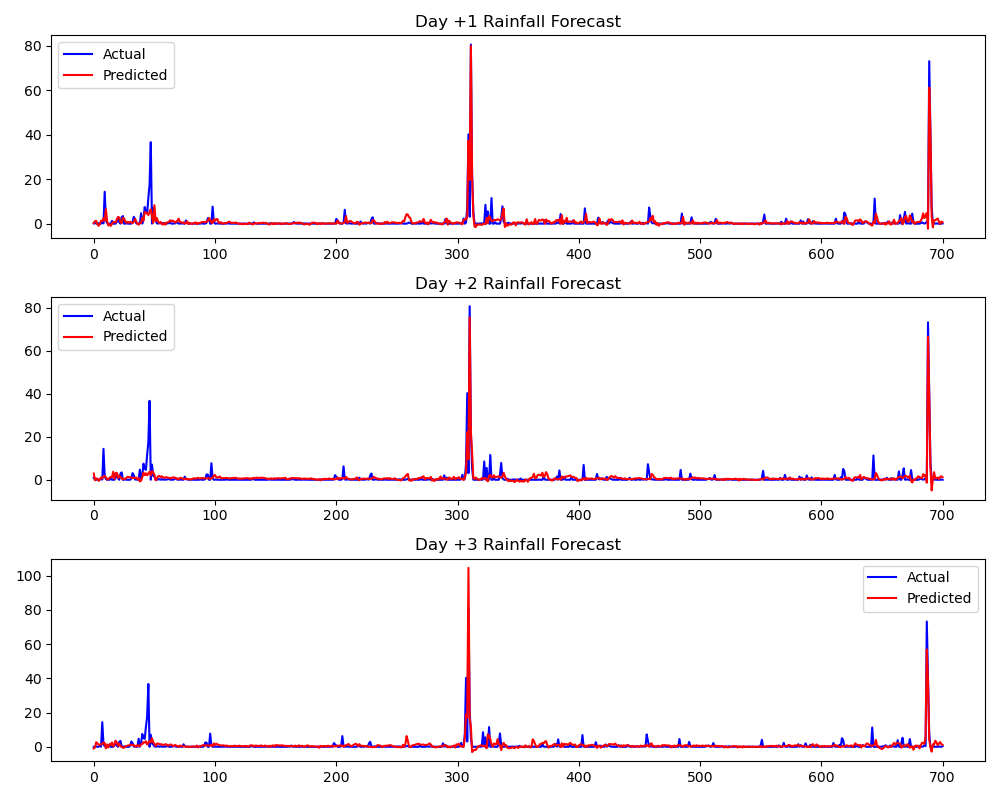
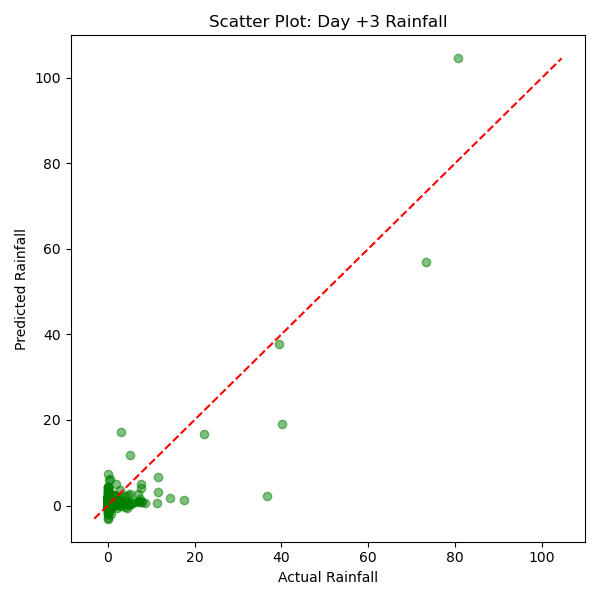
7.2 Evaluation of Rainfall Model

Evaluation metrics analogous to those used in ET prediction—MAE, RMSE, and R²—provided clear insights into model accuracy and reliability for rainfall forecasting.









8.Future Method for Improved Rainfall Prediction

In the future, an enhanced rainfall prediction strategy will be implemented. This innovative method initially classifies each day as either a rain day or a non-rain day. Following this classification step, a regression model such as LSTM will be applied exclusively to the predicted rain days to estimate rainfall amounts. The sequential classification-regression approach is anticipated to significantly enhance rainfall prediction accuracy. This method is currently in the developmental stage and will be refined in subsequent research efforts.

1. Conclusion

This comprehensive analysis demonstrates robust methodologies for forecasting crucial agricultural metrics—ET and rainfall—leveraging LSTM deep learning architectures. Through meticulous data preprocessing, sophisticated feature engineering, and strategic model training, our approach achieves reliable predictive performance. This model serves as a foundation for precise resource management and improved agricultural planning.

Future explorations could involve further hyperparameter tuning, exploring additional weather parameters, and incorporating real-time data for enhanced predictive precision.

End of Report