

Developing machine learning models for vegetation health index forecasting in the Srepok basin, Vietnam

Huynh Cong Luc¹, Viet Van Luong^{1*}

¹ Industrial University of Ho Chi Minh City, Ho Chi Minh City, 700000, Vietnam

* Corresponding author's e-mail: luongvanviet@iuh.edu.vn

ABSTRACT

This study developed machine learning models to predict the vegetation health index (VHI) using predictors derived from the standardized streamflow index (SSI), standardized precipitation index (SPI), and lagged VHI values. Three machine learning approaches were evaluated: stepwise quadratic regression (SQR), random forest (RF), and artificial neural network (ANN). Correlation analysis revealed that SPI at 4–5 month timescales and SSI at 1–3 month timescales exhibited the strongest relationships with VHI, particularly during January–July when vegetation is most sensitive to drought conditions. The models demonstrated reliable forecasting capability for 1–3 month lead times. ANN showed superior performance during the dry season (February–April), achieving correlation coefficients (*R*) of 0.79–0.87, Willmott's index of agreement (*d*) of 0.86–0.93, and RMSE of 11.5–16.4 for one-month lead time forecasts. In contrast, the SQR model yielded comparable accuracy to ANN with lower RMSE values and demonstrated better performance for longer lead times and during the rainy season, while RF yielded the lowest performance among the three models. The relative contribution of predictor variables varied by lead time, with VHI dominating one-month forecasts while SPI and SSI played primary roles in 2–3 month forecasts. These findings provide a scientific foundation for developing early warning systems to monitor meteorological and hydrological drought impacts on vegetation.

Keywords: vegetation health index, artificial neural network, random forest, stepwise quadratic regression, forecast.

INTRODUCTION

Climate variability has emerged as a global challenge with profound impacts on agricultural systems, water resources, and ecosystems worldwide. The Mekong River Basin, spanning six countries across Southeast Asia and China, is particularly vulnerable to climate variability and extreme weather events, including prolonged droughts that threaten food security and economic stability (Commission, 2020). Within this region, the Srepok River Basin in Vietnam represents a critical agricultural area where drought research and forecasting have become increasingly essential for sustainable resource management and disaster risk reduction (Khoi and Nhi, 2021).

Technologies for drought detection and early warning have significantly improved through the integration of remote sensing and modern data

processing approaches. The vegetation health index (VHI), derived from the vegetation condition index (VCI) and temperature condition index (TCI), has proven effective in assessing vegetation stress and drought impacts on agricultural systems (Kogan, 2002; Kogan, 1997). Unlike traditional meteorological drought indices that focus solely on precipitation deficits, VHI provides a more comprehensive assessment of vegetation response to temperature and moisture conditions, thereby enhancing its value for agricultural drought monitoring and crop yield forecasting (Liu et al., 2020; Wan et al., 2004).

The integration of multiple drought indices has shown considerable potential for improving forecast accuracy and reliability. The standardized precipitation index (SPI) (McKee et al., 1993) and standardized streamflow index (SSI) (Modarres, 2007) are among the most frequently

used indices to characterize meteorological and hydrological droughts, respectively. Previous studies have revealed distinct temporal linkages between these drought categories. Typically, meteorological drought emerges several months before agricultural drought (approximately 1–3 months), while hydrological drought tends to develop more slowly, reflecting delayed responses within the catchment system (Sam et al., 2019; Tram et al., 2021). Moreover, vegetation responses to deficits in rainfall and streamflow vary notably according to growth phase and season (Prajapati et al., 2021). The extent of climatic influence on vegetation also depends on regional conditions and environmental characteristics, including land cover, soil type, and vegetation composition (Usman et al., 2013).

In recent years, advances in machine learning (ML) and artificial intelligence (AI) have significantly improved the capacity for drought prediction and assessment. While traditional statistical techniques remain widely used, they often fail to capture the nonlinear relationships and complex interdependencies among multiple environmental variables that influence forecast outcomes (Narmilan et al., 2022). Various ML algorithms, including random forest (RF) (Mokhtari and Akhoondzadeh, 2021), artificial neural networks (ANN) (Kafy et al., 2023), stepwise quadratic regression (SQR) (Rahman et al., 2025), multiple linear regression (MLR), eXtreme gradient boosting (xGBoost), and support vector machines (SVM), have demonstrated strong performance in environmental modeling tasks owing to their ability to process large, multidimensional datasets with improved accuracy and efficiency (Feng et al., 2019; Luong and Bui, 2024). Nevertheless, the effectiveness of each algorithm depends on the characteristics of the dataset, emphasizing the importance of selecting the most suitable model for the specific environmental conditions of the study region.

The Srepok River Basin, covering approximately 18,600 km² in the Central Highlands, experiences a tropical monsoon climate with distinct dry (November–April) and rainy (May–October) seasons, making the region susceptible to seasonal drought that strongly affects agricultural productivity (Sam et al., 2019). The area features diverse land use types including perennial crops, annual crops, rice, and natural forests, each responding differently to drought impacts and requiring appropriate monitoring approaches.

Previous studies in the region have demonstrated strong relationships between drought indices and vegetation health, particularly emphasizing crop sensitivity to climate variability (Luong and Bui, 2023; Van Viet and Thuy, 2023). Nevertheless, studies that simultaneously examine the combined influence of meteorological and hydrological factors on VHI prediction remain scarce, particularly in relation to the differences in sensitivity among vegetation types and across various forecasting lead times.

Given these considerations, this study was conducted with the primary objective of developing and evaluating machine learning approaches for early prediction of the vegetation health index (VHI) within the Srepok Basin in Vietnam. Specifically, the research aims to (1) identify optimal timescales and lag times for input variables (SSI, SPI, and lagged VHI) to improve VHI forecasting accuracy; (2) comparatively evaluate the performance of three machine learning models (ANN, RF, SQR) across different forecast lead times; (3) analyze forecast quality variability by month to identify periods of optimal model performance; and (4) quantify the relative contribution of predictor variables to overall accuracy under different temporal conditions. Through integrating multi-index drought indicators and advanced machine learning techniques, this study seeks to develop a reliable VHI forecasting system, contributing to enhanced early warning capabilities for drought and supporting sustainable agricultural management in the region.

DATA AND METHODOLOGY

Research framework

The methodological framework of this study integrates multi-source datasets from satellite remote sensing, ground-based meteorological observations, and hydrological records into a unified monthly system with a spatial resolution of 463 m. The objective is to develop and evaluate machine learning models for forecasting the VHI across the Srepok River Basin (Figure 1).

The workflow consists of five main components: (i) data acquisition and harmonization; (ii) computation of drought-related indices (SPI and SSI) and vegetation-temperature indices (VCI, TCI, and VHI); (iii) correlation analysis to identify the most influential predictors and their

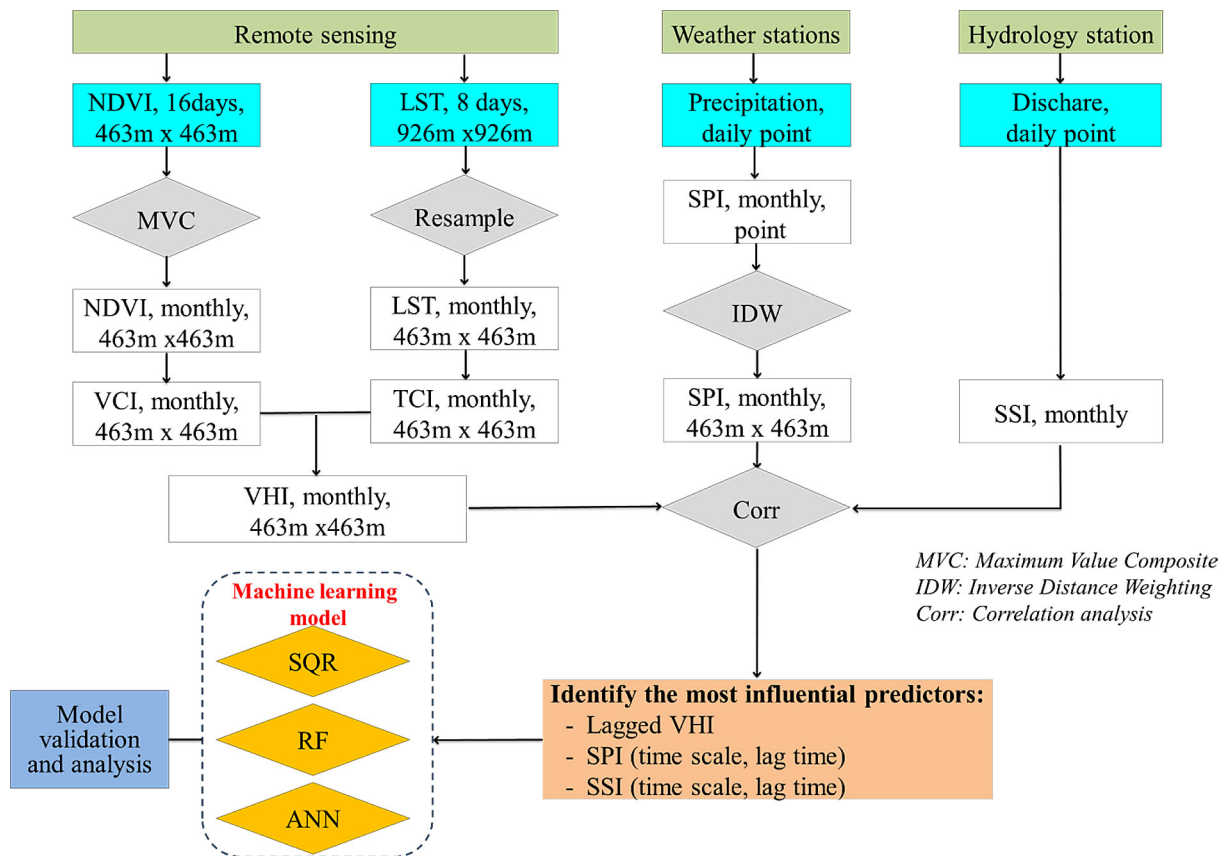


Figure 1. Workflow for developing machine learning models

optimal temporal lags; and (iv) model training and validation using three algorithms – SQR, RF, and ANN.

Data

Data utilized in this study comprised precipitation, streamflow, normalized difference vegetation index (NDVI), and land surface temperature (LST). Precipitation data were obtained from stations on and around the Srepok Basin. A total of 14 precipitation stations were included in the analysis, with 5 stations within the Srepok area and 9 stations in surrounding regions. Streamflow information was obtained from the Ban Don hydrological station, which is situated in the downstream portion of the Srepok Basin (Figure 2). This station monitors 92% of the Srepok Basin area. Data were collected from 2000 to 2022 as monthly averages.

NDVI and LST were employed to compute the VHI. This index was adopted in the study based on the works (Hiep et al., 2023; Van Viet and Thuy, 2023; Luong and Bui, 2023), which confirmed its strong responsiveness to drought

stress and its usefulness for predicting crop productivity in the Central Highlands region.

The NDVI and LST datasets were acquired from MODIS products available through the USGS EarthExplorer portal (<https://earthexplorer.usgs.gov/>). NDVI data were extracted from the MOD13A1 product, offering a spatial resolution of 463 m and a 16-day composite period, while LST data originated from the MOD11A2 product, which provides a 927 m spatial resolution and an 8-day temporal resolution. To ensure comparability between vegetation and drought indicators, all variables were resampled to a monthly time step. The monthly mean LST was computed from the 8-day MOD11A2 composites and then resampled to match the NDVI grid resolution of 463 m using the bilinear interpolation method. Meanwhile, monthly NDVI composites were produced from the 16-day MOD13A1 data using the maximum value composite (MVC) approach, which effectively minimizes the impact of cloud contamination and atmospheric interference (Chu et al., 2019; Holben, 1986; Li et al., 2016).

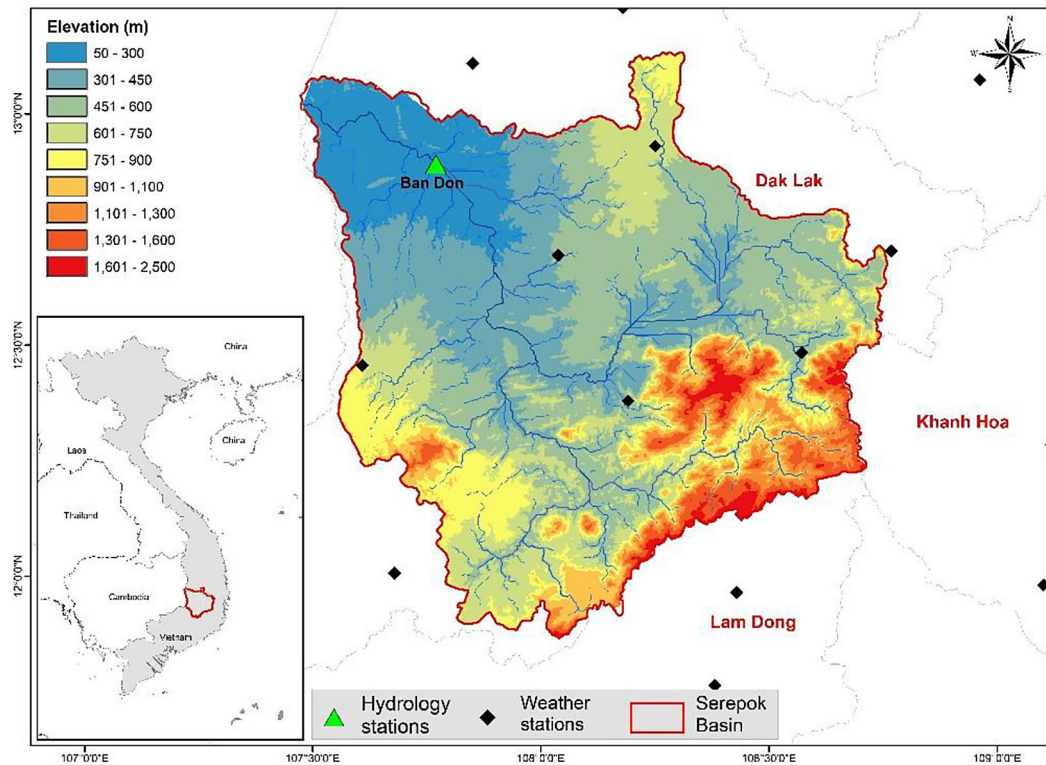


Figure 2. The Srepok River Basin and meteorological and hydrological stations

Methodology

Determination of VHI

The VHI is designed to quantify vegetation health by combining two sub-indices: the vegetation condition index (VCI) and the temperature condition index (TCI). This integration enables VHI to capture both vegetation vitality and the thermal stress influencing plant growth under drought conditions (Kogan, 1997; Kogan, 2002). Mathematically, VHI is expressed as:

$$VHI = \alpha VCI + (1 - \alpha)TCI \quad (1)$$

where: α is a weighting coefficient commonly assigned a value of 0.5. Both VCI and TCI represent the normalized temporal variations of NDVI and LST, respectively. These indices are computed according to the following formulations:

$$VCI = 100 \frac{NDVI_i - NDVI_{min}}{NDVI_{max} - NDVI_{min}} \quad (2)$$

$$TCI = 100 \frac{LST_{max} - LST_i}{LST_{max} - LST_{min}} \quad (3)$$

where: $NDVI_i$ denotes the NDVI value of a given pixel at time i during the study year; $NDVI_{max}$ and $NDVI_{min}$ correspond to the

maximum and minimum $NDVI$ values identified throughout the analysis period; LST_i represents the LST value of a specific pixel at time i for the same period; LST_{max} and LST_{min} indicate the maximum and minimum LST values observed across the temporal span used for computation.

Calculation of the standardized precipitation index (SPI)

The SPI introduced by McKee et al. (1993), is one of the most widely adopted drought indicators used globally. SPI quantifies precipitation anomalies by fitting observed rainfall to a probability distribution function, which is then normalized to produce standardized values representing wet and dry conditions. It can be computed for various timescales, typically ranging from 1 to several months, depending on the purpose of the analysis.

Let x denote the accumulated precipitation for a specific timescale and month. The SPI computation involves the following steps:

- Estimate the parameters of the Gamma distribution:

The shape (β) and scale (α) parameters are determined as:

$$\beta = \frac{1 + \sqrt{1 + 4U/3}}{4U} \quad (4)$$

$$\alpha = \frac{\bar{x}}{\beta}, \quad (5)$$

where U is given by:

$$U = \ln(\bar{X}) - \frac{\sum \ln(X)}{n} \quad (6)$$

and $\bar{x}$ represents the mean precipitation for the selected period.

- Compute the cumulative probability of the Gamma distribution:

$$G(x) = \frac{\int_0^x x^{\alpha-1} e^{-\frac{x}{\beta}} dx}{\beta^\alpha \Gamma(\alpha)} \quad (7)$$

where: $\Gamma(\alpha) = (\alpha - 1)!$, is the Gamma function. Because the gamma function is undefined at $x = 0$, the cumulative probability $H(x)$ is adjusted as:

$$H(x) = q + (1 - q)G(x) \quad (8)$$

with q being the probability of zero precipitation.

- Calculate SPI

The resulting value $H(x)$ is converted into a standardized variable with mean 0 and variance 1, producing the SPI value:

$$SPI = \frac{2.515517 + 0.802583t + 0.010328t^2}{1 + 1.432788t + 0.189269t^2 + 0.001308t^3} - t \quad 0 < H(x) \leq 0.5 \quad (9)$$

$$SPI = t - \frac{2.515517 + 0.802583t + 0.010328t^2}{1 + 1.432788t + 0.189269t^2 + 0.001308t^3} \quad 0.5 < H(x) \leq 1$$

where

$$t = \sqrt{\ln\left(\frac{1}{H(x)^2}\right)} \quad 0 < H(x) \leq 0.5 \quad (10)$$

$$t = \sqrt{\ln\left(\frac{1}{(1-H(x))^2}\right)} \quad 0.5 < H(x) \leq 1$$

SPI was calculated for each station and subsequently interpolated using the inverse distance weighting (IDW) method to generate raster layers at a spatial resolution of 463×463 m, consistent with the VHI data.

Calculation of the standardized streamflow index (SSI)

The SSI was first proposed by (Modarres, 2007) and later refined by (Telesca et al., 2012). It was developed using monthly streamflow values and standardization procedures analogous to those employed for SPI.

Similar to SPI, SSI quantifies hydrological drought conditions, but it substitutes precipitation data with streamflow observations as input. The calculation procedure follows the same statistical framework based on the Gamma probability distribution, allowing for both observed and forecasted streamflow datasets to be evaluated. The SSI provides valuable insight into streamflow variability, identifying periods of low discharge (drought) and high discharge (flood) events within river systems. In this study, SSI was computed at multiple temporal aggregation scales ranging from 1 to several months, to capture the different hydrological response times associated with basin storage and runoff processes.

Machine learning models

Machine learning approaches were employed to predict VHI values using a set of explanatory variables, including SPI, SSI, and antecedent VHI at different aggregation scales representing several months prior to the target month. Additionally, a temporal variable (t) was incorporated into each model to capture long-term trends throughout the study period. This temporal component was defined as the number of months elapsed from the beginning of the available dataset. To minimize the potential influence of seasonal vegetation dynamics, separate predictive models were developed for each month individually. This approach ensures that inter-month variability due to phenological changes is isolated from the effects of meteorological and hydrological factors.

- Stepwise quadratic regression (SQR) model

The SQR approach was applied to forecast the VHI, and its predictive performance was evaluated using independent test datasets. The input predictors included SSI, SPI, and antecedent VHI values.

With the stepwise quadratic regression technique, predictor variables are added or removed through statistical significance testing at each iteration. The VHI forecasting equation is expressed as:

$$Y_s = a + bt + \sum c_i x_i + \sum d_i x_i^2 \quad (11)$$

where: t is a temporal variable employed to remove trend influences from the simulation results, and x_i represents the predictor variables.

- Random forest (RF) model

Random forest (RF) is a machine learning algorithm based on ensemble learning principles, constructed from multiple decision trees (CART – Classification and Regression Trees) to enhance forecasting accuracy and stability (Breiman, 2001). During the training process, each tree is developed from a random bootstrap sample of the dataset, and at each splitting node, only a random subset of predictor variables is selected to determine the optimal splitting threshold. This dual randomization mechanism enables RF to overcome common limitations of individual decision trees, such as overfitting and sensitivity to training sample distribution. The same set of predictor variables used in the SQR model was employed for RF model development.

- Artificial neural network (ANN) model

The ANN approach represents a deep learning framework composed of multiple interconnected layers, each containing a series of neurons that process and transmit information. The training process utilized the backpropagation algorithm to minimize prediction errors and optimize the network's internal weights (Mas and Flores, 2008). In this study, a systematic optimization strategy was adopted to determine the most suitable network architecture, including the type of activation function, number of hidden layers, neurons per layer, and hyperparameters such as epoch count and batch size. The same predictor variables used in SQR and RF models were also employed for ANN development.

The random forest (RF) and artificial neural network (ANN) models were implemented in the Python programming environment using the scikit-learn (sklearn) and TensorFlow/Keras libraries, respectively. The stepwise quadratic regression (SQR) model was developed in the Fortran programming environment to ensure computational efficiency and maintain methodological consistency with previous hydrological and statistical studies.

Evaluation of forecast results and variable contribution

The performance of the VHI forecasting models was quantitatively assessed using three statistical indicators: the Pearson correlation coefficient (R), RMSE, and Willmott's index of agreement (d). These metrics were employed to measure the accuracy, reliability, and consistency of the model predictions. Their formulations are expressed as follows:

$$R = \frac{\sum_{i=1}^n (Y_i - \bar{Y})(Y_{Si} - \bar{Y}_S)}{\sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2} \sqrt{\sum_{i=1}^n (Y_{Si} - \bar{Y}_S)^2}} \quad (12)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_{Si} - Y_i)^2} \quad (13)$$

$$d = 1 - \frac{\sum_{i=1}^n (Y_i - Y_{Si})^2}{\sum_{i=1}^n (|Y_{Si} - \bar{Y}| + |Y_i - \bar{Y}|)^2} \quad (14)$$

where: n is the total number of observations, Y and Y_S represent the observed and predicted VHI values, respectively, and $\bar{Y}$ denote their mean values.

To improve the robustness of the forecast, a variable selection procedure was applied before model training. The preliminary selection was guided by correlation analysis, ensuring that only predictors exhibiting strong relationships with the target variable were retained. The predictive performance was further validated using the leave-one-out cross-validation (LOOCV) method, which minimizes bias in small datasets. Specifically, LOOCV was applied during the variable selection phase to identify the optimal subset of predictors for each month and forecast lead time.

The relative contribution of each predictor variable to the model's predictive capability was evaluated through a variable exclusion approach, in which the decline in the adjusted coefficient of determination (R^2) was examined when a specific variable was excluded from the model.

RESULTS AND DISCUSSION

Preliminary selection of predictor variables and forecast lead time

An excessive number of predictor variables can degrade the accuracy and generalization ability of forecasting models (Mehraein et al., 2022). Therefore, it is essential to select only variables

that provide meaningful and complementary information about the target variable. A predictor's informational relevance is considered high when it exhibits a strong correlation with the variable being predicted. Accordingly, the preliminary screening of predictor variables in this study was conducted based on the Pearson correlation analysis between each potential predictor and the forecasted VHI values.

Correlation coefficient between VHI and SPI

This analysis aims to determine the optimal timescale of the SPI index and the appropriate forecast lead time for the VHI index based on these indices. For SPI timescales (from 1 to 6 months), average correlation coefficients between SPI and VHI were analyzed from two perspectives: variation by month, and variation by lag time. Analysis results are presented in Figure 3.

The correlation analysis between SPI and VHI (Figure 3) exhibited the following characteristics: (1) Monthly analysis (Figure 3a) showed correlation coefficients varied markedly by season, reaching highest values from January to July. SPI4 and SPI5 timescales demonstrated high stability with correlation coefficients maintained in the range of 0.5–0.8 throughout this period. SPI4 for January is calculated from rainfall accumulation from October to February, and SPI4 for July is calculated from March to July rainfall. Due to this calculation method, SPI4 and SPI5 during this period incorporate precipitation from late rainy season of the previous year through early rainy season of the current year (October to July). This represents a critical period that captures water deficit conditions affecting

crop growth. During months 8–12, correlation coefficients decreased significantly, reflecting that vegetation becomes less sensitive when water is abundant; (2) According to Figure 3b, at lag times of 1 to 3 months, the correlation coefficients between SPI and VHI remain relatively strong. The combined analysis of lag time information indicated that SPI4 and SPI5 could be used to forecast VHI at lead times of 1–3 months ahead, corresponding to the forecast period from January to July.

Correlation coefficient between VHI and SSI

The correlation analysis between SSI and VHI (Figure 4) revealed patterns generally consistent with those obtained for SPI, though with several notable differences. (1) Monthly variation: As illustrated in Figure 4a, correlation coefficients between VHI and SSI were highest during February–July. However, short-term SSI timescales (SSI1–SSI3) produced significantly higher correlations than the longer timescales (SSI4–SSI6). This finding suggests that stream-flow variations exert a more immediate influence on vegetation health, whereas the effects of precipitation tend to accumulate over longer periods. (2) Lag-time variation: As shown in Figure 4b, correlations generally decreased with increasing lag time. The SSI1, SSI2, and SSI3 timescales exhibited similar declining trends, indicating a short response time of vegetation to hydrological changes. Overall, the results indicate that SSI1–SSI3 are the most suitable indices for evaluating and forecasting hydrological drought impacts on vegetation dynamics within the Srepok Basin.

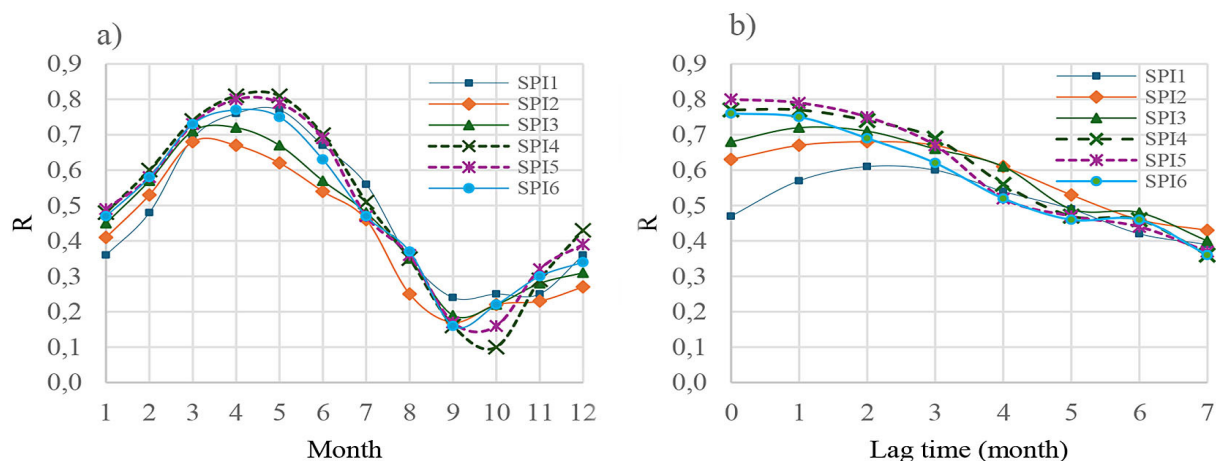


Figure 3. a) Time-lagged correlation between VHI and SPI, and b) maximum correlation at lags 0–7 months

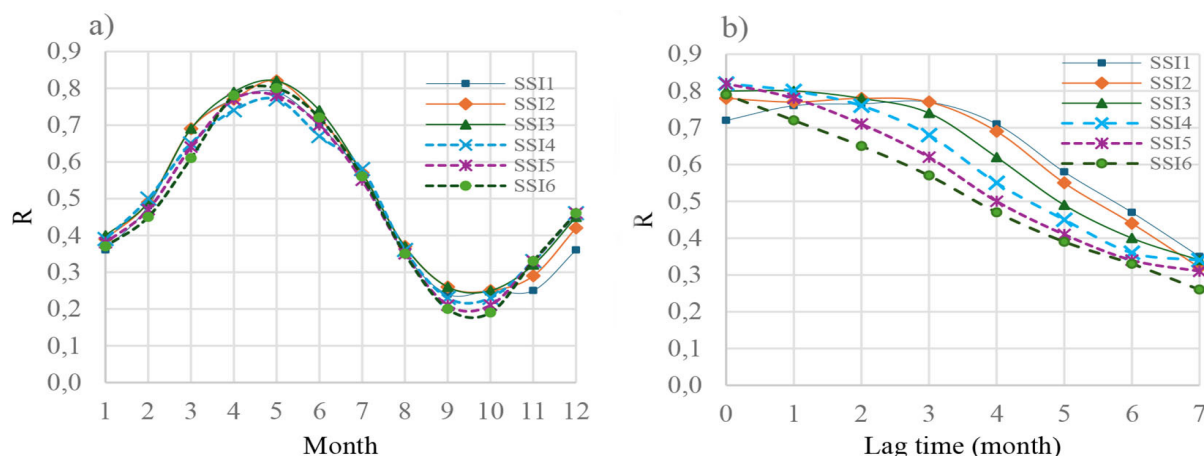


Figure 4. a) Time-lagged correlation between VHI and SSI, and b) maximum correlation at lags 0–7 months

The results indicate that SPI4-SPI5 and SSI1-SSI3 can be combined for VHI forecasting during January to July with forecast lead times of 1 to 3 months.

VHI autocorrelation coefficient

The autocorrelation coefficient represents the ability to forecast VHI based on vegetation health status at previous time steps. Results of the VHI autocorrelation coefficient calculation for the Srepok Basin are presented in detail in Figures 4c and 4d.

Monthly analysis (Figure 5c) showed autocorrelation coefficients varied strongly by season with clear patterns, reaching highest values during March–July. This indicates high forecasting capability during the transition from the late dry season to the early rainy season, when vegetation is particularly vulnerable to moisture stress. Lag time analysis (Figure 5d) provides detailed information about VHI forecasting capability across different time intervals. At 1-month lag, autocorrelation coefficients remained high, indicating good forecasting capability. At 2–3 month lags, correlation coefficients remained moderate, indicating that VHI forecasting based on lagged VHI can extend to 2–3 months. In other words, autocorrelation can be used to forecast VHI for months 3–7 with 1–3 month lead times.

Based on the magnitude of correlation coefficients and lag times, VHI for March and April (late dry season) shows high forecasting capability for VHI from April to July (transition period). January and February have the lowest precipitation amounts in the year (Figure 5a), and March and April are months with minimum

streamflow (Figure 5b); accumulated water stress during these months affects vegetation health in subsequent months.

Results from Figures 2, 3, and 4c–d show that the period from March to May is when correlation coefficients are highest for all index types. This characteristic indicates this is the period when vegetation is most sensitive to meteorological and hydrological drought conditions, and also the period with optimal forecasting capability. Additionally, the distribution of correlation coefficients reflects progressively increasing cumulative time impacts from January to July, then decreasing during the rainy season, reflecting the physiological cycle of vegetation under tropical monsoon climate conditions.

Dataset preparation and model development

As established above, the period from January to July was identified as the most suitable for VHI forecasting; therefore, subsequent analyses focused exclusively on these seven months, with forecast lead times of one to three months. The selected predictor variables included lagged VHI values, SPI4, SPI5, SSI1, SSI2, and SSI3, extracted for one to five months preceding each target forecast month. In total, 31 predictor variables were incorporated into the initial variable pool, including a temporal trend variable (t).

To improve the accuracy and reliability of monthly forecast models, all grid-cell data across the study area were aggregated into a single continuous time-series dataset. The study area encompassed 55,610 grid cells, and with 23 years of monthly data (2000–2022), the resulting dataset consisted of 1,279,030 individual

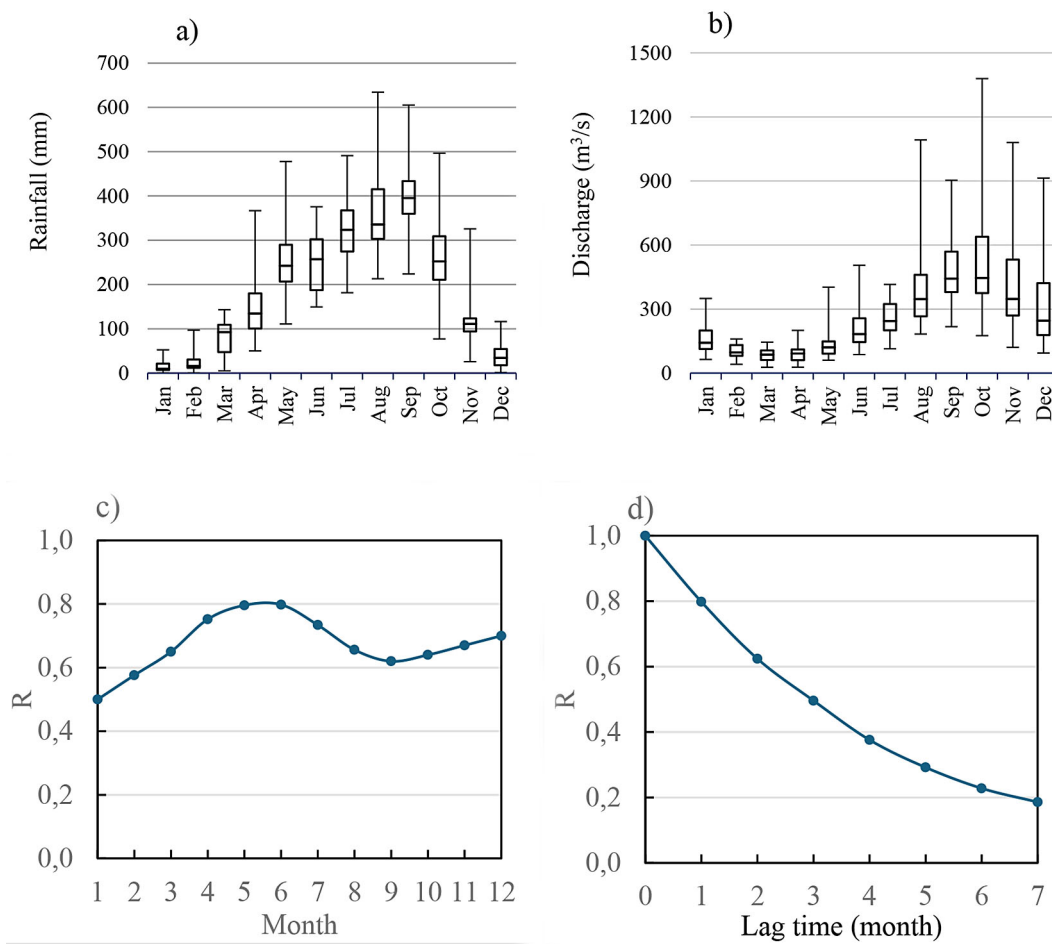


Figure 5. a) Precipitation over Srepok Basin; b) Streamflow observed at Ban Don station; c) Time-lagged autocorrelation of VHI, and d) maximum autocorrelation at lags 0–7 months

records. The data were partitioned temporally, with the period from 2000–2014 (approximately two-thirds of the temporal span) used for model training and calibration, while the remaining period from 2015–2022 (approximately one-third) served as an independent test set for model validation. This temporal splitting approach ensures that models are evaluated on completely unseen future data, thereby providing a realistic assessment of forecast performance.

Given the diversity of crop types and vegetation responses across the basin, standardization was necessary to harmonize the data. Accordingly, VHI values at each grid cell were normalized to a mean of 0 and a standard deviation of 1 using the training period statistics, ensuring comparability across space and time while preventing data leakage from the test set. This preprocessing step enhanced model convergence and allowed the machine learning algorithms to efficiently capture spatio-temporal patterns in vegetation health dynamics.

For each calendar month and forecast lead time, the stepwise variable selection procedure (for SQR), hyperparameter tuning (for RF and ANN), and model training were performed exclusively using the training dataset. The final models were then applied to the independent test set to generate the performance metrics reported in subsequent sections.

Model performance evaluation

This section presents the performance evaluation results for the three machine learning models (SQR, RF, and ANN) across different forecast lead times (1–3 months) and calendar months (January–July). Model performance was assessed using three statistical metrics: Pearson correlation coefficient (R), Willmott's index of agreement (d), and root mean square error (RMSE), calculated on the independent test dataset (2015–2022).

Models using SQR

In the stepwise quadratic regression (SQR) models, predictor variables for VHI forecasting were selected through a sequential regression process, as summarized in Table 1. The resulting coefficients confirmed that the constructed models were statistically robust across all months and forecast lead times. Table 2 presents the model coefficients for April with a one-month lead time. In Table 1, variables are organized according to their order of entry into the model, and the numerical suffixes indicate the lag time associated with each predictor (e.g., VHI.1 represents VHI lagged by 1 month, SSI1.2 represents SSI at 1-month timescale lagged by 2 months). Typically, three to five variables were retained in each model, representing the most influential predictors after the stepwise selection procedure. For the one-month lead time, the lagged VHI (VHI.1) consistently emerged as the most significant predictor, emphasizing the strong temporal persistence of vegetation health conditions. As the lead time increased to two or three months, the SSI and SPI indices gradually gained greater influence, reflecting the growing role of short- and medium-term hydro-meteorological variability in explaining vegetation response dynamics.

Forecast accuracy differed among months, as reflected by the variations in correlation coefficient (R) and Willmott's index of agreement (d) shown in Figure 6. According to this figure, the forecasting performance varied notably among

individual months, and forecast accuracy tended to decline as the lead time increased. The best prediction skill was observed during the late dry season and the onset of the rainy season (February–May). The period from February to April, which includes the final months of the dry season, exhibited the highest forecast quality. This period also coincides with the time when severe drought events commonly occur in the region (Luong and Bui, 2024; Sam et al., 2019). June and July had comparatively lower prediction accuracy, likely due to abundant rainfall and sufficient soil moisture, conditions under which vegetation health remains stable and VHI fluctuations are minimal. For all three forecast lead times, RMSE showed minimal variation and maintained low values during the early rainy season months, which are characterized by low VHI variability (Figure 7).

Models using RF

The RF model was implemented to exploit nonlinear learning capabilities and reduce overfitting commonly encountered in single decision trees. RF hyperparameters were determined through a grid search procedure with the objective of optimizing prediction accuracy, specifically minimizing RMSE while maximizing correlation coefficient (R) and Willmott index (d). The final hyperparameter set selected included: $n_estimators = 200$ (number of decision trees), $max_depth = 10$ (maximum tree depth), $min_samples_split = 2$ (minimum samples to split a

Table 1. Variables participating in forecast equations in the SQR model

Forecast lead time	Jan	Feb	Mar	Apr	May	Jun	Jul
1 month	VHI.1	VHI.1	VHI.1	VHI.1	VHI.1	VHI.1	VHI.1
	SSI1.1	SSI2.2	SSI2.2	SSI3.4	SPI4.4	SPI5.3	SPI4.3
	SPI4.1	SPI4.4	SPI5.3	SPI5.4	SSI1.4	SPI5.2	SPI5.3
	VHI.2	SPI4.1	VHI.4		VHI.3	SPI5.4	SPI4.4
	VHI.4		SSI2.4		VHI.2		VHI.4
2 months	SPI4.5	SSI2.2	VHI.2	VHI.3	VHI.2	VHI.3	SPI4.3
	SSI1.2	SSI3.5	SSI2.2	SSI3.4	SSI1.4	SPI4.4	VHI.2.5
	VHI.4	SPI5.2	SSI1.5	SPI4.5	SPI4.3	SPI5.5	SSI2.4
	SSI2.5	VHI.2	SSI3.5	SSI1.3	VHI.3	VHI.4	VHI.5
		VHI.5	VHI.4	SSI3.3			
3 months	SPI5.5	SSI3.5	SSI1.3	VHI.3	SPI4.3	SPI4.4	SPI4.3
	SPI4.5	SPI4.3	VHI.3	SSI3.4	SSI1.4	VHI.3	SSI1.5
	SSI2.3	SPI5.5	SPI5.5	SSI1.3	VHI.3	SPI5.5	VHI.3
	VHI.3	VHI.3	SSI1.5	SPI4.5	VHI.4	VHI.5	VHI.5
	SPI4.3	SSI1.3			SSI3.5		

Table 2. Coefficients of the SQR model with 1-month forecast time for April with standardized data

R	Adjusted R square	Standard error	F	Significance F
0.88	0.78	0.51	15665.4	0.00E+00
Parameter	Coefficients	Standard Error	t Stat	P-value
Intercept	-0.005	0.004	-1.4	1.5E-01
VHI1	0.768	0.004	174.6	0.0E+00
SSI3.4	0.154	0.005	29.4	4.1E-186
SPI5.4	-0.121	0.006	-18.9	7.7E-79

**Figure 6.** Model performance metrics (R, d, RMSE) for 1-month (a-c), 2-month (d-f), and 3-month (g-i) lead time forecasts

node), and $\text{min_samples_leaf} = 1$ (minimum samples at each leaf) node.

VHI forecasting performance evaluation results for 1–3 month lead times are shown in Figures 6 and 7. Overall, RF forecast quality was lower compared to both SQR and ANN models. R and d values tended to be highest during dry season and early rainy season months (February–May) when VHI fluctuated strongly, and noticeably lower for late rainy season months (June–July). This reflects the basin’s climate-hydrology characteristics: during the dry season, water deficits increase VHI sensitivity to meteorological and hydrological drought. Figure 7 also shows a close relationship between VHI standard deviation and RF forecast quality: months with large VHI variation (dry season) typically had higher R and d values; conversely, the rainy season with low VHI variation resulted in reduced forecast quality.

Despite the hyperparameter optimization, the RF model consistently underperformed compared to SQR and ANN across all months and forecast lead times. This may be attributed to the relatively limited size of the training dataset for each month-specific model, or to the inherent characteristics of vegetation-drought relationships in the study area, which may not align well with the ensemble tree-based approach. The detailed performance metrics are presented in Figure 6 and Table 3.

Models using ANN

The ANN model was established to capture complex nonlinear relationships between drought

indices and vegetation health status. Configuration parameters were determined through systematic experimentation, with optimization criteria of minimum RMSE and maximum R and d values. Results indicated that the sigmoid activation function provided the most suitable performance for this study. The model architecture included two hidden layers, in which the first layer contained 20 neurons and the second layer consisted of 10 neurons. Training was performed using backpropagation with a learning rate of 0.01, batch size of 500, and 128 epochs. Due to the large number of weights and adjustment coefficients in the network, detailed parameter values are not presented here.

Figure 8 illustrates the distribution between actual and predicted VHI values for April. The results revealed that when observed VHI values were relatively low, the predicted values tended to be slightly higher, whereas when VHI values were high, predictions were slightly lower. The smallest errors occurred when VHI values were near the median.

The evaluation of VHI forecasting accuracy for each month, based on the independent test dataset (2015–2022), is summarized in Table 3 and Figure 6.

For 1-month lead time (Figure 6a-c): ANN recorded the highest statistical indices among the three models. April achieved $R = 0.87$, $d = 0.93$, and $\text{RMSE} = 11.5$, which was equivalent to SQR in correlation coefficient (0.87) but superior in d index and RMSE values. In months 2–3

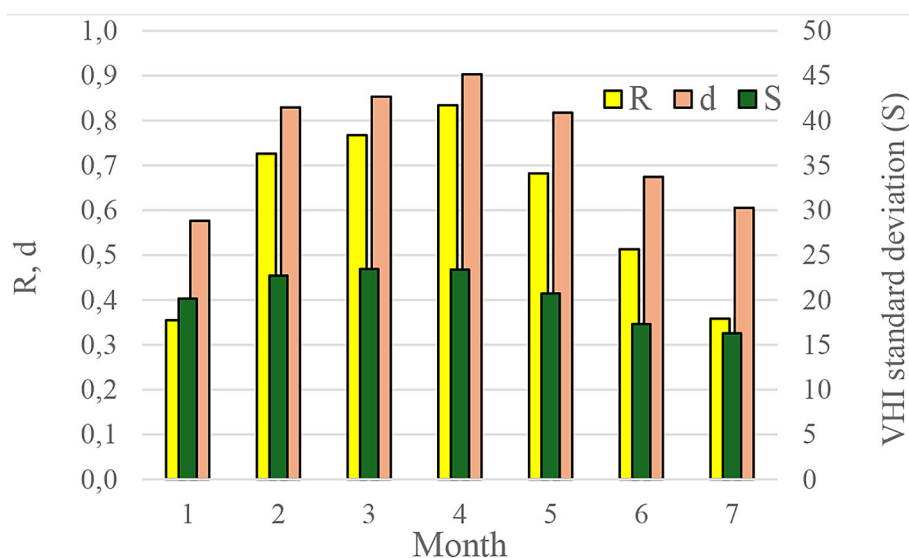


Figure 7. R and d values between actual and predicted results using RF model for 1-month forecast and VHI standard deviation by month

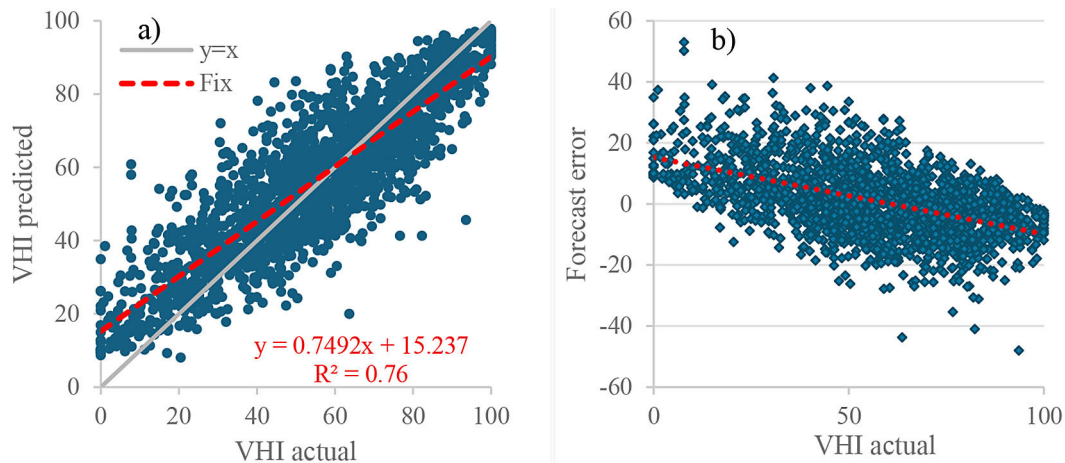


Figure 8. Model performance of ANN for April: (a) Relationship between observed and predicted VHI; (b) Forecast error

(February–March), ANN continued to achieve $R = 0.79$ – 0.81 and $d = 0.86$ – 0.88 , reflecting high sensitivity in detecting VHI variation during drought periods. According to Table 3 which summarizes the percentage of months each model achieved best values for R , RMSE, and d , the ANN model outperformed the other models in 57.1% of months for R , 42.9% for RMSE, and 100% for d index. This demonstrates that ANN is the optimal choice for 1-month lead time forecasting, due to its ability to accurately and stably reproduce VHI variation, although SQR sometimes produced lower absolute errors.

For 2-month lead time (Figure 6d-f): ANN continued to show good forecast quality. March achieved $R = 0.74$, $d = 0.82$, and $RMSE = 16.5$, higher than SQR ($R = 0.73$, $d = 0.72$) and RF ($R = 0.69$, $d = 0.80$). In April, SQR had a slightly higher R (0.75 vs. ANN's 0.73), but ANN still achieved higher d (0.83 vs. 0.74), confirming stability in model agreement with observations. Across all months, ANN led in R for 57.1% of months and d for 85.7% of months, while SQR dominated in RMSE with 85.7% of months. Thus, ANN remains the appropriate choice for 2-month lead time forecasting, especially when the objective is maintaining ability to track VHI variation, while SQR is more suitable when prioritizing absolute error reduction.

For 3-month lead time (Figure 6g-i): The performance of all three models declined significantly as expected for longer lead times. ANN still maintained lower RMSE and higher d than RF, while in some months having performance equivalent to or slightly better than SQR. For example, March achieved $R = 0.56$, $d = 0.70$,

and $RMSE = 20.6$. However, in April and July, SQR recorded higher R (0.69 and 0.49 vs. ANN's 0.67 and 0.33). Across all months, SQR led in R (57.1%) and RMSE (85.7%), while ANN retained its advantage in d (71.4%). This shows that the relative advantage of ANN diminishes when lead time extends, and SQR becomes a more appropriate choice for 3-month lead time forecasting, particularly during periods of low VHI variation.

Monthly analysis showed that ANN was most suitable during the dry season and early rainy season (January–May), with $R > 0.70$, $d > 0.80$, and lower RMSE compared to both SQR and RF. For example, in April with 1-month lead time, ANN achieved $R \approx 0.87$ and $d \approx 0.93$, demonstrating high reliability during the period when vegetation is most sensitive to drought stress. Conversely, during the rainy season (June and July), ANN performance decreased significantly, with $R = 0.51$ – 0.53 and $d = 0.62$ – 0.63 for 1-month lead time, although RMSE remained at low levels (June ≈ 14.4 ; July ≈ 13.8). During the rainy season, SQR achieved higher correlation coefficients, reflecting more stable linear relationships when VHI fluctuation was minimal due to abundant water conditions. Meanwhile, RF exhibited the lowest forecast quality across all months and lead times, especially in July with 3-month lead time when the correlation coefficient only reached 0.25. The degree of performance decline when increasing lead time confirms that ANN is the most stable model, with an average decrease of 15–20% in R per additional month of lead time, lower than SQR (18–25%) and RF (25–35%). This stability is particularly evident during the dry season,

Table 3. Number of months each model achieved best values for R, RMSE, and d by forecast horizon

Forecast lead time	Index	RF (months, %)	ANN (months, %)	SQR (months, %)	Total (months, %)
1 month	R	0 (%)	4 (57.1 %)	3 (42.9%)	7 (100%)
	RMSE	0 (%)	3 (42.9%)	4 (57.1 %)	7 (100%)
	d	0 (%)	7 (100%)	0 (%)	7 (100%)
2 months	R	0 (%)	4 (57.1 %)	3 (42.9%)	7 (100%)
	RMSE	0 (%)	1 (14.3%)	6 (85.7%)	7 (100%)
	d	0 (%)	6 (85.7%)	1 (14.3%)	7 (100%)
3 months	R	0 (%)	3(43%)	4 (57 %)	7 (100%)
	RMSE	0 (%)	1 (14.3%)	6 (85.7%)	7 (100%)
	d	1 (14.3%)	5 (71.4%)	1 (14.3%)	7 (100%)
Total	R	0 (%)	11 (52.4%)	10 (47.6%)	21(100%)
	RMSE	0 (%)	5 (23.8%)	16 (76.2%)	21(100%)
	d	1 (4.8%)	18 (85.7%)	2 (9.5%)	21 (100%)

when ANN maintains high accuracy even with 2–3 month lead times.

Thus, results show ANN is the optimal model for 1–2 month lead time forecasts, especially during the dry season when VHI variation is large and drought impacts are evident. SQR is more suitable for longer 3-month lead time forecasts or during the rainy season, when linear relationships between predictor variables and VHI dominate. In contrast, RF had the lowest forecast quality and poor stability across all conditions, and therefore is not recommended for VHI forecasting in the Srepok Basin.

Spatial performance analysis revealed clear differences in ANN forecasting accuracy among land-cover types (Figure 9). Overall, the ANN model exhibited strong predictive skill across most of the basin, with $R > 0.8$, Willmott's $d = 0.85$ – 0.95 , and $RMSE < 15$. The highest accuracy was observed in agricultural regions (40.9% of the basin), including perennial crops, paddy rice, and annual croplands, where vegetation responds rapidly to rainfall and streamflow variability. These regions showed strong agreement between observed and predicted VHI values. In contrast, forest-dominated areas (44.3% of the basin), comprising protection, special-use, and production forests, displayed moderately lower performance ($R = 0.6$ – 0.8 and $RMSE > 18$), particularly in the southern and western uplands. This difference primarily reflects ecological characteristics: forest ecosystems with dense canopies and deep root systems maintain soil moisture and physiological stability, resulting in weaker and slower vegetation responses to short-term hydroclimatic fluctuations. Meanwhile, agricultural crops with shallow

roots and short growth cycles are more sensitive to water deficits, producing greater VHI variability that the ANN model captures more effectively.

Spatially, the ANN performed best across the central and eastern agricultural plains, where vegetation dynamics are closely linked to rainfall and streamflow variability. These spatial patterns reaffirm the model's capability to capture nonlinear drought–vegetation interactions, particularly in cropland areas directly affected by seasonal water stress.

Role of variable groups in forecasting

The analysis presented above showed that ANN produced higher R values compared to RF and SQR; therefore, it was selected as the basis to determine the relative contribution of predictor variable groups, with results presented in Table 4. This table shows that variable contributions differ clearly between lead times and between months. For the 1-month lead time, lagged VHI made the largest contribution to model performance in most months. During January–April, when vegetation is most affected by drought stress, VHI contributed 74–88% to forecast accuracy. In subsequent months, the contribution declined to 54–63%. SPI ranked second in importance, reaching its highest values in June and July (44–46%), but dropped below 10% in March and April. Meanwhile, SSI exhibited greater influence in the early dry season, accounting for approximately 20% during January–March, and fell below 10% thereafter.

When the forecast time increased to two months, the relative importance of VHI, SPI, and SSI became more balanced, with alternating

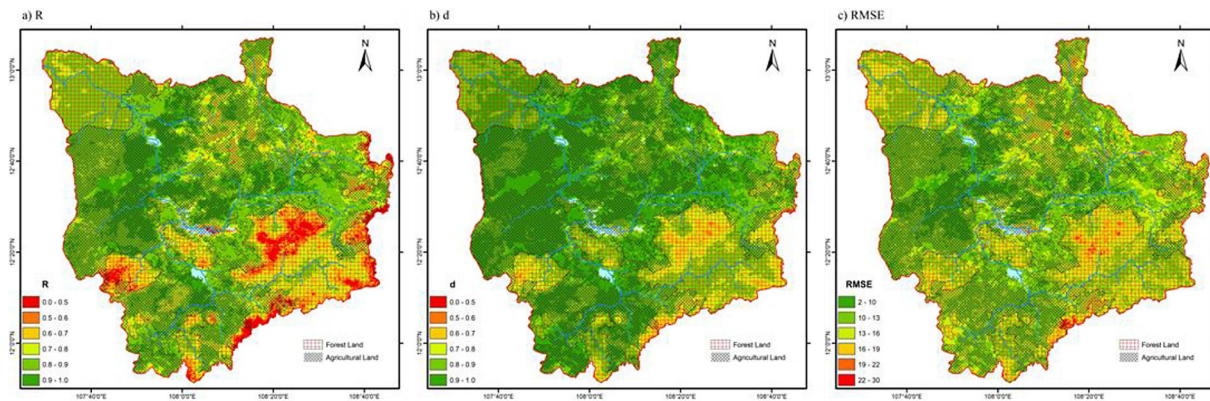


Figure 9. Spatial performance of the ANN model for VHI forecasting across the Srepok Basin, showing (a) R, (b) d, and (c) RMSE, with agricultural and forest areas overlaid

dominance depending on the season. Lagged VHI played a major role in April and May, contributing 48–57%, while SPI became more dominant in June and July (59–67%), indicating its increasing importance under early rainy season conditions. Conversely, SSI showed a stronger influence from January to March (40–57%), reflecting its role in representing accumulated hydrological effects during the early dry season.

For the 3-month forecast, the importance of predictor variables shifted markedly. VHI contribution peaked in April–June (24.8–45.6%), then decreased to around 5–15% in other months. SSI remained relatively stable but showed higher contributions from February to April (51–72%), while SPI was more influential from January to July (59–75%). As forecast lead time increased, the most notable changes were found in SPI (for months 1, 6, and 7) and in SSI (for months 2–4), reflecting each index's sensitivity to lag effects in the forecasting process.

These results demonstrate that predictor variable contribution depends strongly on both forecast time and seasonal characteristics. VHI is the most important variable for short-term forecasting, especially at the end of the dry season; SPI and SSI play more prominent roles when forecast time extends and depend on meteorological-hydrological conditions of each month.

DISCUSSION

The analysis results of relationships between drought indices and VHI showed that these relationships depend strongly on the calendar months of the year. During January–April (mid and late

dry season), VHI was primarily affected by precipitation and streamflow from previous months, with precipitation maintained at low levels and streamflow reaching minimum during February–April. From May–July (early rainy season), the impacts of cumulative rainfall and runoff from earlier periods became more pronounced. Notably, during the transition from the late dry to early rainy season, VHI exhibited joint sensitivity to both SPI and SSI, with the influence of SSI being more dominant during February–March. This finding aligns with those reported by (Sam et al., 2019) in the same basin, emphasizing the role of hydrological memory in vegetation response. Autocorrelation analysis showed that VHI in March–April had significant influence on vegetation status during April–July, reflecting accumulated drought impacts after prolonged water deficit periods. This result is consistent with recent studies in the Central Highlands, Vietnam by (Van Viet & Thuy, 2023).

Regarding predictor variable selection, the study applied a combination of correlation analysis, autocorrelation, and lag time analysis to remove variables with minimal information content or causing noise, retaining only variables with high and seasonally stable correlations. Prominent variables included lagged VHI, SPI4–SPI5, and SSI1–SSI3. This procedure helps optimize model performance and aligns with recommendations in research by (Hao et al., 2018) and (Prodhan et al., 2022), who emphasized the importance of parsimonious variable selection in drought forecasting applications.

Regarding model effectiveness, results showed that ANN and SQR exhibited different strengths in early VHI forecasting in the Srepok

Table 4. Relative contribution (%) of VHI, SPI, and SSI to VHI forecasting at different forecast times

Forecast lead time	Variable	Jan	Feb	Mar	Apr	May	Jun	Jul
1 month	VHI	59.6	60	74	88.1	63.2	54.2	56
	SSI	20.4	19	17.4	8.4	10.6	0	0
	SPI	20	21	8.6	3.5	26.2	45.8	44
2 months	VHI	23.1	23.6	45.3	57	48	40.7	22.8
	SSI	40.3	57	54.7	32.1	35.3	0	10.7
	SPI	36.6	19.4	0	10.9	16.7	59.3	66.5
3 months	VHI	15	5.3	16.3	45.6	30.4	24.8	15.3
	SSI	26	59.3	73.7	51.4	34.9	0	17.6
	SPI	59	35.4	10	3	34.7	75.2	67.1

Basin - a region strongly affected by climate change and drought. The ANN model excelled in 1–2 month lead time forecasts, particularly during the late dry season (January–April) and the onset of the rainy season (May), when VHI exhibited strong fluctuations due to water stress associated with meteorological and hydrological deficits. ANN achieved higher R and d values than SQR and RF in most months, reflecting its capability to capture complex nonlinear relationships between drought indices and vegetation status (Luong and Bui, 2024; Mokhtari and Akhoondzadeh, 2021). Conversely, SQR showed advantages in longer 3-month lead time forecasting and during the rainy season (June–July), when relationships between predictor variables and VHI were more linear and temporal persistence was weaker. Based solely on RMSE, the SQR model had advantages for all three forecast lead times. RF generally performed worse than the other two models, especially during rainy season months and at longer lead times, possibly due to insufficient training data for each month-specific model or suboptimal interaction with the data structure in this application.

Regarding predictor variable contribution, for 1-month lead time forecasting, lagged VHI was the primary factor (especially during March–April, with contribution of 74–88%). When the lead time extended to 2–3 months, the contribution of lagged VHI gradually decreased and was replaced by SPI and SSI, reflecting that in short-term forecasting, current vegetation health status mainly determines the next month, while in longer-term forecasting, factors related to water supply conditions become primary drivers. This shift in predictor importance with increasing lead time is

consistent with the cascading nature of drought propagation from meteorological to hydrological to agricultural drought (Sam et al., 2019; Tram et al., 2021).

Research results have important practical significance by providing a scientific foundation for optimizing model selection by month and forecast horizon, thereby enhancing agricultural drought early warning system effectiveness. Simultaneously, identifying dominant predictor variables at each stage allows streamlining input datasets, reducing data collection and processing costs while maintaining model accuracy. Moreover, the proposed multi-model comparison method can be applied to other basins with similar conditions to develop VHI forecasting systems appropriate for local conditions, contributing to climate variability and change adaptation.

Despite the promising results, several limitations should be acknowledged. First, the study focused on basin-wide aggregated forecasts without stratifying by land cover type, which may obscure vegetation-specific drought responses. Future work should examine whether separate models for different crop types (e.g., coffee, rice, rubber) could improve forecast accuracy. Second, the analysis was limited to linear and moderately nonlinear machine learning approaches; more advanced deep learning architectures such as long short-term memory (LSTM) networks or attention-based models may capture temporal dependencies more effectively. Third, the study period (2000–2022) may not encompass the full range of drought variability under future climate change scenarios, suggesting the need for periodic model recalibration as new data become available.

CONCLUSIONS

This study successfully selected influential variables, developed and evaluated machine learning models for early VHI forecasting in the Srepok River Basin, thereby making important contributions to drought monitoring and early warning in the context of climate change. Correlation analysis identified lagged VHI values along with SPI4-SPI5 and SSI1-SSI3 as the most effective predictor variables, reflecting cumulative impacts of meteorological and hydrological drought on vegetation health. The optimal forecast lead time range of 1–3 months, especially during January–July, demonstrates high application potential for agricultural and water resource management, consistent with crop growth cycles in the region.

Among the methods evaluated, ANN demonstrated superior performance for 1–2 month lead time forecasts during the dry season, when non-linear relationships between drought indices and VHI were clearly expressed. Conversely, SQR was more suitable for 3-month lead time forecasts and for rainy season forecasting, while RF yielded the lowest performance across all conditions. Seasonal variability was also clearly recorded, with highest forecast quality during February–May and decreasing forecast skill during June–July, reflecting characteristics of the region’s tropical monsoon climate with distinct rainy seasons. The combination of low RMSE from SQR and high d and R from ANN suggests potential for ensemble approaches that combine them in future VHI simulation systems.

Predictor variable contribution analysis showed that lagged VHI played a dominant role in 1-month lead time forecasts, especially during the late dry season; when lead time extended to 2–3 months, SPI and SSI gradually replaced this dominance, reflecting the shift from direct vegetation status influence to indirect water supply condition impacts.

Regarding scientific and application significance, the study provides an integrated framework of remote sensing and machine learning, enabling input variable optimization and model selection according to temporal-seasonal conditions. However, the study also acknowledges certain limitations, particularly accuracy decline during the rainy season when VHI has low variation and drought manifestation is weak. Therefore, future research needs to expand predictor variable scope while considering climate change

impacts on drought-vegetation relationships, aiming to maintain and enhance forecast model reliability in the long term.

REFERENCES

- Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5–32.
- Chu, H., Venevsky, S., Wu, C., Wang, M. (2019). NDVI-based vegetation dynamics and its response to climate changes at Amur-Heilongjiang River Basin from 1982 to 2015. *Science of the Total Environment*, 650, 2051–2062.
- Commission, M. R. (2020). Mekong River Commission Annual Report 2019. *Mekong River Commission: Lao PDR, Vientiane*.
- Feng, P., Wang, B., Li Liu, D., Yu, Q. (2019). Machine learning-based integration of remotely-sensed drought factors can improve the estimation of agricultural drought in South-Eastern Australia. *Agricultural Systems*, 173, 303–316.
- Hao, Z., Singh, V. P., Xia, Y. (2018). Seasonal drought prediction: Advances, challenges, and future prospects. *Reviews of Geophysics*, 56(1), 108–141.
- Hiep, N. V., Thao, N. T. T., Viet, L. V., Luc, H. C., Ba, L. H. (2023). Affecting of nature and human activities on the trend of vegetation health indices in dak nong province, Vietnam. *Sustainability*, 15(7), 5695.
- Holben, B. N. (1986). Characteristics of maximum-value composite images from temporal AVHRR data. *International Journal of Remote Sensing*, 7(11), 1417–1434.
- Kafy, A.-A., Bakshi, A., Saha, M., Al Faisal, A., Almulhim, A. I., Rahaman, Z. A., Mohammad, P. (2023). Assessment and prediction of index based agricultural drought vulnerability using machine learning algorithms. *Science of the Total Environment*, 867, 161394.
- Khoi, D. N., Nhi, P. T. T. (2021). Assessment of climate change impact on drought in the Central Highlands of Vietnam. In *Water Security in Asia: Opportunities and Challenges in the Context of Climate Change* 703–713. Springer.
- Kogan, F. (2002). World droughts in the new millennium from AVHRR-based vegetation health indices. *Eos, Transactions American Geophysical Union*, 83(48), 557–563.
- Kogan, F. N. (1997). Global drought watch from space. *Bulletin of the American Meteorological Society*, 78(4), 621–636.
- Li, H., Li, Y., Gao, Y., Zou, C., Yan, S., Gao, J. (2016). Human impact on vegetation dynamics

- around Lhasa, southern Tibetan Plateau, China. *Sustainability*, 8(11), 1146.
13. Liu, Q., Zhang, S., Zhang, H., Bai, Y., Zhang, J. (2020). Monitoring drought using composite drought indices based on remote sensing. *Sci Total Environ*, 711, 134585. <https://doi.org/10.1016/j.scitotenv.2019.134585>
14. Luong, V. V., Bui, D. H. (2023). Determination of the most suitable indicator area and remote-sensing-based indices for early yield warning for winter–spring rice in the Central Highlands, Vietnam. *Journal of Applied Remote Sensing*, 17(1), 014504. <https://doi.org/10.1117/1.JRS.17.014504>
15. Luong, V. V., Bui, D. H. (2024). Evaluation of models and drought-wetness factors contributing to predicting the vegetation health index in Dak Nong Province, Vietnam. *Environmental Research Communications*, 6(4), 045005.
16. Mas, J. F., Flores, J. J. (2008). The application of artificial neural networks to the analysis of remotely sensed data. *International Journal of Remote Sensing*, 29(3), 617–663.
17. McKee, T. B., Doesken, N. J., Kleist, J. (1993). The relationship of drought frequency and duration to time scales. *Proceedings of the 8th Conference on Applied Climatology*, 17(22), 179–183.
18. Mehraein, M., Mohanavelu, A., Naganna, S. R., Kulls, C., Kisi, O. (2022). Monthly streamflow prediction by metaheuristic regression approaches considering satellite precipitation data. *Water*, 14(22), 3636.
19. Modarres, R. (2007). Streamflow drought time series forecasting. *Stochastic Environmental Research and Risk Assessment*, 21, 223–233.
20. Mokhtari, R., Akhoondzadeh, M. (2021). Data fusion and machine learning algorithms for drought forecasting using satellite data. *Journal of the Earth and Space Physics*, 46(4), 231–246.
21. Narmilan, A., Gonzalez, F., Salgadoe, A. S. A., Kumarasiri, U. W. L. M., Weerasinghe, H. A. S., Kulasekara, B. R. (2022). Predicting canopy chlorophyll content in sugarcane crops using machine learning algorithms and spectral vegetation indices derived from UAV multi-spectral imagery. *Remote Sensing*, 14(5), 1140.
22. Prajapati, V., Khanna, M., Singh, M., Kaur, R., Sahoo, R., Singh, D. (2021). Evaluation of time scale of meteorological, hydrological and agricultural drought indices. *Natural hazards*, 109(1), 89–109.
23. Prodhan, F. A., Zhang, J., Hasan, S. S., Sharma, T. P. P., Mohana, H. P. (2022). A review of machine learning methods for drought hazard monitoring and forecasting: Current research trends, challenges, and future research directions. *Environmental modelling & software*, 149, 105327.
24. Rahman, G., Khalid, S., Arshad, S., Moazzam, M. F. U., Kwon, H.-H. (2025). Remote sensing-based spatiotemporal assessment of agricultural drought and its impact on crop yields in Punjab, Pakistan. *Scientific Reports*, 15(1), 20586.
25. Sam, T. T., Khoi, D. N., Thao, N. T. T., Nhi, P. T. T., Quan, N. T., Hoan, N. X., Nguyen, V. T. (2019). Impact of climate change on meteorological, hydrological and agricultural droughts in the Lower Mekong River Basin: a case study of the Srepok Basin, Vietnam. *Water and Environment Journal*, 33(4), 547–559.
26. Telesca, L., Lovallo, M., Lopez-Moreno, I., Vicente-Serrano, S. (2012). Investigation of scaling properties in monthly streamflow and Standardized Streamflow Index (SSI) time series in the Ebro basin (Spain). *Physica A: Statistical Mechanics and its Applications*, 391(4), 1662–1678.
27. Tram, V. N. Q., Somura, H., Moroizumi, T. (2021). Evaluation of drought features in the Dakbla watershed, Central Highlands of Vietnam. *Hydrological Research Letters*, 15(3), 77–83.
28. Usman, U., Yelwa, S., Gulumbe, S., Danbaba, A., Nir, R. (2013). Modelling relationship between NDVI and climatic variables using geographically weighted regression. *Journal of Mathematical Sciences and Applications*, 1(2), 24–28.
29. Van Viet, L., Thuy, T. T. T. (2023). Improving the quality of coffee yield forecasting in Dak Lak Province, Vietnam, through the utilization of remote sensing data. *Environmental Research Communications*, 5(9), 095011.
30. Wan, Z., Wang, P., Li, X. (2004). Using MODIS land surface temperature and normalized difference vegetation index products for monitoring drought in the southern Great Plains, USA. *International Journal of Remote Sensing*, 25(1), 61–72.