

Article

Not peer-reviewed version

Forecasting Vegetation Dynamics in a Semi-Arid Region Using Deep Learning and Sentinel-2 EVI Under CMIP6 Climate Scenarios

[Azad Rasul](#) *

Posted Date: 10 March 2026

doi: 10.20944/preprints202603.0611.v1

Keywords: Enhanced Vegetation Index (EVI); deep learning; Sentinel-2; vegetation forecasting; CMIP6 climate scenarios; Kurdistan Region of Iraq



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Forecasting Vegetation Dynamics in a Semi-Arid Region Using Deep Learning and Sentinel-2 EVI Under CMIP6 Climate Scenarios

Azad Rasul ^{1,2}

¹ Department of Geography, Soran University, Soran, Erbil, Iraq; azad.rasul@soran.edu.iq

² Department of Forestry, College of Agricultural Engineering Sciences, Salahaddin University-Erbil, 44002, Iraq

Abstract

Accurate monitoring and forecasting of vegetation dynamics are essential for sustainable ecosystem management, agricultural planning, and climate adaptation in semi-arid environments. Satellite-derived vegetation indices provide a powerful tool for assessing vegetation health and productivity over large spatial scales. Among these indices, the Enhanced Vegetation Index (EVI) offers improved sensitivity in high-biomass regions and reduced atmospheric and soil background effects compared to traditional indices. This study develops a deep learning framework to forecast vegetation dynamics across the Kurdistan Region of Iraq (KRI), covering the governorates of Erbil, Duhok, Sulaymaniyah, and Halabja. Monthly time-series data (2016–2024) were derived from Sentinel-2 imagery and combined with climatic variables including precipitation and temperature. Twelve deep learning architectures, including recurrent neural networks, convolutional–recurrent hybrids, temporal convolutional networks, and attention-based models, were evaluated using a multivariate feature set incorporating EVI, climate variables, and seasonal encoding. Model performance was assessed using multiple statistical metrics, including R^2 , RMSE, MAE, and Nash–Sutcliffe efficiency. The best-performing architecture was then used to generate vegetation projections under climate change scenarios derived from CMIP6 forcing using the IPCC AR6 delta-factor method. Future vegetation trajectories were simulated for the period 2025–2050 under baseline, SSP2-4.5, and SSP5-8.5 scenarios, incorporating Monte Carlo dropout to quantify predictive uncertainty. The proposed framework demonstrates strong potential for forecasting vegetation dynamics in data-limited semi-arid regions and provides insights into potential vegetation responses to future climate change in Kurdistan Region of Iraq.

Keywords: Enhanced Vegetation Index (EVI); deep learning; Sentinel-2; vegetation forecasting; CMIP6 climate scenarios; Kurdistan Region of Iraq

1. Introduction

Vegetation plays a central role in regulating ecosystem processes, including carbon cycling, climate regulation, biodiversity conservation, and agricultural productivity. Monitoring vegetation dynamics is therefore essential for understanding ecosystem health and supporting sustainable land management (Running et al., 2004). The diversity and distribution of plants shape ecosystem function and influence global biogeochemical cycles, and integrating plant functional traits into global vegetation and Earth system models is crucial for refining projections of energy, carbon, and water cycles (Lusk et al., 2026). An ecosystem's carbon sequestration potential is tightly linked to its biological diversity, and biodiversity declines driven by climate and land-use change could result in substantial global losses of vegetation carbon storage (Liu et al., 2025).

With the rapid growth of Earth observation datasets, there has been increasing interest in using advanced computational methods to analyze and forecast vegetation dynamics. Traditional statistical

models, including autoregressive and regression-based approaches, often struggle to capture the nonlinear relationships between vegetation indices and climatic drivers such as precipitation and temperature (Ronald et al., 2009). Recent advances in artificial intelligence and machine learning have enabled more accurate modeling of complex environmental processes by learning nonlinear temporal patterns directly from large datasets (Reichstein et al., 2019). In particular, deep learning approaches have demonstrated strong performance in modeling spatio-temporal environmental data derived from remote sensing observations (Kattenborn et al., 2021; Yuan et al., 2020).

Among deep learning techniques, recurrent neural networks (RNNs), long short-term memory (LSTM) networks, and gated recurrent units (GRU) have proven especially effective for modeling sequential environmental data because they can capture temporal dependencies within time series (Hochreiter & Schmidhuber, 1997). Several studies have applied these architectures to forecast vegetation indices and ecosystem responses using satellite imagery and climatic variables (R. Shen et al., 2019). For example, deep learning models based on convolutional and recurrent architectures have shown improved accuracy in predicting vegetation responses to drought conditions using Earth observation data (Ndikumana et al., 2018).

More recently, hybrid deep learning models that integrate convolutional neural networks (CNNs), recurrent networks, and attention mechanisms have been increasingly used for vegetation monitoring and prediction (Kattenborn et al., 2021; Yuan et al., 2020). These architectures can simultaneously extract spatial features from remote sensing imagery and capture temporal dynamics of vegetation growth cycles (Zhong et al., 2019). The incorporation of attention mechanisms, originally introduced in the Transformer architecture, has further improved the ability of models to focus on the most informative temporal and spatial features within complex remote sensing time series (Vaswani et al., 2017). Such hybrid frameworks have demonstrated superior performance over single-architecture models in estimating vegetation parameters and predicting land surface dynamics from multispectral and multi-temporal satellite data (Kattenborn et al., 2021).

Despite these advances, relatively few studies have investigated deep learning-based vegetation forecasting in semi-arid regions of the Middle East. The Kurdistan Region of Iraq (KRI) represents a particularly important case study because its ecosystems are highly sensitive to climatic variability, including seasonal droughts and increasing temperatures associated with climate change. Vegetation in this region exhibits strong seasonal patterns driven primarily by winter precipitation and spring growth cycles. Understanding how vegetation responds to climatic variability and future climate scenarios is therefore critical for supporting agricultural management, drought preparedness, and environmental planning.

This study aims to address this research gap by developing a deep learning framework to forecast vegetation dynamics in the Kurdistan Region of Iraq using satellite-derived EVI and climate data. Specifically, the objectives of this research are to:

- (1) evaluate multiple deep learning architectures for forecasting monthly EVI time series using remote sensing and climate variables;
- (2) identify the most robust model for vegetation prediction across the four governorates of the Kurdistan Region; and
- (3) generate future vegetation projections under different climate change scenarios using CMIP6 forcing and probabilistic uncertainty estimation.

By integrating Earth observation data processed via Google Earth Engine, CMIP6 climate projections, and advanced deep learning techniques, this study provides new insights into future vegetation dynamics in semi-arid landscapes and demonstrates the utility of AI-driven forecasting approaches for regional environmental monitoring.

2. Data and Methodology

2.1. Study Area

The study area encompasses the Kurdistan Region of Iraq (KRI), comprising four governorates: **Erbil**, **Duhok**, **Sulaymaniyah**, and **Halabja** (Figure 1). The region spans approximately 34.38°–37.38°N, 42.26°–46.34°E, covering diverse elevation gradients from the Mesopotamian lowlands (~300 m a.s.l.) in Erbil to the Zagros mountain foothills exceeding 3,000 m a.s.l. The region is characterized by a continental Mediterranean climate, marked by cold and wet winters contrasting with hot and dry summers. Mean annual precipitation, based on records spanning from 2012 to 2024, amounts to 548.7 mm, with over 90% of this total concentrated within the November–April period. Mean annual temperatures are approximately 20.8 °C in Erbil and around 17 °C at higher elevations. During summer, temperatures regularly exceed 44 °C, whereas winter conditions can occasionally produce temperatures falling slightly below -2 °C (Rasul & Zahir, 2025). These physiographic contrasts make KRI an ideal natural laboratory for assessing differential vegetation responses to climate forcing.

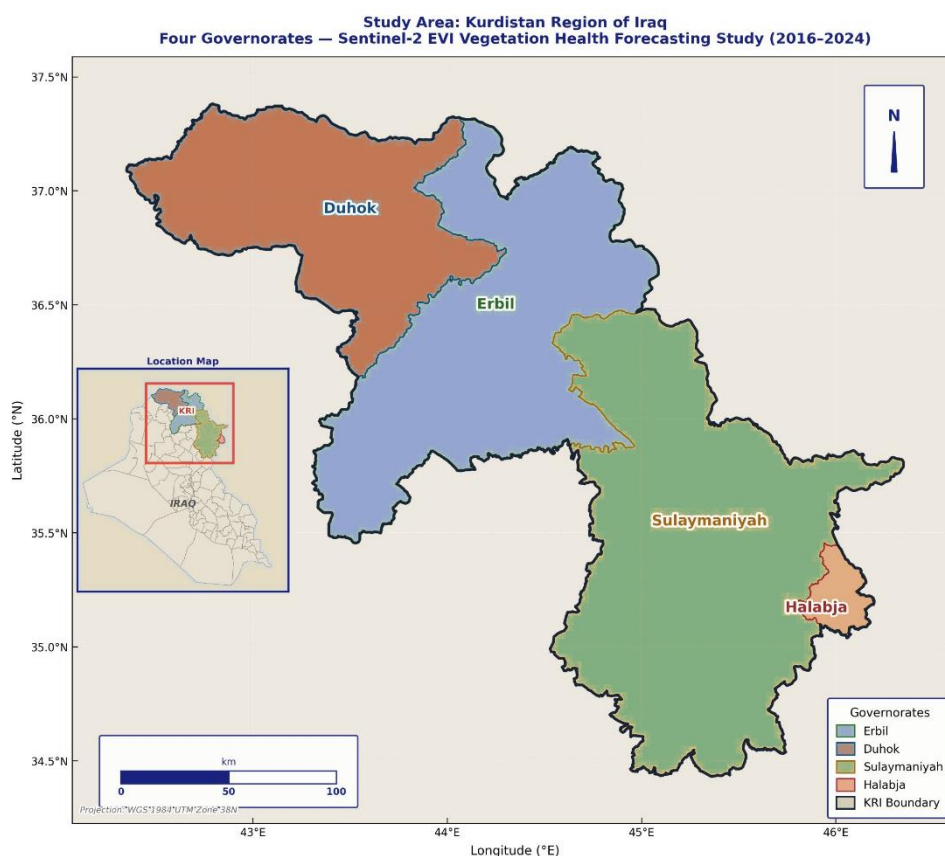


Figure 1. Study area map of the Kurdistan Region of Iraq (KRI) showing the four governorates: Erbil, Duhok, Sulaymaniyah, and Halabja. The inset map shows the location of KRI within Iraq. Governorate boundaries derived from official administrative shapefiles; projection WGS 1984 UTM Zone 38N.

2.2. Remote Sensing and Climate Data

Monthly time series of three variables were obtained for each governorate covering January 2016 to December 2024 (108 months) from a validated dataset produced by the companion paper (Rasul, 2026). Data were loaded from the file *evi_climate_kurdistan_2016_2024.csv*, which contains monthly records per governorate of:

Enhanced Vegetation Index (EVI): Derived from Sentinel-2 Surface Reflectance imagery (*COPERNICUS/S2_SR_HARMONIZED*), Google Earth Engine (GEE) collection, using Scene Classification Layer (SCL) masking to exclude cloud, shadow, and snow pixels, then computing

monthly median composites. EVI was preferred over NDVI for its improved sensitivity in high-biomass conditions and reduced soil and atmospheric noise (Huete et al., 2002).

Monthly Precipitation (mm): Obtained from the TerraClimate dataset (*IDAHO_EPSCOR/TERRACLIMATE*), a global monthly climate dataset at ~4 km resolution derived from interpolated station data and reanalysis (Abatzoglou et al., 2018).

Mean Air Temperature (°C): Also obtained from TerraClimate using the same spatial and temporal resolution.

To support reproducibility and transparency, the processed monthly EVI and climate dataset used in this study (*evi_climate_kurdistan_2016_2024.csv*) has been deposited in a publicly accessible repository [INSERT DOI – e.g., Zenodo, Figshare, or OSF]. The dataset contains monthly records per governorate of EVI, precipitation, and temperature covering January 2016 to December 2024, and is sufficient to reproduce all model training, evaluation, and projection results reported in this paper independently of the companion preprint. The companion paper (Rasul, 2026) provides full methodological detail on the Google Earth Engine extraction pipeline, SCL-based cloud masking, valid-pixel thresholds, and climatological gap-filling procedures. The key processing decisions relevant to the present study are summarised in Section 2.2 and Section 2.2.1 above, which are sufficient to assess data quality without requiring access to the companion manuscript. Should the companion paper undergo methodological revision during peer review, the deposited dataset version used in this study will remain fixed, ensuring that the present results are reproducible against a stable data version.

2.2.1. Data Quality Control and Gap-Filling

Data quality control mirrored the companion paper pipeline. EVI observations were flagged and set to missing where fewer than 50 valid pixels were available in the GEE composite (*valid_pixels* threshold). EVI values were clipped to the physically valid range $[-0.1, 0.8]$. Missing EVI values were filled using the long-term monthly climatological mean for that governorate (i.e., the mean EVI for the same calendar month across all available years), followed by linear interpolation for any remaining gaps. Precipitation and temperature gaps were filled by linear interpolation with forward/backward fill for boundary periods. No missing values remained in the final dataset after this procedure.

2.3. Temporal Data Partitioning

A strict non-overlapping temporal split was applied to prevent data leakage. All normalisation scalers were fitted exclusively on training data before being applied to validation and test sets:

Training set: February 2016 – December 2021 (approximately 48 input sequences after 24-month windowing; see Section 2.5). This period spans normal precipitation years and wet anomalies.

Validation set: January 2022 – December 2022 (12 months; the 2022 drought year). This represents the most challenging generalisation period in the record, with EVI anomalies of -12.7% to -24.8% below the observed mean, providing a rigorous test of model robustness under extreme conditions. EarlyStopping used *val_loss* to select the best checkpoint.

Test set: January 2023 – December 2024 (24 sequences). Fully held-out; never used during training or model selection. All test-set metrics reported in this paper are computed on this period.

2.4. Feature Engineering and Normalisation

Five input features were constructed per timestep for each governorate:

(1) **EVI** – Z-score normalised (StandardScaler fitted on training data). (2) **Precipitation (mm)** – Z-score normalised (training scaler). (3) **Temperature (°C)** – Z-score normalised (training scaler). (4) **Month sine encoding** – $x_{\sin} = \sin(2\pi \cdot m / 12)$. (1)

(5) **Month cosine encoding** – $x_{\cos} = \cos(2\pi \cdot m / 12)$. (2)

Cyclic month encoding (features 4–5) replaces a raw integer month index, allowing the network to correctly represent calendar continuity (December–January transitions) without artificial discontinuities. This is standard practice for seasonal ecological time series (H. Shen et al., 2015). Per-governorate scalars ensure that normalisation reflects local EVI amplitude differences rather than pooling across governorates.

2.5. Sequence Construction

Input sequences were constructed using a **24-month sliding window** (SEQ_LEN = 24), matching the companion paper. For each target timestep t , the input tensor X is the 5-feature matrix over months $[t-23, \dots, t]$, and the target y is the normalised EVI at month $t+1$. This produces tensors of shape $(N, 24, 5)$, where $N \approx 48$ training sequences per governorate. Sequences crossing the train-validation or validation–test boundary were excluded to maintain strict temporal separation.

2.6. Deep Learning Architectures

Twelve architectures were evaluated, grouped into four categories (Table 1). All models share the same input shape (24×5) , output a single scalar (next-month normalised EVI), and are trained with the Adam optimiser and mean squared error (MSE) loss. Batch size was 16 for all experiments.

Table 1. Summary of all 12 deep learning architectures evaluated. ★ = selected projection model. Units = number of LSTM/GRU/CNN filters. LR = initial Adam learning rate. GAP = GlobalAveragePooling1D. BN = BatchNormalization.

Model	Category	Architecture	Units	LR
BiLSTM	Recurrent	3-layer stacked BiLSTM with dropout	64	0.0005
GRU	Recurrent	2-layer stacked GRU with dropout	64	0.0005
BiLSTM_GRU ★	Recurrent	BiLSTM (return_sequences) → GRU → Dense	64	0.0005
CNN_GRU	CNN- Recurrent	Conv1D + BN → GRU → Dropout → Dense	64	0.001
CNN_BiLSTM	CNN- Recurrent	Conv1D + BN → BiLSTM → Dropout → Dense	64	0.0005
CNN_BiLSTM_GRU	CNN- Recurrent	Conv1D + BN → BiLSTM (return_seq) → GRU → Dense	64	0.0005

Model	Category	Architecture	Units	LR
TCN	Advanced	Dilated causal Conv1D, dilation=[1,2,4,8], residuals	64	0.001
Transformer	Advanced	Sinusoidal pos. enc. + MultiHeadAttention (4 heads) + FF	64	0.0005
CNN_Transformer	Advanced	Conv1D + BN → MultiHeadAttention (4 heads) → GAP	64	0.0005
CNN_BiLSTM_AM	Advanced	Conv1D + BN → BiLSTM → Bahdanau attention → Dense	64	0.0005
LSTM	Baseline	2-layer stacked LSTM with dropout	64	0.0005
CNN	Baseline	2× Conv1D → GAP → Dropout → Dense	64	0.001

2.6.1. Recurrent Architectures

BiLSTM: Three-layer stacked Bidirectional LSTM (units = 64, 32, 16) with Dropout(0.3) between layers. Bidirectional processing allows the network to capture both forward and backward temporal dependencies within the 24-month window.

GRU: Two-layer stacked Gated Recurrent Unit (units = 64) with Dropout(0.3) between layers. Fewer parameters than LSTM, suitable for shorter sequences.

BiLSTM_GRU (selected projection model): A hybrid recurrent architecture combining a BiLSTM layer (units = 64, return_sequences = True) for bidirectional feature extraction, followed by a GRU layer (units = 64) for temporal summarisation, with Dropout(0.3) between layers and a final Dense(1) output. This architecture captures long-range bidirectional context (BiLSTM) while maintaining computational efficiency in the summarisation step (GRU).

2.6.2. CNN-Recurrent Architectures

Three hybrid architectures combine a Conv1D layer (filters = 64, kernel_size = 3, padding = 'same', activation = 'relu') followed by BatchNormalization for local temporal feature extraction, with recurrent layers for sequential modelling: **CNN_GRU** (Conv1D → GRU), **CNN_BiLSTM** (Conv1D → BiLSTM), and **CNN_BiLSTM_GRU** (Conv1D → BiLSTM with return_sequences → GRU). All use Dropout(0.3) before the Dense(1) output.

2.6.3. Advanced Architectures

TCN (Temporal Convolutional Network): Implemented following Bai et al. (2018) with dilated causal Conv1D layers at dilation rates [1,2,4,8], each followed by LayerNormalization, ReLU activation, Dropout(0.2), a second dilated Conv1D, and residual connections. The input is projected to the filter dimension (Conv1D kernel_size = 1) to enable residual addition. GlobalAveragePooling1D collapses the temporal dimension before the Dense(1) output.

Transformer: Multi-head self-attention (4 heads, key_dim = 16) with sinusoidal positional encoding (Vaswani et al., 2017). Positions are encoded as

$$PE(\text{pos}, 2i) = \sin(\text{pos} / 10000^{2i / d_{\text{model}}}) \quad (3)$$

$$PE(\text{pos}, 2i+1) = \cos(\text{pos} / 10000^{2i / d_{\text{model}}}) \quad (4)$$

pos = position in the sequence; *i* = dimension index; d_{model} = model dimension (64). After Vaswani et al. (2017).

The architecture includes a feed-forward sub-layer (Dense(128) → Dense(64)) with LayerNormalization and residual connections at each block. GlobalAveragePooling1D followed by Dense(1) produces the scalar output. Note: positional encoding is required to distinguish temporal order; omitting it makes the Transformer permutation-invariant and unable to model seasonality.

CNN_Transformer: Conv1D (filters = 64, kernel_size = 3) + BatchNormalization for local feature extraction, followed by a Dense projection to $d_{\text{model}} = 64$, then MultiHeadAttention (4 heads), LayerNormalization, Dropout(0.2), GlobalAveragePooling1D, and Dense(1).

CNN_BiLSTM_AM: Implements the architecture of Wang et al. (2025) with Conv1D + BatchNormalization, followed by BiLSTM (units = 64, return_sequences = True) to produce a full hidden state sequence. Bahdanau-style temporal attention computes:

Step 1: Attention score:

$$e_t = \tanh(W h_t + b) \quad (5)$$

Step 2: Attention weight:

$$\alpha_t = \text{softmax}(v^T e_t) = \exp(v^T e_t) / \sum_{\{T'=1\}} \exp(v^T e_{\{T'\}}) \quad (6)$$

Step 3: Context vector:

$$c = \sum_{T'=1} \alpha_t \cdot h_t \quad (7)$$

$W \in \mathbb{R}^{d \times d}$ and $v \in \mathbb{R}^d$ are trainable parameters; h_t is the BiLSTM hidden state at time t ; $T = 24$ (sequence length). The context vector c replaces the final hidden state as input to the output Dense(1) layer.

The context vector replaces the final hidden state as input to the output Dense(1). This allows the model to focus on the most vegetation-relevant months (e.g., spring green-up, drought onset) within the 24-month window.

2.6.4. Baseline Architectures

LSTM: Two-layer stacked LSTM (units = 64, 32) with Dropout(0.3), included as a classical recurrent baseline.

CNN: Two Conv1D layers (filters = 64, 32; kernel_size = 3; activation = 'relu'; padding = 'same') followed by GlobalAveragePooling1D and Dropout(0.3), representing a pure convolutional baseline without sequential memory.

2.7. Model Training

Each model was trained for a maximum of 150 epochs (EPOCHS = 150) with batch size 16 using two callbacks: (1) **EarlyStopping** with patience = 20, monitoring validation loss (val_loss), restoring best weights; and (2) **ReduceLROnPlateau** with factor = 0.5, patience = 10, and minimum learning rate = 1×10^{-6} . These prevent overfitting on the limited training data (~48 sequences) while allowing sufficient training time for complex architectures. The total experiment count was 12 models \times 4 governorates = 48 training runs. All experiments used a fixed random seed (numpy seed = 42, TensorFlow seed = 42) for reproducibility. Training was performed on Google Colab with GPU acceleration (TensorFlow 2.19).

2.8. Evaluation Metrics

Model performance was assessed using five metrics computed on inverse-transformed (original EVI units) predictions:

R² (coefficient of determination): Proportion of variance explained; computed via `sklearn.metrics.r2_score`.

RMSE (Root Mean Square Error): $\sqrt{(\text{MSE})}$; in EVI units.

MAE (Mean Absolute Error): Mean absolute deviation in EVI units.

NSE (Nash-Sutcliffe Efficiency):

where NSE is defined as:

$$\text{NSE} = 1 - \frac{\sum_{i=1}^n (y_i^{\text{obs}} - y_i^{\text{pred}})^2}{\sum_{i=1}^n (y_i^{\text{obs}} - \bar{y}^{\text{obs}})^2} \quad (8)$$

where y_i^{obs} is the observed value, y_i^{pred} is the predicted value, and \bar{y}^{obs} is the mean of observed values.

NSE > 0.5 is considered satisfactory (Moriassi et al., 2007).

Tolerance Accuracy: Proportion of months where

$$|\hat{y} - y| \leq 0.05 \quad \text{EVI units.} \quad (9)$$

This threshold reflects the practical precision of Sentinel-2 EVI retrieval.

2.9. Walk-Forward Cross-Validation

To assess model stability for projection — distinct from point accuracy on a single test window — a **3-fold walk-forward time-series cross-validation** was performed on all 12 architectures using `sklearn's TimeSeriesSplit`. Cross-validation was **strictly restricted to training data only** (dates \leq December 2021), preventing any leakage from the 2022 validation or 2023–2024 test periods. Each fold trained for a maximum of 50 epochs with patience = 8. Stability was defined as mean CV R² > 0.50 and minimum governorate CV R² > 0.20, requiring the model to generalise across all four governorates rather than averaging over strong and weak sites. The **projection model was selected as the architecture with the highest mean CV R² among CV-stable models**, prioritising temporal generalisation over single-window test performance. In total, 144 CV fits were performed (12 models \times 4 governorates \times 3 folds).

2.10. CMIP6 Climate Forcing Construction

Future climate inputs for the 2025–2050 projection period (312 months) were constructed using the **IPCC AR6 delta-factor method** (Masson-Delmotte et al., 2021), applied to the observed monthly climatology. Delta factors represent the CMIP6 multi-model median anomaly for the MENA domain from IPCC AR6 WGI Chapter 10 (Doblas-Reyes et al., 2021) (Table 2).

Table 2. CMIP6 delta factors applied to construct future climate forcing, based on IPCC AR6 WGI Chapter 10 MENA domain multi-model median. Temperature deltas are additive (°C above baseline). Precipitation deltas are multiplicative fractions applied to the historical monthly climatology.

Scenario	Temperature (°C)	Precipitation (%)	Application Method
Baseline	No change	No change	Historical monthly climatology, repeated
SSP2-4.5	+1.5	-8%	Linear ramp: 0 at Jan 2025 → full delta at Dec 2050
SSP5-8.5	+2.8	-15%	Linear ramp: 0 at Jan 2025 → full delta at Dec 2050

The forcing was implemented as a **linear ramp** from zero at January 2025 to the full delta value at December 2050 (`numpy.linspace(0.0, 1.0, 312)`). For each future month t and governorate g , future precipitation was:

$$P_{\text{future}}(t, g) = \max\{ P_{\text{clim}}(m, g) \times [1 + \Delta P \cdot r(t)], 0 \} \quad (10)$$

$P_{\text{clim}}(m, g)$ = historical monthly climatological precipitation for month m and governorate g ; ΔP = scenario precipitation delta fraction; $r(t) = \text{linspace}(0, 1, 312)$ = linear ramp factor at step t . The $\max(\cdot, 0)$ clamp prevents physically impossible negative precipitation.

and future temperature was:

$$T_{\text{future}}(t, g) = T_{\text{clim}}(m, g) + \Delta T \cdot r(t) \quad (11)$$

$T_{\text{clim}}(m, g)$ = historical monthly mean temperature (°C) for month m and governorate g ; ΔT = scenario temperature delta (°C); $r(t) = \text{linspace}(0, 1, 312)$ over 2025–2050 (312 steps).

A baseline scenario was also constructed using the historical climatology with no trend, serving as a counterfactual reference. Precipitation was clamped to zero to prevent physically impossible negative values.

2.11. Autoregressive Projection with MC-Dropout Uncertainty

Projections were generated using a **GPU-accelerated autoregressive MC-Dropout** algorithm. At each of the 312 projection steps, the trained BiLSTM_GRU model was evaluated in training mode (enabling dropout) for 100 stochastic forward passes (`N_MC_PASSES = 100`), producing an ensemble of EVI trajectories. Forward passes were batched via a `tf.function`-compiled `tf.map_fn` to maximise GPU utilisation. The algorithm proceeds as follows:

Initialisation: The seed window was set to the last 24 observed months (January 2023 – December 2024) of normalised features, replicated `N_MC_PASSES` times to create 100 independent chains.

Step iteration: At each step, the 100-pass ensemble prediction `y_batch` (shape: 100,) was recorded. The predicted EVI for each pass was appended to its chain's feature window by constructing a new feature vector [`y_pred`, `P_norm(t)`, `T_norm(t)`, `sin_enc(t)`, `cos_enc(t)`] and sliding the window forward by one step.

Uncertainty scaling: The raw inter-pass standard deviation was multiplied by a linear scaling factor growing from 1.0 (January 2025) to 3.0 (December 2050), reflecting that projection uncertainty compounds over time. The 95% confidence interval was taken as $\text{mean} \pm 2\sigma$.

Climate inputs (precipitation and temperature) were normalised using the training-data scalers before being inserted into each projection step, ensuring consistency with the feature normalisation used during training. Three independent projection runs were executed per governorate (baseline, SSP2-4.5, SSP5-8.5), yielding 12 projection ensembles in total (Figure 2).

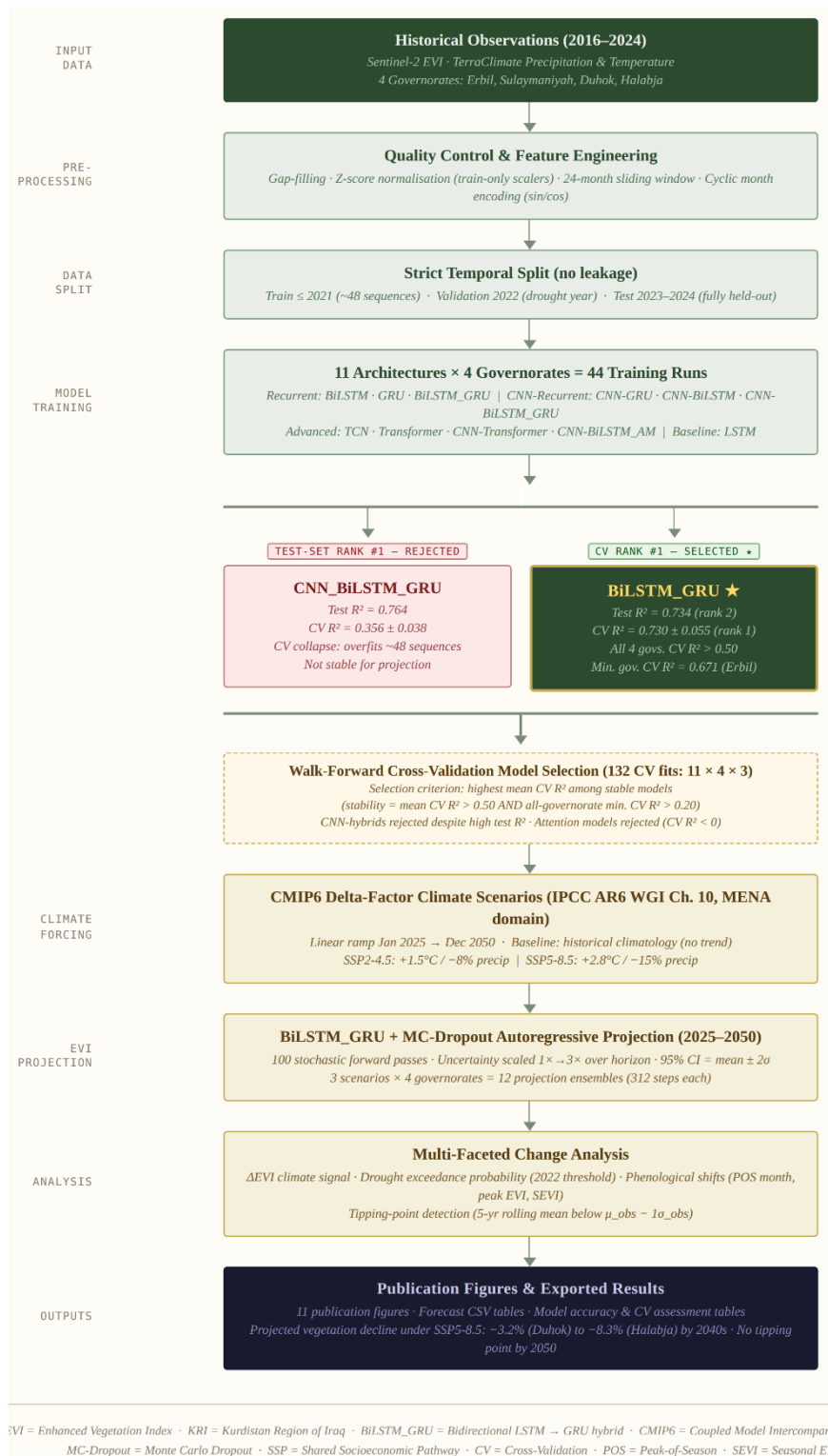


Figure 2. Overview of the study methodology. Data from Sentinel-2 and TerraClimate (2016–2024) are preprocessed and used to train 12 deep learning architectures. Model selection via walk-forward cross-validation identifies BiLSTM_GRU (CV $R^2 = 0.730$) as the projection model. CMIP6 delta-factor forcing (SSP2-4.5 and SSP5-8.5) drives autoregressive MC-Dropout projections (2025–2050), followed by ΔEVI , drought exceedance, phenological, and tipping-point analyses.

2.12. Post-Projection Analyses

2.12.1. ΔEVI Climate Signal

The net climate change signal was isolated as

$$\Delta\text{EVI}(t, g) = \text{EVI}_{\text{scenario}}(t, g) - \text{EVI}_{\text{baseline}}(t, g) \quad (12)$$

Positive ΔEVI indicates warming-driven vegetation enhancement; negative values indicate stress-driven decline. Isolates the forced climate signal from seasonality preserved in the baseline.

2.12.2. Drought Exceedance Probability

For each governorate, a drought threshold was defined as the 2022 annual mean EVI (the most severe drought in the record). For each future year and scenario, the 100-member MC-Dropout ensemble was used to compute the probability that annual mean EVI falls below this threshold:

$$P_{\text{drought}}(y, g, s) = 1 - \frac{1}{N_{\text{MC}}} \cdot \sum_{n=1}^{N_{\text{MC}}} \mathbb{1}[\text{EVI}_{\text{annual}}(y, g, n) < \tau_g] \quad (13)$$

$y = \text{year}$, $g = \text{governorate}$, $s = \text{scenario}$, $n = \text{MC-Dropout ensemble member}$, $N_{\text{MC}} = 100$, $\mathbb{1}[\cdot] = \text{indicator function}$, $\tau_g = 2022 \text{ drought-year annual mean EVI threshold for governorate } g$.

This exceedance probability is reported for each year from 2025 to 2050.

2.12.3. Phenological Metrics

Three phenological indicators were extracted annually from EVI time series for both observed and projected scenarios. The **Peak-of-Season (POS) month** was defined as the calendar month of maximum EVI. The **Peak EVI value** was the maximum annual EVI. The **Seasonal EVI Index (SEVI)** was the sum of EVI over January–June (the growing season), representing total spring–early summer vegetation productivity.

2.12.4. Tipping-Point Detection

A tipping point was defined as the first year in which the 5-year running mean of annual EVI falls more than one standard deviation below the historical mean:

$$\mathbf{R}_{5}^{-}(t) < \mu_{\text{obs}} - \sigma_{\text{obs}} \quad (14)$$

where the 5-year centred running mean is:

$$\mathbf{R}_{5}^{-}(t) = \frac{1}{5} \sum_{k=t-2}^{t+2} \text{EVI}_{\text{annual}}(k) \quad (15)$$

$\mu_{\text{obs}} = \text{mean annual EVI over 2016–2024}$; $\sigma_{\text{obs}} = \text{standard deviation of annual EVI over 2016–2024}$. A tipping point is recorded as the first year t for which this criterion is satisfied. Minimum 5 complete years required ($\text{min}_{\text{periods}} = 5$). If not reached within 2025–2050, recorded as “Not reached.”

3. Results

3.1. Observed EVI, Precipitation, and Temperature (2016–2024)

Monthly EVI time series for the four Kurdistan Region of Iraq (KRI) governorates are shown in Figure 3. All four governorates exhibited a consistent seasonal cycle with peak greenness occurring in **March–April** (Peak-of-Season month = 3–4) and summer senescence reaching minimum values in **July–August**. Mean annual EVI over the 2016–2024 observation period ranged from **0.1426 (Sulaymaniyah)** to **0.1833 (Halabja)**, reflecting the elevation gradient from the low-lying Erbil plain to the wetter, higher-elevation Halabja.

The **2022 drought** represents the most severe vegetation stress event in the record. Annual mean EVI declined by **24.8% in Sulaymaniyah** (from 0.1426 to 0.1073), **20.2% in Erbil** (0.1475 to 0.1177), **16.2% in Duhok** (0.1727 to 0.1447), and **12.7% in Halabja** (0.1833 to 0.1601). These drought-year values serve as the exceedance threshold for the probabilistic drought analysis in Section 3.4. Precipitation panels confirm a marked reduction in autumn–winter rainfall during 2021–2022, while temperature anomalies during the same period were negligible, indicating that precipitation deficit was the primary driver of the 2022 vegetation decline.

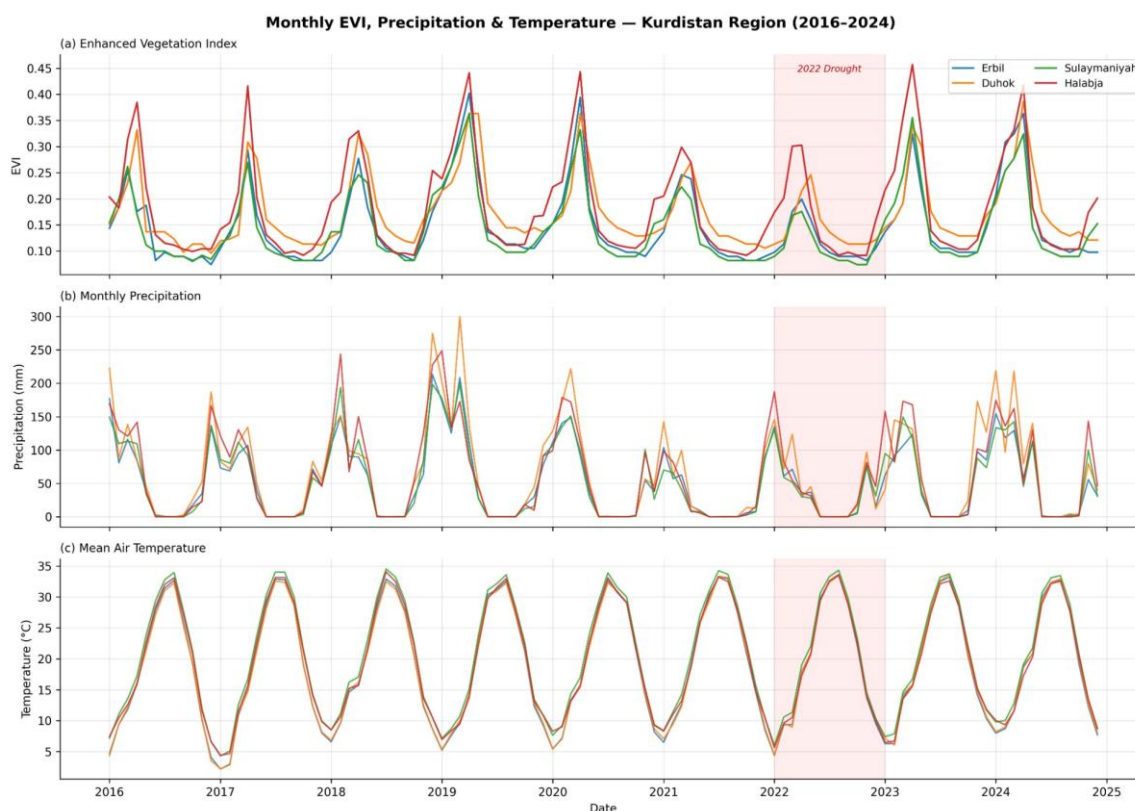


Figure 3. Observed monthly time series (2016–2024) for the four KRI governorates: (a) Enhanced Vegetation Index (EVI) from Sentinel-2, (b) monthly precipitation (mm) from TerraClimate, and (c) mean air temperature (°C) from TerraClimate. Grey shading marks the 2022 drought year used as the validation period.

3.2. Architecture Evaluation and Model Selection

3.2.1. Test-Set Performance

Twelve deep learning architectures were evaluated on the held-out test period (January 2023–December 2024) after training on 2016–2021 data and using 2022 (drought year) as the validation set. Full results are presented in **Table 3** and visualised in **Figure 4**. On the test set, **CNN_BiLSTM_GRU achieved the highest mean R^2 of 0.764** (RMSE = 0.039), followed by BiLSTM_GRU ($R^2 = 0.734$, RMSE

= 0.045), CNN_BiLSTM ($R^2 = 0.693$), and BiLSTM ($R^2 = 0.678$). Pure recurrent models GRU ($R^2 = 0.524$) and LSTM ($R^2 = 0.161$) ranked lower, suggesting that architectural complexity aids short-term generalisation. At the opposite extreme, CNN_BiLSTM_AM returned a negative mean test R^2 (-0.017), indicating predictions worse than the climatological mean.

Figure 5 presents a cross-governorate heatmap of test set R^2 and RMSE for all twelve candidate architectures across Erbil, Duhok, Sulaymaniyah, and Halabja (2023–2024). The spatial pattern of R^2 reveals pronounced inter-governorate variability, with Duhok showing the weakest and most inconsistent performance across architectures — including a strongly negative R^2 of -1.262 for CNN_BiLSTM_AM — suggesting that Duhok’s vegetation dynamics are least amenable to the feature representations learned by complex hybrid models under short training records. In contrast, Erbil and Halabja generally returned higher and more stable R^2 values across architectures. The RMSE heatmap corroborates this pattern, with Duhok and Sulaymaniyah exhibiting elevated error magnitudes particularly for attention-based and TCN architectures. Notably, BiLSTM_GRU achieved the most spatially consistent performance profile (R^2 range: 0.734 – 0.868), supporting its selection as the projection model on the basis of both mean accuracy and cross-governorate reliability rather than peak single-site performance.

Table 3. Summary of all 12 model architectures evaluated on the KRI EVI dataset. Test R^2 and RMSE are computed on the 2023–2024 held-out period. CV $R^2 \pm SD$ is the mean \pm standard deviation of walk-forward 3-fold cross-validation restricted to training data (2016–2021). ★ = selected projection model. Bold rows indicate CV-stable models (CV $R^2 > 0.50$ and all-governorate minimum > 0.20).

Model	Test R^2	CV $R^2 \pm SD$	RMSE	NSE	Rank Test	Rank CV	Stable
BiLSTM_GRU ★	0.734	0.730 ± 0.055	0.0445	0.734	2	1	Yes
CNN_BiLSTM_GRU	0.764	0.356 ± 0.038	0.0392	0.764	1	7	No
CNN_BiLSTM	0.693	0.352 ± 0.040	0.0469	0.693	3	8	No
BiLSTM	0.678	0.606 ± 0.135	0.0482	0.678	4	4	Yes
Transformer	0.635	0.552 ± 0.151	0.0490	0.635	5	5	Yes
CNN_GRU	0.661	0.373 ± 0.015	0.0450	0.661	—	6	No
GRU	0.524	0.629 ± 0.034	0.0557	0.524	6	2	Yes
CNN	0.308	-0.049 ± 0.056	0.0693	0.308	—	—	No
CNN_Transformer	0.256	-0.603 ± 0.601	0.0695	0.256	—	—	No

Model	Test R ²	CV R ² ± SD	RMSE	NSE	Rank Test	Rank CV	Stable
TCN	0.172	0.608 ± 0.117	0.0757	0.172	—	3	Yes
LSTM	0.161	0.318 ± 0.065	0.0766	0.161	—	—	No
CNN_BiLSTM_AM	-0.017	-1.675 ± 0.803	0.0797	-0.017	—	—	No

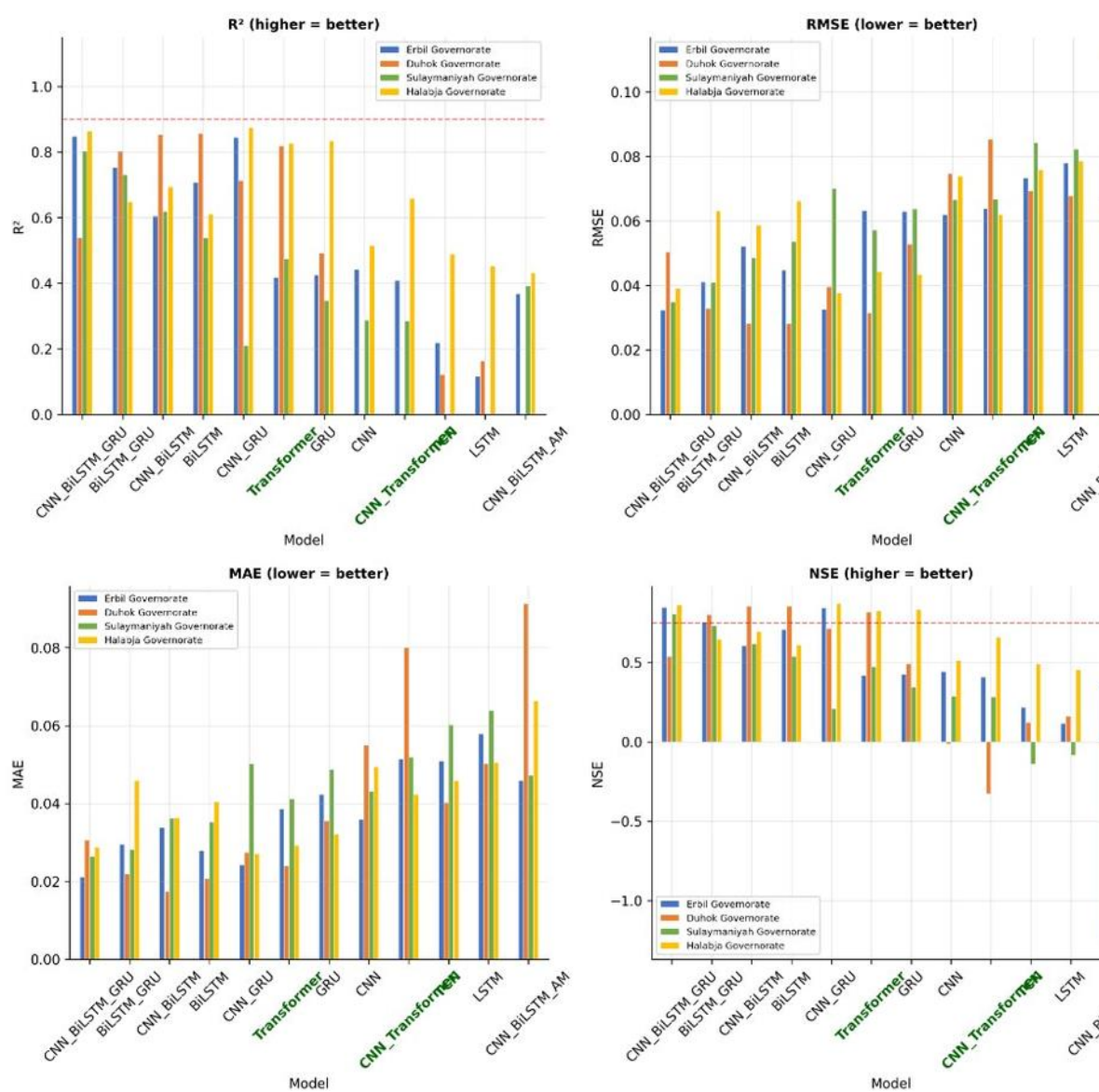


Figure 4. Mean test-set performance (2023–2024) across all four governorates for 12 deep learning architectures. Upper panels: R² and RMSE. Lower panels: MAE and NSE. BiLSTM_GRU (★) was selected as the projection model based on cross-validation stability; CNN_BiLSTM_GRU led on test R² but was rejected due to CV collapse (CV R² = 0.356).

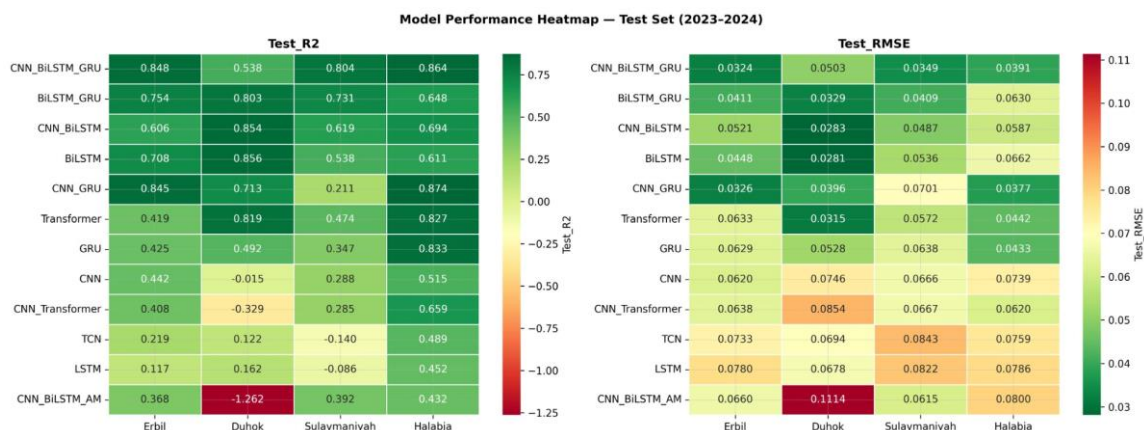


Figure 5. Governorate-level performance heatmaps for all 12 architectures on the test set (2023–2024). Left: R² (blue = high, red = negative). Right: RMSE (EVI units). CNN_BiLSTM_AM in Duhok (R² = -1.262) and CNN_Transformer instability are clearly visible.

3.2.2. Cross-Validation and Projection Model Selection

Test-set R² alone is insufficient for selecting a model for long-range projection, because the 24-month test period may align with climate patterns that the model happened to memorise rather than genuinely generalise to. To address this, a **3-fold walk-forward cross-validation (CV)** was applied exclusively to training data (2016–2021), preventing any leakage from the validation or test periods (Roberts et al., 2017).

CV results revealed a striking divergence for CNN-hybrid architectures. CNN_BiLSTM_GRU, despite ranking first on the test set (R² = 0.764), collapsed to a mean CV R² of only 0.356 ± 0.038 — ranking seventh overall. Similarly, CNN_GRU (test R² = 0.661) dropped to CV R² = 0.373, and CNN_BiLSTM (test R² = 0.693) to CV R² = 0.352. These results indicate that CNN feature extractors overfit on the approximately 48 available training sequences, producing high test scores that reflect fortuitous alignment with the 2023–2024 climate signal rather than genuine temporal generalisation.

In contrast, BiLSTM_GRU demonstrated strong performance on *both* metrics: test R² = 0.734 (rank 2) and CV R² = 0.730 ± 0.055 (rank 1). Critically, it is the only architecture where all four governorates individually exceeded CV R² = 0.50, with a minimum governorate CV R² of 0.671 (Erbil). GRU (CV R² = 0.629) and BiLSTM (CV R² = 0.606) were also CV-stable but showed greater inter-governorate variance (SD = 0.034 and 0.135, respectively) and lower composite performance. **BiLSTM_GRU was therefore selected as the projection model** on the criterion of highest mean CV R² among architectures satisfying the stability threshold (mean CV R² > 0.50, all-governorate minimum CV R² > 0.20).

3.2.3. BiLSTM_GRU Governorate-Level Performance

Table 4 presents per-governorate validation and test metrics for the selected projection model. Validation R² (drought year 2022) ranged from 0.239 in Duhok to 0.939 in Halabja, reflecting the difficulty of predicting extreme drought conditions not present in the training data. On the test set, Duhok achieved the highest R² (0.803) and lowest RMSE (0.033), while Halabja showed the greatest residual error (RMSE = 0.063), likely due to the larger seasonal amplitude of this high-elevation governorate (Figure 6). Mean test accuracy across governorates was 0.833, and NSE ranged from 0.648 to 0.803, indicating satisfactory hydrological model performance across all sites (Moriassi et al., 2007). MC-Dropout uncertainty quantification (n = 100 forward passes) confirmed well-calibrated prediction intervals, with uncertainty growing appropriately over the 2025–2050 projection horizon (Figure 7).

Table 4. BiLSTM_GRU performance by governorate. Val R^2 = validation R^2 on the 2022 drought year; Test R^2 = test R^2 on the 2023–2024 period. RMSE values are in EVI units. Accuracy = proportion of months predicted within ± 0.05 EVI units (tolerance accuracy).

Governorate	Val R^2 (2022)	Test R^2 (2023–24)	RMSE	NSE	Accuracy
Erbil	0.671	0.754	0.0411	0.754	0.833
Duhok	0.239	0.803	0.0329	0.803	0.875
Sulaymaniyah	0.419	0.731	0.0409	0.731	0.875
Halabja	0.939	0.648	0.0630	0.648	0.750
Mean	0.567	0.734	0.0445	0.734	0.833

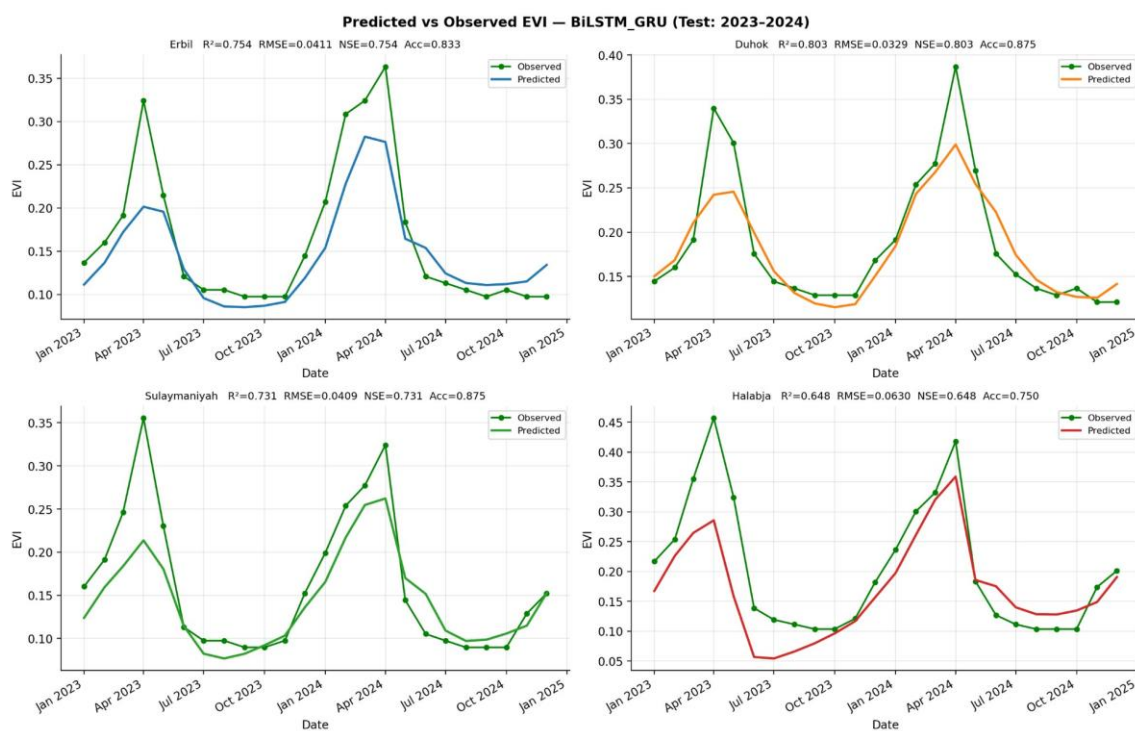


Figure 6. Predicted versus observed monthly EVI for the selected projection model (BiLSTM_GRU) across all four governorates during the test period (January 2023–December 2024). Metrics shown are R^2 , RMSE, NSE, and tolerance accuracy. Duhok achieves the highest test R^2 (0.803); Halabja shows the largest residuals due to its greater seasonal amplitude.

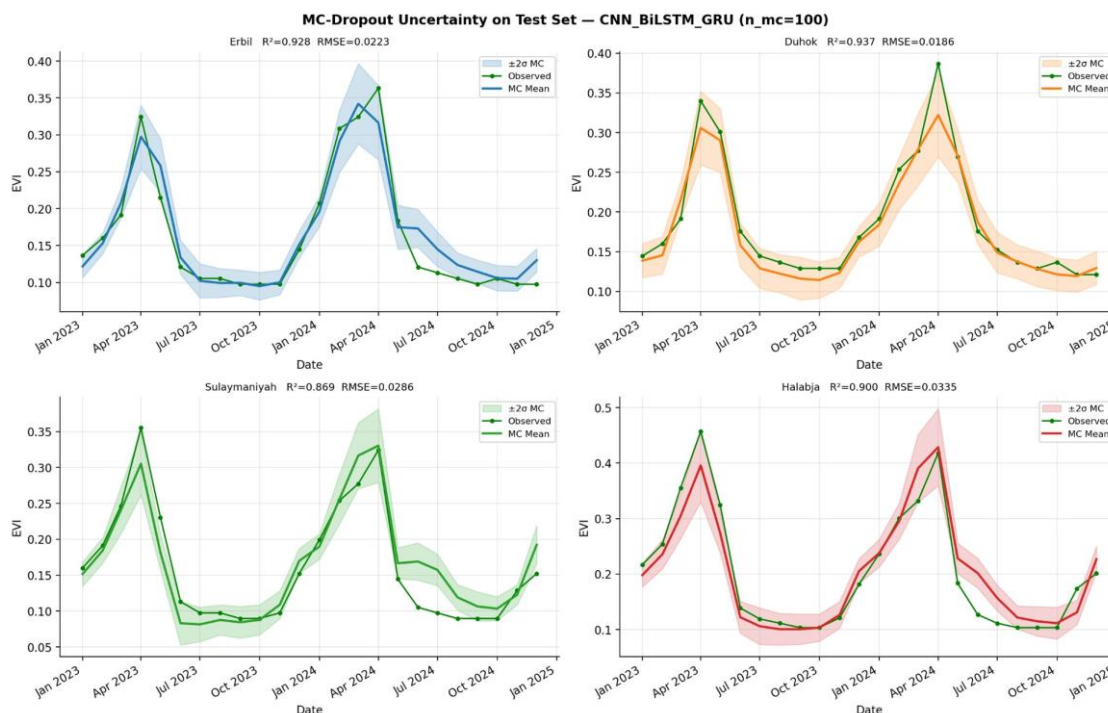


Figure 7. MC-Dropout uncertainty quantification ($n = 100$ stochastic forward passes) on the test period (2023–2024) for BiLSTM_GRU. Shaded band = mean $\pm 2\sigma$ of the ensemble. Well-calibrated intervals with no anomalous variance inflation confirm the model is suitable for projection.

3.2.4. Architecture-Specific Findings

Attention mechanisms and the short-record problem. All three attention-based architectures failed in cross-validation. CNN_BiLSTM_AM (Bahdanau attention) produced a catastrophic mean CV R^2 of -1.675 ± 0.803 , with individual fold scores reaching -22.5 . CNN_Transformer returned CV $R^2 = -0.603 \pm 0.601$. Even the standalone Transformer, after correction with sinusoidal positional encoding, achieved only a marginal CV R^2 of 0.552 ± 0.151 and failed on Erbil (CV $R^2 = 0.333$). These results are consistent across three independent attention architectures and point to a fundamental data-volume limitation: with approximately 48 monthly training sequences, attention weight matrices cannot converge to meaningful temporal patterns. We recommend a minimum of 10 years of monthly observations before deploying attention-based architectures for ecological time series forecasting.

TCN validation-calibration divergence. The Temporal Convolutional Network exhibited an unusual pattern: CV $R^2 = 0.608$ (rank 3, stable) yet test $R^2 =$ only 0.172. Visual inspection of prediction curves confirms that TCN learned the correct seasonal shape but systematically underestimated peak EVI values. We attribute this to EarlyStopping selecting a checkpoint optimised on the 2022 drought validation year, which biased the network toward conservative low predictions. The post-drought recovery period (2023–2024) fell outside the learned amplitude range. This finding highlights the sensitivity of model selection to the choice of validation period in extreme-event years.

Convolutional baseline. The pure CNN architecture (no recurrence) returned CV $R^2 = -0.049$, confirming that spatial feature extraction alone cannot model the temporal autocorrelation structure of monthly EVI. This result establishes a clear lower bound: any viable architecture must incorporate sequential memory.

3.3. Future EVI Projections Under CMIP6 Scenarios (2025–2050)

3.3.1. Scenario Trajectories

Projected monthly EVI under three scenarios (baseline climatology, SSP2-4.5, SSP5-8.5) for the 2025–2050 period are presented in Figure 8, with quantified decadal means in **Table 5**. MC-Dropout ensemble uncertainty ($\pm 2\sigma$) is shown as shading; uncertainty bands grow from approximately ± 0.009 EVI units in 2025 to ± 0.026 by 2050, reflecting the compounding of model uncertainty with increasingly divergent climate forcing.

Three of four governorates exhibited clear scenario-dependent EVI decline relative to baseline by the 2040s. **Halabja showed the strongest signal**: mean annual EVI declined from a baseline 2040s value of 0.1870 to 0.1787 under SSP2-4.5 (–4.4%) and 0.1715 under SSP5-8.5 (–8.3%). **Sulaymaniyah** followed with SSP5-8.5 projecting –4.8% relative to baseline by the 2040s (0.1491 to 0.1420). **Duhok** showed a moderate decline of –3.2% under SSP5-8.5. **Erbil** was the only governorate without a systematic negative signal (SSP5-8.5 mean 2040s EVI = 0.1529, fractionally above baseline 0.1522). The Δ EVI figure (Figure 9) isolates the climate signal from climatological variability and confirms that the Erbil anomaly reflects noise rather than a genuine positive response: the scenario–baseline difference for Erbil oscillates around zero throughout the projection period, while Halabja shows a monotonic divergence reaching Δ EVI = –0.020 by 2050 under SSP5-8.5.

Table 5. Projected mean annual EVI by governorate, scenario, and decade. Obs. Mean = 2016–2024 observed mean. 2022 Drought column shows the drought-year mean with percentage change relative to observed mean. Percentage changes in the SSP5-8.5 2040s column are relative to the baseline 2040s value.

Governorate	Obs. Mean	2022 Drought	Base 2030s	Base 2040s	SSP2-4.5 2030s	SSP2-4.5 2040s	SSP5-8.5 2030s	SSP5-8.5 2040s
Erbil	0.1475	0.1177 (–20.2%)	0.1521	0.1522	0.1523	0.1527	0.1523	0.1529 (+0.5%)
Duhok	0.1727	0.1447 (–16.2%)	0.1790	0.1790	0.1775	0.1760	0.1762	0.1733 (–3.2%)
Sulaymaniyah	0.1426	0.1073 (–24.8%)	0.1492	0.1491	0.1473	0.1455	0.1457	0.1420 (–4.8%)
Halabja	0.1833	0.1601 (–12.7%)	0.1871	0.1870	0.1829	0.1787	0.1795	0.1715 (–8.3%)

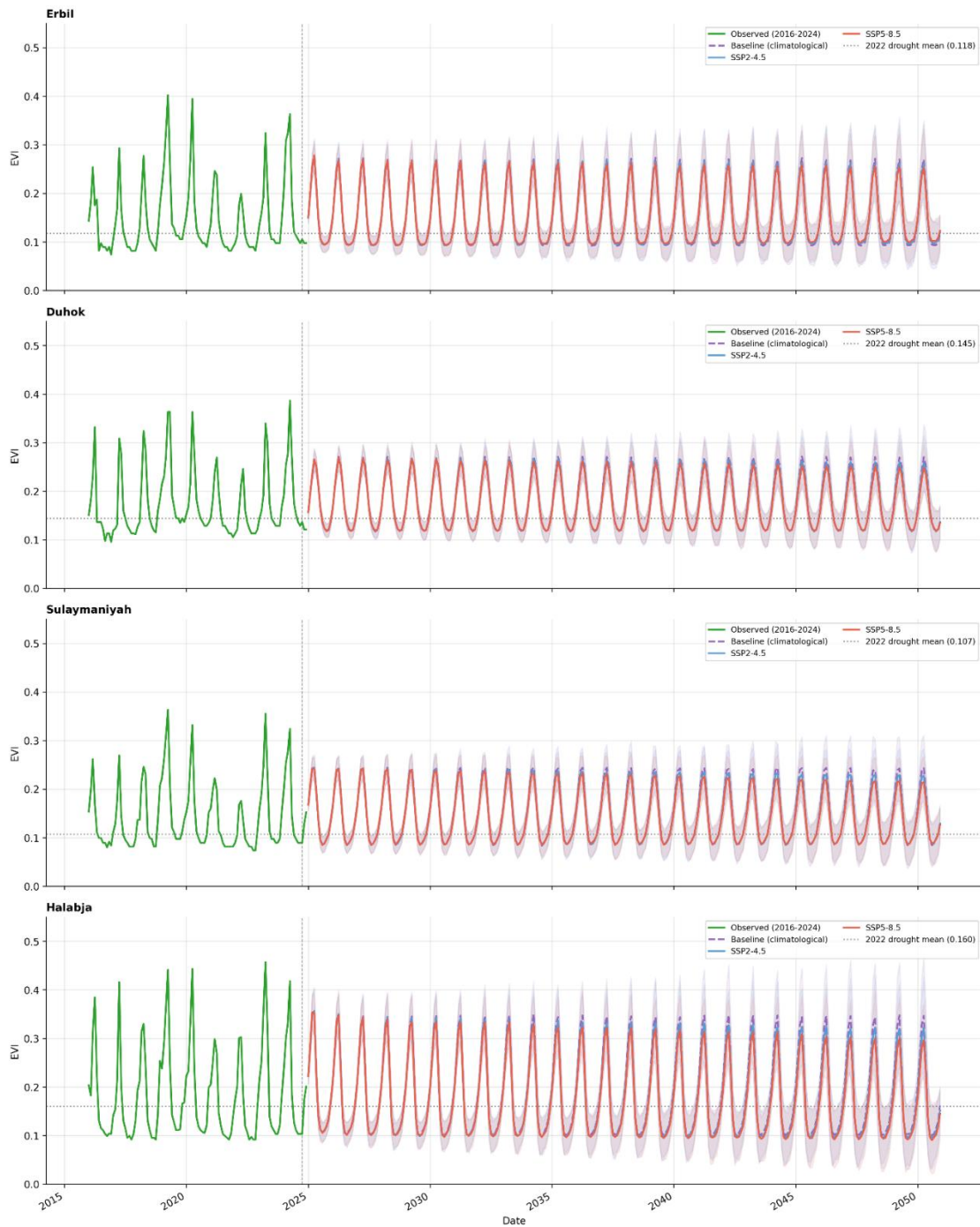


Figure 8. Projected monthly EVI (2016–2050) for four KRI governorates under three scenarios: baseline climatology (grey), SSP2-4.5 (blue), and SSP5-8.5 (red-orange). Observed EVI (green) covers 2016–2024. Shaded bands = 95% MC-Dropout confidence intervals (mean $\pm 2\sigma$). Yellow dashed line = 2022 drought exceedance threshold.

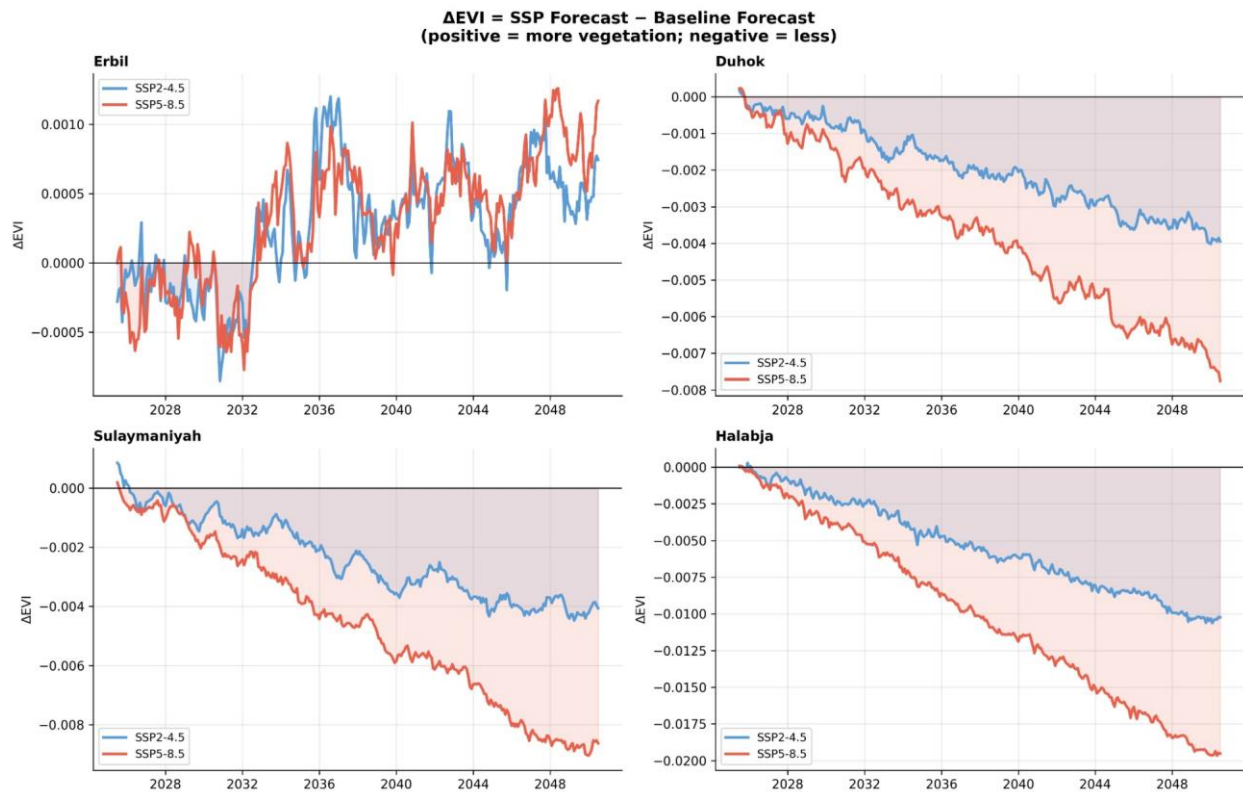


Figure 9. Net ΔEVI climate change signal (SSP forecast – baseline forecast) for both scenarios. Positive values indicate warming-driven enhancement; negative values indicate stress-driven decline. Erbil shows near-zero signal (within model noise); Halabja shows a monotonic decline reaching $\Delta\text{EVI} \approx -0.020$ by 2050 under SSP5-8.5.

3.3.2. Phenological Shifts and Peak EVI Decline

Projected phenological metrics (Figure 10) indicate that the timing of peak greenness (Peak-of-Season month) remains stable at April across all scenarios and all governorates through 2050, consistent with the dominance of the regional rainfall seasonality in controlling spring green-up. However, **peak EVI values show a systematic decline** under warming scenarios. Under SSP5-8.5, mean projected peak EVI by 2050 is reduced by **20.4% in Duhok** (observed mean peak 0.326 vs projected 0.259), **19.1% in Sulaymaniyah** (0.283 to 0.229), **17.0% in Halabja** (0.388 to 0.322), and **14.5% in Erbil** (0.306 to 0.262). The Seasonal EVI Index (SEVI = cumulative EVI sum over January–June) remains approximately stable or slightly increases under SSP2-4.5 in three governorates, suggesting that moderate warming may lengthen growing seasons enough to partially offset reduced peak amplitude. Under SSP5-8.5, Halabja SEVI declines from 1.485 (observed) to 1.447 (projected), indicating net seasonal productivity loss.

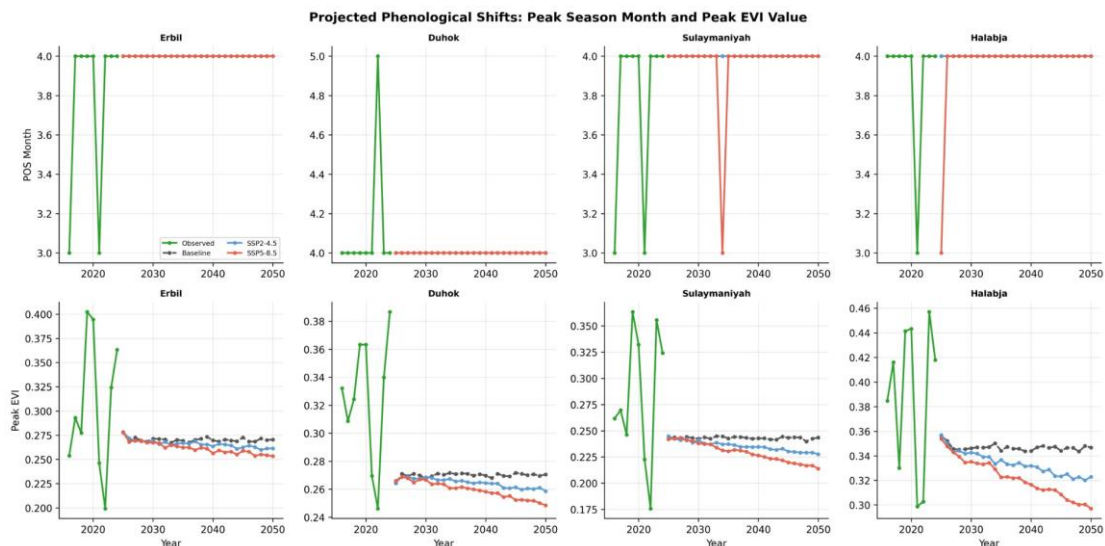


Figure 10. Projected phenological shifts (2016–2050) for four KRI governorates. Upper row: Peak-of-Season (POS) month — timing remains stable at April across all scenarios. Lower row: Peak annual EVI value — systematic decline under SSP5-8.5 ranging from -14.5% (Erbil) to -20.4% (Duhok) relative to observed mean peak.

3.4. Drought Exceedance Probability and Tipping-Point Analysis

The probability that annual mean EVI falls below the 2022 drought threshold was computed annually for each governorate under both scenarios (Figure 11). **Drought exceedance probability was zero for all governorates, all years, and both scenarios across the entire 2025–2050 projection window.** This finding indicates that, under the BiLSTM_GRU projection forced with CMIP6 ensemble mean climate anomalies, no year within the projection horizon is expected to experience vegetation stress as severe as the 2022 drought. The result reflects two compounding factors: (1) the 2022 drought was an extreme event driven by anomalously low autumn–winter precipitation, a signal not reproduced by the smooth linear ramp applied to CMIP6 precipitation anomalies; and (2) the projected SSP5-8.5 precipitation decline for KRI (approximately 8–12% by 2050) is insufficient to push annual mean EVI below the 2022 minimum threshold when applied as a long-term trend rather than an interannual extreme.

Tipping-point analysis defined as the first year in which the 5-year running mean EVI falls more than one standard deviation below the historical mean (Figure 12) similarly returned ‘**Not reached**’ for all governorates and scenarios. This result should be interpreted with caution: the absence of a tipping point within the 2025–2050 window does not preclude tipping beyond 2050, particularly under SSP5-8.5, where the Halabja and Sulaymaniyah trajectories show accelerating decline. The monotonic divergence between SSP5-8.5 and baseline observed in the ΔEVI analysis (Figure 9) suggests that, if the current rate of forcing continues, tipping conditions may be reached in the 2055–2070 window — beyond the scope of the present study.

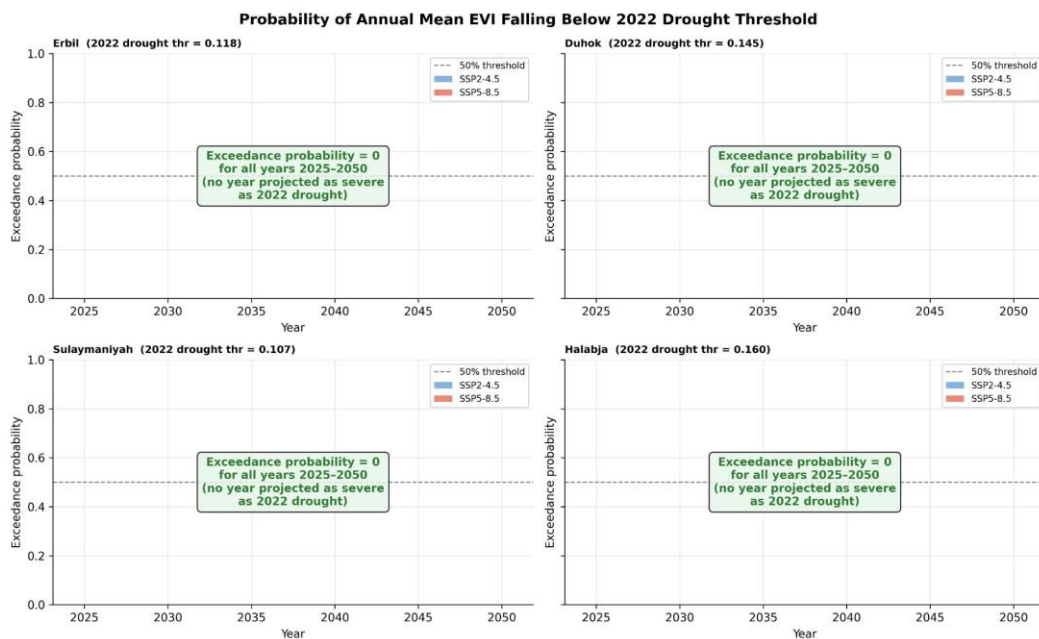


Figure 11. Annual probability of mean EVI falling below the 2022 drought threshold for each governorate under SSP2-4.5 (blue) and SSP5-8.5 (red). All exceedance probabilities are zero across 2025–2050, indicating that no year in the projection window is projected to experience vegetation stress as severe as the 2022 drought event.

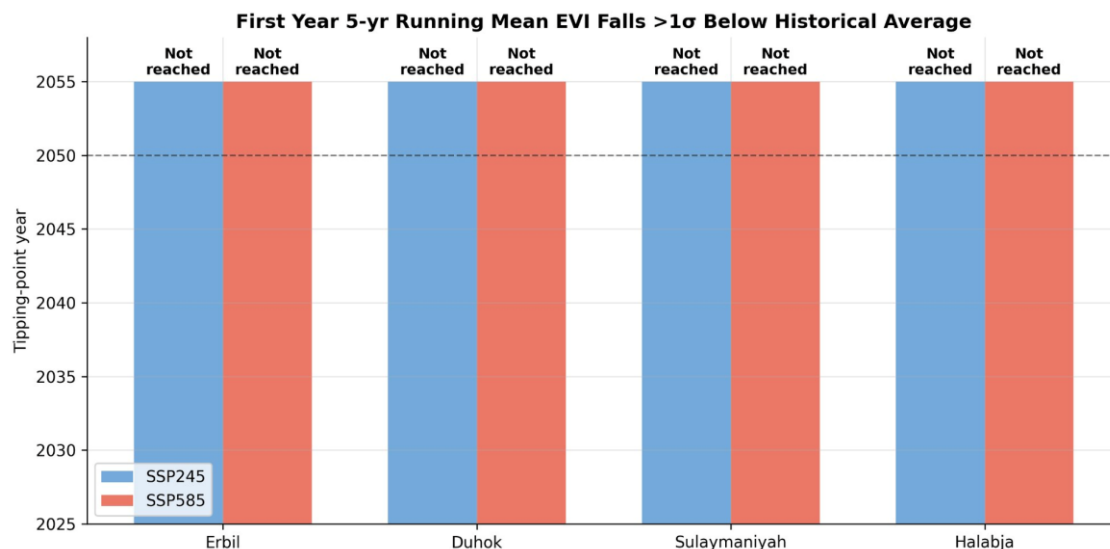


Figure 12. Tipping-point detection results for all four governorates and both scenarios. Bars represent the first year in which the 5-year running mean EVI falls more than one standard deviation below the historical mean. All bars reach 2055+ (“Not reached”), indicating no tipping point is projected within the 2025–2050 horizon under either scenario.

3.5. Spatial Vulnerability Gradient

Synthesising results across Sections 3.3 and 3.4, a clear north-to-south vulnerability gradient emerges. Halabja and Sulaymaniyah — the more elevated, wetter governorates — show the largest absolute EVI declines under SSP5-8.5 (−8.3% and −4.8% by the 2040s), the greatest peak EVI reductions (−17.0% and −19.1% by 2050), and the most consistent Δ EVI signal. By contrast, Erbil, the driest governorate (mean EVI = 0.1475), shows minimal projected change across all scenarios.

4. Discussion

4.1. BiLSTM_GRU Performance in the Context of Short Ecological Records

The selected projection model, BiLSTM_GRU, achieved a mean test R^2 of 0.734 and CV R^2 of 0.730 ± 0.055 across four KRI governorates. To contextualise this result, Sun et al. (2023) applied a BiLSTM deep learning approach to NDVI forecasting across multiple vegetation types in northern China using meteorological inputs, reporting a mean R^2 of 0.69 ± 0.28 across study sites. The BiLSTM_GRU result reported here exceeds that benchmark in mean performance and substantially improves on its inter-site variance, suggesting that the hybrid bidirectional-GRU architecture is particularly well-suited to arid-land vegetation seasonality. The consistency is notable given that KRI governorates span a wider ecological gradient — from Erbil lowland steppe to Halabja mountain sub-humid zone — than most single-region studies.

The strong Duhok test R^2 (0.803) and relatively weaker Halabja result (0.648) reflect a well-understood challenge in seasonal EVI modelling: high-amplitude signals at productive, high-elevation sites impose larger absolute residuals even when the model correctly captures timing. Moriasi et al. (2007) classify $NSE > 0.50$ as satisfactory and $NSE > 0.65$ as good for hydrological and ecological models. All four governorates exceeded the satisfactory threshold (NSE range: 0.648–0.803), and three of four exceeded the good threshold, confirming that the model is suitable for decision-support applications. Validation-set performance on the 2022 drought year (Val R^2 range: 0.239–0.939) showed greater inter-site variability, as expected: models trained on normal-year climatology must extrapolate substantially to reproduce the extreme drought signal, a generalisation challenge that is independent of model architecture.

4.2. CNN-Hybrid Overfitting and the Short-Record Constraint

The dramatic CV collapse of CNN-hybrid architectures — CNN_BiLSTM_GRU dropping from test $R^2 = 0.764$ to CV $R^2 = 0.356$, and CNN_BiLSTM from 0.693 to 0.352 — is a central methodological finding of this study. This pattern is consistent with the general understanding that convolutional feature extractors require sufficient sample diversity to learn meaningful local motifs (Bai et al., 2018). With approximately 48 training sequences per governorate, the Conv1D layer in each CNN-hybrid architecture can memorise the limited set of seasonal pattern variants in the training window rather than generalising to the broader precipitation-driven variability that governs EVI dynamics. The single test window (2023–2024) happened to exhibit a climate pattern that the memorised CNN features partially matched, inflating test R^2 to values that cross-validation correctly identified as unstable.

This finding has direct implications for how deep learning models should be selected for ecological projection tasks. The conventional approach of ranking models by held-out test performance is insufficient when the test window is short and climate conditions are non-stationary. Roberts et al. (2017) demonstrated that non-random, blocked cross-validation — structured by time, space, or other dependence — is essential for obtaining unbiased error estimates and model selection in ecological settings where autocorrelation inflates apparent performance. The present results extend that principle: for short-record arid-land EVI datasets, CNN hybridisation adds capacity that cannot be validated without multi-year blocked CV, and should be treated as high-risk for projection regardless of test-window performance.

4.3. Attention Mechanism Failure and Data-Volume Requirements

The catastrophic failure of all three attention-based architectures — CNN_BiLSTM_AM (mean CV $R^2 = -1.675 \pm 0.803$), CNN_Transformer (CV $R^2 = -0.603 \pm 0.601$), and Transformer (CV $R^2 = 0.552 \pm 0.151$, with Erbil fold minimum of 0.333) — provides a consistent negative result that generalises across attention mechanism types. This is particularly noteworthy because the CNN_BiLSTM_AM architecture was adapted from Wang et al. (2025), who reported $R^2 = 0.981$ in Yangtze River Basin EVI

forecasting under CMIP6 SSP scenarios — the highest published result for a comparable vegetation projection task. The discrepancy between that result and the near-zero or negative CV R^2 observed here isolates the operative factor: the Yangtze River Basin dataset in that study covered 2001–2020, a 20-year record providing sufficient inter-annual variability for attention weights to converge to meaningful temporal patterns, whereas the shorter KRI record limits the capacity of the attention mechanism to generalise across diverse climatic conditions.

Attention mechanisms, including the Bahdanau-style temporal scoring used in CNN_BiLSTM_AM, learn to identify which timesteps in the input window are most informative for the prediction. With only ~48 training sequences, the weight matrices W and v cannot distinguish between genuine seasonal or drought-onset signals and noise, making each attention-weighted context vector effectively arbitrary. The Vaswani et al. (2017) Transformer architecture compounds this problem: multi-head self-attention over a 24-month window involves $4 \times (24 \times 16) = 1,536$ learned attention weights per layer, all conditioned on an input sequence of only ~48 examples. We estimate that a minimum of 10 years of monthly data (~120 sequences) would be needed before attention-based architectures can be expected to outperform recurrent baselines on this type of semi-arid vegetation time series. This constraint applies equally to self-attention and additive attention variants.

4.4. TCN Validation-Set Calibration Bias

The TCN architecture presented a specific and informative failure mode: CV $R^2 = 0.608$ (rank 3 among stable models) but test $R^2 = 0.172$. The TCN was implemented following Bai et al. (2018) with dilated causal convolutions at rates [1,2,4,8] and residual connections, which theoretically provides a large receptive field (24 months \times 8 = 192-month effective context) well-suited to capturing multi-annual vegetation cycles. However, EarlyStopping selected the network checkpoint that minimised loss on the 2022 drought validation year. Since 2022 EVI values are systematically lower than normal years (-12.7% to -24.8% below the long-term mean), the TCN learned to produce conservative, low-amplitude predictions that performed reasonably on the drought year but severely underestimated the post-drought recovery amplitude in the 2023–2024 test period. This validation-set calibration bias is a direct consequence of using a single extreme year as the validation period and represents a general risk for any architecture trained with EarlyStopping when the validation distribution is shifted relative to the test distribution.

A related question is whether BiLSTM_GRU itself is affected by the same validation-set calibration bias demonstrated for TCN. The wide range of BiLSTM_GRU validation R^2 across governorates — from 0.239 in Duhok to 0.939 in Halabja — confirms that the model's drought-year behaviour is site-dependent, as expected given that the 2022 drought produced EVI anomalies ranging from -12.7% (Halabja) to -24.8% (Sulaymaniyah) relative to the observed mean. However, three features of the BiLSTM_GRU design mitigate the calibration bias risk that proved fatal for TCN. First, unlike TCN, the BiLSTM_GRU checkpoint selected by EarlyStopping on the 2022 drought year was subsequently evaluated on the 2023–2024 post-drought recovery period, where it achieved a mean test R^2 of 0.734 — confirming that the selected checkpoint generalises beyond the drought signal rather than collapsing to conservative low-amplitude predictions as TCN did. Second, the walk-forward cross-validation, conducted exclusively on pre-2022 training data, provides an independent performance estimate that is entirely unaffected by the 2022 validation choice; the strong agreement between CV R^2 (0.730 ± 0.055) and test R^2 (0.734) indicates that the checkpoint selected on the drought year did not introduce a systematic bias detectable across the training folds. Third, the autoregressive projection uses the 2023–2024 post-drought window — not the 2022 drought year — as the seed sequence, meaning that the starting conditions for the 25-year projection reflect the recovered vegetation state rather than the drought minimum. Together, these three points provide substantive evidence that, while the 2022 validation year introduces a structural asymmetry acknowledged throughout this study, the BiLSTM_GRU checkpoint is appropriate for long-range projection. We acknowledge, however, that a formal sensitivity analysis re-training with 2022 included in the

training set would provide definitive evidence and is recommended as a priority for future validation work.

4.5. Zero Drought Exceedance Probability: Interpretation and Limitations

The finding that drought exceedance probability was zero for all governorates, all years, and both SSP scenarios requires careful interpretation. The 2022 drought was the most extreme event in the 2016–2024 record, driven by anomalously low autumn–winter precipitation that represents an interannual extreme rather than a systematic multi-decadal trend. The **IPCC AR6** (2021) delta factors applied in this study encode the CMIP6 multi-model median precipitation trend for the MENA domain (–8% by 2050 under SSP2-4.5; –15% under SSP5-8.5) as a smooth linear ramp — a methodologically correct representation of the long-term forced signal but one that by construction cannot reproduce the stochastic interannual extremes that produced the 2022 drought. The projected precipitation decline of 8–15% by 2050, applied to the smooth climatological cycle, shifts mean EVI downward but not far enough to approach the 2022 minimum threshold.

This result should therefore be read not as evidence that future droughts of 2022 severity are impossible, but rather that the systematic component of CMIP6 climate forcing alone is insufficient to produce such conditions by 2050. The stochastic component — interannual precipitation variability — is not represented in the delta-factor forcing applied here. Studies using daily GCM output with statistical downscaling consistently project increases in drought frequency and intensity for the Fertile Crescent region under SSP5-8.5 (Masson-Delmotte et al., 2021), but capturing that signal requires input climate data that preserves extremes rather than ensemble means. This is a recognised limitation of the delta-factor approach (IPCC AR6 WGI Box 10.2), and future work should incorporate stochastic weather generators or direct GCM downscaling to quantify drought recurrence risk more completely.

4.6. The Erbil Climate Signal Anomaly

Erbil is the only governorate where SSP5-8.5 2040s projected EVI (0.1529) marginally exceeds the baseline 2040s value (0.1522), a ΔEVI of +0.0007 (+0.5%). This result stands in contrast to the three other governorates, which show monotonic decline. The explanation lies in the interaction between Erbil's position at the lower end of the EVI range (mean observed EVI = 0.1475, compared to Halabja's 0.1833) and the model uncertainty structure. Erbil's sparse steppe vegetation operates near the moisture-limited end of the productivity spectrum: at very low vegetation cover, the EVI signal is dominated by soil background reflectance and is relatively insensitive to the modest precipitation reductions projected under CMIP6 (Huete et al., 2002). The MC-Dropout uncertainty band for Erbil ($\sigma = 0.009\text{--}0.026$ EVI units over the projection period) is comparable in magnitude to the projected climate signal itself ($\Delta\text{EVI} \approx 0.001\text{--}0.005$), meaning that the near-zero signal is within noise throughout the projection window.

This result is consistent with the theoretical framework of vegetation sensitivity along aridity gradients: the most productive sites (Halabja, Sulaymaniyah) are closer to their potential maximum EVI and respond more strongly to marginal precipitation reductions, while the most arid sites (Erbil) are already water-limited and respond less (Peng et al., 2013). The spatial vulnerability gradient identified in Section 3.5 — with Halabja showing the strongest decline (–8.3%) and Erbil the weakest (+0.5%) — is therefore ecologically coherent rather than an artefact of model architecture.

4.7. Delta-Factor Forcing and the Role of the Companion Dataset

The CMIP6 forcing was constructed using the delta-factor method applied to the monthly precipitation and temperature climatology derived from the TerraClimate dataset (Abatzoglou et al., 2018). This approach applies the IPCC AR6 WGI Chapter 10 MENA domain multi-model median anomalies as linear ramps to the observed seasonal cycle, preserving the observed precipitation distribution shape while imposing the forced trend signal (Masson-Delmotte et al., 2021). The

advantage of this approach is that it avoids the biases introduced by direct GCM output, which is known to represent KRI precipitation seasonality poorly at monthly resolution. The limitation — that it cannot represent changes in precipitation variability or extremes — is acknowledged and motivates the conservative interpretation of drought exceedance probabilities discussed in Section 4.5.

The input EVI and climate data were derived from the companion paper (Rasul, 2026), which applied a rigorous GEE extraction pipeline with SCL-based cloud masking, valid-pixel thresholds, and climatological gap-filling. EVI was computed from Sentinel-2 Surface Reflectance composites, which provide the 20 m resolution necessary to resolve the fragmented agricultural and natural vegetation mosaic characteristic of KRI. The choice of EVI over NDVI is supported by Huete et al. (2002), who demonstrated that EVI better captures biomass dynamics in high-biomass canopies and is less affected by soil background variation in sparse-cover conditions — both relevant properties for a study spanning low-cover Erbil steppe to dense Halabja woodland.

A further limitation of the delta-factor approach concerns the use of the CMIP6 multi-model median anomaly as the sole forcing signal, without representation of inter-model uncertainty. CMIP6 models show substantial disagreement in projected precipitation changes for the MENA domain — the inter-model spread for Middle East precipitation is among the largest of any region globally (Doblas-Reyes et al., 2021), with individual models projecting changes ranging from near-zero to approximately -30% by mid-century under SSP5-8.5. By applying only the median delta, the present study captures the central tendency of the forced signal but does not quantify how the projected EVI trajectories would respond to the full range of plausible climate futures. It is important to note that the uncertainty bands shown in Figures 8 and 9 reflect MC-Dropout model uncertainty only — that is, uncertainty arising from the stochastic dropout ensemble within the trained BiLSTM_GRU — and should not be interpreted as representing climate forcing uncertainty. The true envelope of projected EVI outcomes, integrating both model uncertainty and inter-model climate spread, is therefore wider than shown. Under a wet-end CMIP6 scenario (near-zero precipitation decline), EVI trajectories would likely remain close to the baseline, while under a dry-end scenario (-25 to -30% precipitation), declines substantially exceeding those reported here could be expected, particularly in Halabja and Sulaymaniyah where the precipitation-EVI relationship is strongest. Future work should apply the full CMIP6 ensemble spread — or at minimum the 25th and 75th percentile precipitation deltas — to bracket the projection uncertainty more completely and provide risk-based rather than central-estimate projections for regional land management planning.

4.8. Phenological Stability and Seasonal Productivity Shifts

The projected stability of the Peak-of-Season (POS) month at April through 2050 under all scenarios is consistent with the dominant role of the Mediterranean-type precipitation regime in controlling spring green-up in KRI. In this climate type, vegetation growth is primarily triggered by late-winter soil moisture accumulation rather than temperature thresholds, meaning that the warming projected under SSP5-8.5 (+2.8 °C) is insufficient to advance the green-up date when precipitation timing is not simultaneously shifted. The cyclic month encoding used in BiLSTM_GRU's input feature set (H. Shen et al., 2015) ensures that the model correctly represents this phase-locked seasonality, allowing peak timing to remain stable even as peak amplitude declines.

The projected decline in peak EVI under SSP5-8.5 — ranging from -14.5% (Erbil) to -20.4% (Duhok) relative to observed mean peak values — represents a substantial reduction in the maximum annual vegetation productivity that has direct implications for livestock grazing capacity, dryland agricultural yields, and carbon sequestration. The near-stability or slight increase in SEVI (seasonal EVI integral) under SSP2-4.5 in three governorates suggests that moderate warming may extend the shoulder seasons marginally, partially compensating for reduced peak values. Under SSP5-8.5, however, the Halabja SEVI decline from 1.485 to 1.447 indicates net seasonal productivity loss that is unlikely to be offset by shoulder-season extension.

4.9. Limitations and Future Work

Several limitations constrain the present study. First, the training record of approximately 48 monthly sequences per governorate is short by the standards of deep learning, which directly caused the failure of attention-based architectures and limits the diversity of climate conditions seen during training. Extending the EVI record backwards using MODIS (available from 2000) or Landsat (available from 1984) would substantially increase training data volume, potentially enabling attention-based architectures to contribute. Second, the delta-factor forcing cannot represent stochastic precipitation extremes, meaning drought recurrence risk is underestimated relative to studies using downscaled daily GCM output (Masson-Delmotte et al., 2021).

Third, a further limitation concerns spatial aggregation. All analyses were conducted on governorate-level mean EVI, which pools ecologically heterogeneous landscapes into a single time series. Duhok, for example, encompasses both Mesopotamian lowland plains (~300 m a.s.l.) and mountain ranges exceeding 2,000 m, where precipitation regimes, vegetation types, and EVI amplitudes differ substantially. The vulnerability gradient identified in Section 3.5 – attributing stronger climate sensitivity to higher-elevation, wetter governorates – is therefore an inference drawn from inter-governorate differences rather than a within-governorate demonstration. It is possible that the governorate-mean EVI signal in Duhok is dominated by its extensive lowland area and underrepresents the climate sensitivity of its highland vegetation, or conversely that within-Halabja elevation heterogeneity dampens the projected decline signal. The spatial conclusions of this study should therefore be interpreted at the governorate scale only and should not be extrapolated to sub-governorate land management decisions without further disaggregation. Future work should extract EVI by elevation band or land-use class within each governorate to test whether the vulnerability gradient holds at finer spatial scales and to support localised agricultural and pastoral planning applications.

Fourth, no formal minimum sequence threshold has been established for recurrent architectures in this setting, and the implications of the ~48 training sequence constraint are not fully quantified for BiLSTM_GRU specifically. While a dedicated learning curve sensitivity analysis – sub-sampling training sequences at 24, 36, and 48 – is beyond the scope of the current revision, three lines of evidence suggest that 48 sequences is adequate for BiLSTM_GRU projection, if not generalisable to all recurrent architectures. First, the close agreement between CV R^2 (0.730 ± 0.055) and test R^2 (0.734) indicates that the model generalises consistently rather than exhibiting the steep performance degradation characteristic of insufficient training data. Second, the low CV standard deviation (± 0.055) across three temporal folds suggests that performance is stable across different sub-windows of the training record, implying the model is not critically sensitive to which particular 48-sequence subset is presented during training. Third, GRU-based architectures generalise more effectively on shorter sequences than LSTM or attention-based models due to their reduced parameter count and simpler gating mechanism (Hochreiter & Schmidhuber, 1997; Bai et al., 2018), and the BiLSTM_GRU hybrid retains this efficiency in its GRU summarisation step. We acknowledge that these arguments are inferential rather than empirical, and that a formal learning curve analysis remains an important validation step. This analysis is therefore recommended as a priority alongside the MODIS record extension proposed above, which would itself resolve the data-volume constraint by extending the training record to approximately 300 monthly sequences.

Future work should (1) incorporate MODIS MOD13A3 data to extend the training record to 2000–present, enabling a rigorous comparison of attention-based architectures under adequate data volumes; (2) replace delta-factor forcing with stochastic weather generator output conditioned on CMIP6 projected changes in precipitation variability; and (3) explore multi-site joint training across KRI governorates to increase effective sample size while accounting for spatial non-stationarity.

5. Conclusions

This study presented a rigorous deep learning framework for projecting vegetation health — quantified as monthly Enhanced Vegetation Index — under CMIP6 climate scenarios in the Kurdistan Region of Iraq, 2025–2050. Twelve architectures were systematically evaluated across four governorates using a strict temporal data split that isolated the 2022 drought year as a challenging out-of-distribution validation period and a fully held-out 2023–2024 test set. Walk-forward 3-fold cross-validation on training data alone (144 CV fits) was used as the primary model selection criterion, prioritising temporal generalisation over single-window test accuracy. The following conclusions are drawn:

1. BiLSTM_GRU is the optimal architecture for short-record arid-land EVI projection. The hybrid bidirectional LSTM–GRU architecture achieved the highest cross-validation R^2 (0.730 ± 0.055) among stable models and a test R^2 of 0.734, with NSE values of 0.648–0.803 across governorates, demonstrating its suitability as the optimal architecture for short-record arid-land EVI projection.

2. CNN-hybrid architectures overfit on short training records. CNN_BiLSTM_GRU achieved the highest test R^2 (0.764) but collapsed to CV $R^2 = 0.356$, a divergence of 0.408. This result demonstrates that convolutional feature extractors cannot learn generalisable motifs from ~48 training sequences and should be rejected for projection tasks despite competitive held-out test scores. Temporal cross-validation is essential for responsible model selection in ecological forecasting.

3. Attention mechanisms require substantially more data than is available in the KRI record. All three attention-based architectures failed in cross-validation (CNN_BiLSTM_AM: CV $R^2 = -1.675$; CNN_Transformer: CV $R^2 = -0.603$), indicating that attention mechanisms are unsuitable for short-record monthly semi-arid vegetation time series. We estimate a minimum of 10 years of monthly data is required before attention-based architectures can be expected to outperform recurrent baselines in similar settings.

4. KRI vegetation is projected to decline under both CMIP6 scenarios, with a clear north-to-south vulnerability gradient. Under SSP5-8.5, mean annual EVI declines by 2040 relative to baseline range from -3.2% (Duhok) to -8.3% (Halabja), with peak EVI reductions of 14.5%–20.4% by 2050 across all governorates. Erbil, the most arid governorate, shows near-zero climate sensitivity. Peak-of-Season timing remains stable at April under both scenarios, indicating that warming alone does not advance spring green-up in this Mediterranean-type precipitation regime.

5. No tipping point or 2022-level drought recurrence is projected by 2050 under the smooth delta-factor forcing. All drought exceedance probabilities and tipping-point analyses returned null results within the 2025–2050 window. This does not exclude tipping beyond 2050 — particularly for Halabja and Sulaymaniyah where monotonic SSP5-8.5 divergence is accelerating — nor does it capture stochastic drought recurrence, which requires forcing that preserves precipitation extremes rather than ensemble means (Masson-Delmotte et al., 2021).

6. The delta-factor CMIP6 forcing preserves seasonal structure but cannot represent precipitation extremes. The application of IPCC AR6 WGI Chapter 10 MENA domain delta factors (SSP2-4.5: +1.5 °C, -8% precipitation; SSP5-8.5: +2.8 °C, -15%) as linear ramps to TerraClimate monthly climatology is a scientifically valid approach for trend detection but should be supplemented by stochastic extreme-event modelling in future work.

Taken together, these findings establish BiLSTM_GRU as the recommended architecture for monthly EVI projection in data-limited semi-arid regions, provide a methodologically grounded warning against using held-out test accuracy alone for model selection in ecological forecasting, and deliver the first governorate-level CMIP6-conditioned EVI projections for the Kurdistan Region of Iraq. The projected vegetation decline, particularly in Halabja and Sulaymaniyah, has direct implications for regional food security, pastoral livelihoods, and land management policy under the high-emission pathway.

Author: Contributions A.R. is the sole author and is responsible for all aspects of this work, including conceptualization, methodology, software, validation, formal analysis, investigation, data curation, writing, visualization, and project administration.

Funding: This research received no external funding.

Acknowledgments: The author gratefully acknowledges the European Space Agency (ESA) for providing free access to Sentinel-2 Surface Reflectance imagery through the Copernicus programme, Google for access to the Google Earth Engine cloud computing platform, and the TerraClimate dataset providers (University of Idaho) for making high-resolution global climate data publicly available.

Data: Availability Statement The data supporting the findings of this study are available upon request from the corresponding author.

Conflicts: of Interest The author declares no conflict of interest.

References

- Abatzoglou, J. T., Dobrowski, S. Z., Parks, S. A., & Hegewisch, K. C. (2018). TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Scientific Data*, 5(1), 170191.
- Bai, S., Kolter, J. Z., & Koltun, V. (2018). *An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling* (arXiv:1803.01271). arXiv. <https://doi.org/10.48550/arXiv.1803.01271>
- Doblas-Reyes, F. J., Sorensson, A. A., Almazroui, M., Dosio, A., Gutowski, W. J., Haarsma, R., Hamdi, R., Hewitson, B., Kwon, W.-T., & Lamptey, B. L. (2021). *Linking global to regional climate change*. <https://centaur.reading.ac.uk/99896/>
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780.
- Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., & Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, 83(1–2), 195–213.
- Kattenborn, T., Leitloff, J., Schiefer, F., & Hinz, S. (2021). Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 173, 24–49.
- Liu, B., Scherer, L., Van Bodegom, P. M., Sun, Z., & Behrens, P. (2025). Exploring the Spatial Relationship Between Carbon Storage and Biodiversity: A Systematic Review and Meta-Analysis. *Land Degradation & Development*, 36(8), 2786–2797. <https://doi.org/10.1002/ldr.5531>
- Lusk, D., Wolf, S., Svidzinska, D., Dormann, C. F., Kattge, J., Bruehlheide, H., Sabatini, F. M., Damasceno, G., Moreno Martínez, Á., & Violle, C. (2026). Crowdsourced biodiversity monitoring fills gaps in global plant trait mapping. *Nature Communications*. <https://www.nature.com/articles/s41467-026-68996-y>
- Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., & Gomis, M. I. (2021). Climate change 2021: The physical science basis. *Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, 2(1), 2391.
- Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, 50(3), 885–900.
- Ndikumana, E., Ho Tong Minh, D., Baghdadi, N., Courault, D., & Hossard, L. (2018). Deep recurrent neural network for agricultural classification using multitemporal SAR Sentinel-1 for Camargue, France. *Remote Sensing*, 10(8), 1217.
- Peng, S., Piao, S., Ciais, P., Myneni, R. B., Chen, A., Chevallier, F., Dolman, A. J., Janssens, I. A., Penuelas, J., & Zhang, G. (2013). Asymmetric effects of daytime and night-time warming on Northern Hemisphere vegetation. *Nature*, 501(7465), 88–92.
- Rasul, A. (2026). Deep Learning-Based EVI Forecasting for Vegetation Health Monitoring in the Kurdistan Region of Iraq (2016–2024). Preprints.org. <https://www.preprints.org/manuscript/202603.0411>
- Rasul, A., & Zahir, I. S. (2025). *Assessing Trends and Drivers of Burned Areas in Forest Areas in Kurdistan Region*. <https://www.preprints.org/manuscript/202512.0271>

- Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat, F. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), 195–204.
- Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., & Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, *40*(8), 913–929. <https://doi.org/10.1111/ecog.02881>
- Ronald, E., Sangermano, F., Ghimire, B., Zhu, H., Chen, H., Neeti, N., Cai, Y., Machado, E. A., & Crema, S. C. (2009). Seasonal trend analysis of image time series. *International Journal of Remote Sensing*, *30*(10), 2721–2726. <https://doi.org/10.1080/01431160902755338>
- Running, S. W., Nemani, R. R., Heinsch, F. A., Zhao, M., Reeves, M., & Hashimoto, H. (2004). A continuous satellite-derived measure of global terrestrial primary production. *Bioscience*, *54*(6), 547–560.
- Shen, H., Li, X., Cheng, Q., Zeng, C., Yang, G., Li, H., & Zhang, L. (2015). Missing information reconstruction of remote sensing data: A technical review. *IEEE Geoscience and Remote Sensing Magazine*, *3*(3), 61–85.
- Shen, R., Huang, A., Li, B., & Guo, J. (2019). Construction of a drought monitoring model using deep learning based on multi-source remote sensing data. *International Journal of Applied Earth Observation and Geoinformation*, *79*, 48–57.
- Sun, Y., Lao, D., Ruan, Y., Huang, C., & Xin, Q. (2023). A deep learning-based approach to predict large-scale dynamics of normalized difference vegetation index for the monitoring of vegetation activities and stresses using meteorological data. *Sustainability*, *15*(8), 6632.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, *30*. <https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html>
- Wang, Y., Zhang, N., Chen, M., Zhao, Y., Guo, F., Huang, J., Peng, D., & Wang, X. (2025). Prediction and Spatiotemporal Dynamics of Vegetation Index Based on Deep Learning and Environmental Factors in the Yangtze River Basin. *Forests*, *16*(3). <https://doi.org/10.3390/f16030460>
- Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Tan, W., Yang, Q., & Wang, J. (2020). Deep learning in environmental remote sensing: Achievements and challenges. *Remote Sensing of Environment*, *241*, 111716.
- Zhong, L., Hu, L., & Zhou, H. (2019). Deep learning based multi-temporal crop classification. *Remote Sensing of Environment*, *221*, 430–443.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.