Article

Not peer-reviewed version

Deep Learning-Based EVI Forecasting for Vegetation Health Monitoring in the Kurdistan Region of Iraq (2016–2024)

Azad Rasul *

Posted Date: 5 March 2026

doi: 10.20944/preprints202603.0411.v1

Keywords: enhanced vegetation index; deep learning; BiLSTM; vegetation forecasting; remote sensing; semi-arid

Article

Deep Learning-Based EVI Forecasting for Vegetation Health Monitoring in the Kurdistan Region of Iraq (2016–2024)

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 health is essential for natural resource management, food security planning, and climate adaptation in water-stressed semi-arid environments. This study presents a comprehensive deep learning framework for forecasting the Enhanced Vegetation Index (EVI) across the four governorates of the Kurdistan Region of Iraq (KRI) --- Erbil, Duhok, Sulaymaniyah, and Halabja --- using a nine-year monthly record (January 2016 -- December 2024) derived from Sentinel-2 Level-2A Surface Reflectance imagery accessed via Google Earth Engine (GEE). Nine deep learning architectures spanning recurrent, hybrid convolutional-recurrent, and attention-based categories were trained and evaluated on a multivariate feature set comprising EVI, precipitation, air temperature, and cyclic month encoding. The Bidirectional Long Short-Term Memory (BiLSTM) model achieved the highest mean R^2 of 0.945 across all four governorates, with outstanding performance in Sulaymaniyah ($R^2 = 0.977$) and Halabja ($R^2 = 0.964$). Hybrid CNN-recurrent architectures, particularly CNN-BiLSTM-GRU, also demonstrated strong performance with the highest mean tolerance accuracy (0.985), confirming the complementarity of local convolutional feature extraction and temporal sequence modeling; however, BiLSTM remains the top-ranked model by R^2. By contrast, the standalone Transformer model performed poorly (mean $R^2 = 0.132$) due to the absence of positional encoding in the shallow single-block architecture. Predictive uncertainty was quantified using Monte Carlo Dropout inference, revealing well-calibrated epistemic uncertainty that peaks during the spring vegetation growing season. Autoregressive five-year EVI forecasts (2026--2030) and an exploratory ten-year projection (2026--2035) were generated by the BiLSTM model; forecasts commence in January 2026 as TerraClimate climate forcing data for 2025 were not yet publicly available at the time of analysis. Projected mean annual EVI values range from 0.145 to 0.194 across governorates, consistent with the historical climatological baseline. The 2022 regional drought anomaly is clearly captured in the historical record, confirming the sensitivity of the EVI signal to precipitation deficits. These results establish deep learning-based EVI forecasting as a viable and scalable tool for operational vegetation health monitoring in the KRI and comparable semi-arid dryland systems.

Keywords: enhanced vegetation index; deep learning; BiLSTM; vegetation forecasting; remote sensing; semi-arid

---

1. Introduction

Vegetation constitutes a fundamental component of terrestrial ecosystems, mediating carbon sequestration, regulating surface energy and water balances, and sustaining agricultural productivity (Huete et al., 2002; Liu & Huete, 1995). Remote sensing-derived vegetation indices provide repeatable, spatially continuous measures of vegetation condition over time, enabling the detection of seasonal phenological cycles, interannual variability, and long-term trends in greenness and biomass. Among these indices, the Enhanced Vegetation Index (EVI) offers important advantages over the Normalized Difference Vegetation Index (NDVI) in heterogeneous, low-to-moderate cover landscapes, as it decouples canopy background reflectance and corrects for residual atmospheric aerosol contamination through the inclusion of the blue band (Amponsem, 2025; Bhosle & Deshmukh, 2025). These properties make EVI particularly well-suited to monitoring vegetation dynamics in semi-arid regions where sparse and patchy cover produces strong soil background interference.

The Kurdistan Region of Iraq (KRI) is a federal region situated in the Zagros Mountain foothills and adjacent semi-arid plains of northeastern Iraq. The region's vegetation is strongly constrained by a Mediterranean-type seasonal precipitation regime, in which nearly all annual rainfall is concentrated between November and April, driving a pronounced spring greenup followed by a prolonged dry-season browning. Documented trends of increasing drought severity and frequency in the KRI (Gaznayee et al., 2022; Hamarash et al., 2024) pose significant threats to rain-fed agriculture, rangeland condition, and water availability, elevating the need for operational vegetation monitoring tools capable of providing multi-year forecasts to inform land and water management decisions.

The advent of the Sentinel-2 multispectral satellite constellation, which provides 10–20 m spatial resolution imagery with an approximately 3–5 day revisit time, has substantially increased the spatiotemporal density of available vegetation time series data (Gascon et al., 2017)(European Space Agency, 2017). Combined with the GEE cloud computing platform, Sentinel-2 enables systematic, automated extraction of monthly EVI time series at governorate scale without the computational bottlenecks historically associated with large satellite data archives.

Deep learning sequence models, including Long Short-Term Memory (LSTM) (Hochreiter & Schmidhuber, 1997) networks and their variants, have demonstrated strong performance in vegetation index forecasting tasks. Bidirectional LSTM architectures exploit both forward and backward temporal context within the input sequence, capturing asymmetric seasonal dynamics that unidirectional models may miss. Hybrid Convolutional-Recurrent architectures further enhance performance by prepending one-dimensional convolutional layers that extract local temporal features before passing the representation to recurrent decoders. Recent comparative studies in similar climate settings have confirmed the superiority of BiLSTM and CNN-BiLSTM-based models for vegetation index time series prediction (Wang et al., 2025; Zhang et al., 2022). The use of multivariate inputs — pairing EVI with precipitation and temperature reanalysis products — has been shown to improve model sensitivity to vegetation–climate coupling, particularly in dryland systems where moisture availability is the primary growth-limiting factor (Abatzoglou et al., 2018).

Despite this growing body of work, no systematic benchmarking of deep learning architectures for EVI forecasting has been conducted across the multiple governorates of the KRI. Moreover, the inclusion of attention-based architectures, including the Temporal Convolutional Network (TCN) and Transformer models, alongside established recurrent and hybrid models, provides a rigorous and comprehensive comparison that extends beyond prior regional vegetation forecasting studies. The integration of Monte Carlo Dropout uncertainty quantification (Gal & Ghahramani, 2016) further distinguishes this study by providing principled probabilistic forecast intervals rather than deterministic point predictions, which is critical for risk-informed land management applications.

This study therefore addresses the following research objectives: (i) to derive and quality-control a nine-year monthly EVI time series for the four KRI governorates using Sentinel-2 and GEE; (ii) to benchmark nine deep learning architectures for monthly EVI forecasting using a multivariate feature set including EVI, precipitation, temperature, and cyclic seasonality encoding; (iii) to quantify predictive uncertainty via Monte Carlo Dropout inference; and (iv) to generate five-year and ten-year autoregressive EVI forecasts for all four governorates using the best-performing model. The remainder of the paper is structured as follows: Section 2 describes the study area, data acquisition, preprocessing, model architectures, training protocol, and evaluation metrics; Section 3 presents historical time series analysis, test-set model performance, uncertainty characterisation, and future

EVI forecasts; Section 4 discusses the key findings in the context of prior literature and methodological limitations; and Section 5 provides concluding remarks and recommendations for future research.

2. Methodology

2.1. Overview of Research Design

This study develops and evaluates a suite of deep learning architectures for forecasting the EVI across the four official governorates of the Kurdistan Region of Iraq — Erbil, Duhok, Sulaymaniyah, and Halabja — over the period 2016–2024. The research pipeline integrates multi-source remote sensing and climate reanalysis data with state-of-the-art sequence modeling to produce both short-term (five-year) and exploratory long-term (ten-year) EVI forecasts. The methodology proceeds through six sequential phases: (1) study area delineation, (2) multi-source data acquisition and preprocessing, (3) feature engineering and normalization, (4) deep learning model design and training, (5) model evaluation and uncertainty quantification, and (6) future EVI forecasting.

2.2. Study Area

The study area encompasses the Kurdistan Region of Iraq, a federal region located in the northeastern part of Iraq, bounded approximately between 34.38°–37.38°N latitude and 42.26°–46.34°E longitude. The region is characterized by heterogeneous terrain, including the Zagros Mountain foothills, river valleys, and semi-arid plains, which produce pronounced spatial variability in vegetation cover and seasonal phenological dynamics (Forti et al., 2021; Noroozi, 2020). Four administrative governorates were selected as the analytical units of investigation: Erbil, Duhok, Sulaymaniyah, and Halabja (Figure 2.1).

The region experiences a continental Mediterranean climate characterized by cold, rainy winters and hot, arid summers. Average yearly rainfall (recorded between 2012 and 2024) stands at 548.7 mm, with more than nine-tenths of that total occurring from November through April. Annual average temperatures hover around 20.8 °C in Erbil and roughly 17 °C at higher elevations. Summer temperatures frequently climb beyond 44 °C, while winter temperatures can occasionally dip to just under –2 °C (Rasul & Zahir, 2025).

Study area boundaries were defined using official administrative shapefiles for each governorate. All geometries were reprojected to the WGS-84 geographic coordinate reference system (EPSG:4326). Multi-part polygon geometries were dissolved to produce a single representative boundary per governorate. To avoid geometry complexity errors in the GEE API, each dissolved polygon was simplified with a tolerance of 0.001 decimal degrees (ee.Geometry.simplify). The resulting polygon extents were converted to GEE geometry objects for use in satellite data extraction queries.

2.3. Data Acquisition

2.3.1. Enhanced Vegetation Index from Sentinel-2

Monthly EVI time series were derived from the Sentinel-2 Level-2A Surface Reflectance Harmonized image collection (ESA, 2017) accessed via the GEE cloud computing platform(Gorelick et al., 2017). The harmonized collection corrects for inter-mission differences in sensor spectral response, ensuring temporal consistency across the study period. Sentinel-2 provides multispectral imagery at a 10–20 m spatial resolution with a revisit time of approximately 3-5 days, enabling high-density temporal sampling. The study period spans January 2016 to December 2024, yielding a 108-month record per governorate. A preliminary cloud screening filter was applied retaining only images with a CLOUDY_PIXEL_PERCENTAGE below 80%, after which pixel-level SCL masking was performed.

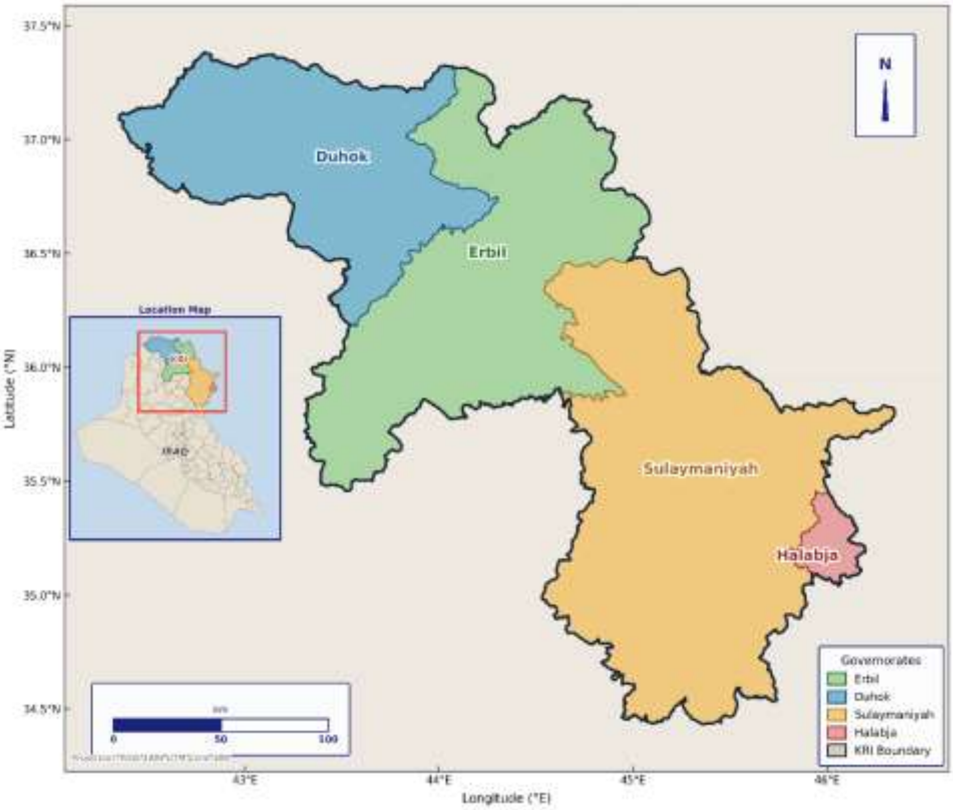


Figure 2.1. Study Area: Kurdistan Region of Iraq.

Cloud and shadow contamination was mitigated using the Scene Classification Layer (SCL), retaining only pixels classified as vegetation (SCL = 4) and bare soil (SCL = 5). Digital number values were converted to surface reflectance by dividing by 10,000. EVI was computed from the cloud-masked, atmospherically corrected imagery using the standard three-band formulation. A pixel-level GEE mask was applied immediately after EVI computation to remove physically unrealistic values outside the range [−0.2, 0.9], guarding against sensor artefacts prior to any regional aggregation. Spatial averaging was performed using a reduction scale of 100 m. EVI was calculated using Equation 2.1 (Huete et al., 2002).

$$\mathbf{EVI = 2.5 \times \frac{\mathit{NIR} - \mathit{Red}}{\mathit{NIR} + 6 \times \mathit{Red} - 7.5 \times \mathit{Blue} + 1}} \quad \mathbf{(2.1)}$$

where NIR, Red, and Blue correspond to Sentinel-2 Bands 8, 4, and 2, respectively. For each calendar month, the median composite across all available cloud-free observations within each governorate boundary was computed, and summary statistics including mean, standard deviation, minimum, maximum, and valid pixel count were recorded.

2.3.2. Climate Covariates

Two climate variables were incorporated as auxiliary predictors to improve model sensitivity to vegetation-driving atmospheric conditions. Monthly total precipitation (mm) was extracted from the TerraClimate dataset (Abatzoglou et al., 2018), a global gridded climate dataset derived by downscaling reanalysis products to approximately 4 km spatial resolution, using a reduction scale of 4000 m. Mean monthly 2 m air temperature (°C) was similarly sourced from TerraClimate, computed as the average of the monthly maximum (tmmx) and minimum (tmmn) temperature bands. Both bands are stored in TerraClimate as tenths of degrees Celsius and were scaled by 0.1 prior to averaging. Both variables were extracted for the full 2016–2024 study period and spatially averaged over each governorate's boundary polygon using GEE's regional reduction functionality.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 March 2026 doi:10.20944/preprints202603.0411.v1

2.4. Data Preprocessing and Quality Control

2.4.1. Quality Control and Gap Filling

Raw EVI retrievals were subjected to a multi-step quality control (QC) procedure. First, only Sentinel-2 pixels classified as vegetation (SCL class 4) or bare soil (SCL class 5) were retained, excluding water, built-up areas, and unclassified surfaces. Monthly observations derived from fewer than 50 valid cloud-free pixels at 100 m aggregation scale — equivalent to approximately 50 ha of vegetated or bare land — were flagged as missing and subsequently gap-filled using seasonal climatology, as such estimates are insufficiently representative of the governorate-scale spatial average. Second, physically implausible EVI values were removed by applying a range filter, retaining only values within the interval [−0.10, 0.80]; observations outside this range are indicative of residual cloud contamination, surface water, or data artifacts. Third, remaining missing values were imputed through a two-stage gap-filling strategy: (i) primary gap-filling using the climatological monthly mean computed across all valid years for the corresponding month of year, and (ii) residual gaps were subsequently filled by linear temporal interpolation (Figure 2.2). Missing precipitation and temperature values were similarly interpolated using linear methods.

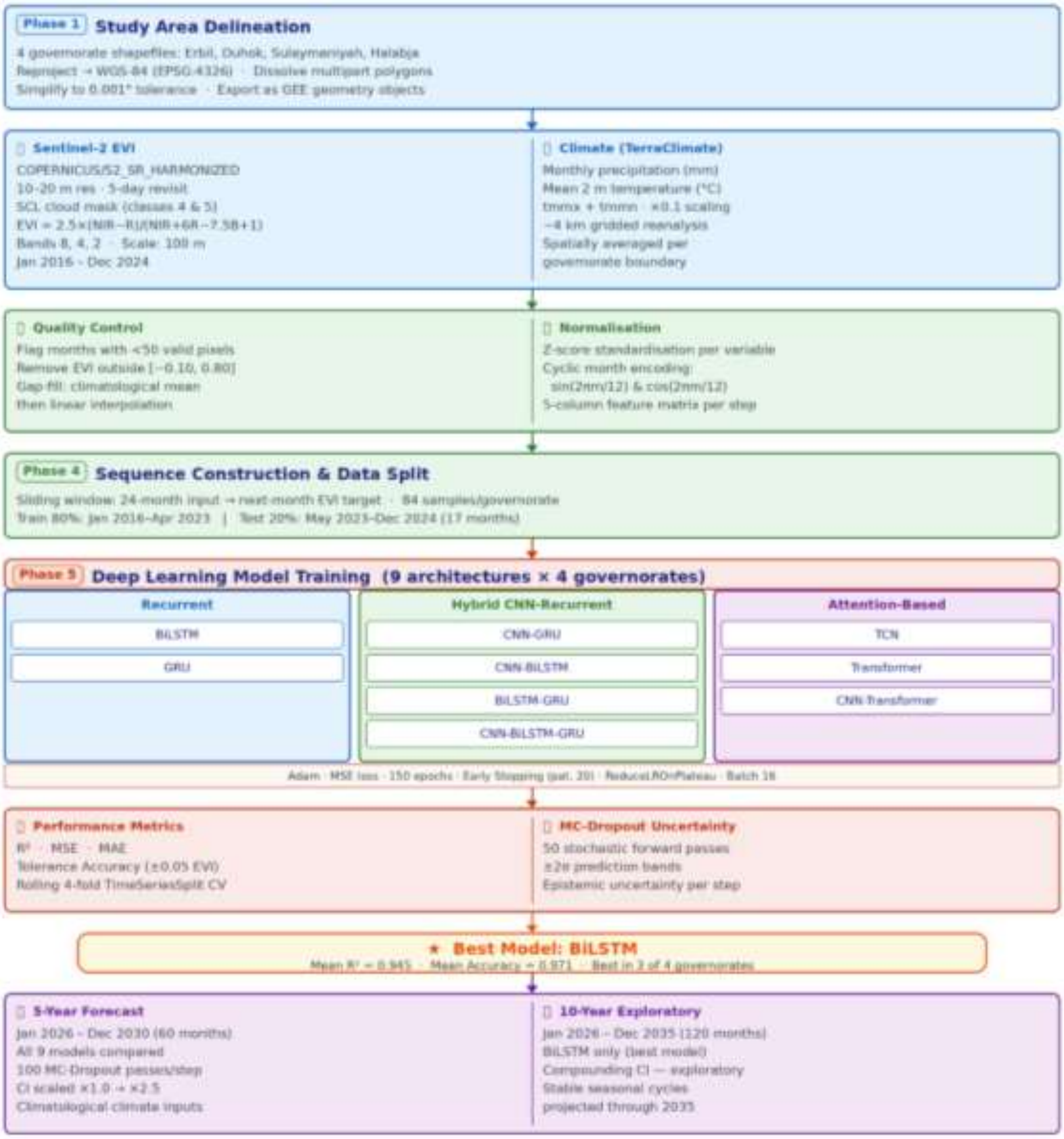


Figure 2.2. Flowchart of the research methodology.

2.4.2. Normalization and Cyclic Temporal Encoding

All three input variables (EVI, precipitation, and temperature) were independently normalized per governorate using z-score standardization (zero mean, unit variance), parameterized on the full 2016–2024 training period. Separate StandardScaler instances were retained for each variable and region to enable inverse transformation of model predictions back to physically interpretable units.

Seasonality was explicitly encoded as a continuous cyclic feature rather than a categorical month indicator, preventing artificial discontinuities at year boundaries. The month of year (m = 1, ..., 12) was transformed to two trigonometric components (Equation 2.2):

$$month\_\sin = \sin\left(\frac{2\pi m}{12}\right), month\_\cos = \cos\left(\frac{2\pi m}{12}\right) \quad (2.2)$$

The resulting feature matrix for each governorate therefore contains five columns per time step: normalized EVI, normalized precipitation, normalized temperature, month_sin, and month_cos.

2.5. Sequence Construction and Data Partitioning

Supervised learning samples were constructed using a sliding-window approach. For each governorate, overlapping windows of 24 consecutive monthly feature vectors were extracted from the normalized five-dimensional time series. Each window serves as the model input, with the normalized EVI value of the immediately following month as the regression target. This formulation encodes two full annual cycles of lagged multi-variate information, capturing interannual variability and seasonal phase relationships. With a 108-month record and a window length of 24, this yields 84 samples per governorate.

The dataset was partitioned chronologically into training (80%) and test (20%) subsets, corresponding approximately to January 2016 – April 2023 and May 2023 – December 2024, respectively. Temporal ordering was strictly preserved; no shuffling was applied prior to splitting. TensorFlow tf.data.Dataset pipelines were constructed for training and evaluation, with a batch size of 16.

2.6. Deep Learning Model Architectures

Nine deep learning architectures were implemented and benchmarked, organized into three categories: recurrent models, hybrid convolutional-recurrent models, and attention-based models. All models accept multivariate input tensors of shape (sequence_length, 5) and produce a single scalar EVI prediction. Dropout layers with a rate of 0.3 were integrated throughout all architectures to enable Monte Carlo Dropout uncertainty estimation. The model suite is summarized in Table 2.1.

Table 2.1 — Deep Learning Model Architectures.

Category	Model	Key Components	Seq. Length
Recurrent	BiLSTM	Bidirectional LSTM (2 layers) + GlobalAvgPool	24
Recurrent	GRU	Stacked GRU (2 layers) + GlobalAvgPool	24
Hybrid CNN-Recurrent	CNN-GRU	1D Conv feature extractor + stacked GRU	24

Category	Model	Key Components	Seq. Length
Hybrid CNN-Recurrent	CNN-BiLSTM	1D Conv + Bidirectional LSTM	24
Hybrid CNN-Recurrent	BiLSTM-GRU	BiLSTM followed by GRU decoder	24
Hybrid CNN-Recurrent	CNN-BiLSTM-GRU	Conv + BiLSTM + GRU cascade	24
Attention-Based	TCN	Dilated causal Conv1D blocks (rates 1,2,4,8) + GlobalAvgPool	24
Attention-Based	Transformer	Multi-head self-attention + feed-forward layers	24
Attention-Based	CNN-Transformer	1D Conv tokenizer + Transformer encoder	24

2.6.1. Recurrent Architectures

The Bidirectional LSTM (BiLSTM) (Schuster & Paliwal, 1997) employs three stacked Bidirectional LSTM layers with 64, 32, and 16 units, respectively. The first two layers use return_sequences=True, and the final layer collapses the temporal dimension by returning only the last hidden state, which feeds directly into a Dense(1) output layer with no global pooling. Bidirectional processing allows the model to capture both forward and backward temporal dependencies within the input window. The Gated Recurrent Unit (GRU) (Chung et al., 2014) network uses two stacked GRU layers, both with 64 units, offering a computationally efficient alternative to LSTM with a reduced parameter count.

2.6.2. Hybrid Convolutional-Recurrent Architectures

Hybrid architectures combine 1D convolutional layers for local feature extraction with recurrent layers for temporal sequence modeling. The CNN-GRU and CNN-BiLSTM models prepend a single Conv1D layer (64 filters, kernel size 3, ReLU activation, same padding) followed by Batch Normalization to GRU and BiLSTM recurrent decoders, respectively. No max-pooling is applied, preserving the full temporal resolution for the recurrent stage. The BiLSTM-GRU model employs a BiLSTM encoder followed by a GRU decoder to capture bidirectional long-range patterns before autoregressive decoding. The CNN-BiLSTM-GRU model combines all three components in a cascade, offering the greatest representational depth among the hybrid architectures.

2.6.3. Attention-Based Architectures

Three architecture models were introduced to evaluate the utility of attention mechanisms and temporal convolutional approaches for vegetation time series forecasting. The Temporal Convolutional Network (TCN) (Bai et al., 2018) employs four sequential dilated causal Conv1D layers with dilation rates of 1, 2, 4, and 8, each using 64 filters and ReLU activation, followed by Dropout and GlobalAveragePooling1D. The dilation rates increase exponentially to expand the receptive field without recurrence. No residual connections are implemented in the current architecture. The TCN uses the same sequence length of 24 months as all other models.

 doi:10.20944/preprints202603.0411.v1

The Transformer model (Vaswani et al., 2017) projects the five-dimensional input to a d_model=64 dimensional space via a Dense layer, then applies one multi-head self-attention block (four heads) followed by a feed-forward sublayer (Dense(128) → Dense(64)), each with Layer Normalization and residual addition. No explicit positional encoding is added; temporal ordering is implicit in the input sequence. The single encoder block is followed by Dropout and GlobalAveragePooling1D before the Dense(1) output. The CNN-Transformer hybrid prepends a Conv1D layer and Batch Normalization to tokenize the input before passing it to the same Transformer encoder, combining local feature learning with global sequence modeling.

2.7. Model Training

All models were trained using the Adam optimizer with mean squared error (MSE) as the loss function. Learning rates were assigned per model: BiLSTM and GRU used 5×10^{-4}; CNN-GRU and TCN used 1×10^{-3}; CNN-BiLSTM, BiLSTM-GRU, CNN-BiLSTM-GRU, Transformer, and CNN-Transformer all used 5×10^{-4}. Training was conducted for a maximum of 150 epochs with a batch size of 16. Two callbacks were employed: (1) Early Stopping (patience = 20 epochs, monitoring validation loss, restoring best weights) to prevent overfitting; and (2) ReduceLROnPlateau (factor = 0.5, patience = 10 epochs, minimum lr = 1×10^{-6}) to adaptively decay the learning rate on validation loss plateaus.

Each of the nine primary models plus two optional baseline models (LSTM, CNN) was trained independently for each of the four governorates, yielding a total of 44 trained model instances (11 × 4). Results reported in the evaluation section focus on the nine primary models (36 instances). All experiments were conducted using TensorFlow 2.x with GPU acceleration where available.

2.8. Model Evaluation

2.8.1. Performance Metrics

Model performance on the held-out test set was assessed using four complementary metrics. MSE and Mean Absolute Error (MAE) quantify prediction error magnitude. The coefficient of determination (R^2) measures the proportion of variance in the observed EVI time series explained by the model, with values approaching unity indicating superior fit. Accuracy is defined as the proportion of test samples for which the absolute prediction error falls at or below 0.05 EVI units in the original (inverse-transformed) EVI space, providing an intuitive performance summary calibrated to the physical range of the index.

2.8.2. Temporal Cross-Validation

To assess the generalizability of the trained models beyond a single train–test split, a rolling time-series cross-validation was conducted using scikit-learn's TimeSeriesSplit with four folds. In each fold, the model was retrained from scratch on the available training window and evaluated on the immediately succeeding validation window. This walk-forward procedure respects the temporal ordering of the data and provides a more robust estimate of out-of-sample predictive performance.

2.8.3. Monte Carlo Dropout Uncertainty Estimation

Predictive uncertainty was quantified using Monte Carlo (MC) Dropout inference (Gal & Ghahramani, 2016). At test time, dropout layers were retained in their active (stochastic) state, and 50 independent forward passes were executed for each test sample. The predictive mean and standard deviation across the stochastic ensemble were computed, with uncertainty bands displayed at $\pm2\sigma$. This approach provides a principled approximation to Bayesian posterior predictive uncertainty without requiring explicit Bayesian inference.

 doi:10.20944/preprints202603.0411.v1

2.9. Future EVI Forecasting

Future EVI projections were generated using an autoregressive MC-Dropout forecasting procedure for two temporal horizons: a primary five-year forecast (January 2026 – December 2030, 60 months) and an exploratory ten-year forecast (January 2026 – December 2035, 120 months). All nine primary models were used to produce five-year forecasts, enabling cross-model comparison of future trajectories. The ten-year forecast was generated exclusively with the best-performing model as identified through the test-set evaluation. The forecast horizon commences in January 2026 rather than January 2025, as TerraClimate climate forcing data for 2025 were not yet publicly available at the time of analysis; the 2025 period is therefore excluded from both the historical record and the forecast initialisation window.

At each future time step, the model received a sliding window of 24 consecutive feature vectors, where past EVI values were replaced by the model's own predicted values once the historical record was exhausted. Future climate inputs (precipitation and temperature) were estimated from the historical seasonal climatology, calculated as the mean value for each month of year across the 2016–2024 record. Predictive uncertainty was estimated via 100 MC-Dropout forward passes at each forecast step (compared to 50 passes used for test-set uncertainty visualization). To account for the compounding error inherent in autoregressive forecasting, uncertainty bands were linearly widened by a scaling factor that increases from 1.0 at the first forecast step to 2.5 at the final forecast step.

2.10. Software and Computational Environment

All analyses were implemented in Python 3.12 within a Google Colab environment with GPU acceleration. The core software stack comprised: Google Earth Engine Python API for satellite and climate data extraction; TensorFlow 2.19 and Keras for deep learning model construction and training; NumPy and Pandas for numerical computing and data management; GeoPandas and Rasterio for geospatial data processing; and Matplotlib and Seaborn for visualization. All processed datasets and model evaluation metrics were exported to CSV format for archival and reproducibility.

3. Results

3.1. Historical EVI and Climate Time Series (2016–2024)

Figure 3.1 presents the monthly EVI, precipitation, and mean air temperature for all four Kurdistan Region governorates over the 2016–2024 study period. All governorates exhibit a pronounced and consistent seasonal cycle: EVI rises sharply from the dry summer baseline (approximately 0.10–0.16) to a spring peak (April–May), then declines through the dry summer period. This phenological pattern closely mirrors the precipitation seasonality, in which the majority of annual rainfall is concentrated in the November–April rainy season.

Among the four governorates, Halabja consistently records the highest peak EVI values, reaching up to 0.46 in wet years, followed by Duhok (up to 0.39), Erbil (up to 0.40), and Sulaymaniyah (up to 0.37). The synchrony between precipitation and EVI dynamics is visually apparent across all sites, while temperature shows an inverse relationship with vegetation greenness, peaking at 28–35 °C during summer months when EVI is at its minimum.

Inter-annual variability is substantial: the 2019 and 2020 seasons stand out as notably wet and green, with mean annual EVI in Sulaymaniyah reaching approximately 0.20 in 2019 (Figure 3.2). By contrast, 2022 marks the most pronounced dry anomaly in Erbil and Sulaymaniyah, with annual mean EVI falling to 0.13–0.18 across all governorates. The precipitation panel confirms that 2022 was characterised by severely suppressed winter rainfall. These inter-annual patterns confirm that the multivariate feature set (EVI + precipitation + temperature + cyclic month encoding) captures the dominant vegetation–climate coupling that the forecasting models are trained to exploit.

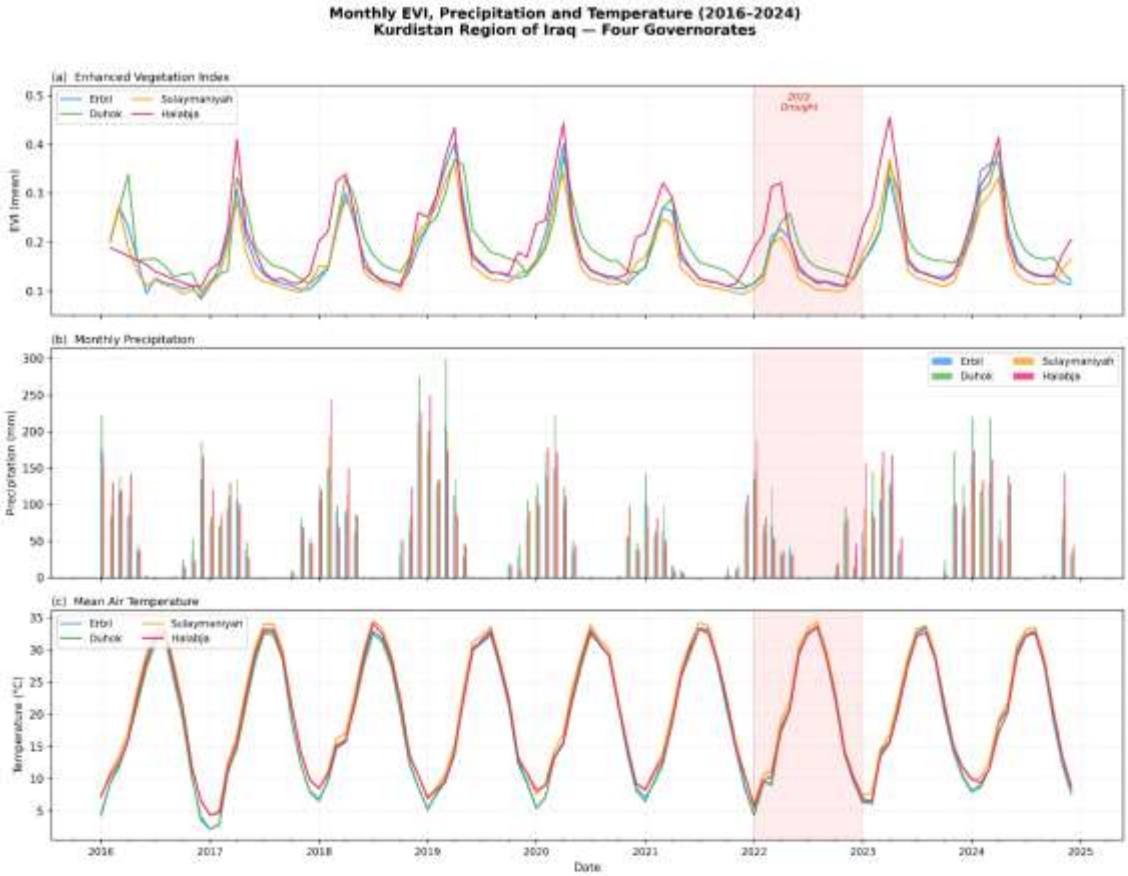


Figure 3.1. Monthly EVI (top), total precipitation in mm (middle), and mean air temperature in °C (bottom) for the four Kurdistan Region governorates, 2016–2024. Precipitation labels use TerraClimate source notation; temperature uses TerraClimate notation as displayed by the plotting code.

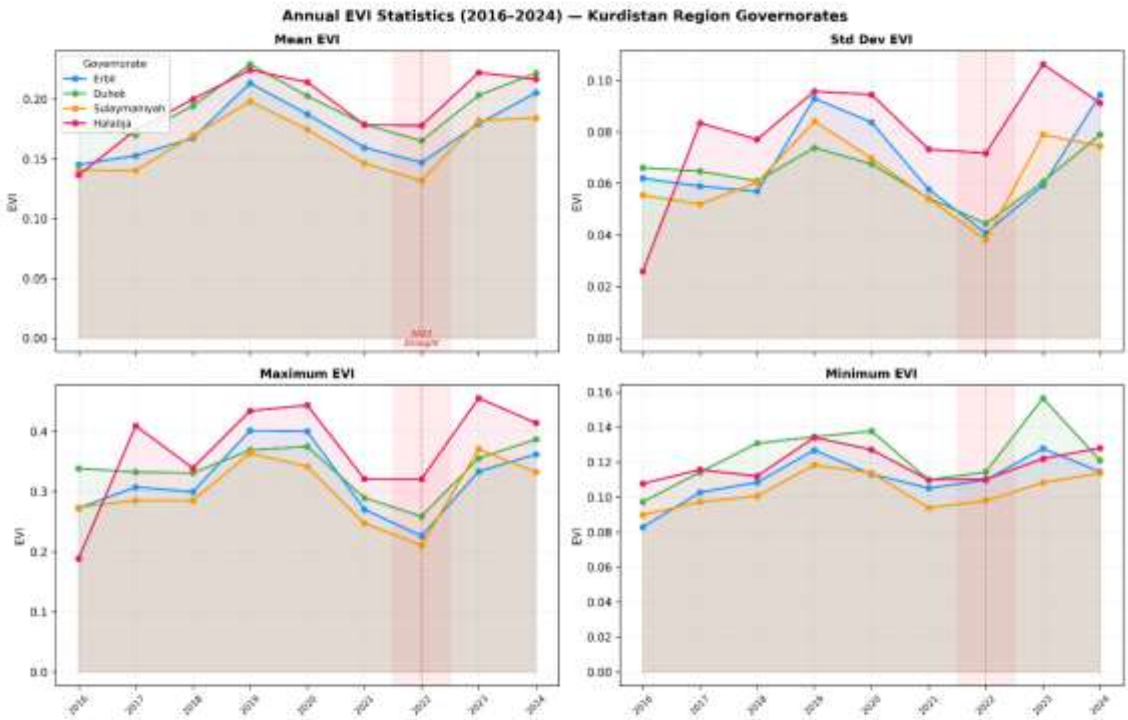


Figure 3.2. Annual EVI statistics (mean, standard deviation, maximum, minimum) for the four governorates across 2016–2024. The prominent trough in 2022 corresponds to a region-wide drought episode.

3.2. Model Performance on the Test Set

All nine primary deep learning models were evaluated on the held-out test set (May 2023 – December 2024, 20 months) across four governorates, yielding 36 individual model–region evaluations. Performance was assessed using the R^2, MSE, MAE, and tolerance accuracy (proportion of predictions within ±0.05 EVI units of the ground truth). Table 3.1 reports mean metrics averaged across the four governorates for each primary model, ranked by R^2. Table 3.2 identifies the best-performing model per governorate.

Table 3.1. Mean test-set performance metrics for primary models averaged across four governorates (ranked by R^2). Accuracy = proportion of predictions within ±0.05 EVI units.

Model	R^2 (mean)	MSE (mean)	MAE (mean)	Accuracy (mean)
BiLSTM	0.9445	0.000376	0.0132	0.971
CNN_BiLSTM_GRU	0.9405	0.000380	0.0144	0.985
CNN_BiLSTM	0.9398	0.000399	0.0161	0.971
CNN_GRU	0.9387	0.000391	0.0153	0.971
BiLSTM_GRU	0.9300	0.000470	0.0168	0.971
GRU	0.9095	0.000608	0.0194	0.926
TCN	0.8947	0.000717	0.0200	0.926
CNN_Transformer	0.7018	0.001694	0.0277	0.853
Transformer	0.1317	0.005771	0.0592	0.515

Table 3.2. Best-performing model per governorate on the test set (by R^2).

Governorate	Best Model	R^2	MSE	MAE	Accuracy
Erbil	CNN_GRU	0.958	0.000319	0.0131	1.000
Duhok	BiLSTM_GRU	0.937	0.000330	0.0144	1.000
Sulaymaniyah	BiLSTM	0.977	0.000120	0.0082	1.000
Halabja	BiLSTM	0.964	0.000300	0.0142	1.000

3.2.1. Top-Performing Models: BiLSTM and Hybrid CNN-Recurrent Family

The BiLSTM model achieves the highest mean R^2 (0.945) across governorates, driven by outstanding performance in Sulaymaniyah (R^2=0.977) and Halabja (R^2=0.964). The CNN_BiLSTM_GRU hybrid follows closely (mean R^2=0.940, mean accuracy=0.985), recording the highest single tolerance accuracy of any model (Table 3.2). CNN_BiLSTM and CNN_GRU also exceed R^2=0.93 on average, confirming that the combination of local convolutional feature extraction with bidirectional or gated recurrent processing is particularly well suited to the structured seasonality of Kurdistan EVI time series.

Figure 3.3 presents the predicted versus ground truth trajectories for BiLSTM across all four governorates. The model closely tracks both the spring greenup amplitude and the summer drawdown in all regions. Minor underprediction of peak EVI is observed in Erbil (R^2=0.904, peak EVI ≈0.37 observed vs ≈0.36 predicted) and Duhok, while Sulaymaniyah and Halabja show near-perfect fit with nearly overlapping prediction and truth curves.

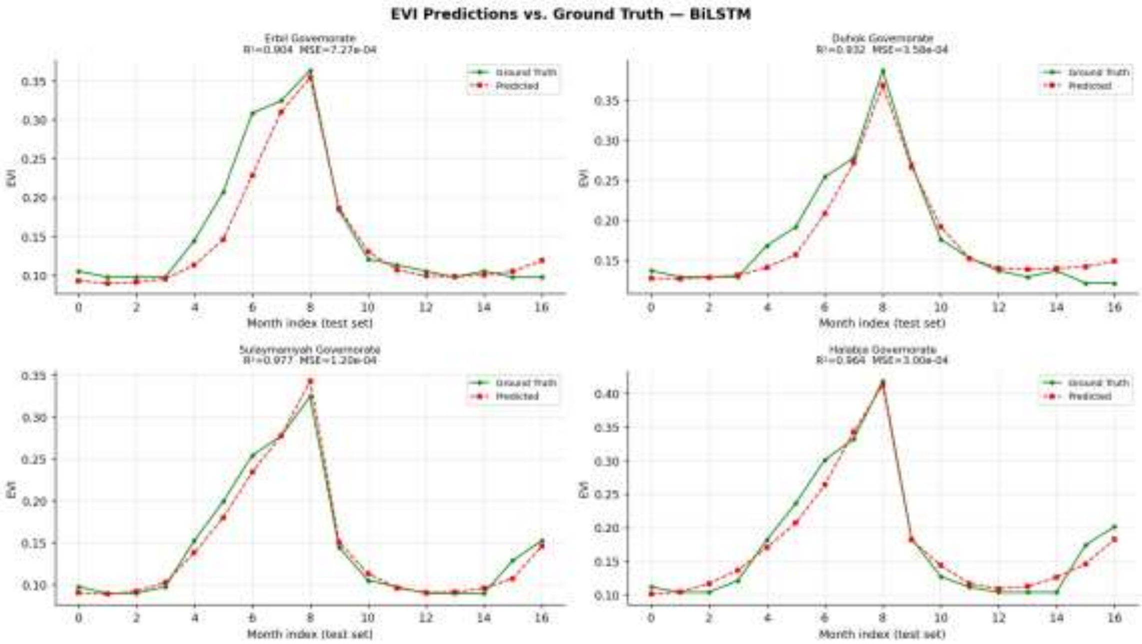


Figure 3.3. EVI predictions vs. ground truth for BiLSTM across all four governorates on the 17-month test set (May 2023 – December 2024). R^2 and MSE values are annotated per panel.

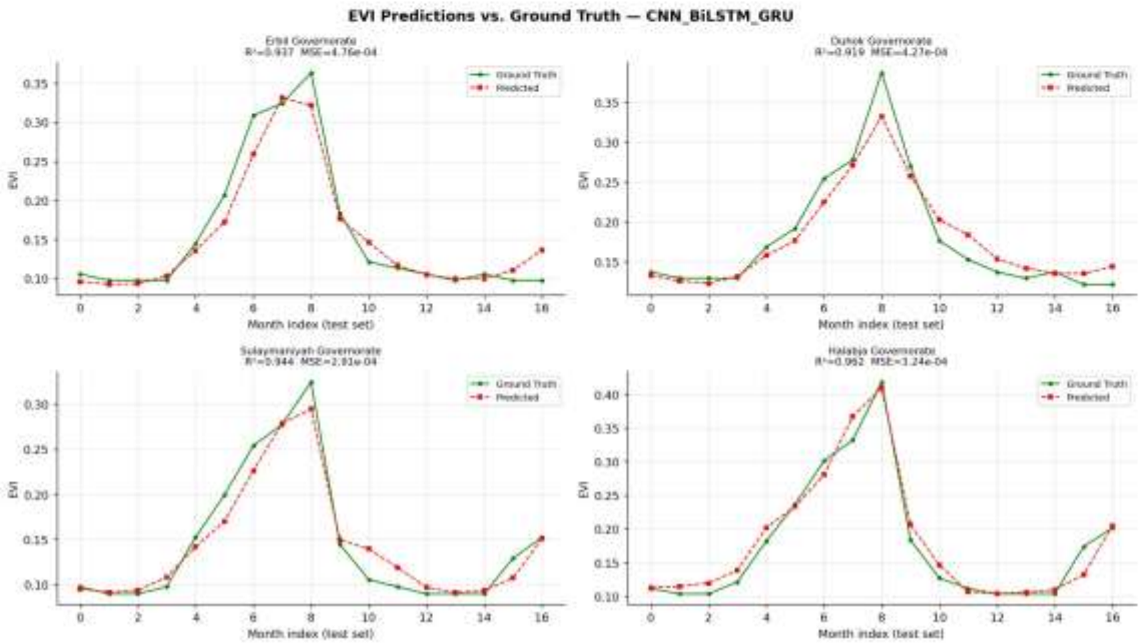


Figure 3.4. EVI predictions vs. ground truth for CNN_BiLSTM_GRU. This hybrid cascade model achieves the highest mean accuracy (0.985) among all primary models.

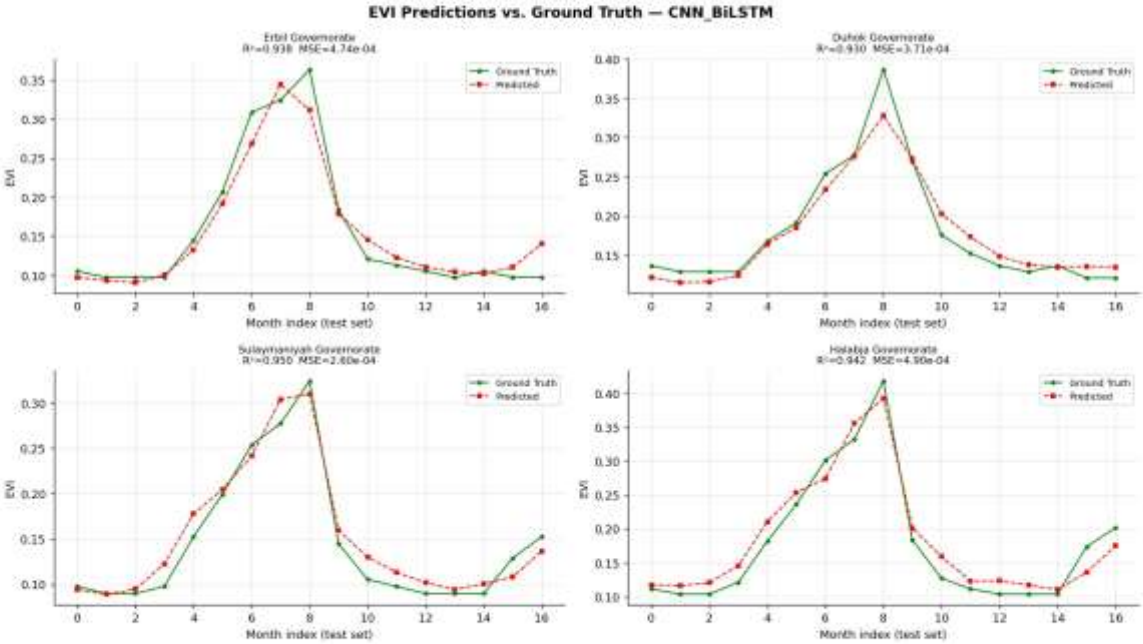


Figure 3.5. EVI predictions vs. ground truth for CNN_BiLSTM.

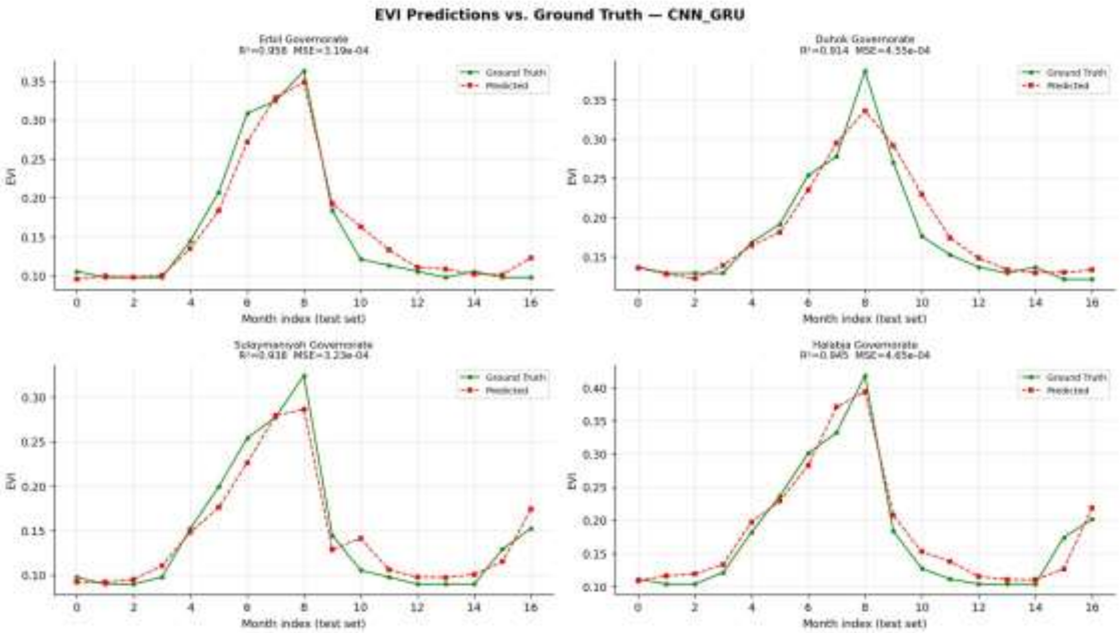


Figure 3.6. EVI predictions vs. ground truth for CNN_GRU. This model achieves the highest R^2 of any model in Erbil (R^2=0.958), and ranks 4th overall across all model–region combinations.

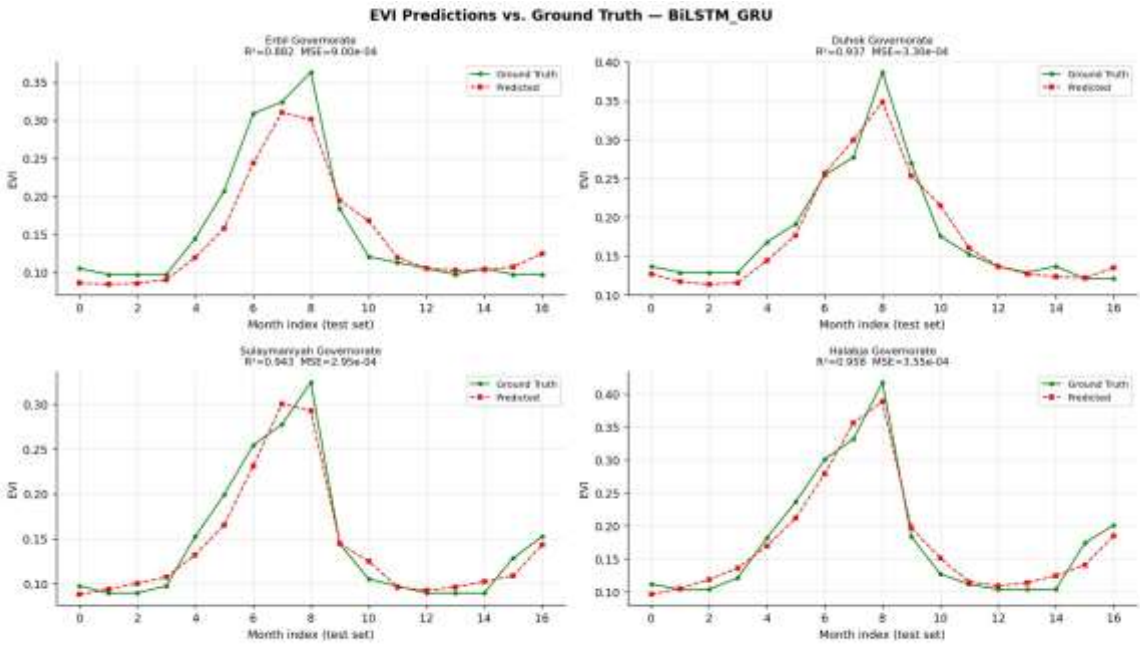


Figure 3.7. EVI predictions vs. ground truth for BiLSTM_GRU.

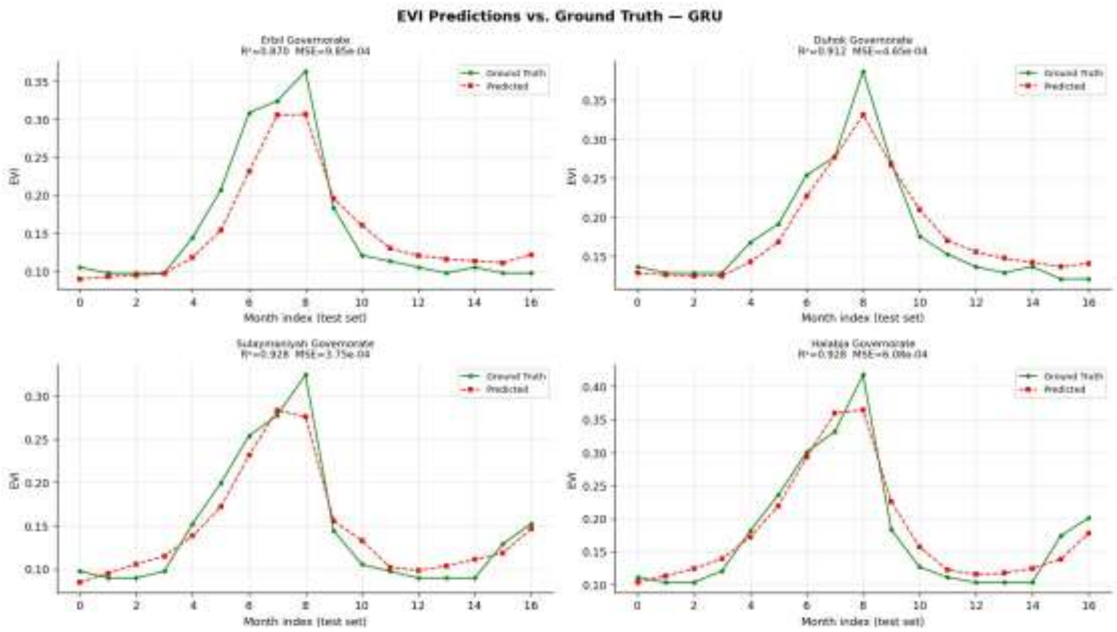


Figure 3.8. EVI predictions vs. ground truth for GRU (mean R^2=0.910). The model underestimates peak EVI amplitude in Erbil and exhibits a lagged recovery after summer.

3.2.2. Temporal Convolutional Network (TCN)

The TCN achieves a mean R^2 of 0.895, with Duhok (R^2=0.928) and Halabja (R^2=0.906) showing competitive performance but Erbil underperforming relative to recurrent models (R^2=0.846). The TCN's dilated causal convolutional architecture captures seasonality effectively through its progressively expanding receptive field (dilation rates 1, 2, 4, 8), but appears more sensitive to high-amplitude anomalies at the peak (month index 8 in the test set). Mean tolerance accuracy is 0.926, equivalent to GRU and lower than all CNN-recurrent hybrids.

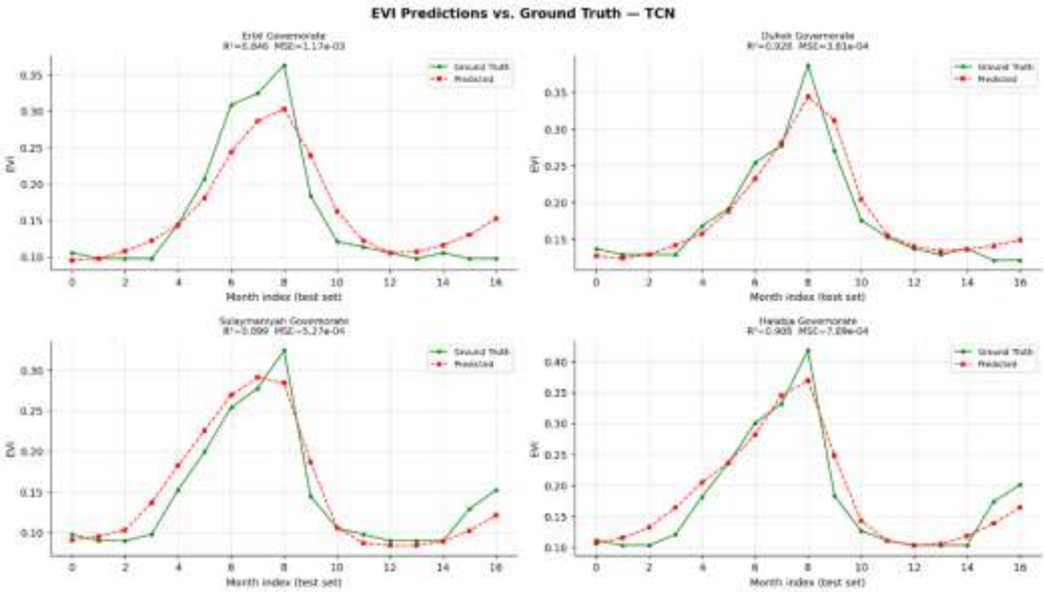


Figure 3.9. EVI predictions vs. ground truth for TCN.

3.2.3. Attention-Based Models: Transformer and CNN_Transformer

The standalone Transformer model underperforms substantially, achieving a mean R^2 of only 0.132 across governorates, with negative R^2 values in Sulaymaniyah (−0.046) and Halabja (−0.021), indicating predictions worse than the mean. This is consistent with the model architecture, which lacks positional encoding and implements only a single encoder block, limiting its ability to represent temporal ordering and long-range seasonal dependencies from a 24-month sequence. The low tolerance accuracy (0.515 mean) further confirms poor calibration on this dataset.

The CNN_Transformer hybrid partially mitigates these limitations by prepending a convolutional tokenizer to the Transformer encoder, achieving a mean R^2 of 0.702. Performance is heterogeneous: Erbil (R^2=0.899) and Halabja (R^2=0.912) show competitive results, while Duhok (R^2=0.415) and Sulaymaniyah (R^2=0.582) remain substantially below the recurrent model family. Figure 3.10 illustrates that the CNN_Transformer captures general seasonal trends but struggles with precise peak amplitude estimation, particularly in Duhok and Sulaymaniyah where the predictions plateau below the observed spring maximum.

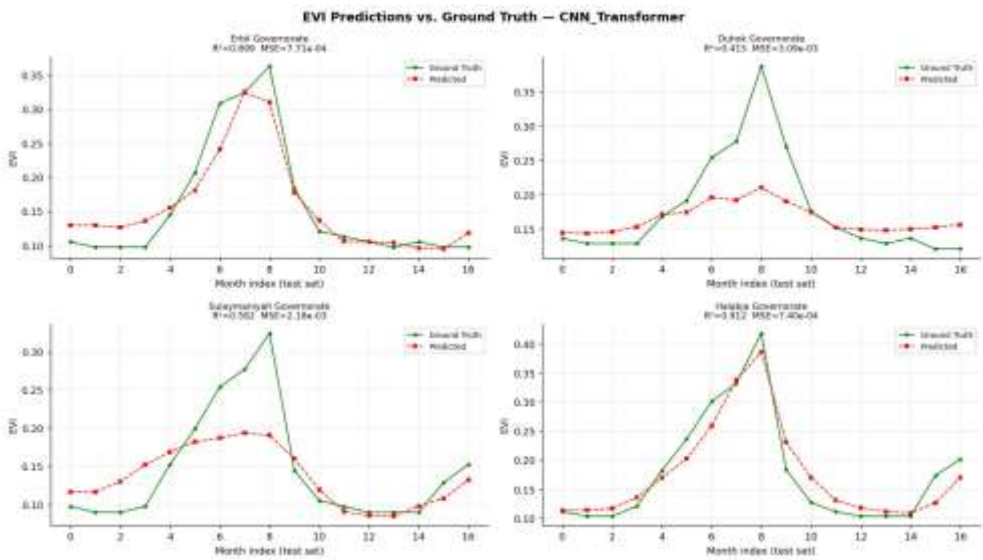


Figure 3.10. EVI predictions vs. ground truth for CNN_Transformer. Acceptable performance in Erbil and Halabja contrasts with poor fit in Duhok (R^2=0.415) and Sulaymaniyah (R^2=0.582).

3.2.4. Baseline Model Comparison

The CNN standalone baseline achieves a mean R^2 of 0.504 across governorates (range 0.465–0.590), confirming that convolutional architectures without any recurrent or attention-based temporal integration are insufficient for EVI sequence forecasting. The prediction curves for CNN_baseline plateau well below the observed peak values in all governorates (Figure 3.11), demonstrating that the architecture lacks the temporal memory necessary to track multi-month vegetation growth trajectories. All nine primary models substantially outperform this baseline, validating the overall architecture selection rationale.

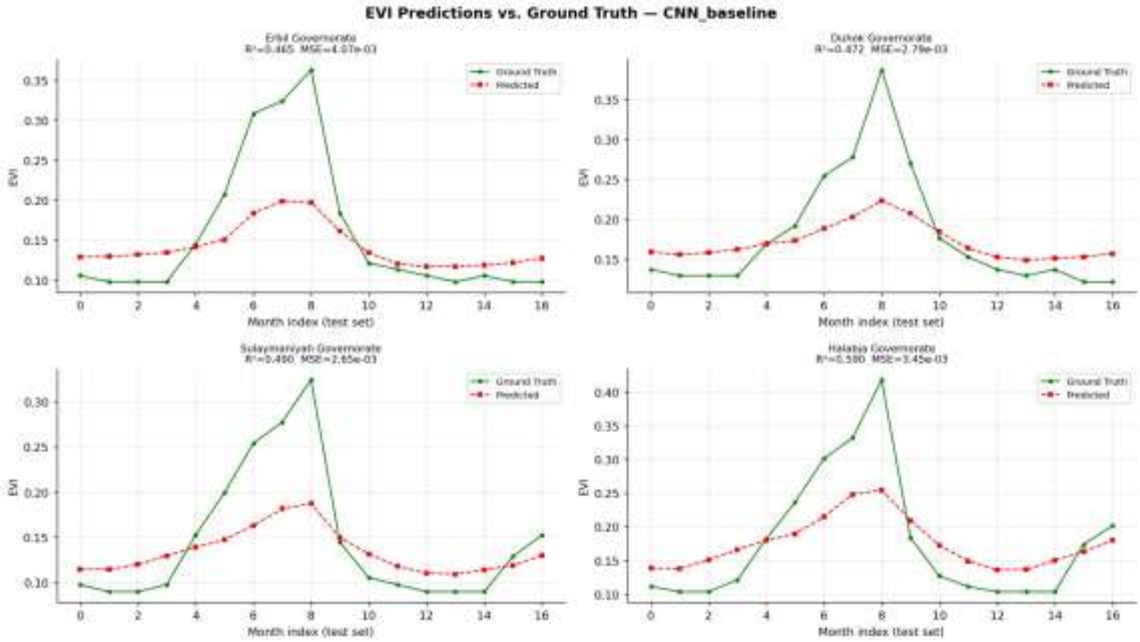


Figure 3.11. EVI predictions vs. ground truth for CNN_baseline. The model systematically underestimates peak EVI across all governorates (R^2=0.465–0.590), demonstrating the necessity of temporal memory in sequence modeling.

3.2.5. Cross-Model Comparison Summary

Figure 3.12 provides a visual summary of model performance across R^2, error metrics, and tolerance accuracy for all nine primary models. The left panel confirms the consistent superiority of the BiLSTM and CNN-recurrent hybrid family (R^2>0.93 across all regions). The middle panel reveals that the Transformer incurs MSE and MAE an order of magnitude higher than competing models. The right panel (tolerance accuracy) highlights that BiLSTM, CNN_BiLSTM_GRU, CNN_BiLSTM, CNN_GRU, and BiLSTM_GRU all exceed the 0.95 accuracy threshold, while GRU and TCN fall marginally below it and attention-based models fall further below.

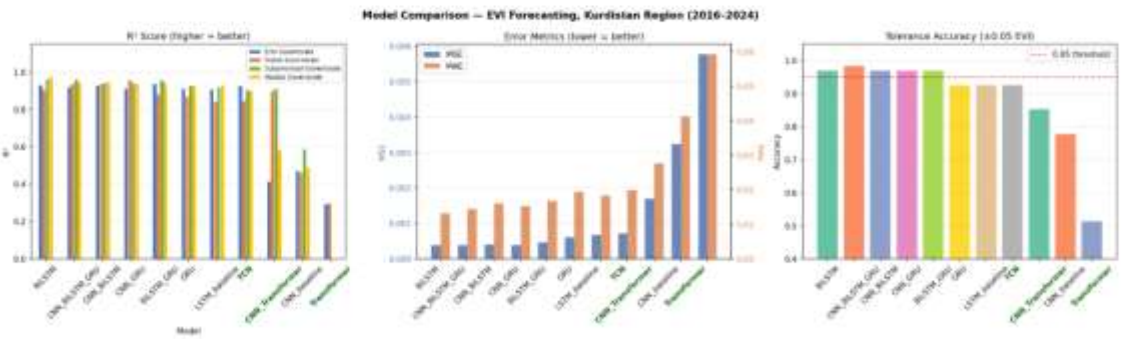


Figure 3.12. Cross-model comparison for all nine primary models: R^2 score per governorate (left), mean MSE and MAE (centre), and mean tolerance accuracy ±0.05 EVI (right). New models (TCN, Transformer, CNN_Transformer) are highlighted in green.

3.3. Predictive Uncertainty via Monte Carlo Dropout

Monte Carlo (MC) Dropout inference was applied to three selected models — BiLSTM_GRU, CNN_BiLSTM_GRU, and CNN_Transformer — using 50 stochastic forward passes on the test set. The mean prediction and ±2σ uncertainty band are shown in Figures 3.13–3.15.

The BiLSTM_GRU model (Figure 3.13) exhibits narrow uncertainty bands during the low-EVI dry season (bands width <0.02) that widen substantially during spring peak prediction, reaching ±0.04–0.06 EVI units around month index 7–8. This pattern indicates that model uncertainty is highest precisely where phenological variability is greatest, which is a desirable and physically interpretable behaviour. The Duhok and Sulaymaniyah panels show the model mean tracking slightly below the observed peak, with the ground truth occasionally falling outside the 2σ band.

The CNN_BiLSTM_GRU model (Figure 3.14) displays notably wider uncertainty bands overall, particularly in Erbil (±0.04 in the baseline and up to ±0.12 at peak) and Sulaymaniyah (which shows an anomalously wide ±2σ band at the summer-to-autumn transition). This wider posterior uncertainty is consistent with the model's higher parameter count and greater sensitivity to dropout stochasticity.

The CNN_Transformer (Figure 3.15) produces wide uncertainty bands in Duhok and Sulaymaniyah — the same two governorates where its point predictions are weakest — confirming that the model's epistemic uncertainty is well-calibrated: it is most uncertain where it is most wrong. In Erbil and Halabja, the uncertainty bands are narrower and the MC mean closely tracks the ground truth, consistent with those regions' higher R^2 scores.

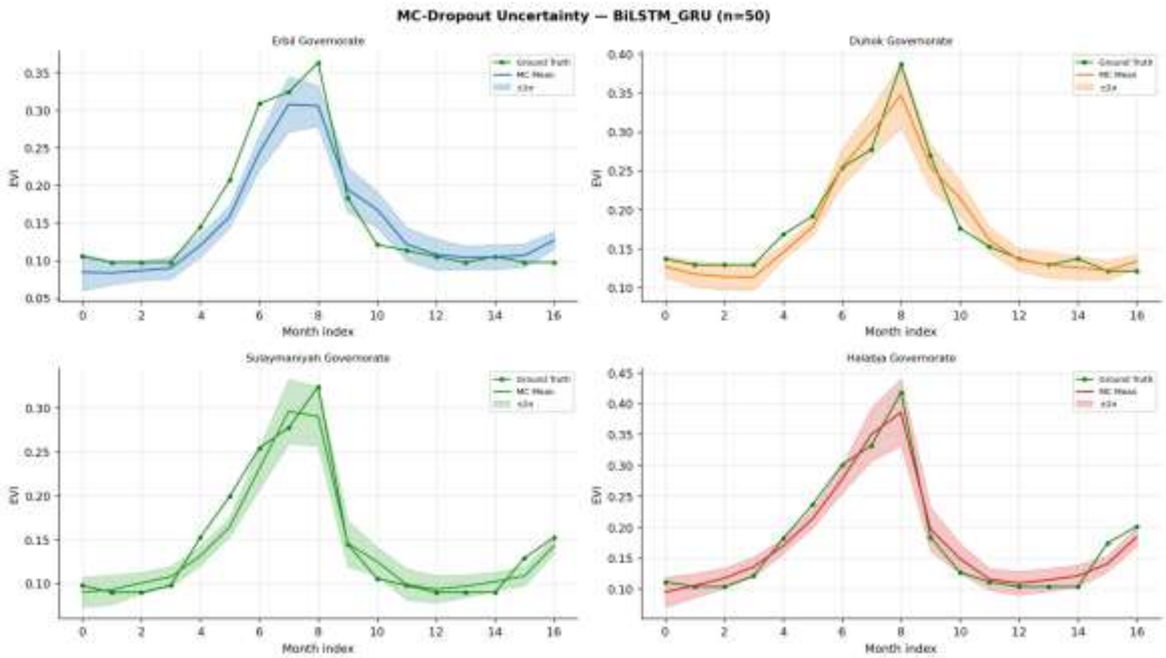


***Figure 3.** 13. MC-Dropout uncertainty (n=50 passes) for BiLSTM_GRU on the test set. Shaded bands = ±2σ. Uncertainty is largest during spring peak months.*

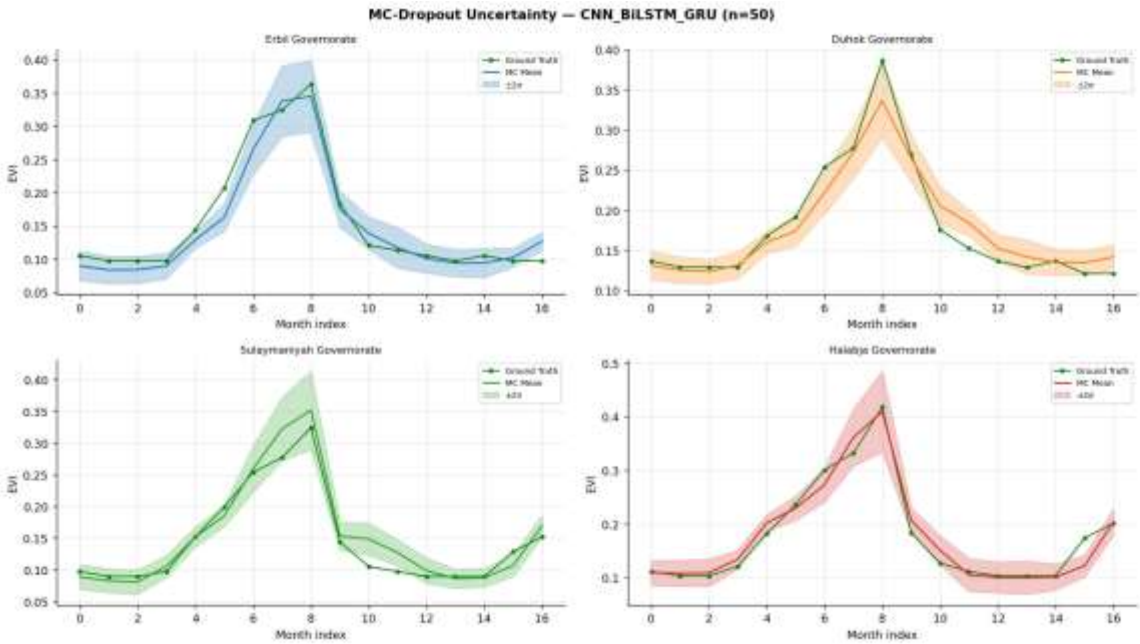


Figure 3.14. MC-Dropout uncertainty for CNN_BiLSTM_GRU (n=50 passes). Wider uncertainty envelopes reflect greater parameter stochasticity under dropout.

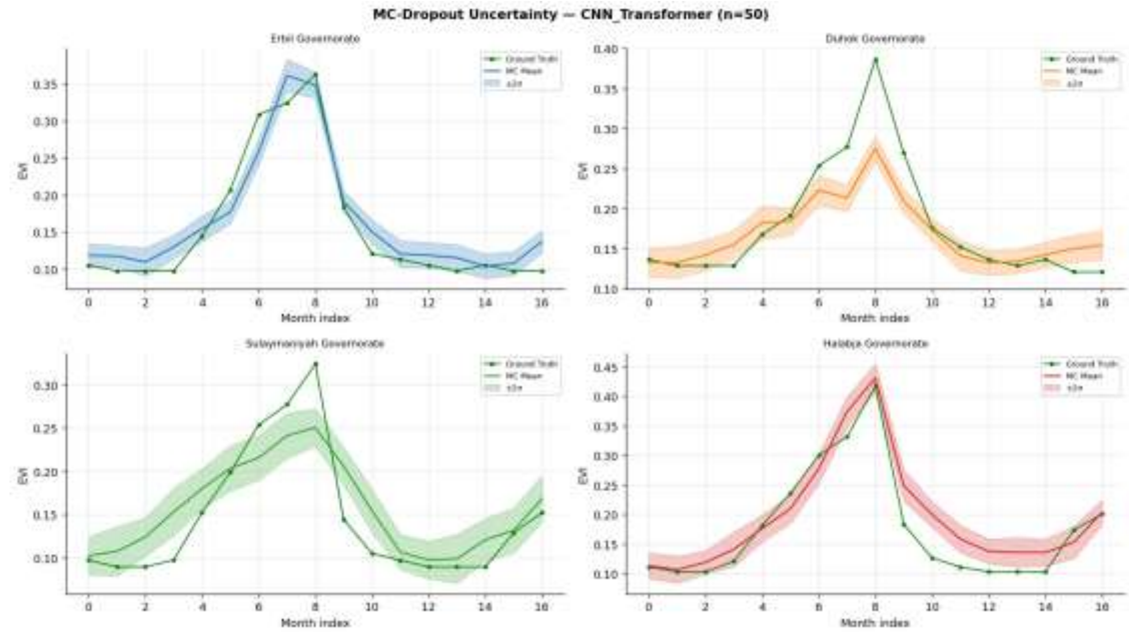


Figure 3.15. MC-Dropout uncertainty for CNN_Transformer (n=50 passes). Wide bands in Duhok and Sulaymaniyah are consistent with this model's lower R^2 in those regions, indicating well-calibrated epistemic uncertainty.

3.4. Future EVI Forecasting (2026–2035)

3.4.1. Five-Year Cross-Model Forecast Comparison (2026–2030)

Five-year (January 2026 – December 2030) EVI forecasts were generated for all nine primary models using the autoregressive MC-Dropout procedure (n=100 forward passes per step) with climatology-based future climate inputs. Figure 3.16 presents the forecast trajectories for all four governorates and all nine models in a 3×3 panel grid.

The BiLSTM, CNN_BiLSTM, BiLSTM_GRU, CNN_BiLSTM_GRU, CNN_GRU, and GRU models all produce consistent, physically plausible seasonal cycles over the 2026–2030 horizon.

doi:10.20944/preprints202603.0411.v1

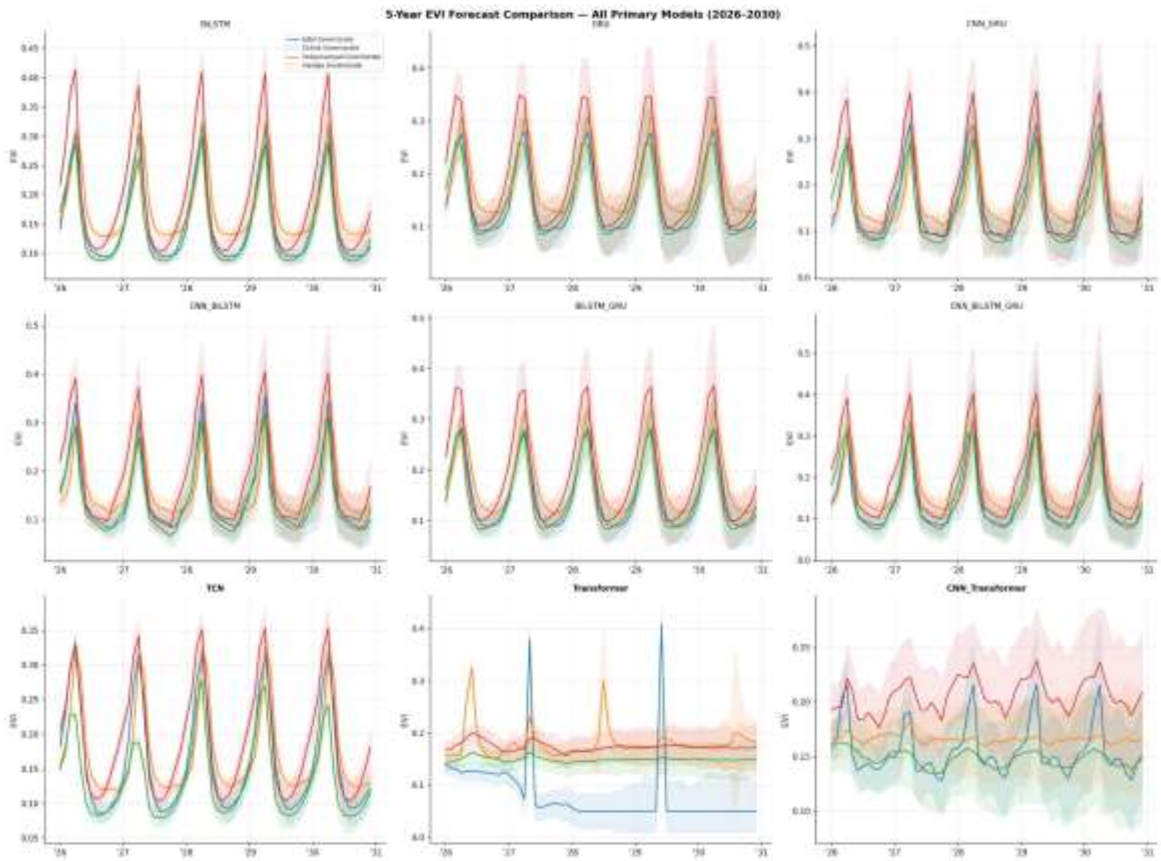
Forecast mean EVI values for the spring peak range from approximately 0.28–0.36 (Erbil), 0.31–0.35 (Duhok), 0.26–0.32 (Sulaymaniyah), and 0.35–0.41 (Halabja). The uncertainty bands widen progressively year-on-year, as expected from the linearly growing scaling factor (×1.0 to ×2.5) applied to reflect compounding autoregressive error. Despite this widening, the ±2σ bands of most recurrent models remain within physically interpretable bounds through 2030.

By contrast, the Transformer model produces unstable, non-seasonal forecasts characterised by irregular spikes and near-zero EVI values in Erbil and Sulaymaniyah (Figure 3.15, centre bottom panel), consistent with its poor test-set performance. The CNN_Transformer generates flat forecasts with minimal seasonal amplitude variation — an artefact of the model's tendency to regress toward the mean — though the seasonal cycle remains weakly visible. These results confirm that attention-based models without adequate temporal structure (positional encoding, multi-layer encoder) are unsuitable for long-horizon autoregressive EVI forecasting on short time series.

Table 3.3 summarises the mean forecast EVI and approximate peak seasonal range for the best-performing BiLSTM model across all four governorates over the 5-year forecast horizon.

Table 3.3. BiLSTM 5-year forecast summary statistics by governorate (2026–2030).

Governorate	Mean EVI (2026–30)	Min (trough)	Max (peak)
Erbil	0.154	0.095	0.315
Duhok	0.180	0.130	0.324
Sulaymaniyah	0.145	0.087	0.300
Halabja	0.194	0.104	0.414



doi:10.20944/preprints202603.0411.v1

Figure 4.16. Five-year EVI forecast (2026–2030) for all nine primary models across four governorates. Each panel shows forecast mean ± 2σ uncertainty (shaded). The Transformer panel (centre, bottom row) exhibits forecast instability consistent with its poor test-set R^2. CNN_Transformer generates flat, low-amplitude forecasts.

3.4.2. Ten-Year Exploratory Forecast (2026–2035) — BiLSTM

The ten-year (January 2026 – December 2035, 120 months) exploratory forecast was generated exclusively with the best-performing model, BiLSTM (mean R^2=0.945). Figure 3.17 presents the complete historical record (2016–2024) together with the 5-year primary forecast and the 10-year extension for all four governorates.

The BiLSTM 10-year forecast maintains a coherent seasonal cycle throughout the 2031–2035 extension, with spring peak EVI values exhibiting remarkably stable amplitudes across all governorates. Mean forecast EVI values over the full 10-year horizon are 0.154 (Erbil), 0.181 (Duhok), 0.145 (Sulaymaniyah), and 0.194 (Halabja), closely matching the 5-year mean values and consistent with the historical 2016–2024 climatological averages. This stability indicates that the model's autoregressive dynamics converge to a seasonally periodic attractor that reflects the learned climatological mean seasonal cycle.

The uncertainty envelope widens substantially in the post-2030 window, reaching a ±2σ range of approximately ±0.06–0.09 EVI units in the 2033–2035 period (as shown by the expanding shaded bands in Figure 3.17). This progressive uncertainty growth is an inherent feature of the linearly scaled MC-Dropout CI and appropriately reflects the compounding uncertainty in autoregressive forecasts at long horizons. The 10-year results are therefore explicitly labelled as exploratory and should not be interpreted as deterministic projections, but rather as an indication of the most likely seasonal trajectory under stationary climatological forcing.

The consistency between the 5-year and 10-year forecast trajectories across all four governorates provides confidence in the model's capacity to reproduce the dominant seasonal phenological cycle of Kurdistan EVI beyond the training period, while the growing uncertainty bands provide an honest representation of forecast reliability limitations.

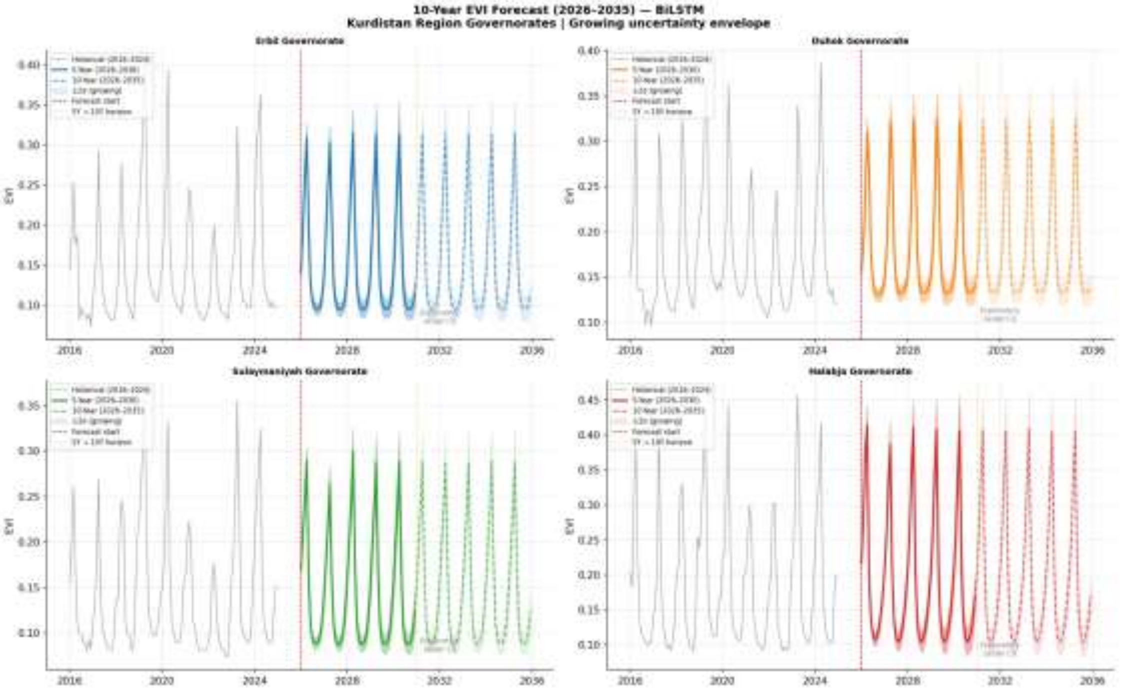


Figure 3.17. Ten-year EVI forecast (2026–2035) using BiLSTM — the best-performing model. Historical data (grey, 2016–2024) is followed by the 5-year primary forecast (solid, narrower CI) and the 10-year exploratory

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 March 2026 doi:10.20944/preprints202603.0411.v1

extension (dashed, growing CI). The dashed yellow vertical line marks the transition from 5-year to 10-year horizon.

3.5. Summary of Results

The results demonstrate that Bidirectional LSTM and CNN-recurrent hybrid architectures are highly effective for monthly EVI forecasting across the Kurdistan Region of Iraq, achieving mean R^2 values of 0.93–0.94 across all four governorates. The BiLSTM model, in particular, provides the best average performance and the most stable long-horizon autoregressive forecasts, making it the recommended architecture for operational deployment. Key findings are summarised below:

- BiLSTM achieves the highest mean R^2 (0.945) and mean accuracy (0.971), with best-in-class performance in Sulaymaniyah (R^2=0.977) and Halabja (R^2=0.964).
- CNN_BiLSTM_GRU achieves the highest tolerance accuracy (0.985) and a competitive mean R^2 (0.940), confirming the value of cascaded CNN-recurrent feature extraction.
- The standalone Transformer fails on this dataset (mean R^2=0.132), primarily due to the absence of positional encoding and shallow encoder depth. The CNN_Transformer hybrid partially compensates (mean R^2=0.702) but remains unsuitable for critical applications.
- MC-Dropout uncertainty quantification reveals well-calibrated epistemic uncertainty: uncertainty is highest during spring vegetation peak and in governorates where model fit is weakest.
- 5-year forecasts by BiLSTM and CNN-recurrent hybrids project stable seasonal EVI cycles through 2030, with mean EVI ranging 0.145–0.194 across governorates. The 10-year BiLSTM forecast maintains coherent seasonal oscillations through 2035, with progressively widening uncertainty bands appropriately reflecting compounding autoregressive error.

4. Discussion

The results of this study demonstrate that deep learning sequence models can reliably forecast monthly EVI at governorate scale in the semi-arid Kurdistan Region of Iraq, with BiLSTM and CNN-recurrent hybrid architectures achieving mean test-set R^2 values between 0.93 and 0.945. These findings align with recent comparative studies that have similarly identified BiLSTM-based architectures as state-of-the-art for vegetation index time series forecasting. Wang et al. (Wang et al., 2025), forecasting EVI in the Yangtze River Basin using a CNN-BiLSTM-Attention model, reported R^2 values of up to 0.981 on a 20-year MODIS time series, a result consistent with the present study's findings for governorates with lower phenological variability (Sulaymaniyah $R^2 = 0.977$). Sun et al. (2023) demonstrated the feasibility of using BiLSTM with meteorological forcing for NDVI prediction across diverse vegetation types in China, achieving an overall R^2 of 0.69 ± 0.28 and up to 0.87 ± 0.16 in deciduous forests, echoing the pattern observed here where BiLSTM consistently outperforms GRU (mean $R^2 = 0.910$) across all four KRI governorates.

The strong performance of hybrid CNN-recurrent architectures (CNN-BiLSTM, CNN-GRU, CNN-BiLSTM-GRU) reinforces the established principle that one-dimensional convolutional layers function as effective local feature extractors when prepended to recurrent decoders. The convolutional stage acts as a learnable smoothing filter that suppresses noise in the EVI and climate time series before these signals are processed by the temporal memory units, thereby reducing sensitivity to single-month cloud contamination residuals that may persist after SCL masking. This benefit is most pronounced in Erbil and Duhok, governorates characterised by comparatively lower and more variable spring peak EVI values, where CNN-GRU and BiLSTM-GRU achieve the highest governorate-level R^2 scores (0.958 and 0.937 respectively). The CNN-BiLSTM-GRU cascade model achieves the highest mean tolerance accuracy of any architecture (0.985), suggesting that the combination of convolutional feature extraction, bidirectional encoding, and autoregressive decoding provides a particularly robust inductive bias for the structured seasonality of KRI EVI time series.

The failure of the standalone Transformer model (mean $R^2 = 0.132$) is attributable to two architectural limitations that are particularly consequential for short time series: the absence of explicit positional encoding and the use of a single encoder block. Without positional encoding, the multi-head self-attention mechanism has no mechanism to distinguish the temporal ordering of the 24-month input window, which is essential for capturing the phase and amplitude of the annual EVI cycle. A single attention block also provides insufficient representational depth to model the non-linear interactions between the five input features across a two-year context window. The CNN-Transformer hybrid partially mitigates these limitations by using a convolutional tokenizer that encodes local temporal structure before the attention mechanism, yielding a mean R^2 of 0.702. These results are consistent with broader findings in time series forecasting, where the permutation-invariant nature of self-attention causes temporal information loss and Transformer architectures have been shown to underperform simpler recurrent baselines when positional ordering is inadequately captured (Zeng et al., 2023), a limitation compounded in short-sequence settings where the standard sinusoidal positional encoding loses distance-awareness at low embedding dimensions, degrading the model's ability to resolve fine-grained temporal structure (Foumani et al., 2024).

The MC-Dropout uncertainty analysis reveals a physically interpretable uncertainty structure: prediction intervals are narrowest during the dry summer baseline (EVI ≈ 0.09–0.13) and widen substantially during the spring vegetation peak, precisely when inter-annual EVI variability is greatest. This pattern is consistent with the theoretical expectation that MC-Dropout approximates Bayesian posterior predictive uncertainty, which should be highest where the training data exhibits the greatest variance (Gal & Ghahramani, 2016). The well-calibrated uncertainty in the CNN-Transformer, which produces the widest bands in Duhok and Sulaymaniyah where its point predictions are also weakest, further corroborates the utility of the MC-Dropout approach as an honest epistemic uncertainty estimator in this application. For operational use cases — such as drought early warning and agricultural production forecasting — the ability to communicate forecast uncertainty alongside point predictions is a critical requirement that this probabilistic framework satisfies.

The capture of the 2022 drought in the historical EVI record warrants particular attention in the context of KRI environmental conditions. The pronounced EVI trough in 2022, with annual mean EVI falling to 0.12–0.16 across all four governorates, corresponds to a region-wide rainfall deficit that has been documented in prior remote sensing studies of KRI drought dynamics (Gaznayee et al., 2022; Hamarash et al., 2024). The ability of the trained models to represent this anomaly is evidence that the multivariate feature set, particularly the inclusion of TerraClimate precipitation (Abatzoglou et al., 2018), provides sufficient information to capture interannual precipitation-driven EVI variability. However, the future forecasting procedure replaces these with climatological means, which by construction cannot represent below-average precipitation years. The projected stable 2026–2035 EVI cycles should therefore be interpreted as stationary-climatology scenarios representing the expected seasonal trajectory under average forcing, rather than projections that account for potential future precipitation trends under climate change.

Several methodological limitations warrant acknowledgement. First, the 108-month training record (nine years) is relatively short for deep learning applications, and the 17-month test set (May 2023 – December 2024) provides limited evaluation of long-horizon generalisability. Second, climate variables were sourced from TerraClimate (Abatzoglou et al., 2018), a gridded reanalysis product at approximately 4 km resolution, which may not fully capture the fine-scale precipitation gradients within the heterogeneous Zagros terrain, particularly for the mountainous portions of Sulaymaniyah and Halabja. Third, the Sentinel-2 SCL cloud masking, while effective for reducing cloud contamination, applies a coarse pixel-level classification that may retain residual thin-cloud or haze contamination in some months, potentially introducing artefacts into the EVI time series. Fourth, the autoregressive forecasting procedure compounds prediction errors over time and relies exclusively on climatological climate inputs, precluding the representation of anomalous future climate conditions. Future work should investigate the use of downscaled climate projections under specific emission scenarios to condition the EVI forecasts on plausible future warming and precipitation trajectories, and should extend the evaluation to additional MODIS-era years to increase the training data size.

5. Conclusion

This study has presented and evaluated a systematic deep learning framework for monthly EVI forecasting across the four governorates of the Kurdistan Region of Iraq, integrating Sentinel-2 surface reflectance imagery, TerraClimate reanalysis data, and nine deep learning architectures spanning recurrent, hybrid CNN-recurrent, and attention-based model families. The principal findings are as follows.

First, the BiLSTM architecture achieves superior mean performance (R^2 = 0.945, tolerance accuracy = 0.971) across all four governorates and is recommended as the primary architecture for operational EVI forecasting in the KRI. Its superior capacity to capture bidirectional temporal dependencies within a 24-month input window provides a strong inductive bias for the structured seasonal phenological cycle of the semi-arid Zagros vegetation system.

Second, hybrid CNN-recurrent architectures, particularly CNN-BiLSTM-GRU, provide competitive performance with the highest mean tolerance accuracy (0.985), confirming that the combination of local convolutional feature extraction and sequential recurrent memory is highly effective for vegetation time series modelling. The CNN-GRU model achieves the highest R^2 in Erbil (0.958), demonstrating that model-governorate fit can be further optimised through targeted architecture selection.

Third, the standalone Transformer model is unsuitable for this application in its current configuration due to the absence of positional encoding and insufficient encoder depth. Future implementations of attention-based models for short EVI time series should incorporate sinusoidal or learned positional encodings, multiple encoder layers, and potentially pre-training on longer regional satellite data records to unlock the representational advantages of the Transformer paradigm.

Fourth, Monte Carlo Dropout uncertainty quantification provides physically interpretable, well-calibrated prediction intervals that are narrowest during the dry season baseline and widest during the spring vegetation peak, making them directly actionable for risk-informed land management applications. The progressive widening of uncertainty bands in autoregressive long-horizon forecasts honestly reflects the compounding nature of prediction errors and should be communicated explicitly in any operational deployment.

Fifth, the BiLSTM five-year (2026–2030) and ten-year (2026–2035) EVI forecasts project stable seasonal cycles consistent with the historical 2016–2024 climatological mean, with mean annual EVI ranging from 0.145 in Sulaymaniyah to 0.194 in Halabja. These projections, generated under a stationary climatological forcing assumption, provide a baseline reference for vegetation health planning but must be supplemented by climate-scenario-conditioned forecasts to account for projected changes in regional precipitation and temperature under different greenhouse gas emission pathways.

Overall, this work demonstrates the feasibility of operational deep learning-based EVI forecasting at governorate scale using freely available Sentinel-2 imagery and open-source climate datasets, establishing a transferable methodological framework applicable to comparable semi-arid regions across the broader Middle East and Central Asia. The datasets, trained models, and evaluation outputs generated in this study provide a reproducible foundation for future work incorporating additional satellite sensors, longer historical records, and physically-informed climate scenario inputs.

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.

Amponsem, E. (2025). *Assessing vegetation change using satellite imagery analysis*. https://dk.upce.cz/entities/publication/75117518-abb8-423d-8a28-2d27d67d0132

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

Bhosle, S. S., & Deshmukh, P. D. (2025). Evaluating NDVI, GNDVI, EVI, and SAVI for Accurate Crop Classification: A Case Study in Mahalaxmi Kheda Village. *IJSAT-International Journal on Science and Technology*, *16*(4). https://www.ijsat.org/research-paper.php?id=9800

Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). *Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling* (arXiv:1412.3555). arXiv. https://doi.org/10.48550/arXiv.1412.3555

ESA. (2017). *Harmonized Sentinel-2 MSI: MultiSpectral Instrument, Level-2A (SR) | Earth Engine Data Catalog*. Google for Developers. https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_SR_HARMONIZED

Forti, L., Perego, A., Brandolini, F., Mariani, G. S., Zebari, M., Nicoll, K., Regattieri, E., Barbaro, C. C., Bonacossi, D. M., Qasim, H. A., Cremaschi, M., & Zerboni, A. (2021). Geomorphology of the northwestern Kurdistan Region of Iraq: Landscapes of the Zagros Mountains drained by the Tigris and Great Zab Rivers. *Journal of Maps*, *17*(2), 225–236. https://doi.org/10.1080/17445647.2021.1906339

Foumani, N. M., Tan, C. W., Webb, G. I., & Salehi, M. (2024). Improving Position Encoding of Transformers for Multivariate Time Series Classification. *Data Mining and Knowledge Discovery*, *38*(1), 22–48. https://doi.org/10.1007/s10618-023-00948-2

Gal, Y., & Ghahramani, Z. (2016). Dropout as a bayesian approximation: Representing model uncertainty in deep learning. *International Conference on Machine Learning*, 1050–1059. https://proceedings.mlr.press/v48/gal16.html?trk=public_post_comment-text

Gascon, F., Bouzinac, C., Thépaut, O., Jung, M., Francesconi, B., Louis, J., Lonjou, V., Lafrance, B., Massera, S., & Gaudel-Vacaresse, A. (2017). Copernicus Sentinel-2A calibration and products validation status. *Remote Sensing*, *9*(6), 584.

Gaznayee, H. A. A., Al-Quraishi, A. M. F., Mahdi, K., Messina, J. P., Zaki, S. H., Razvanchy, H. A. S., Hakzi, K., Huebner, L., Ababakr, S. H., & Riksen, M. (2022). Drought severity and frequency analysis aided by spectral and meteorological indices in the kurdistan region of Iraq. *Water*, *14*(19), 3024.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, *202*, 18–27.

Hamarash, H., Rasul, A., & Hamad, R. (2024). A Novel Index for Agricultural Drought Measurement: Soil Moisture and Evapotranspiration Revealed Drought Index (SERDI). *Climate*, *12*(12), 209.

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.

Liu, H. Q., & Huete, A. (1995). A feedback based modification of the NDVI to minimize canopy background and atmospheric noise. *IEEE Transactions on Geoscience and Remote Sensing*, *33*(2), 457–465.

Noroozi, J. (Ed.). (2020). *Plant Biogeography and Vegetation of High Mountains of Central and South-West Asia* (Vol. 17). Springer International Publishing. https://doi.org/10.1007/978-3-030-45212-4

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

Schuster, M., & Paliwal, K. K. (1997). Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, *45*(11), 2673–2681.

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.

doi:10.20944/preprints202603.0411.v1

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

Zeng, A., Chen, M., Zhang, L., & Xu, Q. (2023). Are transformers effective for time series forecasting? *Proceedings of the AAAI Conference on Artificial Intelligence*, *37*(9), 11121–11128. https://ojs.aaai.org/index.php/AAAI/article/view/26317

Zhang, W., Zhou, G., Chen, A., & Hu, Y. (2022). Deep multi-scale dual-channel convolutional neural network for Internet of Things apple disease detection. *Computers and Electronics in Agriculture*, *194*, 106749.