Journal of Hydrology: Regional Studies 65 (2026) 103376

Contents lists available at ScienceDirect

Journal of Hydrology: Regional Studies

journal homepage: www.elsevier.com/locate/ejrh

Comparative machine learning and deep learning approaches for agricultural drought monitoring: Dual-index modeling in Iran

Mahan Azizi [a,b], Ali Abbasi [b,c,*], Mohammad Reza Asli Charandabi [d]

[a] Department of AI and Land Use Change, Faculty VI, Technische Universität Berlin, Berlin, Germany
[b] Department of Civil Engineering, Faculty of Engineering, Ferdowsi University of Mashhad, Mashhad, Iran
[c] Department of Water Management, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, The Netherlands
[d] Faculty of Civil Engineering, Shahrood University of Technology, Shahrood, Iran

ARTICLE INFO

Keywords:
Agricultural drought monitoring
Soil Moisture Deficit Index
Standardized Precipitation–Evapotranspiration Index
LightGBM
Random Forest
Elastic Net
FT-Transformer

ABSTRACT

Study region: This study considers Iran, encompassing hyper-arid to humid hydroclimates and major agricultural plains. Using 70 synoptic stations (2001–2022), we collocated station observations with satellite/reanalysis predictors from the Global Precipitation Measurement (GPM) mission, the Moderate Resolution Imaging Spectroradiometer (MODIS), the Famine Early Warning Systems Network Land Data Assimilation System (FLDAS), and the Copernicus Climate Change Service (C3S).

Study focus: Agricultural drought monitoring benefits from combining indicators of meteorological forcing and land-surface response, yet many studies rely on a single index or combine indices without an operational integration logic. We propose a dual-index framework for Iran integrating the Soil Moisture Deficit Index (SMDI) and the 3-month Standardized Precipitation–Evapotranspiration Index (SPEI-3).

New hydrological insights for the region: We combine stability selection with leakage-safe forward expanding cross-validation and a held-out most-recent test window to compare Light Gradient Boosting Machine (LightGBM), Random Forest, Elastic Net, and a feature-tokenizer Transformer. SMDI is estimated more reliably (best RMSE = 0.80, R^2 = 0.82) than SPEI-3 (best RMSE = 0.96, R^2 = 0.55). Uncertainty is quantified from held-out test absolute errors via empirical quantiles (50% and 90%); for SMDI, ~50% of predictions fall within ~0.5 index units and ~90% within ~1–1.5 units. These quantile error bands are attached as confidence qualifiers to the monthly drought classes in the monitoring framework, where SMDI anchors severity and SPEI-3 supports early-warning escalation.

1. Introduction

Agricultural drought is a major environmental hazard that emerges when water deficits reduce root-zone soil moisture and constrain crop growth, leading to yield losses and broader impacts on water resources and food security (Orimoloye, 2022; Potopová et al., 2016; Yu et al., 2018). Because its effects can propagate rapidly across farming systems, timely monitoring and early warning are essential for preparedness and decision-making (Dutta et al., 2013; Walker et al., 2024). Drought conditions are commonly quantified

* Corresponding author at: Department of Civil Engineering, Faculty of Engineering, Ferdowsi University of Mashhad, Mashhad, Iran.
E-mail addresses: mahan.azizi@campus.tu-berlin.de (M. Azizi), aabbasi@um.ac.ir, a.abbasi@tudelft.nl (A. Abbasi), m.aslicharandabi@shahroodut.ac.ir (M.R. Asli Charandabi).

https://doi.org/10.1016/j.ejrh.2026.103376
Received 30 August 2025; Received in revised form 16 March 2026; Accepted 21 March 2026
Available online 25 March 2026

using meteorological and soil-moisture indices such as Standardized Precipitation–Evapotranspiration Index (SPEI) (Vicente-Serrano et al., 2010) and Soil Moisture Deficit Index (SMDI) (Narasimhan and Srinivasan, 2005), but conventional assessments can be limited by sparse station coverage and data gaps, motivating complementary remote-sensing-based monitoring for spatially consistent drought characterization (Alam et al., 2017; Alkaraki and Hazaymeh, 2023; En-Nagre et al., 2024; Jung et al., 2020; Vereecken et al., 2008; Zhang et al., 2021). At the same time, drought indices can target different drought types (e.g., meteorological vs. hydrological), and standardized index estimates are sensitive to methodological choices such as the assumed distribution used to standardize the underlying hydro-meteorological variable(s); these factors can yield materially different drought characteristics and therefore motivate careful, context-aware index selection and comparison (Abu Arra et al., 2024a; Abu Arra and Şişman, 2025). In data-scarce regions, this challenge is amplified, and the use—and evaluation—of gridded reanalysis and satellite precipitation products has become a practical pathway to extend drought monitoring and mapping beyond station limitations (Abu Arra et al., 2025).

Remote sensing systems, such as the Moderate Resolution Imaging Spectroradiometer (MODIS), provide critical information on land surface temperature (LST), vegetation health (via vegetation indices like Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI)), and precipitation. These datasets are particularly useful for monitoring drought because they allow for the continuous observation of thermal stress and vegetation conditions, both of which are directly impacted by drought. For instance, it is shown that utilizing MODIS LST and vegetation indices, combined with precipitation data from the Global Precipitation Measurement (GPM) mission, can help develop an improved drought index for mainland China (Wei et al., 2022). This study demonstrated the effectiveness of combining thermal and vegetation data with precipitation for accurate drought monitoring. Similarly, a study demonstrated that applied MODIS-derived LST and NDVI alongside GPM precipitation and soil moisture data from Famine Early Warning Systems Network Land Data Assimilation System (FLDAS) can be utilized to create a comprehensive drought condition indicator using data-driven methods (Alkaraki and Hazaymeh, 2023). However, across Iran's diverse climate regimes, drought classification can vary by region, index type, and even the underlying data source, and Iran-focused studies report both climate-dependent SPI/SPEI behavior and time-lagged propagation from meteorological to soil-moisture drought, motivating an explicitly multi-index and multi-source framework evaluated across climatic zones (Lotfirad et al., 2021; Seifian et al., 2025).

Soil moisture is a key variable in drought monitoring as it directly reflects water availability for crops (Ma et al., 2018; Modanesi et al., 2020). A variety of long-term remotely sensed soil moisture sources are available for drought monitoring, spanning model-based reanalysis and observation-driven multi-sensor records. Among these, FLDAS has shown low uncertainty and strong skill in capturing hydro-thermal conditions and drought signals across drylands, including arid Central Asia and Middle Eastern transboundary basins (Albarakat et al., 2022; Yu et al., 2023). Complementing such models, the Copernicus Climate Change Service (C3S) surface soil moisture (SSM) COMBINED product provides a homogenized, multi-mission observational record whose surface signal has been propagated to the root zone with quantified uncertainties via an exponential-filter framework (Pasik et al., 2023). Taken together, these studies motivate the common practice in regional drought assessment of juxtaposing a model-based reanalysis with an observation-based multi-sensor record as complementary lines of evidence.

Because multi-source drought predictors are complementary but heterogeneous in scale and sensitivity, a growing body of work demonstrates that machine learning (ML) and deep learning (DL) approaches can effectively fuse these inputs and capture the nonlinear, multi-factor controls that underpin drought, complementing index-based frameworks such as SPEI and SMDI (Ahmed Osman et al., 2025; Oyarzabal et al., 2025; Xu et al., 2024). In grassland systems of Inner Mongolia, a head-to-head comparison (Random Forest, Convolutional Neural Network, Support Vector Regression, BP neural network) showed that tree-based ML can provide the strongest out-of-sample skill for integrated drought monitoring, underscoring the robustness of ensemble learners to heterogeneous predictors and interactions (Wang et al., 2024). At the same time, DL adds value by learning cross-variable representations from multisource data: over South Asia, a deep feed-forward network driven by precipitation, vegetation, and soil factors outperformed distributed random forest and gradient boosting baselines for SMDI prediction across crop phenology stages, achieving high cross-validated agreement with observations (Prodhan et al., 2021). For meteorological drought, ML models have also been shown to estimate SPEI reliably across time scales and regions; for example, a study found gradient-boosted trees and random forests effective for SPEI-3 and SPEI-6, with skill improving when humidity and wind speed were included alongside precipitation and temperature (Mokhtar et al., 2021). In parallel, long-term climate-risk planning increasingly relies on robust trend-analysis frameworks that can evaluate both actual trends and plausible risk-conditioned trends in hydro-meteorological variables, supporting proactive agricultural and water-management decisions under climate change (Keskin et al., 2025). Taken together, these results justify adopting a mixed ML/DL strategy, leveraging ensemble trees and gradient-boosting methods alongside deep networks, to model both SMDI and SPEI for operational drought monitoring and prediction in diverse environments.

Recent Iran-focused research further shows why drought monitoring benefits from combining indicators rather than relying on a single signal. Meteorological drought assessments using standardized indices (SPI/SPEI) have been used to characterize drought patterns over key agricultural regions in southwestern Iran and to relate drought variability to rainfed wheat yield, highlighting the operational value of standardized meteorological drought indicators for agricultural decision-making (Adib et al., 2024). In parallel, analyses across Iran's diverse climate regimes indicate that SPI and SPEI behavior (and their agreement) can vary with climate conditions and time scale, and that data-driven models such as Random Forest can provide useful predictive skill for these indices, often with improved performance at longer accumulation scales and with SPEI offering advantages when atmospheric demand effects are important (Lotfirad et al., 2021). More broadly, comprehensive comparisons emphasize that SPI and SPEI can agree in general variability while still diverging in event-level characteristics and classification, and that the data source (station-based vs. gridded/satellite) can meaningfully affect drought classification—supporting the need for explicit multi-index and multi-source comparison in applied drought assessments (Abu Arra and Şişman, 2024a). Moving beyond meteorological-only characterization, recent work in Iran has explicitly investigated how meteorological drought translates into soil-moisture drought using SPI and SMDI,

quantifying lag behavior and estimating the joint probability of co-occurring meteorological and soil-moisture drought across distinct climatic regions, which supports the value of treating atmospheric forcing and soil-moisture response together for agricultural risk management (Seifian et al., 2025). Consistent with this direction, combined-index frameworks have been proposed that integrate precipitation–temperature-derived indices with soil moisture and vegetation indicators (e.g., NDVI/VHI) over multiple time scales while accounting for temporal-lag dependencies, enabling a more holistic representation of drought dynamics over Iran (Raziei et al., 2025). Beyond Iran, the combined use of meteorological and soil-moisture drought indicators has also been applied to diagnose drought evolution and related hydrologic or ecohydrological behavior; for example, SPEI and SMDI were jointly used to characterize soil-drought evolution and identify thresholds in long-term water-budget dynamics in arid grassland systems (Chen et al., 2024). Additionally, long-term vegetation–soil studies have shown that drought response and soil-water deficits can shift across climate patterns and atmospheric-demand regimes, reinforcing the need to interpret meteorological forcing alongside soil-water behavior rather than relying on a single indicator (Liu et al., 2025). Finally, drought characteristics themselves can shift when different baseline periods or period-selection strategies are used, which motivates time-aware evaluation when defining drought events and comparing drought metrics across long records (Abu Arra et al., 2024b).

From a process perspective, coupling a meteorological drought index with a soil-moisture drought index is theoretically well motivated for agricultural drought monitoring. Meteorological anomalies in precipitation and atmospheric demand typically act as the initiating forcing, while soil moisture integrates these anomalies through storage and buffering and represents the condition most directly linked to crop stress (Van Loon et al., 2024; Zhou et al., 2024). As a result, meteorological drought can precede soil-moisture drought with time-lagged propagation that varies across climatic regimes and land-surface conditions. Using SPEI-3 alongside SMDI therefore supports a complementary monitoring logic in which SPEI-3 provides a standardized early signal of developing water-balance anomaly, while SMDI captures realized root-zone stress and persistence that are more directly related to agricultural impacts.

Accordingly, this study develops a nationwide, time-consistent framework for agricultural drought monitoring across Iran by jointly modeling a meteorological drought indicator (SPEI-3) and a soil-moisture drought indicator (SMDI) from a harmonized multi-

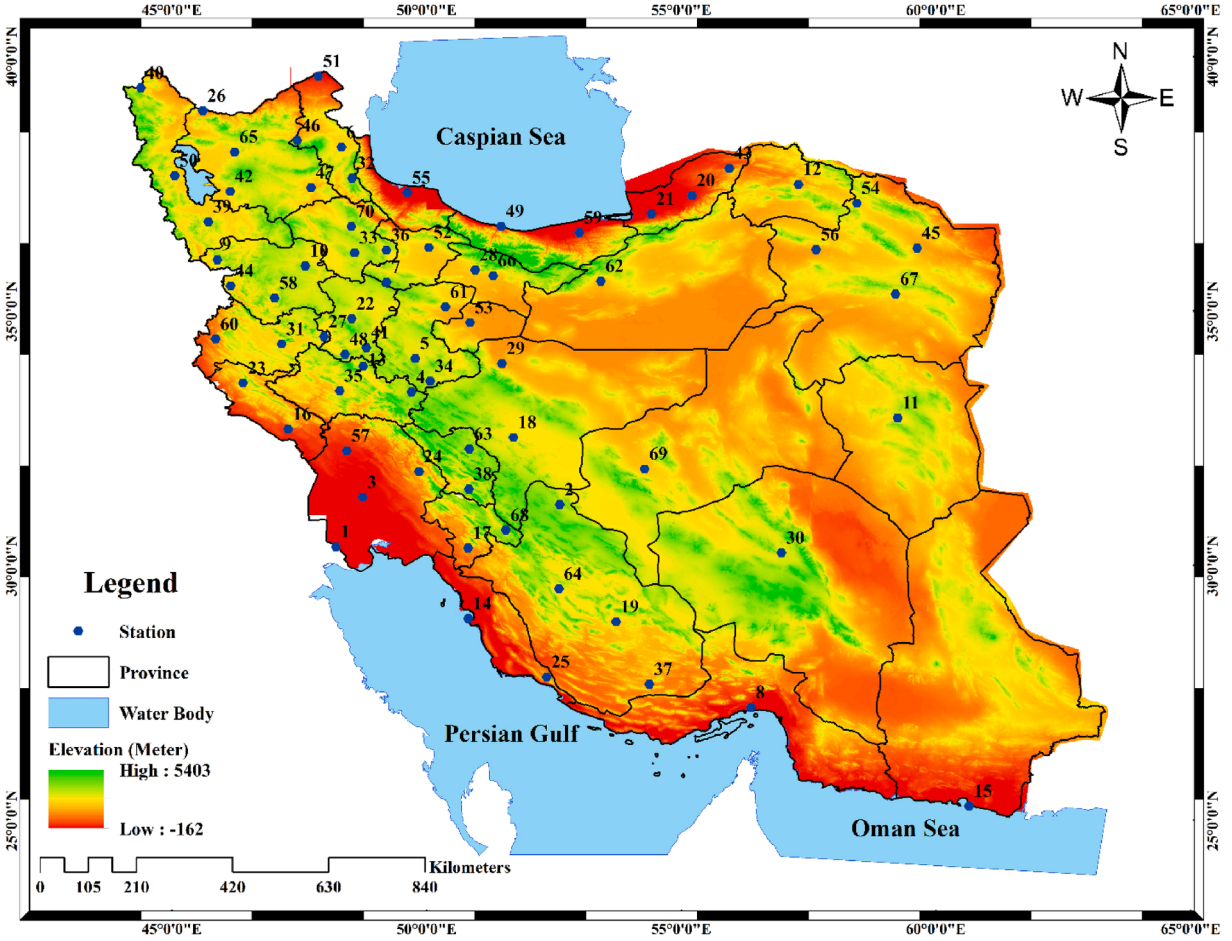

Fig. 1. Study area (Iran) showing provincial boundaries, elevation (m), major water bodies (Caspian Sea, Persian Gulf, and Oman Sea), and the locations of the 70 synoptic stations used in this study.

source predictor set (meteorological stations, MODIS land-surface temperature and vegetation indices, GPM precipitation, and complementary soil-moisture products from FLDAS and C3S). The work is guided by three questions: (Q1) how accurately can SPEI-3 and SMDI be estimated from these multi-source predictors at monthly scale across Iran's diverse climate regimes, (Q2) which model family (Elastic Net, Random Forest, LightGBM, or a Feature Tokenizer Transformer) generalizes most reliably under forward-in-time (time-aware) validation, and (Q3) which predictors and lagged "memory" features most consistently control model skill and interpretability for each drought target. To address these questions, we pursue four objectives: (1) compile and harmonize a nationwide monthly dataset by integrating Earth-observation products with observations from 70 meteorological stations; (2) implement time-aware training/testing to prevent information leakage and provide realistic out-of-sample evaluation; (3) compare complementary ML/DL learners for estimating both SPEI-3 and SMDI using a consistent predictor stack; and (4) develop an operational drought classification that links severity information with early-warning signals derived from the dual-index modeling results. The novelty of this study is fourfold: (i) it provides a national-scale, dual-target modeling system that aligns meteorological and soil-moisture drought on the same temporal scale rather than treating them in isolation; (ii) it introduces a leakage-resistant, forward-chaining validation design for drought ML in Iran to ensure that reported skill reflects true prospective performance; (iii) it leverages complementary soil-moisture evidence streams (reanalysis-based FLDAS and observation-based C3S) within the same framework to reduce single-product dependence and strengthen robustness; and (iv) it couples stability-oriented feature selection (Elastic Net) with interpretable importance profiling and an operational, severity-to-warning classification, yielding a practical and transparent workflow suitable for drought early-warning applications.

2. Methodology

2.1. Study area

Located between 25° and 40° N and 44° and 63° E, Iran is a geographically diverse country of roughly 1.6 million km^2 at the crossroads of Central, South and Middle Asia. The country's landscape—high ridges of the Alborz and Zagros, the salt and sand basins of Dasht-e Kavir and Dasht-e Lut, and more productive corridors along northern and western margins—drives stark differences in climate and agricultural conditions. Those spatial contrasts make uniform drought characterization ineffective, and so accurate, regionally targeted monitoring is essential (Bari Abarghouei et al., 2011).

To address this, we selected a network of 70 weather stations, distributed across Iran's 31 provinces proportionally to each province's cultivation land, and analyzed data from January 2001 to December 2022 at those sites. Using data from Iran's Statistical Centre's, (2023) report, *Share and Rank of Provinces Based on Key Agricultural Sector Variables*, provinces were grouped into four tiers by their share of national agricultural land to prioritize coverage in the most important agricultural areas; accordingly, each province was assigned 4, 3, 2, or 1 meteorological stations, with higher-share provinces receiving more stations. These stations were chosen from a pool of 455 synoptic stations managed by the National Meteorological Organization. The study area and the locations of the selected meteorological stations are shown in Fig. 1. Overall, the final station network represents the best feasible compromise between (i) continuous record availability over 2001–2022, (ii) valid spatial collocation with the remote-sensing products used in this study, and (iii) adequate representation of agriculturally important provinces for agricultural-drought monitoring. As a result, some regions have lower station density due to data and collocation limitations, but the selected set preserves nationwide coverage while prioritizing areas most relevant to agricultural drought impacts.

2.2. Data

Drought results from a web of climatic and environmental drivers. Capturing these drivers in full, then embedding them in predictive models, demands substantial data, computation, and expense. For this work we therefore combined multiple ground observations with satellite-derived drought indices to better represent regional variability across the study area. All variables were collected and evaluated for January 2001–December 2022 which provides a continuous baseline that is sufficient for robust drought assessment at the seasonal SPEI-3 timescale (Abu Arra and Şişman, 2024b).

2.2.1. Remote sensing data

The remote sensing–based datasets used in this study include precipitation, land surface temperature (LST), vegetation/water indices, and soil moisture. Precipitation was obtained from the Global Precipitation Measurement (GPM) IMERG (GSFC, 2023) product with a 30-minute temporal resolution and 0.1° (~11.1 km) spatial resolution using the *PrecipitationCal* band. LST was derived from MODIS (*LST_Day_1km* and *LST_Night_1km*; daily, 1 km) (Wan et al., 2021), and daily mean LST was computed as the average of day and night values. Vegetation/water indices were computed from MODIS surface-reflectance data (daily, 500 m) (Vermote and Wolfe, 2021), including Normalized Difference Vegetation Index (NDVI) (Rouse et al., 1974), Enhanced Vegetation Index (EVI) (Liu and Huete, 1995), Normalized Multi-band Drought Index (NMDI) (Wang and Qu, 2007), and three Normalized Difference Water Index (NDWI) (Gao, 1996) variants calculated separately from the SWIR bands 5, 6, and 7, denoted NDWI5, NDWI6, and NDWI7. Consistently, Normalized Difference Drought Index (NDDI) (Gu et al., 2007) was derived with each NDWI variant, yielding NDDI5, NDDI6, and NDDI7. Soil moisture was characterized using two complementary sources: (i) surface soil moisture (0–10 cm) from the NASA FLDAS Noah v001 product (*SoilMoi00_10cm_tavg*; monthly, 0.1°, volumetric water content, m^3 m^{-3}) (NASA GSFC Hydrological Sciences Laboratory (HSL), 2018), and (ii) root-zone soil moisture represented by the Copernicus Climate Service (C3S) Level-3 Satellite Soil Moisture COMBINED product (*SSMV*; daily, 0.25°, volumetric units, m^3 m^{-3}) (Copernicus Climate Change Service, 2018).

The C3S *SSMV* product was obtained from the Copernicus Climate Data Store (CDS) and collocated to the station coordinates. All other datasets were retrieved through Google Earth Engine (GEE) at the locations of the selected synoptic stations. Gridded predictors were collocated to the station network using station latitude/longitude coordinates. For MODIS-based predictors, values were extracted at each station location using point-based sampling in Google Earth Engine (station-collocated pixel sampling without buffer averaging). In the IMERG precipitation workflow, a *scale* parameter was specified during point extraction to ensure consistent sampling during reprojection. For the C3S soil moisture product provided on a regular latitude–longitude grid, station values were obtained using inverse-distance weighting (IDW) interpolation from the four nearest grid nodes (k = 4, power = 2) based on a KD-tree search, yielding a continuous station-collocated series consistent with the gridded field. To ensure uniformity for modeling, all variables were harmonized to a monthly time step—precipitation was accumulated to monthly totals, and LST, soil moisture, and all indices (NDVI, EVI, NMDI, NDWI5/6/7, NDDI5/6/7) were averaged to monthly means.

2.2.2. Ground-based data

Ground observations were obtained from the data acquisition system of Iran Meteorological Organization. The dataset comprises monthly mean air temperature and monthly accumulated precipitation recorded at the 70 synoptic stations described above, covering the full 22-year study period. A consolidated summary of datasets, sources, spatial/temporal resolution, and preprocessing steps is given in Table 1. Table S1 (Supplementary Material) reports descriptive statistics for the compiled station-collocated monthly variables after outlier screening and quality checks, but before gap filling. At this stage, N may differ slightly among variables because some values were removed as outliers and missing values were not yet filled. In the subsequent preprocessing step, missing values are gap-filled within each station time series using short-gap linear interpolation (≤3 consecutive months) followed by conservative station-median imputation for any remaining gaps.

2.3. Modeling methodology

An overview of the end-to-end workflow, from data acquisition to model outputs and the integrated monitoring step, is shown in Fig. 2.

2.3.1. Clustering

To identify coherent hydroclimatic regimes among the 70 synoptic stations, we clustered three long-term station-level annual summaries: annual precipitation (AP), annual air temperature (AT), and annual soil moisture (ASM). Variables were standardized (zero mean, unit variance) before analysis to avoid scale effects. Details of the 70 synoptic stations are listed in Table S2.

Afterwards, K-means (Hartigan and Wong, 1979) was fit in the standardized space, and candidate solutions K=2–6 were assessed with the silhouette coefficient (Rousseeuw, 1987). The highest silhouette was obtained at K= 3; therefore, a three-cluster solution was adopted. The algorithm was run with multiple random initializations, and the partition with the minimum within-cluster sum of squares was retained. For interpretation, cluster centroids were back-transformed to the original units (AP: mm yr^{-1}; AT: °C; ASM: m^3 m^{-3}). Results are summarized in Table 2.

Distributional contrasts by cluster are summarized in Fig. 3a–c: Cluster 1 corresponds to wetter/ cooler stations with higher ASM; Cluster 2 captures arid to semi-arid stations with low AP, warm AT, and low ASM; and Cluster 3 comprises cooler stations with moderate AP and comparatively higher ASM than the dataset mean. As a complementary visualization, the standardized station data were projected onto the first two principal components, with cluster membership overlaid in Fig. 4, which illustrates clear multivariate separation in the joint AP–AT–ASM space. In this study, clustering was used to capture broad agro-climatic heterogeneity across Iran and to provide the models with a compact representation of regional differences. The resulting cluster memberships were encoded as dummy variables and included as predictors, allowing the learning algorithms to model cluster-specific baselines and relationships and improving nationwide generalization.

Table 1
Summary of datasets used in the study.

Category	Variable(s) / Index	Product (Band)	Provider / Platform	Native resolution (time; space)	Analysis scale	Units
Remote sensing	Precipitation	GPM IMERG (*PrecipitationCal*)	NASA/JAXA – GEE	30 min; 0.1° (~11.1 km)	Monthly totals	mm
	Land Surface Temperature (LST)	MODIS *LST_Day_1km*, *LST_Night_1km*	NASA – GEE	Daily; 1 km	Monthly mean	°C
	NDVI, EVI, NDWI5/6/7, NDDI5/6/7, NMDI	MODIS Surface Reflectance	NASA – GEE	Daily; 500 m	Monthly mean	dimensionless
	Surface soil moisture (0–10 cm)	FLDAS Noah v001 (*SoilMoi00_10cm_tavg*)	NASA/GSFC – GEE	Monthly; 0.1°	Monthly (native)	m^3 m^{-3}
	Surface soil moisture (2-5 cm)	C3S Level-3 SSMV COMBINED	Copernicus CDS	Daily; 0.25°	Monthly mean	m^3 m^{-3}
Ground-based	Air temperature	Synoptic stations	Iran Meteorological Organization	Monthly; point	Monthly (native)	°C
	Precipitation	Synoptic stations	Iran Meteorological Organization	Monthly; point	Monthly (native)	mm

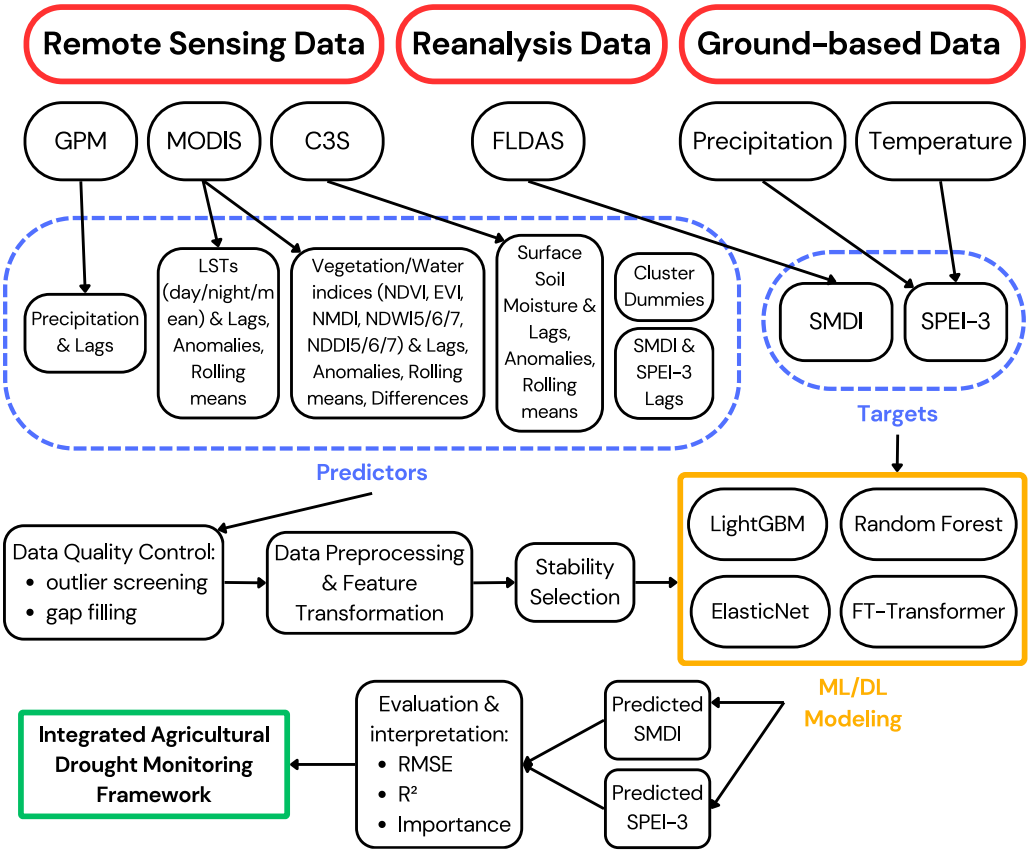

Fig. 2. End-to-end workflow of the methodology used in this study.

Table 2
Mean hydroclimatic characteristics by cluster (22-year averages).

Cluster	Stations (n)	AP (mm)	AT (°C)	ASM ($m^3\ m^{-3}$)
1	6	945.13	15.68	0.22
2	23	227.50	22.26	0.18
3	41	323.29	14.15	0.21

AP: annual precipitation; AT: annual air temperature; ASM: annual soil moisture. Values rounded to two decimals.

2.3.2. *Standardized precipitation–evapotranspiration index (SPEI)*

The Standardized Precipitation–Evapotranspiration Index (SPEI) quantifies drought as anomalies in the climatic water balance evaluated over user-defined time scales and transformed to a standard normal variate, which enables consistent comparison across space and time (Vicente-Serrano et al., 2010). Negative values denote dry conditions and positive values wet conditions; magnitudes ≤ -1 typically indicate the onset of meteorological/agricultural drought (Feng et al., 2019).

Given that only monthly air temperature was available, PET was estimated using the Thornthwaite method, including latitude-based daylength corrections (Thornthwaite, 1948). For each station, the climatic water balance was computed as $D_j = P_j - PET_j$, and then accumulated to the 3-month timescale as $D_j^{(3)} = D_j + D_{j-1} + D_{j-2}$to obtain SPEI-3. To avoid losing early-record months, $D^{(3)}$for January–February 2001 was initialized using station-wise climatological padding for the antecedent months. Standardization followed the canonical SPEI procedure: for each station and calendar month, the distribution of $D^{(3)}$across years was modeled using a log-logistic (Fisk) distribution. When the monthly sample size was insufficient for stable fitting, we used a conservative fallback by pooling adjacent months or fitting using the full station record. The fitted cumulative probability $F(D_j^{(3)})$was then transformed to standard normal space as $\mathrm{SPEI}-3_j = \Phi^{-1}\left(F(D_j^{(3)})\right)$, where $\Phi^{-1}$is the inverse standard normal CDF.

In this study, SPEI is the target variable for one of our modelings. We used SPEI (rather than a precipitation-only index) because it incorporates atmospheric evaporative demand through PET, which is important in arid and warming environments where drought impacts can intensify even under moderate precipitation deficits. We computed station-based monthly SPEI at a 3-month accumulation

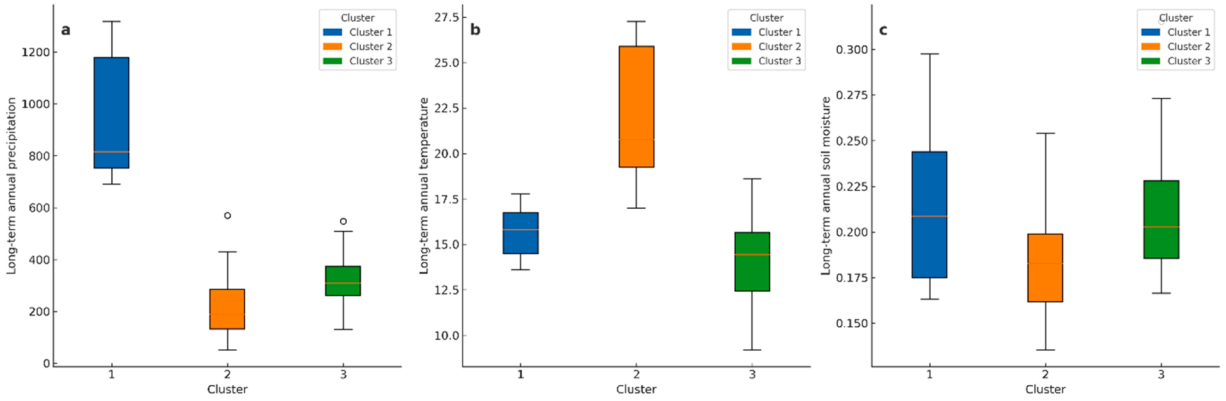

Fig. 3. Distribution of long-term annual (a) precipitation (mm yr^{-1}), (b) air temperature (°C), and (c) soil moisture (m^3 m^{-3}) across the three K-means clusters. Boxes show the interquartile range (Q1–Q3) with the median; whiskers extend to 1.5×IQR; points beyond whiskers denote outliers. Cluster membership is indicated by the plot legend.

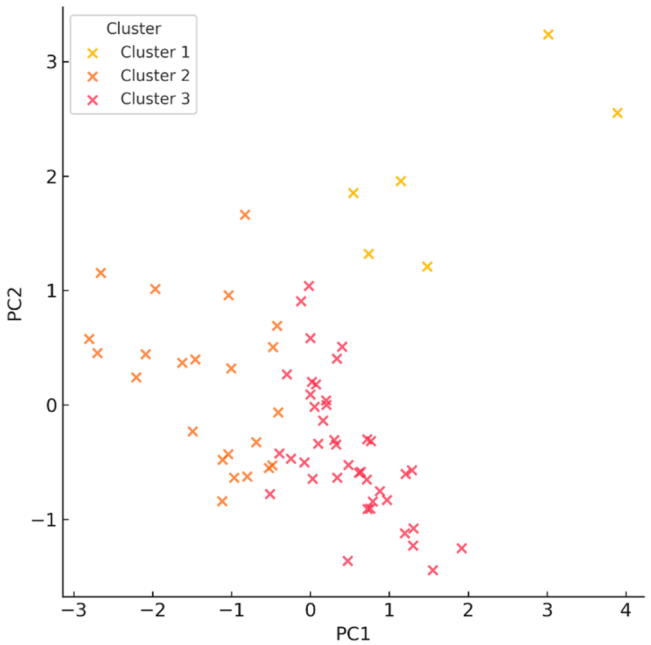

Fig. 4. Station scores on the first two principal components (PC1 and PC2) computed from standardized annual precipitation (AP), air temperature (AT), and annual soil moisture (ASM). Points are grouped by K-means cluster membership (Clusters 1–3), illustrating multivariate separation among climatic regimes.

(SPEI-3) using ground-observed precipitation and temperature. The 3-month Standardized Precipitation Evapotranspiration Index (SPEI-3) is widely recognized as the most effective timescale for monitoring agricultural drought because it shows the strongest statistical correlation with crop yield variations and accurately reflects moisture deficits during critical vegetative growth stages (Labudová et al., 2017).

2.3.3. Soil moisture deficit index (SMDI)

The Soil Moisture Deficit Index (SMDI) quantifies agricultural drought as anomalies in available soil water and incorporates persistence via a recursive update, yielding a dimensionless scale typically spanning −4 (extreme deficit) to + 4 (extreme wetness) (Narasimhan and Srinivasan, 2005). SMDI values nearer to zero reflect conditions close to the long-term norm, while increasingly negative (positive) values represent intensifying dryness (wetness): SMDI ≤ -2 indicates Extreme Dry conditions; −2 to −1.5 Severe Dry; −1.5 to −1 Moderate Dry; −1–1 Near Normal; and > 1 denotes Wet conditions (Prodhan et al., 2021).

In this study, SMDI serves as the target variable for the second modeling. We computed monthly SMDI at the coordinates of the 70 stations using the NASA FLDAS Noah v001 surface soil-moisture series, collocated and harmonized to a monthly time step. FLDAS was

selected for target construction to maximize accuracy and consistency of the dependent variable; the Copernicus C3S soil-moisture product was retained solely as an independent predictor (to avoid target–predictor circularity and to preserve model generalization). The SMDI calculation followed the Narasimhan–Srinivasan procedure adapted to monthly data, producing a continuous index on the −4 to +4 scale used for modeling and interpretation.

2.3.4. Predictive modeling strategy

To evaluate agricultural drought as both an atmospheric forcing and a soil-moisture response, we trained two separate supervised regression models with distinct targets. The first model estimates SPEI-3, a short-to-seasonal meteorological drought indicator that reflects emerging imbalances in climatic water supply and demand and can therefore serve as an early-warning signal. The second model estimates SMDI, which represents realized root-zone moisture stress and is more directly related to agricultural impacts.

Keeping SPEI-3 and SMDI as separate targets improves clarity and avoids conflating processes with different dynamics. In particular, (i) each target can use a process-consistent predictor set (e.g., thermal/vegetation and precipitation signals for SPEI-3 versus moisture and land-surface controls for SMDI), (ii) each model can incorporate appropriate temporal memory through lagged predictors and lagged targets, and (iii) the resulting outputs can be combined into an operational framework in which SMDI anchors drought severity while SPEI-3 provides escalation/early warning when atmospheric forcing indicates impending deterioration.

Across all algorithms, both targets follow the same leakage-resistant temporal evaluation design: predictors are harmonized to monthly station-collocated series, preprocessing and feature transformations are applied in a forward manner (fit on training data and applied to later periods), stability selection is used to identify robust predictors, and final performance is assessed on a chronologically held-out test window (Sections 2.3.5–2.3.11).

2.3.5. Data preprocessing and feature transformation

Building on the dual-model design, we prepared a unified, time-aware input space that preserves physical meaning while enabling robust learning. All variables were harmonized to a monthly time step and aligned per station. Feature engineering emphasized the multi-scale dynamics of agricultural drought: temporal persistence and delayed responses were represented with lagged predictors drawn from meteorological, radiative, vegetation, and soil-moisture fields; seasonal cycles were removed by expressing continuous predictors as departures from each station's long-term monthly climatology; inter-month differences were added to capture emerging stress or recovery; short rolling aggregates summarized recent hydroclimatic context; and the K-means cluster label (Section 2.3.1) was encoded to allow a pooled model to adapt to regime-specific relationships.

To make the engineered time scales explicit, we used i) short lag features at 1- and 2-month lead for selected hydroclimatic predictors to represent delayed propagation from atmospheric anomalies to vegetation and soil-water response, consistent with evidence that meteorological drought signals can translate into soil-moisture deficits with short lags and that land-surface states retain memory over month-scale horizons. ii) We also used a 3-month rolling aggregation (window length = 3 months) for selected predictors to summarize recent seasonal context, which is particularly relevant for agricultural drought monitoring and aligns with the short-to-seasonal time scale emphasized by SPEI-3. These choices are consistent with drought-propagation syntheses and probabilistic propagation analyses that explicitly evaluate short (≈1-month) versus seasonal (≈3-month) meteorological forcing when conditioning soil-moisture drought behavior, and with broader soil-moisture-memory literature showing persistence over monthly time scales (Seifian et al., 2025; Zhou et al., 2024; Zhu et al., 2021).

With the predictor set established, transformation proceeded in two sequential, leakage-safe steps designed to stabilize distributions and place heterogeneous variables on a common scale. First, a Yeo–Johnson power transform (Yeo and Johnson, 2000) was applied to each continuous predictor to mitigate skewness and stabilize variance while accommodating zeros and negatives, which are commonly used in drought and hydrologic time-series preprocessing (Erhardt and Czado, 2018; Kabbilawsh et al., 2022; Marcos Junior et al., 2024). This step reduces the influence of heavy tails and makes linear relationships closer to homoscedastic. Second, features were standardized to zero mean and unit variance so that coefficients are comparable across predictors and Elastic Net's combined L1/L2 regularization can operate even-handedly, improving numerical conditioning and interpretability. Both transforms were fit on the training horizon and then applied forward in time to the hold-out period to preserve temporal integrity.

Within this standardized feature space, the target-specific predictor choices follow process logic. For the SMDI model, contemporaneous and short-lag SPEI-3 were included because SMDI reflects the realized soil-moisture state, which is driven by antecedent imbalances in atmospheric supply and demand; leveraging SPEI-3 thus provides a physically consistent bridge from meteorological to agricultural drought and improves lead information on impending root-zone deficits. For both targets, a one-step lag of the target itself was incorporated ($SMDI_{t-1}$ for SMDI; $SPEI3_{t-1}$ for SPEI-3) to represent inherent persistence arising from storage and memory in soils and the climate system. These autoregressive terms reduce residual autocorrelation, sharpen short-term predictability, and enhance the stability of the learned relationships while using only information available from prior months. Incorporating lagged values of the target index as predictors is a common practice across drought-index forecasting (e.g., SPI/SPEI), autoregressive neural models, and broader hydrologic time-series prediction (e.g., streamflow), where historical target lags are routinely used to stabilize learning and improve forecast skill (Danandeh Mehr et al., 2022; Elbeltagi et al., 2022; Maca and Pech, 2016; Magallanes-Quintanar et al., 2023; Moges et al., 2024; Morid et al., 2007).

2.3.6. Stability-selected predictors and model drivers

To obtain a compact yet robust predictor set for both targets, we embedded stability-selection within an Elastic Net framework and applied it to the harmonized, transformed feature space (Section 2.3.5). Starting from the full library of covariates (lags, monthly anomalies, inter-month differences, short rolling aggregates, vegetation/water indices, soil-moisture metrics, and regime indicators),

we first removed near-constant signals (post-scaling variance $< 1 \times 10^{-5}$) and pruned collinearity by collapsing any pair with $|r| > 0.95$, retaining the higher-variance member. We then fit Elastic Net models with the mixing parameter fixed at 0.5 over a logarithmic grid of penalties tuned to yield ~15 nonzero coefficients. Variable-selection stability was assessed via 50 station-wise subsamples (70% of stations per draw); predictors that received nonzero coefficients in $\geq$ 60% of refits were retained. In each iteration, we sampled 50 stations without replacement at the station level and retained the full monthly time series for those stations, so stability is assessed with respect to spatial resampling rather than resampling individual months. Predictors were retained if they appeared as nonzero coefficients in at least 60% of iterations, which acts as a majority-stability criterion to prioritize robust signals while avoiding overly strict thresholds that can discard meaningful predictors under correlated features. Hydroclimatic regime dummies (Clusters 1–3) were left unpenalized so the models could adjust for spatial heterogeneity.

This procedure yielded parsimonious, physically consistent predictor sets. For both SMDI and SPEI-3 models, the retained features are presented in Table S3. The coefficient magnitudes of SMDI (Fig. 5-a) show a dominant contribution from $SMDI_{t-1}$, highlighting strong soil-moisture memory, with precipitation and daytime-LST anomalies as the next most influential drivers; vegetation and water-index anomalies supply complementary information on canopy condition and surface water availability. Absolute coefficients of SPEI-3 (Fig. 5-b) indicate that SPEI-3_{t-1} is the largest contributor, followed by soil-moisture and land surface temperature anomalies which are consistent with antecedent storage and energy-controlled evapotranspiration shaping the climatic water-balance signal. Short-lag precipitation terms and vegetation/water indices carry secondary weight and refine short-to-seasonal wetness anomalies.

2.3.7. LightGBM (gradient-boosted decision trees)

LightGBM is a gradient-boosted decision-tree method that models the target as an additive ensemble of shallow trees fitted sequentially to reduce residual error. Its histogram-based split finding and leaf-wise growth (with depth constraints) enable efficient learning of nonlinear responses and feature interactions in tabular hydroclimatic datasets (Ke et al., 2017).

In this study, LightGBM was used for regression to estimate both SPEI-3 and SMDI over the full station–month panel (January 2001–December 2022) under a strictly forward-in-time protocol. We reserved the most recent fifth of the record as a single held-out test window and performed all model selection only within the earlier four-fifths. Within the training portion, we used a time-aware expanding cross-validation with five contiguous temporal blocks: at each split, the model was trained on all earlier blocks and validated on the immediately following block. The same temporal partitions were applied simultaneously across all stations to avoid look-ahead and to preserve the chronological structure of the station network.

Hyperparameters were tuned by randomized search under a squared-error objective, with the selection criterion defined as the mean validation RMSE across the expanding folds. The search jointly controlled learning rate and boosting budget, tree complexity (maximum depth and number of leaves), and regularization/variance control (minimum samples per leaf, L2-type penalties on leaf weights, and row/feature subsampling). To prevent overfitting and to standardize training across trials, we enabled early stopping: training stopped if validation error did not improve for 100 consecutive boosting iterations, and the iteration achieving the minimum validation error was retained. The randomized search sampled discrete combinations from the following candidate sets: learning rate {0.03, 0.05, 0.10}, maximum tree depth {no limit, 5, 10}, maximum number of leaves {31, 63, 127}, feature fraction (column subsampling) {0.6, 0.8, 1.0}, bagging fraction (row subsampling) {0.7, 0.9}, bagging frequency fixed at 1, minimum number of samples per leaf {20, 50, 100}, L2 regularization strength {0, 1, 5, 10}, and L1 regularization strength {0, 0.1, 1}. During tuning, the maximum number of boosting iterations was set to 5000 with early stopping of 100 rounds; 15 sampled configurations were evaluated per target using a fixed random seed for reproducibility, and the best configuration was selected by minimizing the mean cross-validated root mean squared error under the time-aware expanding cross-validation scheme.

After identifying the best configuration for each target, we refit LightGBM once on the full training period and evaluated performance a single time on the untouched test window. Reproducibility was ensured by fixed random seeds and deterministic settings.

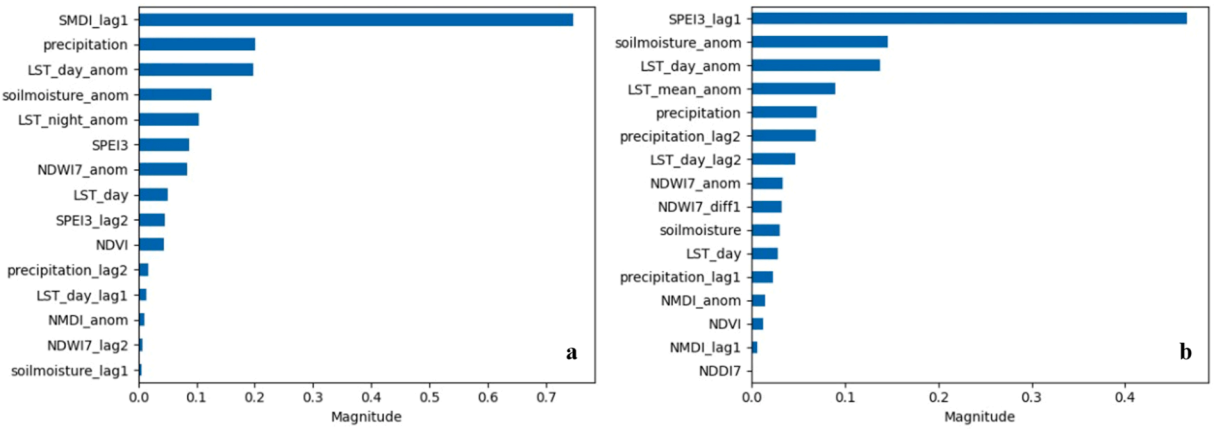

Fig. 5. Absolute coefficient magnitudes from Elastic Net models after stability-selection: (a) SMDI model and (b) SPEI-3 model. Predictors were standardized and bars show |coefficients|.

2.3.8. Elastic net (linear models with mixed L1–L2 regularization)

Elastic Net (ENet) is a penalized linear machine learning regression that combines the sparsity of the lasso with the shrinkage and grouping behavior of ridge regression, making it suitable for correlated hydroclimatic predictors while retaining interpretability (Zou and Hastie, 2005). In words, ENet estimates coefficients by minimizing the usual least-squares error augmented with a penalty that blends an L1 term (encouraging some coefficients to be exactly zero) and an L2 term (shrinking correlated coefficients together). The overall penalty strength and the lasso–ridge mixing ratio govern, respectively, how strongly coefficients are regularized and how much sparsity versus grouping is preferred.

To complement the nonlinear ensemble results, we fit ENet for both targets (SMDI and SPEI-3) under the same strictly forward-in-time protocol as Section 2.3.7. The most recent fifth of the record was reserved once as a held-out test window; all model selection took place only in the earlier four-fifths using a time-aware expanding cross-validation with five contiguous calendar blocks and no embargo. The same calendar partitions were applied simultaneously to all stations to avoid look-ahead. Predictors had already been standardized within training folds, so no additional scaling was applied. Rows with missing values arising from initial lags at the start of each station series were removed prior to fitting to maintain temporal integrity.

Regularization strength and the lasso–ridge mixing ratio were tuned by coordinate descent over a dense, log-spaced grid for the penalty magnitude (from 10^{-4} to 10^{1}) crossed with a broad set of mixing ratios {0, 0.1, 0.3, 0.5, 0.7, 0.9, 1}, including pure ridge and pure lasso. The selection criterion was the mean RMSE across the expanding folds. After selection, coefficients were refit once on the full training portion and evaluated a single time on the held-out window. Convergence safeguards allowed up to 20,000 coordinate-descent iterations with a tight stopping tolerance, and an intercept was retained in all fits to prevent bias when targets are centered but not exactly mean-zero. Because all predictors are on a common scale, ENet coefficients can be read directly for effect size and sign, supporting model interpretation alongside the stability-selection results in Section 2.3.6.

2.3.9. Random forest (bagged decision trees)

Random Forest (RF) is a non-parametric ensemble of decision trees trained on bootstrap samples, with random feature subsetting at each split to decorrelate trees and reduce variance (Breiman, 2001). By aggregating predictions across many diverse trees, RF captures nonlinear responses and higher-order interactions, remains insensitive to monotonic transformations of inputs, and is comparatively robust to multicollinearity and moderate hyperparameter misspecification.

We fit RF regressors for both targets (SMDI and SPEI-3) under the same strictly forward-in-time protocol used elsewhere: an 80/20 chronological split with the most recent fifth reserved once as a held-out test window, and model selection confined to the earlier four-fifths via time-aware expanding cross-validation with five contiguous calendar blocks and no embargo. Each model used the stability-selected predictors identified earlier together with the unpenalized regime indicators (Clusters 1–3); SPEI-3 was retained as a predictor in the SMDI model to provide a physically consistent link from meteorological to agricultural drought. Inputs had already been standardized during preprocessing, though RF does not require scaling; rows with initial lag-induced missing values were excluded before fitting to preserve temporal integrity.

Hyperparameters were tuned only on the training portion by minimizing the mean RMSE across expanding folds. The candidate set covered the number of trees, tree depth/size controls, the fraction of features considered at each split, and bootstrap sampling. The selected configuration was refit on the full training window and evaluated once on the untouched test window. Performance was summarized with RMSE.

2.3.10. Feature-Tokenizer transformer

FT-Transformer is a deep learning architecture that adapts the Transformer to tabular data by first tokenizing features, mapping each numerical feature through a learnable linear embedding and each categorical feature through an embedding lookup, then processing the resulting tokens with an encoder-only Transformer stack. A special [CLS] token aggregates information across features and its final representation is used for prediction. FT-Transformer's multi-head self-attention lets the model capture rich, non-linear feature–feature interactions while residual connections and layer normalization stabilize training (Gorishniy et al., 2021).

We trained an FT-Transformer tailored to monthly tabular time series. Temporal information was represented through the engineered lagged and rolling predictors described in Section 2.3.5, so the Transformer receives a single monthly feature vector per station-month rather than a fixed-length input window. Each monthly feature vector was tokenized feature-wise and passed through the Transformer encoder, with a learnable station identifier embedding included as a fixed token to capture site-specific effects without leaking target information. Modeling followed the same strictly forward-in-time protocol as the other algorithms: the most recent fifth of the record served once as a held-out test window, while all model selection occurred in the earlier four-fifths using expanding-window (walk-forward) time-aware cross-validation applied synchronously across stations (no embargo), so the training set grows with time and validation is always performed on later chronological blocks.

Hyperparameters were tuned with Optuna using pruning on the expanding cross-validation objective, following recent drought and hydrologic studies that employ Optuna to improve forecast skill and efficiency (Ahire et al., 2025; Huang et al., 2025; Islam et al., 2025; Train et al., 2024; Uddin et al., 2024; Xiao et al., 2024). The search explored embedding width {96,128,160,192}; number of encoder blocks {2,3,4}; attention heads {4,8}; dropout in {0.08, 0.12, 0.18, 0.25}; learning rate in $[4 \times 10^{-4}, 2 \times 10^{-3}]$; L_2 weight decay {0, 10^{-5}, 10^{-4}, 5×10^{-4}}; and batch size {256,384,512}. The selected configuration was refit on the full training span and evaluated once on the untouched test window. To aid generalization, we combined dropout, weight decay, and early pruning; training used an adaptive optimizer with the searched learning-rate and weight-decay settings, and inference relied on the [CLS] representation for scalar regression outputs.

2.3.11. Model performance assessment

In this study, predictive skill on the held-out test window was quantified using the Root Mean Squared Error (RMSE; Eq. 1) (Hastie et al., 2009) and the Coefficient of Determination (R^2; Eq. 2) (Draper and Smith, 1998), computed in the original (dimensionless) units of the targets. Let O_i and S_i denote observed and simulated values for $i = 1, \ldots, n$, and $\overline{O}$, $\overline{S}$ their sample means.

$$\mathrm{RMSE} = \sqrt{\left(\frac{1}{n}\right)\sum_{i=1}^{n}(O-S)^2} \tag{1}$$

$$R^2 = \frac{\left(\sum_{i=1}^{n}(O-\overline{O})(S-\overline{S})\right)^2}{\sum_{i=1}^{n}(O-\overline{O})^2 * \quad \sum_{i=1}^{n}(S-\overline{S})^2} \tag{2}$$

Lower values of RMSE indicate smaller typical errors. With this definition, $R^2 \in [0,1]$ measures the strength of linear association between observations and predictions, where higher values indicate better agreement. These same metrics were used during the expanding cross-validation for model selection; final scores are reported once on the untouched test window, with optional summaries by station and by cluster.

3. Results

3.1. Comparative predictive performance on the held-out test window

Table 3 compares generalization performance for four model families (LightGBM, Elastic Net, Random Forest, and FT-Transformer) on the held-out test window (most-recent 20% of the record). Across all methods, predictive skill is consistently higher for SMDI than for SPEI-3. LightGBM yields the best overall accuracy for SMDI (RMSE = 0.800; R^2 = 0.821), while Random Forest, FT-Transformer, and Elastic Net are close behind (RMSE ≈ 0.818–0.825; R^2 ≈ 0.81). For SPEI-3, performance decreases for every algorithm and the ranking compresses: LightGBM and Random Forest are nearly tied (RMSE ≈ 0.96; R^2 ≈ 0.55), and FT-Transformer and Elastic Net trail slightly (RMSE ≈ 1.01; R^2 ≈ 0.50). Because LightGBM and Random Forest are close in Table 3, we tested whether their differences are statistically meaningful using paired comparisons on the same held-out station–month observations. For SMDI, LightGBM has significantly lower error than Random Forest (paired *t*-test on per-sample squared errors: $t = 5.47$, $p = 4.7 \times 10^{-8}$; results are consistent using absolute errors), whereas for SPEI-3 the difference is not statistically significant ($p = 0.62$), indicating essentially tied performance.

This separation between targets is expected and physically interpretable: SMDI integrates storage deficits and thus exhibits stronger persistence and smoother dynamics, whereas SPEI-3 emphasizes short-to-seasonal atmospheric imbalance and contains rarer extremes, which are harder for pointwise regression models trained under squared-error loss to capture.

To directly test regional adaptability, we further evaluated LightGBM performance stratified by the three hydroclimatic clusters on the held-out test window. This analysis is performed without training separate regional models, because cluster membership is already included as dummy predictors in the unified national model; the stratified metrics therefore quantify whether residual skill remains balanced across regimes under the same tuned configuration. Cluster-wise results show consistently strong accuracy for SMDI, with RMSE ranging from 0.738 to 0.834 and R^2 from 0.792 to 0.852 across the three clusters, indicating stable performance over distinct climatic settings. For SPEI-3, cluster-wise RMSE ranges from 0.702 to 0.998 and R^2 from 0.520 to 0.577, with slightly higher errors in Clusters 2–3 but no breakdown in predictive skill. The corresponding observed–predicted panels by cluster are provided in the Supplementary Material to visually confirm these patterns (Figures S3-S8).

Because operational drought monitoring prioritizes rare high-impact conditions, we also assessed predictive behavior under extreme subsets on the held-out test window, namely SMDI ≤ -2 and SPEI-3 < -1.5. Under these subsets, the model remains directionally consistent with observed drought variability but becomes more conservative in magnitude, which is expected when extreme events are infrequent relative to the full record. For SMDI ≤ -2, RMSE is 0.868 ($n = 486$) with a negative mean residual (-0.599), indicating that the most severe soil-moisture deficits tend to be predicted as slightly less extreme. For SPEI-3 < -1.5, RMSE is 2.279 ($n = 449$) with a mean residual of -1.704, reflecting the higher volatility and tail sensitivity of meteorological drought at short accumulation scales. In this context, RMSE and residual bias are the more informative diagnostics than variance-normalized measures within a truncated tail subset.

Table 3
Test-set performance of tuned models on the held-out (most-recent 20%) window. Metrics are dimensionless; boldface marks the best score per target/metric.

Algorithm	SMDI RMSE	SMDI R^2	SPEI-3 RMSE	SPEI-3 R^2
LightGBM	**0.800**	**0.821**	**0.962**	**0.547**
Elastic Net	0.825	0.810	1.011	0.499
Random Forest	0.818	0.813	0.963	0.545
FT-Transformer	0.824	0.810	1.008	0.502

3.2. Error behavior and reliability across drought regimes

Figs. 6–9 & S1-S2 provide complementary diagnostics (observed vs. predicted scatter, residuals through time, and the empirical CDF of absolute errors). For SMDI, all models show dense alignment around the 1:1 line, with dispersion widening primarily in distribution tails, indicating that most month-to-month variability is captured regardless of algorithm. Residual-over-time panels remain centered without sustained drift, supporting time-stable generalization over the held-out period; occasional bursts of dispersion coincide with intervals of stronger interannual variability. Consistently steep CDF curves for SMDI indicate that the majority of predictions fall within modest error bands.

For SPEI-3, the observed-versus-predicted diagnostics show broader spread and a recurring understatement of the most negative values (extreme dry months predicted too mildly). This "regression toward the mean" appears across model families and is consistent with both the rarity of extremes and the use of squared-error objectives. The CDF curves climb more slowly and display heavier right tails than for SMDI, indicating that a smaller fraction of months contributes disproportionately to error.

3.3. Interpreting dominant drivers and cross-model consistency

Despite differences in learning mechanisms, driver attribution converges on a coherent set of controls (Fig. 10 and Fig. 11). For both targets, short-term persistence is dominant ($SMDI_{t-1}$ for SMDI; $SPEI\text{-}3_{t-1}$ for SPEI-3). Next in importance are precipitation and land-surface temperature anomalies, followed by soil-moisture anomalies and vegetation/water indices (e.g., NDWI/NDDI/NDVI), which refine the signal and help represent land-surface response. This hierarchy aligns with process understanding: antecedent conditions provide monthly memory, while net water supply and evaporative demand govern departures from normal, with land-surface indicators adding sensitivity to realized stress.

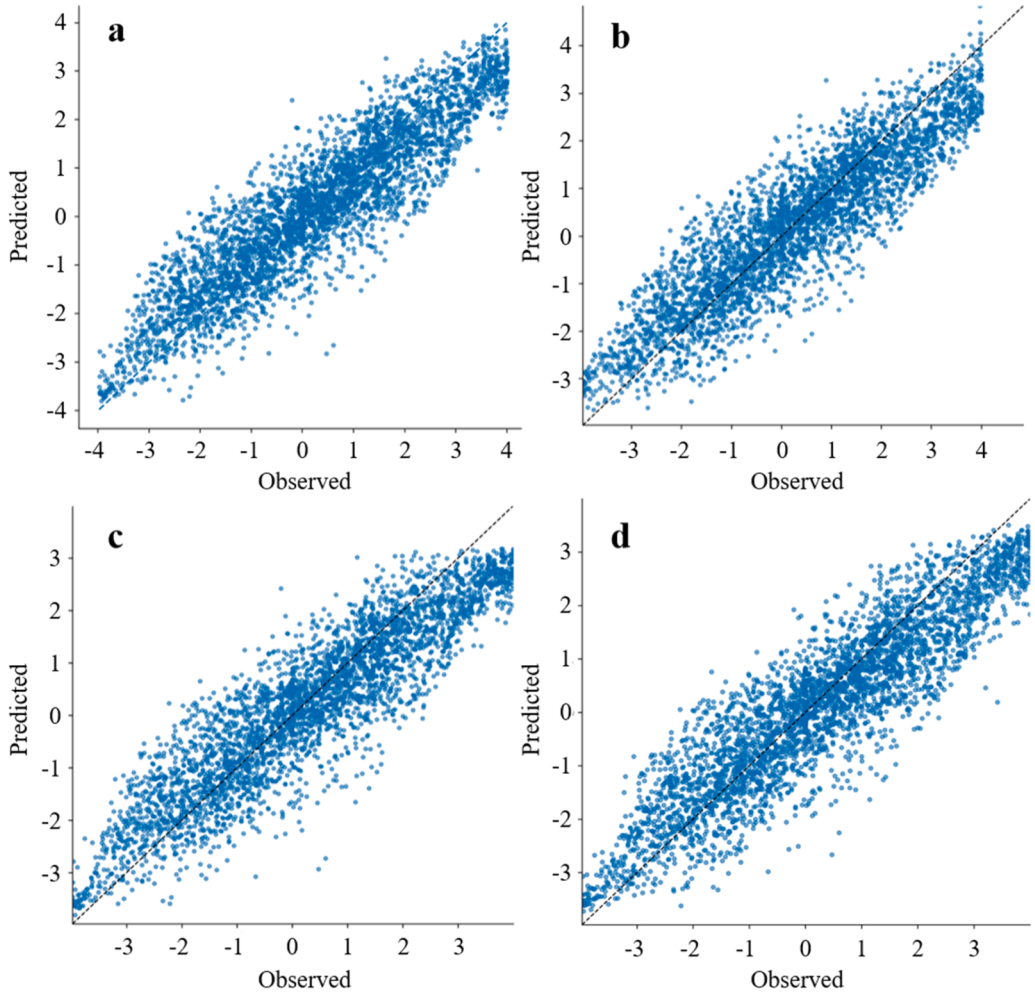

Fig. 6. SMDI observed versus predicted values on the held-out test window (most recent 20% of the monthly record) across all stations. The dashed line indicates the 1:1 reference. Panels: (a) LightGBM, (b) Elastic Net, (c) Random Forest, and (d) FT-Transformer.

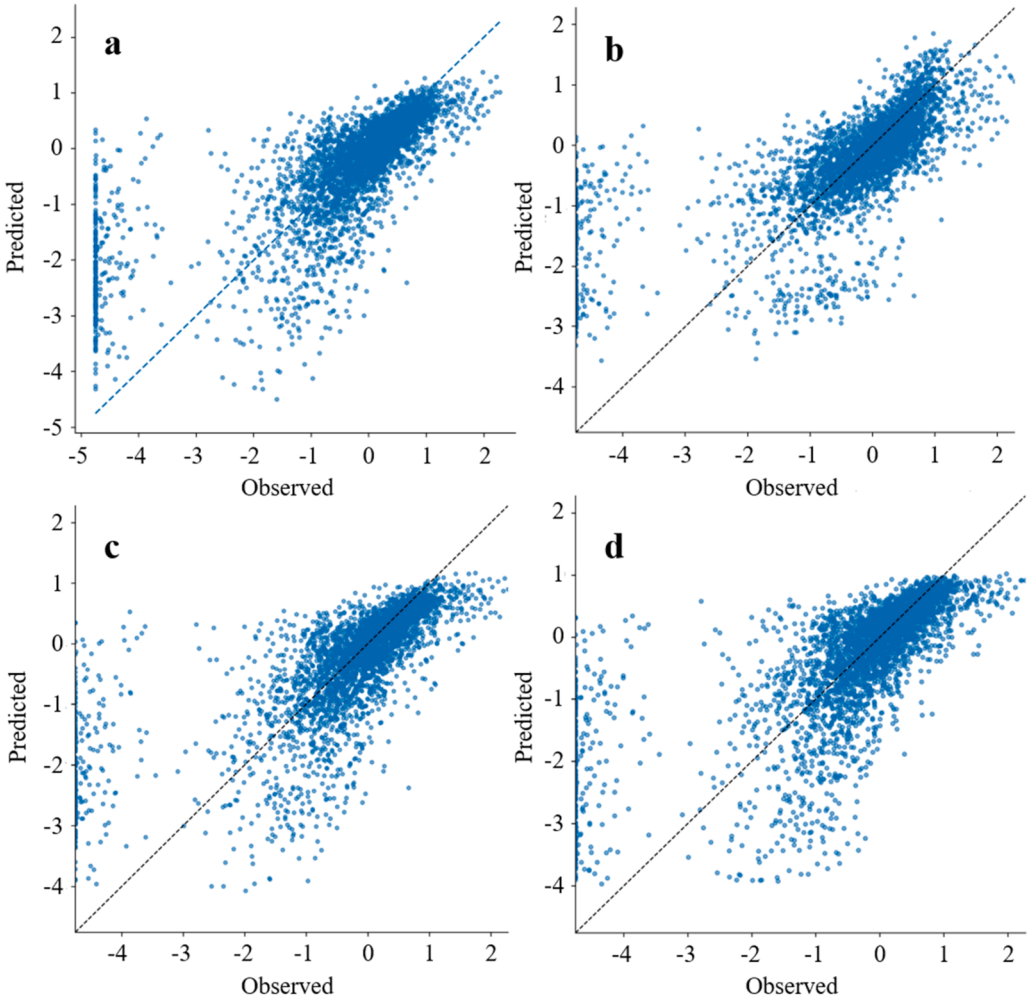

Fig. 7. SPEI-3 observed versus predicted values on the held-out test window (most recent 20% of the monthly record) across all stations. The dashed line indicates the 1:1 reference. Panels: (a) LightGBM, (b) Elastic Net, (c) Random Forest, and (d) FT-Transformer.

Cluster indicators contribute modest but non-zero signal. Their small, consistent contributions suggest that clustering captures spatial heterogeneity without dominating prediction, which is desirable for pooled, station-based monitoring.

3.4. Tuned configurations

3.4.1. LightGBM

For SMDI, the best LightGBM configuration used learning rate 0.03, max depth 5, up to 127 leaves, minimum 50 observations per leaf, predictor subsampling 0.60, row subsampling 0.90, and modest L2 regularization; the cross-validated optimal iteration count averaged ≈ 524 and was fixed for the final refit. For SPEI-3, the final configuration used learning rate 0.03 and depth 5 with 31 leaves, the same minimum leaf size and subsampling, and slightly stronger ridge penalty; the average optimal iteration across folds was ≈ 384.

3.4.2. Elastic net

For SMDI, the optimal setting used a modest penalty with a small lasso component (penalty ≈ 1.76×10^{-2}, $\alpha = 0.1$). For SPEI-3, the search favored essentially pure ridge regularization with a very small penalty (penalty = 10^{-4}, $\alpha = 0$).

3.4.3. Random forest

The tuned RF configuration (used for both targets) comprised 500 trees, maximum depth 15, minimum samples to split 10, minimum samples per leaf 4, 60% of predictors considered at each split, with bootstrap sampling enabled.

3.4.4. FT-Transformer

For SMDI, the best FT-Transformer used embedding dimension 160, 3 encoder blocks, 8 attention heads, dropout ≈ 0.12, learning

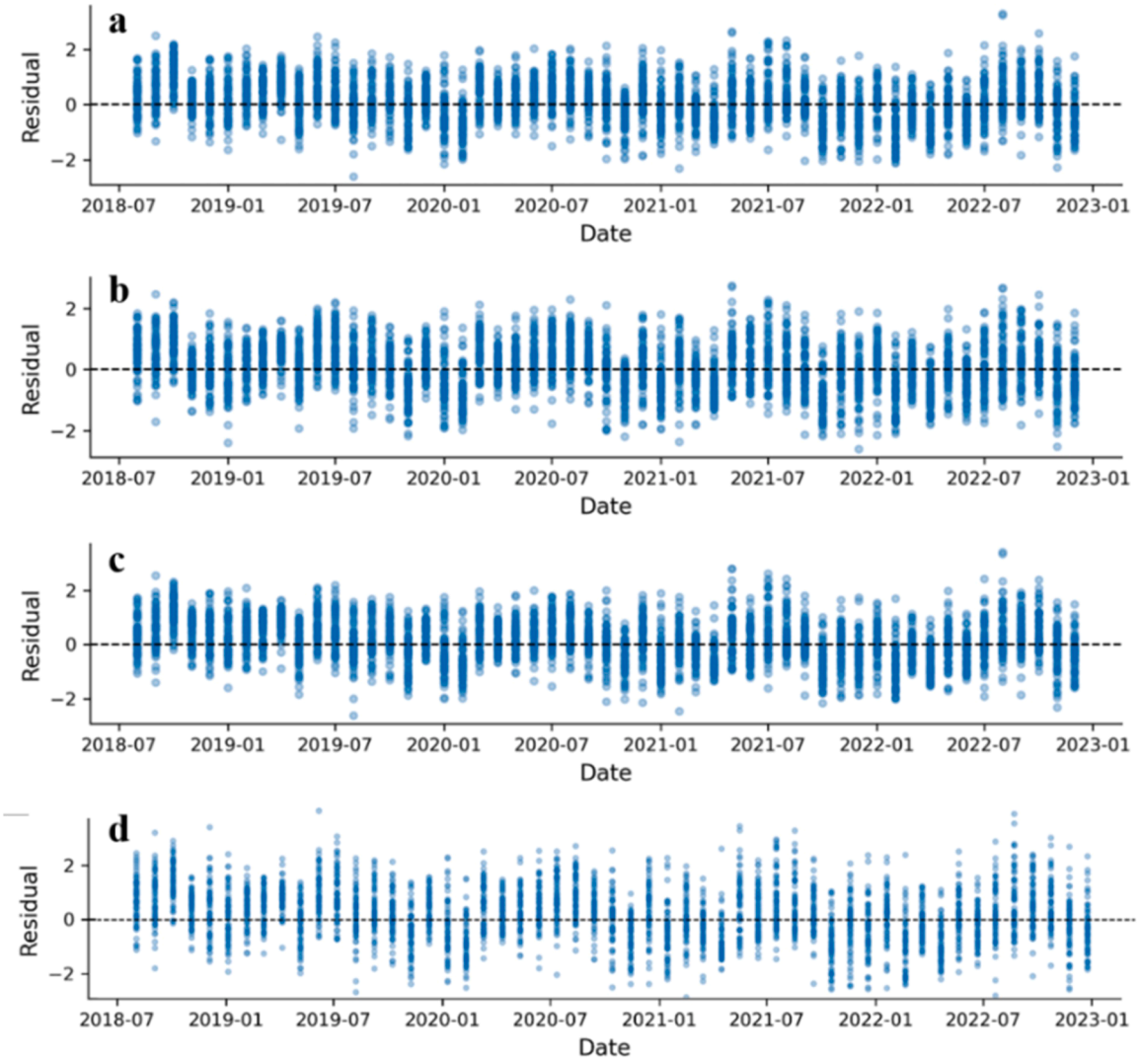

Fig. 8. SMDI residuals over time on the held-out test window across all stations, computed as (observed − predicted). The horizontal line indicates zero residual. Panels: (a) LightGBM, (b) Elastic Net, (c) Random Forest, and (d) FT-Transformer.

rate $\approx 6 \times 10^{-4}$, weight decay 1×10^{-5}, and batch size 256. For SPEI-3, the selected configuration used embedding dimension 128, 3 encoder blocks, dropout ≈ 0.25, learning rate $\approx 4.6 \times 10^{-4}$, weight decay 5×10^{-4}, and batch size 256.

4. Integrated agricultural drought monitoring framework

Grounded in the comparative performance and the diagnostics, we translate the modeling into an operational framework that fuses two complementary signals: SMDI, which reflects realized root-zone stress and is the most predictable target across algorithms, and SPEI-3, which is moderately harder to predict yet consistently provides early warning of emerging atmospheric imbalance. The tight 1:1 alignment, centered residuals, and steep error-CDFs observed for SMDI argue for using SMDI as the severity anchor, while the broader dispersion and heavier error tails for SPEI-3 underline its value as a lead indicator rather than a stand-alone severity metric. Attribution patterns are physically coherent across methods: short-term persistence dominates, followed by precipitation and land-surface temperature anomalies, with vegetation/water indices adding complementary information and cluster indicators providing only modest spatial adjustment.

Operational classes are defined on each index to turn continuous predictions into actionable categories. For SMDI, we retain the native, symmetric scale from Section 2.3.3: *Extreme Dry* (≤ -2), *Severe Dry* (-2 to -1.5), *Moderate Dry* (-1.5 to -1), *Near Normal* (-1–1), and *Wet* (> 1). For SPEI-3, we adopt standard meteorological thresholds: *Normal* (-0.5–0.5), *Moderate Drought* (-1.0 to -0.5), *Severe Drought* (-1.5 to -1.0), and *Extreme Drought* (< -1.5). These bins align with the behavior seen in the observed–versus–predicted panels and with the reliability profiles evident in the error-CDFs.

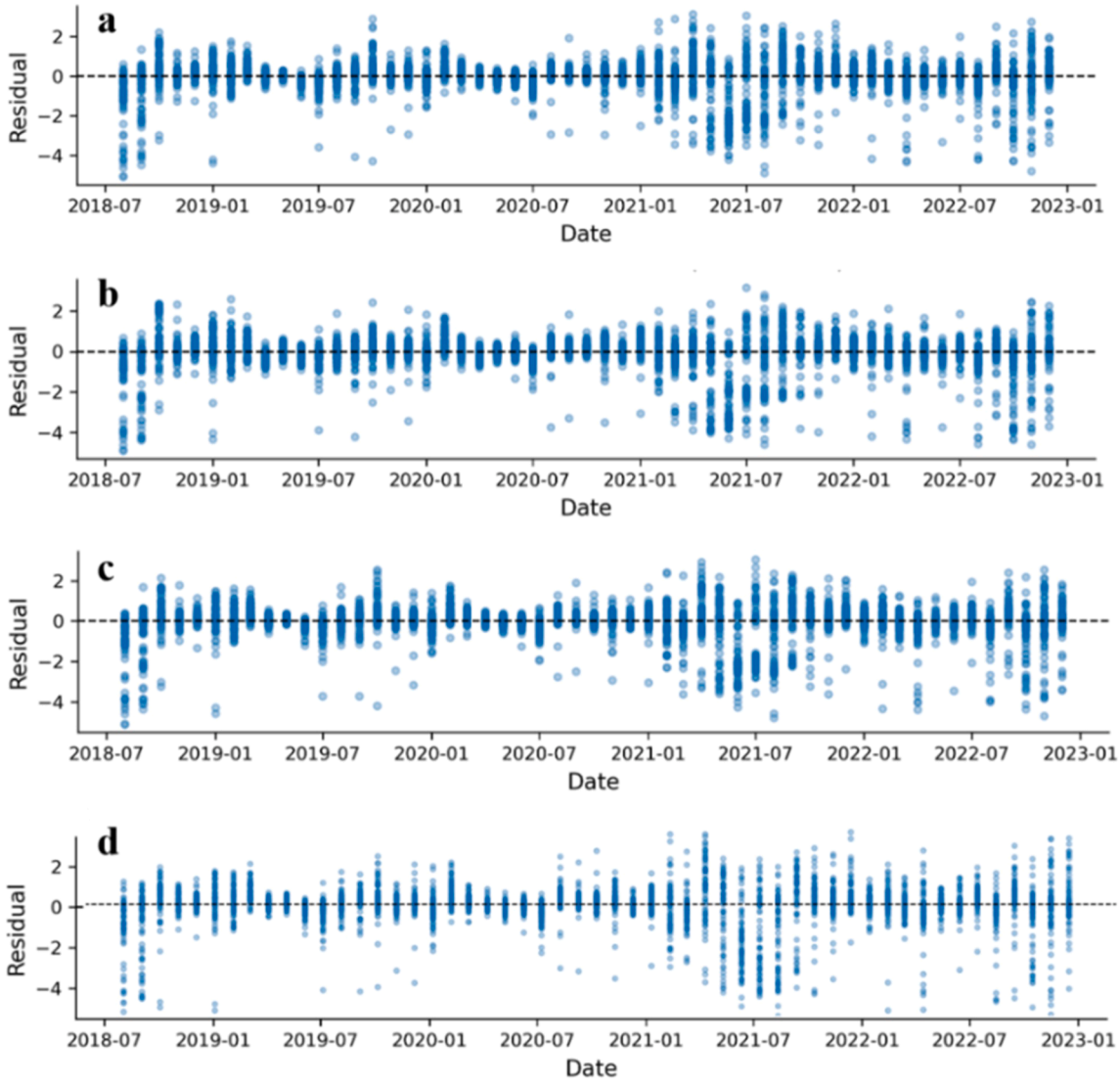

Fig. 9. SPEI-3 residuals over time on the held-out test window across all stations, computed as (observed − predicted). The horizontal line indicates zero residual. Panels: (a) LightGBM, (b) Elastic Net, (c) Random Forest, and (d) FT-Transformer.

Integration follows a conservative, SMDI-anchored policy with SPEI-3-based escalation. When SMDI indicates *Extreme Dry*, the integrated class is *Extreme* regardless of SPEI-3. When SMDI indicates *Severe Dry*, the integrated class is *Severe*, but it escalates to *Extreme* if SPEI-3 is *Severe* or *Extreme*. When SMDI indicates *Moderate Dry*, the integrated class is *Moderate*; it escalates to *Severe* if SPEI-3 is *Moderate* or *Severe*, and to *Extreme* if SPEI-3 is *Extreme*. When SMDI is *Near Normal*, the default is *No drought*; escalation to *Moderate* occurs if SPEI-3 is *Moderate* and to *Severe* if SPEI-3 is *Severe* or *Extreme*. When SMDI is *Wet*, the integrated class remains *No drought*. We implement this policy via the decision matrix in Table 4, which maps SMDI and SPEI-3 classes to the final operational status.

Uncertainty is communicated using the empirical CDFs of absolute errors. For each index, we derive 50% and 90% error quantiles from the test-window CDFs (Figure S1 and Figure S2) and attach confidence qualifiers to alerts: predictions that remain in the same class under the 50% band are flagged as higher confidence; those that remain stable under the 90% but not the 50% band are medium confidence; predictions that change class under the 90% band are low confidence. Because these bands come directly from out-of-sample diagnostics, they provide transparent risk information without complicating the decision rule.

For deployment, monthly updates are natural given data latency (Section 2.2). Reporting integrated classes alongside station-cluster summaries (Section 2.3.1) helps reveal regime-specific behavior without fragmenting the modeling. Given the accuracy ranking in Table 3 and paired tests against Random Forest, LightGBM is a strong default engine for SMDI; for SPEI-3, LightGBM and Random Forest are statistically indistinguishable on the held-out window, so either can be used in routine operation (we retain LightGBM for consistency). Elastic Net offers an interpretable, lightweight fallback, with Random Forest and FT-Transformer retained for sensitivity checks and redundancy. Finally, the integrated classes can be mapped to tiered advisories—ranging from monitoring

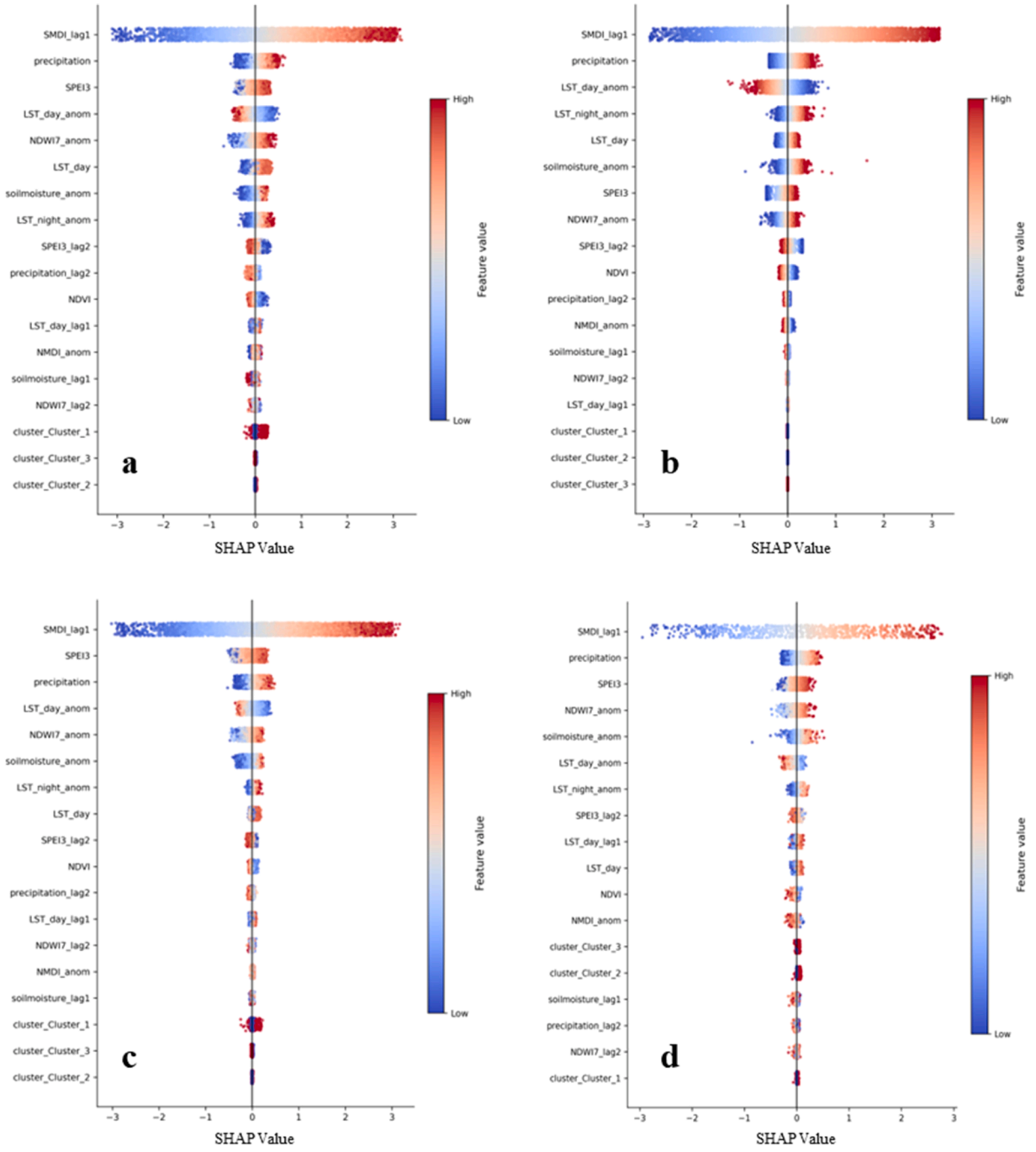

Fig. 10. SMDI SHAP Summary. Feature-attribution beeswarms on the test window. Horizontal spread shows contribution magnitude. Panels: (a) LightGBM, (b) Elastic Net, (c) Random Forest, (d) FT-Transformer.

and irrigation-scheduling reviews at *Moderate* through contingency actions at *Severe* and emergency measures at *Extreme*—so that the system not only diagnoses drought but also supports timely, risk-aware decisions.

5. Discussion

5.1. Key interpretation and main takeaways

The most important takeaway from this study is not only that the models perform differently across targets, but why the

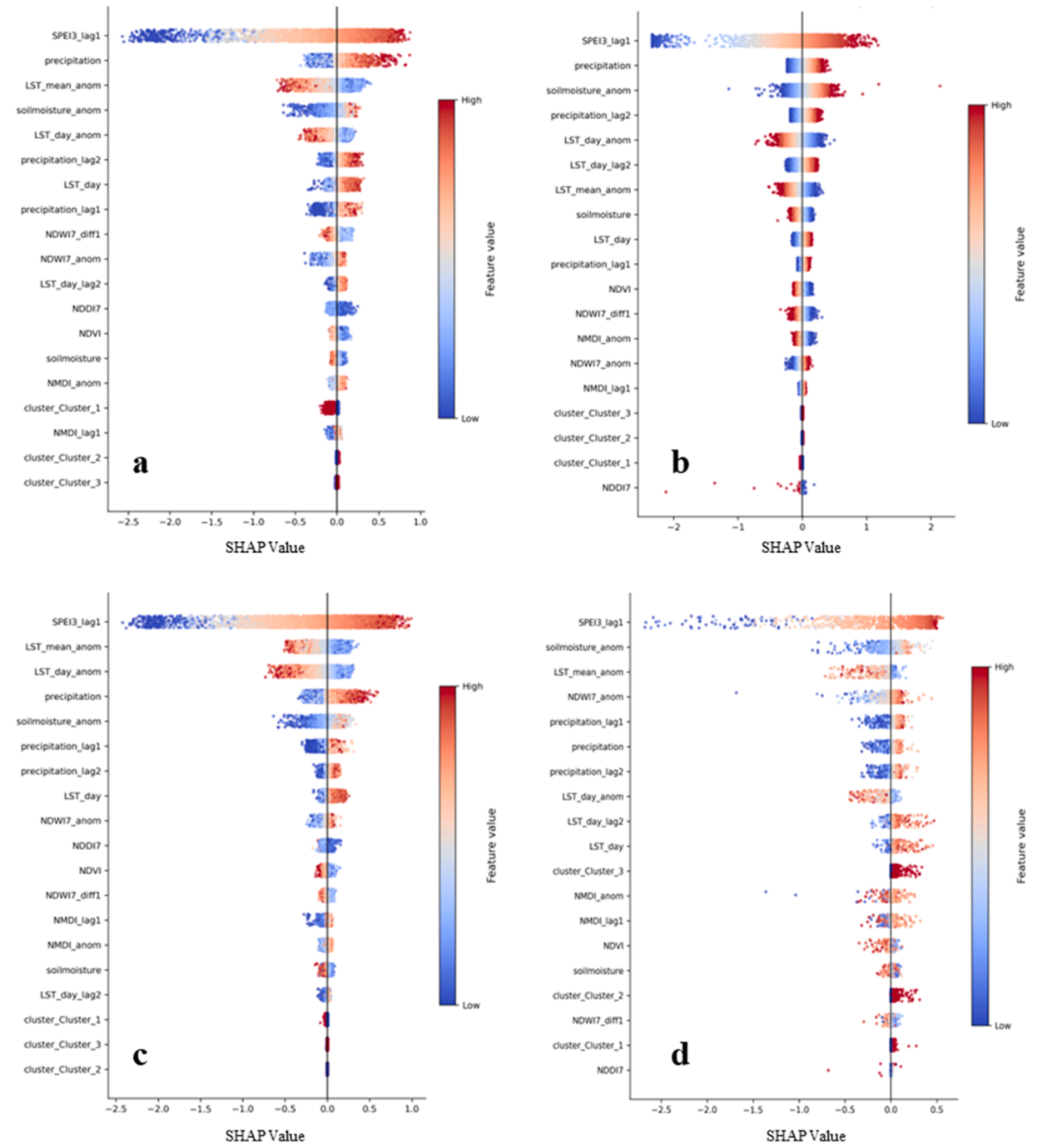

Fig. 11. SPEI-3 SHAP Summary. Feature-attribution beeswarms on the test window. Position shows contribution to prediction. Panels: (a) LightGBM, (b) Elastic Net, (c) Random Forest, (d) FT-Transformer.

performance pattern is so consistent across model families and what it implies for drought monitoring practice. The collective evidence from the scatter diagnostics, residual dynamics, and error-distribution plots indicates that the predictive problem is fundamentally different for the two indices: SMDI behaves as an integrated, memory-rich land-surface stress signal, whereas SPEI-3 is a more volatile atmospheric-balance indicator whose most informative months (extreme deficits) are also the rarest. This difference in index construction and frequency of extremes shapes what can be learned from monthly station–satellite predictors and where error concentrates.

A first interpretive result comes from the observed–predicted plots (Fig. 6 and Fig. 7). For SMDI, the point clouds are consistently

Table 4
Integrated agricultural drought classification (SMDI-anchored with SPEI-3–based escalation).

SMDI class ↓ \ SPEI-3 class →	Normal (−0.5–0.5)	Moderate Drought (−1.0 to −0.5)	Severe Drought (−1.5 to −1.0)	Extreme Drought (< −1.5)
Wet (> 1)	No drought	No drought	No drought	No drought
Near Normal (−1–1)	No drought	Moderate	Severe	Severe
Moderate Dry (−1.5 to −1)	Moderate	Severe	Severe	**Extreme**
Severe Dry (−2 to −1.5)	Severe	Severe	**Extreme**	**Extreme**
Extreme Dry (≤ −2)	**Extreme**	**Extreme**	**Extreme**	**Extreme**

dense around the 1:1 line across all algorithms, with widening primarily at the distribution tails. This pattern indicates that the models are capturing the dominant month-to-month structure of agricultural drought variability and that remaining error is concentrated where the process is intrinsically harder: abrupt transitions, rare extremes, and regimes where satellite proxies are noisier or less directly linked to root-zone stress. For SPEI-3, the same diagnostics show broader spread and a recurring understatement of the most negative values. Importantly, this tail underestimation appears across algorithms, which suggests it is not a "single model issue," but a consequence of the combination of (i) rarity of extreme dry months, (ii) regression objectives that optimize average squared error, and (iii) the practical limitation that monthly predictors often encode extremes less distinctly than moderate anomalies. In other words, the models learn the center of the distribution well, but the driest tail is systematically difficult. To test whether this behavior is region-dependent, we also report cluster-wise test performance across the three hydroclimatic zones (Table S4). Skill remains broadly consistent across clusters for both targets, supporting the regional adaptability of the pooled, regime-aware modeling strategy. In addition, evaluating the models on explicit extreme-drought subsets confirms the same tail behavior seen in the scatter diagnostics (Table S5): predictions are more conservative in the most severe months, which is expected when extremes are rare and models are optimized for overall error minimization.

A second interpretive result is the temporal stability of generalization (Fig. 8 and Fig. 9). The residual-over-time panels remain centered with no sustained drift for either index, implying that the trained relationships are not simply "fitting one sub-period," but are broadly stable across the held-out window. The occasional bursts of larger residual variance are consistent with high-variance hydroclimatic episodes to which drought indices respond nonlinearly through precipitation–temperature forcing and land–atmosphere coupling. To interpret this dispersion more explicitly, we summarized residual spread by month across the station network (Tables S6-S7). The largest bursts occur in a small number of high-variance months—especially for SPEI-3—characterized by (i) widespread extreme deficits and (ii) abrupt month-to-month shifts in observed conditions. For example, mid-2021 contains the highest error months, coinciding with strongly negative mean SPEI-3 and unusually positive LST anomalies with very low precipitation, consistent with short-lived but intense forcing regimes where squared-error training yields conservative estimates. In contrast, SMDI shows weaker coupling between error and drought severity, with larger errors primarily aligned with transition months rather than sustained tail failure. This is a meaningful quality-control finding: it indicates that the evaluation is not dominated by a single anomalous year or by hidden temporal leakage, and it supports the feasibility of operational deployment with periodic re-training.

Third, the empirical CDFs of absolute error (Figure S1 and Figure S2) provide a practical reliability interpretation that scatter plots alone cannot. For SMDI, the steep rise of the CDF indicates that a large fraction of months lie within modest absolute error (on the order of ~1 index unit), whereas SPEI-3 exhibits a slower rise and a heavier right tail, meaning that a smaller subset of months contributes disproportionately to large errors. This is exactly the behavior one would expect when extremes are rare but important. From an operational standpoint, this also provides a concrete basis for uncertainty communication: even when point predictions are used, the system can attach probability-like qualifiers to drought classes (e.g., "high confidence" when error quantiles are small, "caution" when approaching tail regimes), rather than treating all months as equally reliable.

Fourth, the model-family comparison yields a useful methodological interpretation that goes beyond "which model won." The strong performance of gradient-boosted trees for SMDI is consistent with a favorable bias–variance trade-off in medium-sized monthly tabular data: boosting captures nonlinearities and interactions among persistence, precipitation, thermal stress (LST), and land-surface indicators while remaining robust under moderate hyperparameter misspecification. Random Forest's closeness is consistent with variance reduction via bagging, while its slight underperformance relative to boosting is plausible given the absence of residual-fitting and typically coarser function approximation. The fact that Elastic Net remains close to the nonlinear ensembles for SMDI is not trivial. It suggests that, after feature engineering and standardization, a substantial share of predictable SMDI variability is approximately linear (with persistence and hydroclimatic forcing dominating). That is encouraging for interpretability and deployment because it implies the signal is not entirely dependent on complex nonlinear machinery. Finally, the FT-Transformer's competitive but not superior performance is consistent with what is often observed in deep tabular learning: without very large sample sizes or richer temporal structure than monthly station records provide, deep models do not reliably dominate well-tuned gradient-boosted trees.

Fifth, the convergence of feature-attribution patterns across algorithms is one of the strongest indicators that the models are learning physically meaningful relationships rather than dataset quirks (Fig. 10 and Fig. 11). Across model families, short-term persistence terms dominate both targets (antecedent index values as top contributors), followed by precipitation and thermal/energy-related anomalies, with soil moisture and vegetation/water indices providing secondary refinement. This ranking matches process understanding: antecedent conditions provide monthly memory; water supply and evaporative demand modulate departures from normal; and surface/vegetation signals reflect realized stress and soil–vegetation feedbacks. This process-consistent attribution is reinforced by our robustness checks showing broadly comparable predictive skill across the three hydroclimatic clusters (climate

zones) and predictable degradation primarily in the rare extreme tail, where models become more conservative (Section 3.1; Tables S4–S5). Beyond global ranking, TreeSHAP contribution dependence plots (Figure S9) clarify that several key predictors act nonlinearly and conditionally. Precipitation shows a threshold-like response, with diminishing marginal influence away from the wet–dry transition, while positive LST anomalies contribute increasingly negative shifts in both targets, consistent with enhanced evaporative stress under warm conditions. The interaction coloring by antecedent drought state (lag-1) indicates state dependence: the same hydroclimatic anomaly typically exerts a larger negative contribution when the previous month is already dry, reflecting limited buffering under low storage. For SPEI-3, a severity-stratified view further shows that warm anomalies are associated with a wider and more negative contribution range during severe drought months, consistent with amplified evaporative-demand impacts in the driest tail. The modest but non-zero contributions of cluster indicators are also interpretable in a desirable way: they capture residual spatial heterogeneity without overpowering the main hydroclimatic controls, which is consistent with the purpose of pooling stations while still acknowledging that climatic–agricultural behavior is not spatially uniform.

Finally, the structure of errors aligns with the intended roles of the indices and supports a clear operational interpretation. SMDI's integrative nature and the inclusion of relevant predictors enable more stable and symmetric error behavior. SPEI-3, however, is intrinsically harder at the driest tail and is the index most likely to be under-estimated in extreme months. This does not undermine the utility of SPEI-3; rather, it clarifies its best role in a dual-index setting: SPEI-3 is well-suited for escalation/early warning, while SMDI is better suited for anchoring realized severity. In practice, this means that the framework should be conservative in interpreting extreme SPEI-3 point predictions and should incorporate tail-aware enhancements (e.g., alternative objectives, resampling, or probabilistic outputs) if extreme-event fidelity is a priority.

5.2. Comparison with previous studies

Our findings are consistent with (and extend) prior drought forecasting and monitoring research in several ways.

(i) **Machine learning for SPEI forecasting.** Multiple studies have shown that data-driven methods can forecast SPEI at short time scales, often with strong one-step-ahead performance when long records and persistence information are available. For example, deep learning (LSTM) has been demonstrated to achieve high skill for SPEI-1 and SPEI-3 at one-month lead time, particularly when antecedent drought state and hydroclimatic covariates are included (Dikshit et al., 2021). Likewise, ML approaches (including Random Forest and ensembles) have been reported to perform well for SPEI prediction at multiple scales, reinforcing the role of persistence and hydroclimate predictors in drought-index forecasting (En-Nagre et al., 2024; Zhang et al., 2024). Our results align with this literature in identifying antecedent state as a dominant control; however, we additionally show that tail underestimation remains a persistent challenge for SPEI-3 extremes across model families, indicating that objective functions and sampling strategies can matter as much as model class when extreme drought fidelity is the priority.

(ii) **Why a dual-index approach is justified.** Prior methodological work emphasizes that SPEI incorporates evaporative demand (through PET) and can diverge meaningfully from precipitation-only indicators under warming, which is one reason SPEI is widely used to complement SPI-type approaches (Gumus, 2023). Our dual-index framework operationalizes this complementarity in an agricultural context: rather than forcing a single index to serve both "early signal" and "realized stress," the results support using SPEI-3 as an atmospheric imbalance trigger and SMDI as a storage/stress anchor.

(iii) **SMDI and agricultural drought representation.** SMDI was originally proposed to represent soil-moisture deficit dynamics more directly than purely meteorological indices and to better reflect agricultural stress mechanisms. In line with that motivation, our modeling shows that SMDI exhibits higher predictability at monthly resolution, consistent with its stronger memory and smoother dynamics relative to SPEI-3.

(iv) **Regional context and Iran-relevant evidence.** Recent Iran-focused research supports the value of combining meteorological and agricultural perspectives for national-scale drought characterization, and complementary work has shown that machine-learning models trained on multiple standardized drought indices can achieve high forecasting skill across Iran's arid and semi-arid climates, reinforcing the practical benefit of multi-index, data-driven drought assessment in heterogeneous regimes (Amiranipour et al., 2026; Talebi et al., 2025). Our work complements these studies by (i) explicitly emphasizing a leakage-safe, forward-in-time evaluation, (ii) comparing modern tabular learners and a deep tabular transformer under a unified protocol, and (iii) translating the dual-index insight into an operational monitoring framework (next section) that supports decision thresholds and uncertainty communication.

5.3. Limitations and sources of uncertainty

Several constraints affected the analysis design, but they are transparent and importantly do not invalidate the methodological and operational contributions.

(i) **Study period (2001–2022) constrained by station-data availability and access.** The selected period reflects the longest continuous window for which complete monthly station precipitation and temperature records were available to the study at the time of data acquisition. Access to more recent observations was not feasible through the same channels due to institutional release constraints. While this limits "to-present" updating, it preserves an internally consistent dataset needed for robust multi-source integration and forward validation.

(ii) **Station network shaped by completeness and collocation constraints.** Station inclusion required (a) continuous monthly records throughout the analysis window, and (b) valid spatial collocation with all remote-sensing products used. These requirements necessarily exclude many candidate stations with missing periods or invalid satellite retrieval coverage at their coordinates. A further agricultural relevance criterion intentionally increases representation in key farming provinces while maintaining nationwide coverage. This design choice strengthens the study's agricultural focus while ensuring the predictor set is consistently available at all selected locations.

(iii) **Extremes and objective-function bias.** Separate evaluation on extreme-drought subsets confirms that the most severe SPEI-3 deficits are predicted more conservatively across model families, a known consequence when models are optimized for average squared-error behavior and extremes are relatively sparse. This limitation reinforces the value of using SPEI-3 as an escalation signal rather than a sole severity anchor and motivates tail-aware modeling enhancements.

(iv) **Predictor completeness and management effects.** Although the predictor suite captures major hydroclimatic and land-surface signals, some influences (e.g., irrigation management intensity, crop calendars, soil hydraulic parameters, and full PET drivers such as radiation and wind) are only indirectly represented. This can limit explanatory completeness in heavily managed or highly heterogeneous agricultural systems.

5.4. Recommendations and future work

Future extensions can build directly on the present framework while targeting the specific behaviors observed in the diagnostics. For SPEI-3, improving fidelity in rare extreme dry conditions is a natural next step; this can be approached through tail-aware objectives (e.g., Huber/quantile/asymmetric losses), resampling or reweighting strategies that increase emphasis on extreme drought months, and probabilistic formulations that convey risk rather than relying only on point estimates. In parallel, further gains are likely to come from *signal enrichment* more than changing model families, particularly by incorporating additional controls on evaporative demand (e.g., radiation and wind proxies), static land attributes governing buffering capacity (soil properties, rooting depth), and management-related information where feasible (e.g., irrigation footprints and cropping calendars).

Future work would also extend the dual-index framework to wall-to-wall drought mapping by applying the trained models on gridded predictor fields, while maintaining reliability through station-anchored benchmarking and (where needed) bias-correction/calibration of gridded outputs against ground observations.

Operational deployment would also benefit from formalizing uncertainty communication by calibrating drought-class confidence qualifiers using empirical out-of-sample error distributions, allowing end users to interpret predicted classes with transparent reliability context. Finally, robustness can be strengthened through additional generalization tests—such as spatial holdouts or climate-zone cross-validation—to quantify transferability into under-sampled areas and to guide priorities for future monitoring network improvements.

6. Conclusions

This study develops and rigorously evaluates a dual-index framework for monthly agricultural drought monitoring over Iran's 70-station network (2001–2022), combining SMDI to represent realized root-zone stress with SPEI-3 to represent short-to-seasonal meteorological imbalance. Using harmonized station and remote-sensing predictors, leakage-safe preprocessing, regime-aware clustering, and strictly forward-in-time expanding cross-validation, we compared four learning families (LightGBM, Elastic Net, Random Forest, and FT-Transformer) under a unified, stability-selected predictor set.

Three conclusions follow. First, monthly SMDI is consistently more learnable than SPEI-3 because it is governed by root-zone soil-water storage and antecedent drought-state persistence ("soil-moisture memory")—i.e., the tendency for soil moisture deficits to evolve smoothly and carry information from one month to the next. This structure also explains the strong performance of tree ensembles: they naturally capture nonlinear interactions and threshold-like behavior in how precipitation anomalies and thermal stress translate into soil-water deficits under different antecedent wetness conditions. Second, SPEI-3 remains operationally valuable but intrinsically harder at monthly resolution because it reflects short-term hydroclimatic forcing, dominated by recent precipitation deficits and evaporative-demand variability (through PET-related temperature effects), and because the most informative months are rare extremes that squared-error regression tends to predict conservatively. Accordingly, for SPEI-3 the top-performing models are effectively tied on average, and differences between leading methods should be interpreted as small. Third, feature-attribution patterns provide a coherent process narrative across model families: persistence terms dominate both targets, followed by precipitation and land-surface temperature anomalies, with vegetation/water indices adding incremental refinement and regime indicators providing modest spatial adjustment—consistent with drought propagation from atmospheric anomalies into root-zone stress.

Building on these insights, we proposed an operational integration in which SMDI anchors severity while SPEI-3 escalates alerts when meteorological forcing indicates imminent deterioration. This design translates continuous predictions into actionable drought classes and can attach uncertainty qualifiers derived from empirical out-of-sample error distributions, keeping the system transparent and appropriately conservative in high-stakes settings while remaining updateable on a monthly cadence.

Several extensions could further strengthen skill and operational value. On the data side, higher-frequency precipitation, radiation and wind proxies relevant to evaporative demand, and static land/management factors (e.g., soil texture, available water capacity, rooting depth, irrigation footprints) would better represent buffering capacity and improve sensitivity to onset and persistence. On the modeling side, tail-aware objectives and probabilistic formulations would better target rare extremes, while spatial holdouts or climate-zone cross-validation would further quantify transferability into under-sampled regions. Finally, extending this regime-aware

dual-index framework to gridded mapping would enable wall-to-wall drought monitoring while retaining station-anchored benchmarking for reliability.

CRediT authorship contribution statement

Ali Abbasi: Writing – review & editing, Validation, Supervision, Resources, Project administration, Investigation, Data curation, Conceptualization. **Mohammad Reza Asli Charandabi:** Writing – original draft, Visualization, Methodology. **Mahan Azizi:** Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Data curation, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ejrh.2026.103376.

Data availability

Data will be made available on request.

References

Abu Arra, A., Alashan, S., Şişman, E., 2024a. Trends of meteorological and hydrological droughts and associated parameters using innovative approaches. J. Hydrol. 640, 131661. https://doi.org/10.1016/j.jhydrol.2024.131661.

Abu Arra, A., Birpınar, M.E., Şişman, E., 2025. Evaluating ERA5-LAND and IMERG-NASA products for drought analysis: implications for sustainable water resource management. Sustainability 17. https://doi.org/10.3390/su17167529.

Abu Arra, A., Birpınar, M.E., Gazioğlu, Ş.A., Şişman, E., 2024b. Critical drought characteristics: a new concept based on dynamic time period scenarios. Atmosphere (Basel) 15. https://doi.org/10.3390/atmos15070768.

Abu Arra, A., Şişman, E., 2024b. Innovative drought classification matrix and acceptable time period for temporal drought evaluation. Water Resour. Manag. 38, 2811–2833. https://doi.org/10.1007/s11269-024-03793-0.

Abu Arra, A., Şişman, E., 2024a. A comprehensive analysis and comparison of SPI and SPEI for spatiotemporal drought evaluation. Environ. Monit. Assess. 196, 980. https://doi.org/10.1007/s10661-024-13127-7.

Abu Arra, A., Şişman, E., 2025. New insights into meteorological and hydrological drought modeling: a comparative analysis of parametric and non-parametric distributions. Atmosphere (Basel) 16. https://doi.org/10.3390/atmos16070846.

Adib, A., Amiri, M., Lotfirad, M., Farajpanah, H., 2024. Evaluation of meteorological drought using SPI and SPEI on wheat yield in Southwestern Iran. Earth. https://doi.org/10.3390/earth5040053.

Ahire, A., Zade, N., Mujawar, U., Mehta, D., Kotecha, K., 2025. Meteorological drought severity forecasting utilizing blended modelling. MethodsX 15, 103456. https://doi.org/10.1016/j.mex.2025.103456.

Ahmed Osman, A.I., AlDahoul, N., Chong, K.L., Huang, Y.F., Ng, J.L., Elshafie, A., Sherif, M., Ahmed, A.N., 2025. A review on machine learning models for drought monitoring and forecasting. Clim. Risk Manag. 50, 100758. https://doi.org/10.1016/j.crm.2025.100758.

Alam, N.M., Sharma, G.C., Moreira, E., Jana, C., Mishra, P.K., Sharma, N.K., Mandal, D., 2017. Evaluation of drought using SPEI drought class transitions and log-linear models for different agro-ecological regions of India. Phys. Chem. Earth 100, 31–43. https://doi.org/10.1016/j.pce.2017.02.008.

Albarakat, R., Le, M.-H., Lakshmi, V., 2022. Assessment of drought conditions over Iraqi transboundary rivers using FLDAS and satellite datasets. J. Hydrol. Reg. Stud. 41, 101075. https://doi.org/10.1016/j.ejrh.2022.101075.

Alkaraki, K.F., Hazaymeh, K., 2023. A comprehensive remote sensing-based Agriculture Drought Condition Indicator (CADCI) using machine learning. Environ. Chall. 11, 100699. https://doi.org/10.1016/j.envc.2023.100699.

Amiranipour, M., Najafzadeh, M., Mohamadi, S., 2026. Meteorological and agricultural drought assessments using satellite imagery and machine learning models. Adv. Sp. Res. 77, 549–572. https://doi.org/10.1016/j.asr.2025.10.014.

Bari Abarghouei, H., Asadi Zarch, M.A., Dastorani, M.T., Kousari, M.R., Safari Zarch, M., 2011. The survey of climatic drought trend in Iran. Stoch. Environ. Res. Risk Assess. 25, 851–863. https://doi.org/10.1007/s00477-011-0491-7.

Breiman, L., 2001. Random forests. Mach. Learn 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Chen, Y., Liu, X., Zheng, C., Ma, Y., Gao, W., He, J., Hao, L., Liu, Z., Shi, C., Cao, Q., 2024. Estimation of water budget components and its driving factors analysis in arid grassland. Sci. Total Environ. 906, 167654. https://doi.org/10.1016/j.scitotenv.2023.167654.

Copernicus Climate Change Service, 2018. Soil moisture gridded data from 1978 to present. https://doi.org/10.24381/CDS.D7782F18.

Danandeh Mehr, A., Torabi Haghighi, A., Jabarnejad, M., Safari, M.J.S., Nourani, V., 2022. A new evolutionary hybrid random forest model for SPEI forecasting. Water (Switzerland) 14. https://doi.org/10.3390/w14050755.

Dikshit, A., Pradhan, B., Huete, A., 2021. An improved SPEI drought forecasting approach using the long short-term memory neural network. J. Environ. Manag. 283, 111979. https://doi.org/10.1016/j.jenvman.2021.111979.

Draper, N.R., Smith, H., 1998. Applied regression analysis. Wiley Series in Probability and Statistics, 1st ed. Wiley.

Dutta, D., Kundu, A., Patel, N.R., 2013. Predicting agricultural drought in eastern Rajasthan of India using NDVI and standardized precipitation index. Geocarto Int. 28, 192–209. https://doi.org/10.1080/10106049.2012.679975.

Elbeltagi, A., Althobiani, F., Kamruzzaman, M., Shaid, S., Roy, D.K., Deb, L., Islam, M.M., Kundu, P.K., Rahman, M.M., 2022. Estimating the standardized precipitation evapotranspiration index using data-driven techniques: a regional study of Bangladesh. Water (Switzerland) 14, 1–16. https://doi.org/10.3390/w14111764.

En-Nagre, K., Aqnouy, M., Ouarka, A., Ali Asad Naqvi, S., Bouizrou, I., Eddine Stitou El Messari, J., Tariq, A., Soufan, W., Li, W., El-Askary, H., 2024. Assessment and prediction of meteorological drought using machine learning algorithms and climate data. Clim. Risk Manag. 45, 100630. https://doi.org/10.1016/j.crm.2024.100630.

Erhardt, T.M., Czado, C., 2018. Standardized drought indices: a novel univariate and multivariate approach. J. R. Stat. Soc. Ser. C Appl. Stat. 67, 643–664. https://doi.org/10.1111/rssc.12242.
Feng, P., Wang, B., Liu, D.L., Yu, Q., 2019. Machine learning-based integration of remotely-sensed drought factors can improve the estimation of agricultural drought in South-Eastern Australia. Agric. Syst. https://doi.org/10.1016/j.agsy.2019.03.015.
Gao, B., 1996. NDWI - a normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sens. Environ. 58, 257–266. https://doi.org/10.1016/S0034-4257(96)00067-3.
Gorishniy, Y., Rubachev, I., Khrulkov, V., Babenko, A., 2021. Revisiting Deep Learning Models for Tabular Data, in: 35th Conference on Neural Information Processing Systems (NeurIPS 2021), Sydney, Australia. https://doi.org/10.48550/arXiv.2106.11959.
GSFC, P.P.S. (PPS) A.N., 2023. GPM IMERG Final Precipitation L3 Half Hourly 0.1 ° x 0.1 ° V07. https://doi.org/10.5067/GPM/IMERG/3B-HH/07.
Gu, Y., Brown, J.F., Verdin, J.P., Wardlow, B., 2007. A five-year analysis of MODIS NDVI and NDWI for grassland drought assessment over the central Great Plains of the United States. Geophys. Res. Lett. 34. https://doi.org/10.1029/2006GL029127.
Gumus, V., 2023. Evaluating the effect of the SPI and SPEI methods on drought monitoring over Turkey. J. Hydrol. 626, 130386. https://doi.org/10.1016/j.jhydrol.2023.130386.
Hartigan, J.A., Wong, M.A., 1979. Algorithm AS 136: A K-Means Clustering Algorithm. J. R. Stat. Soc. Ser. C. (Appl. Stat. 28, 100–108. https://doi.org/10.2307/2346830.
Hastie, T., Tibshirani, R., Friedman, J., 2009. Springer Series in Statistics. In: The elements of statistical learning. Springer, New York, NY.
Huang, J., Chen, J., Huang, H., Cai, X., 2025. Deep learning-based daily streamflow prediction model for the hanjiang river basin. Hydrology 12. https://doi.org/10.3390/hydrology12070168.
Iran Meteorological Organization, n.d. Weather data acquisition system [WWW Document]. URL ⟨https://data.irimo.ir/⟩.
Iran's Statistical Centre, 2023. Shares and rankings of provinces based on important variables in the agricultural sector.
Islam, A.R.M.T., Mamun, M.A.-A., Hasan, M., Aktar, M.N., Uddin, M.N., Siddique, M.A.B., Chowdhury, M.H., Islam, M.S., Bari, A.B.M.M., Idris, A.M., Senapathi, V., 2025. Optimizing coastal groundwater quality predictions: a novel data mining framework with cross-validation, bootstrapping, and entropy analysis. J. Contam. Hydrol. 269, 104480. https://doi.org/10.1016/j.jconhyd.2024.104480.
Jung, H.C., Kang, D.H., Kim, E., Getirana, A., Yoon, Y., Kumar, S., Peters-lidard, C.D., Hwang, E.H., 2020. Towards a soil moisture drought monitoring system for South Korea. J. Hydrol. 589, 125176. https://doi.org/10.1016/j.jhydrol.2020.125176.
Kabbilawsh, P., Sathish Kumar, D., Chithra, N.R., 2022. Performance evaluation of univariate time-series techniques for forecasting monthly rainfall data. J. Water Clim. Chang 13, 4151–4176. https://doi.org/10.2166/wcc.2022.107.
Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.Y., 2017. LightGBM: A highly efficient gradient boosting decision tree, in: Advances in Neural Information Processing Systems. pp. 3147–3155.
Keskin, M.Z., Abu Arra, A., Akca, S., Şişman, E., 2025. Actual and potential trend analysis under climate change using risk Sen's Slope (RSS) in Western Black Sea Basin in Türkiye. Int. J. Clim. 45, e8703. https://doi.org/10.1002/joc.8703.
Labudová, L., Labuda, M., Takáč, J., 2017. Comparison of SPI and SPEI applicability for drought impact assessment on crop production in the Danubian Lowland and the East Slovakian Lowland. Theor. Appl. Clim. 128, 491–506. https://doi.org/10.1007/s00704-016-1870-2.
Liu, J., Chen, Y., Liu, X., Lu, Y., Zheng, C., Shi, C., Liu, Z., Gao, Y., Zhang, L., 2025. Soil hydrological processes and drought response of typical vegetation in arid regions under long-term climate patterns. J. Hydrol. Reg. Stud. 62, 102919. https://doi.org/10.1016/j.ejrh.2025.102919.
Liu, H.Q., Huete, A., 1995. A feedback based modification of the NDVI to minimize canopy background and atmospheric noise. IEEE Trans. Geosci. Remote Sens 33, 457–465. https://doi.org/10.1109/TGRS.1995.8746027.
Lotfirad, M., Esmaeili-Gisavandani, H., Adib, A., 2021. Drought monitoring and prediction using SPI, SPEI, and random forest model in various climates of Iran. J. Water Clim. Chang 13, 383–406. https://doi.org/10.2166/wcc.2021.287.
Ma, X., He, Q., Zhou, G., 2018. Sequence of changes in maize responding to soil water deficit and related critical thresholds. Front. Plant Sci. 9, 1–12. https://doi.org/10.3389/fpls.2018.00511.
Maca, P., Pech, P., 2016. Forecasting SPEI and SPI drought indices using the integrated artificial neural networks. Comput. Intell. Neurosci. 2016. https://doi.org/10.1155/2016/3868519.
Magallanes-Quintanar, R., Galván-Tejada, C.E., Galván-Tejada, J.I., de Jesús Méndez-Gallegos, S., Blanco-Macías, F., Valdez-Cepeda, R.D., 2023. Artificial neural network models for prediction of standardized precipitation index in central Mexico. Agrociencia 57, 177–207. https://doi.org/10.47163/agrociencia.v57i1.2655.
Marcos Junior, A.D., Silveira, C., da, S., da Costa, J.M.F., Gonçalves, S.T.N., 2024. Combining traditional hydrological models and machine learning for streamflow prediction. Rev. Bras. Recur. Hidr. 29, 1–17. https://doi.org/10.1590/2318-0331.292420230105.
Modanesi, S., Massari, C., Camici, S., Brocca, L., Amarnath, G., 2020. Do Satellite Surface Soil Moisture Observations Better Retain Information About Crop-Yield Variability in Drought Conditions? Water Resour. Res. 56, 1–19. https://doi.org/10.1029/2019WR025855.
Moges, D.M., Virro, H., Kmoch, A., Cibin, R., Rohith, R.A.N., Martínez-Salvador, A., Conesa-García, C., Uuemaa, E., 2024. Streamflow prediction with time-lag-informed random forest and its performance compared to SWAT in diverse catchments. Water (Switz.) 16, 1–21. https://doi.org/10.3390/w16192805.
Mokhtar, A., Jalali, M., He, H., Al-Ansari, N., Elbeltagi, A., Alsafadi, K., Abdo, H.G., Sammen, S.S., Gyasi-Agyei, Y., Rodrigo-Comino, J., 2021. Estimation of SPEI meteorological drought using machine learning algorithms. IEEE Access 9, 65503–65523. https://doi.org/10.1109/ACCESS.2021.3074305.
Morid, S., Smakhtin, V., Bagherzadeh, K., 2007. Drought forecasting using artificial neural networks and time series of drought indices. Int. J. Climatol. 27, 2103–2111. https://doi.org/10.1002/joc.1498.
Narasimhan, B., Srinivasan, R., 2005. Development and evaluation of Soil Moisture Deficit Index (SMDI) and Evapotranspiration Deficit Index (ETDI) for agricultural drought monitoring. Agric. For. Meteor. 133, 69–88. https://doi.org/10.1016/J.AGRFORMET.2005.07.012.
NASA GSFC Hydrological Sciences Laboratory (HSL), 2018. Fldas noah land surface model l4 global monthly 0. 1×0. 1 degre(MERRA-2 and CHIRPS)e v001. https://doi.org/10.5067/5NHC22T9375G.
Orimoloye, I.R., 2022. Agricultural drought and its potential impacts: enabling decision-support for food security in vulnerable regions. Front. Sustain. Food Syst. 6. https://doi.org/10.3389/fsufs.2022.838824.
Oyarzabal, R.S., Santos, L.B.L., Cunningham, C., Broedel, E., de Lima, G.R.T., Cunha-Zeri, G., Peixoto, J.S., Anochi, J.A., Garcia, K., Costa, L.C.O., Pampuch, L.A., Cuartas, L.A., Zeri, M., Guedes, M.R.G., Negri, R.G., Muñoz, V.A., Cunha, A.P.M.A., 2025. Forecasting drought using machine learning: a systematic literature review. Nat. Hazards 121, 9823–9851. https://doi.org/10.1007/s11069-025-07195-2.
Pasik, A., Gruber, A., Preimesberger, W., De Santis, D., Dorigo, W., 2023. Uncertainty estimation for a new exponential-filter-based long-term root-zone soil moisture dataset from Copernicus Climate Change Service (C3S) surface observations. Geosci. Model Dev. 16, 4957–4976. https://doi.org/10.5194/gmd-16-4957-2023.
Potopová, V., Boroneanț, C., Boincean, B., Soukup, J., 2016. Impact of agricultural drought on main crop yields in the Republic of Moldova. Int. J. Climatol. 36, 2063–2082. https://doi.org/10.1002/joc.4481.
Prodhan, F.A., Zhang, J., Yao, F., Shi, L., Sharma, T.P.P., Zhang, D., Cao, D., Zheng, M., Ahmed, N., Mohana, H.P., 2021. Deep learning for monitoring agricultural drought in south asia using remote sensing data. Remote Sens. 13. https://doi.org/10.3390/rs13091715.
Raziei, T., Miri, M., Santos, J.F., Zand, M., Pereira, L.S., 2025. Development and evaluation of a principal component-based composite drought index considering temporal lag dependencies among indices. Water Resour. Manag. 39, 5949–5970. https://doi.org/10.1007/s11269-025-04235-1.
Rouse, J.W., Haas, R.H., Schell, J.A., Deering, D.W., 1974. Monitoring vegetation systems in the Great Plains with ERTS. NASA Spec. Publ. 351, 309.
Rousseeuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. 20, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7.
Seifian, Z., Hooshyaripor, F., Saghafian, B., Mirabbasi, R., 2025. Investigating meteorological drought propagation to soil moisture drought: insights from Iran's diverse climate regions. Theor. Appl. Clim. 157, 15. https://doi.org/10.1007/s00704-025-05924-y.

Talebi, H., Citakoglu, H., Samadianfard, S., Erol, A., 2025. Advanced hybrid machine learning for precise short-term drought prediction: a comparative study of SPI and SPEI indices in Iran's arid and semi-arid regions. Pure Appl. Geophys. https://doi.org/10.1007/s00024-025-03876-y.
Thornthwaite, C.W., 1948. An approach toward a rational classification of climate. Geogr. Rev. 38, 55–94. https://doi.org/10.2307/210739.
Train, J., McBean, E., Gharabaghi, B., 2024. Key factors influencing drinking water advisories on indigenous reserves in Canada: an XGBoost analysis. Water 16. https://doi.org/10.3390/w16243647.
Uddin, M.G., Rana, M.M.S.P., Diganta, M.T.M., Bamal, A., Sajib, A.M., Abioui, M., Shaibur, M.R., Ashekuzzaman, S.M., Nikoo, M.R., Rahman, A., Moniruzzaman, M., Olbert, A.I., 2024. Enhancing groundwater quality assessment in coastal area: a hybrid modeling approach. Heliyon 10. https://doi.org/10.1016/j.heliyon.2024.e33082.
Van Loon, A.F., Kchouk, S., Matanó, A., Tootoonchi, F., Alvarez-Garreton, C., Hassaballah, K.E.A., Wu, M., Wens, M.L.K., Shyrokaya, A., Ridolfi, E., Biella, R., Nagavciuc, V., Barendrecht, M.H., Bastos, A., Cavalcante, L., de Vries, F.T., Garcia, M., Mård, J., Streefkerk, I.N., Teutschbein, C., Tootoonchi, R., Weesie, R., Aich, V., Boisier, J.P., Di Baldassarre, G., Du, Y., Galleguillos, M., Garreaud, R., Ionita, M., Khatami, S., Koehler, J.K.L., Luce, C.H., Maskey, S., Mendoza, H.D., Mwangi, M.N., Pechlivanidis, I.G., Ribeiro Neto, G.G., Roy, T., Stefanski, R., Trambauer, P., Koebele, E.A., Vico, G., Werner, M., 2024. Review article: drought as a continuum – memory effects in interlinked hydrological, ecological, and social systems. Nat. Hazards Earth Syst. Sci. 24, 3173–3205. https://doi.org/10.5194/nhess-24-3173-2024.
Vereecken, H., Huisman, J.A., Bogena, H., Vanderborght, J., Vrugt, J.A., Hopmans, J.W., 2008. On the value of soil moisture measurements in vadose zone hydrology: a review. Water Resour. Res. 46, 1–21. https://doi.org/10.1029/2008WR006829.
Vermote, E., Wolfe, R., 2021. MODIS/Terra Surface Reflectance Daily L2G Global 1km and 500m SIN Grid V061. https://doi.org/10.5067/MODIS/MOD09GA.061.
Vicente-Serrano, S.M., Beguería, S., López-Moreno, J.I., 2010. A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index. J. Clim. 23, 1696–1718. https://doi.org/10.1175/2009JCLI2909.1.
Walker, D.W., Oliveira, J.L., Cavalcante, L., Kchouk, S., Ribeiro Neto, G., Melsen, L.A., Fernandes, F.B.P., Mitroi, V., Gondim, R.S., Martins, E.S.P.R., van Oel, P.R., 2024. It's not all about drought: what "drought impacts" monitoring can reveal. Int. J. Disaster Risk Reduct. 103, 104338. https://doi.org/10.1016/j.ijdrr.2024.104338.
Wan, Z., Hook, S., Hulley, G., 2021. MODIS/Terra Land Surface Temperature/Emissivity Daily L3 Global 1km SIN Grid V061. https://doi.org/10.5067/MODIS/MOD11A1.061.
Wang, Yuchi, Cui, J., Miao, B., Li, Z., Wang, Yongli, Jia, C., Liang, C., 2024. Evaluating performance of multiple machine learning models for drought monitoring: a case study of typical grassland in Inner Mongolia. Land 13. https://doi.org/10.3390/land13060754.
Wang, L., Qu, J.J., 2007. NMDI: a normalized multi-band drought index for monitoring soil and vegetation moisture with satellite remote sensing, 34, 1–5. https://doi.org/10.1029/2007GL031021.
Wei, W., Zhang, H., Ma, L., Wang, X., Guo, Z., Xie, B., Zhou, J., Wang, J., 2022. Reconstruction and application of the temperature-vegetation-precipitation drought index in mainland China based on remote sensing datasets and a spatial distance model. J. Environ. Manag. 323, 116208. https://doi.org/10.1016/j.jenvman.2022.116208.
Xiao, X., Ming, W., Luo, X., Yang, L., Li, M., Yang, P., Ji, X., Li, Y., 2024. Leveraging multisource data for accurate agricultural drought monitoring: a hybrid deep learning model. Agric. Water Manag. 293, 108692. https://doi.org/10.1016/j.agwat.2024.108692.
Xu, X., Chen, F., Wang, B., Harrison, M.T., Chen, Y., Liu, K., Zhang, C., Zhang, M., Zhang, X., Feng, P., Hu, K., 2024. Unleashing the power of machine learning and remote sensing for robust seasonal drought monitoring: a stacking ensemble approach. J. Hydrol. 634, 131102. https://doi.org/10.1016/j.jhydrol.2024.131102.
Yeo, I., Johnson, R.A., 2000. A new family of power transformations to improve normality or symmetry. Biometrika 87, 954–959. https://doi.org/10.1093/biomet/87.4.954.
Yu, T., Guli·Jiapaer, Bao, A., Zhang, J., Tu, H., Chen, B., De Maeyer, P., Van de Voorde, T., 2023. Evaluating surface soil moisture characteristics and the performance of remote sensing and analytical products in Central Asia. J. Hydrol. 617, 128921. https://doi.org/10.1016/j.jhydrol.2022.128921.
Yu, C., Huang, X., Chen, H., Huang, G., Ni, S., Wright, J.S., Hall, J., Ciais, P., Zhang, J., Xiao, Y., Sun, Z., Wang, X., Yu, L., 2018. Assessing the impacts of extreme agricultural droughts in China under climate and socioeconomic changes. Earth's Future. https://doi.org/10.1002/2017EF000768.
Zhang, H., Loaiciga, H.A., Sauter, T., 2024. A novel fusion-based methodology for drought forecasting. Remote Sens. https://doi.org/10.3390/rs16050828.
Zhang, G., Su, X., Ayantobo, O.O., Feng, K., 2021. Drought monitoring and evaluation using ESA CCI and GLDAS-Noah soil moisture datasets across China. Theor. Appl. Clim. 144, 1407–1418. https://doi.org/10.1007/s00704-021-03609-w.
Zhou, Z., Wang, P., Li, L., Fu, Q., Ding, Y., Chen, P., Xue, P., Wang, T., Shi, H., 2024. Recent development on drought propagation: a comprehensive review. J. Hydrol. 645, 132196. https://doi.org/10.1016/j.jhydrol.2024.132196.
Zhu, Y., Liu, ·Yi, Wang, W., Singh, V.P., Ren, L., 2021. A global perspective on the probability of propagation of drought: From meteorological to soil moisture. J. Hydrol. 603, 126907. https://doi.org/10.1016/j.jhydrol.2021.126907.
Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. J. R. Stat. Soc. Ser. B Stat. Method. 67, 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x.