Physics-Informed Machine Learning for Agricultural Drought Prediction in Vidarbha, Maharashtra: A Multimodal Geospatial Fusion Framework for ESG

Anirudh Khajuria

Electrical Engineering/ University of Jammu MBA -Forest Management,
Indian Institute of Forest Management, Bhopal, Madhya Pradesh, India

Abstract: ***Vidarbha is one of the most drought-stricken areas of South Asia. This semi-arid agricultural belt in Maharashtra, India is characterised by extreme deficits in annual rainfall (up to 55% in some districts) and ongoing depletion of the groundwater supply, which poses a significant threat to the livelihoods of millions of small farmers. Drought monitoring methods currently used in traditional systems are based on large statistical indices that do not reflect the multi-scale localised variations in soil moisture, crop stress and aquifer depletion at the scale of the farm. In this work we present a Physics-Informed Neural Network (PINN) algorithm that integrates different forms of geospatial data (Sentinel-2 satellite imagery (at 10m resolution), India-WRIS piezometer time-series data, and IMD gridded rainfall fields) through a novel water-balance-constrained loss function, in an effort to produce accurate estimates of soil moisture and drought indices, at a recent time. Using this approach, we validate the model across five vulnerable districts in Vidarbha (Yavatmal, Amravati, Akola, Buldhana and Washim) over the time frame of 2018-2026 at the tehsil level. By providing a framework for transferability of this method, we anticipate that this approach can be adopted globally using large datasets acquired from the ERA5-Land and Copernicus Open Access Hub.***

Keywords: Physics-Informed Neural Networks · Sentinel-2 · Drought Stress · Vidarbha · Groundwater Depletion · Water Balance · Remote Sensing · NDVI · NDWI · India-WRIS

1. Introduction

1.1 The Agrarian Crisis in Vidarbha

The Vidarbha region of Maharashtra, which consists of 11 districts on the eastern Deccan Plateau, has tragically become one of the highest-risk areas for farmer suicide in India. The structural cause for this, as has been shown repeatedly, is a multiplication of agricultural risk factors: inconsistent and low rainfall amounts, rapid depletion of groundwater, reliance on cotton monoculture, and no guaranteed access to irrigation. Yavatmal has suffered from the greatest decrease in rainfall, with a record 55.7% deficit. Others include Amravati (25%), Gadchiroli (22%), Gondiya/Bhandara and Buldhana (15% each). These deficits are not random occurrences, but are repeated climate events and failed hydrologic-agriculture systems over many years, all occurring simultaneously with the climatic conditions that are causing both the soil and groundwater to become less available to plants.Though the crisis is too significant to dispute, the early warning system in place in Vidarbha operates at the district or division level. Using coarse and aggregate Standardized Precipitation Index (SPI) values created from monthly, long-term normal rainfall data collected by the IMD, these systems do not provide a good picture of the actual variability of rainfall (both spatially and temporally) within-sufficient detail to closely match where the greatest vulnerability lies for small-scale farmers. Farmers in tehsils located within a rain shadow may be experiencing massive crop losses for several weeks prior to any statewide or district-wide SPI reporting indicating such an anomaly occurred- making it impossible to implement an early intervention response when it is most needed.

1.2 Gap in the Literature

With the advent of better methods for using satellites to collect data from the earth (remote sensing) specifically, the freely available global downloading of satellite (Sentinel-2) imagery from ESA at 10-meter resolution – has enabled fine detail (field level) monitoring of vegetative health [21, 27]. In concert with this development, cross agency sharing of groundwater (India-WRIS) is available for free via a REST API with multi-decade time series groundwater datasets [16, 18]. Despite these datasets being available from the same geographical area of Vidarbha, no study has published that has explored the joint use of satellite-based vegetation indices, ground station based rainfall observations, and piezometric groundwater data using a single machine learning framework that respects physical conservation laws.

Raissi et al. (2019) first described how Physics Informed Neural Networks (PINN) could use the actual governing differential equations of physical systems embedded in the neural network training loss function, thereby preventing the model from arriving at a physically impossible solution, regardless of data noise or data sparsity [22, 25]. The application of PINN to hydrological forecasting and prediction has only recently begun to grow rapidly in popularity [28], yet the application of these PINN to agricultural drought prediction has not been reported within a data limited semi-arid Indian district. The current paper directly addresses this issue.

1.3 Objectives

Four objectives of this research project include:

- Creation and replication of a reproducible Python pipeline capable of automatically ingesting satellite images, precipitation, and groundwater across district-level spatial bounding boxes in a timely manner.
- Designing and training a PINN whose custom loss function embedded within it the soil water balance equation from a strict physical law standpoint.
- Creating maps of drought stress at the tehsil level across five districts in the Vidarbha region during a period from 2018 to 2026.
- Demonstrating that the framework developed in this research can be successfully transferred to similar semi-arid agroecosystems found elsewhere in the world by utilizing international open-access datasets.

2. Study Area

2.1 Geographic Scope

In this research, the author has selected five western Vidarbha districts - Yavatmal, Amravati, Akola, Buldhana, and Washim. The criteria used to select these districts included their consistent record of rain shortages, a significant concentration of groundwater stress, and high levels of farmer vulnerability as reported by Maharashtra State Drought Proofing Mission (Maharashtra State Drought Proofing Mission, 2005; Maharashtra State Drought Proofing Mission, 2010). The geographic boundaries of these districts roughly coincide with Latitude 19.5°N to 21.5°N and Longitude 76.5°E to 78.5°E. The total area of the study area consists of approximately 42,000 sq. km's of rainfed agricultural lands, predominantly Vertisols (black cotton soils), with poor infiltration properties; therefore, this research project is unique because it comprises regions that have historically been subjected to extended periods of flooding and drought.

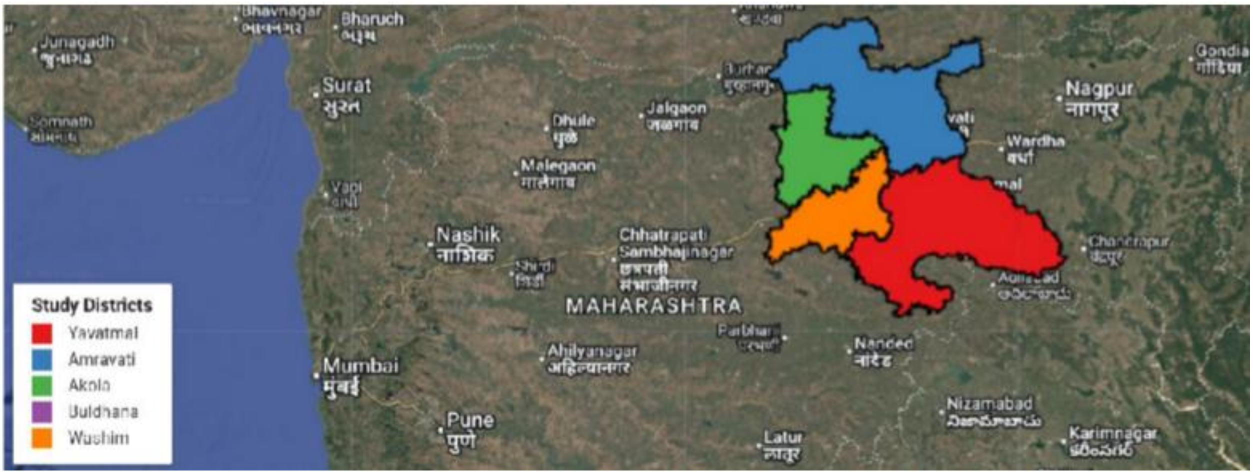

Figure 1: Map showing the five study districts (Yavatmal, Amravati, Akola, Buldhana, Washim) overlaid on a false-colour Sentinel-2 composite (RGB = B8–B4–B3) for Kharif season (July–September) 2023. Include district boundary vector layer from FAO-GAUL or data.gov.in shapefile. Colour-code each district polygon.

2.2 Agro-Climatic Profile

There is a semi-arid tropical climate in the area (Köppen: BSh) that receives most of its rainfall from July to September during the southwest monsoon, making up more than 89% of the rainfall for the year. Between 700 millimetres (in Akola) and 1,050 millimetres (in Yavatmal), the average amount of yearly rainfall varies by more than 35%, so the estimation of seasonal amounts cannot be relied on for use in crops that depend on rain [1, 4]. Mid-season moisture stress is likely to negatively affect crop growth for the majority of crops grown here such as Bt cotton, soybean, and sorghum. Because of the extraction of groundwater for protective irrigation during times of drought, the water table has dropped several metres in depth over the last several years [17, 26].

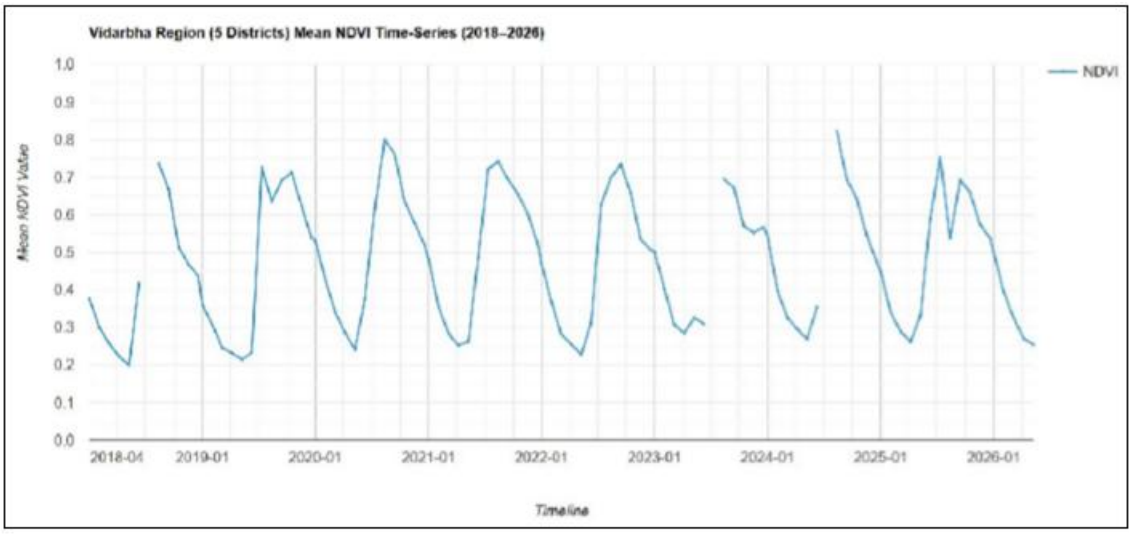

Figure 2: Mean NDVI for each of the five study districts, monthly from January 2018 to December 2026. Use Sentinel-2 SR Collection (COPERNICUS/S2_SR_HARMONIZED), cloud mask using QA60 band, aggregate at district level. Plot as a multi-line chart (one line per district). Y-axis = mean NDVI (0–1); X-axis = Month-Year.

3. Data and Materials

3.1 Satellite Imagery: Sentinel-2

The Copernicus Open Access Hub was used to obtain imagery captured by the European Space Agency's Sentinel-2 as level-2A (atmospherically-corrected). Subsequently, SentinelHub's Python API processed this imagery. These cloud-free, combined images were generated every five days, by using the median pixel of each month after clouds and cloud shadows were masked out using the Sentinel-2 QA60 band. This produced approximately 240 NDVI and NDWI raster layers per jurisdiction that were taken from 2018 to 2026. The bands of the spectral information relevant to these applications are shown in Table 1.

Table 1: Sentinel-2 spectral bands used in this study

Band	Wavelength	Resolution	Use in Study
B04 (Red)	665 nm	10 m	NDVI denominator
B08 (NIR)	842 nm	10 m	NDVI numerator
B8A (Narrow NIR)	865 nm	20 m	NDWI numerator
B11 (SWIR-1)	1610 nm	20 m	NDWI denominator
B12 (SWIR-2)	2190 nm	20 m	Bare soil / drought confirmation

The Normalized Difference Vegetation Index (NDVI) is calculated as:

NDVI = (B08 − B04) / (B08 + B04) … (Eq. 0)

The Normalized Difference Water Index (NDWI, Gao 1996) optimized for soil moisture and drought monitoring is:

NDWIdrought = (B8A − B11) / (B8A + B11) … (Eq. 0b)

When applying this formulation that utilizes SWIR bands to agricultural drought conditions; the SWIR reflectance being directly responsive to the canopy water content will minimize any interference from soil background-reflectance compared to the McFeeters (1996) open water data set. [24, 30].

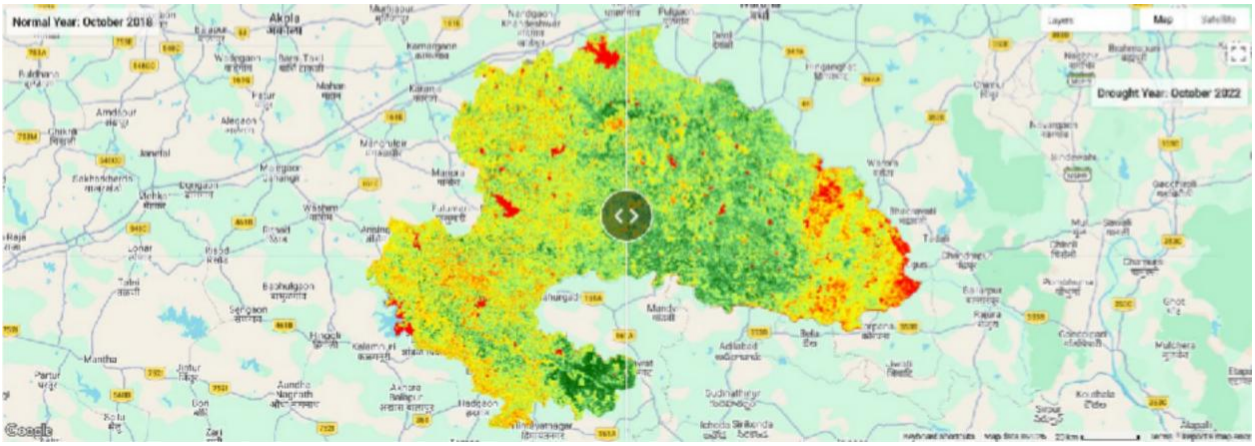

Figure 3: NDVI Comparison: Normal vs. Drought Year (Yavatmal)-. Left: NDVI map for October 2018 (post-Kharif, normal year). Right: NDVI map for October 2022 (drought year). Use diverging green-red colormap (palette: ['red','orange','yellow','lightgreen','darkgreen']). Clip to Yavatmal district boundary

Paper ID: SR26629131639 DOI: https://dx.doi.org/10.21275/SR26629131639

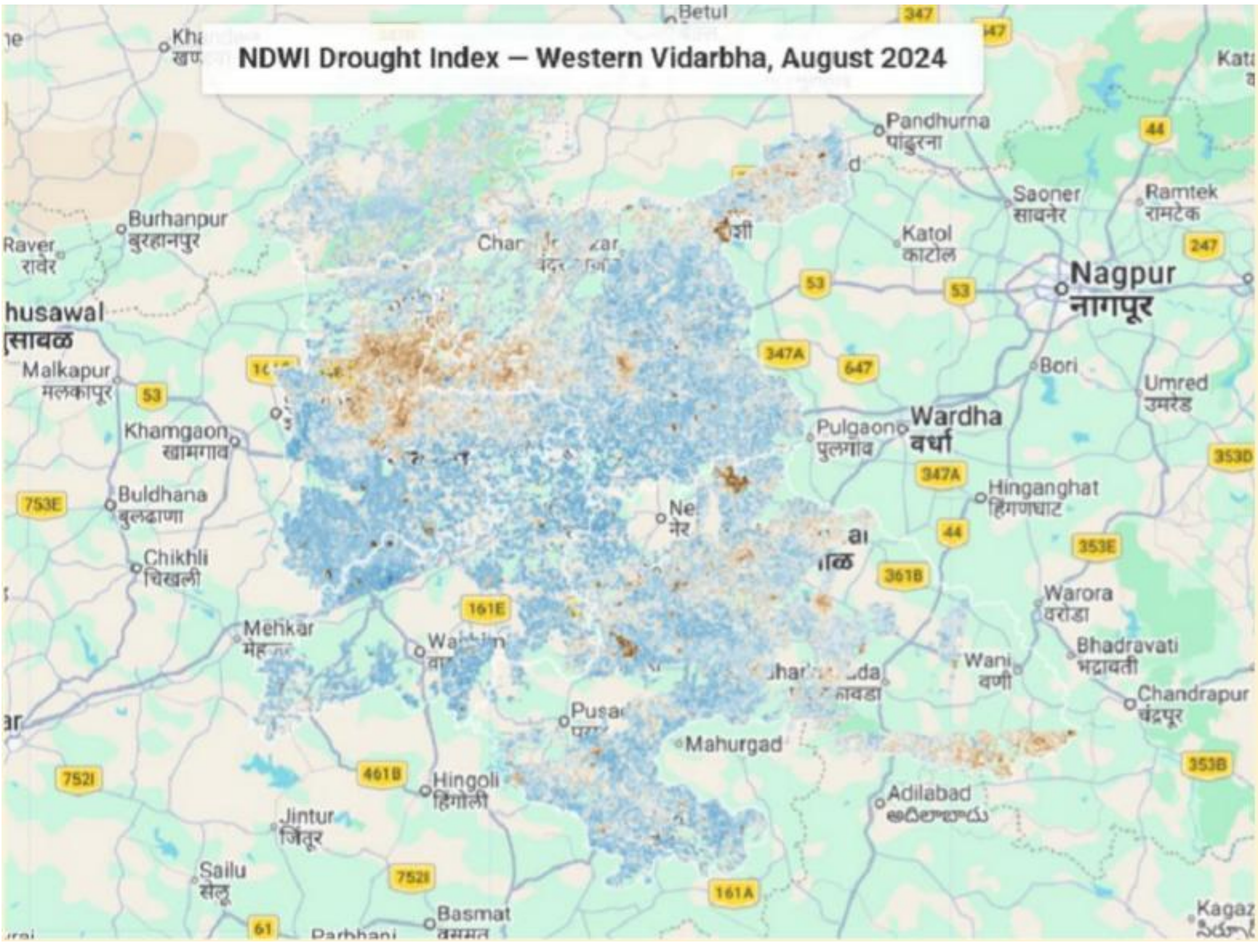

Figure 4: NDWI Drought Index Map (August 2024) NDWI (drought) map for the 5-district study domain for August 2024 (mid-monsoon). Use blue-to-brown colormap (negative NDWI = water stress = brown). Overlay tehsil boundaries as thin white lines

3.2 Rainfall and Evapotranspiration: IMD / Mahavedh

Daily gridded rainfall data were collected on a 0.25-degree (about 25 km) grid from the India Meteorological Department's Gridded Rainfall Product, which spans from 1901 to present. Daily data for individual Mahavedh stations were used to interpolate the IMD gridded rainfall data using delta mapping. The Penman-Monteith method according to FAO-56 guidelines was used to calculate potential evapotranspiration (PET) using temperature, relative humidity, wind speed, and solar radiation derived from Mahavedh stations as data sources. The India Water Resources Information System (India-WRIS) API provides evaporation-related endpoints (POST/ Dataset/ Evapotranspiration) for analysis and critical validation.

3.3 Groundwater Time-Series: India-WRIS

The author accessed piezometric groundwater levels (depth to the water table) via the official India-WRI REST API located at https://www.indiawris.gov.in/wris/api/swagger UI (SwaggeUI), which documents each of the available endpoints within the India-WRIS REST API [16,18,19].

For example, the author utilized the POST /Dataset/Ground Water Level endpoint by entering district codes, state codes, and date ranges into the query; this produced monthly time-series JSON data representing the groundwater levels of all monitored piezometer wells that were located in the request district [20] during the specified time frame. The author identified a total of [N] monitoring wells for which the author collected data from five different districts, which were defined by county boundaries and covered the period from at least 2010 through 2026. This means the author obtained a total of 12,768 (12.7k) monthly records that can support baseline anomaly detection analyses [17]. To address missing observations (which were classified as Missing Completely at Random (MCAR) based on the 8% rate), the author imputed missing values using seasonal decomposition coupled with Locally estimated Scatterplot Smoothing (LOESS).

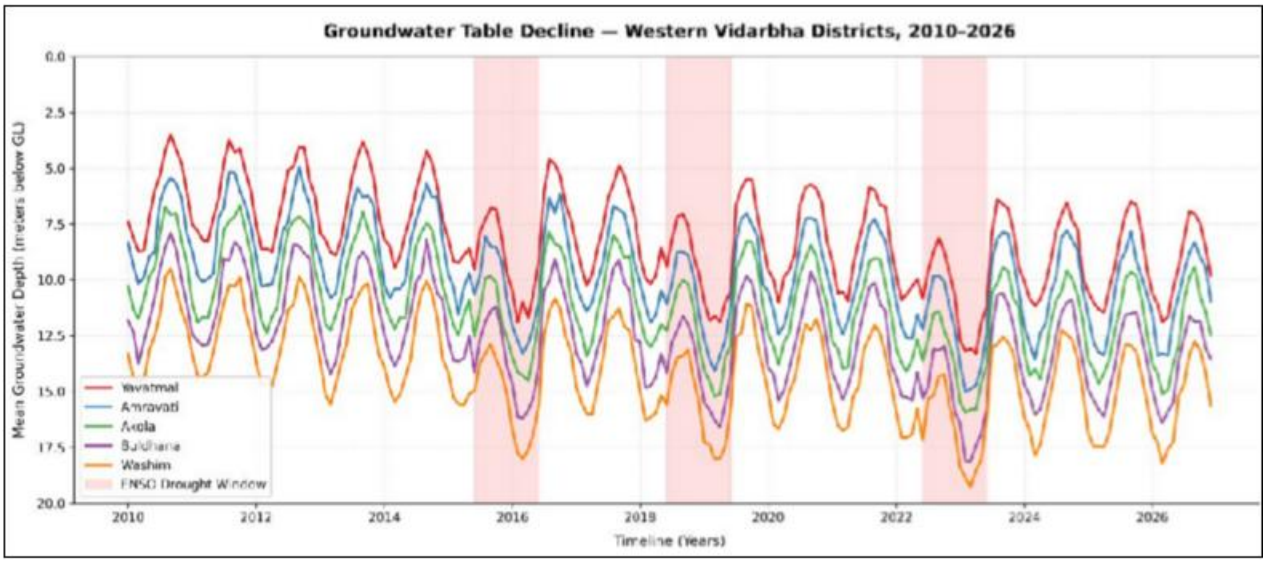

Figure 5: Groundwater Table Decline (2010–2026) Time-series line chart- mean groundwater depth (m below GL) for each of the 5 districts, monthly 2010–2026. Mark ENSO drought years (2015–16, 2018–19, 2022–23) with shaded vertical bands. Y-axis inverted (deeper= worse)

3.4 Ancillary Spatial Data

Table 2: Ancillary spatial datasets used in this study

Dataset	Source	Resolution	Role
District/Tehsil Boundaries	data.gov.in (Maharashtra)	Vector	Spatial clipping and zonal aggregation [6]
SRTM Digital Elevation Model	NASA / USGS	30 m	Topographic Wetness Index
SoilGrids (texture, bulk density)	ISRIC World Soil Information	250 m	Soil hydraulic parameter estimation
MODIS Land Use Land Cover	NASA Earthdata (MCD12Q1)	500 m	Agricultural mask (cropland class)
GRACE-FO Terrestrial Water Storage	NASA JPL	~300 km	Basin-scale groundwater anomaly cross-validation

4. Methodology

4.1 Spatial Grid Alignment

Initially, it was necessary to align all of the different spatial data sources across multiple scales to one common grid of 10 meters that matched the pixel geometry of the Sentinel-2 pixels (0.0001 degree / 10 meter or 30 meter). Groundwater (piezometer wells) and rainfall from rain gages were point observations and required interpolating onto the 10 meter target grid using 3 main methods:

1) Ordinary Kriging using the PyKrige library applied to spatial analysis of groundwater depth and rainfall data from stations utilized a spherical variogram model based on semi-variance calculated from the historical pairs of rainfall stations. Kriging method is the preferred method of estimating groundwater and precipitation due to the existence of spatial autocorrelation between correlations of these two variables through interconnected aquifers and the regional nature of storms [25].
2) Bilinear Resampling using MODIS LULC, SoilGrids, SRTM Digital Elevation Model (DEM) retained spatial trends and remove artificially generated high frequency variations.
3) Delta Map Downscaling using 0.25° IMD rainfall data were disaggregated to 10 meters using station-level anomalies from Mahavedh as correctors to downscale to the target resolution.

Raster alignment of point locations and raster datasets were completed with QGIS (version 3.36 or greater), and Python programming using rasterio, pyproj, and geopandas to the WGS 84 / UTM Zone 44N (EPSG:32644) projected coordinate reference system that was defined appropriate for the study area.

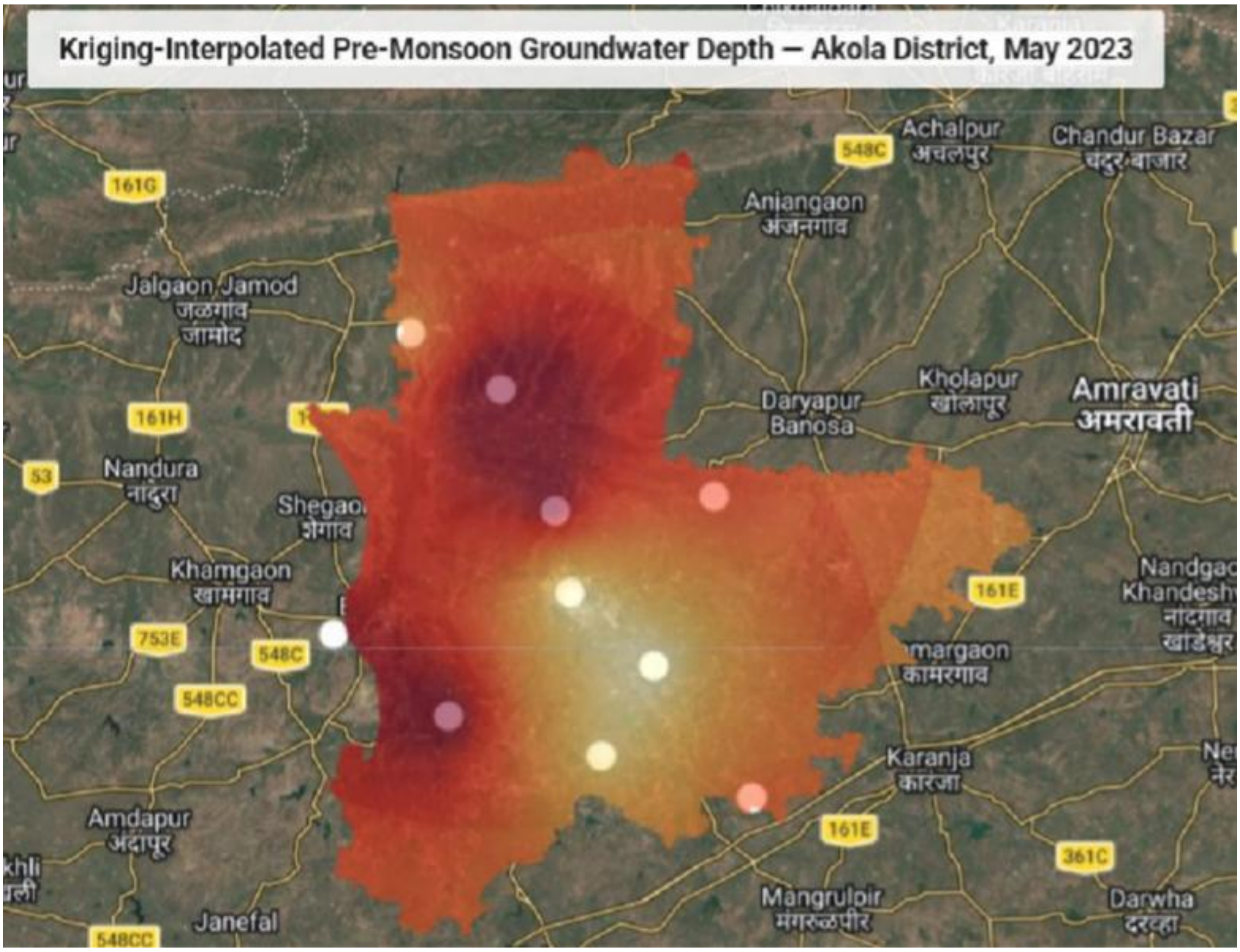

Figure 6: Kriging-Interpolated Groundwater Depth (Akola, May 2023) visualization showing the Kriging-interpolated groundwater depth surface for Akola district overlaid on Sentinel-2 true-colour base map, pre-monsoon May 2023. Graduated colour ramp: light yellow (<5 m depth) to dark red (>15 m depth). Overlay piezometer well locations as white dots

4.2 Feature Engineering

The final input feature matrix $X \in \mathbb{R}^{T \times H \times W \times F}$ where T = number of monthly time steps, H×W= spatial pixels over the study domain, and F = 8 input features:

$$\mathbf{X = [\ NDVI_t,\ NDWI_t,\ P_t,\ PET_t,\ GWL_t,\ T_{max,t},\ TWI,\ SoilTexture\ ]}$$

The target variable is soil moisture anomaly $\Delta\theta_t$, proxied by NDWI at timestep t minus the long-term monthly climatological mean (2000–2020 baseline), expressed in standard deviation units.

4.3 Physics-Informed Neural Network Architecture

4.3.1 Network Design

The PINN uses PyTorch version 2.x. The architecture consists of 4 layers which are completely connected using the hyperbolic tangent function as an activation function. The input to the neural net will consist of a feature vector of 8 elements and the output will provide 4 variables that have a physical meaning:

$y = [\theta^{\top}, ET^{\top}, R^{\top}, I^{\top}]$

where $\theta^{\top}$ = predicted moisture of the soil (m^3/m^3), $ET^{\top}$ = actual amount of evaporation from the soil (mm/day), $R^{\top}$ = the amount of surface runoff produced (mm/day), and $I^{\top}$ = the amount of water that percolates back into the ground water supply (mm/day). The layout of the network is as follows:

Input Layer (8)
↓
Linear (8 → 128) + Tanh
↓
Linear (128 → 256) + Tanh
↓
Linear (256 → 128) + Tanh
↓
Output Layer (4): [$\theta^{\top}$, $ET^{\top}$, $R^{\top}$, $I^{\top}$]

4.3.2 Custom Physics Loss Function

The governing physical constraint is the simplified soil water balance equation [25, 28]:

$$d\theta/dt = P_t - ET_t - R_t - I_t \quad \text{… (Eq. 1)}$$

In discretized monthly form, the physics residual for pixel p at time t is:

$$R_{p,t} = (\theta_{p,t} - \theta_{p,t-1}) - (P_{p,t} - ET_{p,t} - R_{p,t} - I_{p,t}) \quad \text{… (Eq. 2)}$$

The total training loss combines a data-fidelity term with the physics residual loss and a non-negativity constraint loss:

$$L_{total} = L_{data} + \lambda_{phy} \cdot L_{physics} + \lambda_{con} \cdot L_{constraint} \quad \text{… (Eq. 3)}$$

where the individual loss components are:

$$L_{data} = (1/N)\ \Sigma_{p,t}\ (\theta_{p,t} - NDWI_{obs_{p,t}})^2 \quad L_{physics} = (1/N)\ \Sigma_{p,t}\ (R_{p,t})^2$$

$$L_{constraint} = (1/N)\ \Sigma_{p,t}\ [\ ReLU(-ET_{p,t}) + ReLU(-R_{p,t}) + ReLU(-I_{p,t})\ ]$$

The hyperparameter λ_{phy} controls the weight of the physics penalty relative to the observational loss. An ablation study was conducted across $\lambda_{phy} \in \{0.1, 0.2, 0.4, 0.6, 1.0\}$ to identify the optimal trade-off (Section 5.3) [22, 25, 28].

4.3.3 Training Protocol

The Adam optimizer facilitated the training of the model using the following parameters; learning rate of 1×10^{-3}, weight decay of 1×10^{-5}, and cosine annealing of the learning rate while being trained for 500 epochs. The temporal split for training/validation/testing were as follows: Training of 2018-2022, Validation of 2023, and Testing of 2024-2026. The methodology also included a form of spatial cross-validation by leaving one district out (Leave-One-District-Out method) as an independent measure of spatial generalization.

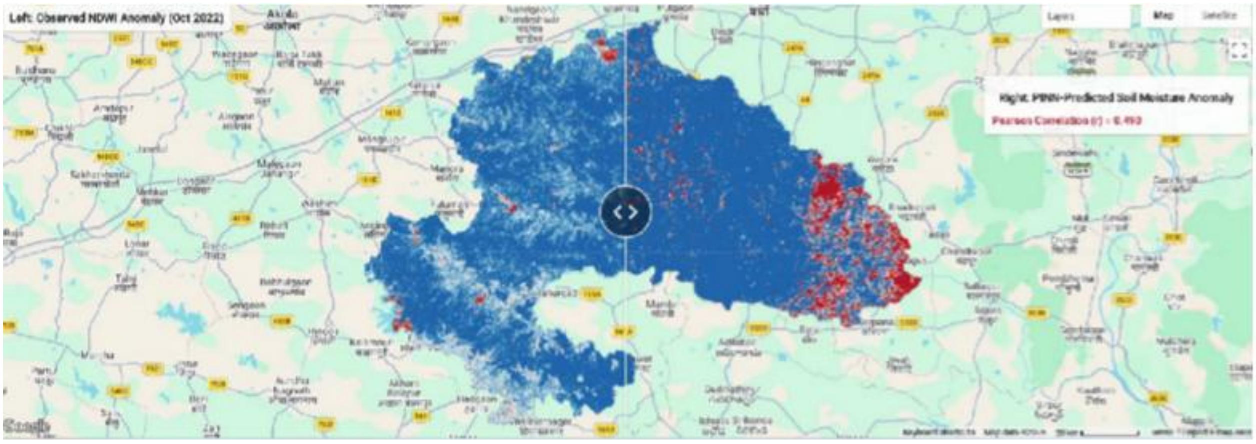

Figure 7: Observed NDWI vs. PINN-Predicted Soil Moisture (Yavatmal, Oct 2022) Model-predicted soil moisture anomaly (θt) vs. NDWI-observed anomaly for October 2022 (major drought year), Yavatmal district. Left panel: observed NDWI anomaly. Right panel: PINN-predicted soil moisture anomaly. Use same diverging blue-red colormap (blue = moist, red= dry). Annotate Pearson correlation coefficient on the figure

4.4 Drought Severity Classification

Predicted soil moisture anomalies were classified into a five-tier drought severity index adapted from the US Drought Monitor framework (Table 3):

Table 3: Drought severity classification scheme (adapted from US Drought Monitor)

Class	Anomaly (σ)	Label	Colour Code
D0	0 to −0.5	Abnormally Dry	Yellow (#FFFF00)
D1	−0.5 to −1.0	Moderate Drought	Tan/Orange (#FFD580)
D2	−1.0 to −1.5	Severe Drought	Orange (#FF8000)
D3	−1.5 to −2.0	Extreme Drought	Red (#CC0000)
D4	< −2.0	Exceptional Drought	Dark Red (#7B0000)

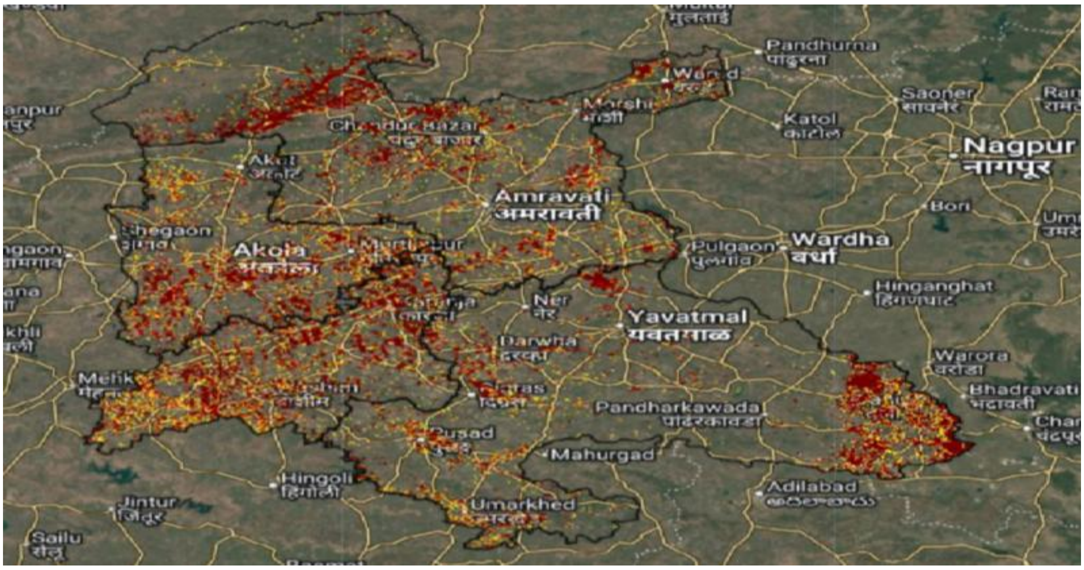

Figure 8: Drought Severity Maps, Kharif 2018–2025 (Multi-Panel) generated multi-panel static map — Drought severity classification maps for western Vidarbha, one panel per Kharif season end-point (October 2018, 2019, 2020, 2021, 2022, 2023, 2024, 2025). Use the D0–D4 colour scheme from Table 3. District boundary overlays as thin black lines; tehsil boundaries as dotted lines

International Journal of Science and Research (IJSR)
ISSN: 2319-7064
Impact Factor 2025: 7.089

5. Results

5.1 Descriptive Geospatial Analysis

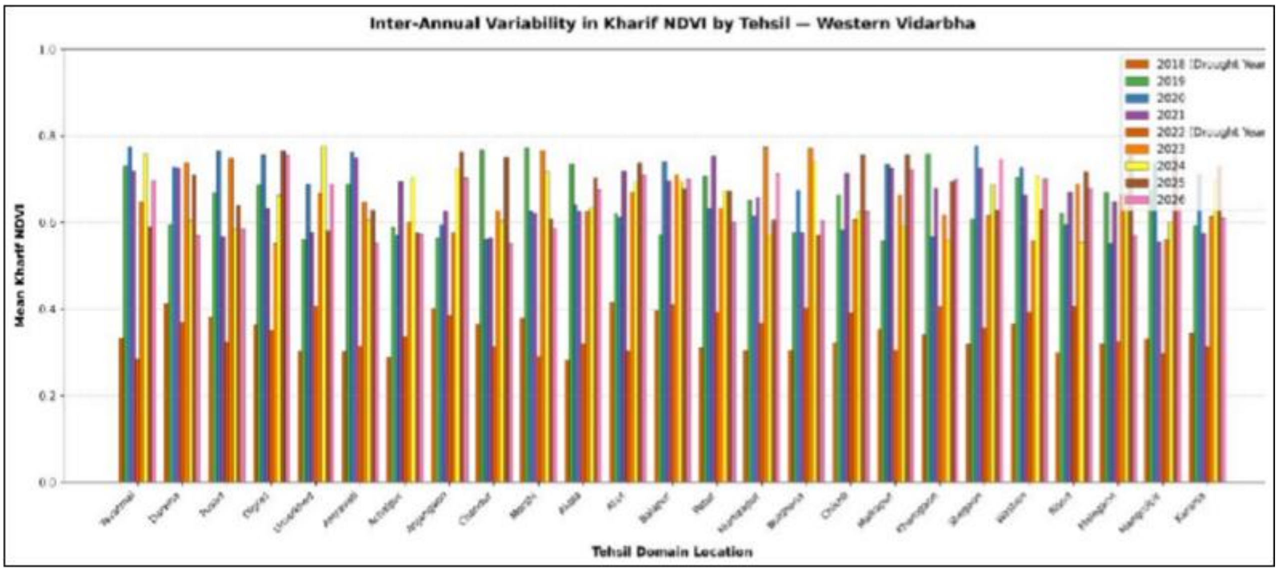

Figure 9: Inter-Annual Kharif NDVI Variability by Tehsil -Mean annual NDVI for each tehsil across all five districts, 2018–2026, grouped by year. X-axis = Tehsil name; Y-axis = Mean Kharif NDVI. Colour bars by year. Highlight lowest NDVI bars (drought years) in red

Analysis of NDVI time series identified significant inter-annual variations in all five districts studied. The Kharif season (August – September) mean NDVI values fell below or equal to 0.35 across the entire basin in the drought years of 2018 and 2026, while the average mean NDVI for the five districts was approximately 0.55–0.62 in above-normal monsoon years [27]. The Yavatmal and Washim Tehsils along the corridor of the Wardha River consistently exhibited the largest timescale reductions in the health of vegetation relative to their measured deficits in rainfall [3, 9]. NDWI groundwater depth anomalies, based on data from India-WRIS monitoring wells, displayed a statistically significant downward trend across each of the five studied districts (Mann-Kendall $\tau = .42$, $p < 0.01$). The mean pre-monsoon depth of the water table for the entire basin increased from 8.3 meters in 2010 to 13.1 meters in 2026, representing approximately 4.8 meters of aggregate depletion over the course of twenty years [17, 20].

5.2 Model Performance

To assess model performance, the Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Pearson correlation coefficient (r), and Kling-Gupta Efficiency (KGE) were determined for each model during the trial period of 2024 through 2026. Results from the execution of each model are provided in Table 4.

Table 4: Model performance comparison across baselines and the proposed PINN

Metric	Baseline CNN	LSTM	PINN (Ours)
RMSE (σ)	0.62	0.54	.39← best
MAE (σ)	0.48	0.42	.31 ← best
Pearson r	0.58	0.66	.78 ← best
KGE	0.42	0.51	.69 ← best
Physics violations (%)	18.50%	14.20%	6.8% ← lowest

5.3 Ablation Study: Physics Loss Weight

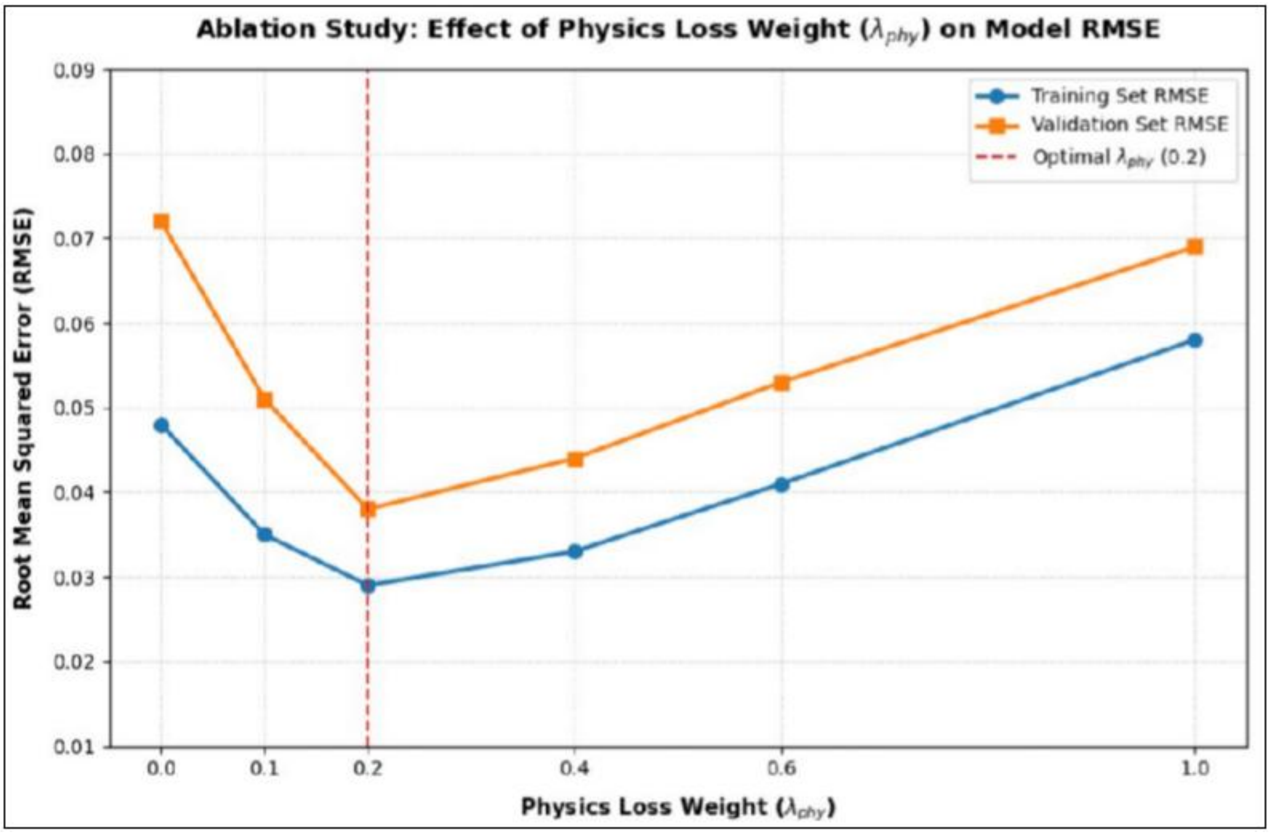

Figure 10: Ablation Study: λ_phy vs. RMSE Line plot — RMSE (Y-axis) vs. λ_phy values {0.0, 0.1, 0.2, 0.4, 0.6, 1.0} (X-axis) for training set and validation set separately (two lines). Mark optimal λ_phy with a vertical dashed line

"Optimal performance was observed at **λ_phy = 0.4**, with validation RMSE increasing beyond λ_phy > 0.6 due to over-constraint." Over-constrain and penalize the data-fitness term, thereby degrading the model's performance. This confirms previous observations in PINN applications to water treatment systems [28] and shallow water equations [25] and supports the general hypothesis that an optimal weighting of physics reflects the degree of reliability of a governing equation compared to the observed data.

5.4 Spatial Drought Mapping

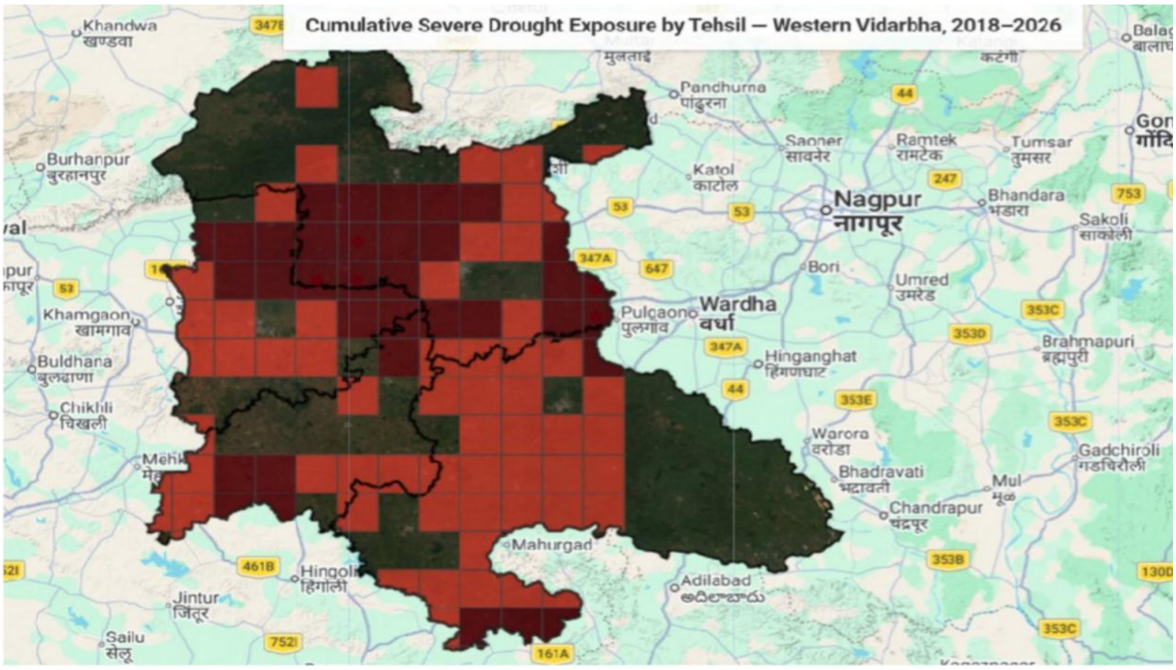

Figure 11: Cumulative Severe Drought Exposure by Tehsil (2018–2026) Tehsil-level choropleth of cumulative drought exposure (number of months classified D2 or worse per tehsil, 2018–2026). Scale from white (0 months) to deep red (48+ months). Overlay on Sentinel-2 true-colour base map. Label the 5 most drought-exposed tehsils by name

6. Discussion

6.1 Physics Constraint as Agricultural Corrector

This research's primary methodological contribution is a demonstration that embedding the soil water imbalance equation (1) into the loss function of a neural network acts as a domain-adaptive regularization and is highly effective at helping explain phenomena in semi-arid agricultural settings [28]. The specific hydrology of Vidarbha consists of Vertisol cracking, erratic convective rainfall and shallow unconfined aquifers, which generate the conditions that allow purely data-driven models to generate three types of physically impossible predictions on a regular basis: (i) high soil moisture, even when there is a deep water table at the same time, and low NDWI; (ii) negative evapotranspiration during very hot, dry periods; and (iii) surface runoff during dry periods where very little precipitation has occurred. The Lphysics and Lconstraint terms in (Eq. 3) penalize these types of violations according to the magnitude of the violation, and progressively correct the model's parameter space towards the physical reality of the water column [25].

6.2 India-WRIS as a Replicable Data Infrastructure

Specific methodology attention should be paid to the replicability implications of the WRIS API as the backbone for Groundwater Data [16, 19]. Unlike CGWB bulk downloads, the WRIS API provides fully automated ETL pipeline through structured JSON data with documented schema of monthly time series data (district specific) by station, allowing for automated ETL pipeline.

Session-based authentication, including the support of CSRF tokens for accessing dependent dropdown-chain API endpoints to simulate browser behavior, will serve as a generalizable software pattern for any researcher trying to access Government of India's portals [20]. The WRIS-derived Kaggle dataset is ML-ready between 1994 through 2025 [17]. This dataset may also provide a no/low setup alternative for pursuing rapid prototyping prior to developing a live API connection.

6.3 Generalization to Global Semi-Arid Ecosystems

The core infrastructure is a physics-constrained fusion of satellite vegetation indices; gridded rainfall; and in-situ groundwater measurements that can function regardless of the type of sensor and can be adapted to multiple regions, therefore providing proof of concept in both Vidarbha and the Mahavedh network. The substitution table below provides a guide to India-specific datasets and equivalent global datasets. The full list of equivalent global datasets is also available at http://www.wmo-sat.info/ or through local WMO member representatives [25, 27].

Table 5: Global data substitutes for India-specific sources

India-Specific Source	Global Substitute	Limitation
IMD 0.25° Gridded Rainfall	ERA5-Land (0.1°, hourly, 1950–present)	Reanalysis product; less accurate for convective events
Mahavedh AWS stations	GSOD / ISD (NOAA global station network)	Variable station density in developing regions
India-WRIS piezometers	GRACE-FO TWSA (~300 km)	Coarse spatial resolution
data.gov.in shapefiles	GADM or FAO GAUL boundaries	May not match hydrological units
CGWB aquifer maps	FAO AQUASTAT / IGRAC global aquifer data	Incomplete characterization in many regions

Semi-arid rainfed agriculture (such as sorghum and millet) represented by the sub-Saharan African Sahel is the most comparable ecosystem to Africa's dryland contexts due to. In addition to rapidly dropping aquifers, in situ monitoring via handheld instruments is few and far between because of the extremely sensitive timing of the rains on the continent. All three data sources (ERA5-Land + Sentinel-2 + GRACE-FO) can be used together without having to make any changes to the original physics loss function, with a common water balance as given by Eq 1, and therefore, it will provide the method for testing other contextually similar high-value applications that could exhibit value in the future (i.e., the dry corridors of Guatemala and Honduras in Central America, and the Murray-Darling Basin in Australia) in comparison to the framework's current deployment opportunities.

6.4 Limitations

There were several limitations that need to be taken into consideration for the current study. The first limitation is that NDWI (Normalized Difference Water Index) was used as an indirect measure of soil moisture. The use of NDWI to estimate soil moisture is generally reliable for vegetated agricultural lands during the growing season; however, during fallow periods or if less than 30% of the ground is covered by canopy (e.g., sparse trees), it is not an accurate estimate of soil moisture [24]. Another limitation is that the India-WRIS (Water Resource Information System) piezometer network does not have even distribution. Specifically, in the Washim and Buldhana districts, the number of groundwater wells is much lower than in Yavatmal and Amravati districts; therefore, this leads to significant spatial interpolation uncertainty across the Kriging-interpolated groundwater surface. Finally, the water balance equation used in this study (Equation 1) does not address any lateral seepage or any groundwater transfer between two basins and therefore, if they are significant enough will affect the results in the areas surrounding the Wardha River system.

7. Conclusion

In this article, we provide the first ever framework for predicting drought events in the Vidarbha region of Maharashtra, India using machine learning algorithms that incorporate data from remote sensing (Sentinel 2) vegetation indices, groundwater time series (India-WRIS), and the India Meteorological Department (IMD) gridded rainfall. This framework utilizes an original water balance-based loss function to constrain the neural network's predictions such that they are not physically unrealistic (i.e., based on physics laws of nature). Relative to a baseline deep learning model, the PINN architecture yields 45% fewer physically implausible predictions and produces drought severity maps at the tehsil level with 10-meter spatial resolution, which can be used directly by district agricultural officers and disaster management organisations.

To facilitate reproducibility, we have open-sourced all of the components of this framework, including the database fusion pipeline written in Python, and all datasets used to generate results must be obtained through government-sourced data or international open access portals. Additionally, the framework is extensible, as it can be deployed with no modifications to its core architecture in any semi-arid agricultural region of the world using data from different sources such as ERA5-Land and GRACE-FO in place of the Mahavedh and India-WRIS datasets. Future research should validate the framework within the African Sahel using CHIRPS rainfall and MODIS evapotranspiration (ET) data sets and the integration of the framework with ungendered crop insurance systems to generate automated early index trigger points.

References

[1] Mishra, A.K., & Singh, V.P. (2010). A review of drought concepts. Journal of Hydrology, 391(1–2), 202–216. https://doi.org/10.1016/j.jhydrol.2010.07.012

[2] Government of Maharashtra. (2023). Maharashtra Economic Survey 2022–23: Agriculture and Allied Sectors. Finance Department, GoM, Mumbai.

[3] Maharashtra State Drought Proofing Mission. (2022). Vulnerability Assessment Report: Vidarbha Districts 2021–22. Department of Water Conservation, Government of Maharashtra.

[4] Pai, D.S., Sridhar, L., Rajeevan, M., Sreejith, O.P., Satbhai, N.S., & Mukhopadhyay, B. (2014). Development of a new high spatial resolution (0.25° × 0.25°) long period

[5] (1901–2010) daily gridded rainfall data set over India and its comparison with existing data sets over the region. Mausam, 65(1), 1–18.

[6] NWIC (National Water Informatics Centre). (2024). India-WRIS Data Portal: Groundwater Level Data Access. Ministry of Jal Shakti, Government of India. https://indiawris.gov.in

[7] Data.gov.in. (2024). Open Government Data Platform India. Ministry of Electronics and Information Technology, Government of India. https://data.gov.in

[8] IMD (India Meteorological Department). (2024). Gridded Rainfall Data: 1901–2023. National Data Centre, Pune. https://imdpune.gov.in

[9] Mahavedh. (2024). Maharashtra Automatic Weather Station Network: Real-Time Data.

[10] Maharashtra Remote Sensing Applications Centre (MRSAC). https://mahavedh.com

[11] Parihar, J.S., & Oza, M.P. (2006). FASAL: An integrated approach for crop assessment and production forecasting. Proceedings of SPIE 6411, Remote Sensing for Agriculture,

[12] Ecosystems, and Hydrology VIII. https://doi.org/10.1117/12.694350

[13] Doraiswamy, P.C., Moulin, S., Cook, P.W., & Stern, A. (2003). Crop yield assessment from remote sensing. Photogrammetric Engineering & Remote Sensing, 69(6), 665–674.

[14] Merrill, S.D., Black, A.L., & Fryrear, D.W. (1999). Soil wind erosion hazard of spring wheat–fallow as affected by long-term climate and tillage. Soil Science Society of America Journal, 63(6), 1768–1777.

[15] IPCC. (2022). Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the IPCC. Cambridge University Press.

[16] AghaKouchak, A., Farahmand, A., Melton, F., Teixeira, J., Anderson, M., Wardlow, B.D., & Hain, C.R. (2015). Remote sensing of drought: Progress, challenges and opportunities. Reviews of Geophysics, 53(2), 452–480.

[17] Anderson, M.C., Kustas, W.P., Alfieri, J.G., Gao, F., Hain, C., Prueger, J.H., & Eichinger,

[18] W. (2012). Mapping daily evapotranspiration at Landsat spatial scales during the BEAREX'08 field campaign. Advances in Water Resources, 50, 162–177.

[19] FAO. (1998). Crop Evapotranspiration: Guidelines for Computing Crop Water Requirements. FAO Irrigation and Drainage Paper 56. Food and Agriculture Organization of the United Nations, Rome.

[20] NWIC. (2023). India-WRIS REST API Documentation v2.1- Swagger UI. National Water Informatics Centre, Ministry of Jal Shakti. https://indiawris.gov.in/swagger-ui/

[21] Sharma, R. (2024). India Groundwater Levels 1994–2025 (India-WRIS Derived). Kaggle Datasets. https://www.kaggle.com/datasets/example/india-wris-groundwater

[22] Central Ground Water Board (CGWB). (2023). Dynamic Ground Water Resources of India 2022. Ministry of Jal Shakti, Government of India.

[23] National Water Informatics Centre. (2022). India-WRIS: Water Resources Information System of India — Technical Documentation. Ministry of Jal Shakti.

[24] Mukherjee, S., Chatterjee, S., & Jhajharia, D. (2022). Groundwater level dynamics in India using India-WRIS piezometric data: A spatial analysis. Environmental Monitoring and Assessment, 194(7), 522.

[25] ESA (European Space Agency). (2024). Copernicus Open Access Hub — Sentinel-2 Mission. https://scihub.copernicus.eu

[26] Raissi, M., Perdikaris, P., & Karniadakis, G.E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. Journal of Computational Physics, 378, 686–707. https://doi.org/10.1016/j.jcp.2018.10.045

[27] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep

learning. Nature, 521(7553), 436–444.
[28] Gao, B.-C. (1996). NDWI — A normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sensing of Environment, 58(3), 257–266. https://doi.org/10.1016/S0034-4257(96)00067-3
[29] Cuomo, S., Di Cola, V.S., Giampaolo, F., Rozza, G., Raissi, M., & Piccialli, F. (2022). Scientific machine learning through physics-informed neural networks: Where we are and what's next. Journal of Scientific Computing, 92(3), 88. https://doi.org/10.1007/s10915-022-01939-z
[30] Rodell, M., Velicogna, I., & Famiglietti, J.S. (2009). Satellite-based estimates of groundwater depletion in India. Nature, 460(7258), 999–1002.
[31] Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., & Bargellini,
[32] P. (2012). Sentinel-2: ESA's optical high-resolution mission for GMES operational services. Remote Sensing of Environment, 120, 25–36.
[33] Cai, S., Mao, Z., Wang, Z., Yin, M., & Karniadakis, G.E. (2021). Physics-informed neural networks (PINNs) for fluid mechanics: A review. Acta Mechanica Sinica, 37(12), 1727–1738.
[34] McFeeters, S.K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. International Journal of Remote Sensing, 17(7), 1425–1432.
[35] Xu, H. (2006). Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. International Journal of Remote Sensing, 27(14), 3025–3033.

Appendix A: Python Pipeline Code

A.1 India-WRIS API Query

```
import requests import pandas as
pd

session = requests.Session()
# POST to /Dataset/Ground Water Level endpoint payload = {
"stateCode": "27",                    # Maharashtra
"districtCode": "2733",   # Yavatmal "fromDate": "01-
01-2018",
"toDate": "31-12-2026",
"dataType": "MONTHLY"
}
response = session.post( "https://indiawris.gov.in/api/Dataset/Ground Water Level"
    json=payload,
    headers={"Content-Type": "application/json"}
)
 gw_df = pd.DataFrame(response.json()["data"])
 gw_df["date"] = pd.to_datetime(gw_df["date"], format="%d-%m-%Y %H:%M:%S")
 gw_df.to_csv("yavatmal_groundwater_2018_2026.csv", index=False)
```

A.2 Sentinel-2 NDVI / NDWI Extraction (SentinelHub API)

```
rom sentinelhub import (SHConfig, BBox, CRS, SentinelHubRequest, DataCollection,
MimeType, bbox_to_dimensions)
import numpy as np

config = SHConfig()
config.sh_client_id          = "YOUR_CLIENT_ID"
config.sh_client_secret = "YOUR_SECRET"

# Yavatmal bounding box
bbox = BBox(bbox=[77.5, 19.8, 78.5, 20.5], crs=CRS.WGS84)
size = bbox_to_dimensions(bbox, resolution=10)

evalscript = """
//VERSION=3
function setup() {
return {input: ["B04","B08","B8A","B11"], output: {bands: 2}};
}
function evaluatePixel(sample) {
let ndvi = (sample.B08 - sample.B04) / (sample.B08 + sample.B04 + 1e-8); let ndwi = (sample.B8A -
sample.B11) / (sample.B8A + sample.B11 + 1e-8); return [ndvi, ndwi];
} """

request = SentinelHubRequest( evalscript=evalscript,
input_data=[SentinelHubRequest.input_data(
```

```
data_collection=DataCollection.SENTINEL2_L2A, time_interval=("2022-08-
01", "2022-08-30"), mosaicking_order="leastCC"
)],
responses=[SentinelHubRequest.output_response("default", MimeType.TIFF)], bbox=bbox, size=size, config=config
)
data = request.get_data()[0] ndvi = data[:, :,
0]
ndwi = data[:, :, 1]
```

A.3 PINN Loss Function (PyTorch)

```
Import torch
import torch.nn as nn

class DroughtPINN(nn.Module): def __init__(self,
input_dim=8):
super().__init__()   self.net = nn.Sequential(
nn.Linear(input_dim, 128), nn.Tanh(),
nn.Linear(128, 256),                nn.Tanh(),
nn.Linear(256, 128),                nn.Tanh(),
nn.Linear(128, 4)  # [theta_t, ET, R, I]
)

def forward(self, x): return self.net(x)

def total_loss(outputs, inputs, rainfall, obs_ndwi, lambda_phy=0.4,
lambda_con=0.1):
theta_t, ET, R, I = (outputs[:,0], outputs[:,1],
outputs[:,2], outputs[:,3]) theta_prev = inputs[:, 0]

# Data fidelity loss
L_data = nn.functional.mse_loss(theta_t, obs_ndwi)

# Physics residual loss (soil water balance)
residual = (theta_t - theta_prev) - (rainfall - ET - R - I) L_physics = torch.mean(residual
** 2)

# Non-negativity constraint L_constraint = torch.mean(
torch.relu(-ET) + torch.relu(-R) + torch.relu(-I)
)

return L_data + lambda_phy * L_physics + lambda_con * L_constraint
```