International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 13 Issue: 11 | Nov 2026 www.irjet.net p-ISSN: 2395-0072

Groundwater Stress Assessment and Spatial Risk Mapping Across India

Mrs. Chandana K. R.[1], Samyak[2], Sanskar Pandey[3], Vaibhav Singh[4], Yashas Allappanavar[5]

[1]*Assistant Professor, Dept. of Computer Science and Engineering, Sir M. Visvesvaraya Institute of Technology, Bengaluru, India*

[2345]*Dept. of Computer Science and Engineering, Sir M. Visvesvaraya Institute of Technology, Bengaluru, India*

Abstract - *India publishes groundwater assessments as backward-looking tables ranked by how much water is pumped, with no spatial structure. We rebuilt a national block-level dataset of 34,114 records (9,605 assessment units, 38 states and union territories) from five assessment cycles and trained five classifiers and a stacking ensemble to predict a five-level stress class. Every model was tested twice: under a random split, and under a spatial split that keeps each district entirely in training or in test data. With block- and district-level predictors, gradient boosting scored best on the spatial split (macro F1 0.498, accuracy 0.748, ROC-AUC 0.821, against 0.165 for an always-Low baseline). The random split overstated macro F1 by 0.099 on average. Adding each unit's class from its previous cycle raised the spatial macro F1 to 0.721 (accuracy 0.863, ROC-AUC 0.941); that variant partly becomes forecasting, so it is reported separately. K-Means gave three risk zones with 33.6%, 10.2% and 2.3% severe units. DBSCAN isolated one hotspot holding 593 of 929 severe units (63.8%), consistent with the 6 cm/yr north-western depletion reported from satellite gravity data. Of the units traceable across cycles (66.6%), 85.9% did not change, and the national fall in Critical share (16.0% to 10.8%) is largely a change in the assessed set.*

Key Words: Density-based clustering, gradient boosting, groundwater stress, spatial cross-validation, risk mapping.

1. INTRODUCTION

Groundwater supplies most of India's irrigation and domestic water. The Central Ground Water Board (CGWB) tracks it through the Ground Water Resource Assessment (GWRA), and each cycle places every block, mandal or taluk in one of five categories: Safe, Semi-Critical, Critical, Over-Exploited or Saline. The category comes mainly from the stage of extraction, meaning annual extraction as a percentage of the annually extractable resource [1], [2]. The reports look backward, lean on a single quantity metric and carry no geometry. They cannot show where stress clusters or how it changes over time.

Prior work is fragmented. Machine-learning studies cover small single-site inventories [3] or aspatial water-quality indices [4], [5]. Density clustering has been used only to screen wells [6]. The closest national study pairs GRACE storage trends with a district-level integrated groundwater footprint (iGF), but it fits no learned model and runs no out-of-sample validation [7]. Geographic data adds a trap of its own. When predictors are district-level and labels are block-level, a random split puts blocks of the same district on both sides, and part of the score then measures memorisation.

This paper makes five contributions: (i) a reproducible pipeline that turns the PDF reports of five GWRA cycles into one dataset; (ii) a measurement of how much optimism random validation adds compared with a district-grouped spatial split; (iii) an explainable five-class classifier with a prior-cycle extension that is reported separately; (iv) joint spatial (K-Means, DBSCAN) and temporal analysis of national stress, cross-checked against the iGF study [7]; and (v) a deployed layer for risk scoring, simulation and monitoring. The system overview is shown in Fig. 1.

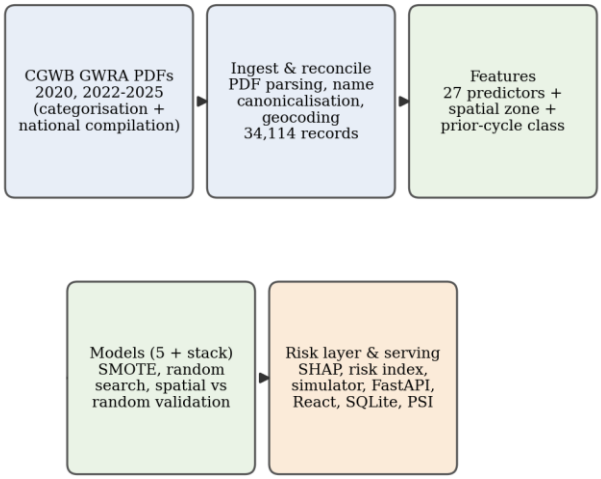

Fig -1: System overview of the CGWB data integration, feature construction, modelling and risk assessment pipeline.

2. RELATED WORK

Sachdeva and Kumar compared eleven classifiers for groundwater potential in Karbi Anglong. Random Forest came out best, on 42 labelled points and a single split [3]. Lawal et al. handled imbalance in binary water-quality classification with resampling and a voting ensemble on 301 samples [4]. Yusri et al. found XGBoost clearly better

than SVM on a four-class task [5]. Gan et al. ran DBSCAN on points-of-interest data to screen unregistered wells, with fixed neighbourhood parameters and no supervised component [6].

At national scale, Pranoy et al. combined GRACE-GLDAS storage anomalies (2002–2022) with CGWB abstraction, recharge and 15 quality parameters to compute a district iGF [7]. Their index extends the groundwater footprint of Gleeson and Wada [8] and Kourgialas et al. [9] with contamination factors. They report a depletion of 6 cm/yr in the north-western alluvial region, a mean GF of 0.675 that rises to a mean iGF of 0.849 once quality is included, 190 districts with iGF/A > 1, and the highest iGF in Jodhpur (9.07), Surendranagar (9.06), Jhajjar (8.68) and Kachchh (8.65). Table I positions the present work. It adds learned models, block-level resolution, five cycles and spatially constrained validation to this line of research.

Table -1: Comparison with closest prior studies.

Study	Data	Target / method	Validation
Sachdeva [3]	42 points, one district	Binary potential; 11 classifiers	Single split
Lawal [4]	301 river samples	Binary WQI; voting ensemble	Random CV
Yusri [5]	1,893 samples	4-class WQI; SVM, XGBoost	70/30 split
Pranoy [7]	730 districts; GRACE + CGWB 2022	iGF index; no learned model	None (descriptive)
This work	34,114 records, 5 cycles	5-class stress; 5 models + stacking; K-Means, DBSCAN	Spatial group + random

WQI = water quality index; CV = cross-validation.

3. METHODOLOGY

The operational workflow of the proposed groundwater stress assessment and spatial risk mapping system is shown in Fig. 2. It proceeds through: (1) CGWB assessment ingestion and record matching; (2) data validation, preprocessing and feature construction; (3) district-grouped model training and validation; (4) five-class groundwater stress assessment with SHAP explanations; and (5) spatial risk-zone identification, hotspot detection and mapping. Decision branches in the flowchart return unmatched records for reconciliation and unsuccessful model configurations for further tuning.

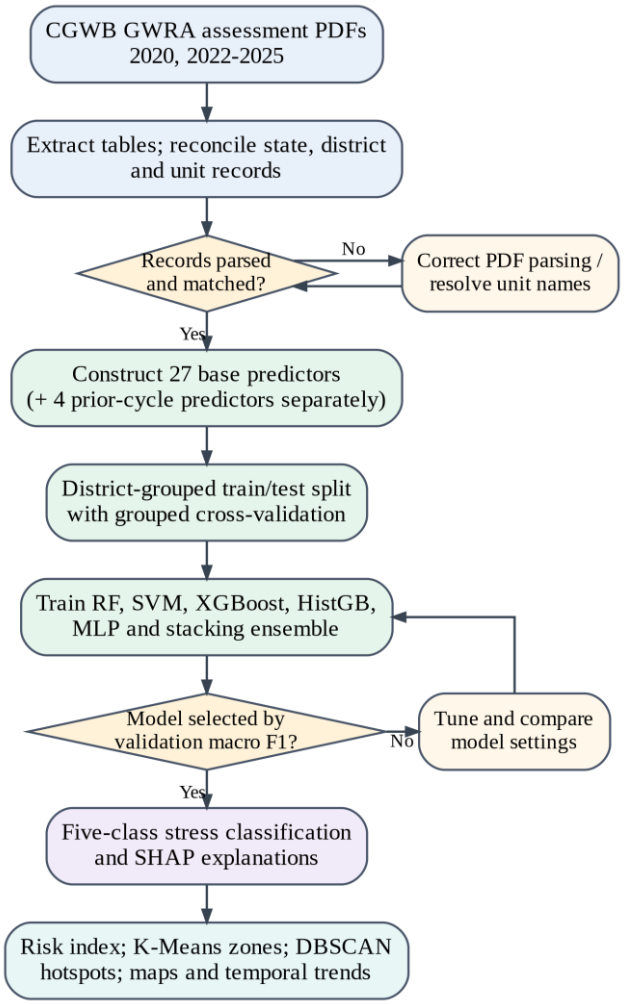

Fig -2: Proposed flowchart showing groundwater data processing, Yes/No data-validation and model-selection decisions, feedback loops, explainable classification, and spatial risk mapping.

3.1 Data and extraction

Block-wise categorisation reports (2020, 2022–2025) supply the label for every unit. The national compilations supply district-level monsoon and non-monsoon recharge (rainfall and other sources), natural discharge, extractable resource, extraction by irrigation, industry and domestic use, future availability and stage of extraction. Block-level fluoride, arsenic and salinity flags come from the quality annexure of the compilations.

The source files had defects, and each needed handling during extraction: reversed text runs in the 2020 compilation, upper-case state names in 2025, 48 spellings for about 36 states and union territories, and a misspelt state header (“CHHATISGARH”) in the 2024 compilation. That last one caused Chhattisgarh’s district rows to be dropped until header detection was routed through the

alias table. State names are canonicalised and districts reconciled by alias table plus within-state fuzzy matching.

Unit centroids came from Nominatim geocoding: 76.7% block-precise, 21.1% district-centroid fallback and 2.2% unresolved. Re-running the pipeline reproduces the target label and quality flags of the master table on all 34,114 rows, and every quantity value wherever both exist.

3.2 Target and features

The five CGWB categories map one-to-one to Low, Moderate, High, Critical and Saline. Saline stays separate because it is a quality condition. Its median stage of extraction (34.2%) sits near Low (43.8%) and far from Critical (124.1%).

There are 27 base predictors: quantity features (stage of extraction, extraction-to-recharge and future-availability ratios); the three contamination flags with an availability indicator; proxies for recharge composition (rainfall share) and extraction composition (irrigation share); raw recharge and extraction volumes; a data-completeness score; the cycle year; and a K-Means spatial-zone indicator fitted on coordinates only. Missing quantity data are kept and flagged, not dropped. The prior-cycle feature set adds four predictors from the same unit's most recent strictly-earlier cycle: class severity, a Saline indicator, stage of extraction and an availability indicator. The task is then partly forecasting, so we report it separately.

3.3 Models, imbalance and validation

Random Forest [10], an RBF SVM, XGBoost [11], histogram gradient boosting (HistGB) and a multilayer perceptron (MLP) are tuned by randomised search, using grouped three-fold cross-validation on the training portion and maximising macro F1. A stacking ensemble combines their probabilities through a class-balanced logistic meta-learner. Imputation and scaling are fitted inside each fold. SMOTE [12] is applied inside training folds for the four tree and neural models, and the SVM uses balanced class weights.

The random protocol is a stratified 80/20 row split (27,291 / 6,823). The spatial protocol uses stratified group splitting on state|district, so every district falls wholly in training (27,286) or test (6,828) data. Metrics are accuracy, macro precision, macro recall, macro F1, one-vs-rest macro ROC-AUC and per-class recall. The test set is 70% Low, which makes macro F1 the primary metric. Table II lists the search grids (eight random draws per model, four for the SVM) and the settings selected under the spatial protocol.

Table -2: Hyperparameter search space and selected optimum (spatial split).

Model	Search space	Selected (base)	Selected (+prior)
Random forest	trees 200/300/500; depth 12/20/30/none; leaf 1/2/4; features sqrt/0.5	500, 12, 2, sqrt	200, 12, 1, sqrt
SVM (RBF)	C 1/5/10/30; gamma scale/0.01/0.05	C=1, scale	C=1, scale
XGBoost	trees 200/300/500; depth 4/6/8; lr 0.03/0.05/0.1; subsample 0.8/1; colsample 0.7/1	500, 4, 0.03, 0.8, 0.7	500, 4, 0.05, 1.0, 1.0
HistGB	iter 200/300/500; depth none/6/10; lr 0.03/0.05/0.1; L2 0/1	300, 10, 0.03, 1.0	500, 6, 0.03, 0.0
MLP	layers (64,32)/(128,64)/(128,64,32); alpha 1e-4/1e-3/1e-2; lr 0.001/0.005	(128,64), 1e-3, 0.001	(128,64), 1e-4, 0.005

3.4 Explanation, risk index and serving

TreeSHAP [13] explains the served XGBoost model class by class. A 0–100 risk index blends the model's class probabilities (weight 0.40) with quantity (0.25), quality (0.15), environmental (0.10) and anthropogenic (0.10) sub-scores. Components without data are excluded and the weights renormalised; they are never scored as clean. Low-confidence scores are pulled toward 50. The weights are illustrative, not calibrated. Stored scores are out-of-fold (five grouped folds).

A scenario simulator rescales extraction and recharge levers and recomputes stage of extraction from the source accounting identities. Models are versioned in SQLite with promote and rollback. A FastAPI service and a React dashboard expose predictions, maps, trends and drift, and the Population Stability Index (PSI) tracks input, label and predicted-class shift between cycles. Automated tests cover feature derivation, split leakage guarantees, simulator accounting identities, the risk index, drift statistics and version promotion.

3.5 Spatial and temporal analysis

K-Means (k = 2–8, chosen by silhouette) partitions the geocoded units after equirectangular projection. DBSCAN [14] with haversine distance runs on High and Critical units only; the radius comes from the elbow of the k-

distance curve, with min_samples = 5. For the temporal analysis, the four quantity classes are ordered by severity and a slope is fitted for each unit seen in at least three cycles. Saline observations are excluded. Mann–Kendall p-values are computed [15], but with at most five points the smallest attained is 0.023, so we read direction from the slope.

4. RESULTS

4.1 Dataset and stress distribution

The dataset has 34,114 records over five cycles, with 6,553–7,089 units per cycle. Class totals are Low 23,895, Moderate 4,110, High 1,136, Critical 4,337 and Saline 636. Quantity attributes exist for 96.5% of records but quality flags for only 8.8%, so the quality dimension rests on a sparse signal. In 2025, 931 of 6,762 units (13.8%) are High or Critical (Fig. 5a). Punjab (79.1%), Rajasthan (78.1%), Haryana (67.8%) and Delhi (61.8%) lead. Median stage of extraction rises monotonically from 43.8% (Low) to 71.1%, 85.2% and 124.1% (Critical).

4.2 Classification performance

Table III lists all results. Under the spatial split with base predictors, HistGB achieves the highest macro F1 (0.498; accuracy 0.748, macro precision 0.542, macro recall 0.506, ROC-AUC 0.821), and XGBoost comes within 0.001 (0.497). Every model gains from the random split. Macro F1 under random validation spans 0.543–0.579, against 0.429–0.498 under spatial validation. The mean optimism is 0.099 (HistGB 0.070; stacking 0.119, the largest). That gap is a direct estimate of the leakage introduced by sharing districts across the split. Stacking does not beat the best single model in either configuration.

The district-majority reference labels every unit with its district's modal class, so it uses test labels. It reaches accuracy 0.825 and macro F1 0.594, an upper bound for district-level information alone. The base-predictor models fall short of it, which fits with district-level predictors being unable to separate blocks within a district.

With the prior-cycle predictor, HistGB reaches macro F1 0.721 (accuracy 0.863, ROC-AUC 0.941), followed by XGBoost 0.710, Random Forest 0.690, stacking 0.668, SVM 0.666 and MLP 0.631. Random-split optimism falls to 0.042 on average (HistGB 0.009). Most of the earlier gap was therefore spatial leakage, not a shortage of predictors. Fig. 4 compares the three validation designs.

Table -3: Classification results (test set; P and R are macro; AUC is one-vs-rest macro).

Model	Acc	P	R	F1	AUC
Random forest	0.707	0.565	0.492	0.472	0.817
SVM (RBF)	0.666	0.423	0.499	0.437	0.871
XGBoost	0.753	0.525	0.497	0.497	0.827
HistGB	0.748	0.542	0.506	0.498	0.821
MLP	0.693	0.423	0.461	0.429	0.805
Stacking	0.697	0.438	0.483	0.443	0.796
Always-Low	0.700	0.140	0.200	0.165	—
Dist-maj.†	0.825	0.681	0.561	0.594	—

Table -3 (continued). Spatial results with prior-cycle features and random-split macro F1.

Model	Acc	P	R	F1	AUC	Rand. base	Rand. prior
Random forest	0.832	0.717	0.708	0.690	0.927	0.569	0.731
SVM (RBF)	0.832	0.629	0.720	0.666	0.943	0.547	0.713
XGBoost	0.860	0.732	0.715	0.710	0.944	0.579	0.730
HistGB	0.863	0.746	0.727	0.721	0.941	0.568	0.730
MLP	0.810	0.600	0.678	0.631	0.929	0.543	0.713
Stacking	0.847	0.642	0.713	0.668	0.913	0.562	0.722
Always-Low	0.700	0.140	0.200	0.165	—	—	—
Dist-maj.†	0.825	0.681	0.561	0.594	—	—	—

† District-majority uses test labels and is not a deployable model. Other methods use no test labels.

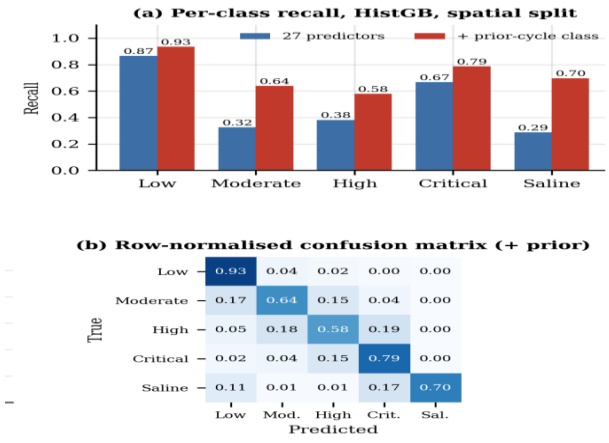

Fig -3: Per-class recall and row-normalised confusion matrix for HistGB, with and without prior-cycle predictors.

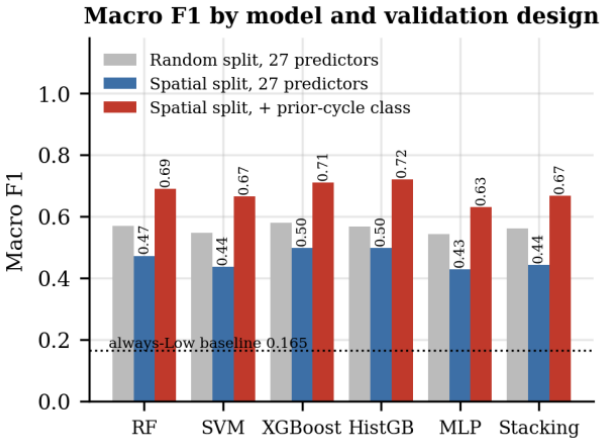

Fig -4: Macro F1 by validation design, with the always-Low baseline indicated.

4.3 Explainability

Table -4: Effect of validation design and prior-cycle predictor (HistGB).

Configuration	Acc	Macro F1
Always-Low	0.700	0.165
Base predictors, random split (HistGB)	0.747	0.568
Base predictors, spatial split (HistGB)	0.748	0.498
+ Prior-cycle class, spatial split (HistGB)	0.863	0.721
+ Prior-cycle class, random split (HistGB)	0.864	0.730

Global TreeSHAP importance for the served XGBoost model (Fig. 6b) is led by prior-cycle severity, the extraction-to-recharge ratio, the prior-cycle availability indicator, the cycle year and the prior Saline indicator. Future availability and stage of extraction follow. The year and availability indicators mainly drive the Saline class, whose counts vary by cycle. That reflects a data artefact in how Saline units are reported, not hydrology, and should not be read physically. The extraction-to-recharge ratio is the strongest contemporaneous driver for Critical units, matching the CGWB definition of over-exploitation. Out-of-fold risk indices rise monotonically with the true class (mean 22.5 Low, 40.2 Moderate, 54.1 High, 75.1 Critical; Saline 42.1), with mean prediction confidence 0.77.

Recall is uneven (Fig. 3a). With base predictors, HistGB recovers 0.87 of Low and 0.67 of Critical units but only 0.38 of High units. High is the narrowest band on the extraction continuum, and confusion with neighbouring classes dominates (Fig. 3b). The prior-cycle predictor lifts High recall to 0.58 and Critical to 0.79, yet High stays the weakest class. High units are where intervention is most useful, so this is the system's main weakness. Saline recall is 0.29 with base predictors and 0.70 with prior-cycle information, consistent with Saline being a quality condition that sparse contamination flags capture poorly.

4.4 Spatial risk zones and hotspots

Silhouette selection gives k = 3 (silhouette 0.50) in 2025 and in every other cycle. The zones are centred on the Indo-Gangetic east (Bihar, West Bengal), the Deccan (Andhra Pradesh, Telangana, Maharashtra) and the north-west (Rajasthan, Madhya Pradesh, Uttar Pradesh), and they contain 2.3%, 10.2% and 33.6% severe units respectively.

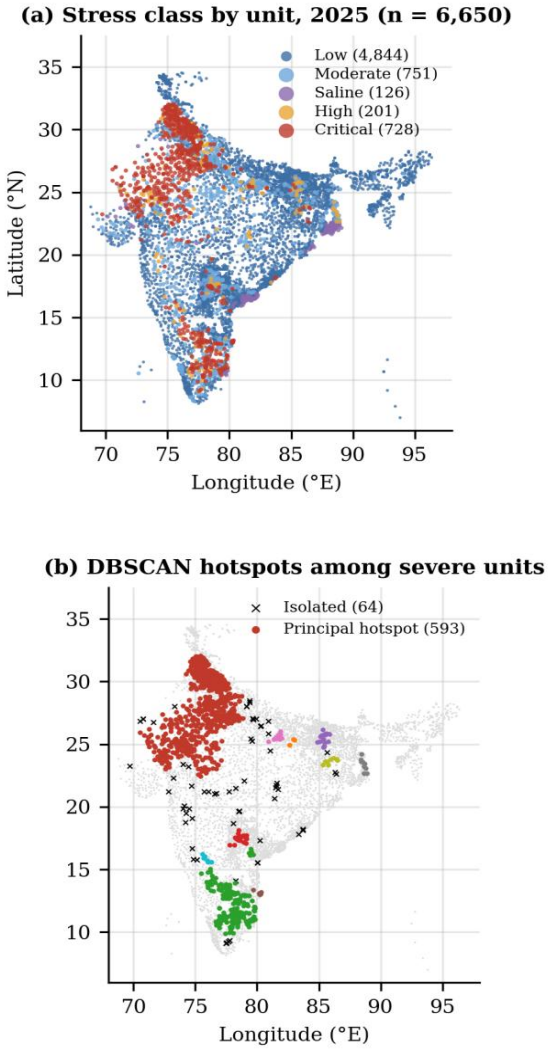

Fig -5: Stress-class distribution of geocoded 2025 units and DBSCAN hotspot clusters among severe units.

DBSCAN finds 11 hotspots and 64 isolated units. The principal hotspot holds 593 of 929 severe units (63.8%):

Rajasthan 229, Punjab 121, Haryana 97, Uttar Pradesh 71, Madhya Pradesh 29. A second large cluster covers the southern peninsula (Fig. 5b). We supplied only coordinates and labels, so recovering the known north-western over-extraction belt supports the validity of the geocoding.

4.5 Temporal behaviour

Of 9,605 units, 6,393 (66.6%) appear in at least three cycles. The rest cannot be traced because administrative reorganisation changes unit names. Among traceable units, 85.9% show no direction change, 8.6% improve and 5.5% worsen, so stress is a persistent condition.

The national Critical share falls from 16.0% (2020) to 10.8% (2025). Restricted to the 22 states whose assessed-unit counts stay within 10% of their own mean in every cycle, it moves only from 12.6% to 11.4% (Fig. 6a). Most of the apparent improvement reflects a change in which units were assessed. PSI between 2020 and 2025 is stable for the class distribution (0.050) and for 21 of 26 predictors. Industrial extraction (0.439) and the salinity flag (0.359) drift critically, three recharge and resource features show warnings, and the served model's predicted-class distribution shows a borderline warning (0.246).

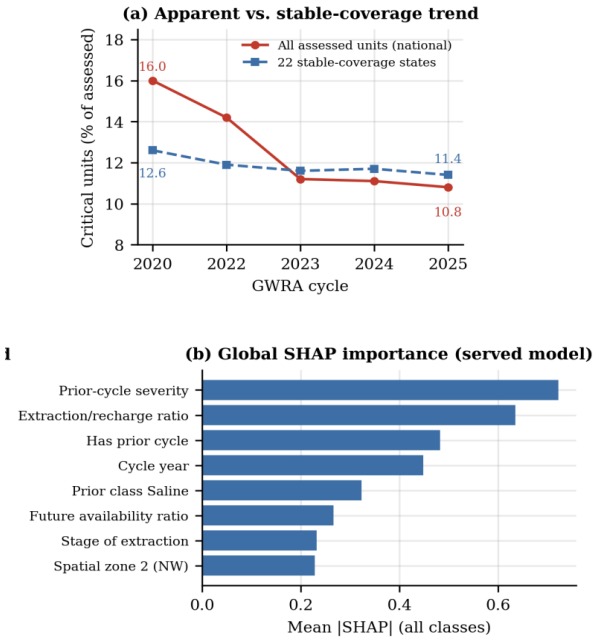

Fig -6: Critical-share trends for assessed and stable-coverage states, and global SHAP feature importance.

4.6 Consistency with the iGF assessment

Table V compares the ten most stressed districts of the iGF study [7] with the CGWB block data used here. Only Jodhpur, Jhunjhunu and Tonk are majority-severe on quantity grounds, and Jodhpur and Jhunjhunu have median stage of extraction above 200%. The other districts, such as Surendranagar, Jhajjar and Muktsar, sit below 100% stage of extraction and so carry no quantity stress in the CGWB categories. Their high iGF most plausibly comes from the contamination factors, which our sparse quality flags (8.8% coverage) cannot reproduce.

Table -5: iGF top districts [7] versus CGWB stress, 2022.

District (state)	iGF [7]	H+C units, 2022 (%)	Median stage (%)
Jodhpur (RJ)	9.07	94	258
Surendranagar (GJ)	9.06	0	49
Jhajjar (HR)	8.68	0	50
Kachchh (GJ)	8.65	30	60
Tonk (RJ)	8.02	67	103
Jhunjhunu (RJ)	7.59	100	215
Tuticorin (TN)	7.36	0‡	28‡
Ahmedabad (GJ)	7.09	20	86
Muktsar (PB)	6.72	0	25
Tiruchirappalli (TN)	6.25	42	78

Source: iGF values from [7]; CGWB block data. H+C = High or Critical. ‡ Thoothukudi district, 2025 cycle (no 2022 records).

The two analyses complement each other. The iGF study ranks districts by combined quantity and quality burden using 2022 district-level data; this work classifies blocks by CGWB quantity category. At regional level they agree: the north-western belt with 6 cm/yr GRACE depletion in [7] contains 447 of the 593 units (75.4%) of our principal hotspot. Units (district versus block) and targets differ, which rules out a numerical score comparison, and we claim none.

5. CONCLUSION

A national classifier built from 34,114 block-year records across five GWRA cycles reaches macro F1 0.498 and ROC-AUC 0.821 under district-grouped validation, using district- and block-level predictors only. Random validation would have reported 0.099 more on average. Adding each unit's previous class lifts macro F1 to 0.721 and ROC-AUC to 0.941, at the cost of turning part of the task into forecasting. High-class recall (0.38; 0.58 with prior-cycle information) is the main limitation, and it traces back to predicting block labels from district-level attributes.

Stress is concentrated, with 63.8% of severe units in one north-western hotspot, and it persists, with 85.9% of traceable units unchanged. The apparent national improvement is mostly a composition effect. The iGF comparison [7] shows that quantity-based categories miss contamination-driven stress in several districts. Well-level quality data with a stricter matching strategy, together with external rainfall and land-use layers, are the next steps. Until quality data reach block level, categories built on quantity alone will keep missing districts such as Surendranagar and Jhajjar.

REFERENCES

1) Central Ground Water Board, National Compilation on Dynamic Ground Water Resources of India, Ministry of Jal Shakti, 2020–2025.
2) Central Ground Water Board, Block wise Categorization as per Ground Water Resource Assessment, Ministry of Jal Shakti, 2020–2025.
3) S. Sachdeva and B. Kumar, "Groundwater potential mapping using machine learning models for Northeastern Karbi Anglong district, Assam, India," in Proc. ICACCCN, 2020, pp. 116–121.
4) Z. K. Lawal et al., "Optimized ensemble methods for classifying imbalanced water quality index data," IEEE Access, vol. 12, pp. 178536–178551, 2024.
5) H. I. H. Yusri et al., "Water quality classification using SVM and XGBoost method," in Proc. IEEE ICSGRC, 2022, pp. 231–236.
6) X. Gan et al., "Machine well screening method based on POI data and DBSCAN clustering algorithm," IEEE Geosci. Remote Sens. Lett., vol. 22, Art. 7504305, 2025.
7) S. Pranoy, S. Srivastava, and A. Singh, "Groundwater stress in India: Integrated analysis of quantity and quality," in Proc. IEEE InGARSS, 2024, doi: 10.1109/InGARSS61818.2024.10983998.
8) T. Gleeson and Y. Wada, "Assessing regional groundwater stress for nations using multiple data sources with the groundwater footprint," Environ. Res. Lett., vol. 8, no. 4, p. 044010, 2013.
9) N. N. Kourgialas et al., "Groundwater footprint methodology as policy tool for balancing water needs in water scarce islands The case of Crete, Greece," Sci. Total Environ., vol. 615, pp. 381–389, 2018.
10) L. Breiman, "Random forests," Mach. Learn., vol. 45, no. 1, pp. 5–32, 2001.
11) T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. ACM SIGKDD, 2016, pp. 785–794.
12) N. V. Chawla et al., "SMOTE: Synthetic minority over-sampling technique," J. Artif. Intell. Res., vol. 16, pp. 321–357, 2002.
13) S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Proc. NeurIPS, 2017, pp. 4765–4774.
14) M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in Proc. KDD, 1996, pp. 226–231.
15) R. Agarwal and P. K. Garg, "Statistical assessment of groundwater resources and long term trend using geospatial techniques," in Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS), 2016, doi: 10.1109/IGARSS.2016.7729464.