To test the stability of the selected branches, a repeated spatial split stability analysis was performed. Ten spatial splits with random seeds of 11, 21, 42, 77, 97, 123, 202, 404, 707, and 999 were used. This additional analysis was implemented as a simplified retrainable branch-fusion experiment consistent with the original GW-RSHF-ML structure. In each split, full-feature branches and modality-specific branches were trained, after which the convex weights were reoptimized on the validation split. The results are shown in Table S1.

Table S1. Stability of GW-RSHF-ML branches under 10 repeated spatial splits.

Branch	Active splits	Mean weight	Median weight	Max weight	Number of predictors	Active frequency
full_hist_gradient_boosting_branch	10	0.5114	0.5221	0.7448	69	1.00
sentinel1_branch	7	0.1588	0.1289	0.4760	13	0.70
full_extra_trees_branch	4	0.1408	0.0000	0.5045	69	0.40
water_balance_branch	3	0.0742	0.0000	0.2660	11	0.30
climate_branch	3	0.0533	0.0083	0.3112	17	0.30
topographic_branch	2	0.0435	0.0000	0.1980	13	0.20
land_cover_branch	1	0.0136	0.0000	0.1207	4	0.10
soil_branch	0	0.0032	0.0000	0.0321	3	0.00
hydro_topographic_branch	0	0.0012	0.0000	0.0124	8	0.00

As shown in the table S1, the exact combination of active branches is not completely invariant with respect to spatial splits. The most stable was the full-feature HistGradientBoosting branch, active in all 10 splits, indicating a robust role for the common nonlinear signal generated by all selected features. The Sentinel-1 branch was active in 7 of 10 splits, confirming the robust informativeness of the SAR predictors. The full-feature Extra Trees branch was active in 4 of 10 splits, and the topographic branch in 2 of 10. Therefore, the final initial configuration with full-feature Extra Trees, topographic, and Sentinel-1 branches should be interpreted as a validation-selected configuration rather than as the only universally stable structure. Repeated spatial splits confirm that the most robust signal sources are the full-feature nonlinear branches and Sentinel-1 SAR information.

The Figure S1 shows the variability of the fusion model performance across 10 spatial splits. The average locked-test PR-AUC was 0.8465 with a standard deviation of 0.0428, and the average locked-test ROC-AUC was 0.9345 with a standard deviation of 0.0216. The average locked-test F1 was 0.7719, balanced accuracy was 0.8556, MCC was 0.6967, and Brier score was 0.0836. These results indicate that the model retains its informative discriminatory power across repeated spatial splits, but the magnitude of the metrics and the composition of active branches remain sensitive to the spatial structure of the train/validation/test splits. Therefore, in the future, the results of GW-RSHF-ML should be interpreted as spatially validation-controlled and competitive, but not as evidence of the absolute stability of a single fixed combination of branches.

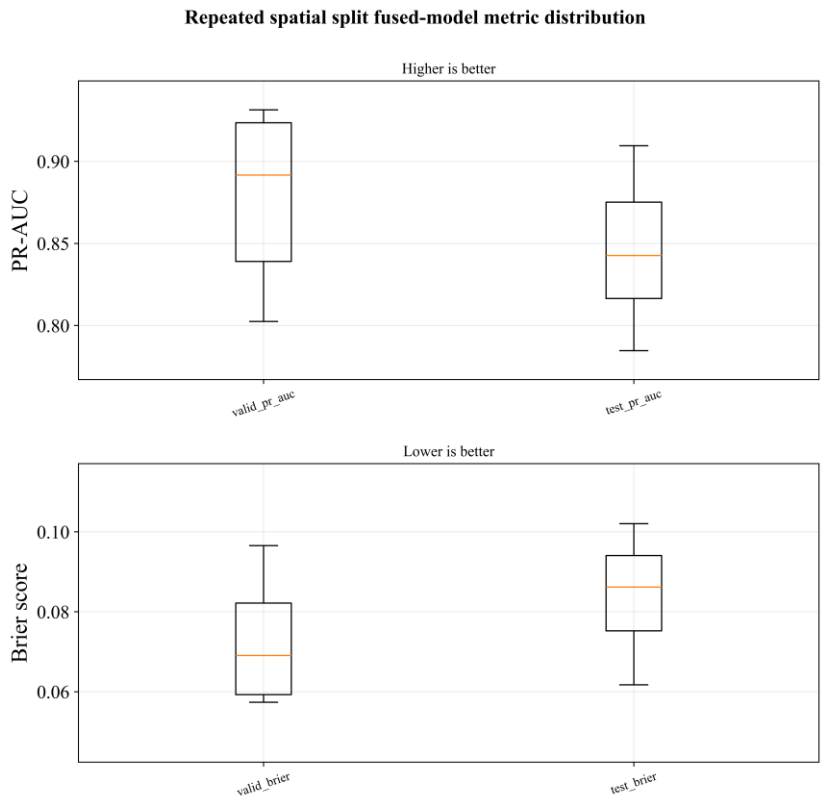


Figure S1. Distribution of PR-AUC and Brier score in repeated spatial split stability analysis.

The Figure S2 shows the final point-supported probabilistic groundwater-related suitability map. The map displays all 2,402 observations within Kazakhstan. Higher probability values are spatially concentrated primarily in the southern, southeastern, and eastern parts of the study area, as well as in isolated localized areas in the west and north. This structure is consistent with the model identifying zones where the combination of topography, climate, SAR, soil, and water-balance features is closer to the conditions of known springs and water wells. However, this map should not be interpreted as a map of confirmed groundwater presence.

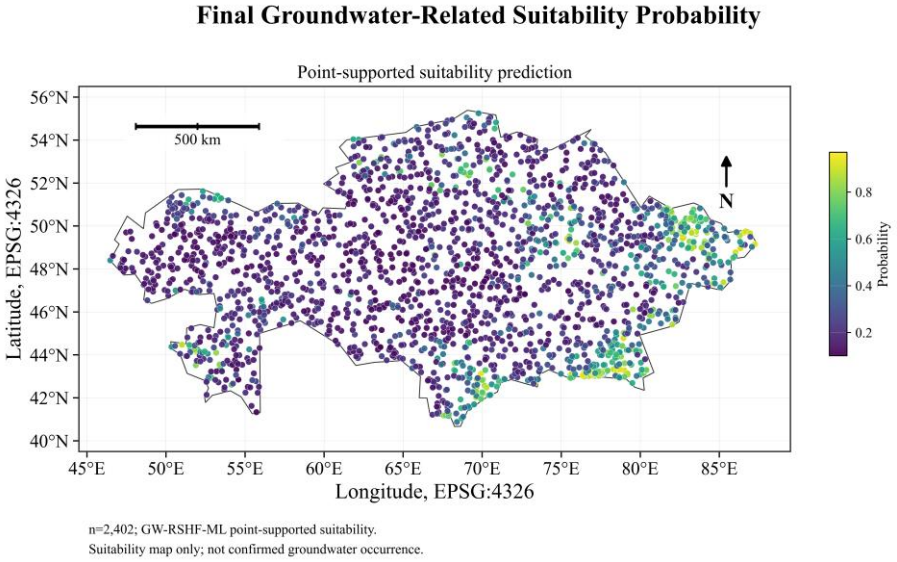


Figure S2. Final Groundwater-Related Suitability Probability.

The Figure S3 represents a classified map of groundwater-related suitability. The low suitability class corresponds to the lower third of the probability distribution, medium suitability to the middle third, and high suitability to the upper third. The classification is used only for visual ranking and preliminary selection of areas of interest. These classes are not operational hydrogeological thresholds and should not be used directly for drilling or water management decisions.

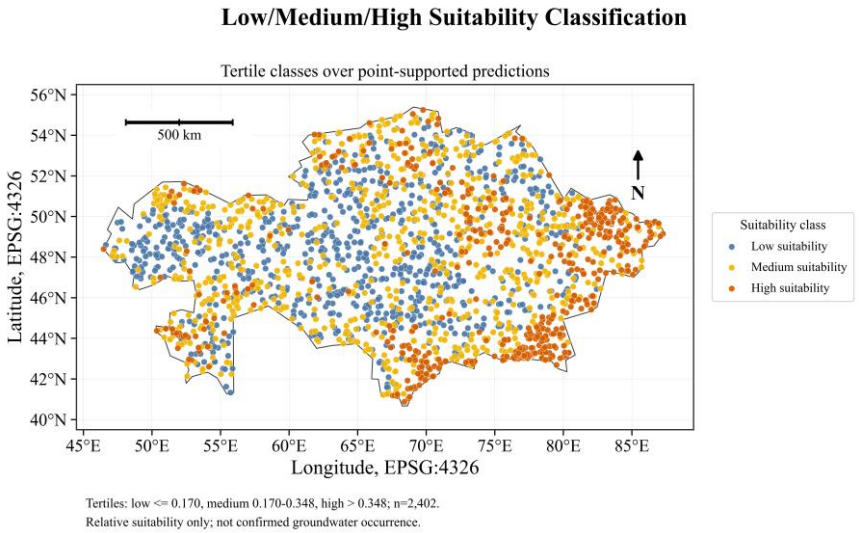


Figure S3. Low/Medium/High Suitability Classification.

The Figure S4 shows the spatial distribution of model uncertainty and model confidence. Higher uncertainty values correspond to points where the active GW-RSHF-ML branches yield less consistent probabilistic estimates. Higher confidence values indicate a more consistent model estimate. However, confidence is not hydrogeological certainty and does not confirm the presence of groundwater. It reflects only the internal consistency of the model with the given set of predictors. Taken together, the three map outputs demonstrate that the study's results are represented not only by numerical classification metrics but also by spatially interpretable maps: probabilistic suitability, classified suitability, and model uncertainty/confidence. All maps should be considered preliminary decision-support layers for prioritizing further hydrogeological surveys. Field verification, well data, hydrogeological maps, and geophysical measurements remain necessary before any decisions regarding drilling, monitoring, or groundwater resource management.

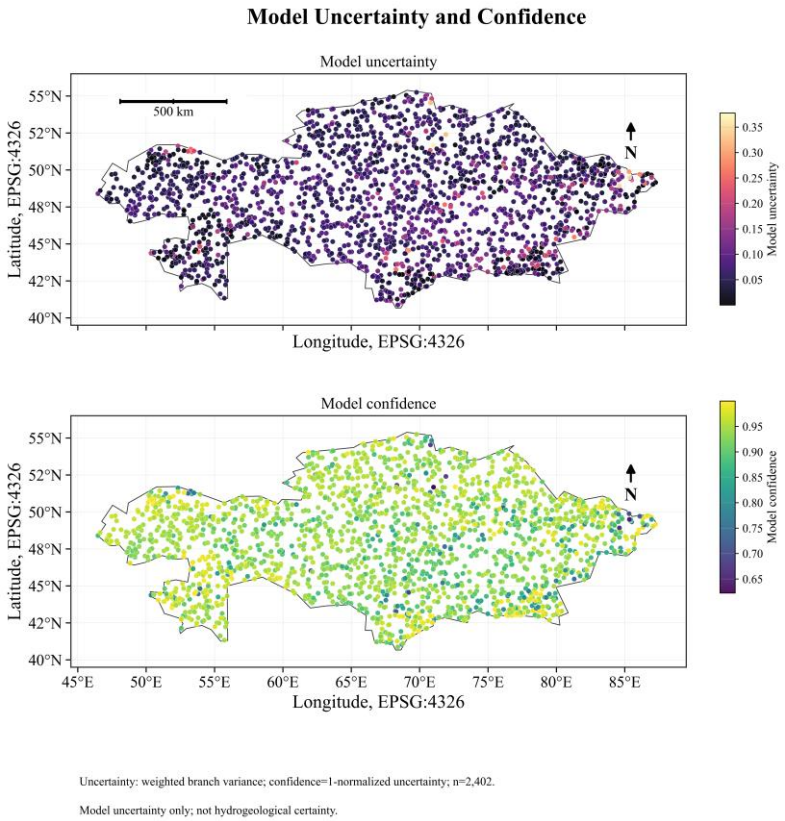


Figure S4. Model Uncertainty and Confidence.

Ablation analysis shows that the contribution of modality-specific branches should not be interpreted as an unambiguous improvement in predictive ranking (Table S2). The "Remove feature-group branches" variant received a validation PR-AUC of 0.8543, while the full GW-RSHF-ML received 0.8363. This means that in this spatial validation split, full-feature branches provided a stronger ranking of the positive class. However, the full model had better validation F1, balanced accuracy, and MCC, indicating a more balanced classification profile. Therefore, modality-specific branches in this architecture are considered primarily as a means of hydrological interpretability and diagnostic decomposition, rather than as a guaranteed source of PR-AUC enhancement. Their presence allows us to verify which physical feature groups retain their independent contribution after validation-selected fusion. In the final configuration, topographic and Sentinel-1 branches received nonzero weights, demonstrating their additional role in explaining the spatial structure of groundwater-related conditions. Nevertheless, the assertion about the predictive superiority of modality-specific branches was relaxed.

Table S2. Main results of the GW-RSHF-ML ablation analysis.

Variant	Valid PR-AUC	Valid F1	Valid BA	Test PR-AUC	Test F1
Remove feature-group branches	0.8543	0.8000	0.8629	0.9004	0.7788
Full GW-RSHF-ML	0.8363	0.8226	0.8847	0.9002	0.7983
Remove linear branch	0.8247	0.8145	0.8792	0.8808	0.7948
Remove nonlinear tree branch	0.8247	0.8145	0.8792	0.8808	0.7948

As can be seen from the Table S3, GW-RSHF-ML maintains high discriminatory performance: its PR-AUC is 0.9002, and ROC-AUC is 0.9535. However, calibration diagnostics show that the absolute probabilities of this model require careful interpretation. The ECE for GW-RSHF-ML is 0.1268, which is higher than that of HistGradientBoosting, ExtraTrees, and FT-Transformer Lite. Furthermore, the calibration slope = 2.1585 indicates heterogeneity of the calibration and a more pronounced discrepancy between predicted probabilities and observed positive fractions in individual probability bins. Therefore, the GW-RSHF-ML probabilities in this paper are interpreted primarily as relative groundwater-related suitability scores, rather than as direct probabilities of the actual presence of groundwater.

Table S3. Reliability metrics and calibration diagnostics on the locked-test split.

Model	**PR-AUC**	**ROC-AUC**	**F1**	**Balanced accuracy**	**MCC**	**Brier**	**ECE**	**MCE**	**Calibration slope**	**Calibration intercept**
GW-RSHF-ML	0.9002	0.9535	0.7983	0.8594	0.7336	0.0865	0.1268	0.1966	2.1585	-0.3702
HistGradientBoosting	0.8921	0.9535	0.8250	0.8830	0.7660	0.0687	0.0314	0.2423	0.8860	0.2131
Extra Trees	0.9002	0.9558	0.8170	0.8733	0.7574	0.0724	0.0590	0.1326	1.6318	0.0941
FT-Transformer Lite	0.8745	0.9421	0.7692	0.8413	0.6945	0.0838	0.0532	0.2973	0.8637	-0.6206