

Optimized TabPFN-gradient boosting multimodel ensemble strategy for predicting soil saturated hydraulic conductivity

LIU Yifei¹, NI Junjun¹, SADEGHI Hamed²

(1. School of Transportation, Southeast University, Nanjing 211189, China; 2. Department of Civil Engineering, Sharif University of Technology, Tehran 1458889694, Iran)

Abstract: Soil saturated hydraulic conductivity is a critical hydraulic parameter in geotechnical engineering and hydrology. Traditional experimental measurement methods are time-consuming and labor-intensive, making it difficult to meet the rapid application requirements of engineering practice. This study constructs prediction models based on easily measurable basic soil properties, using particle composition percentages and bulk density as input features to predict saturated hydraulic conductivity. The study proposes an optimized tabular prior-data fitted network (TabPFN) model that integrates an expert mechanism with a two-stage optimization strategy and compares its performance with traditional gradient boosting models and random forest (RF). A multimodel stratified weighted ensemble strategy is constructed to synthesize the advantages of all models. The results show that the expert mechanism and the two-stage optimization strategy significantly improve TabPFN performance. The optimized TabPFN model achieved a maximum prediction accuracy of $R^2 = 0.9568$ and root mean square error (RMSE) of 0.0011 cm/s , representing a 16.78% improvement in R^2 over the preoptimization model and outperforming all gradient boosting and RF models. The stratified weighted ensemble strategy effectively integrated the prediction advantages of each sub-model across different soil textures, achieving the highest prediction accuracy ($R^2 = 0.9617$, RMSE = 0.0010 cm/s) and effectively reducing prediction errors and heteroscedasticity. The proposed optimized ensemble model provides a new intelligent method for the rapid estimation of saturated hydraulic conductivity in engineering practice.

Key words: soil saturated hydraulic conductivity; machine learning; tabular prior-data fitted network (TabPFN); gra-

dient boosting; ensemble strategy

DOI: 10.3969/j.issn.1003-7985.2026.03.005

Saturated hydraulic conductivity (K_{sat}) is the ratio of the seepage velocity of water to the hydraulic gradient in saturated soil, with dimensions of length/time^[1]. K_{sat} is important in geotechnical engineering, hydrology, geology, agricultural science, and soil science^[2-3]. Accurate K_{sat} determination is crucial for water resource management, geotechnical stability assessment, groundwater recharge evaluation, and natural disaster prediction^[4-5]. In practical applications, K_{sat} can span several orders of magnitude across different soil types and exhibits high spatial variability^[6-7]. The magnitude of K_{sat} is primarily determined by the soil microstructure, including particle size distribution (PSD), bulk density, porosity, organic matter content, particle shape, and pore connectivity^[1-3,8-9]. The hydraulic conductivity of soil cannot be directly measured and must be determined through laboratory or in situ tests, a time-consuming process that makes it difficult to meet the rapid application requirements of engineering practice.

To overcome the difficulties of directly measuring hydraulic conductivity, pedotransfer functions (PTFs) are commonly used in soil mechanics and hydrology^[2,10-11]. PTFs estimate hydraulic conductivity using easily obtainable basic soil physical property parameters (PSD, bulk density, organic carbon content, porosity, water content, etc.). PTF models are mostly constructed using data-driven semiempirical formulas such as the classical Kozeny-Carman (KC) equation^[12]. The KC equation accurately predicts the permeability of sandy soils but performs poorly for soils containing clay particles. Moreover, existing semiempirical formulas generally struggle to account for the influence of adsorbed water films on the pore structure of clay particles, limiting their prediction accuracy for fine-grained soils. Furthermore, the empirical coefficients in these formulas have limited universality and are influenced by numerous factors such as soil type, particle shape, and degree of compaction, which may lead to inaccurate prediction results^[2,11].

Received 2025-11-17, Revised 2026-05-12.

Biographies: LIU Yifei (1997—), male, Ph. D. candidate; NI Junjun (corresponding author), male, doctor, professor, nijunjun@seu.edu.cn.

Foundation items: The National Natural Science Foundation of China (No. U2340227, 52308342), Fundamental Research Funds for the Central Universities (No. RF1028623071, 2242024k30066), Graduate Research and Practice Innovation Program of Jiangsu Province (No. SJCX25_0128).

Citation: LIU Yifei, NI Junjun, SADEGHI Hamed. Optimized TabPFN-gradient boosting multimodel ensemble strategy for predicting soil saturated hydraulic conductivity [J]. Journal of Southeast University (English Edition), 2026, 42(3): 318-328. DOI: 10.3969/j.issn.1003-7985.2026.03.005.

Machine learning techniques show considerable advantages in handling complex, high-dimensional nonlinear data, making them powerful tools in engineering^[13-16]. Combining machine learning methods with PTFs can effectively improve K_{sat} prediction accuracy^[10, 17-18]. Recently, transformer-based deep learning frameworks have developed rapidly in natural language processing, and their ability to process complex data structures has generated new ideas for tabular data prediction. TabPFN achieves approximate Bayesian inference through offline training on synthetic data, enabling it to make predictions without task-specific gradient updates and making it suitable for small- to medium-sized heterogeneous datasets. TabPFN therefore offers a promising technical path for fast, high-accuracy K_{sat} prediction^[19]. Furthermore, ensemble learning strategies are widely used to construct powerful machine learning prediction models. Ensemble learning integrates prediction results from different models to compensate for the prediction shortcomings of individual models and improve overall prediction accuracy^[20].

Current K_{sat} prediction methods typically rely on the most comprehensive input feature sets. However, some of these features, such as organic matter content and porosity, still require complex testing for accurate determination of K_{sat} . Although comprehensive input features can significantly improve model prediction accuracy, they reduce the usability of prediction models. Additionally, traditional gradient boosting algorithms depend on complex tuning processes and have limited K_{sat} prediction accuracy for small feature sets, while the performance of emerging transformer architecture models, such as TabPFN, in predicting K_{sat} remains unclear. Therefore, this study develops an optimized TabPFN-gradient boosting ensemble prediction model by constructing a TabPFN model with a two-stage optimization strategy to explore the performance of this emerging method in K_{sat} prediction. Furthermore, this study compares the TabPFN model with traditional gradient boosting models and random forest (RF). It employs a multimodel stratified weighted ensemble strategy to integrate the local prediction advantages of various models, ultimately constructing a K_{sat} estimation model that surpasses individual models in prediction accuracy and robustness.

1 Materials and Methods

1.1 Dataset and feature selection

This study used the SoilKsatDB database^[21] to construct the training set. This database contains 13 258 K_{sat} measurements, predominantly from North America and Europe, so the model is more applicable to temperate-region soils. Mass fractions of sand, silt, and clay and soil bulk density were selected as input features, as they determine soil pore structure and control saturated seepage, can be rapidly obtained through low-cost laboratory tests

(ASTM D7263, ASTM D2216), and allow quicker K_{sat} range estimates than permeability tests. Other variables, such as organic carbon content, were excluded because their accurate determination requires time-consuming chemical tests, which prevent the rapid prediction of K_{sat} from the most easily and quickly measurable basic soil properties.

When the dataset was constructed, two rounds of data filtering and outlier removal were performed on the SoilKsatDB database. In the first round, 8 266 samples with valid values for the mass fraction of sand, silt, and clay, bulk density, and laboratory-measured K_{sat} were retained. In the second round, a neighborhood similarity method was used to handle anomalous samples. Specifically, a K-nearest neighbor similarity method was applied in the standardized four-dimensional feature space, and a modified Z-score based on median absolute deviation was used as the anomaly score (Eqs. (1) to (6)), with a threshold of 2 to identify anomalous samples.

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & x_{13} & x_{14} \\ x_{21} & x_{22} & x_{23} & x_{24} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & x_{n3} & x_{n4} \end{bmatrix} \quad (1)$$

where n is the total number of samples, and each row represents the four input features of a sample.

$$x_{ij}^* = \frac{x_{ij} - \mu_j}{\sigma_j} \quad (2)$$

where x_{ij} is the feature value, x_{ij}^* is the standardized feature value, μ_j and σ_j are the mean and standard deviation of the j -th feature, respectively.

$$d_{ij} = \sqrt{\sum_{p=1}^4 (x_{ip}^* - x_{jp}^*)^2} \quad (3)$$

where d_{ij} is the Euclidean distance between samples i and j in the standardized feature space.

$$\tilde{K}_i = \text{median}(K_{n_1}, K_{n_2}, \dots, K_{n_k}) \quad (4)$$

$$M_i = \text{median}(|K_{n_1} - \tilde{K}_i|, |K_{n_2} - \tilde{K}_i|, \dots, |K_{n_k} - \tilde{K}_i|) \quad (5)$$

$$Z_i^* = 0.6745 \frac{K_i - \tilde{K}_i}{M_i} \quad (6)$$

where K_i is the K_{sat} value of the i -th sample, and the coefficient 0.6745 is a conversion factor used to balance the data distribution.

After outlier detection and removal, 6 710 samples were ultimately retained to construct the training set. The feature distributions of the filtered samples and the soil texture triangle colored by the K_{sat} distribution are shown in Figs. 1(a)-(e) and Fig. 2, respectively. The sample data statistics are presented in Tables 1 and 2. The original K_{sat} values were uniformly quantified in cm/s. K_{sat} did not conform to a normal distribution but instead exhibited

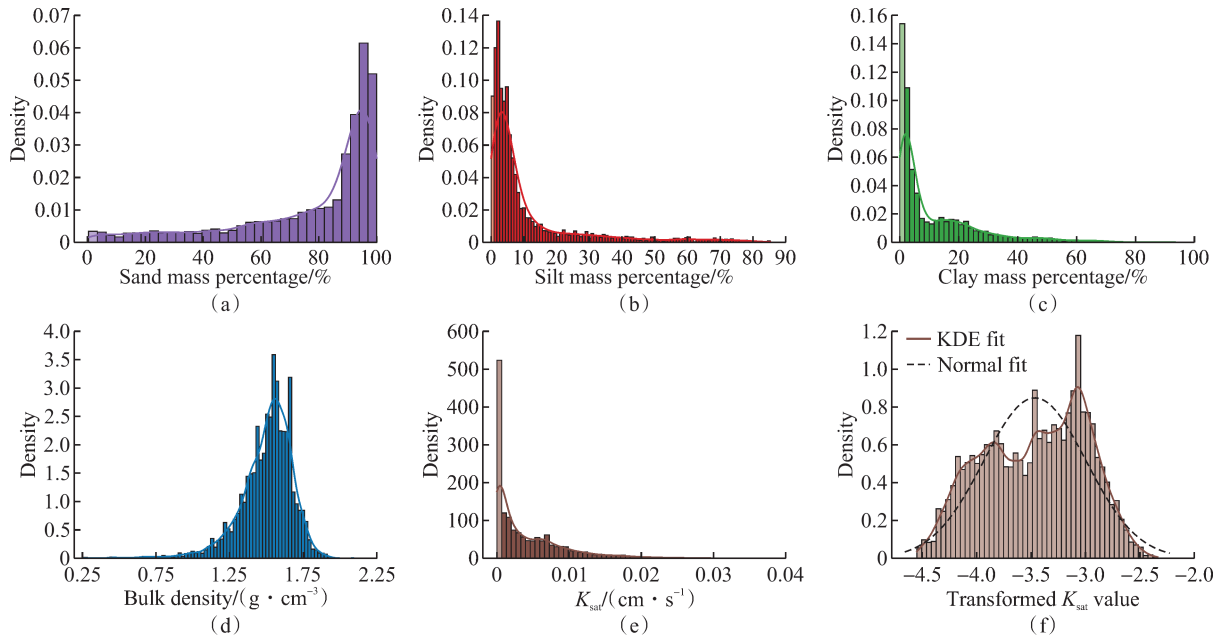


Fig. 1 Feature distribution plots. (a) Mass fraction of sand; (b) Mass fraction of silt; (c) Mass fraction of clay; (d) Bulk density; (e) Original K_{sat} distribution; (f) Box-Cox transformed K_{sat} distribution ($\lambda=0.21$)

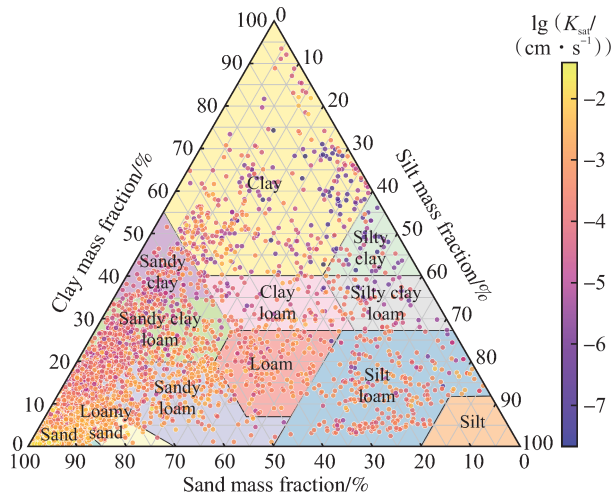


Fig. 2 Soil texture triangle colored by K_{sat} distribution

Table 1 Dataset statistics

Feature	Range	Quartiles (Q_1, Q_2, Q_3)	Mean
Sand	[0.40, 100]	(66.723, 89.400, 95.400)	77.49
Silt	[0, 85.40]	(2.300, 4.608, 10.600)	10.85
Clay	[0, 93.40]	(1.600, 4.500, 17.578)	11.66
Bulk density	[0.26, 2.09]	(1.410, 1.530, 1.620)	1.50
K_{sat}	$[2.15 \times 10^{-8}, 3.89 \times 10^{-2}]$	$(2.78 \times 10^{-4}, 2.01 \times 10^{-3}, 6.57 \times 10^{-3})$	4.15×10^{-3}

Table 2 Soil sample statistics

Soil texture	Number of samples	Soil texture	Number of samples
Sand	3 663	Loam	147
Sandy loam	809	Sandy clay	126
Sandy clay loam	650	Clay loam	95
Loamy sand	500	Silty clay loam	65
Clay	344	Silty clay	45
Silt loam	266		

an extreme right-skewed distribution, spanning several orders of magnitude, which is unfavorable for training regression model algorithms. Thus, the Box-Cox transformation was introduced. The Box-Cox transformation formula is given in Eq. (7), and the transformed distribution (Fig. 1 (f)) closely follows a normal distribution, stabilizing variance and improving model prediction accuracy.

$$K_i^{(\lambda)} = \frac{K_i^\lambda - 1}{\lambda} \quad (7)$$

where λ is the transformation parameter.

1.2 Machine learning methods

The optimal transformation parameter λ was determined using maximum-likelihood estimation. The procedure searches for the value of λ that maximizes the log-likelihood function under which the transformed data follows a normal distribution, yielding an optimal value of $\lambda = 0.21$. To recover the original physical units of the data, the following inverse Box-Cox transformation is applied:

$$y = \begin{cases} (\lambda y^* + 1)^{1/\lambda} & \lambda \neq 0 \\ \exp(y^*) & \lambda = 0 \end{cases} \quad (8)$$

where y^* is the value in the Box-Cox transformed space, and y is the recovered K_{sat} value in cm/s.

A heatmap of Pearson correlation coefficients between features is shown in Fig. 3. The correlations are consistent with soil mechanics principles: K_{sat} is positively correlated with sand content and negatively correlated with silt, clay, and bulk density, reflecting the role of particle composition and packing density in controlling pore structure and seepage resistance.

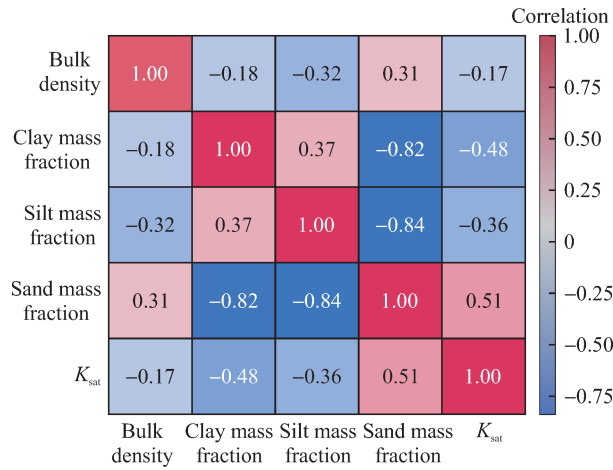


Fig. 3 Feature correlation heatmap

To construct K_{sat} prediction models, this study considered and selected various machine learning methods, optimization strategies, and ensemble strategies for enhanced prediction accuracy. Through extensive preliminary tests, machine learning algorithms, primarily including TabPFN, gradient boosting regression (LightGBM (LGB), XGBoost (XGB), and CatBoost (CB)), and RF, were ultimately selected.

1.2.1 TabPFN

The TabPFN is a foundation model trained on the transformer architecture, suitable for supervised classification and regression tasks on small- to medium-sized tabular data^[19]. When performing inference with real data, TabPFN takes the real training data as context input, feeds it into the pretrained transformer model, and outputs prediction results through a single fast forward computation. Compared with traditional machine learning models, TabPFN significantly reduces training and inference time while achieving prediction performance comparable to or better than traditional algorithms. Furthermore, TabPFN can be automatically tuned through hyperparameter optimization (HPO) frameworks based on Bayesian optimization.

1.2.2 RF and gradient boosting

RF is a classifier that contains multiple decision trees and performs basic ensemble learning based on the bootstrap aggregating principle. Predictions are obtained by averaging across all trees, reducing the high variance inherent in individual decision trees. Gradient boosting decision tree is a sequential ensemble learning method. Gradient boosting corrects the errors of preceding models by iteratively building new models, optimizing along the negative gradient direction of the loss function^[22]. LGB, XGB, and CB are variants of gradient boosting machines^[23]. LGB uses a leaf-wise growth strategy for fast convergence. XGB applies a level-wise growth strategy with L1/L2 regularization. CB uses symmetric decision trees and an ordered target statistics algorithm to prevent overfitting from target leakage.

1.3 Model construction and evaluation

When constructing the prediction models, this study employed various optimization strategies to improve model prediction accuracy. These strategies included TabPFN-HPO based on Bayesian HPO and hyperparameter tuning methods based on the Optuna framework and GridSearchCV. To address the relatively low prediction accuracy of the standard TabPFN, improvements were made by drawing on the mixture of experts (MoE) concept. To further improve the prediction accuracy of the final model, a multimodel ensemble strategy was adopted, integrating the prediction results of all base learners.

Dataset splitting was performed using a standard 8:2 ratio to divide the samples into training and test sets. In this study, no additional independent validation set was constructed; the 20% test set was held out from all model training procedures. To ensure a uniform distribution of soil texture in the two subsets, a stratified sampling method was used. This approach ensures that the proportion of each soil type remains consistent between the training and test sets, thereby avoiding the impact of overall random sampling bias on model robustness and generalization ability.

1.3.1 TabPFN optimized with an expert mechanism

The MoE has been widely used in recent large language models to enhance model performance. This method forms a large model from multiple specialized sub-models (i. e., “experts”). This study proposes a TabPFN strategy optimized with an expert mechanism (hereafter referred to as TabPFN (Exp)). Unlike the standard TabPFN (hereafter referred to as TabPFN (Base)), which inputs the entire dataset, this strategy treats each soil texture category as an independent prediction problem. The dataset is partitioned by soil texture, with each category independently input into a TabPFN model.

Building upon the expert mechanism, this study adopted the automatic HPO method from the TabPFN extension to construct a two-stage optimization strategy. Each sub-model is evaluated against an R^2 threshold of 0.95. Sub-models meeting this threshold are retained directly; otherwise, 50-round hyperparameter tuning is applied. To prevent HPO-induced degradation, the tuned model was adopted only when it outperformed the original. The TabPFN (Exp) model tuned by this two-stage optimization strategy is hereafter referred to as TabPFN (Tuned Exp).

1.3.2 Traditional model optimization strategy

HPO was applied to all models. The Optuna framework, which employs the tree-structured Parzen estimator for efficient Bayesian optimization with early stopping support, was used for the three gradient boosting regressors. GridSearchCV with five-fold cross-validation

was applied for RF, which lacks an intrinsic early stopping mechanism. All models used the same objective (minimizing the root mean square error (RMSE)) over 50 optimization iterations. The specific hyperparameter search space and optimal hyperparameter are shown in Table 3.

Table 3 Optimal hyperparameters of the model

Model	Hyperparameter	Range	Optimal hyperparameter
LGB	Number of estimators	[1 000, 8 000]	4 210
	Learning rate	[0.01, 0.1]	0.077
	Number of leaves	[20, 300]	201
	Maximal depth	[5, 15]	14
	Minimal child samples	[5, 100]	17
	Subsample ratio	[0.6, 1.0]	0.783
	Column sample by tree	[0.6, 1.0]	0.888
XGB	Number of estimators	[1 000, 3 000]	1 000
	Learning rate	[0.01, 0.1]	0.1
	Maximal depth	[9, 11, 12]	12
	Subsample ratio	[0.6, 1.0]	0.926
	Column sample by tree	[0.6, 1.0]	0.860
	Gamma (minimal split loss)	[10 ⁻⁸ , 1.0]	1.14×10 ⁻⁸
CB	Iterations	[1 000, 6 000]	4 784
	Learning rate	[0.01, 0.1]	0.090
	Depth	[4, 10]	10
	L2 leaf regularization	[0.001, 10.0]	0.532
RF	Number of estimators	[5 000, 10 000]	5 000
	Maximal depth	[20, 30]	30
	Minimal samples split	[2, 5]	5
	Minimal samples leaf	[1, 2]	1

1.3.3 Stratified weighted ensemble strategy

Because the number of samples is unevenly distributed across soil texture types, a single globally optimized weight vector may not achieve optimal weighting for all soil textures. Therefore, this study adopted a stratified weighted ensemble strategy to determine an optimized set of weights for each soil texture category individually. The test set is stratified by texture label, and predictions from the seven models—LGB, XGB, CB, RF, TabPFN (Base), TabPFN (Exp), and TabPFN (Tuned Exp)—are inverse Box-Cox transformed back to the original scale. For each texture stratum, the Nelder-Mead simplex algorithm minimizes the mean squared error without non-negativity constraints, after which the weight vector is normalized to sum to one. This formulation allows negative weights, enabling correction of systematic biases in the base models. The resulting local weight vectors are then recombined to evaluate the complete test set and determine the overall ensemble performance.

1.3.4 Performance evaluation metrics

Model performance was evaluated using R^2 , RMSE, and MAE, all computed in original units (cm/s) after inverse Box-Cox transformation. The specific formulas are shown as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (9)$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (10)$$

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (11)$$

where N is the number of samples; y_i is the true value; $\hat{y}_i$ is the predicted value.

2 Results and Discussion

2.1 Comparison of standard TabPFN and optimized TabPFN

Fig. 4 shows the R^2 score, MAE, and RMSE results for the three TabPFN models. The TabPFN (Base) model, trained on the entire dataset, achieved an R^2 of 0.819 3, an RMSE of 0.002 2 cm/s, and an MAE of 0.001 1 cm/s. These results indicate that the standard TabPFN model captures the basic trends in the dataset; however, because of the uneven distribution of soil texture sample counts, it still struggles to accurately predict the K_{sat} in the test set. After the first expert mechanism optimization, the overall model R^2 significantly increased to 0.922 5, representing an improvement of 12.6%, while the RMSE and the MAE decreased to 0.001 5 cm/s and 4.83×10⁻⁴ cm/s, respectively. The prediction R^2 score for some soil categories increased by up to 205.3% (silty clay). The results demonstrate the effectiveness of the stratified strategy, as providing TabPFN with a homogeneous training context helps improve model prediction accuracy. Silt shows a smaller improvement margin than clay. This finding can be attributed to the unique hydraulic behavior of silty soils. Silty soils occupy a transitional position between sandy and clayey soils, and their seepage is jointly controlled by capillary effects from fine particles and pore channel flow through coarser fractions. This characteristic results in highly heterogeneous pore size distributions that are often bimodal,

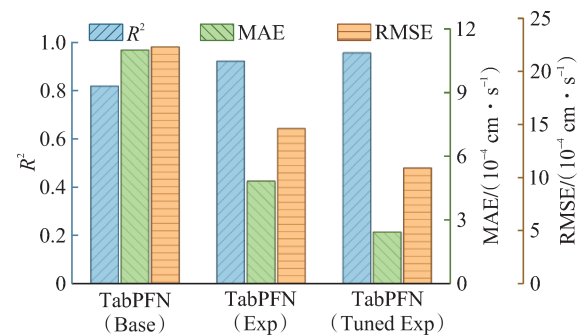


Fig. 4 Comparison of evaluation results for the three TabPFN models

which weakens the deterministic mapping between particle composition and K_{sat} .

The transformer core architecture of TabPFN employs an in-context learning (ICL) mechanism, which uses the training set as context or a prompt to infer data patterns for the regression task in a single forward pass. When the complete dataset is used for training, a highly heterogeneous context is provided to the TabPFN model because of the uneven distribution of sample results, potentially exceeding the scope of data structures favored by its pretrained model priors. Consequently, the prediction accuracy of TabPFN (Base) is lower than that of TabPFN (Exp). The stratification strategy addresses dataset heterogeneity. The data patterns of the decomposed regression tasks become more regular, helping the ICL mechanism of TabPFN to find a match within its pretrained prior space, thus significantly improving model prediction accuracy. This phenomenon is clearly observable in the model prediction accuracy for each soil texture individually, as shown in Fig. 5.

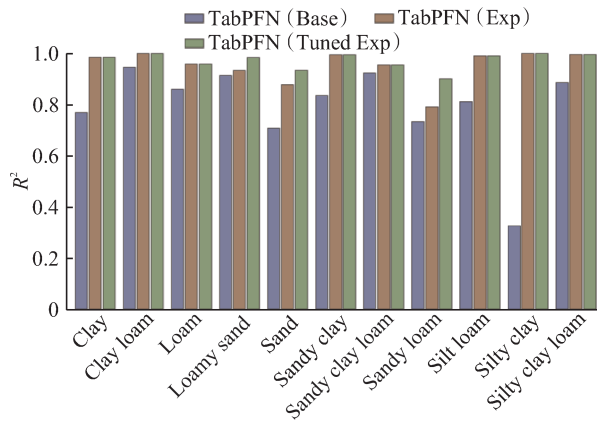


Fig. 5 R^2 score of the three TabPFN models for each soil texture

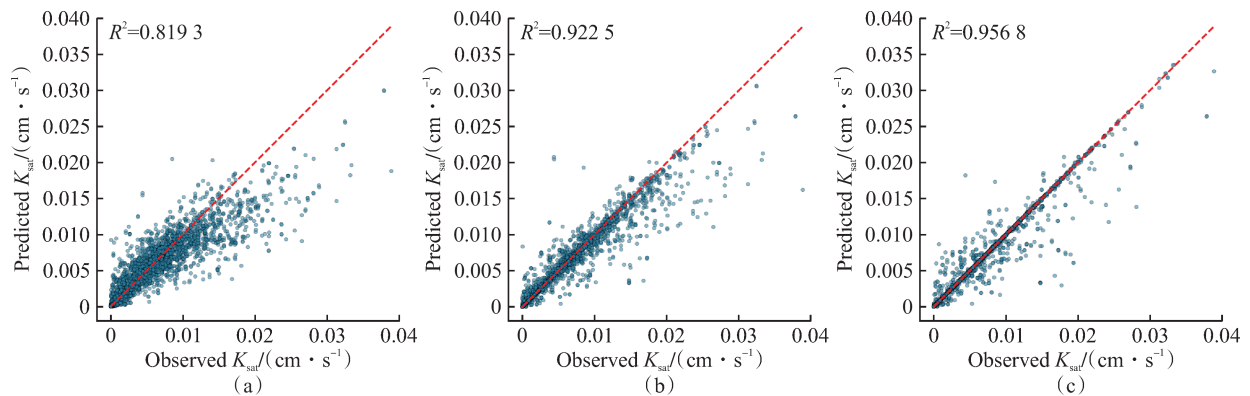


Fig. 6 Predicted vs. observed scatter distributions for the three TabPFN models. (a) TabPFN (Base); (b) TabPFN (Exp); (c) TabPFN (Tuned Exp)

2.2 Comparison of all model prediction results

2.2.1 Hyperparameter optimization results for traditional models

The optimization history and hyperparameter importance plots for the LGB, XGB, CB, and RF models are

shown in Fig. 7. The optimization history plots show how model RMSE changes during the search process. The different hyperparameter combinations cause the RMSE to fluctuate mainly between 0.1 and 0.2. Except for the XGBoost model, the models exhibit a certain sensitivity

Although the overall performance of the models optimized with the stratification strategy significantly improved, room for improvement in individual prediction accuracy persists for a few soil textures, such as sand and sandy loam. After the two-stage optimization strategy was implemented, the K_{sat} prediction accuracy for soil textures like sand and sandy loam was further enhanced, with an 18.9% increase in the maximum R^2 score. The TabPFN (Tuned Exp) model achieved an overall performance R^2 of 0.956 8, representing a 16.78% increase over TabPFN (Base), and the RMSE decreased to 0.001 1 cm/s. This comparison indicates that targeted HPO can further unlock the upper limit of model performance, especially when handling a limited number of input features or complex data distribution patterns within a dataset.

Fig. 6 shows the predicted vs. observed scatter distributions for the three models. As shown in Fig. 6(a), TabPFN (Base) exhibits considerable heteroscedasticity, with a tendency to underestimate high K_{sat} values (> 0.01 cm/s, primarily from sand and loamy sand), while low K_{sat} predictions remain loosely clustered around the 1:1 line. After the stratification strategy is used for optimization, TabPFN (Exp) scatter points converge toward the 1:1 line with notably reduced dispersion in the high-value region. Following the two-stage optimization, the TabPFN (Tuned Exp) model further improves accuracy. Its scatter points are highly concentrated on the 1:1 prediction line, with a substantial reduction in dispersion. The prediction bias for scatter points in the high-value region is significantly reduced, with only a few difficult-to-predict samples scattered on either side of the prediction line.

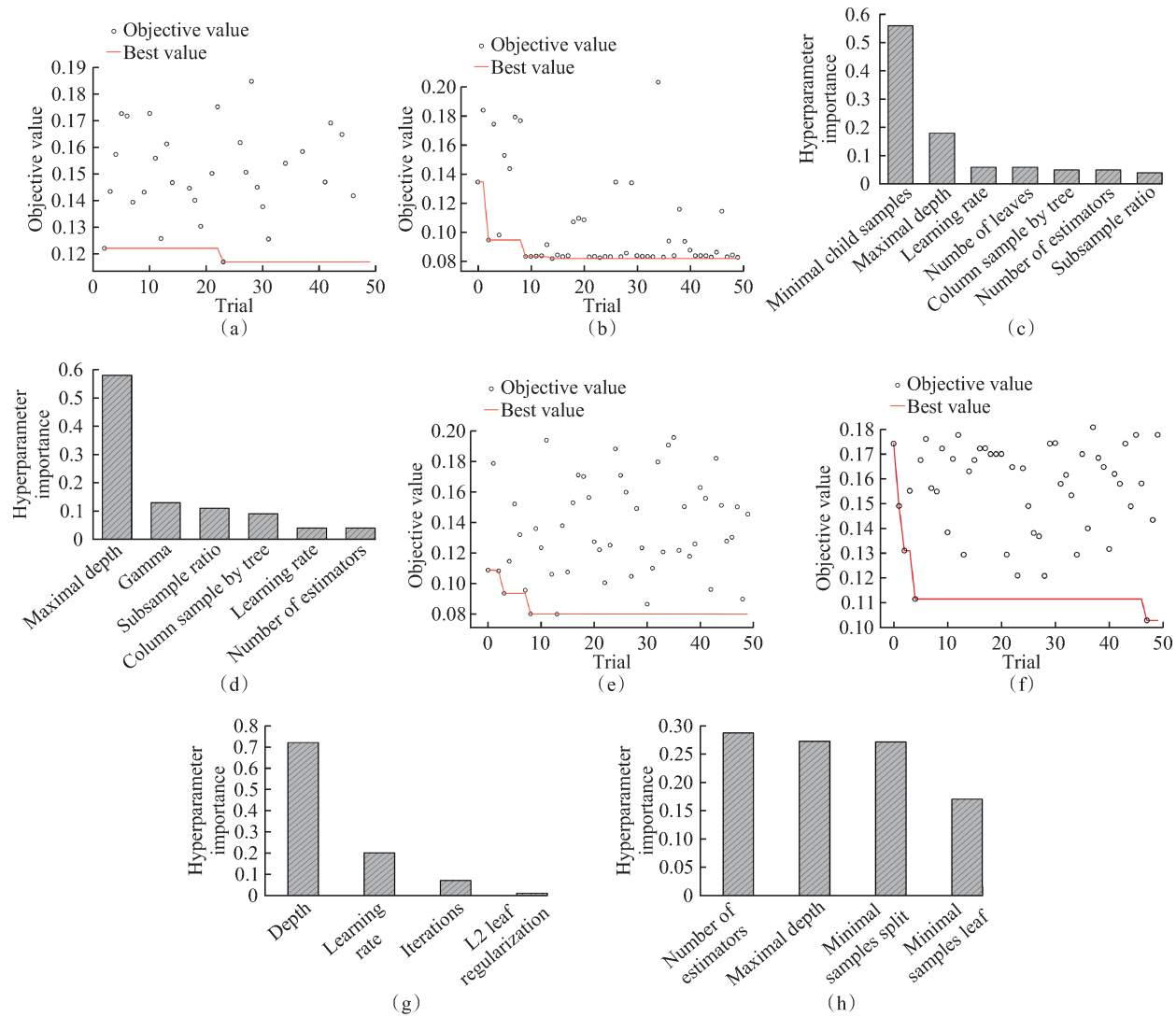


Fig. 7 Traditional model HPO history and hyperparameter importance. (a) LGB HPO history; (b) XGB HPO history; (c) LGB hyperparameter importance; (d) XGB hyperparameter importance; (e) CB HPO history; (f) RF HPO history; (g) CB hyperparameter importance; (h) RF hyperparameter importance

to hyperparameter combinations. Model sensitivity to different hyperparameters also varies, and model accuracy may largely depend on a small subset of hyperparameters. The LGB model is highly sensitive to the minimal child samples, a regularization parameter that controls tree complexity and is related to the leaf-wise growth strategy of the LGB model. The XGB and CB models are highly sensitive to tree depth, the most direct mechanism for controlling complexity. The four hyperparameter categories for the RF model have relatively balanced importance, indicating that the hyperparameters collectively influence model accuracy. After 50 rounds of hyperparameter search, all models obtained the optimal hyperparameter combination that minimized the RMSE, as shown in Table 3.

2.2.2 Stratified ensemble strategy results

Fig. 8 shows the stratified ensemble weight results for 11 soil textures. The ensemble model (EN) assigns dif-

ferent sub-model weights to each soil texture. The stratification-optimized TabPFN (Exp) and TabPFN (Tuned Exp) models receive a weight greater than 0.5 for seven soil textures. The TabPFN (Base) is assigned extremely low or negative weights in all stratified predictions, where a negative weight indicates that the objective function utilized its error structure to correct the systematic bias in the ensemble. For textures such as loam, clay, loamy sand, and sandy clay loam, traditional gradient boosting and RF models outperform TabPFN. This superiority arises because tree-based models better capture the complex nonlinear K_{sat} relationships in fine-grained soils through hierarchical feature partitioning, while the limited context available for these smaller soil texture subsets constrains the predictive capability of TabPFN. After weighting and combining all stratified predictions, the final ensemble model achieves an R^2 of 0.961 7 and an RMSE of 0.001 0 cm/s, outperforming all single

models. The stratified ensemble strategy dynamically adjusts weight allocation based on data characteristics, assigning the optimal sub-model combination for each soil texture and enabling the ensemble to surpass any individual model.

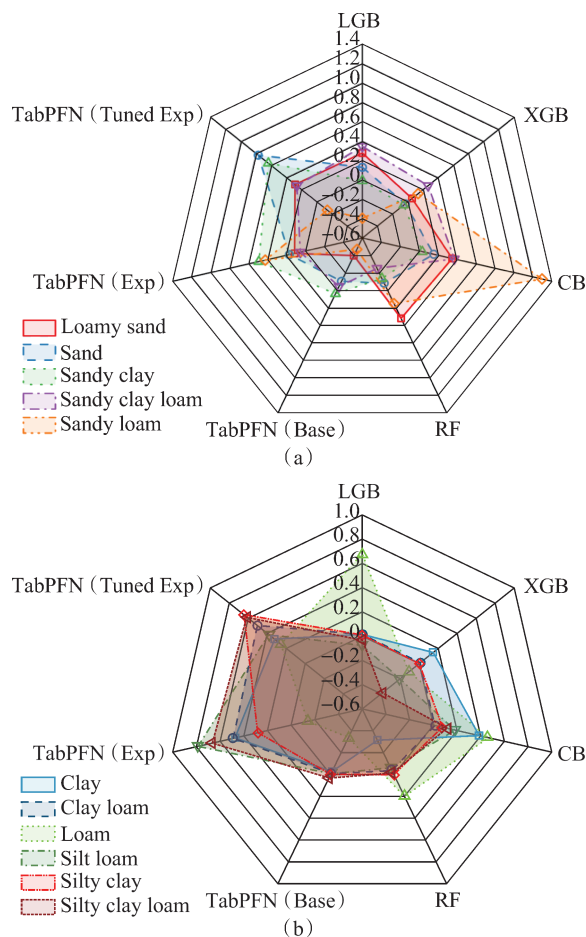


Fig. 8 Radar chart of stratified model weights for the ensemble model. (a) Five types of soil; (b) Six types of soil

2.2.3 Comparison of model evaluation results

Fig. 9 summarizes the R^2 score, MAE, and RMSE evaluation results for all models, including the ensemble model. The three gradient boosting models and the RF model show similar prediction accuracy, with CB and LGB achieving slightly higher R^2 than XGB and RF. TabPFN (Base) performs the worst among all models, reflecting the limitations of unoptimized ICL on tasks with few features and heterogeneous samples. However, this disadvantage is largely overcome by the stratification and HPO, and the optimized models show higher performance potential than the gradient boosting models. The EN model achieves better R^2 and RMSE than the best single model, TabPFN (Tuned Exp), although with a slight increase in MAE, indicating that the ensemble strategy reduces large-error predictions at the cost of a minor systematic bias.

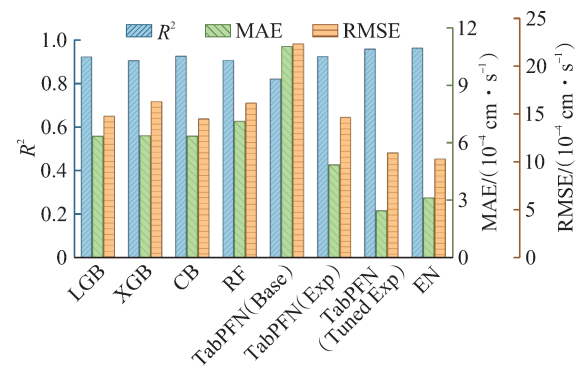


Fig. 9 Comparison of all model evaluation results

Fig. 10 shows the predicted vs. observed scatter distributions for the LGB, XGB, CB, RF, and EN models. Similar to the TabPFN models, the gradient boosting and RF models have insufficient prediction accuracy for K_{sat} values above 0.01 cm/s. In particular, when the XGB and RF models predict extremely high values above 0.03 cm/s, the prediction points deviate significantly from the 1:1 prediction line. For high values, the LGB and CB models have slightly better prediction accuracy than the XGB and RF models. Similar to the TabPFN models, all models tend to underestimate K_{sat} in the high-value range, which is related to the distinct right-skewed distribution of K_{sat} . This underestimation is primarily attributable to two factors. First, high- K_{sat} samples (>0.01 cm/s) predominantly originate from coarse-grained soils where macropore networks and preferential flow paths control seepage^[24]. Second, since particle composition and bulk density cannot characterize macropore connectivity, the sparsity of high- K_{sat} samples also causes the RMSE-based objective function to inherently favor the densely populated central region. The EN model shows a tightly converged distribution of scatter around the 1:1 line with less dispersion than TabPFN (Tuned Exp). This comparison confirms that the ensemble strategy effectively reduces systematic bias and mitigates heteroscedasticity across soil textures.

The heatmap in Fig. 11 shows the prediction R^2 and RMSE for each soil texture across all models. All models exhibit higher errors for sand. For the gradient boosting models without a stratified strategy and the TabPFN models without two-stage optimization, the prediction R^2 is below 0.9. The optimized TabPFN (Tuned Exp) and EN models achieve an R^2 of approximately 0.93, but their RMSE remains slightly high. This result emerges possibly because sand samples in the dataset are relatively dispersed and contain many special values, making it difficult for the current input features to capture variation patterns under high K_{sat} conditions. The heatmap also reveals that no single model has an absolute advantage for all soil textures. For example, for loamy sand, the R^2 of the LGB, XGB, and CB models is sig-

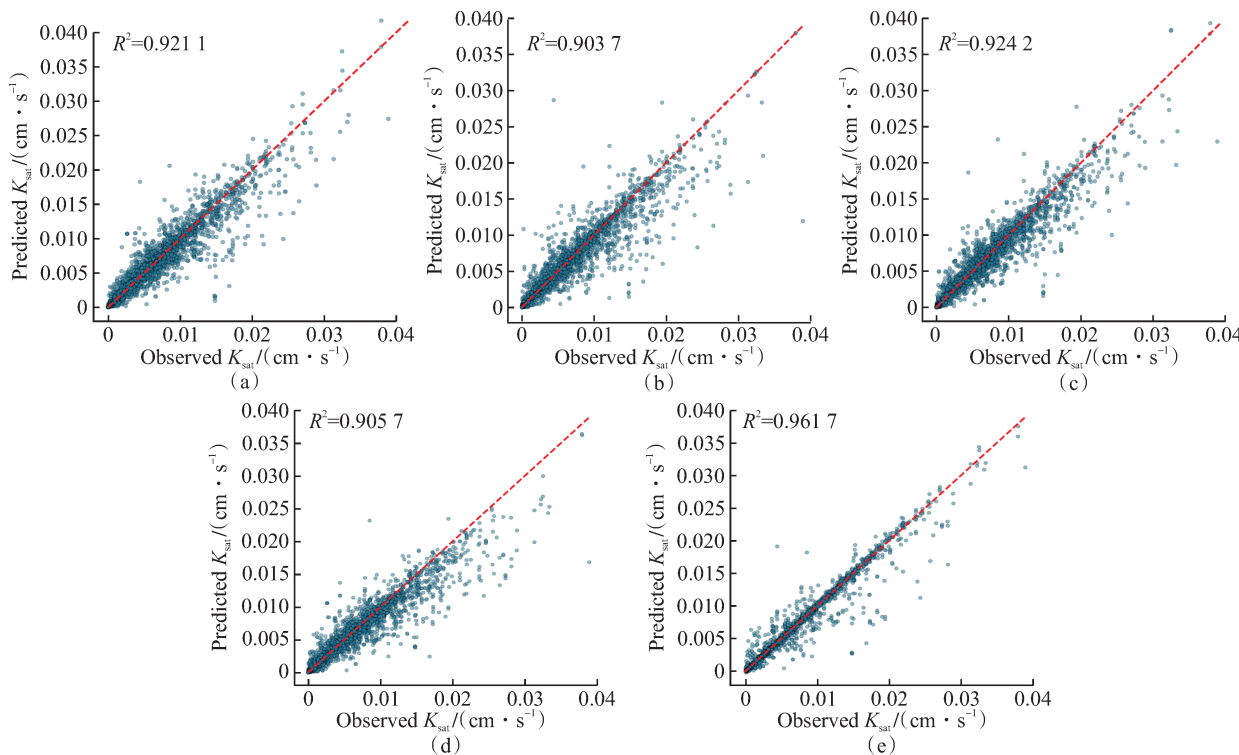


Fig. 10 Predicted vs. observed scatter distributions for the gradient boosting, RF, and EN models. (a) LGB; (b) XGB; (c) CB; (d) RF; (e) EN

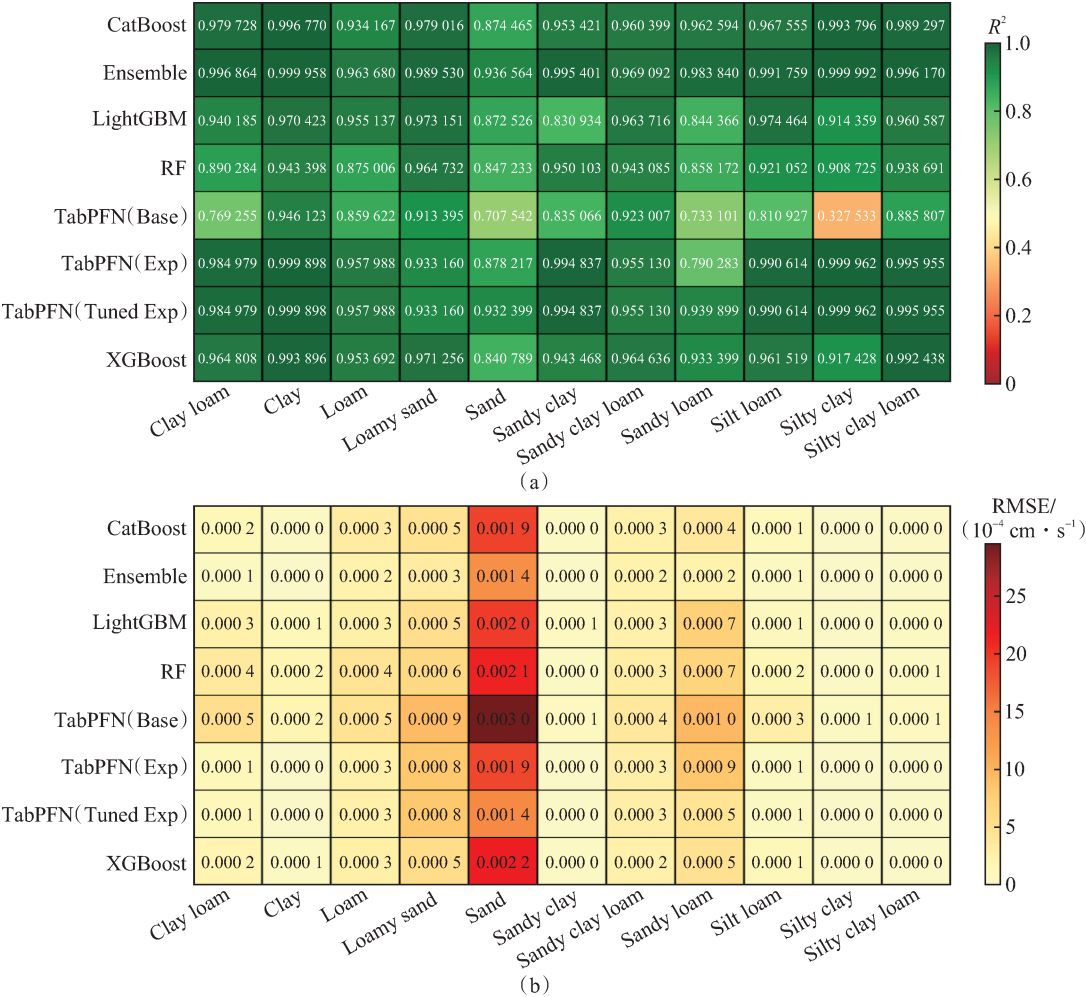


Fig. 11 Prediction accuracy and error heatmap of each model for different soil textures. (a) R^2 score; (b) RMSE

nificantly better than that of TabPFN (Tuned Exp), whereas for sandy loam, CB outperforms all other models except EN. This comparison indicates the necessity of an ensemble strategy. By integrating the predictive strengths of every single model across different soil textures, a superior ensemble prediction model can be constructed. Recent studies^[25] reported that ANN-based PTFs for saturated hydraulic conductivity generally achieved R^2 values of 0.710–0.974. In comparison, the ensemble model proposed in this study achieved a prediction accuracy of $R^2 = 0.9617$, indicating that its performance is within the upper range of reported results.

3 Conclusions

This study built an expert mechanism, optimization strategies, and a stratified ensemble strategy to construct a multimodel ensemble K_{sat} prediction model based on optimized TabPFN and gradient boosting. The main research conclusions are as follows:

(1) The expert mechanism and two-stage optimization strategy can significantly improve the performance of the TabPFN model in predicting soil K_{sat} . After optimization, the TabPFN (Tuned Exp) model had a 17% higher prediction accuracy (R^2 of 0.9568) and, for a single soil texture, a 205.3% higher maximum prediction accuracy. Overall, it outperforms all gradient boosting models and RF models.

(2) The stratified weighted ensemble strategy can dynamically allocate optimal weights to sub-models based on soil texture, effectively integrating the predictive advantages of every single model across different soil textures. This integration improves overall prediction accuracy, reduces heteroscedasticity, and lowers prediction error. The ensemble model achieved an R^2 of 0.9617 and an RMSE of 0.0010 cm/s, outperforming all single models.

(3) Due to the data dispersion of sandy soil samples in the dataset and the right-skewed distribution of K_{sat} , all models exhibited a similar underestimation trend and slightly higher prediction error when predicting sandy soil samples with K_{sat} values greater than 0.01 cm/s. Prediction accuracy can be improved by introducing pore structure characterization parameters or data augmentation strategies.

This study provides a new method for rapidly predicting saturated hydraulic conductivity, but some limitations remain. The current input features cannot directly characterize the pore structure of coarse-grained soils, which limits the prediction accuracy for some samples. In addition, the training data were obtained mainly from laboratory tests, and future studies should integrate multi-source data to improve their applicability to field-scale prediction.

References

- [1] GARCÍA-GUTIÉRREZ C, PACHEPSKY Y, MARTÍN M Á. Technical note: Saturated hydraulic conductivity and textural heterogeneity of soils [J]. *Hydrology and Earth System Sciences*, 2018, 22(7): 3923–3932.
- [2] CHAPUIS R P. Predicting the saturated hydraulic conductivity of soils: A review [J]. *Bulletin of Engineering Geology and the Environment*, 2012, 71(3): 401–434.
- [3] DEXTER A R, CZYŻ E A, GAŹE O P. Soil structure and the saturated hydraulic conductivity of subsoils [J]. *Soil and Tillage Research*, 2004, 79(2): 185–189.
- [4] QI L, LI X, PAN Y, et al. Impacts of forest fire on hydro-sediment using the SWAT model: A case study in the Zhangweinan Canal, China [J]. *Acta Geophysica*, 2025, 73(5): 4761–4780.
- [5] WANG X H, MA C, WANG Y Q, et al. Effect of root architecture on rainfall threshold for slope stability: Variabilities in saturated hydraulic conductivity and strength of root-soil composite [J]. *Landslides*, 2020, 17(8): 1965–1977.
- [6] DOS SANTOS R C V, VARGAS M M, TIMM L C, et al. Examining the implications of spatial variability of saturated soil hydraulic conductivity on direct surface runoff hydrographs [J]. *Catena*, 2021, 207: 105693.
- [7] ZHAO Y L, WANG Y Q, ZHANG X C. Spatial and temporal variation in soil bulk density and saturated hydraulic conductivity and its influencing factors along a 500 km transect [J]. *Catena*, 2021, 207: 105592.
- [8] LIU J Y, WANG X, REN X W. Hydraulic conductivity and particle size of soils: Modeling and experiment [J]. *Acta Geotechnica*, 2025, 20(7): 3395–3409.
- [9] MCDOWELL G, DE BONO J. Relating hydraulic conductivity to particle size using DEM [J]. *International Journal of Geomechanics*, 2021, 21: 06020034.
- [10] PHAM K, WON J. Enhancing the tree-boosting-based pedotransfer function for saturated hydraulic conductivity using data preprocessing and predictor importance using game theory [J]. *Geoderma*, 2022, 420: 115864.
- [11] ZHANG Y G, SCHAAP M G. Estimation of saturated hydraulic conductivity with pedotransfer functions: A review [J]. *Journal of Hydrology*, 2019, 575: 1011–1030.
- [12] REHMAN M, HAFEEZ M B, KRAWCZUK M. A comprehensive review: Applications of the Kozeny-Carman model in engineering with permeability dynamics [J]. *Archives of Computational Methods in Engineering*, 2024, 31(7): 3843–3855.
- [13] DIAZ-GONZALEZ F A, VUELVAS J, CORREA C A, et al. Machine learning and remote sensing techniques applied to estimate soil indicators—Review [J]. *Ecological Indicators*, 2022, 135: 108517.
- [14] LIU Y F, NI J J, ZHANG F, et al. Bayesian-informed DNN and ensemble learning for predicting soil water characteristic curves from easily measurable parameters [J]. *Transportation Geotechnics*, 2026, 56: 101791.
- [15] DONG S, RUI X F, ZUO C. Research progress on intelligent detection and analysis algorithms for pavement damage [J]. *Journal of Southeast University (Natural Sci-*

- ence Edition), 2025, 55(4): 1131-1153. (in Chinese)
- [16] SONG Y, WANG Y. Multi-frame radar HRRP target recognition using MFA-Net [J]. Journal of Southeast University (English Edition), 2025, 41(3): 384-391.
- [17] SINGH V K, PANDA K C, SAGAR A, et al. Novel genetic algorithm (GA) based hybrid machine learning - pedotransfer function (ML-PTF) for prediction of spatial pattern of saturated hydraulic conductivity [J]. Engineering Applications of Computational Fluid Mechanics, 2022, 16(1): 1082-1099.
- [18] ALBALASMEH A, MOHAWESH O, GHARAIBEH M, et al. Artificial neural network optimization to predict saturated hydraulic conductivity in arid and semi-arid regions [J]. Catena, 2022, 217: 106459.
- [19] RUNDEL D, KOBIALKA J, VON CRAILSHEIM C, et al. Interpretable machine learning for TabPFN [C]// Explainable Artificial Intelligence. Cham: Springer, 2024: 465-476.
- [20] DONG X B, YU Z W, CAO W M, et al. A survey on ensemble learning [J]. Frontiers of Computer Science, 2020, 14(2): 241-258.
- [21] GUPTA S, HENGL T, LEHMANN P, et al. SoilKsatDB: Global database of soil saturated hydraulic conductivity measurements for geoscience applications [J]. Earth System Science Data, 2021, 13(4): 1593-1612.
- [22] NATEKIN A, KNOLL A. Gradient boosting machines, a tutorial [J]. Frontiers in Neurorobotics, 2013, 7: 21.
- [23] HU S Y, XIE Z N, YANG Y. Interference wind pressure prediction of high-rise buildings with square section based on machine learning [J]. Journal of Southeast University (Natural Science Edition), 2024, 54(6): 1425-1437. (in Chinese)
- [24] CHEN R, CHEN H, FAN Z H, et al. Experimental study on the unsaturated shear strength of stabilized aeolian sand considering geopolymer content [J]. Journal of Southeast University (Natural Science Edition), 2026, 56(2): 302-311. (in Chinese)
- [25] MOZAFFARI H, PAKJOO M, NEMATOLLAHI M A, et al. Predicting soil hydraulic conductivity: A review of artificial neural networks applications [M]// Artificial Intelligence Applications for a Sustainable Environment. Cham: Springer, 2025: 441-462.

预测土饱和导水率的优化 TabPFN-梯度提升多模型集成策略

刘翼飞¹, 倪钧钧¹, SADEGHI Hamed²

(1. 东南大学交通学院, 南京 211189; 2. Department of Civil Engineering, Sharif University of Technology, Tehran 1458889694, Iran)

摘要: 土的饱和导水率是岩土工程和水文学中的关键水力参数, 传统试验测定方法耗时费力, 难以满足工程实践中快速应用的需求。本文基于易于测得的土壤基本性质构建预测模型, 以颗粒组成质量分数和容重作为输入特征, 对饱和导水率进行预测。提出了一种优化的表格先验数据拟合网络 (TabPFN) 模型, 该模型融合了专家机制和两阶段优化策略, 并将其性能与传统梯度提升模型和随机森林模型进行了比较。构建了多模型分层加权集成策略, 以综合各类模型的优势。结果表明, 专家机制和两阶段优化策略显著提升了 TabPFN 的性能。优化后的 TabPFN 模型最高预测精度 R^2 为 0.956 8, RMSE 为 0.001 1 cm/s; 与优化前模型相比, R^2 提高了 16.78%, 且优于所有梯度提升模型和随机森林模型。分层加权集成策略能够有效整合各子模型在不同土壤质地条件下的预测优势, 取得了最高的预测精度 (R^2 为 0.961 7, RMSE 为 0.001 0 cm/s), 并有效降低了预测误差和异方差性。本文提出的优化集成模型为工程实践中快速估算饱和导水率提供了一种新的智能化方法。

关键词: 土壤饱和导水率; 机器学习; TabPFN; 梯度提升; 集成策略

中图分类号: TU42