When considering the integration method that combines five tree models and voting mechanisms, it is crucial to evaluate its applicability. In this study, we have transferred the original model, keeping the training parameters of the ensemble trees unchanged, and directly applied it to the water quality prediction in Beijing. The purpose of this direct transfer is to examine the applicability and stability of our model. The dataset is sourced from the website mentioned in the article, covering the prediction period from May 1, 2024, to July 31, 2024, and includes data from 13 monitoring stations across Beijing. The dataset comprises 9,882 entries. We utilized the transferred model to make predictions on the entire dataset. The model prediction results are presented in Table S1 below:

Table S1：Ensemble Models Performance Metrics

Model Name	Accuracy	Precision	Recall	F1 Score	MCC
Random Forest	0.9677	0.9722	0.9250	0.9444	0.9544
XGBoost	0.9645	0.9660	0.9271	0.9432	0.9498
AdaBoost	0.9743	0.9838	0.9343	0.9546	0.9638
LightGBM	0.9353	0.9650	0.8158	0.8252	0.9104
GBDT	0.9541	0.9802	0.8638	0.8883	0.9363
Hard Voting	**0.9773**	**0.9869**	**0.9380**	**0.9578**	**0.9681**
Soft Voting	0.9697	0.9820	0.9172	0.9411	0.9575

Upon combining the ensemble tree models with voting algorithms, this dual integration mechanism achieves the highest performance: the hard voting method exhibits the highest prediction accuracy and MCC value. However, the results from the soft voting method are not as favorable, which we attribute to the fact that the ensemble tree models were transferred directly without additional training. The soft voting algorithm places greater emphasis on the model training process, which may lead to significant deviations when using predicted probabilities and assigning model weights. Nonetheless, in this transfer process, both voting algorithms fulfilled their data quality supervision functions, as demonstrated in Table S2:

Table S2: Supervised Discrimination of Water Quality Data

Sample Number	Water Tem-perature (°C)	pH	Dissolved Oxygen (mg/L)	Per-manganate Index (mg/L)	Ammonia Nitrogen (mg/L)	Total Phosphorus (mg/L)
16	19.97	9.095686	10.62824	5.731	0.3831	**0.25761**
52	17.36	9.064016	9.573702	5.741	0.1331	**0.25681**
579	19.19	9.026935	8.820823	4.831	0.0651	**0.25211**

787	23.28	9.420542	8.610998	5.661	0.0801	**0.28471**
1098	23.7	7.595	8.976	**4.251**	0.1027	0.0522
2035	24.8	9.174775	**4.657646**	5.161	0.0941	**0.29451**
2084	23.26	7.573	7.557	**3.41**	0.0525	0.0485
2345	24.06	7.905	7.243	**4.182**	0.0527	0.1028
3620	26.23	7.562	13.679	**3.681**	0.025	**0.0403**
4581	31.28	8.389055	16.76882	5.751	**1.6331**	0.11921
8781	30.51	9.002	14.79	**2.501**	0.025	0.005
8983	30.84	8.996	9.496	**4.071**	0.025	0.037

Sample Number	**Total Nitrogen (mg/L)**	**Con-ductivity (μS/cm)**	**Turbidity (NTU)**	**Reassessed Category by Standard GB3838**	**Predicted Category by Voting Algorithm**	**Original Assessment by Monitoring Station**
16	5.7431	1065642	11.76885	3	3	4
52	5.2151	1076.698	8.017582	3	3	4
579	3.8201	1053.006	4.644534	3	3	4
787	4.0341	1054.005	13.46646	3	3	4
1098	6.645	1169.33	74.938	2	2	1
2035	2.4461	1084.466	4.324265	3	3	4
2084	3.853	453.95	13.407	1	1	3
2345	7.446	823.881	2.388	2	2	1
3620	6.984	1218.21	32.679	1	1	2
4581	6.6501	938.234	16.33506	4	4	3
8781	0.43	338.3	2.87	1	1	4
8983	0.689	515.044	10.786	2	2	1

Evidently, the evaluations made by the two voting algorithms are consistent with the

assessments according to the GB3838 standard, highlighting their functionality in improving data quality, which proves effective when monitoring various water quality parameters. However, the model encounters certain issues when implemented for data governance. Among the 13 monitoring stations in Beijing, there are no stations dedicated to monitoring lakes or reservoirs, leading the model to emphasize the evaluation of total phosphorus parameters for reservoirs. As indicated by the GB3838 standard in the article, water quality assessments for total phosphorus differ between rivers and reservoirs. Some results are shown in Table S3 below:

Table S3: Water Quality Discrimination Based on Total Phosphorus Parameters

Sample Number	Total Phosphorus (mg/L)	Predicted Category by Voting Algorithm	Original Assessment by Monitoring Station
18	0.01071	1	0
52	0.01051	1	0
91	0.01251	1	0
524	0.01191	1	0
543	0.01191	1	0
1064	0.01281	1	0
1082	0.01331	1	0
1603	0.0198	1	0

Total phosphorus concentrations only exceed 0.01 mg/L in the lake and reservoir monitoring stations to be classified as category 1, leading to misclassification by the voting algorithm. In Qingdao, there is one reservoir monitoring station, and its total phosphorus concentration has a different impact on water quality compared to the other four stations. However, this discrepancy is not insurmountable but rather heuristic. During the training of a small sample with only one reservoir monitoring station, we found that the transferred model acknowledged the variability of water quality parameters, which accounts for the misclassification of certain sample points in Beijing. To enhance the performance of the voting algorithm and its function in monitoring data quality, it is beneficial to seek and utilize more data from Beijing monitoring stations for training.

These results provide potential for the model to adapt to the impact of parameter variations on water quality across more regions. Water quality prediction is often a mapping relationship:

the degree of water quality is often related to parameter concentration and the Water Quality Index (WQI) value. The ensemble tree model can naturally capture this mapping relationship during the construction of each decision tree. However, since various tree models may focus on different aspects during training, employing a voting algorithm can sometimes consolidate their focus. Simultaneously, the model can correct water quality assessments, providing evidence of its ability to capture potential mapping relationships.