

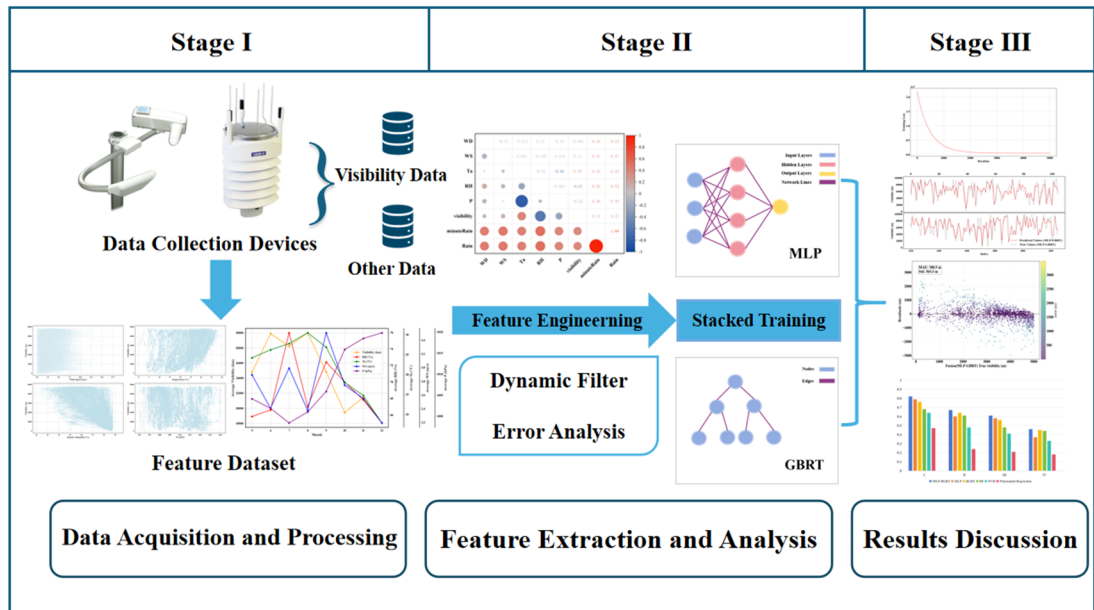
Application-Oriented Stacked MLP-GBRT Modeling for Atmospheric Visibility Prediction

Yuguo Ni^{a,b}, Chenbo Xie^{a*}, Jian Huang^b, Tongzhan Xue^b, Zichen Zhang^{a,b}, Jianfeng Chen^a

^a State Key Laboratory of Laser Interaction with Matter, Anhui Institute of Optics and Fine Mechanics, HFIPS, Chinese Academy of Sciences, Hefei 230031, China

^b School of Environment and Energy Engineering, Anhui Jianzhu University, Hefei 230009, China

*Correspondence author.



Graphical Abstract

Highlights:

- High-resolution, low-cost in-situ observations enable fine-scale visibility modeling after dynamic filtering.
- A stacked MLP-GBRT with a linear meta-learner (Huber loss) consistently outperforms single models.
- Gains are largest under low-visibility (<5 km), improving both continuous scores and graded verification.

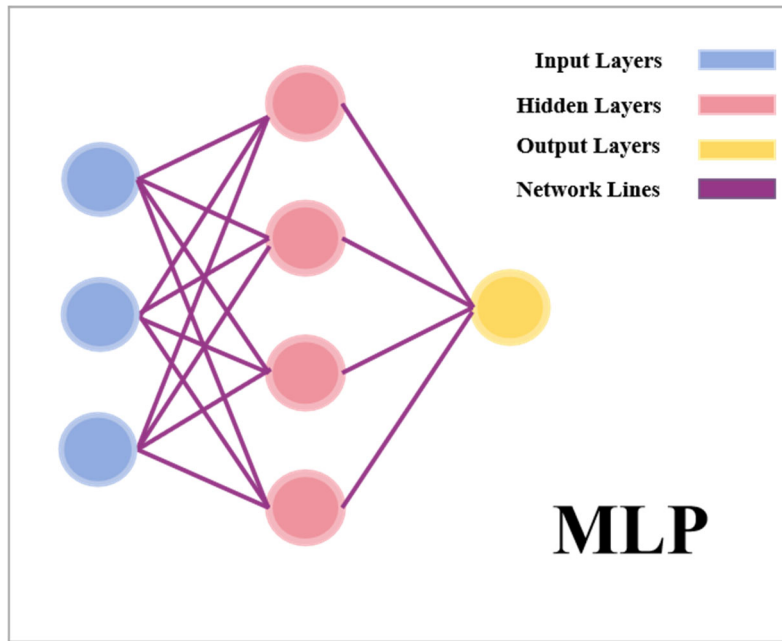


Figure S1. The structure of MLP

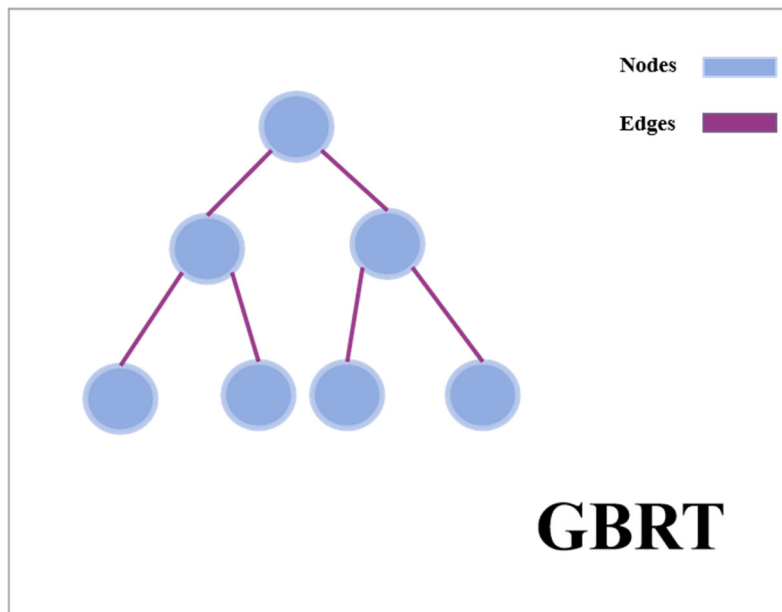


Figure S2. The structure of GBRT



Figure S3. PWD50



Figure S4. WXT5360

Table S1 GBRT Parameters Table

Model	Tree Structure Control	Regularization	Dropout	Others
GBRT	max_depth: 5	Subsample:0.8	validation_fraction:0.1	loss=Huber Loss

min_samples_split: 10	max_features:sqrt	n_iter_no_change:10	learning_rate=0.1
min_samples_leaf: 5	ccp_alpha:0.01	Tol:1e-4	n_estimators=200
max_leaf_nodes: 30	min_impurity_decrease:0.001	-	-

The GBRT model is optimized using key hyperparameters to balance performance and generalization. Maximum tree depth (max_depth) is set to 5 to limit complexity. Each split requires at least 5 samples (min_samples_split), with a minimum of 5 samples per leaf (min_samples_leaf). The maximum number of leaf nodes is capped at 30 (max_leaf_nodes), and subsampling is set to 0.8 to reduce variance. Feature randomness is introduced by limiting the number of features considered per split (max_features). Huber loss (loss) with a quantile threshold of 0.9 ensures robustness to outliers. Minimum impurity decrease (min_impurity_decrease) is set to 0.001 to allow only meaningful splits. Cost-complexity pruning (ccp_alpha) is applied at 0.01 for tree simplification. Training uses 200 trees (n_estimators) with a learning rate of 0.1. A 10% validation set (validation_fraction) enables early stopping if no improvement occurs over 10 iterations (n_iter_no_change), with a tolerance of 1e-4 (tol). Model training employs 5-fold cross-validation; details are summarized in Table 4.

Table S2 MLP Parameters Table

Model	MLP
Activation	LeakyReLU
BatchNorm	BatchNorm1d
Weight Initialization	Xavier Normal
Structure Design	Input_layer=7
	Hidden_layer1=2048
	Hidden_layer2=1024
	Output_layer=1
Dropout	Hidden_layer1=0.3
	Hidden_layer2=0.2
Others	Optimizer:adam
	Loss Function:HuberLoss

The input dimension is 7, corresponding to the following meteorological features: air temperature (Ta), relative humidity (RH), atmospheric pressure (P), wind speed (WS), wind direction components (Wind_X, Wind_Y), and minute-level precipitation (minuteRain). To accelerate convergence, linear layers are initialized using Xavier normal distribution, with biases set to 0.01, ensuring stable gradients. The Huber loss ($\delta = 1.0$) is used for robustness to outliers, combining quadratic and linear penalties. Model optimization is performed with Adam (learning rate = 0.001). Training runs for 5000 epochs, with loss recorded every 100 epochs. Although the full dataset is used per iteration, the procedure follows mini-batch semantics.