

Article

Feature-Based Machine Learning Framework for Multi-Source NH₃ Dataset Analysis

Ata Jahangir Moshayedi ^{1,2,3} , Babar Hussain Shah ^{3,4} , Amir Sohail Khan ^{3,5} , Seyyed Ali Eftekhari ^{3,6} ,
Amin Kolahdooz ^{3,7*}  and David Bassir ^{1,8} 

- ¹ Smart Structural Health Monitoring and Control Laboratory, DGUT-CNAM, Dongguan University of Technology, D1, Daxue Rd., Songshan Lake, Dongguan 523808, China; aj.moshayedi@lecnam.net (A.J.M.); david.bassir@ens-paris-saclay.fr (D.B.)
 - ² DGUT-CNAM Institute, Dongguan University of Technology, D1, Daxue Rd., Songshan Lake, Dongguan 523808, China
 - ³ Robotics & Automation Resreach Lab (RARL), Leicester LE1 9BH, UK; babarshah0525@gmail.com or 2520250041@mail.jxust.edu.cn (B.H.S.); mrsohail21@gmail.com (A.S.K.); s.ali.ef@gnu.ac.kr (S.A.E.)
 - ⁴ School of Information Engineering, Jiangxi University of Science and Technology, Ganzhou 341000, China
 - ⁵ School of Computer Science and AI, Wuhan Textile University, Wuhan 430200, China
 - ⁶ Aerospace Systems Research Center Gyeongsang National University, Jinju 52828, Republic of Korea
 - ⁷ School of Engineering, Infrastructure and Sustainability, Faculty of Technology, Arts & Culture, De Montfort University, Leicester LE1 9BH, UK
 - ⁸ Centre Borelli, ENS-Paris-Saclay University, UMR CNRS 9010, 91190 Gif-sur-Yvette, France
- * Correspondence: amin.kolahdooz@dmu.ac.uk

Abstract

Accurate monitoring of atmospheric ammonia (NH₃) is important for air-quality assessment and sustainable agriculture, but available datasets differ in spatial resolution, temporal coverage, units, and physical meaning. This study presents a feature-based machine-learning framework with train–test leakage-controlled preprocessing to evaluate relative NH₃ classification consistency across three datasets over China: CAMS GEI, CAMS EAC4, and MEIC. Statistical descriptors were extracted for four spatial–temporal cases: Province–Year, Zone–Year, Province–Season–Year, and Zone–Season–Year. Six classifiers were evaluated using chronological testing and cross-validation. CAMS GEI provided the broadest spatial–temporal coverage, while MEIC showed comparatively stable classifier behavior. The best-performing ensemble models commonly achieved accuracies between 0.97 and 0.99 in the main province-level cases. Mean Absolute Value (MAV) was the leading feature in several province-level analyses, with a maximum reported contribution of 64.26%. These scores describe the separability of threshold-derived reference classes and should not be interpreted as an independent physical prediction of NH₃ from external atmospheric drivers.

Keywords: sustainable farming; ammonia (NH₃); atmospheric pollutant; MEIC model; machine learning framework



Academic Editor: Chenliang Liu

Received: 12 June 2026

Revised: 16 July 2026

Accepted: 4 August 2026

Published: 7 August 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Ammonia (NH₃) is an important atmospheric pollutant. It contributes to the formation of fine particulate matter (PM_{2.5}), which affects air quality and ecosystem health [1]. Agriculture is the main source of NH₃ emissions. The largest contributions usually come from livestock production and fertilizer use. Major emission hotspots have been reported in the USA, China, India, and Brazil [2]. However, NH₃ emissions are still difficult to estimate accurately. The main reasons are the non-linear atmospheric chemistry, the wide range

of emission sources, and the effect of weather and local environmental conditions. These factors also create uncertainty in bottom-up inventories and satellite-based estimates [3].

Data-driven methods provide another way to study NH_3 patterns. They can capture spatial and temporal changes that are difficult to describe with simple statistical models [4]. Combining emission inventories with atmospheric reanalysis products can also improve the study of large-scale pollution patterns. However, datasets may still disagree because they have different spatial resolutions, time periods, and modelling assumptions [5]. NH_3 also changes with season, location, farming activity, and industrial activity. This creates non-linear patterns in the data. For this reason, machine-learning methods have become useful for environmental prediction and monitoring [6–8]. They can help identify relationships in the data and support NH_3 assessment [9].

NH_3 information mainly comes from satellite observations, ground measurements, and atmospheric models. Satellite instruments such as the Infrared Atmospheric Sounding Interferometer (IASI), Cross-track Infrared Sounder (CrIS), and Tropospheric Emission Spectrometer (TES) provide repeated observations over large areas [10]. Ground stations provide accurate measurements at specific locations and often have high temporal resolution. Emission inventories and atmospheric models estimate NH_3 from agricultural, industrial, and urban sources. Satellite observations are especially useful for regional and global studies because they support long-term monitoring and the study of remote areas [11].

Several studies have evaluated atmospheric reanalysis and satellite products. Wang et al. (2020) assessed the Copernicus Atmosphere Monitoring Service (CAMS) global atmospheric trace-gas reanalysis using aircraft observations. They reported good agreement for ozone (O_3 , $R^2 = 0.93$) and carbon monoxide (CO), but lower performance for some short-lived species [12]. Wagner et al. (2021) studied the temporal consistency of CAMS global reanalysis using observations from 2003–2018. CAMS represented global O_3 and CO patterns well, but short-lived pollutants such as NO_2 remained more difficult to model [13]. These studies show that reanalysis products are useful, but they also have clear limits.

Lei et al. (2021) combined IASI observations, high-resolution Google Earth images, and the YOLOv4 model to detect NH_3 sources in Northeast China. The detected sources included agricultural, industrial, and energy-related sites. The study also showed that the final accuracy depends strongly on the quality of the labelled training data [14]. Luo et al. (2022) estimated global NH_3 emissions using IASI data and GEOS-Chem simulations. Their estimates were higher than traditional bottom-up inventories, and they reported large uncertainties in regions such as East China and India [11]. Chen et al. (2023) compared satellite-based inversion results to bottom-up inventories in China. They found that NH_3 emissions increased after 1980 and later became more stable or declined slightly as agricultural practices improved [15].

Ding et al. (2024) estimated NH_3 emissions over Europe using CrIS and TROPOMI observations with the DECSO inversion method. Their results agreed well with bottom-up inventories, with $R^2 > 0.95$. However, cloud effects, retrieval limits, and temporal sampling remained important sources of uncertainty [16]. Johansson et al. (2024) evaluated CAMS NH_3 in the upper troposphere and lower stratosphere using GLORIA airborne measurements. CAMS reproduced the large-scale Asian monsoon transport, but it underestimated NH_3 in biomass-burning plumes [17]. Sun et al. (2025) combined CrIS measurements with ground observations from the NNDMN network. They reported an increase in NH_3 concentrations in China between 2013 and 2023, mainly in eastern regions with strong agricultural and urban activity [18].

Recent studies have also used updated inversion methods and machine learning. Kumar et al. (2025) developed a global NH_3 emission estimate using updated emission factors and inversion methods. Their results agreed better with observations, but separating biomass burning, agriculture, and other sources remained difficult [19]. Moshayedi et al. (2025) studied agricultural NH_3 emissions in China using CAMS data and machine-learning models, including Support Vector Machine (SVM), XGBoost, and LightGBM. XGBoost gave the highest accuracy and identified major emission hotspots in northern and eastern China. The study also showed that model results were sensitive to normalization and lacked independent threshold-based validation [20].

Other environmental studies also show the value of machine learning for complex spatial data. Wan et al. (2025) used deep learning for nighttime urban flood detection and showed that neural networks can extract useful patterns from difficult image data [21]. Tang et al. (2025) studied the role of NH_3 emissions in $\text{PM}_{2.5}$ formation in eastern China. Their optimized NH_3 emissions were higher than the original MEIC values, but uncertainty remained in the vertical distribution of NH_3 and in satellite retrievals [22]. Choi et al. (2025) used IASI observations to study regional and seasonal NH_3 patterns over East Asia. They confirmed the effects of fertilizer use and temperature, but also reported uncertainty in mountainous areas and in separating natural and human-related sources [23]. Sheng et al. (2025) studied daytime and nighttime NH_3 changes and found good agreement between satellite and ground-based FTIR measurements. However, GEOS-CF still showed large bias for short-term changes [24]. Zhang et al. (2026) developed high-resolution livestock emission maps for China, but their results still depended on census quality and statistical assumptions [25]. Cao et al. (2026) developed a dynamic NH_3 emission and nitrogen-deposition model using WRF and Noah-MP-CN. Their model captured seasonal changes better than static inventories [26].

These studies show clear progress in NH_3 monitoring and emission estimation. However, several problems remain. Many studies focus on one dataset, one region, or one type of analysis. Differences in spatial resolution, time coverage, retrieval quality, and model assumptions also make dataset comparison difficult. Few studies compare several machine-learning models across the same province, zone, and seasonal structure. The role of simple statistical features has also not been studied in the same way across different datasets and classifiers.

This study addresses these points by comparing three NH_3 datasets over China. The analysis uses several machine-learning classifiers and four spatial–temporal cases. The aim is not to make a direct physical comparison between datasets with different units. Instead, the study compares their relative classification behavior after dataset-specific processing. It also examines which statistical features are most useful for separating Low, Medium, and High NH_3 reference classes.

To address these gaps, this study uses a feature-based machine-learning framework with train–test leakage-controlled preprocessing. Three datasets are examined: CAMS Global Emission Inventories (GEI), CAMS Global Reanalysis EAC4, and the Multi-resolution Emission Inventory for China (MEIC). The main objectives are to (1) compare six machine-learning classifiers at different spatial and temporal scales; (2) compare the relative NH_3 class patterns obtained from the three datasets; (3) study the effects of dataset properties and classifier behavior using feature contribution and Leave-One-Out Cross-Validation (LOO-CV); and (4) identify the most useful statistical features for NH_3 class separation.

The reviewed studies show that satellite observations, atmospheric models, inversion methods, and machine-learning tools provide different but useful views of NH_3 sources and trends. They also show common limits, including retrieval bias, model assumptions, spatial resolution, temporal coverage, and limited ground validation. The present study

focuses on a narrower question. It compares the classification behavior of three NH_3 datasets using the same province, zone, and seasonal case structure, and it examines how simple statistical features affect the results.

The remainder of this paper is organized as follows. Section 2 describes the datasets, machine-learning models, evaluation metrics, feature extraction, and validation procedure. Section 3 presents the experimental results, spatial and seasonal patterns, and feature-contribution analysis. Section 4 summarizes the main findings, limitations, and future research directions.

2. Materials and Methods

2.1. Data Sources and Processing

This study utilizes three NH_3 datasets: the Copernicus Atmosphere Monitoring Service Global Emission Inventories (CAMS GEI) as Dataset 1 (D1), the CAMS Global Reanalysis EAC4 (CAMS EAC4) as Dataset 2 (D2), and the Multi-resolution Emission Inventory for China (MEIC) as Dataset 3 (D3). These datasets differ in physical meaning, unit, spatial resolution, and temporal coverage. D1 and D3 represent emission inventory products, whereas D2 represents atmospheric concentration/mixing-ratio information from reanalysis. Therefore, the datasets were not treated as directly interchangeable physical quantities. Instead, each dataset was analyzed within its own normalized statistical distribution, and the comparison was performed at the relative classification-consistency level.

The full available temporal span of each dataset was used to preserve the maximum sample size and information content for classifier benchmarking. Harmonizing all datasets to a common temporal window would have substantially reduced the number of samples, especially for D1, and weakened the evaluation of long-term and seasonal patterns. This choice preserves dataset-specific statistical power, but it also introduces temporal heterogeneity; therefore, direct point-to-point physical comparison among datasets was not attempted. A summary of the datasets is provided in Table 1.

Table 1. Dataset characteristics and processing summary. P denotes the number of provinces or administrative units, and DP denotes the number of data points.

Dataset	P	Period	DP	Coverage	Data Type	Unit	Vertical Level	Native Resolution and Processing
D1	35	2015–2026	420	All of China	Emission flux	$\text{kg m}^{-2} \text{s}^{-1}$	Surface	Monthly; gridded values aggregated to province/zone level
D2	34	2019–2025	238	Excludes Macau	Concentration/mixing ratio	kg kg^{-1}	60 model levels	3-hourly native data aggregated to monthly province/zone values
D3	31	2017–2023	217	Excludes Macau, HK, Taiwan, Zhoushan	Emission inventory	Inventory-based NH_3 emissions	Surface	Monthly/yearly provincial inventory data

(1) Dataset 1: CAMS Global Emission Inventories (CAMS GEI)

Dataset 1 provides gridded NH_3 emission fluxes at the surface level in $\text{kg m}^{-2} \text{s}^{-1}$, with a spatial resolution of $0.1^\circ \times 0.1^\circ$ and monthly temporal resolution. The data for China cover 2015–2026. The NetCDF files were processed in Python V5, and the China vector boundary file was used to extract and aggregate the gridded NH_3 values at province and zone levels [27].

(2) Dataset 2: CAMS Global Reanalysis EAC4 (CAMS EAC4) Dataset 2 provides atmospheric NH_3 concentration/mixing-ratio information in kg kg^{-1} , with a native spatial resolution of $0.75^\circ \times 0.75^\circ$, 60 model levels, and 3-hourly temporal resolution. The data for China cover 2019–2025. The NetCDF files were processed in Python and aggregated to monthly province- and zone-level values for the case analyses [28].

(3) Dataset 3: Multi-resolution Emission Inventory for China (MEIC) Dataset 3 provides bottom-up anthropogenic NH_3 emission-inventory information at the provincial level. The data cover 2017–2023 and include monthly or yearly sectoral emission information. Because MEIC is already available at province level, no gridded-to-provincial spatial aggregation was required [29].

Because the datasets represent different physical quantities and units, direct unit-to-unit conversion was not performed. Each dataset was processed within its own statistical distribution using dataset-specific normalization and threshold-derived reference classes. The comparison therefore concerns relative within-dataset classification consistency, not direct physical equivalence among the three products.

2.2. Used Classifier

To analyze the data, a classification-based analysis was performed to examine NH_3 levels across all three datasets. Six widely used classifiers were employed, namely Support Vector Machine (SVM), Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), Long Short-Term Memory Network (LSTM), Random Forest (RF), and Geographically Weighted Random Forest (GWRF) [10] as shown in Figure 1. These models were selected to represent complementary learning mechanisms: SVM as a margin-based classifier, RF as a bagging ensemble, XGBoost and LightGBM as boosting-based tree models, LSTM as a neural-network benchmark, and GWRF as a geographically augmented Random-Forest implementation.

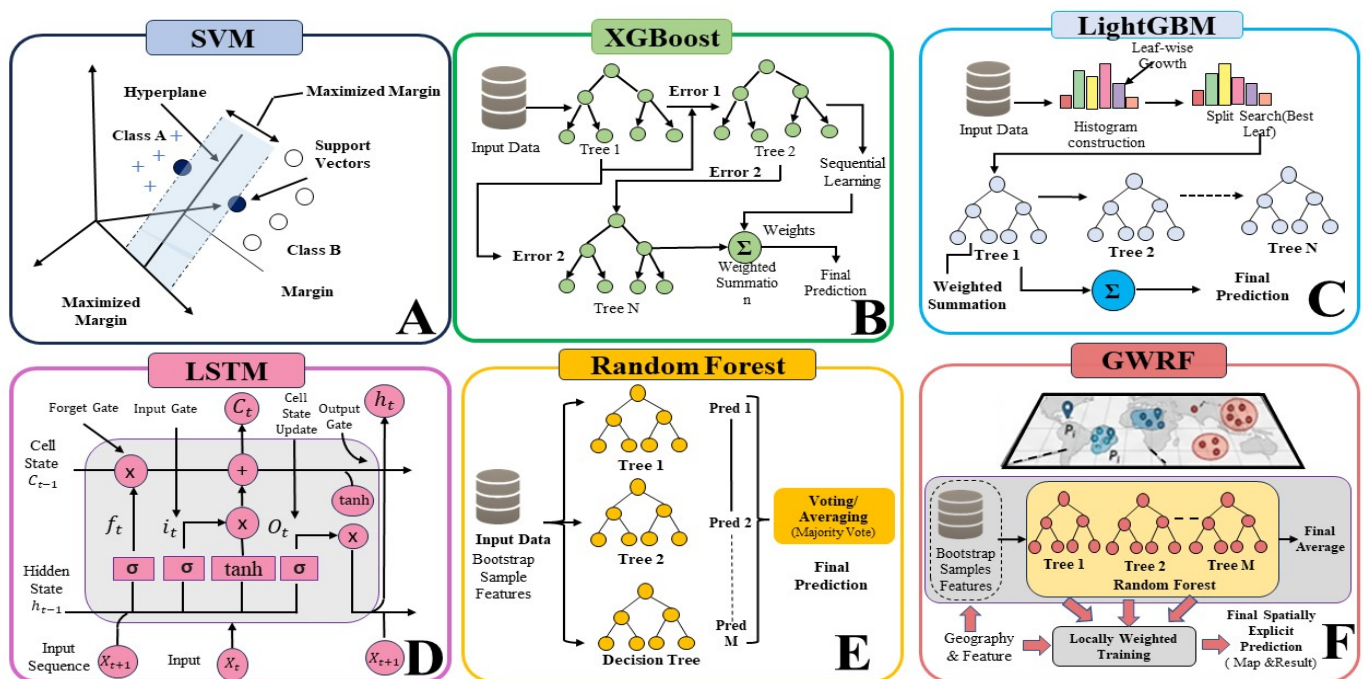


Figure 1. Classifier structure: (A) SVM, (B) XGBoost, (C) LightGBM, (D) LSTM, (E) Random Forest, (F) GWRF.

- **Support Vector Machine (SVM)** separates classes by finding a decision boundary with the largest possible margin. It can also model non-linear class boundaries through kernel functions. In this study, the tested kernels were radial basis function (RBF), linear, polynomial, and sigmoid. The regularization parameter C was searched from 0.01 to 50, while γ was set to either `scale` or `auto` [20,30].
- **Light Gradient Boosting Machine (LightGBM)** is an efficient gradient boosting framework that grows trees leaf-wise and uses sampling and feature reduction to

handle large datasets quickly. Its hyperparameters, number of leaves (3, 4, 5, 6), number of trees (50, 100, 150, 200), subsample fraction (0.5, 0.6, 0.7, 0.8), and column sampling fraction (0.5, 0.6, 0.7, 0.8), are optimized to balance predictive accuracy and computation speed [31].

- **Extreme Gradient Boosting (XGBoost):** XGBoost is an improved version of gradient boosting decision trees, excelling at incremental error correction and regularisation, that aims to improve predictive performance. It works very well in regressing as well as in classification. The number of estimators (20–200), maximum depth (1–4), gamma (0–2) and subsample/colsample_bytree fractions (0.4–0.8) are tuned to boost accuracy without causing overfitting of the model [32].
- **Long Short-Term Memory (LSTM):** LSTM was included as a neural-network benchmark for comparison with classical and ensemble classifiers. The input consisted of aggregated statistical features rather than a continuous long-sequence time series for each province or zone. Therefore, the feature vector was treated as a single-step sequence, and the LSTM results were not interpreted as evidence of long-term temporal-dependency learning. The tested settings included 16, 32, or 64 LSTM units, regularization, dropout, batch size, epochs, and early stopping [33].
- **Random Forest (RF)** is an ensemble method that uses multiple trees to boost performance of the decision-making process with the aim of reducing overfitting. It is resistant to noise, and can work with high dimensional data. The number of trees, the maximum depth and the types of fractions of the subsample are carefully optimized in this study in order to achieve a better classification performance (50, 100, 150, 200 trees, 2–6 maximum depth and subsample fraction from 0.4 to 0.8) [34].
- **Geographically augmented Random Forest (GWRF):** In the present implementation, province or zone identifiers were encoded and combined with the numerical features used by a Random-Forest model. This allows the classifier to represent broad regional differences, but it is not a full local-kernel GWRF with explicit bandwidth optimization. The tuned settings included the number of trees, maximum features, and maximum samples. Because the same geographic units can appear in training and testing years, spatial dependence may remain; therefore, the results do not demonstrate transfer to unseen regions [35].

As shown in Table 2, to further enhance model performance, all classifiers were adaptively scaled according to the dataset size. The hyperparameter ranges were kept moderate because several zone-level and seasonal cases contain only a small number of samples. Wider ranges could increase the risk of overfitting. Optuna was then used to select the best setting for each dataset and case from these predefined ranges. For SVM, smaller datasets (≤ 50 rows) used simpler kernels (linear or sigmoid) with lower regularization values ($C = 0.01$ – 10), while larger datasets (> 50 rows) employed the more flexible RBF kernel with $C = 10$ – 50 and $\gamma = 0.5$. Similarly, XGBoost and LightGBM adjusted the number of estimators and tree depth based on dataset size, where smaller datasets used fewer trees and shallower depths, while larger datasets utilized more trees and deeper structures.

To stabilize the learning process irrespective of the size of datasets, more regularization techniques were used for the LSTM such as Dropout (0.3) and Early Stopping to restore the best weights learned for saving them from potential overfitting. Random Forest (RF) had to be trained with balanced weight across the classes and according to the size of datasets, the Gini or entropy splitting criterion was used to train the RF model. Fixed random states were used where applicable for reproducibility. GWRF, on the other hand, integrated spatial data by using the concept of encoded geographic features (Province/Zone), and in the case of the larger datasets, hyperparameters key to the optimal spatial predictive power were tuned using Optuna. Finally, the performance of all models was validated by the proper

cross validation techniques, namely the Leave-One-Out (LOO), five-fold and 10-fold cross validation for small, medium and larger datasets, respectively, to guarantee the robustness and reliability of the results.

Table 2. Hyperparameter search space and model settings for SVM, XGBoost, LightGBM, LSTM, Random Forest, and the geographically augmented Random-Forest implementation.

Parameter	SVM	XGBoost	LightGBM	LSTM	RF	GWRF
Objective	Classification (SVC)	multi:softprob	multiclass	Sparse Categorical Crossentropy	—	Multiclass
Number of Classes	3(L,M,H)	3(L,M,H)	3(L,M,H)	3(L,M,H)	3(L,M,H)	3(L,M,H)
Kernel/Booster	rbf, linear, poly, sigmoid	gbtree booster	gbdt	LSTM (Sequential Model)	—	—
C/Regularization	C = trial.suggest_float(0.01, 50)	reg_alpha = log(0.1, 5.0)	reg_alpha = 1.0, reg_lambda = 2.0	kernel_regularizer = l2(0.01)	—	—
Gamma/Learning n_estimators/Trees	gamma = scale, auto	gamma = 0.0, 2.0	—	—	—	—
max_depth/Leaves/units	—	trial.suggest_int(20, 200)	50, 100, 150, 200	—	50, 100, 150, 200	150, 200, 250, 300
Subsample/batchsize	—	trial.suggest_int(1, 4)	3, 4, 5, 6	16, 32, 64	2, 3, 4, 5, 6	8, 10, 12, 15
colsample_bytree/epochs	—	trial.suggest_float(0.4, 0.8)	0.5, 0.6, 0.7, 0.8	—	0.4, 0.5, 0.6, 0.7, 0.8	max_samples (0.8)
Probability/Output	probability = True	multi:softprob	—	Softmax (activation)	Balanced Sub-sample	max_features (0.5–0.8)
Random State	42	42	123	42	42	777
Feature Scaling	StandardScaler	—	—	StandardScaler	—	—
Features Used	Encoded cat + num	Encoded cat + num	Encoded cat + num	Encoded cat + num	Encoded cat + num	Geog + num
n_jobs/CPU Cores	Auto-detect	Auto-detect	Auto-detect	—	Auto-detect	Auto-detect

2.3. Evaluation Metrics

Five metrics (as shown in Table 3) were used consistently in the reported results: Accuracy, macro-Precision, macro-Recall, macro-F1-score, and macro-AUC. Accuracy measures overall correctness, while the macro-averaged metrics give equal weight to each class and are therefore more informative when class frequencies differ. AUC was reported only when the required classes were present in the test labels and predictions.

Table 3. Evaluation metrics used for model assessment. TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively.

Metric	Formula	Description
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$	Overall proportion of correctly classified samples.
Precision	$\frac{TP}{TP+FP}$	Proportion of predicted positive samples that are correct.
Recall	$\frac{TP}{TP+FN}$	Proportion of actual positive samples that are correctly identified.
F1-score	$2 \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$	Harmonic mean of precision and recall.
AUC	—	Area under the one-vs-rest receiver operating characteristic curve; reported only when mathematically defined.

2.4. Features Extraction

For each province or zone, the statistical features were computed from the temporal sequence of monthly NH₃ values. Each spatial unit was represented by a fixed-length vector describing signal magnitude, variability, distributional shape, trend, and local extremes. The features were Mean Absolute Value (MAV), Root Mean Square (RMS), Standard Deviation (Std), Interquartile Range (IQR), Kurtosis, Slope, Skewness, and maximum Z-score, as listed in Table 4. Z-score max was calculated relative to the mean and standard deviation of the corresponding province or zone time series and represents the largest local standardized anomaly. Because the low, medium, and high labels were defined from the NH₃ distribution, this task measures how well the reference classes can be separated. It

should not be treated as an independent physical prediction problem. High accuracy was therefore interpreted cautiously, especially for magnitude-based features such as MAV and RMS.

Table 4. Features and their respective mathematical formulas. In all features, N is the total number of data points and X_i is each data point in the dataset. In Std, Kurtosis, and Skewness, μ is the Mean of the dataset. In IQR, Q_1 is the First Quartile and Q_3 is the Third Quartile, representing IQR as 50% of data between 25% and 75%. In Kurtosis and Skewness, s represents the Standard Deviation of the dataset. In Slope, X_i and Y_i are the data points on the x-axis and y-axis.

Feature Name	Formula
Mean Absolute Value (MAV)	$MAV = \frac{1}{N} \sum_{i=1}^N x_i $
Root Mean Square (RMS)	$RMS = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2}$
Standard Deviation (Std)	$\sigma = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2}$
Interquartile Range (IQR)	$IQR = Q_3 - Q_1$
Kurtosis	$Kurt = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^4}{(\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2)^2}$
Slope	$m = \frac{n \sum(xy) - \sum x \sum y}{n \sum(x^2) - (\sum x)^2}$
Skewness	$Skew = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^3}{(\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2)^{3/2}}$
Z_Score_Max	$Z_{max} = \max\left(\frac{x_i - \mu}{\sigma}\right)$

2.5. Leave-One-Out Cross-Validation (LOO-CV)

To refine feature selection, LOO-CV was employed in which one observation is held out as the test set while the remaining $n - 1$ samples form the training set. This process is repeated for all n observations, and overall performance is averaged to provide a reliable estimate of the model's accuracy [36]. In feature selection, LOO-CV assesses the contribution of each feature by iteratively removing one feature at a time, retraining the model, and observing the change in performance. Features are considered important when their removal causes a clear drop in accuracy. Features with little effect on accuracy are treated as less important. LOO-CV evaluates each sample once as the test case and uses the remaining samples for training. This gives a detailed estimate of model performance, especially for small datasets. It helps identify the most impactful features by measuring performance changes when each feature is removed. LOO is especially useful for small or imbalanced datasets, ensuring reliable assessment and model generalization [37]. It also enables fair comparison across algorithms while improving confidence in generalization, feature importance, and dataset reliability [38].

3. Experimental Test and Results

The workflow of the study is shown in Figure 2. As shown in the figure, the extracted data from all datasets were organized into four main divisions based on province, zone, year, and season [38].

- **Case 01: Province–Year**—Data divided by individual provinces in China.
- **Case 02: Zone–Year**—Data grouped by regions (North, Central, South).
- **Case 03: Province–Season–Year**—Data divided by provinces and seasons (Spring, Summer, Autumn, Winter).
- **Case 04: Zone–Season–Year**—Data divided by zones and seasons.

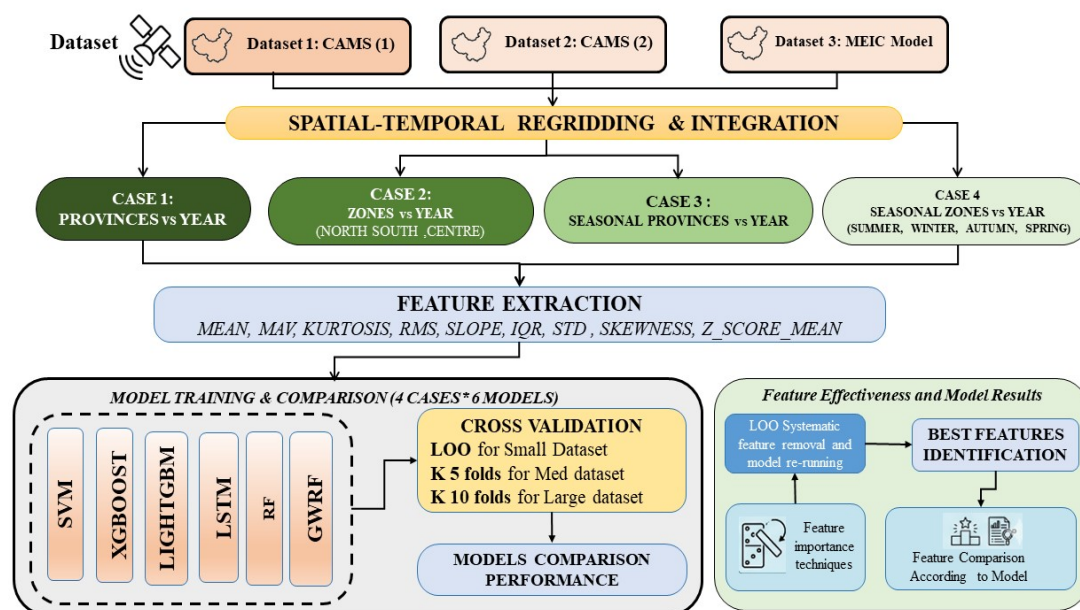


Figure 2. Study workflow for dataset preparation, case construction, feature extraction, model evaluation, and feature-contribution analysis.

As shown in Figure 3, the investigated provinces are used as reference regions. In all cases, the labels presented in the figure correspond to the provinces analyzed in this study. Seasons were categorized as Spring (March, April, and May), Summer (June, July and August), Autumn (September, October, and November), and Winter (December, January, and February). The monthly NH_3 values were averaged for each season to evaluate the general trend of NH_3 values throughout the season. This helps smooth out the month-to-month variability and makes seasonal variability more apparent. The comparison was performed at the classification-consistency level after aggregation and class labeling, rather than as a direct physical unit-to-unit comparison.

A tabular summary of all datasets showing detailed data used in the four cases is given in Table 5. Each dataset is described by the total number of samples, the number of classes, the number of samples in each class (low, medium, high), and the number of years covered; for all datasets, the number of features was eight, which did not vary. Case 3 is a spring, summer, autumn, winter season breakdown of the cases at the provincial level and Case 4 provides a breakdown of the cases by season at the zonal level. This case has limited sample size and reduced seasonal variability, so its results should be interpreted as exploratory. This structure enables readers to view temporal and spatial distribution of datasets, and to see the class balance. The table also shows that sample size and class balance vary substantially across datasets and cases. Province-level cases contain more observations, whereas several zone-level and seasonal subsets are small or imbalanced. Some zone-level and seasonal subsets, particularly Case 4, contain limited and imbalanced samples. Therefore, their results were interpreted cautiously and considered exploratory. To reduce bias from class imbalance, macro-averaged precision, recall, F1-score, and confusion matrices were reported together with accuracy.

As mentioned from these partitions, eight statistical features (Table 4) were extracted and used to train six machine learning models. LOO was applied for small datasets, while five-fold and 10-fold cross-validation were used for medium and large datasets, respectively, to prevent overfitting and allow model comparison. Based on the best-performing dataset, LOO was further applied to evaluate the effectiveness of each feature and report its

contribution percentage. Finally, the prediction accuracy of all models was computed and compared across all cases. The classifier codes (as its shown in Appendix A Algorithms A1 and A2) were executed in Google Colab. This Turing-based GPU is equipped with 16 GB of GDDR6 memory and 2560 CUDA cores, providing optimized performance for AI training and inference. However, RF and GWRF are classical classifiers and do not inherently benefit from GPU acceleration. These are typically implemented using CPU-based operations.

Table 5. Overview of the datasets used in the four experimental cases. **D** is the dataset identifier, **NS** is the number of samples, **NC** is the number of classes, **SPC** is the number of samples per class (Low, Medium, and High), and **NY** is the number of years represented. There are 3 classes and 8 input features to all datasets

Dataset	NS	NC	SPC	NY
Case 1				
D1	420	3	Low = 155, Medium = 165, High = 100	12
D2	238	3	Low = 76, Medium = 80, High = 61	7
D3	217	3	Low = 89, Medium = 87, High = 62	7
Case 2				
D1	36	3	Low = 12, Medium = 10, High = 14	12
D2	21	3	Low = 7, Medium = 7, High = 7	7
D3	21	3	Low = 13, Medium = 1, High = 7	7
Case 3				
D1 Spring	420	3	Low = 142, Medium = 146, High = 132	12
D1 Summer	420	3	Low = 163, Medium = 145, High = 112	12
D1 Autumn	420	3	Low = 151, Medium = 159, High = 110	12
D1 Winter	420	3	Low = 140, Medium = 192, High = 88	12
D2 Spring	238	3	Low = 84, Medium = 98, High = 56	7
D2 Summer	238	3	Low = 66, Medium = 106, High = 66	7
D2 Autumn	238	3	Low = 76, Medium = 68, High = 60	7
D2 Winter	238	3	Low = 96, Medium = 89, High = 53	7
D3 Spring	217	3	Low = 74, Medium = 88, High = 62	7
D3 Summer	217	3	Low = 81, Medium = 76, High = 67	7
D3 Autumn	217	3	Low = 76, Medium = 85, High = 63	7
D3 Winter	217	3	Low = 75, Medium = 87, High = 62	7
Case 4				
D1 Spring	36	3	Low = 12, Medium = 10, High = 14	12
D1 Summer	36	3	Low = 12, Medium = 9, High = 15	12
D1 Autumn	36	3	Low = 12, Medium = 11, High = 13	12
D1 Winter	36	3	Low = 12, Medium = 12, High = 12	12
D2 Spring	21	3	Low = 7, Medium = 7, High = 7	7
D2 Summer	21	3	Low = 7, Medium = 3, High = 11	7
D2 Autumn	21	3	Low = 7, Medium = 7, High = 7	7
D2 Winter	21	3	Low = 14, Medium = 0, High = 7	7
D3 Spring	21	3	Low = 13, Medium = 1, High = 7	7
D3 Summer	21	3	Low = 13, Medium = 1, High = 7	7
D3 Autumn	21	3	Low = 13, Medium = 1, High = 7	7
D3 Winter	21	3	Low = 13, Medium = 1, High = 7	7

As mentioned, this study aims to investigate two approaches, classifier-based analysis on three different datasets over four cases, and analysis based on the contribution of each feature in four cases and six classifiers. During the evaluation, a train–test leakage-controlled preprocessing procedure was adopted by restricting threshold calculation, scaling, and normalization to the training data. The complete dataset was first partitioned into training and testing subsets based on chronological order. Statistical measures required for target classification, including the mean and standard deviation, were calculated solely

from the training subset. The corresponding thresholds defining the low, medium, and high classes were subsequently transferred to the test subset without modification. In the same manner, feature scaling and normalization parameters were learned from the training data and then applied to the test data. Consequently, no information from the test subset was used during model training, preprocessing, or threshold determination. This procedure controls direct train–test preprocessing leakage; however, it does not eliminate all possible dependence caused by threshold-derived labels or repeated geographic units across years. Therefore, the reported results are interpreted as classification consistency within the available data structure rather than unrestricted spatial generalization. The chronological split evaluates temporal generalization within the available province/zone structure. Although chronological splitting was used to reduce direct temporal leakage, spatial and temporal autocorrelation may still remain because neighboring regions and adjacent years can have similar NH_3 patterns. Therefore, the reported results should be interpreted as performance under a chronological validation setting. Future work should include spatial-block and temporal-block cross-validation to provide a stricter assessment of model generalization. However, it does not fully assess transferability to unseen geographic units. Therefore, spatial-block and leave-province-out validation are recommended as additional tests in future work to evaluate spatial generalization and reduce the effect of spatial autocorrelation. Geographic identifiers were used only to represent spatial grouping effects within the available province/zone structure. Therefore, the reported results mainly reflect temporal generalization within known geographic units, not transferability to unseen provinces or zones. Future work should include leave-province-out or spatial-block validation to evaluate spatial generalization more strictly.

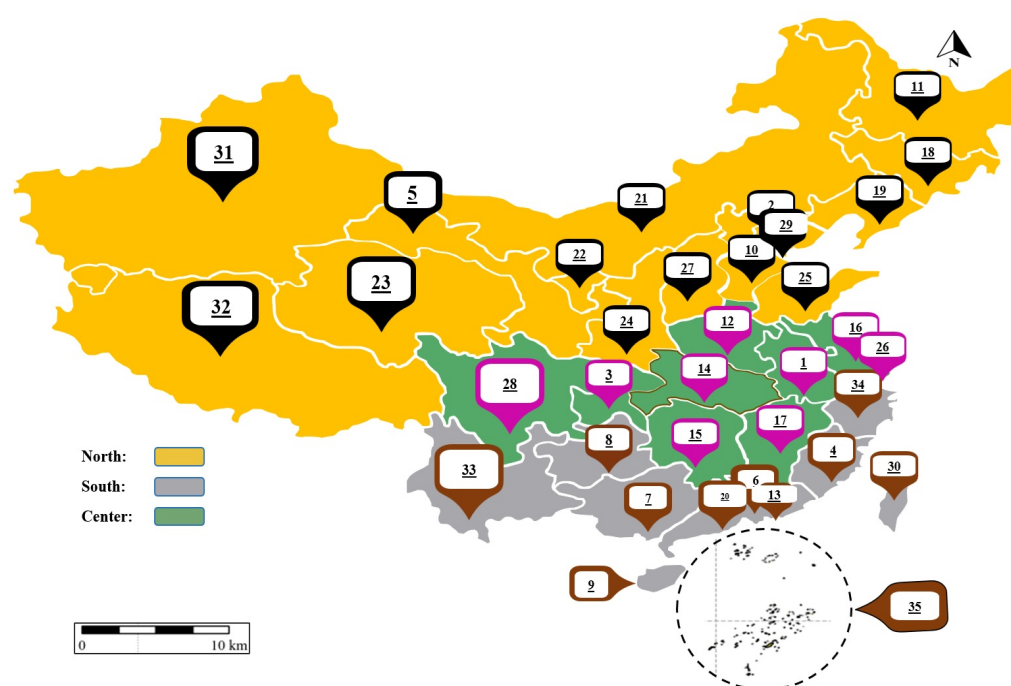


Figure 3. China province indexed by ID: 1-Anhui, 2-Beijing, 3-Chongqing, 4-Fujian, 5-Gansu, 6-Guangdong, 7-Guangxi, 8-Guizhou, 9-Hainan, 10-Hebei, 11-Heilongjiang, 12-Henan, 13-Hong Kong, 14-Hubei, 15-Hunan, 16-Jiangsu, 17-Jiangxi, 18-Jilin, 19-Liaoning, 20-Macau, 21-Inner Mongolia, 22-Ningxia Hui, 23-Qinghai, 24-Shaanxi, 25-Shandong, 26-Shanghai, 27-Shanxi, 28-Sichuan, 29-Tianjin, 30-Taiwan, 31-Xinjiang, 32-Xizang, 33-Yunnan, 34-Zhejiang. 35-the group of small islands is called Xisha Qundao and Nansha Qundao.

3.1. Classifier Performance and Result

The performance of all six classifiers was evaluated across the three datasets in four distinct cases, as summarized in Table 6 and illustrated below for each case. Simple majority-class, year-only, and province/zone-ID-only baselines were not available in the supplied result set. Their absence limits the interpretation of the high classification scores, and these comparisons should be included in future validation to separate feature-based performance from class imbalance, time trends, and repeated geographic identifiers. For several models in Case 4, AUC-ROC could not be computed because seasonal aggregation and the small zone-level test set produced single-class or highly imbalanced predictions. Case 4 is therefore treated as an exploratory zone–season analysis, and its results are interpreted using macro-averaged metrics and the regional prediction patterns rather than accuracy alone.

Table 6. The six-classifier performance across four cases (C1–C4) over the evaluation metrics: 1: Accuracy, 2: Precision, 3: Recall, 4: F1-Score and 5: AUC. Models: M1 (SVM), M2 (XGBoost), M3 (LightGBM), M4 (LSTM), M5 (Random Forest), and M6 (GWRF). Datasets: D1, D2, and D3.

CASE 01		1	2	3	4	5	CASE 2		1	2	3	4	5
D1	M1	0.53	0.53	0.5	0.5	0.68	D1	M1	0.47	0.79	0.47	0.41	0.55
	M2	0.99	0.99	0.99	0.99	0.99		M2	0.67	0.5	0.67	0.56	0.99
	M3	0.99	0.99	0.99	0.99	0.99		M3	0.99	0.99	0.99	0.99	0.99
	M4	0.55	0.37	0.47	0.4	0.69		M4	0.67	0.5	0.67	0.56	0.89
	M5	0.97	0.97	0.97	0.97	0.99		M5	0.99	0.99	0.99	0.99	0.99
	M6	0.99	0.99	0.99	0.99	0.99		M6	0.99	0.99	0.99	0.99	0.99
D2	M1	0.5	0.5	0.5	0.5	0.63	D2	M1	0.99	0.99	0.99	0.99	0.99
	M2	0.89	0.9	0.89	0.89	0.93		M2	0.99	0.99	0.99	0.99	0.99
	M3	0.89	0.9	0.89	0.89	0.93		M3	0.99	0.99	0.99	0.99	0.99
	M4	0.52	0.36	0.5	0.42	0.7		M4	0.33	0.99	0.33	0.99	0.83
	M5	0.87	0.89	0.86	0.86	0.92		M5	0.99	0.99	0.99	0.99	0.99
	M6	0.89	0.9	0.89	0.89	0.93		M6	0.99	0.99	0.99	0.99	0.99
D3	M1	0.53	0.54	0.54	0.53	0.72	D3	M1	0.83	0.67	0.5	0.56	NA
	M2	0.99	0.99	0.99	0.99	0.99		M2	0.99	0.99	0.99	0.99	0.99
	M3	0.99	0.99	0.99	0.99	0.99		M3	0.99	0.99	0.99	0.99	0.99
	M4	0.5	0.5	0.51	0.5	0.73		M4	0.67	0.33	0.5	0.4	NA
	M5	0.96	0.97	0.97	0.96	0.99		M5	0.93	0.99	0.99	0.99	0.99
	M6	0.99	0.99	0.99	0.99	0.99		M6	0.99	0.99	0.99	0.99	0.99
CASE 3		1	2	3	4	5	CASE 4		1	2	3	4	5
D1	M1	0.42	0.39	0.44	0.36	0.63	D1	M1	0.75	0.73	0.7	0.65	0.82
	M2	0.99	0.99	0.99	0.99	0.99		M2	0.88	0.78	0.83	0.8	0.95
	M3	0.99	0.99	0.99	0.99	0.99		M3	0.97	0.9	0.92	0.91	0.99
	M4	0.47	0.31	0.41	0.35	0.66		M4	0.82	0.74	0.77	0.74	0.85
	M5	0.97	0.97	0.97	0.97	0.99		M5	0.97	0.90	0.92	0.91	0.99
	M6	0.99	0.99	0.99	0.99	0.99		M6	0.97	0.9	0.92	0.91	0.99
D2	M1	0.39	0.32	0.39	0.26	0.51	D2	M1	0.96	0.9	0.92	0.9	0.89
	M2	0.94	0.95	0.94	0.94	0.98		M2	0.88	0.73	0.79	0.75	0.95
	M3	0.94	0.94	0.94	0.93	0.98		M3	0.96	0.9	0.92	0.9	0.95
	M4	0.54	0.36	0.48	0.41	0.69		M4	0.46	0.18	0.38	0.24	0.82
	M5	0.90	0.90	0.90	0.90	0.96		M5	0.96	0.9	0.92	0.9	0.96
	M6	0.94	0.94	0.94	0.94	0.97		M6	0.96	0.9	0.92	0.9	0.98
D3	M1	0.50	0.44	0.47	0.43	0.69	D3	M1	0.83	0.67	0.5	0.56	NA
	M2	0.94	0.94	0.94	0.94	0.98		M2	0.99	0.99	0.99	0.99	NA
	M3	0.94	0.94	0.94	0.94	0.99		M3	0.99	0.99	0.99	0.99	NA
	M4	0.48	0.39	0.45	0.41	0.73		M4	0.67	0.33	0.5	0.4	NA
	M5	0.95	0.96	0.95	0.95	0.98		M5	0.99	0.99	0.99	0.99	NA
	M6	0.98	0.99	0.98	0.98	0.99		M6	0.99	0.99	0.99	0.99	NA

3.1.1. Case 01: Province vs. Year

The performance of six classifiers under Case 01 was evaluated using Accuracy and AUC across three datasets (D1–D3). For Dataset 1, ensemble-based models (XGBoost, LightGBM, and GWRF) achieved the highest performance with accuracy of 0.99, AUC of 0.99, and balanced macro metrics (Precision = 0.99, Recall = 0.99, F1 = 0.99). Random Forest also performed strongly (accuracy = 0.97, AUC = 0.99). SVM and LSTM showed lower performance, with SVM achieving 0.53 accuracy and 0.50 F1-score, and LSTM achieving slightly higher accuracy of 0.55 but lower macro-precision (0.37) and F1-score (0.40). In Dataset 2, XGBoost, LightGBM, and GWRF again showed stable results with accuracy and F1-score of 0.89, and AUC of 0.93. Random Forest achieved slightly lower results (accuracy = 0.87, AUC = 0.92). SVM and LSTM exhibited moderate performance, with SVM achieving 0.50 accuracy and 0.50 F1-score, and LSTM achieving 0.52 accuracy and 0.42 F1-score. In the case of Dataset 3, ensemble methods consistently outperformed other models, with XGBoost, LightGBM, and GWRF achieving 0.99 accuracy, 0.99 F1-score, and AUC of 0.99. Random Forest also performed strongly (accuracy = 0.96, F1-score = 0.96, AUC = 0.99). SVM and LSTM showed moderate results, with SVM at 0.53 accuracy and LSTM at 0.50 accuracy.

The near-perfect values in Table 6 should be interpreted cautiously. The labels are threshold-derived; magnitude-related features are closely aligned with those thresholds, and the same geographic units occur across years. The results therefore indicate strong reference-class separability within the available data structure, not independent physical prediction or transfer to unseen regions.

3.1.2. Case 02: Zone vs. Year

For Dataset 1, LightGBM, Random Forest, and GWRF achieved the best performance with accuracy of 0.99 and AUC of 0.99, along with balanced macro-precision, recall, and F1-score of 0.99. XGBoost and LSTM showed moderate performance with an accuracy of 0.67 and F1-score of 0.56. SVM showed lower performance with an accuracy of 0.47 and F1-score of 0.41, despite a relatively higher macro-precision of 0.79. Across Dataset 2, most models including SVM, XGBoost, LightGBM, Random Forest, and GWRF achieved near-perfect performance with an accuracy of 0.99 and AUC of 0.99. In contrast, LSTM showed inconsistent behavior, with low accuracy (0.33) but high macro-precision and macro-F1-score values (0.99), indicating unstable classification performance. When considering Dataset 3, GWRF and ensemble-based models such as XGBoost and LightGBM achieved the best results with an accuracy of 0.99 and AUC of 0.99. Random Forest also performed strongly with an accuracy of 0.93 and AUC of 0.99. SVM showed moderate performance (accuracy = 0.83, F1-score = 0.56), while LSTM again produced lower results (accuracy = 0.67, F1-score = 0.40). Overall, ensemble learning methods consistently outperformed SVM and LSTM across all datasets, demonstrating higher stability and classification accuracy.

3.1.3. Case 03: Province–Season vs. Year

For Dataset 1, ensemble-based methods showed the strongest performance. XGBoost, LightGBM, and GWRF achieved the highest results, with an accuracy of 0.99 and AUC of 0.99, along with consistent macro-precision, recall, and F1-score of 0.99. Random Forest also performed strongly with an accuracy of 0.97 and AUC of 0.99. In contrast, SVM showed weak performance (accuracy = 0.42, F1-score = 0.36), while LSTM achieved low to moderate performance with an accuracy of 0.47 and AUC of 0.66. Concerning Dataset 2, SVM again produced low performance (accuracy = 0.39, AUC = 0.51). XGBoost, LightGBM, and GWRF achieved strong and stable results, with an accuracy ranging from 0.94 to 0.95

and AUC up to 0.98. Random Forest achieved slightly lower performance (accuracy = 0.90, AUC = 0.96). LSTM remained weaker compared to ensemble models, with an accuracy of 0.54 and F1-score of 0.41. Under Dataset 3, GWRF achieved the best overall performance with an accuracy of 0.98, macro-precision of 0.99, macro-recall of 0.98, macro-F1-score of 0.98, and AUC of 0.99. XGBoost and LightGBM also showed high and consistent results (accuracy = 0.94, AUC up to 0.99). Random Forest achieved slightly lower but stable performance (accuracy = 0.95, AUC = 0.98). SVM showed moderate performance (accuracy = 0.50), while LSTM again produced lower results (accuracy = 0.48, AUC = 0.73). Overall, ensemble learning methods consistently outperformed SVM and LSTM across all datasets, with GWRF, XGBoost, and LightGBM providing the most stable and highest classification performance.

3.1.4. Case 04: Zone–Season vs. Year

For Dataset 1, LightGBM, Random Forest, and GWRF gave the best results. Their accuracy was 0.97 and their AUC was 0.99. The corresponding macro-precision, macro-recall, and macro-F1-score were 0.90, 0.92, and 0.91. XGBoost reached 0.88 accuracy and 0.95 AUC. SVM gave 0.75 accuracy, while LSTM reached 0.82 accuracy and 0.85 AUC. For Dataset 2, SVM reached 0.96 accuracy, but its AUC was lower at 0.89. LightGBM, Random Forest, and GWRF also reached 0.96 accuracy, with AUC values up to 0.98. XGBoost gave 0.88 accuracy and 0.95 AUC. LSTM gave the weakest result, with a macro-F1-score of 0.24 despite an AUC of 0.82. For Dataset 3, XGBoost, LightGBM, Random Forest, and GWRF reached 0.99 for the main performance metrics. SVM gave 0.83 accuracy and a 0.56 F1-score. LSTM gave the lowest result, with 0.67 accuracy and a 0.40 F1-score. AUC values were not available for Dataset 3. Overall, ensemble methods consistently produced higher and more stable performance compared to SVM, XGBoost, and LSTM across all datasets. In Case 4, AUC-ROC was not reported for several models because the seasonal zone-level aggregation produced single-class or highly imbalanced predictions. This class-collapse behavior is a limitation of the small zone–season setting and does not support strong discrimination claims.

3.1.5. Comparative Analysis Across Cases and Datasets

Across the four cases, ensemble-based models consistently achieved the best overall performance across all datasets. In Dataset 1, ensemble methods remained the most stable and accurate across all cases. Random Forest showed slightly lower performance compared to other ensemble models, but still outperformed non-ensemble methods. SVM and LSTM consistently performed at a lower level, with noticeable variation across cases. For Dataset 2, ensemble models again demonstrated strong and stable performance across all cases. Random Forest followed a similar trend but remained slightly below the top-performing ensemble methods. SVM showed inconsistent behavior depending on the case, while LSTM produced unstable results with clear performance fluctuations. Under Dataset 3, ensemble methods maintained consistently high performance across all cases, with minimal variation between models. Random Forest showed competitive results but remained slightly lower than the best-performing ensembles. SVM achieved moderate performance, while LSTM remained the weakest model overall. Overall, ensemble-based methods consistently outperformed SVM and LSTM across all cases and datasets. Random Forest performed competitively but was generally below the top ensemble models. SVM and LSTM showed lower stability and higher sensitivity to dataset and case variations. As the results show, considering both data quality and model performance, Dataset 3 (MEIC v2.0) is the most suitable for analysis due to its high stability and consistent results across all classifiers. Although Dataset 3 contains fewer provinces and data points than Dataset 1,

its results are more consistent across the tested models. Dataset 1 (CAMS GEI) is still useful for national-scale studies and long-term trend analysis.

3.1.6. Dataset 1 Reference-Label-Based Performance Analysis over the Best Models

The reported results compare predicted class labels from each classifier with the corresponding threshold-derived reference labels in Dataset 1. The models evaluated include SVM, XGBoost, LightGBM, LSTM, Random Forest, and GWRF. Among them, XGBoost, LightGBM, and GWRF show the most consistent performance, correctly classifying almost all labels (H, M, and L) across most provinces and years, as summarized in Table 6 and Figures 4–6. Based on obtained data in Case 1, (XGBOOST, LightGBM, and GWRF) across Chinese provinces and municipalities from 2022 to 2026, the results are highly consistent across the three models, with no noticeable classification differences for any province or year. A clear spatial pattern can be observed. Major municipalities and several developed regions such as Beijing, Shanghai, Tianjin, Chongqing, Hebei, Henan, Shanxi, and Macau consistently show high performance throughout the entire period. By comparison, provinces such as Anhui, Guangxi, Guizhou, Hubei, Hunan, Jiangsu, Jiangxi, Jilin, Liaoning, Ningxia, Shaanxi, Shandong, Sichuan and Tibet are still at a medium level, having had no change over five years. The low-class regions are mainly in Fujian, Gansu, Guangdong, Hainan, Heilongjiang, Hong Kong, Inner Mongolia, Qinghai, Taiwan, Xinjiang, Yunnan, Zhejiang and Zhoushan Islands. For most of these areas, there is little change, only slight improvement in GWRF in Zhoushan Islands in the last two years (from low to medium level). Overall, the overall cross-temporal stability is good and the cross-time differences are primarily spatial. For Case 2, in the North zone, the trend is stable with XGBOOST always achieving a high performance and LIGHTGBM, RF, GWRF medium performance all year round. Within this zone, no temporal variation is noticed. All models in the South zone are low and for all years. The results are the same for all four methods showing consistently poor performance in this area. For the Central zone, the performance remains consistently at high level across all the models (XGBOOST, LIGHTGBM, RF and GWRF) and this has not varied in the past 5 years. This situation shows high spatial stability in the results. The North is mixed (high/medium), the South is persistently low and the Central zone is persistently high. Temporal variation is not noted and differences are primarily due to geographic zones and not to the model type or year. There are similar trends across the five years for each model in Case 3: The results remain stable over the five years. In most cases, these values are the same for LSTM, XGBOOST and LIGHTGBM and GWRF follows its values with only small differences seen seasonally. These values are often greater in spring and summer reaching H/M based on the model and year. In several instances, autumn and winter reached M or L. This trend is seen over all the years, with mostly high values in spring and summer as well as mostly low values in autumn and winter. Most of the values are medium with recurring medium values throughout all the seasons. Changes in model behavior over time are not significant. Differences are primarily associated with seasonal variability and not interannual or model variations. Case 4 denotes the performance of XGBOOST, LIGHTGBM and GWRF for the central, north and south zone from 2022 to 2026, where the central zone is unchanged across all the models and zone for all the seasons. A continuous consistency of high occurs across all four seasons (spring, summer, autumn, winter) for the period 2022–2026 across all models. A seasonal pattern is repeated in the North zone during the whole period of the present investigation—low for all models (spring), high for XGBOOST and medium for LIGHTGBM and GWRF (summer), medium for all models (autumn), low for all models (winter). This is the same throughout five years. The south zone also remains stable over time. Spring, summer, and autumn are consistently low for all models, while winter is consistently medium from 2022 to 2026. No changes

appear across years or between models. The results indicate clear spatial consistency. The central zone maintains consistently high values, the north zone shows mixed seasonal behavior depending on the model and season, and the south zone remains mostly low with a slight increase in winter.

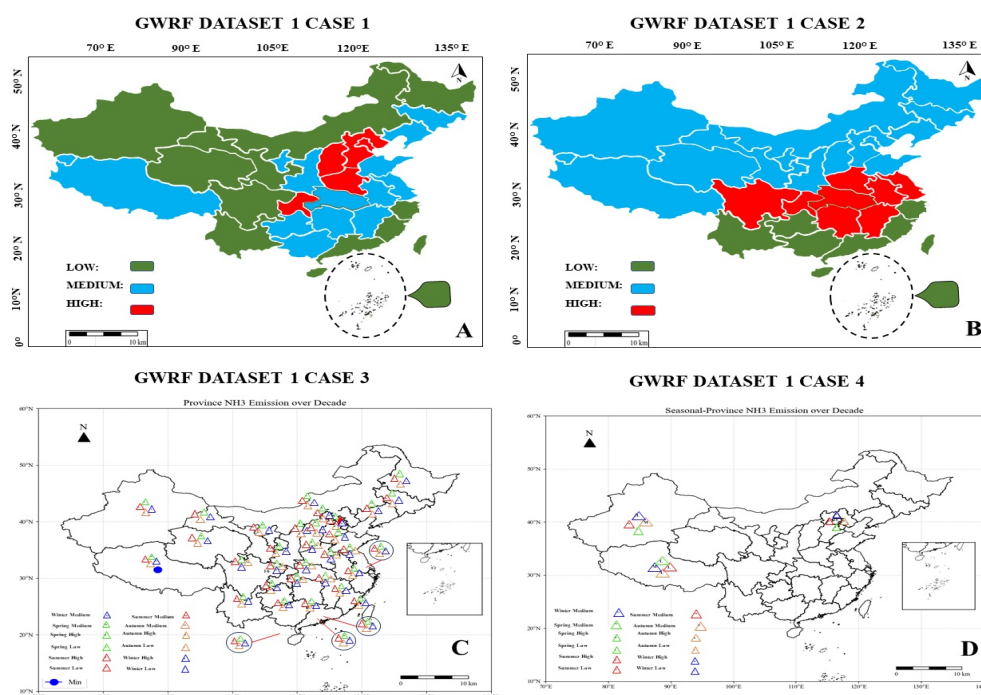


Figure 4. The GWR machine learning model over four cases, (A) Case 1, (B) Case 2, (C) Case 3, (D) Case 4.

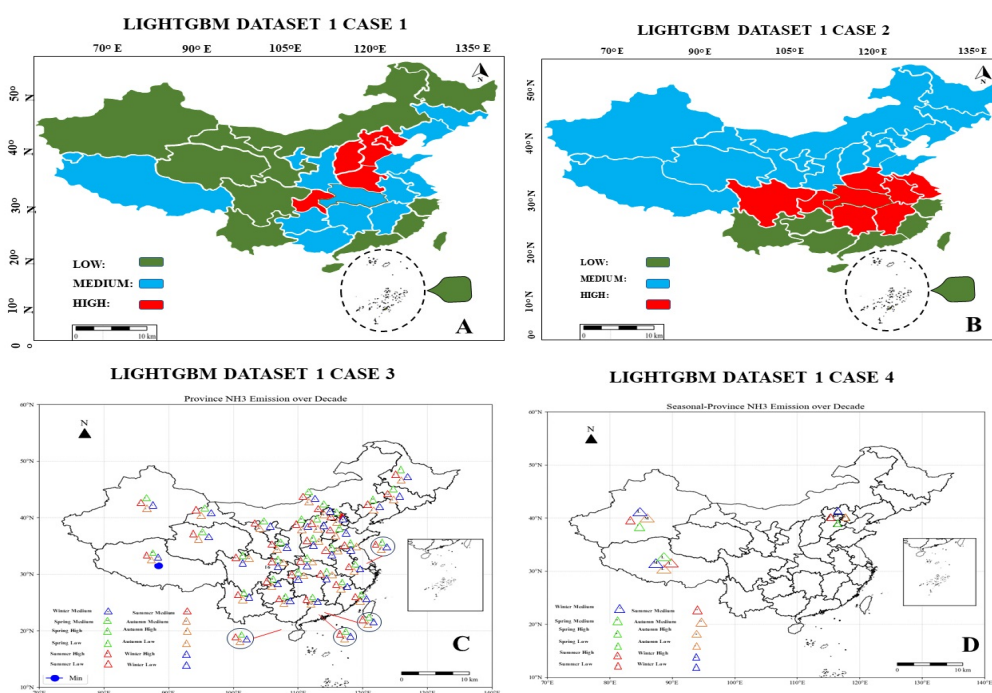


Figure 5. The LightGBM machine learning model over four cases, (A) Case 1, (B) Case 2, (C) Case 3, (D) Case 4.

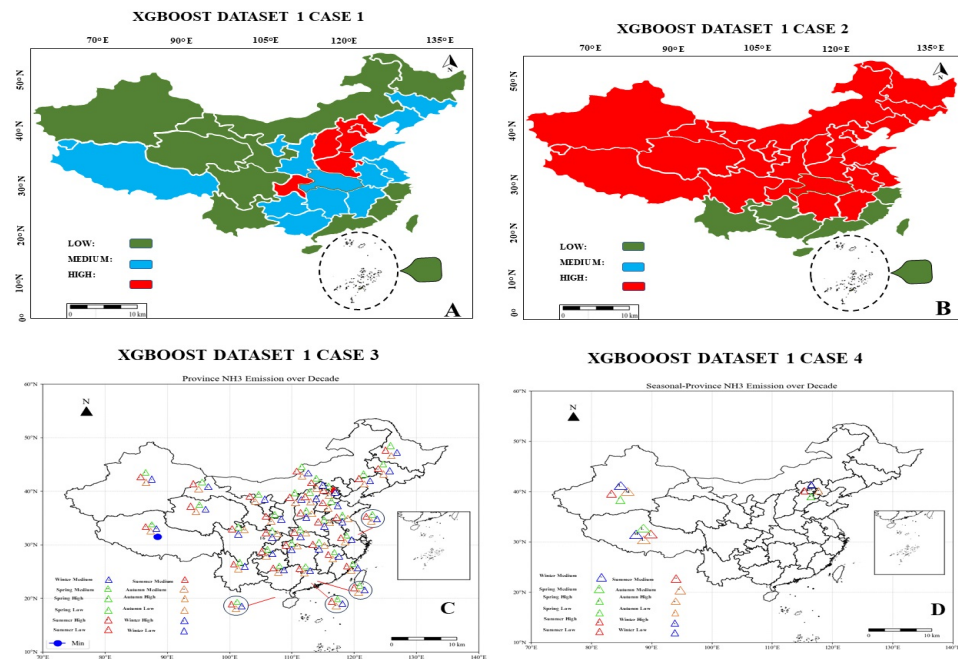


Figure 6. The XGBOOST machine learning model over four cases, (A) Case 1, (B) Case 2, (C) Case 3, (D) Case 4.

To evaluate prediction accuracy, confusion matrices were generated by comparing predicted values with the actual threshold-derived reference. The results are presented in Figure 7.

As it shown, from the heatmap of province-wise classification accuracy for Case 1, there is considerable variation in province-wise classification accuracy of the models. XGBoost, LightGBM and GWRF had an accuracy of 100% for all the provinces and administrative regions, which proved to have a high within-sample classification consistency. Random Forest also performed very well, and it achieved 100% in almost all provinces except in the Inner Mongolia Autonomous Region, where the accuracy was 0%. By contrast, SVM showed a wide variation with an accuracy from 0% to 100%, depending on the data properties of the region.

The performances of LSTM fluctuated among provinces/municipalities, showing 0% in some provinces/municipalities, including Henan, Beijing, Chongqing, Shanghai, Hebei, Tibet, etc., and 100% in several provinces/municipalities, including Fujian, Hubei, Hunan, Liaoning, Shaanxi, Shandong, Xinjiang, Yunnan, Zhejiang, Zhoushan Islands, etc. The obtained results show that the ensemble-based methods (i.e., XGBoost, LightGBM, and GWRF implementation used in this study) have the highest and most consistent classification performance for all geographical regions, while the methods of SVM and LSTM are less robust and more sensitive to the region-dependent variations in the dataset.

The heatmap of zone-wise classification accuracy for Case 2 reveals varying performances between the assessed machine learning models, highlighting a notable variation in the three geographic zones. The GWRF implementation used in this study and LightGBM had 100% accuracy in all zones, showing good classification accuracy and reliability. This was achieved with 100% accuracy in the central, north and south zones by Random Forest. For XGBoost, 100% samples were classified as central/south zone, whereas the samples in north zone were not classified, hence an accuracy of 0%. The SVM showed an ability of maximum 100% accuracy with a minimum of 20% accuracy in north and south zones. Also, LSTM achieved 100% accuracy in the central and south zones, while it failed to achieve any accuracy in the north zone. The results suggest that the three classifiers—GWRF, LightGBM,

and Random Forest—show the most stable performance to all zones, and SVM, XGBoost, and LSTM in less stable when performing on the north zone.

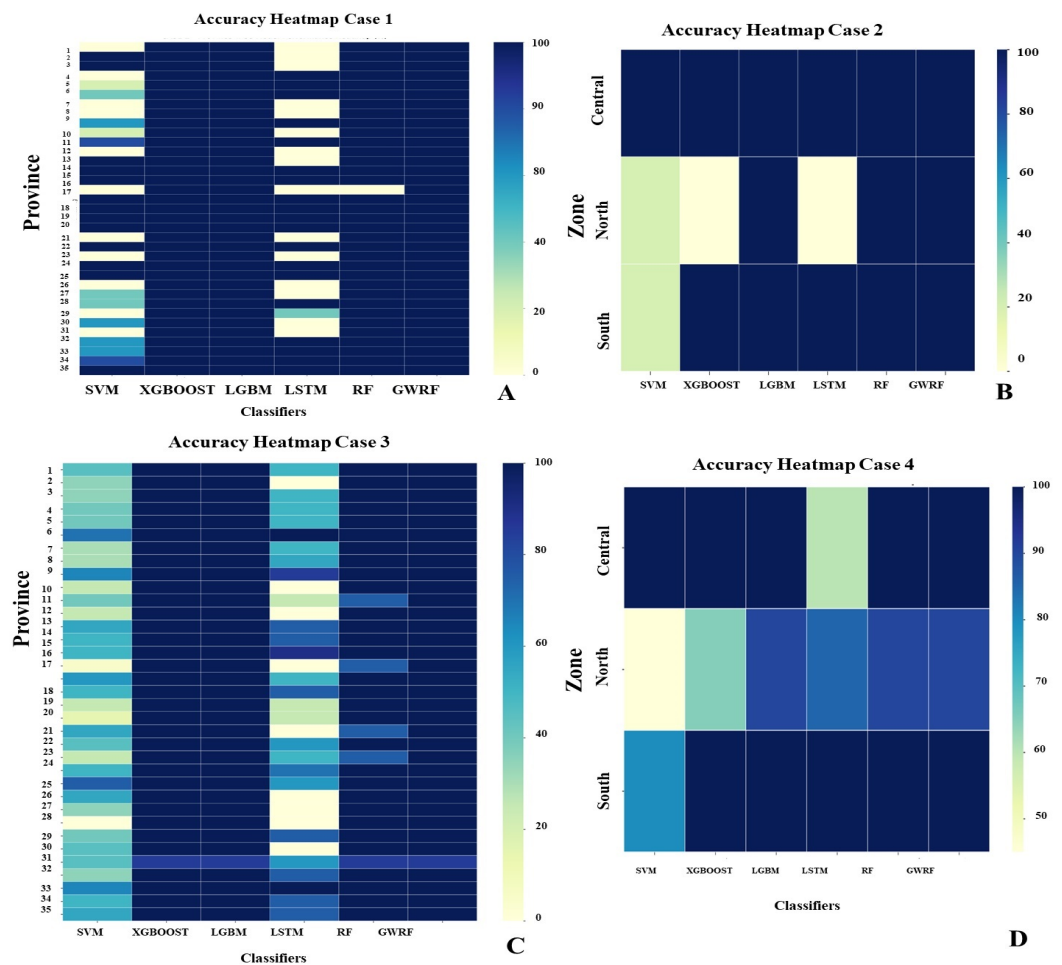


Figure 7. Province-wise and zone-wise classifier-accuracy heatmaps for the four experimental cases. Panels (A–D) correspond to Cases 1–4. Rows represent provinces or zones, columns represent classifiers, and color intensity indicates accuracy.

The Province-wise Classification Accuracy heatmap of Case 3 shows the strong variations between the robustness of the machine learning models evaluated in each province and administration area. The models of XGBoost, LightGBM and the GWRF implementation had nearly 100% accuracy for almost all provinces except the Tibet Autonomous Region where it fell to 85%. Random Forest had fairly good performances as well, with 100% accuracy for most provinces or regions, and 75–85% for those of Heilongjiang Province, Inner Mongolia Autonomous Region, Macau SAR, Qinghai Province and Tibet Autonomous Region. SVM, in contrast, had much more variable and lower performance, with the accuracy varying from low of 0% in Sichuan Province to high of 75% in Shandong Province. LSTM was able to moderate classification with various municipalities/provinces having 0%, such as Beijing, Hebei, Henan, Shanghai, Shanxi, Sichuan, and up to 100% in Guangdong Province and Yunnan Province. Higher LSTM accuracies were also found in Hainan Province, Hunan Province, Hong Kong SAR and some provinces in the eastern regions, and results were not found to be consistent throughout regions. In general, it can be concluded that XGBoost, LightGBM, and GWRF have the most consistent and stable classification performance across all provinces, and Random Forest has good performance with slightly less stability, while SVM and LSTM are more sensitive to regional variation in the dataset.

The zone-wise classification accuracy heatmap for Case 4 reveals a clear disparity among the central, north and south zones in terms of the accuracy of the models. In the central and south zones, the consistently high accuracy of LightGBM and the GWRF implementation attained 100%, and 90% in the north zone, respectively, as a result of their high stability. Random Forest also had high performance, having 100% accuracy in the central and south sub-zones and 90% in the north zone. The overall results indicated that XGBoost classified all the data samples in the central and south zones and with a slight deviation 65% in the north zone. SVM had varying results in central zone with 100%, south zone 80% and north zone 45%. There was moderate and inconsistent performance with LSTM achieving 100% accuracy in the South, 85% in the North, and 60% in the Central zone. Overall, LightGBM, GWRF and Random Forest perform with the most reliability when compared to other models, whereas SVM, XGBoost and LSTM models exhibit less solidity especially in the North zone.

In all four cases: XGBoost, LightGBM, and the GWRF implementation showed a very high and stable classification accuracy across both provinces and zones and revealed high classification consistency within the evaluated regions. Random Forest exhibited generally high performance with some decreases in performance observed in some of the provinces and in northern zones. Both SVM and LSTM had variable and inconsistent performance ranging between 0% and 100% with dependence of the region where they were used. All in all, the ensemble and the gradient boosting models delivered the most consistent classification especially in all geographical regions, while the SVM and LSTM models were less robust.

3.1.7. Feature Contribution Results

Since Dataset 1 showed the best overall performance in the previous stage, it is selected as the reference dataset for further analysis. This decision was made after analyzing the performances across the three datasets, where Dataset 1 (CAMS) consistently yielded the best results across all six classifiers and four cases. As the next step, to better understand the effectiveness of the extracted features and their influence on the classifiers, the contribution of each feature to model performance is systematically evaluated across the four cases using Dataset 1. Forward Sequential Feature Selection and Leave-One-Out Cross-Validation (LOO-CV) were applied to the six classifiers. At each step, candidate features were evaluated individually, and the feature producing the largest increase in LOO-CV accuracy was retained. The process continued until all features were ranked. A normalized hybrid contribution score was then calculated within each model and case. Because the classifiers provide different importance measures, the score combines model-specific quantities after normalization. For SVM and LSTM, permutation importance and absolute Pearson correlation were used. For XGBoost and LightGBM, native gain, permutation importance, and correlation were combined. For Random Forest and GWRF, Gini importance, permutation importance, correlation, and SHAP-based information were combined. The resulting values are descriptive relative rankings and should not be interpreted as causal physical effects. The results of this feature-wise analysis are summarized in Tables 7 and 8 and Figure 8. As the feature-wise values for six models (M1–M6) across the four cases (Case1 to Case 4) shown, each case corresponds to a different analysis scenario: Case 1 (Province–Year), Case 2 (Zone–Year), Case3 (Province–Season–Year), and Case 4 (Zone–Season–Year). The features include MAV, Kurtosis, RMS, Slope, IQR, Standard Deviation, Skewness, and Z-score. For Case 1 and Case 3, which use province-level data, MAV reaches up to 64.26 (M3, Case 1) and RMS up to 32.58 (M4, Case 3), while slope values reach 5.87 (M4, C3), indicating that these models capture the magnitude and trends of the input signals more strongly. Kurtosis ranges from 1.03 to 7.42, and skewness varies between 0.53 and 7.25 across models, reflect-

ing sensitivity to signal shape and asymmetry. Z-score values show relative differences, with M1 and M4 reaching up to 7.92 (Case 3). For Case 2 and Case 4, which use zone-level data, MAV varies sharply from 11.40 to 48.05 (Case 4), and RMS ranges from 6.38 to 37.72, showing the effect of seasonal aggregation on signal representation. Slope values reach 27.61 (M1, Case 4) and IQR ranges from 4.42 to 31.96, indicating higher heterogeneity in seasonal data. Overall, feature distributions differ notably across cases and models, which likely contributes to variations in classification performance.

Table 7. Performance summary across cases (Low, Medium, High Categories).

Case	Low Category	Medium Category	High Category
Case 1	Fujian, Gansu, Guangdong, Hainan, Heilongjiang, Hong Kong SAR, Inner Mongolia, Qinghai, Taiwan, Xinjiang, Yunnan, Zhejiang, Zhoushan Islands	Anhui, Guangxi, Guizhou, Hubei, Hunan, Jiangsu, Jiangxi, Jilin, Liaoning, Ningxia, Shaanxi, Shandong, Sichuan, Tibet	Beijing, Shanghai, Tianjin, Chongqing, Hebei, Henan, Shanxi, Macau SAR
Case 2	South Zone	North Zone (except XGBOOST High in Spring; others Medium)	Central Zone
Case 3	Autumn (partially), Winter (mainly Low), some seasonal dips across models	Most values across seasons (dominant level across dataset)	Spring and Summer (frequent peaks across models)
Case 4	South Zone (Spring, Summer, Autumn), North Zone (Spring, Winter)	South Zone (Winter), North Zone (Autumn + LIGHTGBM/GWRF seasons), mixed North patterns	Central Zone (all seasons), North Zone (XGBOOST Summer)

Table 8. Normalized relative feature-contribution scores (%) for six classifiers across four cases. Features are MAV, Kurtosis, RMS, Slope, IQR, Std, Skewness, and Z-score_{max}. Models are M1 (SVM), M2 (XGBoost), M3 (LightGBM), M4 (LSTM), M5 (Random Forest), and M6 (GWRF).

Case	Feature	M1	M2	M3	M4	M5	M6	Case	M1	M2	M3	M4	M5	M6
C1	MAV	27.04	56.46	64.26	30.61	47.93	47.13	C2	22.70	49.23	59.12	20.40	28.22	23.75
	Kurtosis	5.06	1.65	1.69	2.26	2.13	2.59		0.62	5.21	9.66	2.07	4.91	3.89
	RMS	25.28	19.66	16.93	30.36	21.47	21.97		22.62	11.77	13.28	20.12	27.14	31.48
	Slope	2.07	1.77	0.86	1.28	2.71	1.98		4.97	4.84	10.54	16.56	6.91	8.56
	IQR	12.17	7.61	7.19	13.35	10.92	11.54		29.36	4.42	6.09	18.03	10.79	10.84
	Std	16.79	10.79	8.12	14.05	12.11	11.77		17.86	9.45	5.79	16.55	11.14	13.11
	Skewness	7.25	1.56	0.53	4.59	1.48	1.68		1.31	7.79	2.18	4.35	5.50	4.54
	Z-score	4.35	0.49	0.40	3.40	1.25	1.19		0.57	7.20	3.33	1.91	5.38	3.82
C3	MAV	26.02	50.56	59.07	33.21	43.77	46.23	C4	36.80	48.05	30.96	44.57	11.40	14.17
	Kurtosis	7.42	2.21	1.03	3.08	1.67	1.58		3.56	10.36	9.46	3.65	9.38	11.70
	RMS	24.96	17.28	20.40	32.58	25.30	23.98		10.59	10.05	6.38	6.97	37.72	15.53
	Slope	5.29	3.35	3.88	5.87	5.34	4.86		27.61	8.61	5.92	25.25	6.59	11.44
	IQR	11.13	4.19	5.30	7.23	7.37	7.27		10.10	8.40	31.96	8.31	11.56	15.66
	Std	14.37	6.61	5.74	9.12	11.06	10.57		3.63	8.02	6.54	3.63	7.42	10.23
	Skewness	2.89	4.33	2.12	3.71	2.89	2.73		3.84	3.27	4.43	3.84	8.74	11.59
	Z-score	7.92	2.48	2.47	5.20	2.60	2.77		3.78	3.23	4.36	3.78	7.19	9.67

The numerical results indicate a clear shift in feature dominance across the four cases. In C1 and C2, the feature space is largely MAV-driven, where higher MAV values (up to 64.26 in C1 and 48.05 in C4) dominate the signal representation. In contrast, C3 and C4 show a more balanced contribution of features, where RMS (up to 37.72), slope (up to 27.61), and variability measures such as IQR and standard deviation become more influential. Tree-based models such as XGBoost, LightGBM, Random Forest, and GWRF consistently align with higher MAV and RMS ranges, while models like SVM and LSTM

are more affected by variability-related features, including slope, skewness, and kurtosis. Overall, the results indicate that feature importance is case-dependent and varies with both temporal aggregation and the learning mechanism of each classifier.

Feature Contribution Results Over Cases for Dataset 1

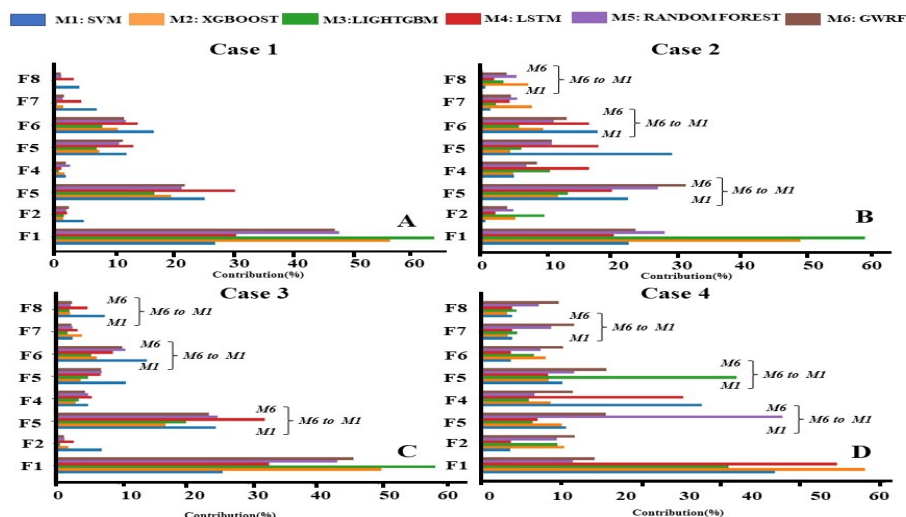


Figure 8. LOO performance feature contribution results over cases for Dataset 1, M1 SVM, M2 XGBoost, M3 LightGBM, M4 LSTM, M5 Random Forest, M6 GWRF, F1: MAV, F2: Kurtosis, F3: RMS, F4: Slope, F5: IQR, F6: Std, F7: Skewness, F8: Zscore max.

4. Conclusions

This study evaluated relative NH_3 reference classes across three multi-source datasets and four spatial–temporal cases: Province–Year, Zone–Year, Province–Season–Year, and Zone–Season–Year. Six machine-learning classifiers were compared using chronological testing, cross-validation, and feature-contribution analysis. Both data quality and model performance are taken into account, Dataset 3 (MEIC v2.0) emerges as the most stable one and possesses more uniform performance with each one of the classifiers and hence is considered the most appropriate one for analysis. It contains less provinces and data points than Dataset 1, but allows for more reliable and uniform behavior in prediction across models and experimental configurations. Dataset 1 (CAMS GEI), on the contrary, is still suitable for running on a larger scale, such as in national-level assessments or long-time trend analysis. XGBoost, LightGBM and GWRF exhibit the improvements in accuracy and the most stable results on all datasets and cases. Random Forest is generally effective, but there are some differences in certain areas. The performance of SVM and LSTM is less stable and inferior under high spatial aggregation. The four cases show different spatial–temporal behaviors. Province-level analysis captures detailed regional differences, zone-level analysis reflects broader spatial patterns, and seasonal cases show temporal variation in emission levels. Ensemble and boosting models remain stable across all cases. As the study shows, in Case 1 (Province vs. Year), low NH_3 levels occur in coastal or high-altitude provinces because of low agricultural/livestock density, and cooler/wetter climate. Moderate farming and fertilizer application in the mid latitudes give medium NH_3 in the provinces. High NH_3 in intensive fertilization areas in the tropics and subtropics with warm temperatures. Seasonal and spatial variations are an impact of combined effects of crop cycles, livestock, and urban–industrial activities. Case 2 (Zone vs. Year): Low NH_3 in South Zone because of high rainfall and the reduced volatilization from fertilizer. Medium NH_3 in North Zone (moderate agriculture, livestock, seasonal fertilization—XGBoost detects NH_3 peaks in spring). Intensive agriculture, animal husbandry, and dry warm weather

when volatilization is enhanced all contribute to the high NH_3 in Central Zone. There is significant interaction between human activities and climate zones in patterns. Case 3 (Province—Season vs. Year): Low NH_3 in autumn, winter as temperatures are cooler and minimal fertilizer is used. Moderate NH_3 concentrations during most seasons, indicative of moderate agricultural use and animal sources. High NH_3 in spring/summer because of high fertilizer application, warm temperature and active livestock manure management. Seasonal peaks coincide with agricultural activities and climatic factors related to NH_3 volatilization. Case 4 (Zone—Season vs. Year): Low NH_3 at south zone (spring–autumn) and North Zone (winter) because of rainfall or cold surface limiting NH_3 volatilization. Moderate NH_3 levels occur in south zone (winter) and north zone (autumn/mixed)). Intensive agriculture, livestock, high fertilizer usage: high NH_3 in central zone (all seasons) and north zone (summer). The patterns show the overall effects of seasonality, intensity of farming, temperature, and fertilization practices. Feature analysis shows clear variation in importance across cases. MAV and RMS are the most important features in province-level analysis. Slope, IQR, standard deviation, and skewness become more important in zone- and season-level cases. Z-score contributes to separating relative differences across regions.

Overall, MEIC combined with XGBoost, LightGBM, and GWRF showed the most stable classification behavior, while CAMS GEI remained useful for broader national-scale and long-term analysis. Because the labels are threshold-derived and the datasets differ in physical meaning, the high scores should be interpreted as reference-class separability rather than independent physical prediction. Future work should include independent datasets, physically harmonized products, simple baselines, spatial-block validation, and external predictors such as fertilizer use, livestock density, meteorology, and land use.

Author Contributions: A.J.M.: Methodology, Software, Validation, Formal analysis, Investigation, Writing—original draft, Writing—review and editing, Visualization, Project administration. B.H.S.: Methodology, Validation, Formal analysis, Investigation, Writing—original draft, Writing—review and editing, Visualization. A.S.K.: Methodology, Validation, Formal analysis, Investigation, Writing—original draft, Writing—review and editing, Visualization. S.A.E.: Validation, Formal analysis, Investigation, Writing—original draft, Writing—review and editing, Visualization, Project administration. A.K.: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing—review and editing, Visualization, Supervision, Project administration. D.B.: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing—review and editing, Visualization, Supervision, Project administration. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding

Data Availability Statement: Data collected for this study are available by request through the corresponding author.

Acknowledgments: The Authors thank Robotics and Automation Research Lab (RARL, <https://rarl-lab.com/>) and Smart Structural Health Monitoring and Control for resilient, safe infrastructure (SHMC: <https://www.sshmc.org/>) for their help and support.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
ANN	Artificial Neural Network
AUC	Area Under the Curve
D	Dataset

D1	Dataset 1
D2	Dataset 2
D3	Dataset 3
DT	Decision Tree
F1-Score	Harmonic Mean of Precision and Recall
HCS	Hybrid Contribution Score
KNN	K-Nearest Neighbors
L	Low Class
M	Medium Class
H	High Class
ML	Machine Learning
NC	Number of Classes
NS	Number of Samples
NY	Number of Years
PCA	Principal Component Analysis
RF	Random Forest
ROC	Receiver Operating Characteristic
R²	Coefficient of Determination
SHAP	SHapley Additive exPlanations
SPC	Samples per Class
SVM	Support Vector Machine
XGBoost	Extreme Gradient Boosting

Appendix A

Algorithm A1 Unified Time-Series Classification Pipeline for Multiple Methods with Classifier Settings (Part I)

Require: Dataset $\mathcal{D}$, Train Years, Test Years, K-Folds K , Optuna Trials N

Ensure: Prepared Data and Selected Classification Pipeline

- 1: **Step 1: Data Acquisition**
- 2: Load dataset $\mathcal{D}$
- 3: **Step 2: Auto-Detect Structure**
- 4: Identify geographic ID column and chronological Year column
- 5: **Step 3: Data Cleaning**
- 6: Drop null/sparse rows and convert features to numeric types
- 7: **Step 4: Chronological Split for Train–Test Leakage Control**
- 8: Split data into training and testing sets based on years
- 9: **Step 5: Compute Statistics (Train Only)**
- 10: Compute mean μ and standard deviation σ of target feature
- 11: **Step 6: Categorize Target**
- 12: Convert continuous target to classes $C \in \{0, 1, 2\}$

$$C = \begin{cases} 0, & x \leq \mu - 0.5\sigma \\ 1, & \mu - 0.5\sigma < x \leq \mu + 0.5\sigma \\ 2, & x > \mu + 0.5\sigma \end{cases}$$
- 13: Apply the same thresholds to test data
- 14: **Step 7: Encode Geographic Features**
- 15: Encode categorical geographic labels into numeric IDs
- 16: **Step 8: Feature Scaling (Train Only)**
- 17: Fit scaler using training data ONLY
- 18: Transform training data
- 19: Apply the same scaler to test data
- 20: **Step 9: Chronological Split**
- 21: Split training and testing data according to years

Algorithm A1 *Cont.*

```

22: Step 10: Select Classifier Pipeline
23: if model = LSTM then
24:     Standardize features
25:     Reshape to (samples, 1, features)
26: else if model = SVM then
27:     Standardize features
28: else if model = XGBoost then
29:     Skip feature scaling
30: else if model = LightGBM then
31:     Skip feature scaling
32: else if model = Random Forest then
33:     Skip feature scaling
34: else if model = GWRF then
35:     Encode spatial-weight features
36: end if

```

Algorithm A2 Unified Time-Series Classification Pipeline for Multiple Methods with Classifier Settings (Part II)

```

1: if model = LSTM then
2:     Initialize LSTM with L2 regularization, recurrent dropout, batch normalization
3:     Dense Softmax output

```

$$\hat{y}_i = \frac{e^{z_i}}{\sum_{j=1}^C e^{z_j}}$$

```

4: else if model = SVM then
5:     Optuna search: C, kernel, gamma, degree

```

$$f(x) = \text{sign} \left(\sum_{i=1}^N \alpha_i y_i K(x_i, x) + b \right)$$

```

6: else if model = XGBoost then
7:     Optuna search: max_depth, min_child_weight, n_estimators
8:     Ensemble prediction

```

$$\hat{y} = \text{softmax} \left(\sum_{k=1}^{n_{\text{estimators}}} f_k(x) \right)$$

```

9: else if model = LightGBM then
10:    Optuna search: num_leaves, max_depth, n_estimators

```

$$\hat{y} = \sum_{l=1}^L w_l$$

```

11: else if model = Random Forest then
12:    Train ensemble of decision trees

```

$$\hat{y} = \text{mode} \{f_1(x), f_2(x), \dots, f_n(x)\}$$

```

13: else if model = GWRF then
14:    Train geographically weighted random forest

```

$$\hat{y} = \text{Weighted Local Tree Aggregation}$$

```

15: end if
16: Step 9: Evaluation

```

Algorithm A2 *Cont.*

```

17: Apply K-Fold Cross-Validation
18: Train final model on full training data
19: Predict on test set
20: Compute Accuracy
21: Compute Macro Precision
22: Compute Macro Recall
23: Compute Macro F1-score
24: Compute Macro ROC-AUC when mathematically defined
25: if model = GWRP then
26:     Extract OOB score
27: end if
28: Generate classification report
29: Generate confusion matrix
30: Plot training/validation curves
31: return trained model  $\mathcal{M}$  and evaluation metrics

```

References

1. Moshayedi, A.J.; Sohail Khan, A.; Hu, J.; Nawaz, A.; Zhu, J. E-nose-driven advancements in ammonia gas detection: A comprehensive review from traditional to cutting-edge systems in indoor to outdoor agriculture. *Sustainability* **2023**, *15*, 11601.
2. Shah, S.B.; Westerman, P.W.; Arogo, J. Measuring ammonia concentrations and emissions from agricultural land and liquid surfaces: A review. *J. Air Waste Manag. Assoc.* **2006**, *56*, 945–960.
3. Moshayedi, A.J.; Khan, A.S.; Geng, K.; Hu, J.; Kolahdooz, A. Advancing Agricultural Practices: Analyzing the Role of Robotics in Corn Farming. *Int. J. Eng.* **2025**, *38*, 1517–1532. <https://doi.org/10.5829/ije.2025.38.07a.07>
4. Tavousi, A.; Shariati-Rad, M. Ammonia Detection Based on Image Analysis of a Colorimetric Paper Sensor. *Anal. Bioanal. Chem. Res.* **2026**, *13*, 283–300.
5. Moosa, M.Y.; Sathyakam, P.U. From structure to sensing: Metal Organic Framework for ammonia gas detection. *Mater. Today Phys.* **2026**, *61*, 102029. <https://doi.org/10.1016/j.mtphys.2026.102029>.
6. McGinn, S.M.; Janzen, H.H. Ammonia sources in agriculture and their measurement. *Can. J. Soil Sci.* **1998**, *78*, 139–148.
7. Chiang, C.L.; Huang, Y.Y.; Liu, T.Y.; Chen, Y.Y.; Soppera, O.; Chen, L.Y.; Meng, H.F.; Zan, H.W. Two-Minute Onsite Detection of Agricultural Water Nitrogen Pollution with an Ammonia Gas Sensing System. *ACS ES&T Water* **2025**, *5*, 2965–2974.
8. Li, T.; Zhang, T.; Ji, W.; Huang, J.; Zheng, Y.; Zhang, W.; Li, Z.; Shen, W. Decadal spatiotemporal dynamics of agricultural ammonia emissions in the yangtze river delta (2000–2019): Patterns, drivers, and sustainability implications. *Environ. Dev. Sustain.* **2026**, *1*–23. <https://doi.org/10.1007/s10668-026-07933-x>
9. Moshayedi, A.J.; Khan, A.S.; Shuxin, Y.; Kuan, G.; Jiandong, H.; Soleimani, M.; Razi, A. E-Nose design and structures from statistical analysis to application in robotic: a compressive review. *EAI Endorsed Trans. AI Robot.* **2023**, *2*, e1.
10. Dammers, E.; McLinden, C.A.; Griffin, D.; Shephard, M.W.; Van Der Graaf, S.; Lutsch, E.; Schaap, M.; Gainairu-Matz, Y.; Fioletov, V.; Van Damme, M.; et al. NH₃ emissions from large point sources derived from CrIS and IASI satellite observations. *Atmos. Chem. Phys.* **2019**, *19*, 12261–12293.
11. Luo, Z.; Zhang, Y.; Chen, W.; Van Damme, M.; Coheur, P.F.; Clarisse, L. Estimating global ammonia (NH₃) emissions based on IASI observations from 2008 to 2018. *Atmospheric chemistry and physics*. *Atmos. Chem. Phys.* **2022**, *22*, 10375–10388.
12. Wang, Y.; Ma, Y.-F.; Eskes, H.; Inness, A.; Flemming, J.; Brasseur, G.P. Evaluation of the CAMS global atmospheric trace gas reanalysis 2003–2016 using aircraft campaign observations. *Atmos. Chem. Phys.* **2020**, *20*, 4493–4521. <https://doi.org/10.5194/acp-20-4493-2020>.
13. Wagner, A.; Bennouna, Y.; Blechschmidt, A.-M.; Brasseur, G.; Chabrilat, S.; Christophe, Y.; Errera, Q.; Eskes, H.; Flemming, J.; Hansen, K.M.; et al. Comprehensive evaluation of the Copernicus Atmosphere Monitoring Service (CAMS) reanalysis against independent observations. *Elem. Sci. Anthr.* **2021**, *9*, 00171. <https://doi.org/10.1525/elementa.2020.00171>.
14. Lei, M.; Cheng, T.; Li, X.; Shi, S.; Zuo, X.; Guo, H.; Wu, Y. Atmospheric ammonia point source detection technique at regional scale using high resolution satellite imagery and deep learning. *Atmos. Res.* **2021**, *257*, 105587. <https://doi.org/10.1016/j.atmosres.2021.105587>.
15. Chen, J.; Cheng, M.; Krol, M.; de Vries, W.; Zhu, Q.; Liu, X.; Zhang, F.; Xu, W. Trends in anthropogenic ammonia emissions in China since 1980: A review of approaches and estimations. *Front. Environ. Sci.* **2023**, *11*, 1133753. <https://doi.org/10.3389/fenvs.2023.1133753>.

16. Ding, J.; van der A, R.; Eskes, H.; Dammers, E.; Shephard, M.; Kruit, R.W.; Guevara, M.; Tarrason, L. Ammonia emission estimates using CrIS satellite observations over Europe. *Atmos. Chem. Phys.* **2024**, *24*, 10583–10599. <https://doi.org/10.5194/acp-24-10583-2024>.
17. Johansson, S.; Höpfner, M.; Friedl-Vallon, F.; Glatthor, N.; Gulde, T.; Huijnen, V.; Kleinert, A.; Kretschmer, E.; Maucher, G.; Neubert, T.; et al. Ammonia (NH₃) in the UTLS: GLORIA airborne measurements for CAMS model evaluation in the Asian Monsoon and in biomass burning plumes above the South Atlantic. *EGUsphere* **2024**. <https://doi.org/10.5194/egusphere-2024-77>.
18. Sun, F.; Cui, Y.; Su, J.; Shephard, M.W.; Kharol, S.K.; Zhang, Y.; Shi, X.; Zhang, J.; Liu, H.; Xiao, Q.; et al. Decadal changes in atmospheric ammonia and dry deposition across China inferred from space-ground measurements and model simulations. *Atmos. Chem. Phys.* **2025**, *25*, 17907–17932. <https://doi.org/10.5194/acp-25-17907-2025>.
19. Kumar, P.; Broquet, G.; Hauglustaine, D.; Beaudor, M.; Clarisse, L.; Van Damme, M.; Coheur, P.; Cozic, A.; Zheng, B.; Romero, B.R.; et al. Global atmospheric inversion of the anthropogenic NH₃ emissions over 2019–2022 using the LMDZ-INCA chemistry transport model and the IASI NH₃ observations. *Atmos. Chem. Phys.* **2025**, *25*, 12379–12407. <https://doi.org/10.5194/acp-25-12379-2025>.
20. Moshayedi, A.J.; Abbasi, R.; Teng, M.; Chen, Y.; Dahoo, P.R.L.; Bassir, D. Decadal Analysis of Agricultural Ammonia Distribution in China Using CAMS Data and Machine Learning Classifiers. In *Proceedings of the 2025 2nd Asia Conference on Advances in Electrical and Power Engineering (ACEPE)*; IEEE: New York, NY, USA, 2025; pp. 1–8. <https://doi.org/10.1109/ACEPE67873.2025.11405537>.
21. Wan, J.; Wang, X.; Cheng, Y.; Zhang, C.; Xue, F.; Yang, T.; Tong, F.; Wang, Q.J. Identification of nighttime urban flood inundation extent using deep learning. *Nat. Hazards Earth Syst. Sci.* **2025**, *25*, 4361–4373. <https://doi.org/10.5194/egusphere-2025-77>.
22. Tang, K.; Zhang, H.; Xu, G.; Chang, F.; Xu, Y.; Miao, J.; Cui, X.; Jin, J.; Li, B.; Li, K.; et al. Optimizing Ammonia Emissions for PM_{2.5} Mitigation: Environmental and Health Co-Benefits in Eastern China. *EGUsphere* **2025**. <https://doi.org/10.5194/egusphere-2025-1407>.
23. Choi, H.; Park, M.E.; Bae, J.-H. Distinct Regional and Seasonal Patterns of Atmospheric NH₃ Observed from Satellite over East Asia. *Remote Sens.* **2025**, *17*, 2587. <https://doi.org/10.3390/rs17152587>.
24. Sheng, M.; Zhou, R.; Hua, J.; Han, S.; Liu, S.; Zhang, L.; Wang, W.; Dang, R.; Cao, H.; Chen, Z.; et al. Geostationary observations of atmospheric ammonia over East Asia: spatio-temporal variations revealed by three years of FY-4B/GIIRS measurements. *EGUsphere* **2025**. <https://doi.org/10.5194/egusphere-2025-5699>.
25. Zhang, Y.; Brasseur, G.; Kanakidou, M.; Granier, C.; Daskalakis, N.; Poulidis, A.P.; Qu, K.; Vrekoussis, M. Towards an integrated inventory of anthropogenic emissions for China. *Geosci. Model Dev.* **2026**, *19*, 217–237. <https://doi.org/10.5194/gmd-19-217-2026>.
26. Cao, Y.; Ren, C.; Zhang, H.; Wei, Z.; Guo, Y.; Cai, X. Dynamic Modeling of Ammonia Emissions and Nitrogen Deposition via Online Coupling of WRF-Chem and Noah-MP-CN. *J. Geophys. Res. Atmos.* **2026**, *131*, e2025JD044260. <https://doi.org/10.1029/2025jd044260>.
27. Sindelarova, K.; Markova, J.; Simpson, D.; Huszar, P.; Karlicky, J.; Darras, S.; Granier, C. High-resolution biogenic global emission inventory for the time period 2000–2019 for air quality modelling. *Earth Syst. Sci. Data* **2022**, *14*, 251–270.
28. Block, K. *Cloud Condensation Nuclei (CCN) Numbers Derived from CAMS Reanalysis EAC4 (Version 1)*; WDC Climate: Hamburg, Germany, 2023.
29. He, K. Multi-resolution Emission Inventory for China (MEIC): Model Framework and 1990–2010 Anthropogenic Emissions. In *Proceedings of the AGU Fall Meeting Abstracts*; American Geophysical Union: Washington, DC, USA, 2012; Volume 2012, p. A32B–05.
30. Achmad, A.; Syarif, S. Prediction of Ammonia Contamination Levels in Wastewater Management Plant Using the SVM Method. In *2021 International Seminar on Intelligent Technology and Its Applications (ISITIA)*; IEEE: New York, NY, USA, 2021; pp. 250–255.
31. Xu, T.; Alkhayyat, A.; Hamodi, Z.A.; Yadav, A.; Shreenidhi, H.S.; Kamil, H.; Kumar, A.; Singh, D.; Jain, V.; Shomurotova, S.; et al. Accurate Data-Driven Frameworks to Estimate Solubility of Ammonia in Ionic Liquids. *J. Chem. Eng. Jpn.* **2025**, *58*, 2560443.
32. Zhang, Y.; Mao, J.; Li, W.; Wu, C.; Huang, K. XGBoost machine learning-coupled dual-mode sensing platform for on-site quantification of sulfide ions in oxygen-deficient combustion. *Microchem. J.* **2025**, *219*, 116249.
33. Wang, Y.; Xu, D.; Li, X.; Wang, W. Prediction model of ammonia nitrogen concentration in aquaculture based on improved AdaBoost and LSTM. *Mathematics* **2024**, *12*, 627.
34. Madani, A.; Hagage, M.; Elbeih, S.F. Random Forest and Logistic Regression algorithms for prediction of groundwater contamination using ammonia concentration. *Arab. J. Geosci.* **2022**, *15*, 1619.
35. Neethling, J.B.; Willey, B.; Nichols, A.; Hetherington, M.; Heinz, D.; Wade, R.; Sevenandt, W. Converting BOD to Nitrification/-Denitrification in Same Reactor While Maintaining Capacity. In *Proceedings of the Proceedings of the Water Environment Federation WEFTEC*; Water Environment Federation: Alexandria, VA, USA, 2011; pp. 6627–6637.
36. Arlot, S.; Celisse, A. A survey of cross-validation procedures for model selection. *Stat. Surv.* **2010**, *4*, 40–79. <https://doi.org/10.1214/09-SS054>.

37. Silva, L.A.; Zanella, G. Robust leave-one-out cross-validation for high-dimensional Bayesian models. *J. Am. Stat. Assoc.* **2024**, *119*, 2369–2381.
38. Moshayedi, A.J.; Aghda, A.M.; Eftekhari, S.A.; Andani, M.E.; Salahshour, S. Respiratory parameter estimation using pharyngeal phonetics and machine learning: Breaking free from spirometry. *Eng. Appl. Artif. Intell.* **2025**, *160*, 111989.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.