AGU ADVANCING EARTH AND SPACE SCIENCES

Water Resources Research

RESEARCH ARTICLE
10.1029/2025WR040137

An Automated Data Efficient Morphometric Approach to Define Global Lentic and Lotic Inland Waters

Ankit Sharma[1], **Mukund Narayanan**[1], and **Idhayachandhiran Ilampooranan**[1,2]

[1]Department of Water Resources Development and Management, Indian Institute of Technology Roorkee, Roorkee, Uttarakhand, India, [2]Centre for Sustainable Energy, Indian Institute of Technology Roorkee, Roorkee, Uttarakhand, India

Key Points:

- An accurate framework for defining global lentic and lotic systems up to 0.09 ha using multi-dimensional morphometric interpretation
- Allows automated, data-efficient, and explainable classification of inland waters with equal performance to machine and deep learning
- This framework advances upon previous global classification products and maintains high accuracy during 1991–2021

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

I. Ilampooranan,
idhaya@wr.iitr.ac.in

Citation:

Sharma, A., Narayanan, M., & Ilampooranan, I. (2026). An automated data efficient morphometric approach to define global lentic and lotic inland waters. *Water Resources Research*, *62*, e2025WR040137. https://doi.org/10.1029/2025WR040137

Received 3 FEB 2025
Accepted 5 JAN 2026

Author Contributions:

Conceptualization: Ankit Sharma, Mukund Narayanan, Idhayachandhiran Ilampooranan
Data curation: Ankit Sharma
Formal analysis: Ankit Sharma, Mukund Narayanan, Idhayachandhiran Ilampooranan
Funding acquisition: Idhayachandhiran Ilampooranan
Investigation: Ankit Sharma, Mukund Narayanan, Idhayachandhiran Ilampooranan
Methodology: Ankit Sharma, Mukund Narayanan, Idhayachandhiran Ilampooranan
Project administration: Idhayachandhiran Ilampooranan

Abstract Defining lentic and lotic system types is critical for understanding hydrological, ecological, and biochemical processes. Traditional classification methods rely on non-generalizable site-specific parameters such as visual characteristics, historical inventory, and residence time. While machine learning and deep learning models address these challenges to some extent, they are limited by high data requirements, unverified training data sets, computational demands, and the inability to accurately detect inland waters smaller than 3 ha. To address this gap, this study introduces a novel Automated Data Efficient Morphometric Approach (ADEMA) that classifies inland waters into lentic and lotic system types globally up to 0.09 ha (33 times smaller than previous studies) using multi-dimensional morphometric interpretations. ADEMA was developed and validated using 17,391 expert-labeled inland waters spanning 66 globally diverse locations and compared against state-of-the-art, comprehensively optimized machine learning, deep learning, and global models. Results show ADEMA equivalently performed to the machine learning and deep learning models, achieving $F1$ scores of 92%, 95%, and 71% in small, medium, and large inland waters, respectively. Across 17,391 expert-labeled samples, ADEMA maintained a high performance with a precision of 89%, a recall of 99%, and an $F1$ score of 94%. Analysis across four decadal intervals (1991–2021) demonstrated ADEMA's temporal invariance, with consistently high $F1$ scores (90%–93%) and negligible omission errors (0%–2%). Further, ADEMA surpassed global classification products (average $F1$ score: 97% vs. 62%). These findings emphasize ADEMA's potential for accurately classifying global inland waters into lentic and lotic system types.

Plain Language Summary Understanding the differences between lentic (still) and lotic (flowing) inland waters is important for studying water flow, aquatic life, and the environment. However, traditional methods for finding these differences rely on site-specific traits that are difficult to apply everywhere. While machine learning and deep learning models help, they require large data sets, high computing power, and cannot differentiate small inland waters. To overcome these limitations, this study introduces a new method called ADEMA (Automated Data Efficient Morphometric Approach), which was developed and tested using 17,391 expert-labeled inland waters across 66 locations worldwide. Unlike previous studies, ADEMA can find differences in inland waters up to 33 times smaller than existing methods and models by looking at their shape and structure in multiple ways. ADEMA performed on par with machine learning and deep learning models, achieving high accuracy, particularly for small and medium-sized inland waters. It also remained accurate over four decadal intervals (1991–2021), showing that it is reliable across time. These results highlight ADEMA's ability to classify inland waters efficiently and accurately, filling a critical gap in global hydrography.

1. Introduction

Defining lentic and lotic system types in Inland Waters (IWs; see Appendix A for all abbreviations) is fundamental to understanding their hydrological, limnological, ecological, and biogeochemical functions (Allen & Pavelsky, 2018; Jones et al., 2017; Messager et al., 2016; Qu et al., 2018). Lentic system types (still IWs), such as lakes, ponds, and reservoirs, retain water for extended periods, creating relatively still conditions that favor the accumulation of sediments and support biological communities (Dodson et al., 2000; Downing et al., 2008; Post et al., 2000). In contrast, lotic system types (flowing IWs) behave in a fluvial manner, such as streams, creeks, and rivers, with shorter residence time and greater downstream transport of sediments and nutrients (Fisher et al., 2012; Liu et al., 2021). While these definitions do not consider multiple key differentiating factors such as shape, size, salinity, etc., they are extended to classify all IWs binarily based on how similar the system is to a lake-like (lentic) or a river-like (lotic). Fundamentally, the major differences in lentic and lotic system types are often characterized by residence time and morphometry (Jones et al., 2017; Luo et al., 2021; Van Der Werff &

Resources: Idhayachandhiran Ilampooranan
Software: Ankit Sharma, Mukund Narayanan
Supervision: Idhayachandhiran Ilampooranan
Validation: Ankit Sharma, Mukund Narayanan, Idhayachandhiran Ilampooranan
Visualization: Ankit Sharma, Mukund Narayanan, Idhayachandhiran Ilampooranan
Writing – original draft: Ankit Sharma, Mukund Narayanan
Writing – review & editing: Ankit Sharma, Mukund Narayanan, Idhayachandhiran Ilampooranan

Van Der Meer, 2008). However, defining these system types was also based on historical inventory and visual interpretation-based segregation (Liu et al., 2022) as well as satellite image segmentation techniques (De Lucia Lobo et al., 2012; Jiang et al., 2014; Martínez-Santos et al., 2021; Meng et al., 2019). Most typically, residence times and flow rates have been previously proposed as indicators for distinguishing lentic and lotic system types (Andersen & Shafroth, 2010; Błędzki & Ellison, 2000; Pellett et al., 1983). For example, Rennella and Quirós (2006) set a 15-day residence time threshold for differentiating lentic and lotic system types in the Salado river basin of Argentina. Unfortunately, such simple residence time thresholds are site- and time-specific, limiting their global applicability. To overcome this limitation, Jones et al. (2017) developed a more comprehensive freshwater continuum classification framework using integrated residence time, inflow, discharge, and volume to classify lentic, intermediate, and lotic system types. Nevertheless, this methodology focused solely on hydrologically connected IWs and was unsuitable for defining individual IW systems. Furthermore, residence time-based classifications are hindered by data availability, as inflow, discharge, and volume records are sparse for many IWs (Cael et al., 2017; Messager et al., 2016; Verpoorter et al., 2014). Alternatively, using satellite images, Jiang et al. (2014) developed an automated method using band normalization, threshold segmentation, and image intensity change analysis to extract rivers (lotic) and lakes (lentic). Nonetheless, the limitations to this approach were that (a) the method was validated for a single time step and within a specific region (China), and (b) the smallest IW classifiable was 5.4 ha.

At a global scale, studies like Pi et al. (2022a) and Khandelwal et al. (2022) sought to separate lentic and lotic system types using the global surface water data set (Pekel et al., 2016a) through deep learning (DL) and erosion-based thresholding operations. While these global classification products improved classification accuracy, they were data-intensive and unable to classify IWs smaller than 3 ha. A simpler and data-efficient alternative approach is morphometry-based classification, which differentiates the IW systems based on the topology, such as roundness, elongation, compactness, curvature, etc. (Liu et al., 2021). In previous attempts to classify IWs using morphometry, Van Der Werff and Van Der Meer (2008) used three morphometric factors derived from a single LANDSAT image to classify lentic and lotic system types in the Koyukuk River, Alaska, USA. Similarly, Jiao et al. (2012) delineated lentic and lotic system types in Changjiang County, Hainan Province, China, using 10 shape factors from a single tile of SPOT-5 imagery with 12 scenes. However, both studies had certain limitations: (a) the smallest classifiable IW was 45 ha, (b) the methods were region-specific, and (c) they lacked significant validation. As an improvement, Luo et al. (2021) used eight shape factors and machine learning (ML) models to classify rivers (lotic) and lakes (lentic) in the Yangtze River basin in China. Yet, this method was limited to IWs larger than 3 ha. Further, this method used a static (single-timestep) and unverified data set for training and validation, potentially reducing classification reliability. Therefore, there is a need for a data-efficient, globally applicable, and robust morphometric-based approach to define lentic and lotic system types. Hence, this study proposed a novel Automated Data Efficient Morphometric Approach (ADEMA) for defining lentic and lotic system types globally up to 0.09 ha (equivalent to a single LANDSAT pixel) using multi-dimensional morphometric interpretations. ADEMA was developed and validated using expert-labeled data from 66 spatially diverse locations comprising 17,391 lentic and lotic system types. As a first attempt, ADEMA is benchmarked against previous well-established global classification products and state-of-the-art, comprehensively optimized ML and DL models for different size classes, that is, small (<10 ha), medium (10–1,000 ha), and large (>1,000 ha).

2. Materials and Methods

2.1. Inland Waters Data Acquisition and Labeling

ADEMA was developed using the Global Surface Maximum Water extent (GSMW) data set from the Joint Research Centre, Europe (Pekel et al., 2016a). The GSMW data set provides comprehensive global data spanning 60°S to 80°N, covering all areas ever detected as IWs from 1984 to 2021 at a 30-m spatial resolution (Pekel et al., 2016b). This data set provides a pixel-based maximum IW extent, capturing seasonal variations, long-term trends, and changes in IWs' presence over 38 years (1984–2021). With an Omission Error (OE) of 5% and a false detection rate of 1% (Pekel et al., 2016a; Pi et al., 2022a), the GSMW data set offers highly accurate IW maps. Therefore, it was selected to define precise and consistent system type labels for the development of ADEMA and other models in this study. A total of 66 locations across the globe were selected to generate labels for lentic and lotic system types (Figure 1). To ensure representative sampling of IWs types, the following factors were considered: (a) sufficient availability of lentic and lotic system types, (b) inclusion of continuous and fragmented lotic system types to account for possible misclassifications, (c) incorporation of elongated lentic IWs resembling

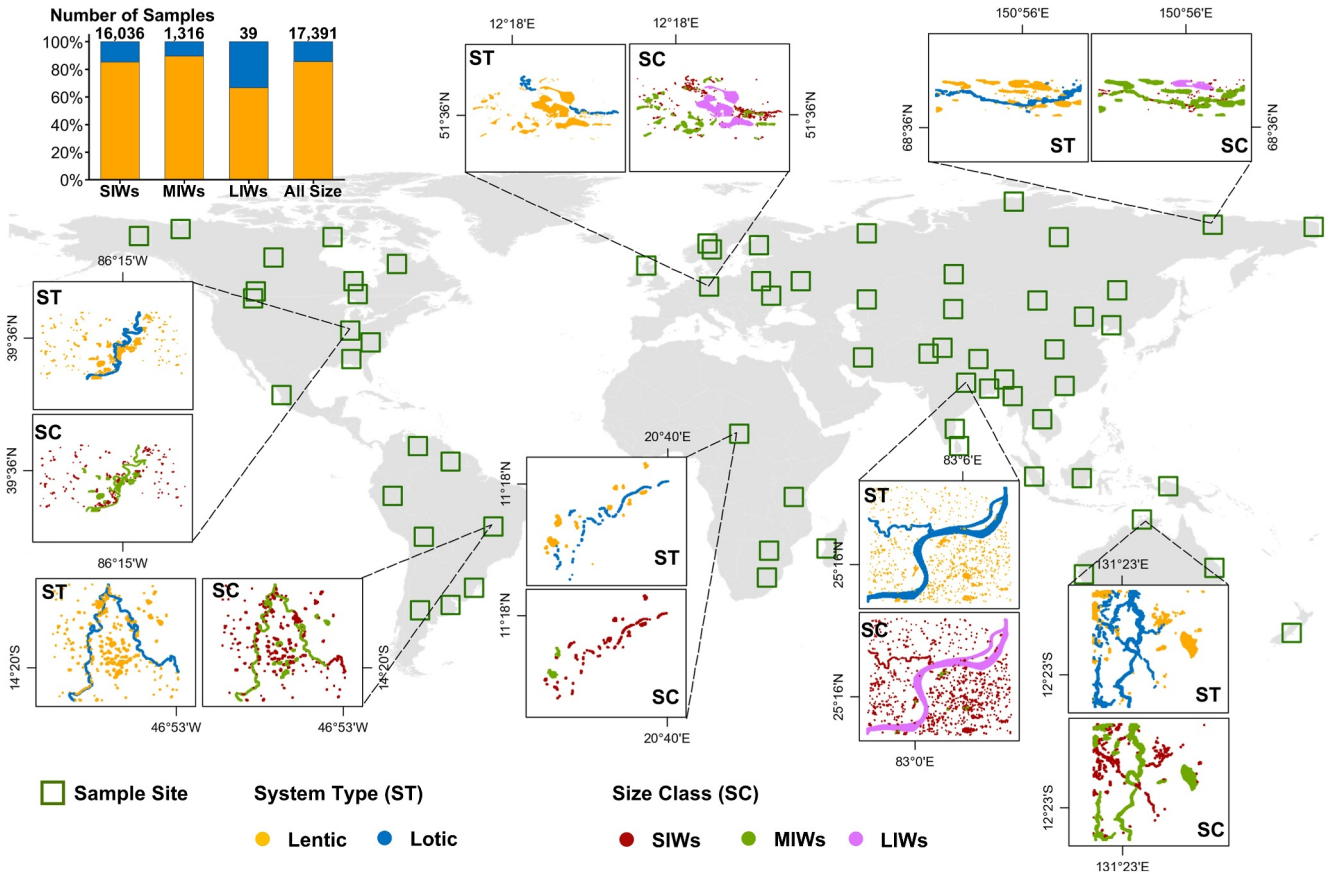

Figure 1. Sample sites for training and validation of ADEMA with zoomed versions of lentic and lotic System Types (ST) as well as Size Classes (SC) across 66 global locations. Small Inland Waters (SIWs), Medium Inland Waters (MIWs), and Large Inland Waters (LIWs) refer to Small (<10 ha), Medium (10–1,000 ha), and Large (>1,000 ha) Inland Waters.

rivers, and (d) consistency with global distribution of lentic (500 million ha; ~86% of surface water area (Verpoorter et al., 2014)) and lotic system types (78 million ha; ~14% of surface water area (Allen & Pavelsky, 2018)), by selecting locations with more lentic system types than lotic system types through visual inspection. However, in its existing (raw) format, the GSMW data set provides only IW pixels, making it unsuitable for system type labeling. To address this, the GSMW data set was preprocessed through the "*BUfferDissolveSplit*" (*BUDS*) function (developed in this study) to generate accurate and spatially connected polygon boundaries. The *BUDS* function was implemented in Google Earth Engine (GEE) (Gorelick et al., 2017), utilizing its powerful cloud computing capabilities to effectively preprocess the GSMW data in two stages.

In the first stage, the *BUDS* function converted the IW pixels from the raster format by aggregating adjacent pixels with the same value into vector polygons. However, the resulting polygons were fragmented into smaller and discrete entities, and islands were formed within the vector polygons, failing to capture the continuous spatial extent of the IW polygons accurately. For example, as illustrated in Figure S1a in Supporting Information S1, converting the GSMW raster layer of the lotic system type into vector polygons resulted in 23 fragmented polygons, while the lentic system type conversion yielded six discrete IW polygons, failing to capture the full extent. To address this issue, the second stage of the *BUDS* function buffered each IW polygon proportionally to its area, where the buffer radius (r) was computed using Equation 1.

$$r = \sqrt{\frac{\mathrm{IW}_{\mathrm{area}}}{\pi}} \tag{1}$$

This area-based buffering ensured that neighboring and island polygons of varied sizes were merged within the same lentic or lotic entity without coalescing with unrelated features. The resulting buffered polygons were then

dissolved into a single unified polygon, eliminating fragmentation. However, since this process consolidated multiple polygons into a single spatial entity, the *BUDS* function further decomposed the single spatial entity into its constituent components. This decomposition was achieved by using the "*.toList()*" and "*.geometries()*" functions within the *BUDS*, transforming each component back into a standalone entity (Figure S1a in Supporting Information S1). This overall process ensured that while overlaps were eliminated during the first stage, the individual identities of the IW entities were retained in the second stage.

After processing the GSMW data set using the *BUDS* function, the derived IW polygons were mere geometries without lentic and lotic system type labels. A major challenge in demarcating these system types across global scales and multiple time steps is their complex and dynamic morphologies, influenced by seasonality, topography, connectivity, and anthropogenic modifications. To overcome this challenge and annotate these labels, this study used an expert group that performed a well-established and validated visual analytics and an evidential reasoning approach (Brown et al., 2022; Pekel et al., 2016a). The expert groups, functioning as non-parametric classifiers, integrate domain knowledge to enable interactive exploration and interpretation of system type morphometry while effectively handling data variability and quality issues in the GSMW data set. In the expert group, six annotators with expertise in domains such as agriculture, hydrology, and eco-hydrology, and basic knowledge of remote sensing and image processing, annotated IW polygons as lentic (label: "1") or lotic (label: "0"). Lotic system types include rivers, streams, and tributaries (Figure S2a in Supporting Information S1), while lentic system types include the rest of the polygons (Figure S2b in Supporting Information S1). The 66 global locations were evenly distributed among the six annotators, with each annotator independently labeling the polygons within their assigned location using ArcGIS 10.8 and very high-resolution Esri, Maxar satellite imagery. Each polygon was initially labeled by a single annotator and then cross-checked by the other annotators. These labeled polygons formed the basis for subsequent ML, DL, and ADEMA development (more details in Sections 2.3 and 2.4). A total of 17,391 IW polygons were labeled globally with 86% lentic and 14% lotic system types (Figure 1).

While the *BUDS* function improved same-system type connectivity, some edge cases remained due to GSMW data set limitations, such as sensor resolution, vegetation interference, and cloud cover. To evaluate *BUDS* function's effectiveness in those edge cases, we compared ADEMA classifications (refer to Section 2.3 for methodology) with *BUDS* and NO *BUDS* across nine representative samples (see Text S1 in Supporting Information S1).

2.2. Estimation of Morphometric Variables

The data for developing ML, DL, and ADEMA used the expert-annotated 17,391 IWs polygons. For all these polygons, four morphometric variables, that is, Circularity Ratio (CR), Lemniscate Ratio (LR), Elongation Ratio (ER), and Compactness (CT), were estimated using GEE (Table 1). The area (A_{iw}) and perimeter (P_{iw}) of each IW polygon were estimated using the "*geometry.area()*" and "*geometry.perimeter()*" functions. The maximum length for each polygon was estimated using the "*convexHull*" function, which creates the smallest convex boundary enclosing the polygon. The vertices of this convex boundary were then extracted, and the distance between each pair of adjacent vertices was calculated. The longest distance between two adjacent vertices was determined and assigned as the maximum length of the polygon (L_{max}).

After estimating the morphometric variables, the data were subsetted into three size classes based on the area of each IW polygon, namely, Small Inland Waters (SIWs) (<10 ha), Medium Inland Waters (MIWs) (10–1,000 ha), and Large Inland Waters (LIWs) (>1,000 ha). Further, these subsets were also combined into a superset size class called "All Size," which included IWs from all three size classes. Post subsetting and combining, 60% of each size class was split into training, and the remaining 40% was used for testing to develop the ADEMA, ML, and DL models. Moreover, the training and testing subsets were spatially uniformly distributed across 66 global locations to minimize spatial clustering and reduce regional biases (Figure S3 in Supporting Information S1).

2.3. Development of Automated-Data-Efficient-Morphometric Approach (ADEMA)

A novel data-efficient approach for classifying lentic and lotic system types was proposed through multi-dimensional morphometric interpretation of CR, ER, LR, and CT spaces (Figures 2a–2c). This approach involved delineating decision boundaries based on the points of minimal separation between system types in each dimensional space of the training data set (Figure S3 in Supporting Information S1). To elaborate, decision

Table 1
The Definitions, Interpretations, Formulae, and Ranges of Morphometric Variables

Morphometric variable	Definition	Mathematical formula	Interpretation	Range
Circularity Ratio (CR)	It is the ratio of the IW polygon area to the circle area with the same perimeter	$CR = \frac{4\times\pi\times A_{iw}}{P_{iw}^2}$	Higher values mean the shape is more *"circle-like"*	[0.0007 to 0.79]
Lemniscate Ratio (LR)	It is the ratio of the area of the lemniscate having the same length as the IW polygon to the actual IW area	$LR = \frac{\pi\times(L_{max})^2}{4\times A_{iw}}$	Larger values mean the polygon extends farther in at least one direction relative to its overall area	[0.08 to 113.87]
Elongation Ratio (ER)	It is the ratio of the diameter of a circle with the same area as the IW polygon to the polygon's maximum length	$ER = \frac{2\times\sqrt{\frac{A_{iw}}{\pi}}}{L_{max}}$	Lower values indicate the shape is more stretched out; a higher value suggests it is less elongated	[0.09 to 3.45]
Compactness (CT)	It is the ratio of the IW polygon's perimeter to its area	$CT = \frac{P_{iw}}{A_{iw}}$	Smaller values mean the shape is more compact; larger values mean it is more irregular or elongated	[0.75 to 324.80]

Note. Where, π is a constant (3.14), A_{iw} is the area of the polygon in km^2, P_{iw} is the perimeter of the polygon in km, L_{max} is the maximum length of the polygon in km.

boundaries are determined heuristically for all six morphometric dimensional spaces (CR vs. ER, CR vs. CT, CR vs. LR, ER vs. LR, ER vs. CT, and LR vs. CT). Then, for each space, points that most effectively distinguish lentic and lotic system types guide the selection of the primary boundaries. Misclassifications from these primary boundaries were then analyzed visually, and additional boundaries in secondary or tertiary spaces were established to reduce/eliminate misclassification. As a result of following these steps, the ADEMA structure was delineated to classify IWs as a lentic or a lotic system type and is outlined in Figure 2d and Equation 2. Elaborating on the structure, a system is classified as lotic if the IW's shape is minimally circular or distinctly non-circular, characterized by CR < 0.0008. Conversely, IWs with a mildly circular shape (0.0008 ≤ CR < 0.08)

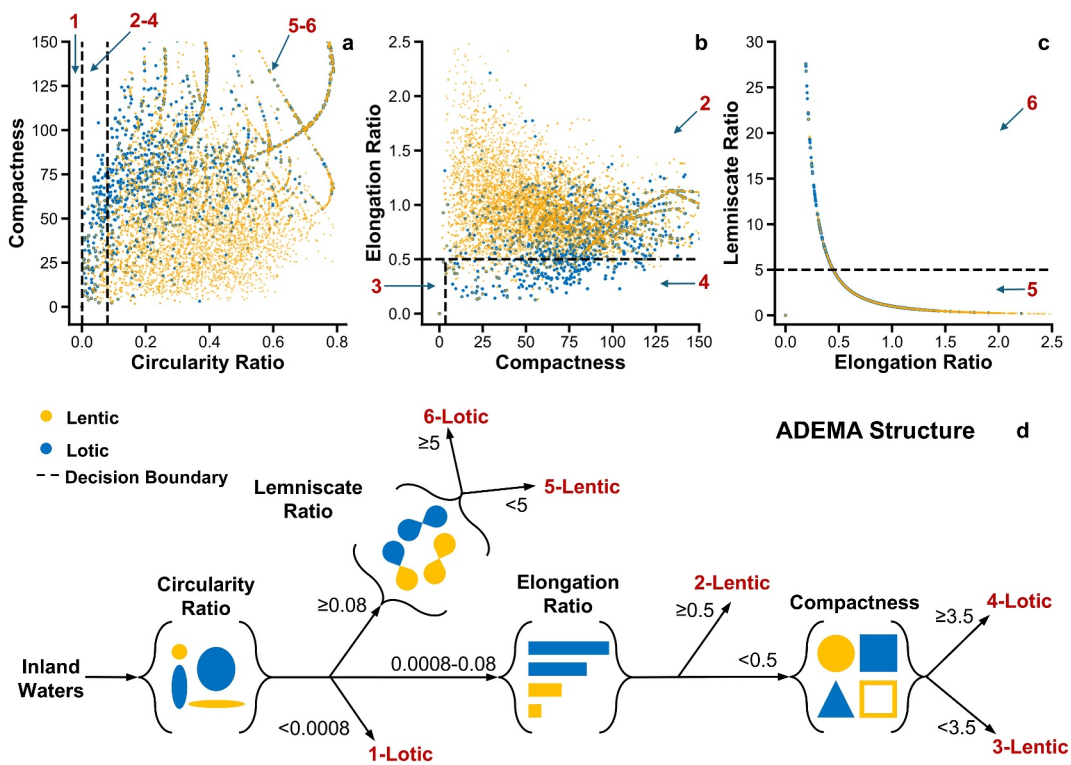

Figure 2. ADEMA framework for defining lentic and lotic IWs. Subplots (a–c) show multi-dimensional morphometric interpretation of lentic (gold) and lotic (dark blue) IWs with decision boundaries as dashed lines across circularity ratio, elongation ratio, lemniscate ratio, and compactness spaces. Subplot (d) presents a flow chart of the classification rules (structure) in the ADEMA framework, including threshold values for all six cases.

are categorized as lentic if they are not highly elongated (ER > 0.5). For IWs with a mildly circular shape ($0.0008 \leq CR < 0.08$) that are highly elongated (ER < 0.5), compactness further refines their classification. In this category, IWs who show moderate compactness (CT < 3.5) were classified as lotic, while those with higher compactness (CT ≥ 3.5) were deemed lentic. When the IW's shape is moderately circular (CR > 0.08), the lemniscate ratio provides the final classification criterion, where IWs with LR < 5 are categorized as lentic and those with LR ≥ 5 are classified as lotic. The robustness of these thresholds was evaluated through proportional k-fold cross-validation (86% lentic and 14% lotic system types were maintained in all folds), using 10 randomized folds on the training data set. ADEMA showed consistent $F1$ scores across all folds (mean ± standard deviation: 92.41 ± 0.03% for training, 92.41 ± 0.28% for testing) (Table S2 in Supporting Information S1), confirming its robustness for classification.

$$\text{System Type} = \begin{matrix} \text{Lotic,} \\ \\ \\ \text{Lentic,} \end{matrix} \begin{cases} CR < 0.0008 \text{ OR} \\ 0.0008 \leq CR < 0.08 \text{ AND } ER < 0.5 \text{ AND } CT \geq 3.5 \text{ OR} \\ CR \geq 0.08 \text{ AND } LR \geq 5 \\ \text{Otherwise} \end{cases} \tag{2}$$

2.4. Machine and Deep Learning Models

Using the generated morphometric variables for 17,391 expert-labeled IWs, three ML and one DL models were trained and tested, namely, Decision Tree (DT) (Breiman et al., 1984), Random Forest (RF) (Breiman, 2001), eXtreme Gradient boosting (XG) (Chen & Guestrin, 2016), and artificial Neural Networks (NN) (Rosenblatt, 1958). A total of four independent variables were used in these models (Equation 3).

$$ST = f_m(CR, LR, ER, CT) \tag{3}$$

where, ST is the IW system type, that is, lentic or lotic, and $f_m()$ is the machine or deep learning (DT, RF, XG, and NN) model used under the mth size class (SIWs, MIWs, LIWs, or All Size), $CR, LR, ER,$ and CT are the morphometric variables mentioned in Table 1. During implementation, these independent variables were scaled using the "*StandardScaler*" function in the "*preprocessing*" module of the "*scikit-learn*" library (Pedregosa et al., 2011) to maintain uniformity. Further, the models were tuned for their hyperparameters, whose definitions and ranges are listed in Table 2.

In total, 13 hyperparameters across all models were tuned (Table 2). To avoid local optima and to have a broader exploration of the hyperparameter ranges, two optimization algorithms (instead of just one conventionally), namely, Gaussian process minimization and gradient-boosted tree minimization, were used for tuning. Gaussian process minimization applies a Bayesian model that predicts the objective function and estimates uncertainty in unexplored regions (Head et al., 2021). Gradient boosted tree minimization constructs an ensemble of iteratively refined trees, correcting errors from predecessors (Head et al., 2021). These optimization algorithms were performed by defining an objective function (Obj$F$) whose value is calculated during each iteration for a trial configuration of hyperparameters (Equation 4).

$$\text{Obj}F = (F1_{tr} - F1_{ts}) - \frac{(F1_{tr} + F1_{ts})}{2} \tag{4}$$

where, $F1_{tr}$ is the $F1$ score (Equation 9) of the training data set and $F1_{ts}$ is the $F1$ score of the testing data set. In Obj$F$, training and testing $F1$ scores are averaged to maximize accuracy and subtracted to reduce overfitting. The Obj$F$ was evaluated for multiple iterations till convergence. Since the optimization techniques were run for minimization, the trial configuration of the hyperparameters giving the lowest Obj$F$ value among the optimization methods was selected (Table S3 in Supporting Information S1). Overall, the ML and DL models were developed in Python 3.12.4 by importing the "*tree*," "*ensemble*," and "*neural_network*" modules from the "*scikit-learn*" library alongside the "*xgboost*" module, and the models were optimized using the "*scikit-optimize*" library (Head et al., 2021). The details on the best sets of hyperparameters obtained after optimization are presented in Table S3 in Supporting Information S1.

Table 2
Hyperparameter Definitions and Ranges Used During ML and DL Model Optimization

Hyperparameter	ML/DL Model(s)	Definition	Type	Range
max_depth	DT, RF, & XG	The maximum depth of the tree	Integer	[1 to 17,391 (total observations)]
max_features	DT & RF	The number of features to consider when looking for the best split	Integer	[1 to 4 (total variables used)]
min_samples_split	DT & RF	The minimum number of samples required to split an internal node	Integer	[2 to 17,391 (total observations)]
min_samples_leaf	DT & RF	The minimum number of samples required to be at a leaf node	Integer	[1 to 17,391 (total observations)]
criterion	DT & RF	The function to measure the quality of a split	Categorical	["gini," "entropy"]
n_estimators	RF & XG	The number of trees	Integer	[10 to 1,000]
learning_rate	XG	Step size shrinkage is used in the update to prevent overfitting	Float	[0.01 to 0.5]
subsample	XG	Subsample ratio of the training instances. Setting it to 0.5 means that XG would randomly sample half of the training data prior to growing the tree to prevent overfitting	Float	[0.5 to 1.0]
colsample_bytree	XG	The subsample ratio of columns when constructing each tree	Float	[0.0 to 1.0]
hidden_layer_sizes	NN	The number of hidden layers	Integer	[10 to 200]
alpha	NN	Strength of the L2 regularization term. The L2 regularization term is divided by the sample size when added to the loss.	Float	[0.0001 to 0.01]
solver	NN	Approach used in NN for arriving at the best accuracy	Categorical	["adam," "sgd"]
learning_rate_init	NN	The initial learning rate used. It controls the step-size in updating the weights. Only used when solver = "sgd" or "adam"	float	[0.001 to 0.1]

2.5. Accuracy Assessment Metrics

The classification performance of the ADEMA, ML, and DL models was assessed using a comprehensive set of accuracy assessment metrics, including the $F1$ score (Hripcsak & Rothschild, 2005), discrete Area Under the receiver operating characteristic Curve (dAUC), precision (user accuracy), recall (producer accuracy), Commission error (CE) and Omission error (OE) (Equations 5–9). These metrics were derived from the confusion matrices constructed for each model (ML, DL, and ADEMA), following standard definitions of true and false classifications. Specifically, True Lentic (TLe) includes the lentic polygons correctly identified as a lentic system type (true positive), True Lotic (TLo) includes lotic polygons that are correctly identified as a lotic system type (true negative), False Lentic (FLe) includes the lotic polygon incorrectly classified as lentic (false positive) and False Lotic (FLo) includes the lentic polygons incorrectly classified as lotic (false negative). The dAUC (%) was calculated using binary values obtained from each model through the "*roc_auc_score*" function from the "*metrics*" module of the "*sklearn*" library in Python 3.12.4. Since the distribution of lentic (86%) and lotic (14%) system types in our data set reflects the global proportions, no explicit preprocessing was performed to balance system type distribution. However, to provide a more unbiased evaluation, the models were also checked for weighted $F1$ scores (Equation 10). The weight (w_{ST}) for lentic (Le) system types is 0.86, and the w_{ST} for lotic (Lo) system types is 0.14. In Equation 10, the suffix ST indicates system type specific.

$$\text{Precision (\%)} = \frac{\text{TLe}}{\text{TLe} + \text{FLe}} \times 100 \tag{5}$$

$$\text{Recall (\%)} = \frac{\text{TLe}}{\text{TLe} + \text{FLo}} \times 100 \tag{6}$$

$$\text{Commission Error (\%)} = 100 - \text{Precision (\%)} \tag{7}$$

$$\text{Omission Error (\%)} = 100 - \text{Recall (\%)} \tag{8}$$

$$F1\ \text{score}\ (\%) = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{9}$$

$$\text{Weighted}\ F1\ \text{score}\ (\%) = \sum_{\text{ST=Le}}^{\text{Lo}} w_{\text{ST}} \times \frac{2 \times \text{Precision}_{\text{ST}} \times \text{Recall}_{\text{ST}}}{\text{Precision}_{\text{ST}} + \text{Recall}_{\text{ST}}} \tag{10}$$

2.6. Explainability-Based Variable Ranking

To examine how the independent variables explain the system type classifications, a jack-knifing test was conducted to quantitatively rank each independent variable. This test involved creating a baseline model using all four independent variables and recording its $F1$ score as the baseline performance. Next, an iterative process was implemented where each independent variable was excluded from the model one at a time. For each iteration, the model was retrained using the reduced set of independent variables, and the $F1$ score was recalculated. This approach ensures that the observed changes in performance are directly attributable to the excluded variable, eliminating any correlations among independent variables. The drop in $F1$ score after each iteration is calculated as the difference between the baseline $F1$ score ($F1_n$) and the $F1$ score after excluding a variable ($F1_{n-i}$). Mathematically, this drop ($\Delta\ F1_i$) in performance for variable i can be expressed as Equation 11:

$$\Delta\ F1_i = F1_n - F1_{n-i} \tag{11}$$

where, n is the total number of independent variables ($n = 4$). While this method can be performed for ML and DL models, in the case of ADEMA, due to hierarchical connections among decision boundaries (e.g., CR thresholds are predecessors to ER, LR, and CT thresholds (Figure 2d)), this jack-knifing method could not be performed. To get an equivalent method for the explainability of independent variables in ADEMA, this study developed the Jack-knifing for Explaining Hierarchical variables in the Lentic and Lotic system type classification (JEHLL) method. JEHLL utilizes the hierarchical ranking of independent variables in ADEMA's classification framework (Figure 2d). The variables are ranked as CR > ER > LR > CT, reflecting their sequential evaluation. Specifically, CR is assessed first, followed by the subordinate stages for ER and LR, while CT is evaluated last. Based on these ranks, the following steps of JEHLL were implemented: (a) the lowest hierarchical variable (CT) was first excluded, and the $\Delta\ F1_i$ for CT is estimated. Similarly, the $\Delta\ F1_i$ for LR was calculated by excluding it from ADEMA (b) next, the $\Delta\ F1_i$ for ER is calculated by excluding ER from the hierarchy and making CT (the subordinate of ER) the sole decision criterion for conditions involving ER in ADEMA, and (c) lastly, $\Delta\ F1_i$ for CR is calculated by excluding CR from the hierarchy, and in its place, the decision criteria were based on all other subordinate variables, such as ER, LR, and CT in ADEMA. For more clarity on the criteria used under each exclusion in JEHLL, refer to Table S4 in Supporting Information S1. Further, these individual drops in performance ($\Delta\ F1_i$) calculated from jack-knifing and JEHLL were then summed across all variables grouped by each model to obtain the total performance drop ($\Delta\ F1_{\text{total}}$) as represented in Equation 12:

$$\Delta\ F1_{\text{total}} = \sum_{i=1}^{n} \Delta\ F1_i \tag{12}$$

Finally, the explainability score of each variable was calculated as the proportion of its individual performance drop relative to the total drop. This explainability score (η_i) for variable i is given by Equation 13:

$$\eta_i = \left(\frac{\Delta\ F1_i}{\Delta\ F1_{\text{total}}} \right) \times 100 \tag{13}$$

2.7. Temporal Invariance Analysis Across Decadal Intervals

The accuracy of ADEMA for defining system type was tested at four decadal intervals: 1991, 2001, 2011, and 2021. The IW data for all four years were collected from the "*Yearly History*" layer of the global surface water data set (Pekel et al., 2016a). This layer provides annual maps of the IW seasonality from 1984 to 2021 based on the water occurrence values detected throughout each year. The "*Yearly History*" layer includes 38 images (one image for each year) under the attribute "*waterClass*," which categorizes each pixel into one of four classes: "*no observations*," "*not water*," "*seasonal water*," and "*permanent water*." For this analysis, only pixels identified as

"*seasonal water*" or "*permanent water*" were retained to avoid the inclusion of any misclassified IW pixels in the data set. The filtered data were collected for all 66 grids at four decadal intervals. Further, this data underwent preprocessing using the two-step methodology detailed in Section 2.1, which includes the creation of spatially connected polygons using the *BUDS* function. Once preprocessed, the IW samples for each year were labeled using the expert annotation group (described in Section 2.1). For this analysis, ADEMA was chosen over ML and DL models due to its robustness and lower computational requirements. The temporal accuracy of ADEMA was assessed using precision, recall, CE, OE, and the *F*1 score (Equations 5–9).

Additionally, we selected four different transition sites (TS1-TS4), beyond the initial 66 global locations, from two geographically and hydrologically contrasting regions to evaluate ADEMA's capability in capturing transitions between lentic and lotic system types. Transition sites TS1 and TS3 are located near the Aral Sea (in Kazakhstan), while TS2 and TS4 are situated in India, with TS2 in the Brahmaputra delta region of West Bengal and TS4 in the Indo-Gangetic plains of Uttar Pradesh. These locations were chosen as they represent complex hydro-geomorphic environments where transitions between lentic and lotic system types are common. For instance, large-scale irrigation-driven flow diversions in the Aral Sea region have disrupted hydrological connectivity, while monsoon-driven flooding and sediment delivery shape the floodplains of the Indian transition sites (Goodbred & Kuehl, 1999; Micklin, 2016; Raff et al., 2023; Yang et al., 2020). Each transition site was preprocessed using the *BUDS* function and subsequently labeled by our expert annotation group for each decadal interval between 1991 and 2021. ADEMA was then applied, and *F*1 scores were computed to assess classification performance over time.

2.8. Comparison With Global Classification Products

Since no global data sets enable a direct comparison with ADEMA, as a contextual benchmark, the performance of ADEMA was evaluated by comparing its outputs with two well-established global classification products from the Global Lakes data set (known as GLAKES) (Pi et al., 2022a) and the Reservoir and Lake Surface Area Time Series (known as ReaLSAT) data set (Khandelwal et al., 2022). The reasons for selecting these global classification products were: (a) GLAKES and ReaLSAT utilized the Global Surface Water data set (Pekel et al., 2016a) to identify IWs worldwide, similar to the current study, which used it as a base layer to generate lentic and lotic system type labels for ADEMA, ML, and DL model development, and (b) both GLAKES and ReaLSAT data sets converted pixel-based IWs from the Global Surface Water data set into a defined polygon, aligning with the *BUDS* function framework of the current study (see Section 2.1). Furthermore, to the best of the authors' knowledge, these studies are the only global-scale attempts that performed lentic and lotic system type classifications as part of their methodologies. Additionally, global IW products like the HydroLAKES (Messager et al., 2016), GRanD (Lehner et al., 2011), and GLWD (Lehner & Döll, 2004), while offering a comprehensive global inventory of IWs, do not explicitly apply river-removal methodologies, and were therefore not included in direct comparisons with ADEMA.

GLAKES data set maps 3.4 million lakes globally (60°S to 80°N) and includes IWs with an area greater than 3 ha. GLAKES developed a U-Net deep learning model to separate the lentic (lakes) and lotic (rivers) system types. Similarly, the ReaLSAT data set provides the location and monthly water extent of ∼0.6 million lakes and reservoirs of IW size greater than 10 ha, covering regions south of 50°N. To remove lotic system types from lentic system types, ReaLSAT applied an erosion-based thresholding operation with a fixed morphological score of 0.05, followed by the Ordering Based Information Transfer (ORBIT) algorithm (Khandelwal et al., 2017). In our comparison, polygons missing from these global classification products were treated as lotic system types. This interpretation was necessary to enable consistent and meaningful comparison with ADEMA, and was based on two considerations: (a) both the global classification products explicitly applied river-removal procedures, so omissions likely correspond to rivers or lotic system types, and (b) all three classification approaches (ADEMA, GLAKES, and ReaLSAT) use the Global Surface Water data set (Pekel et al., 2016a) as the base data set, which only includes IW pixels that could be either lentic or lotic, thus the assumption of missing polygons to be lotic is valid. Further, to ensure a comprehensive and fair comparison, these global classification products were evaluated using two approaches: (a) considering all IWs without applying any filtering (hereafter referred to as no filtering), and (b) evaluating only those IWs that meet the area thresholds defined by each global classification product (>3 ha for GLAKES and >10 ha for ReaLSAT) (hereafter referred to as with area filtering). To perform this analysis, the metrics used for the evaluation were precision, recall, CE, OE, and the *F*1 score (Equations 5–9). Since these are not direct comparisons of models (ADEMA vs. U-Net Deep learning model for GLAKES vs.

Table 3
Accuracy Assessment of ML, DL, and ADEMA in Defining Lentic and Lotic IWs

Size class	Model	Precision (%)	Recall (%)	$F1$ score (%)	dAUC (%)
SIWs	DT	87%	98%	92%	53%
	NN	86%	99%	92%	53%
	RF	85%	100%	92%	50%
	XG	88%	99%	93%	53%
	ADEMA	86%	100%	92%	54%
MIWs	DT	90%	100%	95%	50%
	NN	93%	99%	96%	55%
	RF	93%	99%	96%	50%
	XG	91%	100%	95%	58%
	ADEMA	94%	96%	95%	72%
LIWs	DT	67%	100%	80%	50%
	NN	92%	88%	90%	52%
	RF	67%	100%	80%	56%
	XG	85%	88%	87%	50%
	ADEMA	77%	65%	71%	63%
All Size	DT	86%	100%	92%	50%
	NN	87%	99%	93%	54%
	RF	86%	100%	92%	50%
	XG	87%	99%	93%	55%
	ADEMA	89%	99%	94%	55%

ORBIT for ReaLSAT), these metrics represent the overall completeness and usability of the classification, not just the performance of the core classifier. In this context, a method that identifies more system types (lentic or lotic) accurately is considered more valuable in our framework.

3. Results

3.1. Model Performance Across Various Size Classes

The performance of ML and DL models, along with ADEMA, was evaluated across three size classes using metrics such as precision, recall, $F1$ score (Overall), and dAUC (Table 3), including Commission Error (CE), Omission Error (OE), and the counts of True Lentic (TLe), False Lentic (FLe), True Lotic (TLo), False Lotic (FLo) , training $F1$ Score ($F1_{tr}$), and testing $F1$ score ($F1_{ts}$) (Table S5 in Supporting Information S1). In Small Inland Waters (SIWs), all models had demonstrated superior precision, recall, and $F1$ score of ~85–88%, ~98–100%, and ~92–93%, respectively. However, ADEMA achieved a moderate dAUC value of ~54%, which was slightly higher than DT, NN, RF, and XG, all of which ranged from ~50 to 53%. For Medium Inland Waters (MIWs), similar to SIWs, all models had shown superior precision, recall, and $F1$ scores of ~90–94%, ~96–100%, and ~95–96%, respectively. Nevertheless, ADEMA showed a relatively higher dAUC of 72% compared to DT (50%), NN (55%), RF (50%), and XG (58%). In the Large Inland Waters (LIWs) category, ADEMA faced challenges, achieving a precision of 77% and a recall of 65%, both of which were lower than NN (92% precision, 88% recall) and XG (85% precision, 88% recall) and recall in RF (100% recall) and DT (100% recall). However, ADEMA's precision is better than that of RF and DT, which were 67% each. Along with lower precision, ADEMA maintained an $F1$ score of 71%, which was lower than DT (80%), RF (80%), NN (90%), and XG (87%). Nonetheless, ADEMA's dAUC value of 63% was relatively higher than DT (50%), NN (52%), RF (56%), and XG (50%) in LIWs. In the All Size category, ADEMA maintained consistent performance with a precision of 89%, slightly higher than other models (at 86%–87%). The recall for ADEMA was 99%, matching NN and XG and slightly less than DT and RF (both at 100%). With an $F1$ score of 94%, ADEMA was on par with NN and XG and marginally better than DT and RF (both at 92%). ADEMA's dAUC value of 55% was on par with XG and higher than DT (50%), NN (54%), and RF (50%). Overall, ADEMA showed superior performance to other models among all size classes. In general, ADEMA and all ML and DL models have had comparable $F1_{tr}$ and $F1_{ts}$ to the overall $F1$ score under each size class (Table S5 in Supporting Information S1). Further, ADEMA retained an on par weighted $F1$ score (82%), comparable to ML and DL models (range: 79%–85%), even when accounting for class imbalance (Table S6 in Supporting Information S1).

Table 4
Average Explainability Scores (η_i) of Morphometric Variables From ML, DL, and ADEMA for Defining System Types Based on Size Class

Category	Type	CR (%)	CT (%)	ER (%)	LR (%)
Size Class (All Models)	SIWs	23	19	11	47
	MIWs	31	17	26	26
	LIWs	20	36	18	26
Model (All Size)	DT	37	8	1	54
	NN	27	19	17	37
	RF	17	42	25	16
	XG	30	20	32	18
	ADEMA	32	2	64	2

3.2. Explainability-Based Ranking of Variables in Defining Lentic and Lotic System Types

The average explainability scores of all four independent variables were analyzed across different size classes and models (Table 4; Table S7 in Supporting Information S1). For size classes, the explainability score of CR is highest for MIWs (31%), followed by SIWs (23%) and LIWs (20%), with an overall score of 29% for the All Size category. The CT demonstrates the highest explainability (η_i) for LIWs (36%), while SIWs and MIWs show relatively lower scores of 19% and 17%, respectively. The ER has its highest η_i for MIWs (26%), with LIWs (18%) and SIWs (11%) showing progressively lower contributions, leading to an All Size category explainability score of 28%. The LR demonstrates its highest η_i for SIWs (47%), significantly higher than MIWs and LIWs (both score 26%), with an overall All Size category value of 25%. While evaluating the models for the All Size category, in the DT model, LR exhibited the highest η_i (54%), followed by CR (37%). Further,

CT and ER contribute minimally, with the η_i of 8% and 1%, respectively. The NN model shows relatively balanced explainability across variables, with LR at 37%, CR at 27%, CT at 19%, and ER at 17%. In the RF model, CT has the highest η_i (42%), while CR, ER, and LR have η_i of 17%, 25%, and 16%, respectively. For the XG model, ER is the most important variable, with a η_i of 32%, followed by CR (30%), CT (20%), and LR (18%). Finally, in the ADEMA model, ER has the highest η_i (64%), CR scores 32%, and CT and LR contribute minimally at 2% each.

3.3. Temporal Invariance of ADEMA

Given ADEMA's robustness, computational efficiency, and high accuracy in classifying lentic and lotic system types, it was evaluated across decadal intervals to assess its temporal invariance. ADEMA consistently performed across all four decadal intervals, with higher $F1$ scores and lower error rates across 66 grids (Figure 3a; Table S8 in Supporting Information S1). The median $F1$ score remained high, with values of 93% in 1991 (range: 25%–100%), 91% in 2001 (3%–100%), 92% in 2011 (20%–100%), and 90% in 2021 (16%–100%). CE showed slight variability, with a median of 10% in 1991, increasing to 16% in 2001, 14% in 2011, and 18% in 2021. Conversely, the OE was negligible (median = 0%) across all years.

As part of this analysis, two diverse geographical locations (i.e., location-10 in Brazil and location-3 in Australia) were selected from the 66 locations to test the ADEMA's performance in regions undergoing significant changes in IW polygons (Figure 3b). In location-10 (Figure 3b), the total number of IW polygons decreased from 152 to 126 during 1991–2021, with lentic system polygons decreasing from 142 to 107 (Table S8 in Supporting Information S1), while the number of lotic system polygons increased from 10 to 19 during the same period. In contrast, in location-3 (Figure 3b), the total number of IW polygons doubled from 77 in 1991 to 141 in 2021. Specifically, the number of lotic system polygons slightly increased from 4 in 1991 to 7 in 2021, while lentic system polygons increased two-fold from 73 to 134 during the same period. In location-10, with decreasing IW polygons, the performance of ADEMA remains consistent, with high $F1$ ranging from 72% to 93% during 1991–2021. Similarly, in location-3, with increasing IW polygons, the $F1$ score ranged from 74% to 84%. Further, the ADEMA showed superior performance in capturing transitions between lentic and lotic system types across four transition sites (TS1–TS4). At TS1 (Aral Sea region), $F1$ scores were 91% (1991), 83% (2001), 83% (2011), and 94% (2021) (Figure 3c). TS3 (Aral Sea region) showed a decline from 100% (1991) to 92% (2001) and 88% (2011), followed by an improvement to 97% in 2021 (Figure 3c). TS2 (West Bengal region) remained clearly defined morphologically with consistent $F1$ scores of 100% across all years. TS4 (Uttar Pradesh region) recorded $F1$ scores of 100% (1991), 90% (2001), 92% (2011), and 89% (2021) (Figure 3c). These results support ADEMA's temporal invariance and its ability to capture system type transitions, reinforcing its reliability for long-term segregation of lentic and lotic IWs.

3.4. Performance of ADEMA and Global Classification Products

ADEMA demonstrated higher agreement with the expert-labeled data set than the global classification products (GLAKES and ReaLSAT) in defining lentic and lotic system types under both evaluation approaches (Figure 4; Figure S4 and Tables S9–S11 in Supporting Information S1). Under the first approach (no filtering), ADEMA achieved a median $F1$ score of 94% (range: 29%–100%) across all 66 global locations. In comparison, GLAKES and ReaLSAT showed much lower median $F1$ scores of 12% (0.2%–52%) and 5% (1%–44%), respectively. Following the second approach (area filtering), 55 global locations were used for GLAKES and 34 for ReaLSAT, excluding locations not within their intended scope (>3 ha for GLAKES and >10 ha for ReaLSAT). Under this approach, the median $F1$ scores improved for both global classification products, with GLAKES reaching 50% (range: 7%–100%) and ReaLSAT reaching 74% (13%–100%). Nevertheless, ADEMA continued to perform better than both global classification products under GLAKES (>3 ha) and ReaLSAT (>10 ha) area thresholds, achieving higher median $F1$ scores of 96% (63%–100%) and 97% (67%–100%), respectively. Overall, ADEMA maintained higher average $F1$ scores (94% and 97%) under the no filtering and area filtering approaches, compared to 9% and 62% for the global classification products.

In terms of precision and recall, ADEMA showed a remarkable performance under both evaluation approaches. Under no filtering (Table S9 in Supporting Information S1), ADEMA achieved a median precision of 89% and a median recall of 100%. While GLAKES and ReaLSAT also recorded high precision values with a median of 100% for both, their recall values were considerably lower, with medians of 6% and 3%, respectively. ADEMA

10.1029/2025WR040137

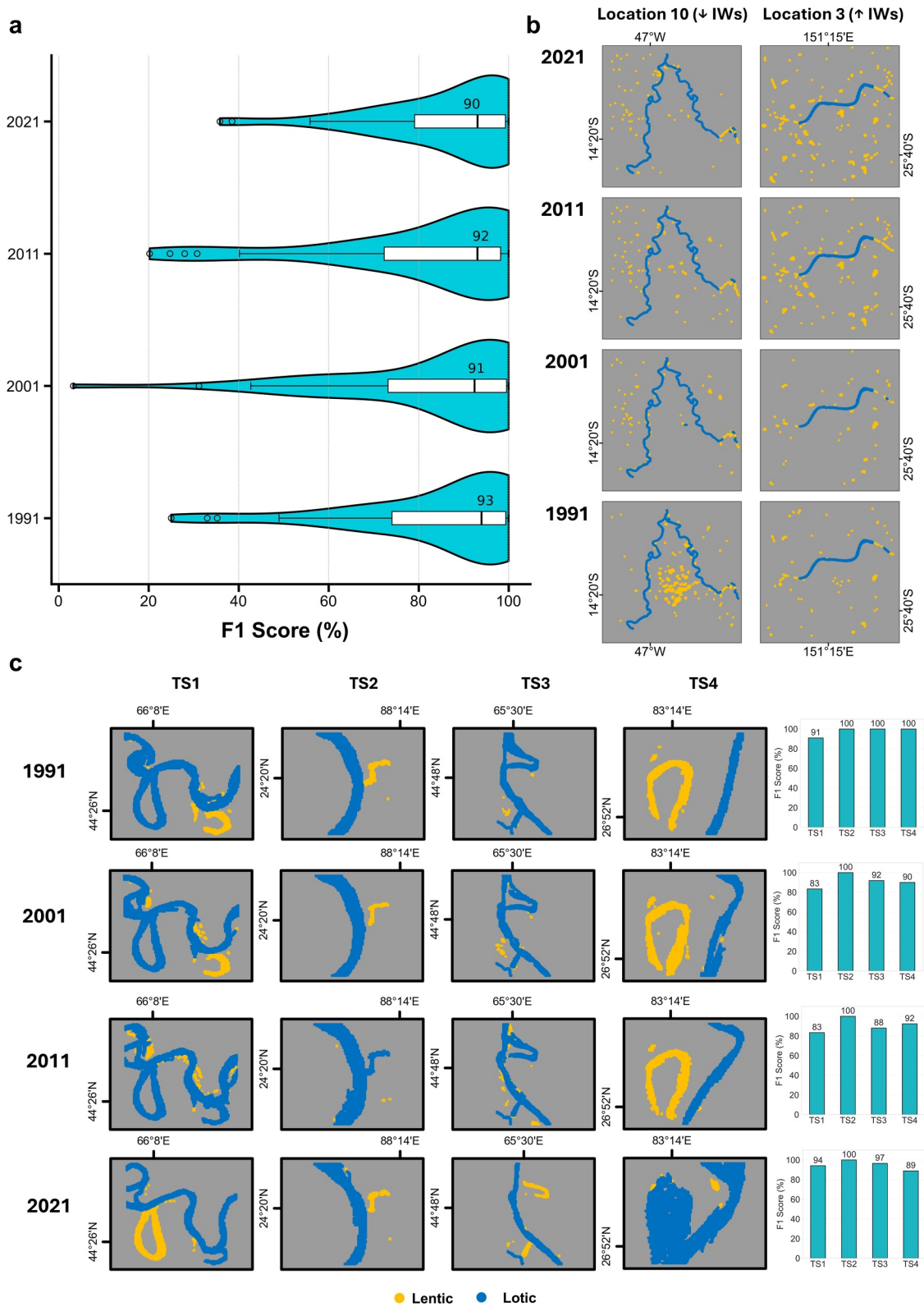

Figure 3. Temporal evaluation of ADEMA's classification performance and ability to capture transitions between lentic and lotic system types from 1991 to 2021. Violin plots (a) showing the distribution of *F*1 scores across 66 locations for each decadal interval. Spatial classification results (b) for two representative locations: one with a consistent decline in IWs (location 10) and another with a gradual increase (location 3). Transition sites (TS1–TS4) showing shifts (c) between lentic and lotic system types over time, along with the corresponding *F*1 scores for each year. Lentic and lotic system types are shown in gold and dark blue, respectively.

10.1029/2025WR040137

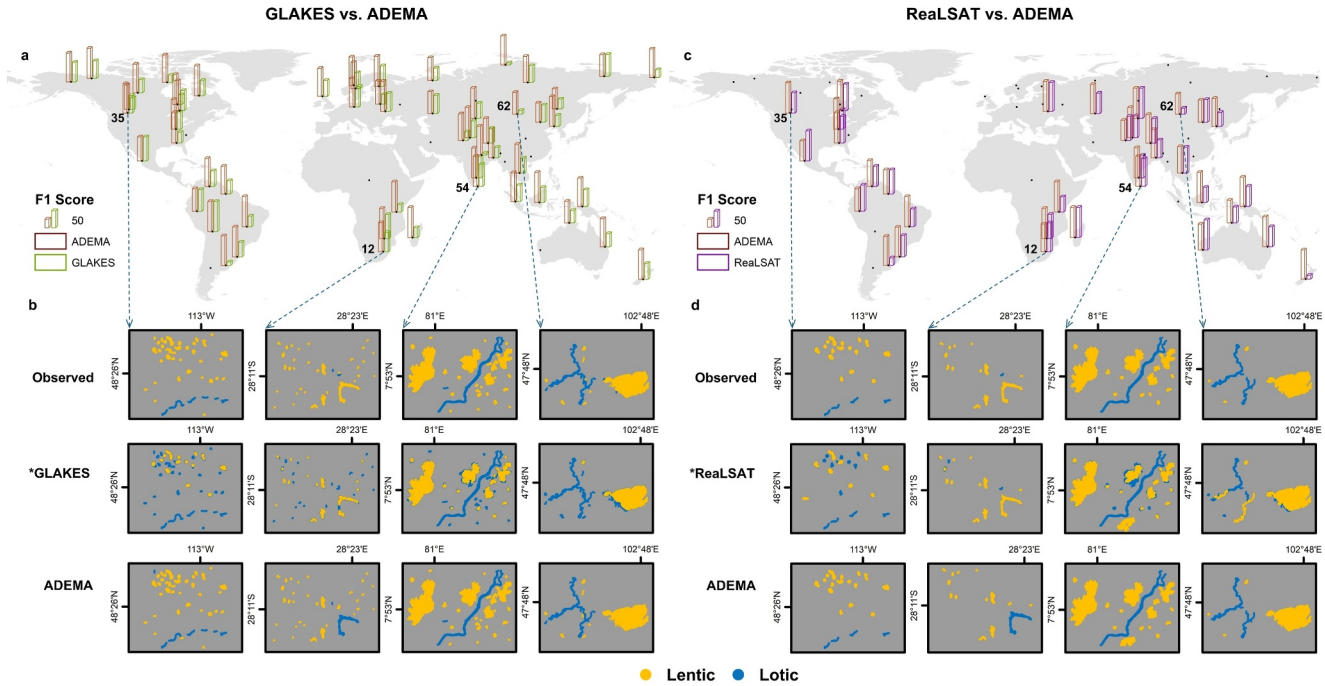

Figure 4. Comparison of ADEMA against GLAKES and ReaLSAT for identifying system types under the area filtering approach (>3 ha for GLAKES and >10 ha for ReaLSAT). In (a, c), bar charts of *F*1-scores are shown across global locations for GLAKES, ReaLSAT, and ADEMA. Maps in (b, d) show expert-annotated labels (top row) for four example locations (IDs 35, 12, 54, 62, respectively) and results from GLAKES (middle row in b), ReaLSAT (middle row in d), and ADEMA (bottom row), with lentic system types in gold and lotic system types in dark blue. Black dots indicate global locations not within the intended scope of global classification products (refer to Figure S4 in Supporting Information S1 for the no filtering based comparison). *Note: missing polygons in ReaLSAT and GLAKES are considered lotic system types for contextual benchmarking.

effectively minimized errors, with a low CE (median = 11%) and negligible OE (median = 0%), reducing both false lentic and lotic system types. In contrast, although GLAKES and ReaLSAT recorded a negligible CE (median = 0%), their OEs were high, with GLAKES achieving a median OE of 94% and ReaLSAT getting a median OE of 97%.

With area filtering (Tables S10 and S11 in Supporting Information S1), ADEMA maintained superior performance, achieving a median precision of 96% and recall of 99% for the GLAKES threshold, and both precision and recall of 100% for the ReaLSAT threshold. GLAKES maintained a high precision (100%) but a low recall of 33%. While the CE for GLAKES was negligible, it achieved a high OE of 67%. On the other hand, ReaLSAT showed 100% precision and moderate recall (59%), with no CE and a high OE of 41%.

In specific locations, under the no filtering (Figure S4a in Supporting Information S1), ReaLSAT and GLAKES showed lower accuracy than ADEMA. In location-35 (in the United States), location-12 (in South Africa), and location-54 (in Sri Lanka), ReaLSAT and GLAKES achieved low *F*1 scores in the range of 10%–17% and 13%–26% (Table S9 in Supporting Information S1), respectively. The lowest *F*1 Score from these global classification products was recorded at location 62 (in Mongolia), with a value of 1% (Figures S4j–S4q and Table S9 in Supporting Information S1). In contrast, ADEMA demonstrated superior performance in location-35, location-12, and location-54 with a high *F*1 score ranging between 94% and 99% (Figures S4e, S4i, S4m, and Table S9 in Supporting Information S1). Similarly, in location 62, the *F*1 score dropped to 74% due to a few lotic misclassifications (Figure S4q in Supporting Information S1), though it remained significantly better than both GLAKES (Figure S4p in Supporting Information S1) and ReaLSAT (Figure S4o in Supporting Information S1).

Under area filtering, the global classification products performed better than those with no filtering. For example, at location 12, ReaLSAT slightly performed better than ADEMA (*F*1 scores: 96% vs. 92%), while GLAKES had a moderate *F*1 score of 64% (Figures 4a–4d; Tables S10 and S11 in Supporting Information S1). At location 54, both ADEMA and ReaLSAT performed well (*F*1 scores: 100% and 92%, respectively), whereas GLAKES achieved a moderate *F*1 score of 73%. However, across most locations (>74% of locations for ReaLSAT and

>95% for GLAKES), ADEMA consistently showed a higher $F1$ score compared to both global classification products, even within their intended scope (Figures 4a and 4c). For instance, at location 35, ADEMA achieved $F1$ scores of 97% and 100%, compared to 58% for GLAKES and 67% for ReaLSAT. At location 62 (Mongolia), both GLAKES and ReaLSAT continued to perform poorly, with $F1$ scores of 12% and 20%, respectively, while ADEMA maintained moderate scores of 72% and 67% respectively, under GLAKES and ReaLSAT thresholds (Figures 4b and 4d; Tables S10 and S11 in Supporting Information S1). These findings highlight ADEMA's consistent accuracy and reliability in defining lentic and lotic system types, with higher performance than GLAKES and ReaLSAT in both evaluation approaches and across all metrics.

4. Discussion

4.1. ADEMA Versus ML and DL

ADEMA's performance across size classes demonstrates its clear advantages over traditional comprehensively optimized ML and DL models such as DT, NN, RF, and XG, with a few exceptions. These exceptions, such as a high recall score of ML and DL models (Table S5 in Supporting Information S1), are major factors in misclassifying IWs as lentic. For instance, the RF model for SIWs shows 100% recall despite having a moderate dAUC because all the lotic SIWs are misclassified as lentic (Table S5 in Supporting Information S1). Although dAUC values are reasonable for SIWs, ADEMA's on par performance with ML and DL models supports its suitability for SIWs. Based on the dAUC, ADEMA is relatively better in MIWs and LIWs and is on par in the All Size categories compared to ML and DL models. The main reason for disparities in dAUC is hypothesized as the data set size (while comprehensive) may not be enough for ML and DL models to derive intricate patterns among independent variables (Rather et al., 2024) and system types to achieve the same level of performance as ADEMA. One of the key reasons for ADEMA's performance may be that the framework of ADEMA uses a heuristic approach based on multi-dimensional morphometric interpretation, which is decided based on expert judgment and general physics of the system types. This ensures accurate, data-efficient, and explainable classifications by ADEMA.

Further, IWs often exhibit a wide range of shapes, from highly circular to elongated or compact forms (Luo et al., 2021; Van Der Werff & Van Der Meer, 2008), providing clear demarcations for ADEMA's decision boundaries to capture intricate distinctions between system types effectively. For instance, its ability to classify mildly circular IWs based on elongation (ER > 0.5) or compactness (CT ≥ 3.5) ensures accurate differentiation between lentic and lotic system types. In contrast, traditional ML and DL models rely on automatically derived thresholds that can be skewed by mixed regions, resulting in poorer classifications of system types (Luo et al., 2021; Martínez-Santos et al., 2021; Pi et al., 2022a). Although ADEMA has a moderate $F1$ score in classifying LIWs, its dAUC is slightly better than ML and DL models, indicating that the drop in $F1$ for large IWs is mainly due to a lower number of true lentic cases in that category (Table S5 in Supporting Information S1).

Since the class imbalance is representative of natural lentic and lotic systems distribution, there were no explicit preprocessing techniques performed for balancing the distribution of system types (e.g., class weights) during ML and DL models optimization (Section 2.4). Moreover, ADEMA and all ML and DL models showed similar training, testing, and overall $F1$ score under each size class (Table S5 in Supporting Information S1), showing that the optimization has made generalized models. However, to give a more unbiased evaluation, weighted $F1$ scores were computed on the training data set and testing to assess the effect of class imbalance. Even under class imbalance conditions, ADEMA achieved a weighted $F1$ score that was comparable to those obtained by ML and DL models (Table S6 in Supporting Information S1).

ADEMA also outperforms other models in terms of computational efficiency. Traditional ML and DL models like RF, NN, and XG require significant computational resources to optimize hyperparameters and train on complex data sets (Thompson et al., 2020; Zhang et al., 2017). In contrast, ADEMA does not need these computational demands, as it employs a direct classification based on pre-defined decision boundaries. This efficiency makes ADEMA scalable and well-suited for applications in resource-constrained environments where computational power and data availability are often limited. Therefore, ADEMA is recommended for large-scale classifications of lentic and lotic system types.

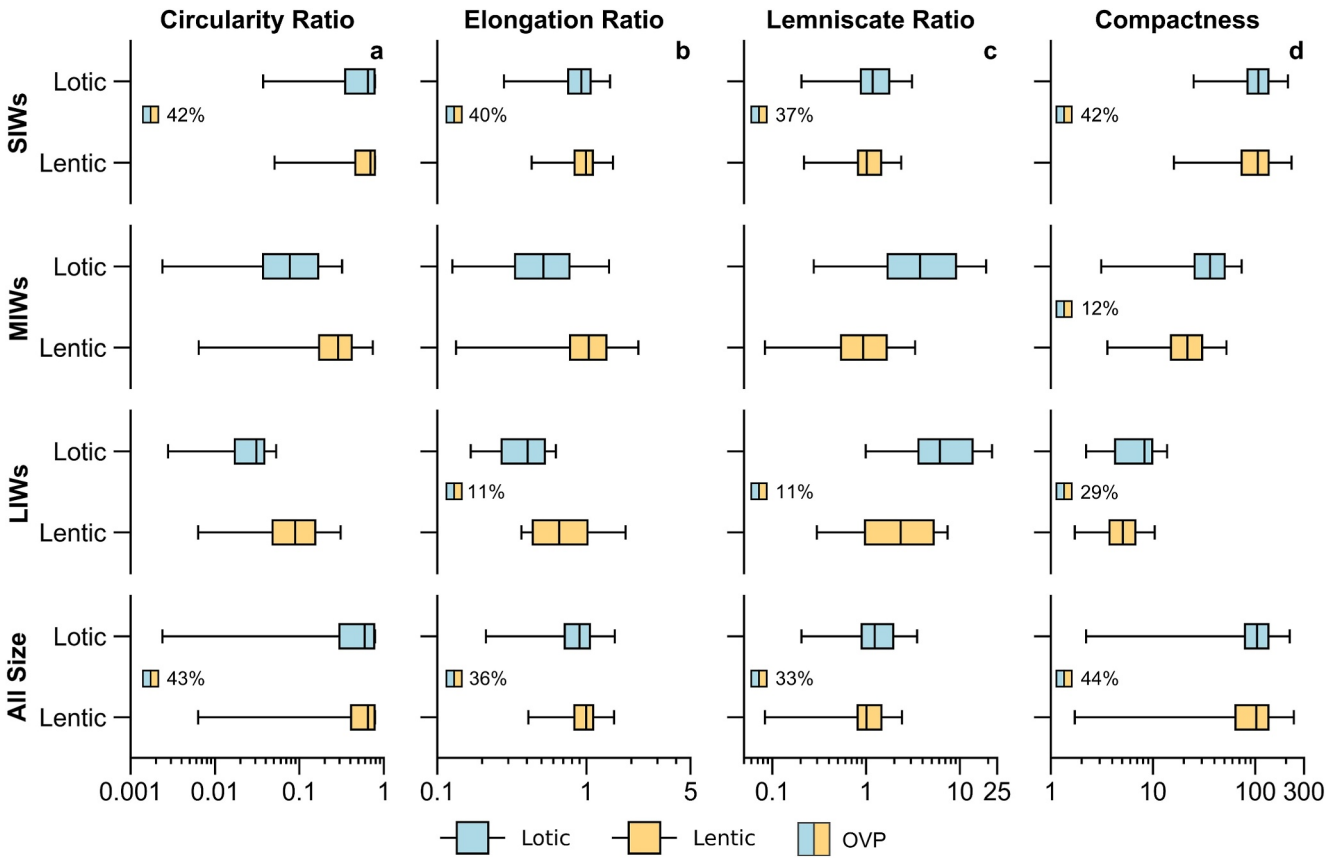

Figure 5. Size class-wise independent variable distributions with OVP values in lentic and lotic system types across 66 locations of the study. Subplots (a–d) represent distributions in circularity ratio, elongation ratio, lemniscate ratio, and compactness, respectively.

4.2. Explainability-Based Ranking of Independent Variables

The η_i reveals the influence of independent variables in defining lentic and lotic system types. These scores reflect the underlying morphometric characteristics of the system types as well as how different size classes and models use these variables for classification. Since there is a lack of literature explaining the importance of morphometric variables in system type classifications, this study developed Figure 5 and used Equation 14 (% overlap—OVP). These representations of independent variable distributions (Figure 5) justify why certain variables have greater η_i among size classes and models.

$$\mathrm{OVP} = \left(\frac{Q3_j - Q1_i}{Q3_i - Q1_i + Q3_j - Q1_j}\right) \times 100 \tag{14}$$

where, OVP is the percentage overlap between interquartile range (IQR), i is the variable with the larger IQR, and j is the variable with the smaller IQR. Q1 and Q3 are the first and third quartiles in each variable i or j.

Among size classes, the CR shows a notably high η_i for MIWs, likely due to the significant disparity in the IQR of CR values for lentic (IQR: 0.17–0.41) and lotic system types (IQR: 0.04–0.17), with no OVP within this category (Figure 5a). Similarly, the ER demonstrates its peak contribution among MIWs with declining importance for LIWs and SIWs. This pattern arises because of no OVP between lentic and lotic system types (Figure 5b), unlike in LIWs (11% OVP) and SIWs (40% OVP), where the ranges are overlapping. Intriguingly, the LR has the highest η_i when classifying SIWs compared to MIWs and LIWs (Table 4), due to high irregularities in morphometry in the SIWs category (Read & Rose, 2013; Winslow et al., 2015). Moreover, there are relatively distinct variations in median values of LR for SIWs (lentic: 1.01 and lotic: 1.17) compared to other independent variables (Figure 5c). In contrast, CT highly explains classifications in LIWs (Table 4), as this size class has relatively defined

boundaries and less irregular shapes (Anandharuban et al., 2019; Yao et al., 2023). Furthermore, the medians of CT in LIWs are significantly different between lentic and lotic system types, with values of 5.08 and 8.24, respectively (Figure 5d).

When comparing ML and DL models with ADEMA in the All size category, the η_i assigned to each independent variable shifts appreciably. These disagreements in η_i are common and seen in a previous study (Martínez-Santos et al., 2021), which found that support vector machine models rely on altitude, thermal floor, and classic landforms. Martínez-Santos et al. (2021) also report that RF favors morphometry, water table depth, and stream power index for classifying lentic and lotic system types (aquatic ecosystems). In the present study, for the DT model, LR attains the highest η_i (Table 4), which could be due to less OVP (33%) compared to other morphometric variables (36%–44%) (Figure 5). After LR, CR contributes the highest due to greater differences in the median values of lentic and lotic system types (CR in lentic vs. lotic: 0.65 vs. 0.59). These outcomes reflect that DT favors variables that offer clear demarcations in IW systems, which is consistent with its algorithm (Breiman et al., 1984). Further, these η_i's indicate that IW curvature (CR and LR) plays a major role for DT in classifying system type and may be more suitable for classifying SIWs (due to higher η_i to LR).

Aside from DT, the NN model ranks independent variables in the following order: LR > CR > CT > ER, indicating that NN models are suitable for classifying SIWs and MIWs (Table 4). The reason for such preference for LR may be because of the skewed distribution of IW samples, with SIWs having the most numbers, followed by MIWs and LIWs (Naik et al., 2025). Conversely, XG ranks ER > CR > CT > LR, indicating that XG is more suitable for classifying MIWs (Table 4). This preference of ER is hypothesized as optimization of XG's hyperparameters, likely making it more physics-informed, subtly capturing patterns aligned with ADEMA's expert-labeled morphometric insights. The RF model identifies CT as the most influential variable in the All Size category, favoring LIWs over MIWs and SIWs. The preference for CT is attributed to the RF model's hyperparameter optimization, which is influenced by the distribution of IWs and, as a result, favors LIWs over other size classes. Lastly, the ADEMA places significant emphasis on ER, with moderate importance on CR, while CT and LR remain comparatively minor contributors (Table 4). The higher importance of ER and CR enables ADEMA to effectively identify elongated and round IW systems, such as fragmented rivers or elongated reservoirs. The lower CT and LR ranks in ADEMA are consistent with the hierarchical structure (CR > ER > LR > CT) (Figure 2d). Therefore, ADEMA is suitable for classifying IWs due to their higher performance (refer to Section 3.1).

4.3. Temporal Invariance of ADEMA

ADEMA's overall high $F1$ score and consistently low OE and CE across four decades highlight its temporal invariance in defining system types. The main reason for its temporal robustness is that ADEMA is developed based on the expert-labeled data set collected from GSMW. The GSMW data set provides an aggregated history of seasonal variations, long-term trends, and changes in lentic and lotic systems over 38 years. ADEMA's temporally consistent performance across the 66 locations shows its suitability for long-term monitoring of system type dynamics and trends. For instance, in location-3 (Figures S5a–S5e in Supporting Information S1), ADEMA performed well despite the significant increase in lentic polygons alongside a minor increase in lotic polygons, demonstrating its consistent performance amid increasing IW polygons. Similarly, in location-10 (Figures S5f–S5j in Supporting Information S1), ADEMA effectively captured the decline in lentic polygons and the rise in lotic polygons. The strong performance in these locations is largely due to the presence of continuous, elongated lotic polygons, which are morphometrically distinct from lentic system types, reducing the potential for misclassification. However, temporal variations in $F1$ scores across a few regions reveal limitations in accurately delineating system type polygons. For instance, ADEMA underperformed in location-37 (in China), where $F1$ scores remained consistently low (1991: 33%, 2001: 31%, 2011: 25%, 2021: 36%). This reduced accuracy is attributed to the dominance of a very small and fragmented lotic system type resembling lentic morphology (Figures S5k–S5o in Supporting Information S1).

In system type transition sites (TS1–TS4), ADEMA showed consistently high performance, maintaining $F1$ scores above 83% across all decadal intervals. At TS1 and TS3 (Aral Sea region), dynamic transitions, including lentic-to-lotic and lotic-to-lentic shifts, led to moderate performance drops during active transition phases, followed by improvements as system type boundaries became more clearly defined. In contrast, TS2 (West Bengal region) offered ideal conditions for classification due to the persistent spatial separation of system types, resulting

in perfect classification across all years. TS4 (Uttar Pradesh region) presented a scenario where a lotic channel gradually merged with a lentic system type, increasing morphometric complexity and leading to a relatively slight drop in classification performance of ADEMA compared to TS2. This shows ADEMA's ability to adapt to system type transitioning regions. These outcomes highlight ADEMA's robustness in reliably capturing both stable and transitioning inland water systems over time, indicating its suitability for assessing long-term changes in global lentic and lotic system types.

4.4. ADEMA Versus Global Classification Products

While the comparison with GLAKES and ReaLSAT assumes missing IWs as lotic, ADEMA's higher *F*1 scores, precision, recall, and lower error rates (both CE & OE) across both evaluation approaches (no filtering and area filtering) highlight its reliability and robustness in defining the system types on a global scale. With a high *F*1 score, ADEMA significantly improves classification accuracy compared to GLAKES and ReaLSAT, even within their intended scope (>3 ha for GLAKES and >10 ha for ReaLSAT). The lower accuracy of GLAKES under the no filtering can be attributed to two main factors: (a) moderate average CE (∼29%) and high average OE (∼58%) reported in the original GLAKES classification product by Pi et al. (2022a) (vs. negligible CE but very high OE (∼94%) in the present study) when detecting SIWs, and (b) not including the IWs smaller than 3 ha in the data set. Similarly, ReaLSAT's reliance on a single morphological score (0.05) also created challenges in distinguishing narrow and compact lentic IWs from lotic system types, further reducing classification accuracy. Additionally, ReaLSAT focused on IWs larger than 10 ha, contributing to its lower *F*1 score.

Upon comparison, within their intended scope, global classification products still showed weaker performance than ADEMA. This persistence of lesser performance highlights that both global classification products could not accurately identify IW system types even under area filtering. In contrast, ADEMA effectively addresses these limitations by employing a multi-dimensional morphometric interpretation of variable spaces, enabling it to classify lentic and lotic system types up to 0.09 ha (33 times finer than previous global studies). ADEMA's high recall and precision reflect its ability to identify these systems accurately. While GLAKES and ReaLSAT achieved high precision values, their recall was extremely low, leading to a high rate of false lotic classification and exclusion of many lentic and lotic system types. ADEMA's lower CE and OE further highlight its robustness in minimizing both false lentic and false lotic system types. By comparison, while GLAKES and ReaLSAT have negligible CE, their significantly higher OE values demonstrate their inability to detect a significant proportion of IWs globally. These results not only demonstrate the size limitations of existing global classification products but also highlight how the ADEMA would fill the gaps by including these SIWs, offering consistent and accurate classification across a broader range of IW sizes.

ADEMA's adaptability across diverse geographic and hydrological conditions is further evidenced by its strong performance under both evaluation approaches in locations where GLAKES and ReaLSAT underperformed. For instance, in South Africa (location-12) and Sri Lanka (location-54), where GLAKES and ReaLSAT showed limited performance under the no filtering approach, ADEMA maintained high *F*1 scores. These results reflect its ability to capture the local complexities of hydrological systems, including fragmented and seasonally variable IWs that GLAKES and ReaLSAT misclassified. Upon applying area filtering, ReaLSAT showed marginally better performance than ADEMA at a few locations (location-12 and location-54), suggesting that its classification is more effective when limited to its intended size scope. However, GLAKES continued to show lower performance after area filtering. In the United States (location-35), ADEMA maintained high classification accuracy under both evaluation approaches, while both global classification products showed reduced accuracy. Even in Mongolia (location-62), where the *F*1 score dropped to ∼70% in both evaluation approaches, ADEMA still significantly performed better than both global classification products, demonstrating its ability to mitigate classification errors. ADEMA is more data efficient by using just 17,391 samples compared to GLAKES's 161,682 samples (∼10 times more). Although ReaLSAT used 5,000 samples for validation, the *F*1 score per sample was significantly lower than ADEMA (ReaLSAT: 0.001% vs. ADEMA: 0.005%), indicating ADEMA is more data efficient. By addressing the limitations of global classification products, ADEMA provides a more refined and valuable method for filling a critical gap in global hydrography through system-type classification.

4.5. Limitations and Future Recommendations

While the ADEMA framework significantly improves the classification of IWs, certain limitations persist. These limitations originate largely from computational constraints and data set characteristics. To optimize the computational efficiency of ADEMA in GEE, a grid size of 1° × 1° was chosen. This decision was informed by extensive testing of smaller (0.25° × 0.25°, 0.5° × 0.5°) and larger grid sizes (2° × 2°, 3° × 3°, and 4° × 4°), which either failed to export or required excessive processing time. Despite the 1° × 1° grid size offering a practical balance between computational feasibility and classification accuracy, it limits the algorithm's ability to fully capture larger IWs. Nevertheless, the selected grid size ensures that a substantial portion of IWs can still be effectively classified within the existing framework. Additionally, a 10,000 ha threshold was employed to reduce the misclassification of large lotic and lentic system types within the 1° × 1° grid. While this threshold excluded IWs larger than 10,000 ha, it was essential for maintaining the computational viability of the framework. Nonetheless, this threshold adequately represents global IWs as most inland waterbodies are below 10,000 ha (Downing et al., 2006; Messager et al., 2016; Pi et al., 2022a; Verpoorter et al., 2014). With greater computational resources, the 1° × 1° grid structure could potentially be expanded to include larger IWs (>10,000 ha). Although ADEMA demonstrates the ability to classify IWs as small as a single LANDSAT pixel (∼0.09 ha), morphometric reliability at this scale may be compromised by sensor noise, ephemeral flooding, fragmented systems, or false detections (∼1%) in the GSMW data set. Further limitations of the GSMW data set pertain to its inability to recognize water surfaces below vegetation, such as those found in wetlands (dense canopies, marshes, and floating vegetation). Therefore, an elbow detection method (Figure S6 in Supporting Information S1; minimum area threshold) was performed to identify a more meaningful area threshold that accounts for these uncertainties. This elbow detection method revealed that classification accuracy ($F1$ score) for ADEMA stabilizes at a minimum size of ∼0.5 ha (>5 LANDSAT pixels), offering a more reliable morphometric representation and accounting for the aforementioned uncertainties of the GSMW data set (more details in Text S2 in Supporting Information S1). While ADEMA still performs well at the 0.09 ha threshold, researchers should employ supplementary validation methods, such as temporal consistency checks, multi-temporal image analysis, or higher-resolution ancillary data sets, to confirm the reliability of single-pixel detections. To further address this, future work should consider deriving data sets that integrate both optical and Synthetic Aperture Radar (SAR) observations for detecting water surfaces below vegetation. In this context, advanced AI-based models such as DeepAqua (Peña et al., 2024), which employ cross-modal learning by using optical-derived water masks to train SAR-based models, offer improved detection of both open and vegetated water surfaces. The outputs generated from such models could provide more comprehensive inputs and enhance ADEMA's classification performance in wetland-dominated regions.

Another challenge in the existing framework is misclassifying fragmented lotic (river) segments and elongated lentic (long reservoirs, lakes, etc.) systems. The GSMW layer includes small river fragments with morphometry resembling lentic system types due to their short length and greater width, which led to their misclassification. Similarly, long reservoirs and lakes with elongated shapes and narrow widths were occasionally misclassified as lotic by the algorithm (Figure S7 in Supporting Information S1). Despite these limitations, high $F1$ scores, negligible OE, and low CE demonstrate that such misclassifications constitute a relatively small percentage of cases (<1% of the data set). Furthermore, incorporating the *BUDS* function within the framework has already addressed the issue of fragmented IWs to some extent. While fragmentation still poses challenges, future enhancements of the *BUDS* algorithm, such as integrating topographical data sets like digital elevation models, can improve connectivity among river fragments.

Finally, another limitation is that AUC is primarily designed for probabilistic outputs, and its application to deterministic models may not be ideal. In this study, dAUC was used due to the absence of equally robust alternatives for evaluating deterministic outputs. Accordingly, dAUC-based interpretations should be viewed as secondary. Developing equally robust evaluation metrics for deterministic models remains an important direction for future research. Overall, the ADEMA framework demonstrates significant advancements in IW system type classification with scope for future refinements. Continuous refinements addressing computational and dataset-related limitations will further enhance its robustness and applicability on a global scale.

5. Conclusions

The ADEMA framework offers a robust, efficient, and scalable solution for globally classifying lentic and lotic system types across diverse size classes and temporal scales. Its superior performance, as demonstrated through comprehensive comparisons with traditional ML and DL models (DT, NN, RF, and XG) and global classification products (GLAKES and ReaLSAT), establishes its reliability and accuracy. ADEMA consistently delivers high precision, recall, and *F*1 scores, particularly excelling in small and medium-sized inland waters while maintaining competitive performance in large inland waters. ADEMA's use of multi-dimensional morphometric interpretations, guided by expert judgment, ensures explainable and data-efficient classifications. The temporal invariance of ADEMA across decades underscores its robustness for long-term monitoring of inland water systems, adapting seamlessly to changes in hydrological and geographic conditions. Its consistent performance across diverse global locations further highlights its adaptability and suitability for various hydrological, limnological, and ecological contexts. Importantly, ADEMA achieves these results with significantly lower computational demands, making it particularly advantageous for resource-constrained environments. While ADEMA demonstrates substantial advancements, its limitations, such as the exclusion of IWs larger than 10,000 ha and challenges with fragmented or elongated inland water systems, indicate areas for future improvement. Addressing these challenges through enhanced computational resources, more accurate data sets, and the integration of topographical data will not only improve ADEMA's accuracy but also strengthen its scalability and reduce classification errors.

Appendix A: List of Descriptions of Abbreviations and Acronyms Used in This Study

Abbreviation/Acronym	Description
ADEMA	Automated Data Efficient Morphometric Approach
BUDS	BUfferDissolveSplit
JEHLL	Jack-knifing for Explaining Hierarchical variables in Lentic and Lotic system type classification
CR	Circularity Ratio
LR	Lemniscate Ratio
ER	Elongation Ratio
CT	Compactness
DT	Decision Tree
NN	Neural Network
RF	Random Forest
XG	eXtreme Gradient Boosting
IWs	Inland Waters
SIWs	Small Inland Waters
MIWs	Medium Inland Waters
LIWs	Large Inland Waters
GSMW	Global Surface Maximum Water Extent
GLAKES	Global Lakes (Pi et al., 2022a)
ReaLSAT	Reservoir and Lake Surface Area Time Series (Khandelwal et al., 2022)
CE	Commission Error
OE	Omission Error
dAUC	Discrete Area Under Curve
ST	System Type
GEE	Google Earth Engine
SC	Size Class

Appendix A
Continued

Abbreviation/Acronym	Description
ML	Machine Learning
DL	Deep Learning
OVP	% Overlap
IQR	Interquartile Range
IoU	Intersection Over Union
GRanD	Global Reservoir and Dam (Lehner et al., 2011)
GLWD	Global Lakes and Wetlands Database (Lehner & Döll, 2004)
ORBIT	Ordering Based Information Transfer

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

All data sets and code used in this study are available in the Zenodo repository (Sharma et al., 2026). The following are the references to the publicly available data sets, codes, and software:

1. Global Surface Maximum Water extent (GSMW) layer and yearly water history layer (Pekel et al., 2016b).
2. Global Lakes (GLAKES) data set (Pi et al., 2022b).
3. Reservoir and Lake Surface Area Time series (ReaLSAT) data set (Khandelwal et al., 2023).
4. Scikit-learn (Pedregosa et al., 2011).
5. Scikit-optimize (Head et al., 2021).
6. Google Earth Engine (Gorelick et al., 2017).

Acknowledgments
All authors thank the Greenhouse Gas Emissions and Nutrients (GeN) lab members, Department of Water Resources Development and Management, Indian Institute of Technology Roorkee, for assisting in labeling lentic and lotic system types. The maps generated in this study were created using ArcGIS version 10.7.1, licensed under the Institute Computer Centre, Indian Institute of Technology Roorkee. We thank the two anonymous reviewers, the associate editor, and the editor (Prof. Georgia Destouni) for providing suggestions that significantly enhanced the manuscript.

References

Allen, G. H., & Pavelsky, T. M. (2018). Global extent of rivers and streams. *Science*, *361*(6402), 585–588. https://doi.org/10.1126/science.aat0636
Anandharuban, P., La Rocca, M., & Elango, L. (2019). A box-model approach for reservoir operation during extreme rainfall events: A case study. *Journal of Earth System Science*, *128*(8), 229. https://doi.org/10.1007/s12040-019-1258-7
Andersen, D. C., & Shafroth, P. B. (2010). Beaver dams, hydrological thresholds, and controlled floods as a management tool in a desert riverine ecosystem, Bill Williams River, Arizona. *Ecohydrology*, *3*(3), 325–338. https://doi.org/10.1002/eco.113
Błędzki, L., & Ellison, A. (2000). Effects of water retention time on zooplankton of shallow rheolimnic reservoirs. *Internationale Vereinigung Für Theoretische Und Angewandte Limnologie: Verhandlungen*, *27*(5), 2865–2869. https://doi.org/10.1080/03680770.1998.11898192
Breiman, L. (2001). Random forest. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324
Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Routledge. https://doi.org/10.1201/9781315139470
Brown, C. F., Brumby, S. P., Guzder-Williams, B., Birch, T., Hyde, S. B., Mazzariello, J., et al. (2022). Dynamic world, near real-time global 10 m land use land cover mapping. *Scientific Data*, *9*(1), 251. https://doi.org/10.1038/s41597-022-01307-4
Cael, B. B., Heathcote, A. J., & Seekell, D. A. (2017). The volume and mean depth of Earth's lakes. *Geophysical Research Letters*, *44*(1), 209–218. https://doi.org/10.1002/2016GL071378
Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). ACM. https://doi.org/10.1145/2939672.2939785
De Lucia Lobo, F., Novo, E. M. L. D. M., Barbosa, C. C. F., & Galvão, L. S. (2012). Reference spectra to classify Amazon water types. *International Journal of Remote Sensing*, *33*(11), 3422–3442. https://doi.org/10.1080/01431161.2011.627391
Dodson, S. I., Arnott, S. E., & Cottingham, K. L. (2000). The relationship in lake communities between primary productivity and species richness. *Ecology*, *81*(10), 2662–2679. https://doi.org/10.1890/0012-9658(2000)081[2662:TRILCB]2.0.CO;2
Downing, J. A., Cole, J. J., Middelburg, J. J., Striegl, R. G., Duarte, C. M., Kortelainen, P., et al. (2008). Sediment organic carbon burial in agriculturally eutrophic impoundments over the last century. *Global Biogeochemical Cycles*, *22*(1), 2006GB002854. https://doi.org/10.1029/2006GB002854
Downing, J. A., Prairie, Y. T., Cole, J. J., Duarte, C. M., Tranvik, L. J., Striegl, R. G., et al. (2006). The global abundance and size distribution of lakes, ponds, and impoundments. *Limnology & Oceanography*, *51*(5), 2388–2397. https://doi.org/10.4319/lo.2006.51.5.2388
Fisher, W. L., Bozek, M. A., Vokoun, J. C., & Jacobson, R. B. (2012). Freshwater aquatic habitat measurements. In *Fisheries Techniques, 3rd Edition. American Fisheries Society, Bethesda, Maryland* (pp. 101–161).
Goodbred, S. L., Jr., & Kuehl, S. A. (1999). Holocene and modern sediment budgets for the Ganges-Brahmaputra river system: Evidence for highstand dispersal to flood-plain, shelf, and deep-sea depocenters. *Geology*, *27*(6), 559–562. https://doi.org/10.1130/0091-7613(1999)027<0559:hamsbf>2.3.co;2

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, *202*, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Head, T., Kumar, M., Nahrstaedt, H., Louppe, G., & Shcherbatyi, I. (2021). scikit-optimize/scikit-optimize (version v0.9.0) [Software]. *Zenodo*. https://doi.org/10.5281/ZENODO.1157319

Hripcsak, G., & Rothschild, A. S. (2005). Agreement, the f-measure, and reliability in information retrieval. *Journal of the American Medical Informatics Association*, *12*(3), 296–298. https://doi.org/10.1197/jamia.m1733

Jiang, H., Feng, M., Zhu, Y., Lu, N., Huang, J., & Xiao, T. (2014). An automated method for extracting rivers and lakes from Landsat imagery. *Remote Sensing*, *6*(6), 5067–5089. https://doi.org/10.3390/rs6065067

Jiao, L., Liu, Y., & Li, H. (2012). Characterizing land-use classes in remote sensing imagery by shape metrics. *ISPRS Journal of Photogrammetry and Remote Sensing*, *72*, 46–55. https://doi.org/10.1016/j.isprsjprs.2012.05.012

Jones, A. E., Hodges, B. R., McClelland, J. W., Hardison, A. K., & Moffett, K. B. (2017). Residence-time-based classification of surface water systems. *Water Resources Research*, *53*(7), 5567–5584. https://doi.org/10.1002/2016wr019928

Khandelwal, A., Karpatne, A., & Kumar, V. (2017). ORBIT: Ordering based information transfer across space and time for global surface water monitoring (version 1). arXiv. https://doi.org/10.48550/ARXIV.1711.05799

Khandelwal, A., Karpatne, A., Ravirathinam, P., Ghosh, R., Wei, Z., Dugan, H. A., et al. (2022). ReaLSAT, a global dataset of reservoir and lake surface area variations. *Scientific Data*, *9*(1), 356. https://doi.org/10.1038/s41597-022-01449-5

Khandelwal, A., Karpatne, A., Ravirathinam, P., Ghosh, R., Wei, Z., Dugan, H. A., et al. (2023). ReaLSAT, a global dataset of reservoir and lake surface area variations (2.0) [Dataset]. *Zenodo*. https://doi.org/10.5281/zenodo.7614815

Lehner, B., & Döll, P. (2004). Development and validation of a global database of lakes, reservoirs and wetlands. *Journal of Hydrology*, *296*(1–4), 1–22. https://doi.org/10.1016/j.jhydrol.2004.03.028

Lehner, B., Liermann, C. R., Revenga, C., Vörösmarty, C., Fekete, B., Crouzet, P., et al. (2011). High- resolution mapping of the world's reservoirs and dams for sustainable river- flow management. *Frontiers in Ecology and the Environment*, *9*(9), 494–502. https://doi.org/10.1890/100125

Liu, Y., Kwan, M.-P., & Wu, Z. (2022). Visualizing and quantifying the spatiotemporal expansion of the Blue Lentic Belt in Alabama and Mississippi. *Water Research*, *217*, 118444. https://doi.org/10.1016/j.watres.2022.118444

Liu, Y., Liu, H., Wang, L., Xu, M., Cohen, S., & Liu, K. (2021). Derivation of spatially detailed lentic habitat map and inventory at a basin scale by integrating multispectral Sentinel-2 satellite imagery and USGS Digital Elevation Models. *Journal of Hydrology*, *603*, 126876. https://doi.org/10.1016/j.jhydrol.2021.126876

Luo, J., Chen, Z., & Chen, N. (2021). Distinguishing different subclasses of water bodies for long-term and large-scale statistics of lakes: A case study of the Yangtze River basin from 2008 to 2018. *International Journal of Digital Earth*, *14*(2), 202–230. https://doi.org/10.1080/17538947.2020.1810338

Martínez-Santos, P., Aristizábal, H. F., Díaz-Alcaide, S., & Gómez-Escalonilla, V. (2021). Predictive mapping of aquatic ecosystems by means of support vector machines and random forests. *Journal of Hydrology*, *595*, 126026. https://doi.org/10.1016/j.jhydrol.2021.126026

Meng, L., Zhang, Z., Zhang, W., Ye, J., Wu, C., Chen, D., & Song, C. (2019). An automatic extraction method for lakes and reservoirs using satellite images. *IEEE Access*, *7*, 62443–62456. https://doi.org/10.1109/ACCESS.2019.2916148

Messager, M. L., Lehner, B., Grill, G., Nedeva, I., & Schmitt, O. (2016). Estimating the volume and age of water stored in global lakes using a geostatistical approach. *Nature Communications*, *7*(1), 13603. https://doi.org/10.1038/ncomms13603

Micklin, P. (2016). The future Aral Sea: Hope and despair. *Environmental Earth Sciences*, *75*(9), 844. https://doi.org/10.1007/s12665-016-5614-5

Naik, S. M., Chakraborty, T., Panja, M., Hadid, A., & Chakraborty, B. (2025). Skew-probabilistic neural networks for learning from imbalanced data. *Pattern Recognition*, *165*, 111677. https://doi.org/10.1016/j.patcog.2025.111677

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Pekel, J. F., Cottam, A., Gorelick, N., & Belward, A. S. (2016a). High-resolution mapping of global surface water and its long-term changes. *Nature*, *540*(7633), 418–422. https://doi.org/10.1038/nature20584

Pekel, J. F., Cottam, A., Gorelick, N., & Belward, A. S. (2016b). High-resolution mapping of global surface water and its long-term changes [Dataset]. Retrieved from https://global-surface-water.appspot.com/download

Pellett, S., Bigley, D., & Grimes, D. (1983). Distribution of Pseudomonas aeruginosa in a riverine ecosystem. *Applied and Environmental Microbiology*, *45*(1), 328–332. https://doi.org/10.1128/aem.45.1.328-332.1983

Peña, F. J., Hübinger, C., Payberah, A. H., & Jaramillo, F. (2024). DeepAqua: Semantic segmentation of wetland water surfaces with SAR imagery using deep neural networks without manually annotated data. *International Journal of Applied Earth Observation and Geoinformation*, *126*, 103624. https://doi.org/10.1016/j.jag.2023.103624

Pi, X., Luo, Q., Feng, L., Xu, Y., Tang, J., Liang, X., et al. (2022a). Mapping global lake dynamics reveals the emerging roles of small lakes. *Nature Communications*, *13*(1), 5777. https://doi.org/10.1038/s41467-022-33239-3

Pi, X., Luo, Q., Feng, L., Xu, Y., Tang, J., Liang, X., et al. (2022b). Mapping global lake dynamics reveals the emerging roles of small lakes: Code and data [Dataset]. *Zenodo*. https://doi.org/10.5281/zenodo.7016548

Post, D. M., Pace, M. L., & Hairston, N. G. (2000). Ecosystem size determines food-chain length in lakes. *Nature*, *405*(6790), 1047–1049. https://doi.org/10.1038/35016565

Qu, Y., Wu, N., Guse, B., & Fohrer, N. (2018). Riverine phytoplankton shifting along a lentic-lotic continuum under hydrological, physiochemical conditions and species dispersal. *Science of the Total Environment*, *619–620*, 1628–1636. https://doi.org/10.1016/j.scitotenv.2017.10.139

Raff, J. L., Goodbred, S. L., Pickering, J. L., Sincavage, R. S., Ayers, J. C., Hossain, M. S., et al. (2023). Sediment delivery to sustain the Ganges-Brahmaputra delta under climate change and anthropogenic impacts. *Nature Communications*, *14*(1), 2429. https://doi.org/10.1038/s41467-023-38057-9

Rather, I. H., Kumar, S., & Gandomi, A. H. (2024). Breaking the data barrier: A review of deep learning techniques for democratizing AI with small datasets. *Artificial Intelligence Review*, *57*(9), 226. https://doi.org/10.1007/s10462-024-10859-3

Read, J. S., & Rose, K. C. (2013). Physical responses of small temperate lakes to variation in dissolved organic carbon concentrations. *Limnology & Oceanography*, *58*(3), 921–931. https://doi.org/10.4319/lo.2013.58.3.0921

Rennella, A. M., & Quirós, R. (2006). The effects of hydrology on plankton biomass in shallow lakes of the pampa plain. *Hydrobiologia*, *556*(1), 181–191. https://doi.org/10.1007/s10750-005-0318-y

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, *65*(6), 386–408. https://doi.org/10.1037/h0042519

Sharma, A., Narayanan, M., & Ilampooranan, I. (2026). An automated data efficient morphometric approach to define global lentic and lotic inland waters (version v1). *Zenodo*. https://doi.org/10.5281/zenodo.18277872

Thompson, N. C., Greenewald, K., Lee, K., & Manso, G. F. (2020). The computational limits of deep learning (version 2). arXiv. https://doi.org/10.48550/ARXIV.2007.05558

Van Der Werff, H. M. A., & Van Der Meer, F. D. (2008). Shape-based classification of spectrally identical objects. *ISPRS Journal of Photogrammetry and Remote Sensing*, *63*(2), 251–258. https://doi.org/10.1016/j.isprsjprs.2007.09.007

Verpoorter, C., Kutser, T., Seekell, D. A., & Tranvik, L. J. (2014). A global inventory of lakes based on high-resolution satellite imagery. *Geophysical Research Letters*, *41*(18), 6396–6402. https://doi.org/10.1002/2014GL060641

Winslow, L. A., Read, J. S., Hansen, G. J. A., & Hanson, P. C. (2015). Small lakes show muted climate change signal in deepwater temperatures. *Geophysical Research Letters*, *42*(2), 355–361. https://doi.org/10.1002/2014GL062325

Yang, X., Wang, N., Chen, A. A., He, J., Hua, T., & Qie, Y. (2020). Changes in area and water volume of the Aral Sea in the arid Central Asia over the period of 1960–2018 and their causes. *Catena*, *191*, 104566. https://doi.org/10.1016/j.catena.2020.104566

Yao, F., Livneh, B., Rajagopalan, B., Wang, J., Crétaux, J.-F., Wada, Y., & Berge-Nguyen, M. (2023). Satellites reveal widespread decline in global lake water storage. *Science*, *380*(6646), 743–749. https://doi.org/10.1126/science.abo2812

Zhang, H., Si, S., & Hsieh, C.-J. (2017). GPU-acceleration for large-scale tree boosting (version 1). arXiv. https://doi.org/10.48550/ARXIV.1706.08359