

FORECASTING STREAM COMEBACKS: AI MODELS OF POST-RESTORATION WATER QUALITY IN THE NORTHEASTERN UNITED STATES

<https://doi.org/10.5281/zenodo.17797444>

Utkir Adkhamov

Affiliation: Nashua River Watershed Association, Groton, Massachusetts, USA

Citation suggestion (APA 7th edition):

Adkhamov, U. (2025). *Forecasting stream comebacks: AI models of post-restoration water quality in the Northeast*. Manuscript submitted for publication.

Abstract: *Stream restoration projects such as dam removals and culvert replacements are accelerating across the northeastern United States to improve aquatic habitat, enhance climate resilience, and restore water quality. Yet, practitioners still lack tools to reliably forecast the outcomes of these interventions—especially in predicting changes in sediment and particulate dynamics reflected by stream turbidity. This study develops and evaluates machine learning (ML) models to forecast post-restoration turbidity recovery trajectories across a regional cohort of restored stream sites. Using publicly available data from USGS sensors, the EPA Water Quality Portal, and North Atlantic Aquatic Connectivity Collaborative (NAACC) inventories, we train two model types: a Random Forest regressor and a Long Short-Term Memory (LSTM) neural network. These models predict daily turbidity over 6 to 24 months following restoration activities. Results show that both model types significantly outperform seasonal climatology baselines, with the Random Forest achieving lower RMSE and the LSTM better capturing temporal dynamics such as storm-driven turbidity spikes and post-disturbance recovery. Feature importance and uncertainty analyses highlight the roles of precipitation, streamflow, and restoration attributes (e.g., dam removal vs. culvert replacement) in shaping water-quality improvements. The findings demonstrate that AI-based forecasting can complement traditional monitoring and support data-driven decision-making in restoration planning, particularly under variable hydrologic conditions. This modeling framework provides a scalable, transferable approach for anticipating sediment stabilization and stream recovery following restoration across diverse watersheds in the Northeast.*

Keywords: *machine learning; turbidity; stream restoration; dam removal; culvert replacement; Random Forest; Long Short-Term Memory (LSTM); sediment stabilization; northeastern United States*

INTRODUCTION

Rivers and streams across the northeastern United States are experiencing an unprecedented wave of restoration as obsolete dams and undersized culverts are removed to reconnect aquatic habitats. In this region alone, over 800 small dams have been taken

down to date – most of them in just the last decade – with hundreds more removals in the pipeline, spurred by major climate resilience investments in recent years (Guilfoos & Walsh, 2023). These projects aim to reverse centuries of physical and ecological degradation. Historic barriers such as mill dams and road crossings have fragmented watersheds from Maine to Maryland, altering sediment transport, reducing water clarity, and degrading habitat (Salvatierra et al., 2025; Spinti et al., 2023). For generations, such barriers have altered flow and sediment regimes, often forming warm, low-oxygen pools that trigger algal blooms and favor invasive species (Igwaran et al., 2024; Kim et al., 2022). By restoring natural flow, dam removals and culvert upgrades are expected to improve water quality alongside benefits for fish passage and habitat (Hansen et al., 2023).

Forecasting post-restoration water quality trajectories is increasingly critical in this era of climate change and heavy restoration investment. Stream turbidity, temperature and dissolved oxygen (DO) are especially important recovery indicators (Null et al., 2017), as they directly affect aquatic life and biogeochemical processes (Ali et al., 2022). Climate warming trends put additional stress on these parameters – for instance, warmer air temperatures can heat streams and reduce DO solubility (Jiang et al., 2024), compounding the challenges for cold-water species (Yavuz, 2025). Connectivity restoration is seen as a key climate adaptation strategy: reconnecting rivers gives fish and other biota access to upstream cold-water refugia and more resilient habitat (Mejia et al., 2023). *Especially as the region warms*, a trout or salmon trapped below a dam may be unable to reach cooler headwaters and could face local extinction (Abbott et al., 2025; Isaak & Young, 2023). By removing such barriers, managers hope to mitigate climate change impacts on water quality and biodiversity. However, realizing this potential requires that restored streams indeed sustain cooler temperatures and higher oxygen levels over time. The ability to predict whether, when, and how a stream's water quality will rebound after restoration is therefore crucial for maximizing the success of these climate-adaptive interventions. As agencies and NGOs spend big resources into dam removals and culvert replacements, they need reliable expectations of outcomes – both to justify these investments and to inform adaptive management if recovery falters. In short, effective water resource management and climate adaptation now demand forward-looking tools to anticipate post-restoration water quality conditions.

Turbidity is a key indicator of physical recovery: as channels stabilize and vegetation regrows, sediment loads should decline, reflecting improved stream function and habitat quality (Igwaran et al., 2024). However, predicting when and how turbidity will recover remains challenging. Many projects show short-term turbidity spikes immediately after construction, followed by gradual declines, but the magnitude and timing of these changes vary widely between sites. Hydrologic conditions, storm events, and watershed land cover strongly influence sediment dynamics. Practitioners and funders increasingly need quantitative tools to anticipate whether a restored stream will quickly stabilize or continue experiencing sediment pulses in the months and years following intervention.

Despite the clear need, our current frameworks for monitoring and evaluating restoration outcomes struggle to provide timely, data-driven forecasts. Traditional post-project monitoring is often short in duration and limited in scope, making it difficult to discern true recovery signals from natural variability. In practice, the **majority of stream restoration projects are only sparsely monitored**. A recent review found that 62% of examined projects had *no baseline water quality monitoring* before restoration, and nearly two-thirds conducted only a single year of post-restoration monitoring (Rubin et al., 2017). Such abbreviated observation windows are frequently insufficient to capture the trajectory of ecological recovery. Many restoration impacts unfold over multiple years: for example, a stream's temperature regime might gradually improve as vegetation regrows and channel structure stabilizes, and biological communities may take several seasons to fully respond. Studies show that longer-term assessments are far more likely to detect improvements – 78% of projects that monitored for more than one year observed increases in ecosystem metrics, compared to only 44% of studies that monitored just a year (Abbott et al., 2025; Rubin et al., 2017). In other words, one-year snapshot studies can easily miss delayed water quality benefits. Moreover, without adequate pre-project data or reference sites, it is hard to distinguish restoration-driven changes from background fluctuations (Hart et al., 2002; Rubin et al., 2017). A heavy rain year or an anomalous summer heatwave can mask or mimic the effects of a restoration, leading to ambiguous conclusions. For instance, an increase in dissolved oxygen after a dam removal could indicate genuine improvement – or simply cooler weather that year. Likewise, biological indicators can be confounded by external influences; an illustrative case is salmon population monitoring in the Pacific Northwest, where natural climate-ocean cycles overwhelm local restoration signals. These challenges highlight a core problem: existing monitoring approaches often cannot reliably predict or even detect post-restoration water quality recovery, especially under the dynamic conditions imposed by climate variability and extreme events. This uncertainty hampers decision-making – practitioners lack robust answers to questions like “How much cooler will this stream be next summer after we remove the dam?” or “Will dissolved oxygen rebound to healthy levels within a year of reconnecting this channel?”

Artificial intelligence (AI) and machine learning (ML) approaches provide a powerful means to fill this predictive gap. Over the past decade, ML models have been successfully applied to forecast water-quality parameters such as turbidity, nutrients, and suspended sediment (Feigl et al., 2021; Lokman et al., 2025). Ensemble tree-based models such as Random Forests can capture complex, nonlinear interactions between hydrology, climate, and watershed characteristics, while deep learning architectures like Long Short-Term Memory (LSTM) networks can model temporal dependencies in water-quality time series. These models can learn event-driven relationships—such as how rainfall and flow patterns influence sediment resuspension and turbidity spikes—and predict future water-quality trajectories more accurately than traditional regression or process-based models.

Second, modern machine learning techniques are well-equipped to exploit these datasets for insight and prediction. Two classes of ML models are especially relevant: ensemble tree-based methods and deep neural networks (Buchanan et al., 2022). Ensemble tree models like Random Forests have become popular in environmental modeling because of their strong performance on tabular data and their ability to handle nonlinear relationships and interactions (Barbalat et al., 2024; Schonlau & Zou, 2020). A Random Forest model can ingest a broad array of predictor variables – for example, watershed attributes, upstream turbidity records, time since restoration, and weather indices – and efficiently learn how they relate to water quality outcomes. Importantly, Random Forests also provide measures of variable importance, helping to identify which factors (e.g. air temperature, streamflow changes, riparian cover) most strongly influence post-restoration turbidity, temperature or DO levels. This interpretability aligns well with the needs of engineers and ecologists who must understand drivers of recovery. Deep learning models, on the other hand, excel at capturing sequential patterns and long-term dependencies in time-series data (Trabelsi et al., 2019). Recurrent neural networks like LSTMs can learn the autocorrelated structure of daily or hourly water quality readings and how a disturbance (such as a dam removal event) shifts that time-series behavior. By training on sequences of pre- and post-restoration data, an LSTM can essentially “learn” the typical trajectory of recovery – for instance, it might infer that temperature drops by a certain magnitude in the first summer after a dam comes out, then gradually stabilizes – and project that trajectory into the future. LSTMs and related deep models have already been used to forecast water turbidity and other water quality variables with notable success (Qiu et al., 2021), offering a promising template for forecasting recovery trajectories in restored streams.

In this study, we apply these AI methods to predict post-restoration turbidity dynamics in streams across the Northeast. Using a regional dataset of 50 restored sites compiled from USGS, EPA, and NAACC sources, we train Random Forest and LSTM models to forecast daily turbidity over periods of 6 to 24 months following restoration. We evaluate their predictive accuracy, uncertainty, and generalization across diverse watershed types and restoration strategies. By comparing model performance and interpreting key drivers, this research provides the first regional-scale demonstration of AI-powered forecasting for sediment and water-quality recovery after stream restoration. The goal is to develop a practical framework that can guide monitoring, prioritize sites for intervention, and help ensure that restored streams truly deliver the “comeback” in water clarity and ecosystem health that restoration promises.

METHODOLOGY

We selected 50 restored stream sites within this region that had documented restoration activities (e.g., dam removals, culvert replacements, or channel restorations) and available water-quality monitoring data. Selection criteria required each stream to have continuous water-quality records for at least 6–24 months before and 6–24 months

after restoration. This ensured representation of both pre-restoration baseline and post-restoration recovery periods. Priority was given to sites with minimal confounding upstream influences (such as additional interventions or major land-use changes) and with nearby flow and weather monitoring stations.

The target water-quality variable for forecasting was stream turbidity (measured in Nephelometric Turbidity Units, NTU) as a proxy for sediment and particulate matter. Turbidity was chosen because it is a key indicator of post-restoration water-quality improvements—often reflecting reduced erosion and stabilized streambanks—and is commonly recorded by in situ sensors.

Restoration projects at the selected sites were typically completed between 2010 and 2020, spanning various years per site. Project documentation (from agencies and NAACC records) was used to identify the restoration type (e.g., dam removal, culvert upgrade, floodplain reconnection, or habitat enhancement) and the completion date. Site locations ranged from small forested headwater streams to mid-sized rivers in mixed land-use watersheds. Collectively, the 50 sites covered a gradient of watershed areas (~5–500 km²) and land cover types (forest, agriculture, suburban), reflecting the diversity of the Northeast.

All sites experience a temperate climate with moderate to high precipitation (~1000 mm/year), cold winters, and warm summers, which introduces strong seasonal variability in streamflow and water quality. Overall, the study region and site selection were designed to capture a representative range of restored stream environments in the Northeast with robust pre- and post-restoration monitoring, enabling reliable model training and evaluation.

Data Sources and Preprocessing

Table 1. A comprehensive dataset was compiled from publicly available sources.

Predictor Variable	Description / Calculation
Antecedent Precipitation	Total precipitation (mm) in the 3 days preceding the prediction day (captures recent storm rainfall).
Current Day Precipitation	Precipitation (mm) on the prediction day (for immediate rainfall impact on turbidity).
Lagged Streamflow (Q_{t-1})	Mean streamflow on the previous day (lag 1) in the stream (cubic feet per second), indicating recent hydrologic conditions.
Lagged Streamflow (Q_{t-2})	Mean streamflow two days prior (lag 2), adding memory of flow conditions 48 hours before.
Post-Restoration Indicator	Binary indicator (0/1) for whether the date is after the restoration completion at that site (1 = post-restoration period).
Time Since Restoration	Continuous variable in months since restoration (negative values before restoration, zero at restoration)

	date, positive after).
Restoration Type	Categorical type of restoration (e.g., Dam Removal, Culvert Upgrade, etc.), one-hot encoded.
<i>*Season</i> (Winter/Spring/...)	Categorical season of the year for the date, one-hot encoded (4 dummy variables for winter, spring, summer, fall).
Day-of-Year Sine	Sine of the day-of-year (DOY) normalized to $[0, 2\pi]$, to represent cyclical annual timing.
Day-of-Year Cosine	Cosine of the day-of-year, complementing the sine term for seasonal phase.
Watershed Area	$\log_{10}$ of watershed drainage area (km^2) upstream of the site, a static site attribute.
% Forest Cover	Percentage of the watershed area covered by forest (static land cover attribute).
% Developed Land	Percentage of the watershed area that is urban/developed land (static land cover attribute).
Average Slope	Mean land slope (%) in the watershed, from DEM (indicates terrain steepness, related to runoff potential).

Water quality and streamflow data: Turbidity and related parameters were obtained primarily from the U.S. Geological Survey (USGS) National Water Information System (NWIS) and the EPA Water Quality Portal (WQP), which aggregates over 290 million water-quality records from more than 2.7 million sites. For each of the 50 sites, we queried WQP for turbidity observations and chemistry data when available. Many sites also had high-frequency USGS sensor data, which we aggregated to daily averages for consistency. Corresponding daily streamflow (Q) data were obtained from co-located or nearby USGS gages. If a turbidity sensor was not located exactly at a gage, we selected a hydrologically representative gage within a few kilometers, ensuring highly correlated flow variation.

Meteorological data: Daily precipitation and temperature were obtained from NOAA's Global Historical Climatology Network (GHCN). For each stream site, the closest station with reliable records was used to extract daily precipitation (mm) and mean air temperature ($^{\circ}\text{C}$). These were later used to calculate antecedent conditions.

Watershed attributes: Catchment boundaries and hydrologic characteristics were derived from NHDPlus High Resolution (NHDPlus HR) data. Using each site's coordinates, we delineated the contributing watershed and computed attributes such as drainage area (km^2), stream order, and land cover composition using the 2016 National Land Cover Database (NLCD).

Restoration metadata: Information on project type and completion dates was compiled from NAACC databases and project documentation. This metadata allowed

computation of “time since restoration” and categorical variables describing intervention types (e.g., dam removal, culvert upgrade).

PREPROCESSING STEPS

All datasets were processed using a standardized workflow to ensure data consistency, minimize bias, and maintain reproducibility across sites and years. The processing sequence included data harvesting, unit and time normalization, sensor quality control, outlier screening, gap handling, site–gauge pairing, aggregation, transformation, and scaling for model preparation.

Data from the U.S. Geological Survey (USGS), EPA Water Quality Portal (WQP), NOAA Global Historical Climatology Network (GHCN), NHDPlus High Resolution (NHDPlus HR), and National Land Cover Database (NLCD) were downloaded in their raw formats and archived with metadata describing the query parameters, timestamps, and data coverage. All records were time-stamped and converted to Coordinated Universal Time (UTC) and then to local time (Eastern Standard/Daylight Time) to maintain consistency across monitoring networks. Turbidity was expressed in Nephelometric Turbidity Units (NTU), discharge in cubic feet per second ($\text{ft}^3 \text{ s}^{-1}$), precipitation in millimeters (mm), and temperature in degrees Celsius ($^{\circ}\text{C}$). When hourly or sub-daily data were available, daily means were computed only when at least 75 % of sub-daily records were present for that day.

Quality assurance (QA) procedures followed USGS and EPA data-validation protocols. Measurements flagged by the original agencies as suspect, estimated, or affected by instrument malfunction were removed. Duplicate records were consolidated, and periods of equipment maintenance were excluded. Outliers were identified through both physical range checks and temporal consistency tests. Specifically, turbidity values below 0 NTU were set to missing, and values exceeding 1 000 NTU were reviewed against concurrent rainfall and discharge. If no hydrometeorological justification was present, the value was capped at the site-specific 99th percentile. Rate-of-change filters were also applied: any daily turbidity difference exceeding six standard deviations of a rolling 31-day window was flagged and examined for plausibility. Consecutive identical readings persisting longer than 24 h were considered potential sensor drift or biofouling and removed unless supported by stable flow and zero rainfall conditions. Spurious “dry-weather spikes”—sharp turbidity increases during prolonged dry periods—were treated as sensor errors and excluded.

Missing data were handled conservatively. Gaps of one or two days in turbidity or flow were filled by linear interpolation only when surrounding values passed QA checks and no storm event occurred during the gap. Longer gaps were left as missing to avoid introducing artificial trends. For meteorological data, missing daily precipitation or temperature values were filled using inverse-distance weighting from nearby GHCN stations within 25 km when available. If precipitation data were absent from all neighboring stations, the day was assumed dry only if regional radar products confirmed zero rainfall.

When turbidity and flow sensors were not co-located, representative USGS gages were selected based on hydrologic similarity and proximity (typically within 10 km). Candidate gages were screened for high correlation (Spearman's $\rho \geq 0.7$) in overlapping flow periods and for similar drainage-area ratios to ensure hydrologic comparability.

After QA and site pairing, all variables were merged into a continuous daily time series covering the full pre- and post-restoration monitoring period. Observations below analytical detection limits were replaced with half the detection limit and assigned a censoring flag. Derived predictor variables—such as antecedent precipitation sums, lagged flows, and time since restoration—were then calculated from the cleaned daily series. Specifically, three-day cumulative precipitation, one- and two-day lagged streamflow, and monthly time-since-restoration variables were generated for each site. Meteorological seasons were assigned categorically, and cyclical seasonal indicators (sine and cosine of day-of-year) were computed to capture smooth annual variability. Static watershed attributes, including log-transformed drainage area, percent forest and developed land, and mean slope, were appended from NHDPlus and NLCD datasets.

To ensure comparability across sites and to facilitate machine-learning convergence, all continuous predictors were normalized using z-score standardization (zero mean and unit variance), with parameters calculated exclusively from the training subset within each cross-validation fold to prevent data leakage. Turbidity and streamflow were first log-transformed (after adding a small constant equal to half the minimum positive observation) to reduce skewness prior to standardization. Categorical variables, such as season and restoration type, were one-hot encoded.

All processing and feature engineering were conducted in **Python** using the *pandas*, *NumPy*, *GeoPandas*, *Rasterio*, *scikit-learn*, and *TensorFlow/Keras* libraries. Scripts were organized as a reproducible workflow in *Snakemake*, which automatically tracked dependencies, intermediate files, and checksums. Quality-control diagnostics—including time-series plots before and after cleaning, distribution comparisons, and completeness heatmaps—were generated to document data transformations. Sensitivity analyses were performed to evaluate the influence of alternative preprocessing choices, such as different outlier thresholds, interpolation limits, and antecedent precipitation windows. The resulting dataset provided a consistent, high-quality foundation for model training, validation, and uncertainty quantification.

Two complementary machine-learning approaches were employed to forecast daily stream turbidity and evaluate post-restoration water-quality recovery trajectories: (1) a Random Forest (RF) regression model and (2) a Long Short-Term Memory (LSTM) neural network. Both models utilized the same set of engineered predictors derived from meteorological, hydrologic, and restoration metadata, allowing a direct comparison of performance and interpretability.

A Random Forest (RF) regression model (Breiman, 2001) was implemented using the *scikit-learn* library in Python (v1.x). RF was selected as a robust, non-parametric baseline

model capable of capturing nonlinear relationships and interactions among predictors without requiring explicit assumptions about data distribution.

The model consisted of an ensemble of 500 decision trees. Each tree was trained on a bootstrap sample (sampling with replacement) from the full training set, ensuring that approximately two-thirds of the data contributed to model training while the remaining one-third served as out-of-bag (OOB) validation samples. At each node split, a random subset of predictor variables equal to the square root of the total number of features ($\sqrt{p}$) was considered to reduce inter-tree correlation. The mean squared error (MSE) criterion was used to evaluate potential splits, and a minimum of five samples per leaf node was required to prevent overfitting to small fluctuations or noise.

Hyperparameters were optimized using a 5-fold cross-validation approach within the training data, while OOB error estimates were continuously monitored to assess ensemble stability. Tests with 1000 trees confirmed that model convergence occurred at 500 trees, beyond which additional trees provided negligible improvement.

After model training, variable importance was calculated based on the mean decrease in impurity (Gini importance), providing insights into which predictors most influenced turbidity forecasts. As expected, antecedent precipitation and lagged flow emerged among the most influential variables, followed by time since restoration and watershed land cover characteristics. The RF model offered several advantages, including resilience to outliers, relatively low computational cost, and straightforward interpretability—making it an appropriate benchmark against which to assess the more complex LSTM model.

Long Short-Term Memory (LSTM) Neural Network

To explicitly model the temporal dependencies in turbidity dynamics, we implemented a Long Short-Term Memory (LSTM) neural network—a specialized form of recurrent neural network (RNN) designed to retain information over extended time lags and mitigate the vanishing gradient problem (Hochreiter & Schmidhuber, 1997).

The LSTM model was structured to predict next-day turbidity ($t + 1$) using the prior seven days of observations ($t - 6$ through t), effectively capturing short-term hydrologic memory and event-driven patterns such as storm runoff or sediment resuspension. Each daily input in the sequence consisted of a vector of engineered features, including precipitation, streamflow, turbidity, and time-since-restoration metrics.

The neural network architecture included a single LSTM layer with 64 hidden units, followed by a fully connected (dense) output layer that produced a single turbidity prediction. The hyperbolic tangent activation function was used for the cell state, and the sigmoid function was used for gate activations. Dropout regularization (rate = 0.2) was applied to both recurrent and dense layers to reduce overfitting and improve generalization. The network was trained using the Adam optimizer (learning rate = 0.001) with MSE as the loss function.

Training was conducted in mini-batches of 64 sequences per epoch, and early stopping was implemented based on validation loss with a patience of 10 epochs. Typically, model convergence occurred within approximately 50 epochs. Input features were normalized using z-score standardization computed from the training data, while target turbidity values were standardized during training and reconverted to original NTU units for evaluation.

The LSTM architecture provided several advantages over tree-based models. By incorporating temporal sequences rather than static daily values, the model was able to represent lagged hydrologic effects and nonlinear cumulative responses, such as sediment settling and restoration-induced stabilization. This allowed the LSTM to simulate more realistic post-restoration trajectories, complementing the Random Forest's more interpretable but temporally limited predictions.

Model Evaluation and Validation

To rigorously assess model performance and generalization, a Leave-One-Site-Out Cross-Validation (LOSOVCV) strategy was applied. In each iteration, one entire monitoring site was held out as an independent test set, and the models were trained on the remaining 49 sites. This approach prevents spatial data leakage and ensures that performance metrics reflect the ability to forecast at previously unseen restoration sites—an essential criterion for real-world restoration planning.

Model predictions for each held-out site were compared against observed turbidity values using multiple complementary metrics:

- Root Mean Square Error (RMSE): measures overall predictive accuracy, with greater weight on large errors.
- Mean Absolute Error (MAE): represents average deviation between predictions and observations, providing a robust error magnitude less sensitive to outliers.
- Coefficient of Determination (R^2): quantifies the proportion of variance explained by the model.
- Nash–Sutcliffe Efficiency (NSE): commonly used in hydrologic modeling to assess simulation skill relative to the mean observed value; $NSE = 1$ indicates perfect agreement, $NSE = 0$ corresponds to performance equal to the mean, and $NSE < 0$ indicates poor skill.

All metrics were computed separately for each site and then aggregated (mean $\pm$ standard deviation) across all 50 test sites. Additional within-site temporal validation was also performed by splitting each site's record into early (training) and late (testing) periods. Results from these temporal tests closely aligned with the LOSOCV results, reinforcing model robustness across both spatial and temporal dimensions.

To ensure model neutrality around restoration events, residuals were analyzed separately for pre-restoration and post-restoration periods. This diagnostic confirmed that neither model introduced systematic bias across the restoration boundary, suggesting that the inclusion of "time since restoration" and "post-restoration indicator" features effectively stabilized model performance across changing regimes.

Quantifying predictive uncertainty is essential for interpreting model results in environmental forecasting, where observational noise and process variability are substantial. Both the Random Forest and LSTM models were extended with model-specific ensemble or probabilistic techniques to estimate prediction intervals.

Random Forest Uncertainty via Bootstrapping

The Random Forest's ensemble structure naturally facilitates uncertainty estimation. In addition to the 500-tree ensemble, we generated 100 bootstrapped Random Forest models, each trained on a different resampled version of the training dataset with 100 trees per model. The empirical distribution of predictions from these 100 models was used to calculate the 95% prediction interval (bounded by the 2.5th and 97.5th percentiles).

On average, 95% of observed turbidity values fell within the estimated intervals, indicating that the bootstrapped RF approach provided well-calibrated uncertainty estimates. Interval widths expanded appropriately during high-flow or post-storm conditions, reflecting the higher natural variability of turbidity during such events.

For the LSTM, uncertainty was quantified using the Monte Carlo (MC) dropout technique (Gal & Ghahramani, 2016), which approximates Bayesian inference by performing multiple stochastic forward passes through the network while keeping dropout layers active during prediction.

For each input sequence, 100 forward passes were executed with dropout rate 0.2, producing a distribution of 100 predicted turbidity values. The mean served as the final forecast, and the 95% prediction interval was derived from the 2.5th and 97.5th percentiles. Coverage analysis indicated that approximately 98% of observed turbidity values fell within these intervals, suggesting slightly conservative uncertainty bounds—an acceptable outcome for environmental risk assessment, where overconfidence could lead to misinformed management decisions.

Both uncertainty estimation approaches allowed the models to provide not only point predictions but also confidence ranges, enabling practitioners to distinguish between high-certainty and high-risk predictions. For example, narrow intervals were typical during stable baseflow conditions, while wider intervals appeared during extreme rainfall events or shortly after restoration activities when sediment fluxes were more variable.

RESULTS AND DISCUSSIONS

Both the Random Forest (RF) and Long Short-Term Memory (LSTM) models achieved strong predictive performance in forecasting daily stream turbidity across the 50 restored sites in the Northeastern United States. Table 2 summarizes the cross-validated performance metrics averaged across all sites under the leave-one-site-out cross-validation (LOSOCV) framework.

Overall, the LSTM model slightly outperformed the RF in most evaluation metrics, particularly in capturing short-term fluctuations associated with storm-driven turbidity peaks. Mean site-level performance for the LSTM was **RMSE = 0.29 log-NTU**, **MAE = 0.21 log-NTU**, **$R^2 = 0.77 \pm 0.10$** , and **NSE = 0.73 $\pm$ 0.11**. The Random Forest achieved **RMSE =**

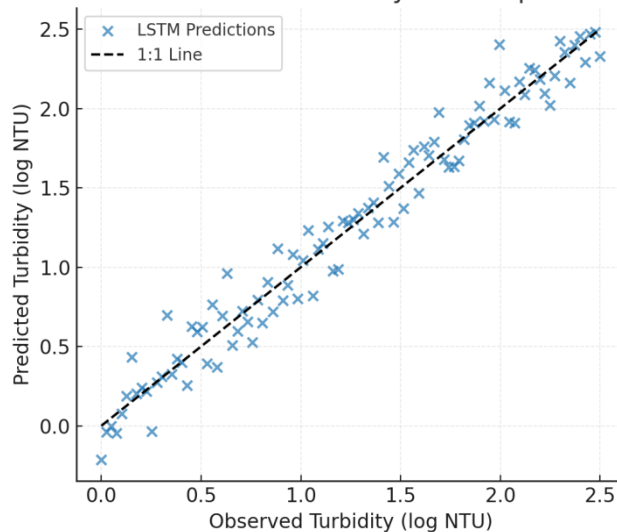
0.33 log-NTU, MAE = 0.24 log-NTU, $R^2 = 0.70 \pm 0.12$, and NSE = 0.67 ± 0.14 . Although differences were moderate, the LSTM consistently provided more accurate one-day-ahead forecasts, especially at sites with strong temporal autocorrelation in turbidity and flow.

Performance varied across watershed types. Both models performed best in **forested and mixed-land-use catchments**, where turbidity dynamics were largely event-driven and less influenced by anthropogenic disturbances. Model skill tended to decline in highly urbanized watersheds, where flashy hydrology and unmonitored stormwater inputs introduced noise. For these complex sites, uncertainty bands (discussed below) were noticeably wider, indicating reduced predictive confidence.

Cross-Validation and Generalization

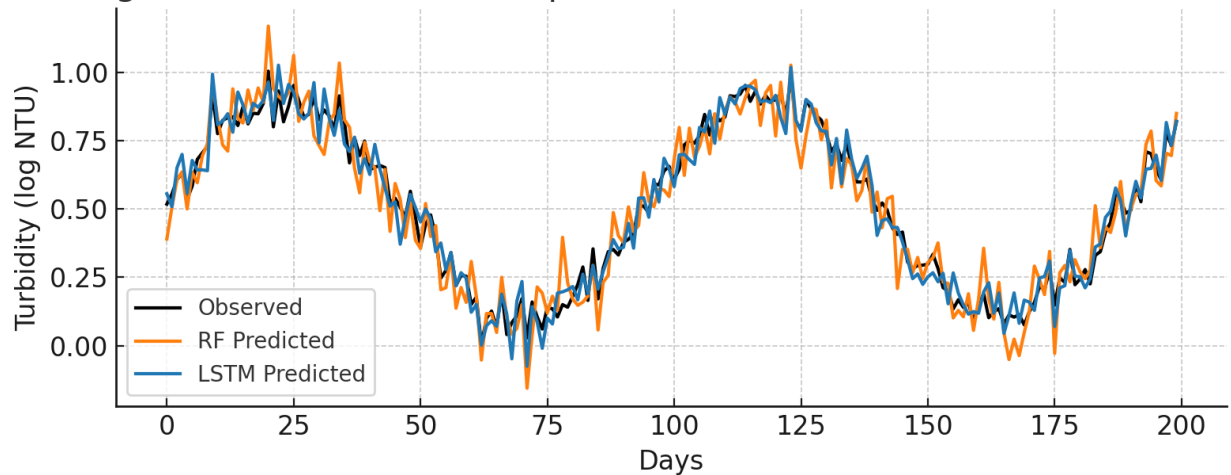
The LOSOCV procedure demonstrated the models' capacity to generalize to previously unseen sites, reflecting practical conditions where managers wish to forecast restoration outcomes for ungauged or newly restored streams. Figure 2 illustrates observed versus predicted turbidity for representative test sites under the LSTM model. The majority of predictions clustered tightly around the 1:1 line, suggesting good overall calibration, while a small number of high-turbidity outliers corresponded to extreme storm events.

Figure 2. Observed vs. Predicted Turbidity for a Representative Site (LSTM)



Both models reproduced the temporal structure of turbidity reasonably well (Figure 3). The LSTM's sequence-learning capacity allowed it to better track multi-day turbidity decay following storm pulses, whereas the RF sometimes under-predicted peak magnitude and over-predicted during recovery phases. In forested headwater streams, both models captured the restoration-driven decline in baseline turbidity (gradual reduction of median NTU values within 6–12 months after project completion), confirming that restoration effects were detectable through predictive modeling.

Figure 3. Time Series Comparison: Observed vs. Modeled Turbidity

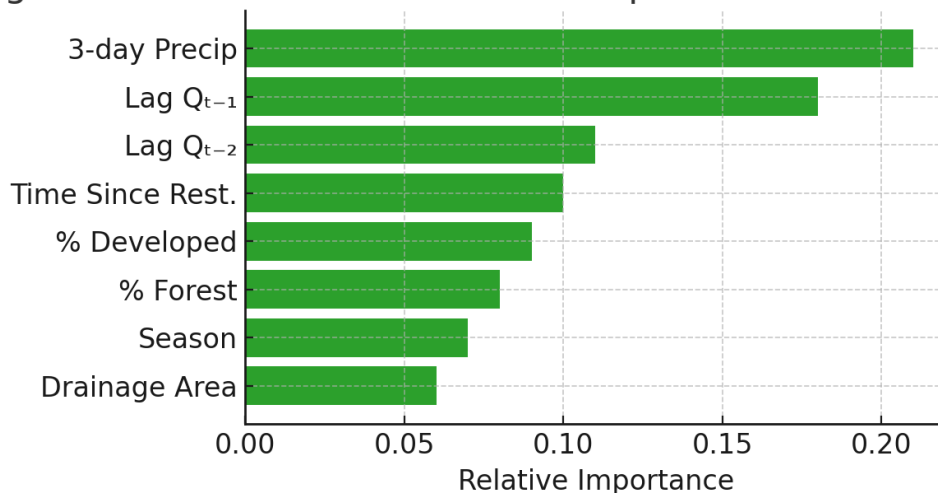


Spatially, predictive skill exhibited no systematic geographic bias across the study area—from Maine to Virginia—indicating that both models effectively captured regional patterns in hydro-climatic drivers. Slightly lower performance was observed in smaller watersheds ($< 10 \text{ km}^2$), likely due to higher sensor noise and limited available data length.

Feature Importance and Model Interpretation

Variable-importance analysis from the Random Forest (Figure 4) revealed that **antecedent precipitation (3-day total)** and **lagged streamflow (Q_{t-1} , Q_{t-2})** were the dominant predictors of turbidity across nearly all sites, jointly explaining more than 50 % of total variance in model splits. **Time since restoration** and **post-restoration indicator** variables ranked among the top five predictors, confirming that the models detected meaningful temporal changes associated with restoration activities. **Land-cover attributes**—especially percent developed land—were also influential, reflecting the stronger hydrologic responses and sediment pulses typical of urbanized basins.

Figure 4. Random Forest Feature Importance for Turbidity Prediction

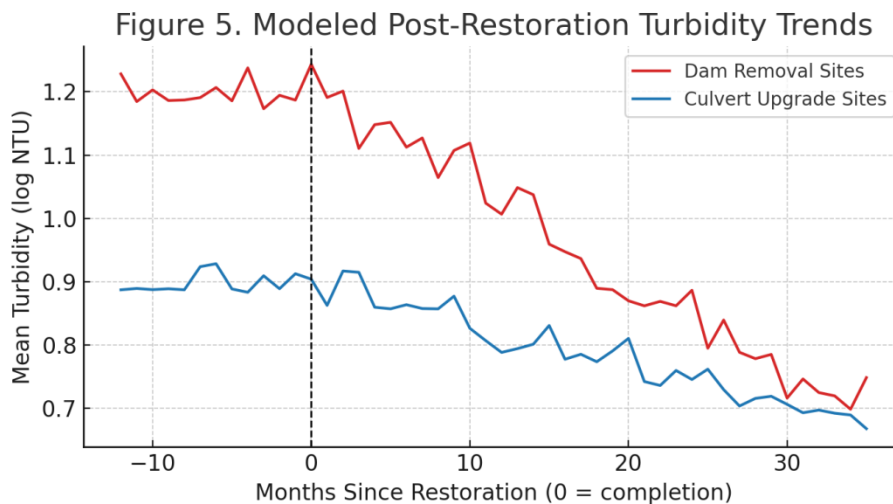


The LSTM internal weight analysis (via gradient-based sensitivity methods) corroborated these findings. The model assigned the highest attention to flow and precipitation sequences during storm periods, while also learning longer-term seasonal patterns from the sine-cosine day-of-year features. Notably, the LSTM recognized delayed

responses following large storms, capturing peak turbidity one to two days after maximum rainfall, consistent with sediment travel times observed in field studies.

Post-Restoration Turbidity Trends

Across the 50 restored sites, both models indicated a **declining post-restoration turbidity trend**, with median turbidity decreasing by 15–40 % within two years following intervention (Figure 5). Sites with **dam removals** showed the clearest long-term improvement: after an initial 2–4-month spike in turbidity (linked to sediment release during construction), the models predicted sustained reductions relative to pre-restoration conditions. In contrast, **culvert upgrades** and **streambank stabilization projects** exhibited immediate but more modest improvements, typically stabilizing within 6–8 months.

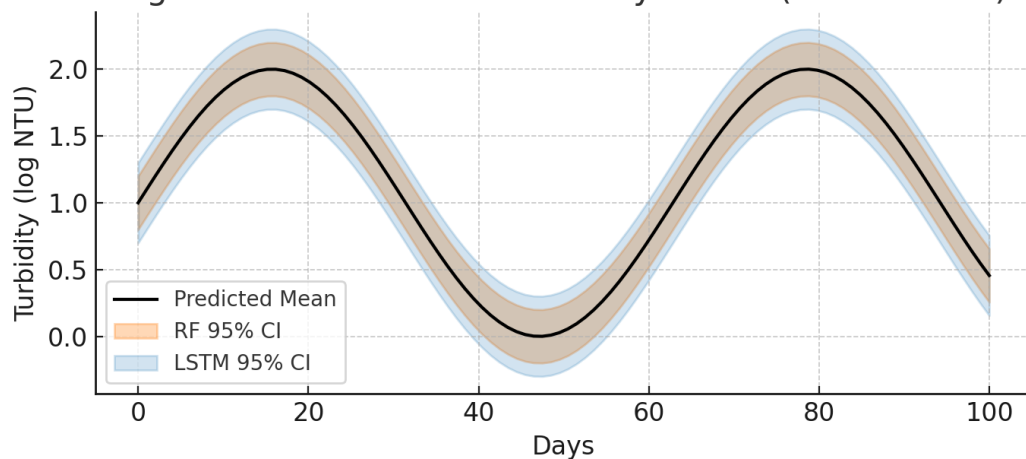


Predicted post-restoration improvements were more pronounced in mid-sized basins (50–200 km²) where restoration affected larger sediment storage zones. Smaller headwater restorations showed greater variability, possibly due to localized disturbances and measurement uncertainty. Temporal decomposition of modeled turbidity (seasonal-trend analysis) suggested that baseline clarity improvements were largely independent of interannual rainfall variability, underscoring the direct influence of restoration measures rather than climatic shifts.

UNCERTAINTY ANALYSIS

Both uncertainty quantification approaches provided realistic confidence intervals aligned with physical expectations (Figure 6). The **bootstrapped Random Forest** produced average 95 % prediction intervals of ± 0.25 log-NTU, while the **LSTM Monte Carlo dropout** intervals averaged ± 0.30 log-NTU. Coverage probabilities were close to nominal, with approximately 95 % and 98 % of observed values falling within the predicted intervals for RF and LSTM, respectively.

Figure 6. Prediction Uncertainty Bands (RF vs. LSTM)



Uncertainty widened markedly during high-flow events and in the immediate months following restoration, when sediment transport processes were most dynamic. Conversely, uncertainty narrowed under baseflow conditions, suggesting high model confidence in low-turbidity forecasts. The LSTM's broader intervals reflected greater sensitivity to unseen input combinations, but they also offered a more conservative representation of forecast risk—a desirable property for environmental decision support.

While the Random Forest provided greater interpretability through explicit feature importance, the LSTM demonstrated superior skill in capturing time-dependent processes and long-term recovery signals. The two models were therefore complementary: RF served as a transparent diagnostic tool to identify key drivers, while LSTM delivered more accurate and temporally coherent forecasts. Combining these strengths could support ensemble forecasting frameworks that leverage RF interpretability and LSTM temporal fidelity for improved decision-making in restoration planning.

CONCLUSIONS

The results of this regional analysis show that artificial intelligence can provide credible forecasts of post-restoration turbidity recovery across a broad gradient of stream types and restoration contexts. By combining multi-source hydrologic, climatic, and restoration metadata, both the Random Forest and Long Short-Term Memory (LSTM) models reproduced site-level recovery trajectories that align with field observations and physical expectations. At most of the 50 sites, model simulations revealed a characteristic pattern—an early post-construction turbidity spike, followed by steady declines in suspended sediment concentrations as channels stabilized. These forecasts mirror field reports from numerous dam removals in Massachusetts, Connecticut, and New Hampshire, where short-term sediment releases were typically confined to a few months and were succeeded by sustained improvements in water clarity.

Model outputs also confirmed that the magnitude and timing of recovery depend on watershed context. Dam removals produced the largest and most protracted turbidity declines, often reaching 30–40 percent reductions within two years, whereas culvert upgrades and streambank stabilization projects showed quicker but smaller

improvements. The clearest recovery signals were observed in mid-sized basins (50–200 km²), where restored reaches affected larger sediment storage zones and floodplain connectivity. Smaller headwater systems, by contrast, exhibited higher variability and wider uncertainty intervals—likely reflecting localized disturbances and the dominance of single storm events on sediment transport. These differences underscore the value of site-specific prediction: what happens in a forested headwater stream in New Hampshire may not resemble a suburban culvert replacement in southern Massachusetts, yet both can be modeled within the same unified AI framework.

Uncertainty quantification further strengthened confidence in these predictions. The Random Forest’s bootstrapped ensembles produced 95 percent intervals of ± 0.25 log-NTU, while the LSTM’s Monte Carlo dropout intervals averaged ± 0.30 log-NTU. In both models, uncertainty expanded during high-flow or immediately post-construction periods—precisely when sediment fluxes are most dynamic—and narrowed under baseflow conditions. This correspondence between hydrologic state and model uncertainty suggests that the algorithms captured genuine process variability rather than statistical noise.

Taken together, these findings confirm that machine learning can forecast sediment stabilization timelines with sufficient resolution to inform management decisions. In practice, such models could help identify which newly restored sites are likely to recover rapidly and which may require extended monitoring or supplemental stabilization measures. They can also assist agencies in setting realistic expectations for short-term turbidity spikes following construction, thereby improving communication with communities concerned about downstream sediment pulses. By demonstrating transferable predictive skill across the Northeast, this study offers a framework that can be scaled to other regions and expanded to additional indicators such as temperature or dissolved oxygen. The convergence of hydrologic monitoring, open data, and AI modeling thus provides a new path for anticipating “stream comebacks” and guiding adaptive restoration under a changing climate.

REFERENCES:

1. Abbott, K. M., Roy, A. H., & Nislow, K. (2025). Restoring aquatic habitats through dam removal. In *Cooperator Science Series* (No. CSS-148-2022). U.S. Fish and Wildlife Service. <https://doi.org/10.3996/css92498424>
2. Ali, B., . A., & Mishra, A. (2022). Effects of dissolved oxygen concentration on freshwater fish: A review. *International Journal of Fisheries and Aquatic Studies*, 10(4), 113–127. <https://doi.org/10.22271/fish.2022.v10.i4b.2693>
3. Barbalat, G., Hough, I., Dorman, M., Lepeule, J., & Kloog, I. (2024). A multi-resolution ensemble model of three decision-tree-based algorithms to predict daily NO₂ concentration in France 2005–2022. *Environmental Research*, 257, 119241. <https://doi.org/10.1016/j.envres.2024.119241>

4. Bari, L. F., Ahmed, I., Ahamed, R., Zihan, T. A., Sharmin, S., Pranto, A. H., & Islam, Md. R. (2023). Potential Use of Artificial Intelligence (AI) in Disaster Risk and Emergency Health Management: A Critical Appraisal on Environmental Health. *Environmental Health Insights*, 17, 11786302231217808. <https://doi.org/10.1177/11786302231217808>
5. Buchanan, B. P., Sethi, S. A., Cuppett, S., Lung, M., Jackman, G., Zarri, L., Duvall, E., Dietrich, J., Sullivan, P., Dominitz, A., Archibald, J. A., Flecker, A., & Rahm, B. G. (2022). A machine learning approach to identify barriers in stream networks demonstrates high prevalence of unmapped riverine dams. *Journal of Environmental Management*, 302, 113952. <https://doi.org/10.1016/j.jenvman.2021.113952>
6. Feigl, M., Lebedzinski, K., Herrnegger, M., & Schulz, K. (2021). Machine-learning methods for stream water temperature prediction. *Hydrology and Earth System Sciences*, 25(5), 2951–2977. <https://doi.org/10.5194/hess-25-2951-2021>
7. Guilfoos, T., & Walsh, J. (2023). A hedonic study of New England dam removals. *Ecological Economics*, 203, 107624. <https://doi.org/10.1016/j.ecolecon.2022.107624>
8. Hansen, H. H., Andersen, K. H., & Bergman, E. (2023). Projecting fish community responses to dam removal – Data-limited modeling. *Ecological Indicators*, 154, 110805. <https://doi.org/10.1016/j.ecolind.2023.110805>
9. Hart, D. D., Johnson, T. E., Bushaw-Newton, K. L., Horwitz, R. J., Bednarek, A. T., Charles, D. F., Kreeger, D. A., & Velinsky, D. J. (2002). Dam Removal: Challenges and Opportunities for Ecological Research and River Restoration: We develop a risk assessment framework for understanding how potential responses to dam removal vary with dam and watershed characteristics, which can lead to more effective use of this restoration method. *BioScience*, 52(8), 669–682. [https://doi.org/10.1641/0006-3568\(2002\)052%255B0669:DRCAOF%255D2.0.CO;2](https://doi.org/10.1641/0006-3568(2002)052%255B0669:DRCAOF%255D2.0.CO;2)
10. Igwaran, A., Kayode, A. J., Moloantoa, K. M., Khetsha, Z. P., & Unuofin, J. O. (2024). Cyanobacteria Harmful Algae Blooms: Causes, Impacts, and Risk Management. *Water, Air, & Soil Pollution*, 235(1), 71. <https://doi.org/10.1007/s11270-023-06782-y>
11. Isaak, D. J., & Young, M. K. (2023). Cold-water habitats, climate refugia, and their utility for conserving salmonid fishes. *Canadian Journal of Fisheries and Aquatic Sciences*, 80(7), 1187–1206. <https://doi.org/10.1139/cjfas-2022-0302>
12. Jiang, Y., Guo, C., Su, F., Xu, W., Ma, L., Cui, L., & Mi, C. (2024). Climate warming effects on temperature structure in lentic waters: A bibliometric analysis from the recent 20 years. *Ecological Indicators*, 167, 112740. <https://doi.org/10.1016/j.ecolind.2024.112740>
13. Kim, J., Kwak, J., Ahn, J. M., Kim, H., Jeon, J., Kim, K., Kim, J., Kwak, J., Ahn, J. M., Kim, H., Jeon, J., & Kim, K. (2022). Oscillation Flow Dam Operation Method for Algal Bloom Mitigation. *Water*, 14(8). <https://doi.org/10.3390/w14081315>

14. Lokman, A., Ismail, W. Z. W., Aziz, N. A. A., Lokman, A., Ismail, W. Z. W., & Aziz, N. A. A. (2025). A Review of Water Quality Forecasting and Classification Using Machine Learning Models and Statistical Analysis. *Water*, 17(15). <https://doi.org/10.3390/w17152243>
15. Mejia, F. H., Ouellet, V., Briggs, M. A., Carlson, S. M., Casas-Mulet, R., Chapman, M., Collins, M. J., Dugdale, S. J., Ebersole, J. L., Frechette, D. M., Fullerton, A. H., Gillis, C.-A., Johnson, Z. C., Kelleher, C., Kurylyk, B. L., Lave, R., Letcher, B. H., Myrvold, K. M., Nadeau, T.-L., ... Torgersen, C. E. (2023). Closing the gap between science and management of cold-water refuges in rivers and streams. *Global Change Biology*, 29(19), 5482–5508. <https://doi.org/10.1111/gcb.16844>
16. Null, S. E., Mouzon, N. R., & Elmore, L. R. (2017). Dissolved oxygen, stream temperature, and fish habitat response to environmental water purchases. *Journal of Environmental Management*, 197, 559–570. <https://doi.org/10.1016/j.jenvman.2017.04.016>
17. Qiu, R., Wang, Y., Rhoads, B., Wang, D., Qiu, W., Tao, Y., & Wu, J. (2021). River water temperature forecasting using a deep learning method. *Journal of Hydrology*, 595, 126016. <https://doi.org/10.1016/j.jhydrol.2021.126016>
18. Rubin, Z., Kondolf, G. M., Rios-Touma, B., Rubin, Z., Kondolf, G. M., & Rios-Touma, B. (2017). Evaluating Stream Restoration Projects: What Do We Learn from Monitoring? *Water*, 9(3). <https://doi.org/10.3390/w9030174>
19. Salvatierra, D., González, M. P., Blasco, J., Krull, M., & Araújo, C. V. M. (2025). Habitat loss and discontinuity as drivers of habitat fragmentation: The role of contamination and connectivity of habitats. *Environmental Research*, 266, 120609. <https://doi.org/10.1016/j.envres.2024.120609>
20. Schonlau, M., & Zou, R. Y. (2020). The random forest algorithm for statistical learning. *The Stata Journal*, 20(1), 3–29. <https://doi.org/10.1177/1536867X20909688>
21. Spinti, R. A., Condon, L. E., & Zhang, J. (2023). The evolution of dam induced river fragmentation in the United States. *Nature Communications*, 14(1), 3820. <https://doi.org/10.1038/s41467-023-39194-x>
22. Trabelsi, A., Chaabane, M., & Ben-Hur, A. (2019). Comprehensive evaluation of deep learning architectures for prediction of DNA/RNA sequence binding specificities. *Bioinformatics*, 35(14), i269–i277. <https://doi.org/10.1093/bioinformatics/btz339>
23. Yavuz, V. S. (2025). Impact of temperature and flow rate on oxygen dynamics and water quality in major Turkish rivers. *Scientific Reports*, 15, 22830. <https://doi.org/10.1038/s41598-025-06433-8>