Journal of Environmental Management 392 (2025) 126592

Contents lists available at ScienceDirect

Journal of Environmental Management

journal homepage: www.elsevier.com/locate/jenvman

Research article

Optimizing models for the prediction of one step ahead extreme flows to wastewater treatment plants using different synthetic sampling methods

Isaac G. Musaazi [a], Lu Liu [b], Andrew Shaw [c], Marta Zaniolo [a], Lauren B. Stadler [d], Jeseth Delgado Vela [a,*]

[a] Department of Civil and Environmental Engineering, Duke University, BOX 90287, Durham, NC, 27708, USA
[b] Department of Civil, Construction and Environmental Engineering, Iowa State University, Ames, IA, 50011, USA
[c] Black and Veatch, 4400 Post Oak Pkwy, Suite 1200, Houston, TX, 77027, USA
[d] Department of Civil and Environmental Engineering, Rice University, 6100 Main Street, Houston, TX, 77005, USA

ARTICLE INFO

Keywords:
High flows
Water resource recovery facilities
Wet weather events
Machine learning
ADASYN
SMOTE
Antecedent dry days

ABSTRACT

High-flow events that significantly impact Water Resource Recovery Facility (WRRF) operations are rare, but accurately predicting these flows could improve treatment operations. Data-driven modeling approaches could be used; however, high flow events that impact operation are an infrequent occurrence, providing limited data from which to learn meaningful patterns. The performance of a statistical model (logistic regression) and two machine learning (ML) models (support vector machine and random forest) were evaluated to predict high flow events one-day-ahead to two plants located in different parts of the United States, Northern Virginia and the Gulf Coast of Texas, with combined and separate sewers, respectively. We compared baseline models (no synthetic data added) to models trained with synthetic data added from two different sampling techniques (SMOTE and ADASYN) that increased the representation of rare events in the training data. Both techniques enhanced the sample size of the very high-flow class, but ADASYN, which focused on generating synthetic samples near decision boundaries, led to greater improvements in model performance (reduced misclassification rates). Random forest combined with ADASYN achieved the best overall performance for both plants, demonstrating its robustness in identifying one-day-ahead extreme flow events to treatment plants. These results suggest that combining sampling techniques with ML has the potential to significantly improve the modeling of high-flow events at treatment plants. Our work will prove useful in building reliable predictive models that can inform management decisions needed for the better control of treatment operations.

1. Introduction

The increasing frequency of intense rainfall events caused by climate change presents significant challenges to wastewater treatment plants. These events can result in flows that exceed plant capacity, placing considerable strain on sewer infrastructure, particularly in aging systems (Abdellatif et al., 2014; Botturi et al., 2021; Schertzinger et al., 2019). As a result, these challenges increase the risk of discharging partially treated or untreated water to the environment, which poses health risks and potential violations of regulatory permits (McLellan et al., 2007; Newton et al., 2013; Sibanda et al., 2015). To mitigate these risks and maintain treatment efficacy, plant operators must anticipate and respond to threshold-exceeding flows that trigger specific operational interventions. For instance, flows exceeding critical thresholds such as the maximum capacity of secondary treatment processes often force operators to provide minimum treatment involving primary clarification followed by disinfection (Peters and Zitomer, 2021). The effective management of such strategies can be greatly enhanced by predictive models serving as early warning systems.

Machine learning (ML) models have recently emerged as robust tools for various flow prediction tasks that necessitate adaptive management strategies, often framing the problem as a regression task that yields continuous numerical estimates (Kanneganti et al., 2022; Zhou et al., 2022). This is because ML models are simple to build and can expose patterns from historical data without the need for prior knowledge of the physical, biological, or chemical processes involved. However, precise flow predictions do not always translate directly into actionable plant decisions, especially under extreme conditions where small variations in

* Corresponding author.
E-mail address: jeseth.delgadovela@duke.edu (J. Delgado Vela).

https://doi.org/10.1016/j.jenvman.2025.126592
Received 11 April 2025; Received in revised form 10 July 2025; Accepted 12 July 2025
Available online 21 July 2025

estimated values have little operational significance. Instead, classification-based models have potential for a more effective decision-support framework by categorizing flows into operationally relevant levels ensuring that predictions align with plant management needs. Unlike regression, which focuses on numerical precision, classification models inherently support threshold-based decision-making and can better support the control of plant operations.

Extreme flow classification can be challenging to predict in that these events are rare; the resulting scarcity in data creates a significant imbalance between normal and extreme flow cases. To overcome the data challenge, data sampling approaches that increase representation of rare extreme events by generating additional synthetic samples, have gained considerable attention in recent years (Agrawal and Petersen, 2021; Almeida and Coelho, 2023; Han et al., 2019; Sharififar and Sarmadian, 2023). Integrating data sampling with ML models has proven effective in addressing imbalanced datasets in water research (Bourel et al., 2021; Xu et al., 2020). Meiska et al. (2023) demonstrated that the Synthetic Minority Over-sampling Technique (SMOTE) significantly improved recall and F1-scores. Recall measures the proportion of actual positive cases correctly identified by the model, while the F1-score represents the harmonic mean of precision (the proportion of predicted positive cases that are correct) and recall. These improvements were observed across several models used to assess drinking water quality. Similarly, Jeong et al. (2022) applied SMOTE to predict harmful algal blooms (HABs) in South Korean reservoirs. There was a marked improvement in the recall and F1-scores which were critical in detecting bloom conditions amidst predominantly non-bloom cases. Another notable example was combining a Deep Cascade Forest (DCF) model with SMOTE to identify water that fails to meet established quality standards for consumption defined by the exceedance of acceptable thresholds (Chen et al., 2021). This study reported F1-scores as high as 94.7%, significantly outperforming baseline models without data sampling. To our knowledge, there have been no published examples of the application of SMOTE to identify future flows to WRRFs.

In addition to SMOTE, Adaptive Synthetic Sampling (ADASYN) has been applied to improve predictions in water quality management such as predicting exceedances of Fecal Indicator Bacteria (FIB) in water (Xu et al., 2020). ADASYN increased the ability of the models to predict above-threshold FIB values to over 75%, while maintaining a high specificity of over 90% for below-threshold predictions. A comparative study evaluated the performance of SMOTE and ADASYN for predicting the presence of *Escherichia coli* MG1655 in spinach wash water (Stanosheck et al., 2024). SMOTE outperformed ADASYN, achieving 90% accuracy and 93.8% sensitivity with a Random Forest (RF) model, compared to ADASYN which achieved 86.5% accuracy and 87.5% sensitivity. However, both techniques significantly improved model performance compared to baseline.

We coupled ML with training data enriched with information from data sampling techniques since it remains underexplored in the context of wastewater systems. This gap presents a valuable opportunity to investigate their effectiveness in improving the prediction of extreme flow events. As such, our study evaluates the effectiveness of SMOTE and ADASYN in improving the one-day-ahead prediction of extreme flow events at plants. To this end, the study focuses on addressing the following research questions: (1) Which statistical or machine learning (ML) models perform best at identifying high flow events? (2) Can the performance of these models improve when coupled with data sampling techniques? The outcomes of this research will inform the development of reliable predictive models capable of assisting operators in better managing extreme flow events, thereby reducing the risk of operational disruptions and regulatory violations.

2. Materials and methods

2.1. Wastewater facilities and data collection

Data was obtained from regular monitoring of two WRRFs, one located in Northern Virginia (referred to as Plant I hereinafter), and another in the Gulf Coast of Texas (referred to as Plant II hereinafter). These two plants are located in regions that are anticipated to encounter more extreme precipitation events in the future (Easterling et al., 2017). Plant I has a design flow capacity of 54 million gallons per day (MGD) and is served by a combined sewer. The plant managers provided a total of 1461 consecutive historical average daily flow records and rainfall data. The cumulative time for the flow data was 48 months (i.e., from January 2017 to December 2020). Plant II has a design flow capacity of 7.25 MGD, a design 2-h peak flow of 50 MGD, and is served by a separate sewer. The plant managers provided 1125 consecutive historical average daily flow records and rainfall. The cumulative number of days for the flow data was 37 months (i.e., from June 2016 to June 2019).

The flow data was classified into three distinct categories: normal, high, and very high based on statistical thresholds derived from the distribution of flow. Table S1 shows the summary statistics (average, median, standard deviation, and maximum values) of the data provided for both Plants I and II. *Very high flows* were defined as the top 3% of the flow data and represented the extreme events (Fig. S1). *High flows* were defined as the next 15% of flows, capturing significant, but less extreme, flow events. The remaining 82% of flows were classified as *normal*, representing typical operating conditions. This methodology was consistent with hydrological practices, where percentiles and flow-duration curves are commonly used to define flow regimes (Vogel and Fennessey, 1994). In addition, for Plant I, these classifications were appropriate for example, the very high flow, ranged between 60 and 99.5 MGD, which was approximately 1.64–2.71 times the average flow (36.68 MGD). Additionally, we verified that these classifications were appropriate because storm events categorized as 10-year storms and more severe were observed at Plant I in our dataset and resulted in flows above 60 MGD (Table S2). For Plant II, the minimum value for the very high flow (12.06 MGD) was approximately 3.31 times the average flow (3.64 MGD) and 1.66 times the permitted 2-h peak flow (7.25 MGD). The very high flow classification was 3.45 times higher than a previously established 3.5 MGD threshold used to distinguish dry from wet weather flow for this particular plant (Liu et al., 2025). As such, these classifications aligned with operational thresholds that might necessitate activation of different wet weather management protocols.

Rainfall data was used to calculate the antecedent dry days (ADD) using Equation (1), where x_t is the rainfall amount at the current timestamp, x_{t-1} is the rainfall amount at the previous timestamp, and n is the total number of timestamps. However, daily rainfall below 1 mm (~0.04 inches) was set as the threshold. This is because light rain can either be absorbed by the soil or evaporate into the atmosphere before stormwater reaches runoff levels that can drain to storm sewers (Groisman and Knight, 2008). A count register of the ADD was reset to zero whenever the rainfall at x_t was greater than 0.04 inches (i.e., each rain event led to a start over in the count of ADD, in days).

$$ADD_i = \begin{cases} 1 + ADD_{i-1} & 0 < x_t \leq 0.04 \\ 0 & x_t > 0.04 \\ 0 & \mathrm{t} = 1 \text{ and } x_1 > 0.04 \end{cases} \tag{1}$$

Timestamps were used to extract features that capture flow variation over different time intervals (Zhou et al., 2019). The month of the year was extracted from each timestamp to capture the meteorological seasons that have been shown to affect flow in combined sewers (Andreides et al., 2022; Borzooei et al., 2019). This could be due to snow melting in the spring, storm events in the summer, and seasonal changes in water consumption patterns. The seasonal variable was treated as categorical and comprised of four seasons: summer (June 1–August 31), fall (September 1–November 30), spring (March 1–May 31), and winter

(December 1–February 28). Rolling statistics were also calculated for flow and rainfall using window sizes from two to six days to capture short-term trends and variability. These smoothed the time series data by reducing the influence of daily fluctuations and highlighted localized patterns relevant to flow prediction. In summary, the model inputs included: ADD derived from rainfall data, seasonal variations categorized by meteorological seasons, and rolling statistics for flow and rainfall. These inputs were designed to represent key hydrological and temporal factors influencing flow dynamics. For each classification task, these inputs were at time, t with the goal of predicting the flow class at time, $t+1$. This approach ensured that the models used current conditions while predicting future flow events (one-day-ahead), providing operators with actionable lead time to implement mitigation strategies.

2.2. Data sampling techniques used in this study

We applied two data-level methods, SMOTE and ADASYN, to generate synthetic data for the minority class. SMOTE interpolates between a minority class instance and its k-nearest neighbors (Chawla et al., 2002). However, since our data was a time series, we had to account for the temporal characteristics of the data and generate synthetic samples that respect the temporal relationships inherent in the data. Briefly, the total number of synthetic samples, G_s, was calculated using Equation (2):

$$G_s = m_s \times (R-1) \tag{2}$$

where G_s is the total number of synthetic samples to generate, m_s is the number of minority class samples, and R is the imbalance ratio. In this study, $R = \frac{m_l}{m_s}$, where m_l is the number of majority class samples. The goal was to generate synthetic samples such that $m_l = m_s$ such that Equation (3) is satisfied.

$$m_l = m_s + G_s \tag{3}$$

For each minority sample, x_i a set of valid neighbors (K) in the feature space was determined based on N_i which represented the set of valid neighbors (x_j) for x_i. This set was defined using the Euclidean distance, d, the absolute temporal index difference, $|j- \quad i|$, and the temporal window size, ΔT, between the feature vectors x_i and x_j. In simpler terms, the selection of neighbors considered the temporal closeness. The synthetic samples, S_i were generated using Equation (4):

$$S_i = x_i + (x_j - x_i) \times \lambda \tag{4}$$

Where λ is a random interpolation number uniformly distributed between 0 and 1.

Adaptive Synthetic Sampling (ADASYN) algorithm was another method we tested to address class imbalance in our study. ADASYN focused on instances of the minority class that are harder for the classifier to learn. It determines the learning difficulty of each minority instance by analyzing the ratio of majority class samples in its local neighborhood (Gameng et al., 2019; He et al., 2008). However, we introduced a temporal constraint, ΔT, to synthetically generate samples based on temporally relevant data points. Briefly, the total number of synthetic samples, G_A was calculated using Equation (5):

$$G_A = (m_l - m_s) \times \beta \tag{5}$$

Where G_A is the total number of synthetic samples to be generated and β was determined dynamically based on the imbalance ratio, R. However, we implemented a capping mechanism to limit the number of synthetic samples. While this cap may result in the minority class size being slightly smaller than the majority class ($m_s + G_A < m_l$), it was introduced to address challenges associated with oversampling techniques. Specifically, data sampling strategies often perform poorly near the boundaries of the minority class distribution and may inadvertently duplicate existing minority samples, which can reduce the quality and diversity of the generated data (Gameng et al., 2019; Sakho et al., 2024).

For each minority sample, x_i k-nearest neighbors were identified using the Euclidean distance in feature space, and the r_i ratio calculated using Equation (6):

$$r_i = \frac{\Delta_i}{K}, i = 1, 2, \ldots, m_s \tag{6}$$

where Δi is the number of minority samples in the generated cluster from x_i by K number of neighbors. A temporal window constraint was applied to ensure that only neighbors x_j satisfying Equation (7) were considered valid:

$$|j - i| \leq \Delta T \tag{7}$$

The r_i values were normalized to ensure proportional weighting of minority samples using Equation (8):

$$\widehat{r}_i = \frac{r_i}{\sum_{i=1}^{m_s} r_1} \tag{8}$$

The number of synthetic data samples generated for each minority sample, x_i were calculated using Equation (9):

$$g_i = \widehat{r}_i \times G_A \tag{9}$$

Synthetic data, s_i was generated using Equation (4).

2.3. Machine learning model development

We compared a statistical logistic regression (LR) model alongside two machine learning models: support vector machines (SVM) and random forest (RF). A description of each of these models can be found in Supplementary Information, Text S1. Seventy percent of the data was used for training the models and the remaining 30% was used for model testing. This split ratio is commonly used in wastewater prediction models, where it has been associated with lower prediction errors (Hamada et al., 2024; Nallakaruppan et al., 2024). We used a time series train and test split to ensure that future data did not inform past predictions, and that model performance was not inflated beyond the actual predictive capability of the model (Hyndman and Athanasopoulos, 2018). The set of variables in the training data set were also scaled to prevent biased estimation and achieve consistency across all models, as detailed in Text S2.

The grid search method was applied to optimize the model hyperparameters (Injadat et al., 2020). A predefined hyperparameter space comprising a set of values was obtained by trying different combinations of all the hyperparameter values presented in Table S3. To avoid any potential overfitting issues (when the error on the internal test set was significantly greater than the training error) and problems induced by the inappropriate partition of datasets that are common with ML models, a time series cross validation analysis with an expanding window was conducted to determine the hyperparameters. Validation data used to generate hyperparameters was iteratively used for model training, but model testing was independent of hyperparameter optimization (Fig. 1) (Nielsen, 2019). The best set of hyperparameters that minimized the training error was used for model training (Table S4). By adopting these optimized hyperparameters, the performance of the model was evaluated based on the internal test set (validation) before testing on unseen data.

The trained models were then applied to a test set to assess the generalization performance (ability to predict unknown datasets). The performance of the models was conducted using a combination of visual and quantitative metrics to comprehensively assess performance (see Text S3). Confusion matrices were plotted for each model providing insights into true positives, true negatives, false positives, and false negatives (Powers, 2011). Mathews Correlation Coefficient (MCC) was used to provide a balanced assessment of performance across all classes

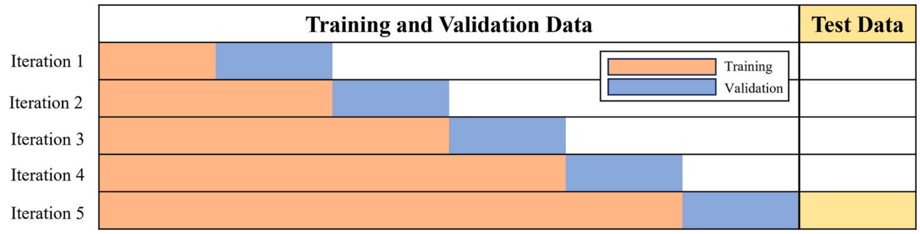

Fig. 1. A time series cross validation on a rolling basis involved starting with a small training set, predicting on the internal test set (validation), and then checking the model performance during training. The prior training and internal test sets were then included as part of the next training set and subsequent test set was predicted. The data points from each training were before the corresponding internal test set containing unique data points of the same size. This led to five possible groups of training sets and validation sets.

(Chicco and Jurman, 2020). These combined approaches ensured a thorough and interpretable evaluation of model performance and facilitated comparison with findings from other classification studies. All data analysis was conducted in Google Colaboratory.

3. Results and discussion

3.1. Models failed to predict the very high flow class at both plants because of data scarcity

The performance evaluation across models for Plant I and Plant II highlighted differences in predicting the three flow classes considered in our study: normal, high, and very high. For the normal flow class, all models consistently demonstrated strong performance during testing. LR and SVM achieved precision (0.94–095), recall (0.72–0.74), F1-scores (0.82–0.83), and MCC (0.33–0.34) at Plant I (Fig. S2). These results show the capacity of the models to accurately learn patterns for normal flow events, which are well-represented during training. In contrast, the very high flow events (11 in total) posed the most significant challenge for all models due to the small sample size. LR performed poorly with a low precision (0.10), recall (0.36), F1-score (0.16), and MCC (0.15). RF also exhibited poor performance, with precision (0.10), recall (0.18), F1-score (0.13), and MCC (0.10). This highlighted the inability of RF to correctly classify very high flow events. The confusion matrices for the very high flow class revealed a consistent trend: all models frequently misclassified these events as high flow (Fig. S3). For example, RF misclassified three very high flow events as high. LR misclassified five and two very high flow events as normal and high, respectively. The overall failure of RF to generalize the very high flow classes likely stems from its inherent bias toward the high flow class predictions that influenced decision-making across trees. Similar findings were observed in a prior study where RF failed to predict rare events due to a relatively small sample size (Ai et al., 2023; Shin et al., 2021).

For Plant II, a similar trend emerged as observed in Plant I. The normal flow class predictions achieved excellent results across all models (Fig. S4). For example, LR predicted 209 out of 261 instances (Fig. S5), yielding a recall (0.80), F1-score (0.87), and MCC (0.60). However, predictions for the very high flow class remained problematic across all models. For example, SVM – precision (0.18), recall (0.38), F1-score (0.24), and MCC (0.20) and RF – precision (0.11), recall (0.31), F1-score (0.16), and MCC (0.11) demonstrate poor performance in predicting the very high flow class. RF misclassified 10 out of 16 very high flow events as high flows and one event was classified as normal indicating a significant bias against the high flow class (Fig. S5). Taken together, the findings emphasize robust model performance, particularly for very high flow conditions, depends on sufficient representation in the training data.

3.2. SMOTE and ADASYN effectively addressed class imbalance in the training data by generating synthetic samples for the underrepresented minority class

Given the challenges with all models in predicting critical but rare very high flow events at plants, we tested SMOTE and ADASYN as data sampling techniques in our study. Both methods were able to generate synthetic samples for the minority class by interpolating between existing minority class instances. However, as demonstrated in Fig. 2, SMOTE and ADASYN differed significantly in how they prioritized the placement of the synthetic samples across both treatment plant datasets.

For both plants, the synthetic samples generated with SMOTE were distributed more broadly across the feature space (red crosses, Fig. 2a and c). This was expected because SMOTE interpolates between all available minority class samples. We used temporal constraints and a capping mechanism in our study to limit the synthetic samples generated based on valid neighbors (x_j). This approach helped mitigate a common issue in SMOTE, where synthetic data may reinforce existing minority class patterns, resulting in ambiguous decision boundaries (Li et al., 2024; Salehi and Khedmati, 2024; Sreejith et al., 2020).

In contrast to SMOTE, ADASYN generated synthetic samples that were tightly clustered around existing minority class instances, focusing specifically on regions where class boundaries were difficult to learn (red crosses, Fig. 2b and d). This localized sampling approach allowed ADASYN to more aggressively address areas of imbalance, particularly where minority samples were located near decision boundaries. While this made ADASYN effective in improving recognition of hard-to-classify minority cases, it also introduced limited diversity in the synthetic data created, leading to over-representation in certain regions. Moreover, the dynamic weighting mechanism used in ADASYN to determine the number of synthetic samples generated for each minority instance based on the proportion of majority class neighbors further concentrated sampling near class boundaries. This limitation has previously been reported in the literature (Gameng et al., 2019; Khan et al., 2023; Kurniawati et al., 2018). In addition, feature-space overlap between classes was observed in the ADASYN-generated data. While the temporal window constraint we implemented partially mitigated this issue, it did not explicitly account for class separation in the feature space. This highlights the need for future work to explore hybrid approaches and parameter tuning strategies to improve the effectiveness of ADASYN in generating more diverse and separable synthetic samples.

Despite these limitations, both SMOTE and ADASYN improved class balance at both plants, offering value for downstream classification tasks. However, it is important to recognize that class imbalance was particularly severe for the very high flow events at Plant I, which had only 32 data points compared to 125 high flow events. Therefore for both approaches, synthetic samples were generated for the very high flow classes, but ADASYN generated fewer total synthetic samples (Fig. 3). This skewed distribution posed a significant predictive challenge for both the statistical and ML models at both plants, as discussed in section 3.1. The integration of synthetic sampling methods substantially altered the representation of flow classes during model training.

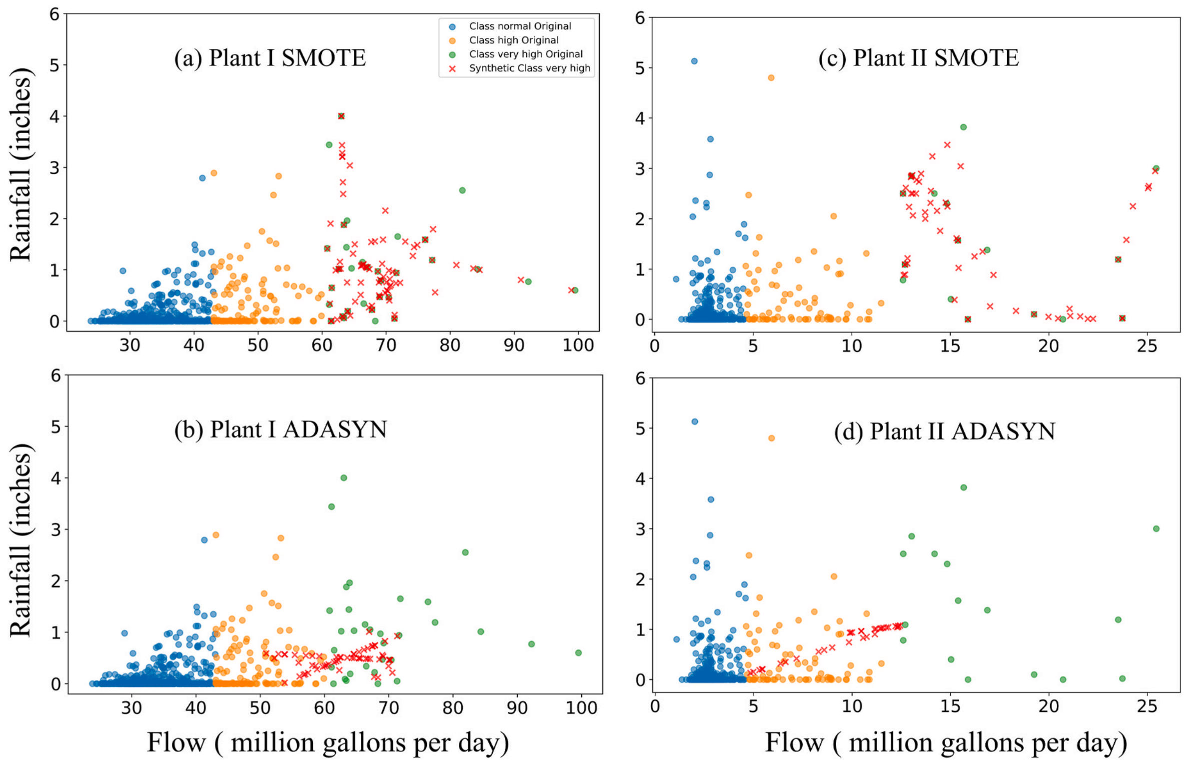

Fig. 2. Scatterplots show the distribution of synthetic data generated by different sampling methods. Flow values correspond to time (t+1), while rainfall is from time (t). High or very high flows under low or no rainfall conditions may reflect lagged hydrologic responses or upstream rainfall not captured by the local rain gauge. Subfigures (a) and (b) represent SMOTE and ADASYN for Plant I, while (c) and (d) represent SMOTE and ADASYN for Plant II. Red crosses indicate the synthetic samples generated by both methods, which improved the balance of the training data. Feature values for synthetic samples were clipped to observed minimum and maximum ranges to prevent unrealistic or out-of-bound values.

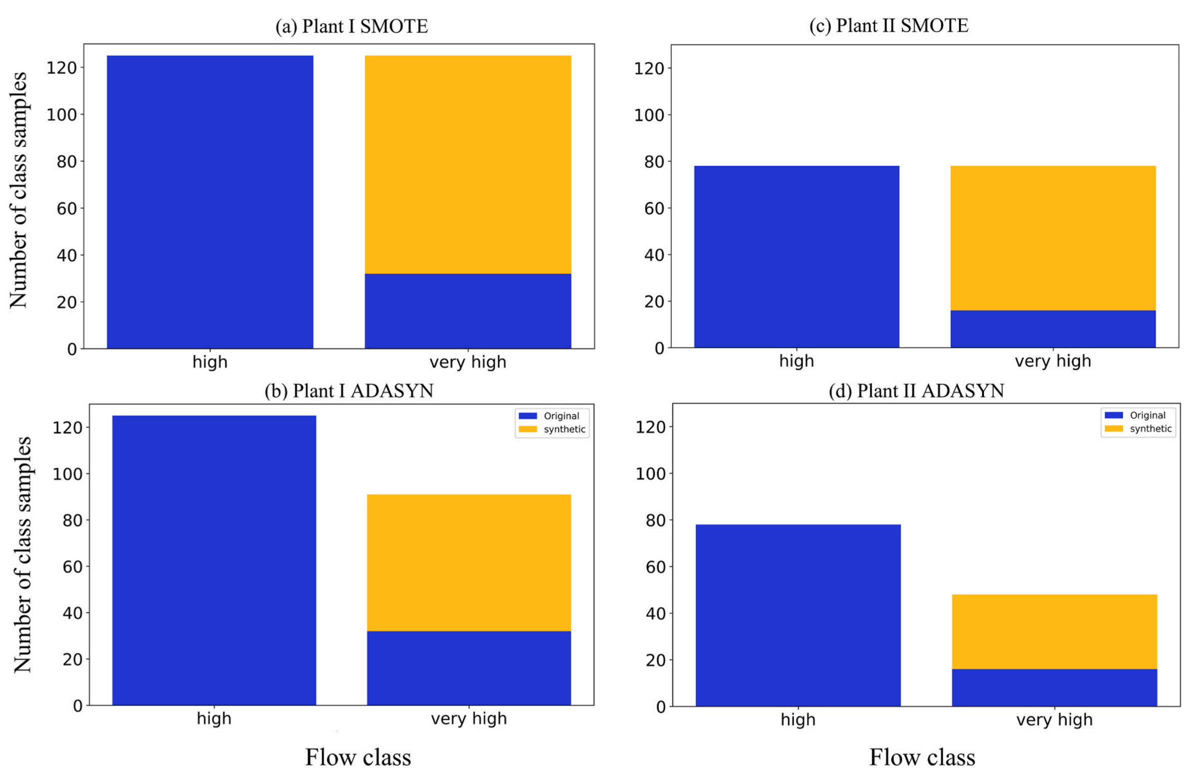

Fig. 3. Class distribution of original and synthetic data using SMOTE and ADASYN at Plant I [(a) and (b)] and Plant II [(c) and (d)], respectively.

For example, SMOTE addressed this issue by increasing the very high flow class count to 125 (Fig. 3a). This ensured that both high and very high-flow classes were equally represented. While this addressed one aspect of the problem, the reliance on the number of nearest neighbors (K) to guide interpolation introduced another trade-off: a small K may fail to identify meaningful neighbors in sparse regions, while a large K risks incorporating noisy or irrelevant samples. Future work should aim to improve the distribution and quality of synthetic samples. This can involve enhancing the current sampling method by classifying minority class samples into "safe" and "noise" points. Qu and Zhang (2024) discussed this approach to help prioritize reliable samples for sampling and leveraging the minority-class centroid to guide synthetic sample generation.

ADASYN targeted the minority class in a more adaptive manner, and initially, the high (majority) to very high (minority) class comparison was defined by the following: R = 0.26 and β = 1.488. Here, R quantified the global imbalance, while β determined the number of synthetic samples (G_A) to be generated. Unlike SMOTE, which focused on evenly increasing minority class representation, ADASYN dynamically identified sparse regions in the minority class that were closer to decision boundaries (high vs very high boundary), computed the required number of synthetic samples using β, and then allocated those samples preferentially to "hard-to-learn" minority instances specifically, those surrounded by many majority class neighbors. An additional 59 synthetic samples were generated, making a total of 91 very high flow classes in the training data (Fig. 3b). For Plant II, a similar challenge emerged, with the very high flow class having only 16 samples compared to the high flow class (78 samples). SMOTE was able to increase the very high flow class count to 78 (Fig. 3c). ADASYN again took a dynamic approach and by focusing on sparse regions closer to decision boundaries ($R = 0.21; \beta = 1.59$) and generated 32 synthetic samples. These were not distributed uniformly but were instead concentrated around difficult decision boundaries where misclassification rate was highest, which improved the local representation of this class making a total of 48 samples in the training data (Fig. 3d). In both cases, the use of β ensured that the augmentation volume matched imbalance severity, while the neighborhood-based weighting directed attention to the most informative areas. However, while ADASYN weighted minority samples based on nearby majority class instances, it may have introduced redundant synthetic points that focused too narrowly on overlapping regions between the high and very high flow classes. In addition, the reliance on K to identify suitable neighbors for synthetic sample generation and the temporal window constraint, ΔT might have led to instances with limited neighbor availability, particularly for the very high flow samples located farther away from the class boundaries. As a result, no synthetic samples were generated for these instances. Future work could explore advanced sampling strategies that prioritize the generation of diverse and well-distributed synthetic samples. This will support the idea that strategically placed synthetic samples are critical for correcting imbalance and enhancing predictive performance. For example, incorporating preference mechanisms designed to resolve ambiguities in neighbor selection and prioritizing specific values based on class distribution have been shown to improve class separability (Kurniawati et al., 2018).

3.3. ADASYN consistently outperformed SMOTE for very high-flow predictions across all models at both plants, with the most notable improvements observed in ML models

The prediction of the very high flow class varied significantly across models and data sampling techniques. This variability highlighted how sensitive minority class performance is to both the sampling strategy and the model architecture. Differences in how effectively SMOTE and ADASYN addressed the minority class based on the performance of the RF model are highlighted in Fig. 4, which shows that during model testing, overall model performance was improved more using ADASYN,

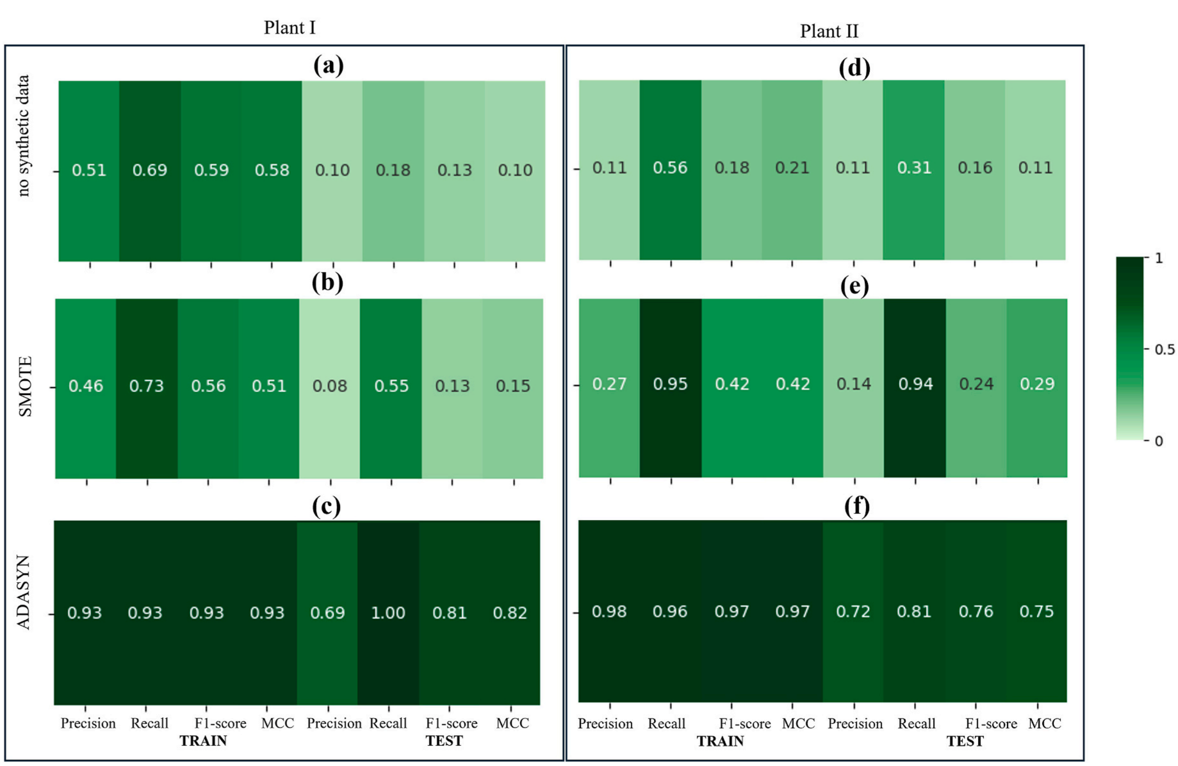

Fig. 4. Comparison of training and testing performance metrics (Precision, Recall, F1-Score, MCC) for Random Forest (RF) models during very high flow events at Plant I and Plant II. Model performance without data sampling during training for Plant I and Plant II, respectively (first row). Performance when SMOTE is used to balance the training set (second row). The performance when ADASYN is used for data sampling. The best RF performance is achieved when RF is coupled with ADASYN for both plants (third row).

as compared to both baseline (no synthetic data) and when SMOTE was used to generate synthetic samples.

At Plant I, the RF model exhibited low performance in identifying very high flow events in the test data when the training data included synthetic samples generated by SMOTE (Fig. 4b). This was likely due to suboptimal hyperparameter settings that led the model into poorly performing regions of the grid search space (see Table S3). Grid search is known to exhaustively test parameter combinations without adaptively focusing on promising regions, which can limit model performance (Yang and Shami, 2020). Additionally, SMOTE is limited in its ability to differentiate between representative minority samples and noisy or outlier points (Imani et al., 2024; Matharaarachchi et al., 2024). Also, the reliance on a randomly generated λ might have contributed to potential redundancy by producing synthetic samples that were similar to the original data. As a result, SMOTE may have generated synthetic samples around noisy instances, thereby distorting the overall data distribution and negatively impacting model performance. This issue is particularly problematic for tree-based models like RF, which rely on recursive splitting. When trained on synthetic data with poorly defined boundaries, certain trees in the ensemble may overfit to these artificial patterns, especially when synthetic points reinforce class overlap or noise (Usman-Hamza et al., 2024). Other models trained with SMOTE showed similarly limited performance. For instance, LR achieved recall (0.36), F1-score (0.09), and MCC (0.08) (Fig. S6), with the confusion matrix revealing misclassification of two very high flow instances as high (Fig. S7).

A similar trend was observed at Plant II, where RF performance with SMOTE during model testing was also low, illustrated by precision (0.14), recall (0.94), F1-score (0.24), and MCC (0.29) values (Fig. 4e). These metrics suggest that while the model captured most positive instances (high recall), it failed with precision, leading to poor overall class discrimination. LR performed poorly with recall (0.31), F1-score (0.34), and MCC (0.32) (Fig. S8), where eight very high flow instances were misclassified as high (Fig. S9). The SVM model also had a poor performance, with recall (0.25), F1-score (0.15), and MCC (0.10), correctly identifying five out of 16 observed very high flow events, while misclassifying the remaining 11 as high. The optimal SVM configuration selected was a linear kernel, which is effective for linearly separable data and offers computational efficiency. In contrast, other kernels such as RBF (Gaussian), polynomial, and sigmoid are designed to handle nonlinear boundaries by projecting data into higher-dimensional feature spaces, enabling the model to capture more complex patterns within the data (Savas and Dovis, 2019). However, the choice of kernel should not be fixed and must depend on specific data characteristics (Patle and Chouhan, 2013). The linear kernel was chosen based on the characteristics of our data and the ability to balance complexity and performance. Overall, these results from both plants indicate that SMOTE was insufficient for robust minority class learning, particularly for models that are sensitive to noise in the input space. Future work should adopt noise-filtering mechanisms that remove noisy or borderline examples to ensure that synthetic samples are representative of the true minority class distribution. The effectiveness of generating near-class-border samples (beyond the scope of our work), synthesizing additional samples along temporal trajectories, and employing weighted sampling to mitigate noise have previously been shown to be effective (Zhao et al., 2022). Integrating such approaches during sampling to further leverage the data characteristics should be adopted in future studies.

The application of ADASYN led to a significant improvement in identifying the very high flow across all models, which demonstrated its advantage over SMOTE. ADASYN at Plant I substantially improved performance across all models, with RF demonstrating the best performance with precision (0.69), recall (1.00), F1-score (0.81), and MCC (0.82) (Fig. 4c). Similarly, RF coupled with ADASYN at Plant II was the best model with recall (0.81), F1-score (0.76), and MCC (0.75) (Fig. 4f). The adaptive strategy helped shift the decision boundary toward the challenging regions of the feature space. By focusing on difficult-to-learn instances, ADASYN enhanced the ability of the model to generalize and improve model classification performance, particularly for the minority class. LR also achieved recall (0.91), F1-score (0.77), and MCC (0.77), while SVM excelled with recall (0.82), F1-score (0.67), and MCC (0.67) (Fig. S10). These results underscore the utility of ADASYN in leveraging a linear margin-based learning approach to achieve balanced and improved classification outcomes. In a previous study, ADASYN significantly enhanced the sensitivity (recall) of SVM models which enabled it to better identify the minority class (Pristyanto et al., 2022). Additional model performance details are provided in Fig. S12 and S13. In sum, these results emphasize the superiority of ADASYN over SMOTE for addressing severe class imbalance, caused by rare events at both plants. Tailored data sampling, therefore, offers a promising approach to improve the robustness of both statistical and machine learning models for underrepresented classes.

4. Conclusions

The prediction of rare or extreme events is increasingly important in the face of global climate change. Our study demonstrated the substantial challenges posed by data class imbalance in predicting very high flow events in wastewater treatment systems. Models trained without synthetic data from sampling had limited accuracy predicting these rare events. Sampling techniques, particularly ADASYN, proved to be important in model learning for better generalization of one-step ahead very high flow class predictions. This work will not only support decisions on improved upstream operations geared toward efficiently managing the fluctuating loads to plants but also serve to better manage flow to the plant unit processes usually needed to maintain desired treatment performance. However, there are limitations associated with our study that should be addressed in future studies. Expansion of the study area to consider different climate conditions both within and outside the U.S. would be interesting as both plants were in relatively temperate climates. With additional case studies where data on rare events is quite relevant, our study findings can further be validated.

The key findings are:

1. All models (LR, SVM, and RF) faced significant challenges in one-step ahead prediction of the very high flow class, with high misclassification rates caused by the small sample size of this class during training.
2. Sampling techniques increased the representation of the very high flow class in the training data, with ADASYN consistently outperforming SMOTE across all models and both plants.
3. RF combined with ADASYN delivered the best overall performance and proved to be the most effective model for identifying very high flow events in one-day-ahead prediction.

This work will contribute to the broader discussion on the benefits and limitations of data sampling while aiming to develop robust models for extreme flow prediction to treatment plants. The findings are expected to support informed decision-making in wastewater treatment management, strengthening plant resilience to variability and extreme weather conditions driven by climate change. Future research should focus on advancing synthetic sampling methods to improve the quality, diversity, and placement of synthetic samples, particularly in overlapping or sparse regions of the feature space. Additionally, tuning sampling parameters (e.g., $K, \Delta T, \lambda$) and incorporating temporal trajectories could further enhance model generalization and ensure that synthetic samples are generated in regions that reinforce class separability. This will improve applicability and effectiveness of sampling techniques across various real-world imbalanced datasets.

CRediT authorship contribution statement

Isaac G. Musaazi: Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Lu Liu:** Writing – review & editing, Validation, Software. **Andrew Shaw:** Writing – review & editing, Funding acquisition. **Marta Zaniolo:** Writing – review & editing, Validation, Software, Methodology. **Lauren B. Stadler:** Writing – review & editing, Project administration, Funding acquisition, Data curation. **Jeseth Delgado Vela:** Writing – review & editing, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was funded in part by the National Science Foundation Grant Opportunities for Academic Liaison with Industry (GOALI) under Grants 1931937 and 1932000. Additional support for this work was provided by the Gulf Coast Research Program. This work used Bridges-2 high-performance computing infrastructure system at the Pittsburgh Supercomputing Center (PSC) through allocation EVB210001 from Extreme Science and Engineering Discovery Environment (XSEDE), which was supported by the National Science Foundation Grant 1548562. The authors thank Dr. Dylan Christenson at Garver, Dr. Priyanka Ali at Black and Veatch, Jarret Morrison at Iowa State University, and Moriah Brown at University of Nebraska–Lincoln for the assistance and valuable discussions on developing the modeling framework. We are also grateful to Dr. Caleb Buahin at Xylem for initial insights and helpful comments during the early stages of model selection.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jenvman.2025.126592. The code used to run the models is available at https://github.com/Delgado-Vela-Research-Group/extremeflowmodeling

Data availability

Data will be made available on request.

References

Abdellatif, M., Atherton, W., Alkhaddar, R., 2014. Assessing combined sewer overflows with long lead time for better surface water management. Environ. Technol. 35, 568–580. https://doi.org/10.1080/09593330.2013.837938.

Agrawal, A., Petersen, M.R., 2021. Detecting arsenic contamination using satellite imagery and machine learning. Toxics 9. https://doi.org/10.3390/toxics9120333.

Ai, H., Zhang, K., Sun, J., Zhang, H., 2023. Short-term Lake Erie algal bloom prediction by classification and regression models. Water Res. 232, 119710. https://doi.org/10.1016/j.watres.2023.119710.

Almeida, M.C., Coelho, P.S., 2023. Modeling river water temperature with limiting forcing data: Air2stream v1.0.0, machine learning and multiple regression. Geosci. Model Dev. (GMD) 16, 4083–4112. https://doi.org/10.5194/gmd-16-4083-2023.

Andreides, M., Dolejš, P., Bartáček, J., 2022. The prediction of WWTP influent characteristics: good practices and challenges. J. Water Proc. Eng. 49, 103009. https://doi.org/10.1016/j.jwpe.2022.103009.

Borzooei, S., Teegavarapu, R., Abolfathi, S., Amerlinck, Y., Nopens, I., Zanetti, M.C., 2019. Data mining application in assessment of weather-based influent scenarios for a WWTP: getting the Most out of plant historical data. Water Air Soil Pollut. 230. https://doi.org/10.1007/s11270-018-4053-1.

Botturi, A., Ozbayram, E.G., Tondera, K., Gilbert, N.I., Rouault, P., Caradot, N., Gutierrez, O., Daneshgar, S., Frison, N., Akyol, Ç., Foglia, A., Eusebi, A.L., Fatone, F., 2021. Combined sewer overflows: a critical review on best practice and innovative solutions to mitigate impacts on environment and human health. Crit. Rev. Environ. Sci. Technol. 51, 1585–1618. https://doi.org/10.1080/10643389.2020.1757957.

Bourel, M., Segura, A.M., Crisci, C., López, G., Sampognaro, L., Vidal, V., Kruk, C., Piccini, C., Perera, G., 2021. Machine learning methods for imbalanced data set for prediction of faecal contamination in beach waters. Water Res. 202. https://doi.org/10.1016/j.watres.2021.117450.

Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. 16, 321–357. https://doi.org/10.1613/jair.953.

Chen, X., Liu, H., Xu, X., Zhang, L., Lin, T., Zuo, M., Huang, Y., Shen, R., Chen, D., Deng, Y., 2021. Identification of suitable technologies for drinking water quality prediction: a comparative study of traditional, ensemble, cost-sensitive, outlier detection learning models and sampling algorithms. ACS EST Water 1, 1676–1685. https://doi.org/10.1021/acsestwater.1c00037.

Chicco, D., Jurman, G., 2020. The advantages of the matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genom. 21. https://doi.org/10.1186/s12864-019-6413-7.

Easterling, D.R., Arnold, J.R., Knutson, T., Kunkel, K.E., LeGrande, A.N., Leung, L.R., Vose, R.S., Waliser, D.E., Wehner, M.F., 2017. Precipitation change in the United States. In: Climate Science Special Report: Fourth National Climate Assessment, vol. I. U.S. Global Change Research Program. https://doi.org/10.7930/J0H993CC.

Gameng, H., Gerardo, B., Medina, R., 2019. Modified adaptive synthetic SMOTE to improve classification performance in imbalanced datasets. In: 2019 IEEE 6th International Conference on Engineering Technologies and Applied Sciences (ICETAS), pp. 1–5. https://doi.org/10.1109/ICETAS48360.2019.9117287.

Groisman, P.Y., Knight, R.W., 2008. Prolonged dry episodes over the conterminous United States: new tendencies emerging during the last 40 years. J. Clim. 21, 1850–1862. https://doi.org/10.1175/2007JCLI2013.1.

Hamada, S.M., Zaqoot, A.H., Sethar, A.W., 2024. Using a supervised machine learning approach to predict water quality at the Gaza wastewater treatment plant. Environ. Sci. Adv. 3, 132–144. https://doi.org/10.1039/D3VA00170A.

Han, X., Cui, R., Lan, Y., Kang, Y., Deng, J., Jia, N., 2019. A Gaussian mixture model based combined resampling algorithm for classification of imbalanced credit data sets. International Journal of Machine Learning and Cybernetics 10, 3687–3699. https://doi.org/10.1007/s13042-019-00953-2.

He, H., Bai, Y., Garcia, E.A., Li, S., 2008. ADASYN: adaptive synthetic sampling approach for imbalanced learning. In: 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence). Hong Kong, China, pp. 1322–1328. https://doi.org/10.1109/IJCNN.2008.4633969.

Hyndman, R.J., Athanasopoulos, G., 2018. Forecasting: Principles and Practice, 2nd ed. https://otexts.com/fpp2/. Melbourne, Australia.

Imani, M., Ghaderpour, Z., Joudaki, M., Beikmohammadi, A., 2024. The impact of SMOTE and ADASYN on random Forest and advanced gradient boosting techniques in Telecom customer churn prediction. In: 2024 10th International Conference on Web Research, ICWR 2024. Institute of Electrical and Electronics Engineers Inc., pp. 202–209. https://doi.org/10.1109/ICWR61162.2024.10533320

Injadat, M.N., Moubayed, A., Nassif, A.B., Shami, A., 2020. Systematic ensemble model selection approach for educational data mining. Knowl. Base Syst. 200. https://doi.org/10.1016/j.knosys.2020.105992.

Jeong, B., Chapeta, M.R., Kim, M., Kim, J., Shin, J., Cha, Y.K., 2022. Machine learning-based prediction of harmful algal blooms in water supply reservoirs. Water Quality Research Journal 57, 304–318. https://doi.org/10.2166/wqrj.2022.019.

Kanneganti, D., Reinersman, L.E., Holm, R.H., Smith, T., 2022. Estimating sewage flow rate in Jefferson County, Kentucky, using machine learning for wastewater-based epidemiology applications. Water Supply 22, 8434–8439. https://doi.org/10.2166/ws.2022.395.

Khan, S.M., Nath, T.D., Hossain, M.M., Mukherjee, A., Hasnath, H.B., Manhaz Meem, T., Khan, U., 2023. Comparison of multiclass classification techniques using dry bean dataset. International Journal of Cognitive Computing in Engineering 4, 6–20. https://doi.org/10.1016/j.ijcce.2023.01.002.

Kurniawati, Y.E., Permanasari, A.E., Fauziati, S., 2018. Adaptive synthetic–nominal (ADASYN–N) and adaptive Synthetic–KNN (ADASYN- KNN) for multiclass imbalance learning on laboratory test data. In: Presented at the 2018 4th International Conference on Science and Technology (ICST), pp. 1–6. https://doi.org/10.1109/ICSTC.2018.8528679.

Li, H., Wang, S., Jiang, J., Deng, C., Ou, J., Zhou, Z., Yu, D., 2024. Augmenting the diversity of imbalanced datasets via multi-vector stochastic exploration oversampling. Neurocomputing 583. https://doi.org/10.1016/j.neucom.2024.127600.

Liu, L., Morrison, J., Stadler, L., Shaw, A., Vela, J.D., Christenson, D., 2025. Modeling the resilience performance of Houston's wastewater treatment plant under wet weather conditions. Journal of Sustainable Water in the Built Environment 11, 04025002. https://doi.org/10.1061/JSWBAY.SWENG-609.

Matharaarachchi, S., Domaratzki, M., Muthukumarana, S., 2024. Enhancing SMOTE for imbalanced data with abnormal minority instances. Machine Learning with Applications 18, 100597. https://doi.org/10.1016/j.mlwa.2024.100597.

McLellan, S.L., Hollis, E.J., Depas, M.M., Van Dyke, M., Harris, J., Scopel, C.O., 2007. Distribution and fate of Escherichia coli in Lake Michigan following contamination with urban stormwater and combined sewer overflows. J. Great Lake. Res. 33, 566–580. https://doi.org/10.3394/0380-1330(2007)33[566:DAFOEC]2.0.CO;2.

Meiska, I., Naf'an, S.M., Dwi IgP., Wibawa, Kallista, M., Aina, B.F., Awaludin, A., 2023. Improving k-Nearest neighbors model using SMOTE with bagging ensemble. In: 2023 IEEE Asia Pacific Conference on Wireless and Mobile (Apwimob), pp. 195–201. https://doi.org/10.1109/APWiMob59963.2023.10365600.

Nallakaruppan, M.K., Gangadevi, E., Shri, M.L., Balusamy, B., Bhattacharya, S., Selvarajan, S., 2024. Reliable water quality prediction and parametric analysis using explainable AI models. Sci. Rep. 14, 7520. https://doi.org/10.1038/s41598-024-56775-y.

Newton, R.J., Bootsma, M.J., Morrison, H.G., Sogin, M.L., McLellan, S.L., 2013. A microbial signature approach to identify fecal pollution in the waters off an urbanized Coast of Lake Michigan. Microb. Ecol. 65, 1011–1023. https://doi.org/10.1007/s00248-013-0200-9.

Nielsen, A., 2019. Practical Time Series Analysis: Prediction with Statistics and Machine Learning. O'Reilly Online Learning. URL. https://www.oreilly.com/library/view/practical-time-series/9781492041641/ (accessed 5.15.24).

Patle, A., Chouhan, D.S., 2013. SVM kernel functions for classification. In: 2013 International Conference on Advances in Technology and Engineering (ICATE). Presented at the 2013 International Conference on Advances in Technology and Engineering (ICATE 2013). IEEE, Mumbai, pp. 1–9. https://doi.org/10.1109/ICAdTE.2013.6524743.

Peters, P.E., Zitomer, D.H., 2021. Current and future approaches to wet weather flow management: a review. Water Environ. Res. 93, 1179–1193. https://doi.org/10.1002/wer.1506.

Powers, D.M.W., 2011. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. Int. J. Mach. Learn. Technol. 2, 37–63. https://doi.org/10.48550/arXiv.2010.16061.

Pristyanto, Y., Nugraha, A.F., Dahlan, A., Wirasakti, L.A., Zein, A.A., Pratama, I., 2022. Multiclass imbalanced handling using ADASYN oversampling and stacking Algorithm. In: Proceedings of the 2022 16th International Conference on Ubiquitous Information Management and Communication, IMCOM 2022. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/IMCOM53663.2022.9721632.

Qu, H., Zhang, Z., 2024. A time series data augmentation method based on SMOTE. In: Proceedings of the 36th Chinese Control and Decision Conference, CCDC 2024. Institute of Electrical and Electronics Engineers Inc., Xi'an, China, pp. 5336–5341. https://doi.org/10.1109/CCDC62350.2024.10587816.

Sakho, A., Malherbe, E., Scornet, E., 2024. Do we need rebalancing strategies? A theoretical and empirical study around SMOTE and its variants. https://doi.org/10.48550/arXiv.2402.03819. arXIV Preprint.

Salehi, A.R., Khedmati, M., 2024. A cluster-based SMOTE both-sampling (CSBBoost) ensemble algorithm for classifying imbalanced data. Sci. Rep. 14. https://doi.org/10.1038/s41598-024-55598-1.

Savas, C., Dovis, F., 2019. The impact of different kernel functions on the performance of scintillation detection based on support vector machines. Sensors (Switzerland) 19. https://doi.org/10.3390/s19235219.

Schertzinger, G., Zimmermann, S., Sures, B., 2019. Predicted sediment toxicity downstream of combined sewer overflows corresponds with effects measured in two sediment contact bioassays. Environ. Pollut. 248, 782–791. https://doi.org/10.1016/j.envpol.2019.02.079.

Sharififar, A., Sarmadian, F., 2023. Coping with imbalanced data problem in digital mapping of soil classes. Eur. J. Soil Sci. 74. https://doi.org/10.1111/ejss.13368.

Shin, J., Yoon, S., Kim, Y., Kim, T., Go, B., Cha, Y., 2021. Effects of class imbalance on resampling and ensemble learning for improved prediction of Cyanobacteria blooms. Ecol. Inform. 61, 101202. https://doi.org/10.1016/j.ecoinf.2020.101202.

Sibanda, T., Selvarajan, R., Tekere, M., 2015. Urban effluent discharges as causes of public and environmental health concerns in south Africa's aquatic milieu. Environ. Sci. Pollut. Control Ser. 22, 18301–18317. https://doi.org/10.1007/s11356-015-5416-4.

Sreejith, S., Khanna Nehemiah, H., Kannan, A., 2020. Clinical data classification using an enhanced SMOTE and chaotic evolutionary feature selection. Comput. Biol. Med. 126. https://doi.org/10.1016/j.compbiomed.2020.103991.

Stanosheck, J.A., Castell-Perez, M.E., Moreira, R.G., King, M.D., Castillo, A., 2024. Oversampling methods for machine learning model data training to improve model capabilities to predict the presence of Escherichia coli MG1655 in spinach wash water. J. Food Sci. 89, 150–173. https://doi.org/10.1111/1750-3841.16850.

Usman-Hamza, F.E., Balogun, A.O., Nasiru, S.K., Capretz, L.F., Mojeed, H.A., Salihu, S.A., Akintola, A.G., Mabayoje, M.A., Awotunde, J.B., 2024. Empirical analysis of tree-based classification models for customer churn prediction. Sci. Afr. 23, e02054. https://doi.org/10.1016/j.sciaf.2023.e02054.

Vogel, R.M., Fennessey, N.M., 1994. Flow-duration curves. I: new interpretation and confidence intervals. J. Water Resour. Plann. Manag. 120, 485–504. https://doi.org/10.1061/(ASCE)0733-9496(1994)120:4(485).

Xu, T., Coco, G., Neale, M., 2020. A predictive model of recreational water quality based on adaptive synthetic sampling algorithms and machine learning. Water Res. 177. https://doi.org/10.1016/j.watres.2020.115788.

Yang, L., Shami, A., 2020. On hyperparameter optimization of machine learning algorithms: theory and practice. Neurocomputing 415, 295–316. https://doi.org/10.1016/j.neucom.2020.07.061.

Zhao, P., Luo, C., Qiao, B., Wang, L., Rajmohan, S., Lin, Q., Zhang, D., 2022. T-SMOTE: temporal-oriented synthetic minority oversampling technique for imbalanced time series classification. In: Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence. Vienna, Austria, pp. 2406–2412. https://doi.org/10.24963/ijcai.2022/334.

Zhou, P., Li, Z., Snowling, S., Baetz, B.W., Na, D., Boyd, G., 2019. A random forest model for inflow prediction at wastewater treatment plants. Stoch. Environ. Res. Risk Assess. 33, 1781–1792. https://doi.org/10.1007/s00477-019-01732-9.

Zhou, P., Li, Z., Snowling, S., Goel, R., Zhang, Q., 2022. Multi-step ahead prediction of hourly influent characteristics for wastewater treatment plants: a case study from North America. Environ. Monit. Assess. 194, 389. https://doi.org/10.1007/s10661-022-09957-y.