

<https://doi.org/10.1038/s41545-025-00461-7>

Viral particle prediction in wastewater treatment plants using nonlinear lifelong learning models



Jianxu Chen¹, Ibrahima N'Doye^{1,2}✉, Yevhen Myshkevych^{1,3}, Fahad Aljehani², Mohammad Khalil Monjed⁴, Taous-Meriem Laleg-Kirati⁵ & Pei-Ying Hong^{1,3}

Predicting new unseen data using only wastewater process inputs remains an open challenge. This paper proposes lifelong learning approaches that integrate long short-term memory (LSTM), gated recurrent unit (GRU) and tree-based machine learning models with knowledge-based dictionaries for real-time viral prediction across various wastewater treatment plants (WWTPs) in Saudi Arabia. Limited data prompted the use of a Wasserstein generative adversarial network to generate synthetic data from physicochemical parameters (e.g., pH, chemical oxygen demand, total dissolved solids, total suspended solids, turbidity, conductivity, NO₂-N, NO₃-N, NH₄-N), virometry, and PCR-based methods. The input features and predictors are combined into a coupled dictionary learning framework, enabling knowledge transfer for new WWTP batches. We tested the framework for predicting total virus, adenovirus, and pepper mild mottle virus from WWTP stages, including conventional activated sludge, sand filter, and ultrafiltration effluents. The LSTM and GRU models adapted well to new data, maintaining robust performance. Tests on total viral prediction across four municipal WWTPs in Saudi Arabia showed the lifelong learning model's value for adaptive viral particle prediction and performance enhancement.

Reclaimed wastewater is becoming increasingly important to circumvent water scarcity¹, particularly in regions with constrained access to natural water sources. As a climate-resilient resource, treated wastewater can help reduce dependency on traditional water sources. Accurately predicting and assessing the removal efficiency of viral particles in municipal wastewater treatment plants is necessary for effective health risk management, the prevention of food contamination, and the maintenance of the ecological integrity of water resource recovery processes. Strict water quality monitoring is necessary to guarantee its safety when reusing it. However, viral particles are very small, which makes them more difficult to remove than other biological contaminants in conventional wastewater treatment, and they must be monitored to ensure the effectiveness of the treatment and the safety of the reclaimed water. Methods used in wastewater-based epidemiology are tedious, highly specific, and time consuming^{2,3}, underscoring the urgent need for more advanced and efficient monitoring approaches. These methods also do not reflect the measurement of the wastewater

process variables in real-time streaming, as the water samples are collected on-site and then measured in off-site laboratory analysis^{4,5}. Soft-sensor algorithms based on accessible physicochemical water quality measurements must be developed to accurately predict and monitor viral particles and to support plant operators in water resource recovery facilities (WRRFs) in alleviating these issues.

Efforts have been made to develop model-based and learning-based methods to describe the physicochemical and biological interactions between water-quality parameters and bacterial cells or viral particle concentrations, and to assess these concentrations^{5–11}. However, developing model-based estimation approaches presents several challenges, including difficulties related to accurate system identification and process model representation. To address these issues, data-driven approaches can bypass traditional analytical and model-based design techniques by learning input–output relationships and capturing dominant patterns in wastewater treatment plants (WWTPs) (see, e.g.^{5,8–10,12–15}, and references therein).

¹Environmental Science and Engineering Program, Biological and Environmental Science and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal, 23955-6900, Saudi Arabia. ²Electrical and Engineering Program, Computer, Electrical and Mathematical Sciences and Engineering Division (CEMSE), King Abdullah University of Science and Technology (KAUST), Thuwal, 23955-6900, Saudi Arabia. ³Center of Excellence on Smart Health, King Abdullah University of Science and Technology, Thuwal, 23955-6900, Saudi Arabia. ⁴Biology Department, Faculty of Science, Umm Al-Qura University, Makkah, Saudi Arabia. ⁵National Institute for Research in Digital Science and Technology (INRIA), Paris-Saclay, France. ✉e-mail: ibrahima.ndoye@kaust.edu.sa

Machine learning (ML) algorithms have found extensive application in diverse domains and contexts^{16–21}. These ML algorithms, including linear regression, ensemble learning, decision trees, and neural networks, have proven particularly promising for processing input–output relationships from data and performing predictions. In the context of WWTPs, some studies have specifically focused on the quantification and estimation of bacterial concentrations and viral particle estimation using data-driven methods. For instance, the authors in refs. 5,10 proposed ML models based on ensemble learning, decision trees, and neural network algorithms to effectively estimate bacterial concentrations. The authors in ref. 11 proposed data-driven modeling based on a linear regression model, artificial neural network, and random forest to estimate two indigenous viruses—PMMoV and Norovirus GII—in a pilot-scale municipal anaerobic membrane wastewater bioreactor to verify their corresponding log reduction value (LRV). These data-driven methods have been demonstrated to improve water quality monitoring and improve the estimation performance of micro-organism concentrations from influent and effluent wastewater processes.

In previous studies, model development has often been limited by an isolated learning framework that relies on training, testing, and validating models using only historical data while intrinsically assuming a strong correlation between input and output datasets. This assumption often breaks down when applied to new datasets that were not used during model development. Moreover, isolated learning strategies are vulnerable to process drifts caused by distribution shifts and differing time scales of unseen datasets in various WWTPs. Factors such as environmental variability, uncertainty in sensing (e.g., sensor drift and faulty sensors in different treatment processes), variability in water quality input parameters, and equipment degradation introduce significant uncertainties and distribution shifts. Model generalization must be enhanced during deployment to ensure consistent and reliable performance.

Model generalization plays a critical role in streaming process prediction and the estimation of machine learning algorithms. It enables the mitigation of distribution shifts and process drifts in diverse systems and ensures reliable forecasting of previously unseen datasets. In wastewater process modeling, the development of soft-sensing models that learn local and global time series remains an open problem²¹. Recently²², introduced a calibration framework employing an out-of-distribution (OOD) generalization approach to address unbalanced distribution problems across different WWTPs by enhancing the data's statistical distributional shift. This approach effectively handles unseen datasets not encountered during model development, thereby enhancing the model's generalizability across different WWTPs. However, this OOD method relies on confidence intervals of a high probability measure of the data that need to be optimized to ensure that the retraining phase occurs when new incoming samples contain new distribution information.

An alternative approach is streaming process prediction through an incremental learning of predictive models on a batch-by-batch basis, leveraging knowledge base adaptation. The lifelong learning framework (see Text 2.2, Supporting Information) is a procedure that combines predictive models and knowledge base adaptation^{23–25}. The efficient lifelong learning algorithm (ELLA) and streaming modeling are well-known and effective traditional lifelong learning methods that have been proposed for regression, classification, and decision-making problems involving online modeling and control adaptation tasks (see^{26–32} and references therein). These lifelong learning methods based on local and global knowledge transfer tasks performed well, assuming that input and target output data are known. However, the output measurement information is generally hard for most industrial process applications to measure in real time. For instance, the viral particle counts in wastewater treatment plants (WWTPs) are often measured offline via laboratory analysis and take a few hours. Streaming modeling can adapt well in tracking the changes in system dynamics present in most process dynamic characteristics, including wastewater matrices in WWTPs, by updating the predictive model state and parameters compared to ELLA^{33–36}. However, the streaming prediction performance fails when the output is unknown and significantly degrades when the measurement

output is delayed due to its inherent local nature and nonadaptive design²⁵. The authors in ref. 25 proposed an effective lifelong learning based on linear regression models. The core idea of this framework lies in integrating input features in a dictionary learning that are updated over time, thereby achieving the unsupervised knowledge transfer prediction of new data batches without immediate output data²⁵. This method allows new batches to be predicted quickly using only process inputs before the corresponding output measurements arrive, making it suitable for most industrial process modeling and prediction characterized by time-varying dynamics and feedback delays²⁴. However, the linear regression models might not fully capture the inherent complex dependencies and nonlinear dynamic changes such as transients and process shifts in WWTPs. Therefore, developing an accurate and online adaptive predictive model that learns only new input batches by knowledge transfer before the actual measurement output is received, thereby providing a long-term time dependency while capturing underlying nonlinear patterns, is essential for viral particle prediction in WWTPs.

The main objective of this work is to enhance the model generalization capabilities of the isolated learning models for new unseen WWTP datasets that might encounter distribution shifts and process drifts in different treatment processes using the lifelong learning framework. First, we extended the existing linear lifelong learning to nonlinear lifelong learning by integrating long short-term memory (LSTM, see Text 2.1.4, Supporting Information), gated recurrent unit (GRU, see Text 2.1.5, Supporting Information) and tree-based ML algorithms as predictors and base models. This extension utilized the model adaptation mechanism by tailoring the input features of the new unseen datasets to the LSTM and tree-based knowledge base models to estimate the concentration with their corresponding predictor models. Second, we conducted a comprehensive comparison of four ML models—partial least square (PLS, see Text 2.1.1, Supporting Information), extreme gradient boosting (XGB, see Text 2.1.2, Supporting Information), category boosting (CatBoost, see Text 2.1.3, Supporting Information), GRU, and LSTM—to accurately predict total virus (TV), adenovirus (AdV), and pepper mild mottle virus (PMMoV) concentrations across two distinct wastewater matrices. Third, we validated the model adaptation performance based on LSTM for predicting TV, AdV, and PMMoV across different wastewater matrices, including conventional activated sludge (CAS) process effluent (aerobic), sand filter effluent (sand), and ultrafiltration effluent (MBR). Finally, we conducted model adaptation tests for predicting TV particle concentrations across three wastewater treatment processes (i.e., influent, aerobic, and sand) in four different WWTPs (Plant A, Plant H, Plant P1, and Plant P2) that were geographically located at different sites.

This work makes significant contributions to the existing literature. The proposed lifelong learning-based LSTM and GRU models demonstrated considerable adaptability across various sewage treatment plant datasets and conditions. Despite the evident disparities in feature distribution among datasets from different sewage treatment plants, the model progressively comprehended and adjusted to the data distribution of each plant through multiple iterations of adaptive learning, thereby achieving knowledge transfer and predictive optimization across different datasets. This capability allows the model to sustain strong predictive accuracy and flexibility in complex situations characterized by significant variations in process drifts and data distributions. The findings of this work address the potential to develop soft-sensors capable of handling static datasets while simultaneously self-updating in response to dynamic and evolving data environments, hence ensuring robust and efficient real-time viral particle prediction across WWTPs.

Results

Brief overview of the lifelong learning framework

The lifelong learning framework comprises two key components: a local machine learning (ML) predictor module and a knowledge-based adaptation module incorporating a coupled dictionary learning representation (Fig. 1). The limited available real data will be expanded using the WGAN

data generator (Fig. 1A) and fed to the local predictor. The local ML predictor module is constructed with the historical water quality input and virus particle output batches (Fig. 1B). This module is used to encode the water quality inputs as feature vectors. The local ML predictor used historical input-output batches to build a predictive model. Genetic algorithm and Lasso were used to find the best feature combination. However, the optimal feature combination that contains all the input features was retained and selected for the model adaptation phase due to its advantage of streamlining the viral particle prediction from the influent to different effluent wastewater matrices and WWTPs. The knowledge transfer scheme for local predictor reconstruction occurs when new batches arrive, and only process inputs are provided²⁵. The knowledge-based adaptation module operates in batches and adjusts the feature representation of the input data through coupled dictionary learning, thereby indirectly affecting the prediction ability and learning process of the local predictor. The predictor parameters and input feature vectors are decomposed into a sparse combination of a shared dictionary to facilitate knowledge transfer (Fig. 1B). When a new batch arrives, the local predictor is trained again only on the process input data, and the loss index is evaluated. The knowledge transfer scheme for local predictor reconstruction occurs when the above is completed. The model is dynamically adjusted according to the batch by continuously updating the sparse coefficients and the corresponding matrix to complete the adaptive adjustment and dictionary update. The dictionary is used to represent the feature space associated with each task, and to provide shared knowledge and mapping between tasks. Each column of its matrix represents a feature combination of a specific task. When it is updated, it changes the representation of the input data in the feature space, thereby affecting the way the local prediction model interprets the data in the

subsequent adaptation process. The input feature vector and prediction parameters are transformed through a sparse matrix to generate a new representation for sparse linear combinations. The local predictor model is used to calculate a sparse coefficient vector, which represents the importance and weight of the input features or prediction parameters in the sparse space.

The input features and prediction parameters are represented as the product of a sparse matrix and sparse coefficients to achieve sparse linear combination decomposition. After a few batches have been learned in a supervised manner using the full water quality input and virus concentration datasets, we then construct the knowledge transfer scheme (Fig. 1B) for newly arriving batches, providing only the water quality input data that relies on a shared dictionary learning scheme that is updated over time (Fig. 1C).

In the first model adaptation case, the influent and CAS effluent (aerobic) were considered for the performance comparison of the lifelong learning models for predicting various viral particle concentrations. Here, the MODON influent is utilized for the model development and the aerobic dataset was used as new unseen testing set data to evaluate the online prediction (i.e., model adaptation) performance of the proposed lifelong learning models. In the second case, we generated additional sand and ultrafiltration effluent datasets generated by the WGAN through their actual datasets to develop the model adaptation models across the wastewater matrices (aerobic, sand, and MBR) and test their performance results. In the third case, the lifelong learning models were developed via the baseline MODON AEMBR based on various wastewater matrices (influent, aerobic, sand) to perform model adaptation on TV concentrations across different WWTPs (Plant A, Plant H, Plant P1, Plant P2).

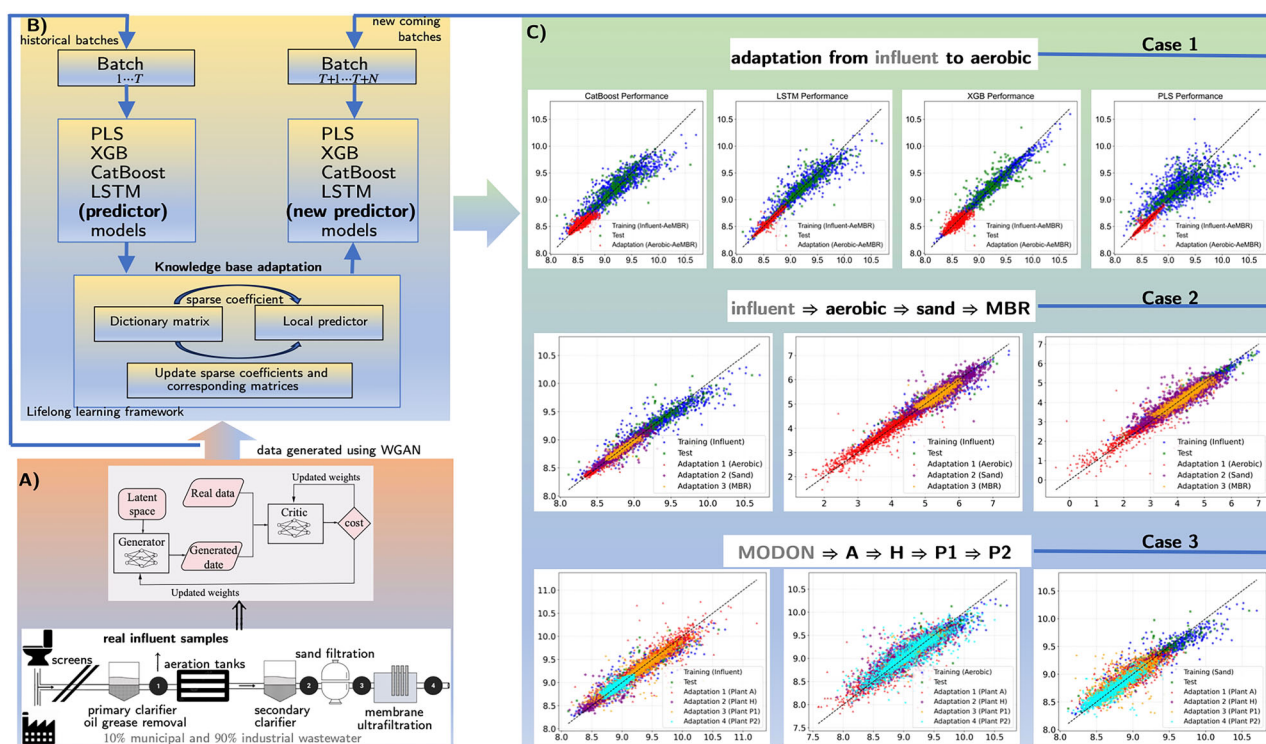


Fig. 1 | Flowchart detailing the lifelong learning strategy for online prediction of total virus particles and specific viral subpopulation concentrations across different wastewater matrices and WWTPs. The flowchart includes three components: **A** the data generation of the water quality and virus particle concentration of the influent treatment process using the WGAN, **B** the streaming process that combines the local predictors, the knowledge base models, and the reconstructed local predictors, and **C** the unseen testing datasets and the model adaptation results with the baseline influent treatment process for the three case studies: case 1

illustrates the comparative results of the lifelong model adaptation performance from the baseline MODON influent to the MODON aerobic treatment process using the four local ML predictor algorithms; case 2 shows the model adaptation performance results for predicting a) TV, b) AdV and c) PMMoV viral particles from the MODON AEMBR influent across different MODON wastewater effluent matrices; and case 3 illustrates the model adaptation performance results for predicting TV viral particles from the MODON AEMBR treatment processes across different WWTPs.

Table 1 | Performance comparison of different lifelong local predictors for predicting total virus and specific viral genera across the aerobic treatment process

Model	Set	Total Virus				AdV				PMMoV			
		R ²	RMSE	MAE	SMAPE	R ²	RMSE	MAE	SMAPE	R ²	RMSE	MAE	SMAPE
PLS	Train	0.514	0.263	0.198	2.139	0.814	0.304	0.226	2.092	0.754	0.426	0.322	2.282
	Test	0.602	0.231	0.171	1.854	0.841	0.276	0.202	1.876	0.770	0.436	0.336	1.790
	Adaptation (aerobic)	0.970	0.017	0.012	0.203	0.910	0.235	0.181	0.259	0.938	0.201	0.149	0.255
XGB	Train	0.971	0.064	0.044	0.479	0.989	0.073	0.052	0.805	0.988	0.093	0.066	0.509
	Test	0.726	0.191	0.132	1.422	0.864	0.255	0.187	1.765	0.851	0.351	0.252	1.123
	Adaptation (aerobic)	0.629	0.065	0.050	0.613	0.659	0.470	0.358	0.624	0.673	0.551	0.408	0.613
CatBoost	Train	0.820	0.160	0.115	1.237	0.920	0.199	0.147	1.140	0.918	0.246	0.186	1.265
	Test	0.723	0.192	0.136	1.464	0.855	0.347	0.261	1.368	0.853	0.347	0.261	1.125
	Adaptation (aerobic)	0.719	0.057	0.042	0.523	0.741	0.412	0.313	0.572	0.732	0.505	0.362	0.563
GRU	Train	0.863	0.140	0.093	0.973	0.925	0.193	0.129	0.938	0.944	0.203	0.138	0.941
	Test	0.832	0.150	0.096	1.002	0.929	0.185	0.127	0.934	0.943	0.217	0.151	1.047
	Adaptation (Aerobic)	0.974	0.016	0.012	0.214	0.960	0.148	0.108	0.207	0.948	0.184	0.132	0.307
LSTM	Train	0.850	0.146	0.094	0.991	0.925	0.193	0.129	0.958	0.944	0.203	0.138	1.001
	Test	0.831	0.150	0.095	1.028	0.929	0.185	0.127	0.896	0.943	0.217	0.151	0.947
	Adaptation (aerobic)	0.973	0.015	0.011	0.208	0.960	0.148	0.108	0.207	0.947	0.184	0.131	0.285

The best model adaptation performance results for predicting total virus and specific viral genera (i.e., AdV and PMMoV) across the aerobic treatment process, with a higher coefficient of determination R^2 value and a lower (RMSE, MAE, and SMAPE) value between the local predictors (PLS, XGB, CatBoost, GRU, and LSTM), are highlighted in bold.

Performance comparison of the lifelong local predictor models for predicting various viral particle concentrations

We comprehensively compared the lifelong local predictor models—linear regression (PLS), tree-based machine learning models (eXtreme Gradient Boosting (XGB) and Category Boosting (CatBoost)), GRU and LSTM models—for predicting TV, AdV, and PMMoV particle concentrations across two different wastewater matrices. The tree-based ML, GRU and LSTM models capture nonlinear relationships and interaction effects in the data, demonstrating their advantages in modeling complex relationships. Overall, the training, testing, and model adaptation results of each ML model, along with TV, AdV, and PMMoV particle concentrations, are presented in Table 1. These performance results are further supported by the predicted values versus true values figures (see Figs. S1, S3, and S5) and the evaluation of the training, testing, and model adaptation performances (see Figs. S2, S4, and S6) of each ML model and each viral community. The results showed that the tree-based ML models XGB and CatBoost and GRU LSTM models exhibited superior testing performance and outperformed the PLS model (XGB and CatBoost $R^2 = 0.72$, GRU and LSTM $R^2 = 0.83$) (see Table 1). However, the model adaptation capability of the tree-based models (XGB and CatBoost) did not extend to the knowledge-based adaptation phase or online prediction across different virus particle concentrations. The resulting poor performance in the model adaptation phase of these tree-based ML models was mainly due to the distribution shift and multiple time scales observed in the CAS effluent process.

The PLS, GRU and LSTM models displayed significantly better model adaptation performance results across (TV, AdV, and PMMoV) viral particle concentrations in the streaming lifelong learning framework with an average coefficient of determination of ($R^2 = 0.65$ for XGB, $R^2 = 0.73$ for CatBoost, $R^2 = 0.94$ for PLS, and $R^2 = 0.96$ for GRU and LSTM) (Table 1). The root mean square error (RMSE), mean absolute error (MAE), and symmetric mean absolute percentage error (SMAPE) values shown in Table 1 also confirm the effectiveness of the model adaptation approach. LSTM and GRU models exhibited the highest performance across the different virus particle communities (Table 1). Despite its lower performance in the testing set, PLS exhibited significant flexibility in adapting to data distribution shifts induced by the unseen testing test data. LSTM and GRU showed consistently very similar performance results across all the viral particles in the model development and model adaptation stages. Notably,

LSTM and GRU were inherited from the multiple time scales and nonlinear dependencies of the water quality input and virus concentration output, and their predictions outperformed the other three models. This implies that supervised learning techniques based on LSTM and GRU predictors can potentially improve the streaming lifelong learning framework for predicting viral particle communities in WWTPs. LSTM and GRU are powerful nonlinear base learners in streaming lifelong learning frameworks. The next section describes the experiments conducted on several unseen test datasets using the LSTM model as a local predictor in the knowledge-based adaptation analysis.

Model adaptation performance for predicting total virus and specific viral genera across different wastewater effluent matrices

We evaluated the model adaptation performance based on the LSTM local predictor for predicting TV, AdV, and PMMoV across different wastewater matrices, such as conventional activated sludge effluent (aerobic), sand filter effluent (sand), and ultrafiltration effluent (MBR) (see Text 1.1 and Fig. 1.1.1, Supporting Information). The MODON AeMBR influent was used as a baseline model for the model adaptation framework. The wastewater effluent matrices (aerobic, sand, and MBR) were considered unseen testing test data. The online prediction (i.e., model adaptation) of (TV, AdV, and PMMoV) concentrations significantly improved across all three wastewater matrices. The prediction results across the aerobic treatment process showed R^2 values of 0.95 for TV, 0.94 for AdV, and 0.94 for PMMoV (Table 2). Similarly, we obtained R^2 values of 0.92 for TV, 0.86 for AdV, and 0.87 for PMMoV for the model adaptation across the sand treatment process (Table 2). The root mean square error (RMSE), mean square error (MSE), and mean absolute error (MAE) values shown in Table 2 also confirm the effectiveness of the model adaptation approach. Moreover, the online prediction across the MBR treatment process resulted in R^2 values of 0.97 for TV, 0.91 for AdV, and 0.94 for PMMoV. A strong prediction performance of the model during the training, testing, and adaptation phases of the TV concentrations was observed across the different wastewater matrices (Fig. 2). Figure 3a, b, and c shows the prediction performance results of the TV, AdV and PMMoV, respectively, across the different wastewater treatment matrices (see Figs. S7, S8, and S9 for TV concentrations; Figs. S10, S11, and S12 for AdV concentrations; Figs. S13, S14, and S15 for PMMoV concentrations).

Table 2 | Model adaptation performance results for predicting online TV, AdV, and PMMoV concentrations across different wastewater matrices

Model	Set	Total Virus				AdV				PMMoV			
		R ²	RMSE	MAE	MSE	R ²	RMSE	MAE	MSE	R ²	RMSE	MAE	MSE
LSTM	Train	0.903	0.118	0.074	0.014	0.980	0.100	0.070	0.010	0.955	0.182	0.126	0.033
	Test	0.861	0.136	0.081	0.019	0.937	0.174	0.110	0.030	0.942	0.219	0.152	0.048
	Adaptation 1 (aerobic)	0.955	0.020	0.015	0.001	0.940	0.186	0.138	0.038	0.942	0.213	0.155	0.050
	Adaptation 2 (sand)	0.930	0.039	0.028	0.002	0.859	0.164	0.118	0.028	0.867	0.247	0.183	0.065
	Adaptation 3 (MBR)	0.965	0.016	0.012	0.0008	0.907	0.080	0.059	0.007	0.935	0.111	0.082	0.014

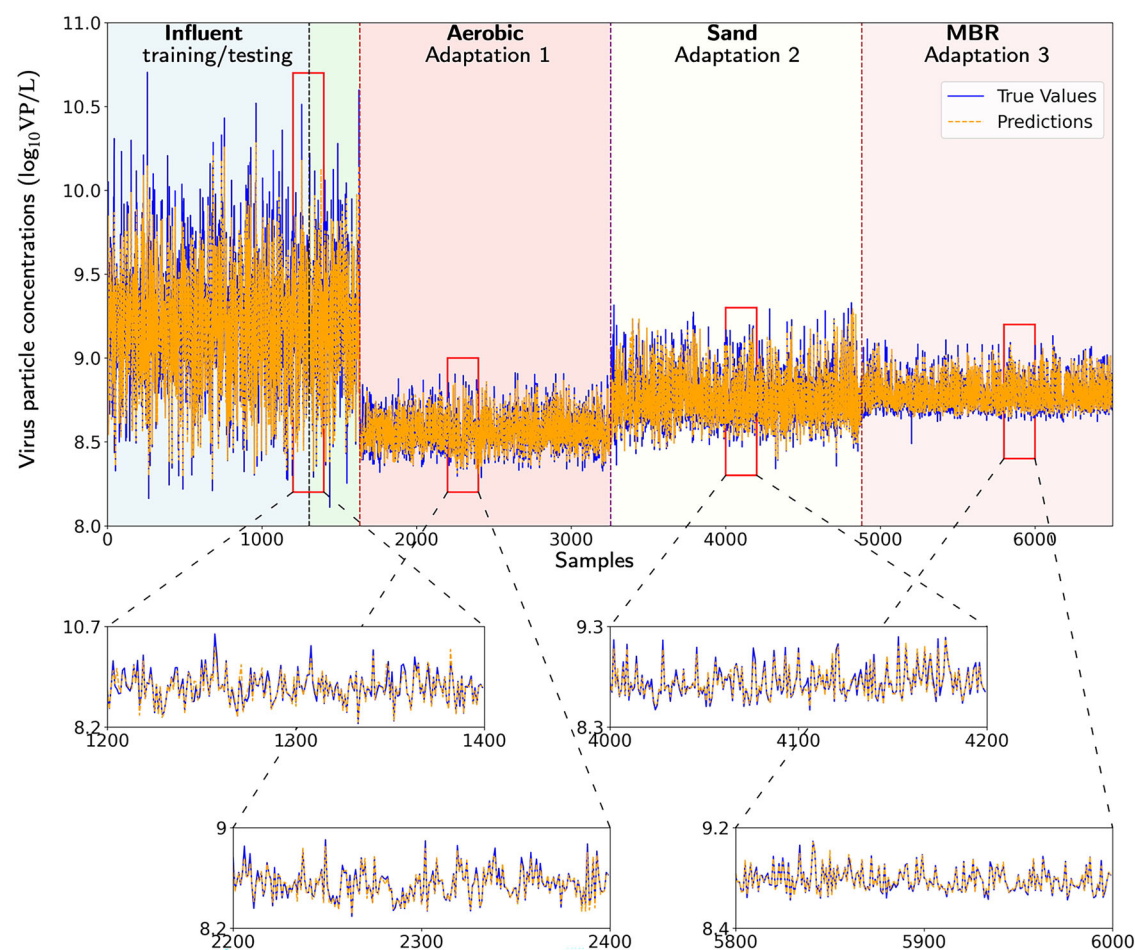


Fig. 2 | High accuracy of the lifelong learning based LSTM model in viral particle prediction across various wastewater matrices. Graph illustrating the online prediction of the total virus particle concentrations from the influent treatment process across the aerobic, sand and MBR treatment processes. The influent treatment process is used for the model development (i.e., training and testing) and the different wastewater matrices are used to perform model performance enhancement on unseen test datasets (i.e., model adaptation).

Overall, the prediction results showed excellent performance in all tests, from the initial baseline training stage to the subsequent baseline testing model and model adaptation across the different wastewater stages. Additionally, the proposed lifelong learning model maintained a high degree of prediction accuracy for different viral-type particle concentrations and across different wastewater matrices.

Model adaptation performance for predicting total virus particle concentrations across different WWTPs

We conducted model adaptation tests to predict total virus (TV) particle concentrations from three wastewater treatment processes (influent, aerobic, and sand) and across four different WWTPs (Plant A, Plant H, Plant P1, and Plant P2) (see Text 1.2 and Fig. 1.2.1, Supporting Information). These

WWTPs were located in Makkah and Madinah and were designed to treat municipal wastewater using a process similar to the MODON AeMBR. As outlined in previous sections, the model was initially trained and tested on the baseline AeMBR in the MODON WWTP datasets, maintaining the same data partitioning strategy. Following this, the developed model underwent a sequential adaptation process using the four distinct WWTP datasets with respective treatment processes, each comprising approximately 1800 samples. Our results showed a strong correlation between the predicted and actual values of the TV concentrations across the four wastewater plants for each treatment process (influent, aerobic, and sand) (Fig. 4). In addition, a strong correlation between the predicted and actual values of the TV concentrations across the different WWTPs was observed in all three treatment processes (Fig. 5).

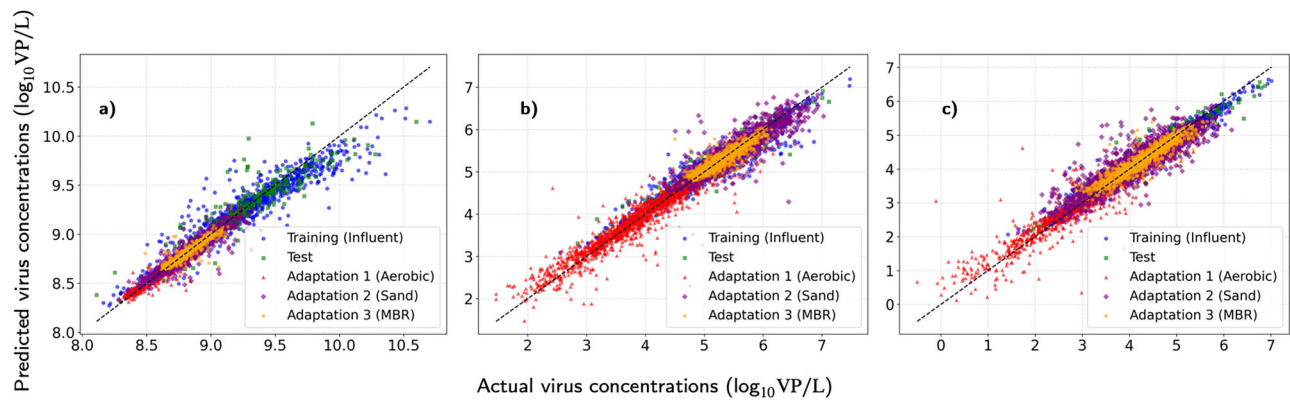


Fig. 3 | Predicted versus actual viral particle concentrations across various wastewater matrices. Scatterplots illustrating the predicted versus actual values of the (a) TV, (b) AdV, and (c) PMMoV particle concentrations across different wastewater matrices using the lifelong learning based LSTM model. The blue and

green colors represent the training and testing of the baseline MODON influent treatment process, respectively. The red, purple and yellow colors represent the model adaptation performance of the unseen aerobic, sand and MBR test samples, respectively.

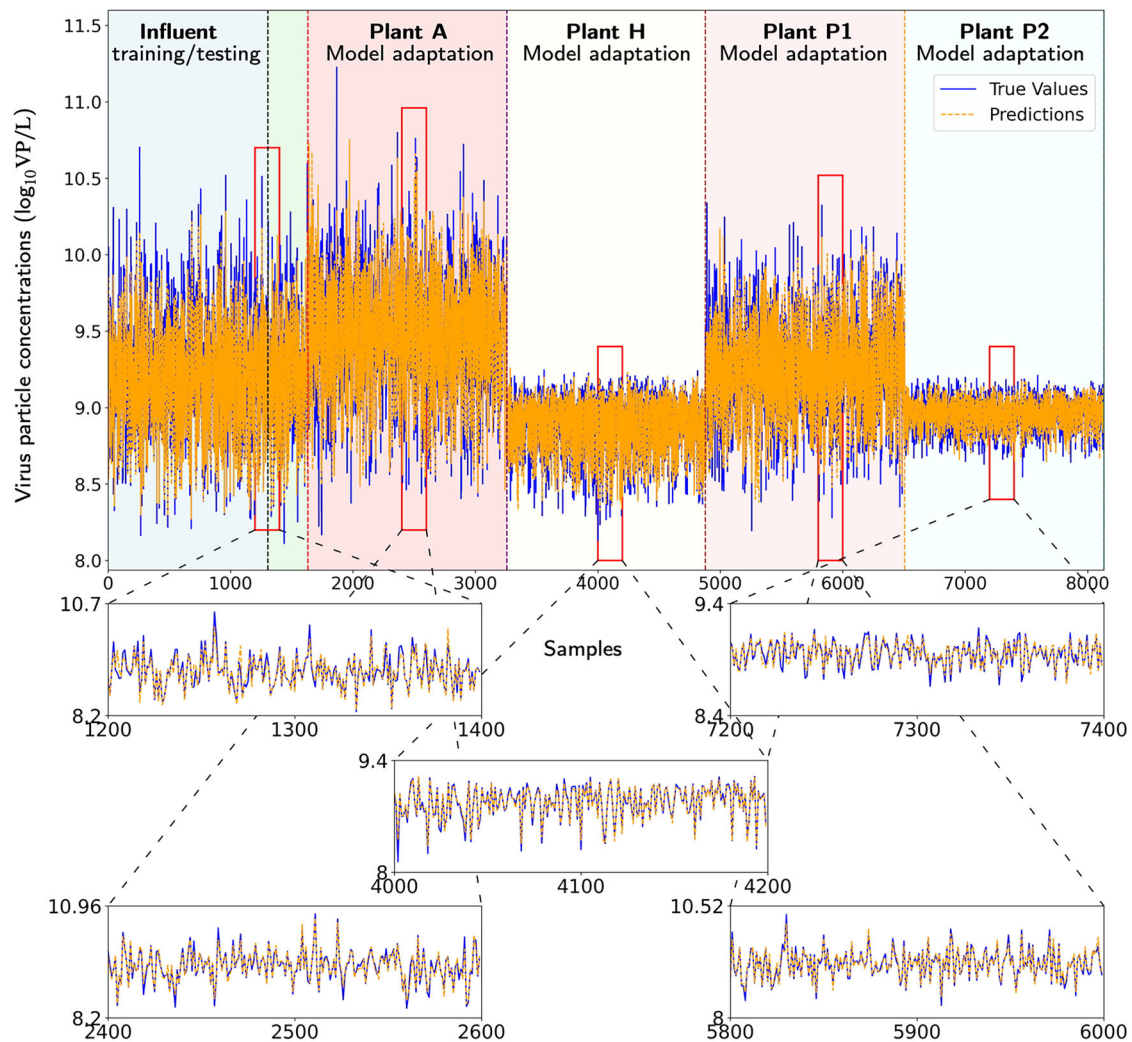


Fig. 4 | High accuracy of the lifelong learning based LSTM model in predicting total virus (TV) particle concentrations across different WWTPs. Model adaptation performance for predicting TV concentrations from the MODON influent

treatment process across various influent treatment processes of different WWTPs (Plant A, Plant H, Plant P1, Plant P2).

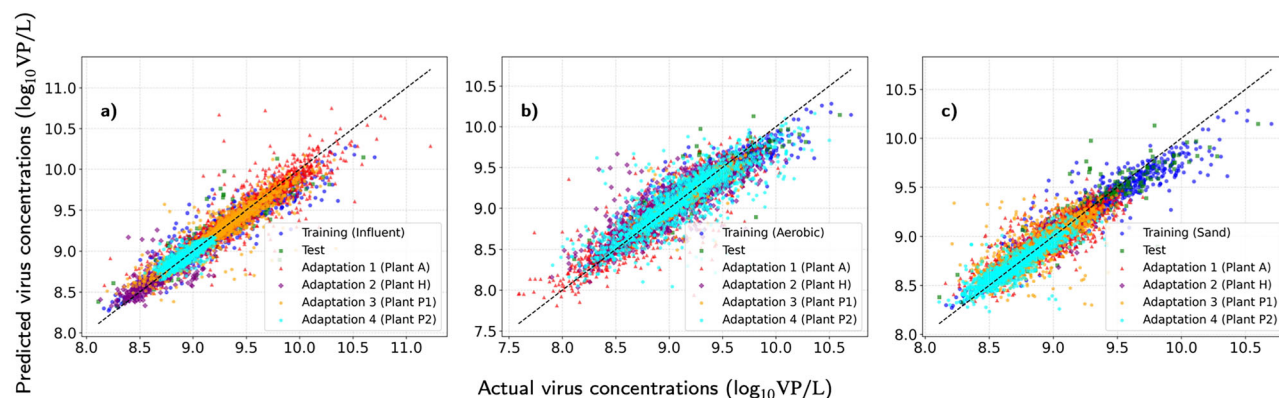


Fig. 5 | Predicted versus actual total virus (TV) particle concentrations across different WWTPs. Scatterplots illustrating the predicted versus actual values of the TV particle concentrations in the (a) influent, (b) aerobic, and (c) sand treatment processes across different WWTPs. The blue and green colors represent the training

and testing of the baseline treatment process, respectively. The red, purple, yellow, and cyan colors represent the model adaptation performance of the unseen test data for Plant A, Plant H, Plant P1, and Plant P2, respectively.

Table 3 | Model adaptation results for predicting TV concentrations from the three treatment processes (influent, aerobic, sand) and across different WWTPs (Plant A, Plant H, Plant P1, Plant P2)

Model	Set	Influent				Aerobic				Sand			
		R ²	RMSE	MAE	MSE	R ²	RMSE	MAE	MSE	R ²	RMSE	MAE	MSE
LSTM	Train	0.903	0.118	0.074	0.014	0.903	0.074	0.049	0.075	0.912	0.107	0.075	0.012
	Test	0.861	0.136	0.081	0.019	0.835	0.143	0.092	0.020	0.874	0.061	0.044	0.004
	Adaptation 1 (A)	0.911	0.109	0.077	0.013	0.882	0.131	0.094	0.018	0.903	0.106	0.077	0.013
	Adaptation 2 (H)	0.956	0.036	0.027	0.002	0.840	0.119	0.091	0.015	0.956	0.021	0.014	0.001
	Adaptation 3 (P1)	0.944	0.071	0.054	0.006	0.902	0.059	0.041	0.004	0.886	0.094	0.070	0.009
	Adaptation 4 (P2)	0.852	0.038	0.029	0.002	0.837	0.153	0.108	0.025	0.874	0.061	0.044	0.004

The mean R² value of the model adaptation-based LSTM predictor for the four WWTPs was 0.915 for the influent treatment process, 0.865 for the aerobic treatment process effluent, and 0.904 for the sand filter treatment process effluent (Table 3). The results and advantages in the RMSE, MAE, and MSE of the model adaptation approach compared to the testing results of the MODON AeMBR baseline also confirmed that the model adaptation performed well and improved overall model performance (Table 3).

The model adaptation performance results for predicting TV concentrations from the three treatment processes of the MODON AeMBR-based WWTP across different WWTPs (Plant A, Plant H, Plant P1, Plant P2) are also included in the supplementary information (see Figs. S16, S17, and S18 for the influent treatment process; Figs. S19, S20, and S21 for the aerobic treatment process; and Figs. S22, S23 and S24 for the sand).

Overall, the model adaptation performance for predicting the total virus particle concentrations from the baseline wastewater treatment processes (influent, aerobic, and sand) in the MODON AeMBR WWTP across the four WWTPs indicated the utility of the lifelong learning-based LSTM approach for multiple time scales and distribution shifts in the wastewater datasets. Since the derived model transits between treatment processes, including datasets in the model adaptation phase, a slight variation typically occurred between the dataset transition phases. This shift was mainly due to the different underlying treatment processes of the effluents. However, after a few batch cycles, due to the robustness feature of the model adaptation framework, the knowledge-based adaptation self-corrected and adjusted the learning strategy to achieve better performance.

Discussion

State-of-the-art methods for viral particle estimation are time-consuming and tedious and hamper the decision-making process, thus decreasing the efficiency of wastewater-based epidemiology (WBE). The current machine-

learning methods of estimating viral particles are limited to developing optimal models based on training, validation, and testing, but they do not infer an adaptation mechanism for unseen testing sets, preventing their model generalization capabilities. Additionally, existing methods do not account for distribution shifts and process drifts in unseen testing datasets, underscoring their limitations in simulating real experimental conditions, which are common in real experimental scenarios. In particular, predicting viral particles in real time across varying wastewater matrices or treatment plants presents substantial challenges due to rapid shifts in treatment conditions. To address these challenges, we proposed a lifelong learning approach based on the LSTM, GRU and tree-based machine learning frameworks that incorporate nonlinear local predictors and base models to predict viral particles across different wastewater matrices and plants in real time.

The proposed approach provides solutions to real-time prediction processes by combining knowledge and local ML predictors. As the model adaptation phase progressed in predicting viral particles, particularly in later stages, its performance showed substantial improvement due to the accumulation of prior knowledge. The model becomes increasingly adept at identifying complex patterns and situations as the volume of data grows, enhancing its ability to generalize across various data distributions. Moreover, the model's prior learning significantly improves its adaptation performance for new unseen dataset, exemplifying the core advantage of lifelong learning models. Unlike traditional models, which often suffer from the phenomenon of forgetting, the lifelong learning model continues to self-update and refine in response to new data.

The proposed approach provides solutions to real-time prediction processes by combining knowledge and local ML predictors. As the model adaptation phase progressed in predicting viral particles, particularly in later stages, its performance showed substantial improvement due to the

accumulation of prior knowledge. The model becomes increasingly adept at identifying complex patterns and situations as the volume of data grows, enhancing its ability to generalize across various data distributions. Moreover, the model's prior learning significantly improves its adaptation performance for new unseen dataset, exemplifying the core advantage of lifelong learning models. Unlike traditional models, which often suffer from the phenomenon of forgetting, the lifelong learning model continues to self-update and refine in response to new data.

We conducted a comprehensive comparison of four local predictors (a linear PLS, GRU, LSTM models, and two tree-based models, XGB and CatBoost) in the lifelong learning module, with GRU and LSTM proving to be the best predictors. LSTM and GRU models were best suited to process drifts due to their inherent properties for coping with the treatment process drifts and time scales of unseen testing datasets across different wastewater matrices and viral concentrations. As illustrated in Table 1, LSTM and GRU models demonstrated superior performance to the PLS, XGB, and CatBoost models.

In the lifelong learning module, LSTM and GRU models contributed to the development of a nonlinear knowledge base model, which helped integrate new input features without output data. These base models enabled an augmented dictionary between the local predictor and input features and achieved knowledge transfer through the updated dictionary. The PLS model, as a linear predictor, maintained stable performance despite treatment process drifts and distribution shifts across all cases, contributing positively to model adaptation, as demonstrated in previous work²⁵, and shown in Table 1. The tree-based models exhibited remarkably good performance results in the model development as illustrated in Table 1; however, they were limited by insufficient knowledge-based model adaptation capabilities. Overall, LSTM and GRU models achieved remarkable performance in the viral particle prediction, with an average of R^2 value of 0.96 across all viral particles. Both approaches effectively captured the treatment process drifts and multiple time scales of the wastewater matrices and plants in these viral prediction tasks. Next, we used LSTM as a local predictor to evaluate the performance of the lifelong learning method and to keep consistency in the detailed processing of the lifelong learning methods.

The performance of the lifelong learning-based LSTM model was evaluated across different effluent treatment processes and WWTPs using the evaluation metrics presented in the Methods section. Tables 2 and 3 provide a comprehensive summary of the model's performance during the training, testing, and model adaptation stages. The three adaptation phases for predicting TV, AdV, and PMMoV from the influent treatment process across different wastewater matrices (aerobic, sand, and MBR) are illustrated in Figs. 2 and 3. The strong generalizability of the proposed framework in predicting TV concentrations across different WWTPs at various wastewater matrices is shown in Figs. 4 and 5. These results collectively reinforce the significant value of the lifelong learning-based LSTM model in diverse operational settings.

The proposed lifelong learning-based LSTM demonstrated remarkable performance in predicting viral particle concentrations across various wastewater matrices and WWTPs. Existing research has been limited to developing ML models to quantify and estimate viral particle concentrations. The authors of ref. 11 proposed estimated PMMoV and Norovirus GII viral particle concentrations based on neural networks and linear regression models. However, they needed to infer calibration or adaptation mechanisms to unseen test datasets from different WWTPs, which limited performance for model generalization purposes. Our analysis of the lifelong learning approach revealed the importance of combining an adaptation mechanism with the development of most data-driven regression modeling to overcome process drifts and ensure model generalization ability without requiring the best optimal or dominant input features of the isolated learning method. The additional layer of the knowledge-based models constitutes the key component in incorporating only new water quality input samples to adaptively and accurately predict the viral particle concentrations. The proposed knowledge transfer approach is based on a dictionary that arises from a mathematical representation compared to the

OOD generalization tasks that are often defined within good heuristic samples with the requirement of a few output samples and a condition to retrain the ML model, which might limit the model generalization task. In the model adaptation stage, we implemented the multi-head attention-based gated recurrent unit (MAGRU, see Text 2.1.6, Supporting Information) for comparison purposes. MAGRU combines a multi-head attention module with GRU module³⁷. As an attention-based mechanism with a GRU local predictor, it can balance the input segments to capture relationships and dependencies, thereby ensuring model performance enhancement. Although MAGRU effectively can perform well in the model development of time series data, it does not possess an adaptation mechanism to handle process drifts, which limits its generalization capability across various wastewater matrices and WWTPs. As we can observe, Figure S25 illustrates the poor prediction performance of the viral particles for the MODON AeMBR-based WWTP across various wastewater matrices using MAGRU. Estimating the abundance of viral particles in the wastewater distribution system systematically requires an adaptation mechanism that simultaneously adjusts the predictive model parameter and the designed adaptation law. The MAGRU framework can effectively work locally, but it might lack the model adaptation and robustness features to convey generalization capability across dynamic WWTP changes. Compared to other transfer learning techniques based on attention mechanisms such as MAGRU and calibration framework, lifelong learning as an emerging paradigm has the unique potential to stream process monitoring with delayed output. Our findings demonstrated that the proposed lifelong learning approach could streamline the rapid prediction and monitoring of viral particle concentrations across different WWTPs, facilitating better community health protection.

Methods

Sampling sites description and data collection

The pilot MODON aerobic membrane bioreactor (AeMBR)-based WWTP treats a mix of municipal and industrial wastewater which is collected from food production and light industries with high chemical oxygen demand (COD) content. Approximately 10% of influent comes from local communities and the rest 90% are supplemented by nearby enterprises and factories. Primary influent undergoes a screening process in order to remove bulky debris and meet water quality standards. Smaller particles are settled in a grit removal tank. Primary sedimentation and grease removal are subsequently applied to influent wastewater treatment. An activated sludge process with hydraulic retention time (HRT) of 72 hours is applied to remove COD, phosphate, and ammonium. Clarifier effluent undergoes sand filtration (sand) and membrane ultrafiltration (MBR) with a pore size of 0.02 μm (Fig. 1.1.1). Samples for model validation were collected from four points: influent, aerobic (wastewater after the conventional activated sludge process), sand, and MBR (Fig. 1.1.1).

The four AeMBR WWTPs (A, H, P1, P2) located in Makkah and Medinah regions were designed to treat municipal wastewater, with a process similar to that of the MODON AeMBR, though with some modifications³⁸. The HRT was set to 12 hours, compared to 72 hours in the MODON plant. Ultrafiltration and reverse osmosis (RO) steps were omitted. However, a UV treatment step was introduced between the sand filtration and chlorination stages to ensure effluent safety and to minimize chlorine consumption (Fig. 1.2.1). Samples for model validation were collected from three points: influent, aerobic (wastewater after the conventional activated sludge process), and sand (wastewater after sand filter) (Fig. 1.2.1).

Physicochemical water quality parameters such as pH, total dissolved solids (TDS), electroconductivity (EC), total suspended solids (TSS), turbidity, ammonium nitrogen $\text{NH}_4\text{-N}$, nitrate nitrogen $\text{NO}_3\text{-N}$, nitrite nitrogen $\text{NO}_2\text{-N}$, and chemical oxygen demand (COD) concentration and flow virometry and PCR-based methods (RT-qPCR) including TV, AdV, and PMMoV concentrations were appropriately measured from the MODON AeMBR and the four WWTPs (A, H, P1, P2) (Table 2.5.1, Supporting Information). Human total virus (TV) and adenovirus (AdV) were chosen as a target for enteric viral pathogens. TV concentration reflects

overall viral diversity regardless of viral genera as a predictive parameter. Pepper mild mottle virus (PMMoV) was chosen to be detected as a viral indicator. For more details related to the equipment and the collection of the initial samples, we refer the readers to³⁸ that provided a detailed analysis and processing of all the parameters involved in the source-tracking microbial in WWTPs.

Dataset preprocessing and statistical analyses

To evaluate the lifelong learning framework in the three model adaptation cases highlighted in the Results section, we first selected real samples from the MODON AeMBR-based WWTP. Then, additional datasets from the four WWTPs (A, H, P1, P2) were utilized for further model adaptation validation from the available measurements of water quality and flow virometry-PCR. All values of the viral particle concentration are converted to the logarithmic scale (i.e., $\log_{10}$ VP/L). The real samples were used to generate synthetic datasets using the Wasserstein generative adversarial network (WGAN) due to the lack of large available WWTP datasets. The WGAN was quantitatively evaluated based on the Wasserstein critic loss, which minimizes the training and validation loss within stopping criteria for generated synthetic data (Text 2.4 and Figs. 2.4.1, and 2.5.1, Supporting Information). We conducted a series of experiments to select the critic and generator architectures of the WGAN to generate synthetic data from the available measurements of water quality and flow cytometry-PCR (nine input variables and three virus particle output concentrations; both input-output variables contained approximately 8–15 real samples) (Table S1). More details on the architecture and hyperparameters of the WGAN are presented in the supplementary material (Text 2.4 and Figs. 2.4.1, and 2.5.1). Figure 2.5.1 in Supporting Information illustrates the convergence curves of the training and validation losses within the stopping criteria, demonstrating high fidelity in quantitative measurements of the similarity between generated and real data in all experiments. In all model adaptation scenarios, we generated 1800 samples for the selected baseline model and the model adaptation datasets.

Pearson's correlation metric has been performed in the combined real and synthetic data to analyze the linear dependency between variables. It is formulated as follows

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

where x_i and y_i are the samples of features x and y , and $\bar{x}$ and $\bar{y}$ are the mean values of the features x and y . Two input features are highly correlated when r is near to 1. Fig. 2.5.3 illustrates the correlation between features of the real and generated influent and effluent datasets. Pearson correlation helps identify strong linear relationships between input features and reduces necessary redundancy between input variables in the data preprocessing stage. r -value between two input variables across the real and all generative models is not greater than the specific threshold value $r = 0.99$ (Fig. 2.5.3), as highlighted in ref. 8 to eliminate features in the subsequent model development stage. These correlation results demonstrated that the proposed generative models performed well, which indicates the accuracy of these models in avoiding multicollinearity and preserving the distributions between the original and synthetic generated input features (Fig. 2.5.3). The latter is reminiscent of preserving all the input features in Table 1, which is crucial for the model development stage and specifically for achieving remarkable model adaptation performance across various wastewater treatment matrices and WWTPs.

The isolation forest (IF) algorithm was used to detect and remove outliers of the generated synthetic data in all the three proposed model adaptation cases to ensure data quality and make different features comparable. The parameters and detailed description of IF were provided in the supplementary information (Text 2.1.8). The baseline dataset was partitioned into 80% training and 20% testing sets and used to develop the baseline training and testing model in all cases. The mean and standard

deviation of the WGAN-generated water quality parameters and viral concentrations for the MODON influent and effluent (aerobic, sand, and MBR) datasets were given in the supplementary information (Table 2.5.1). As an illustrative example, Fig. 2.5.2 depicts the statistical summary of the WGAN-generated water-quality parameters and viral particle concentrations of the baseline MODON influent treatment process.

Overview of the lifelong learning framework

Lifelong learning methodologies present a compelling solution for managing continuous data streams in real-time industrial settings²⁵. In these processes, data are generally produced in the form of sequential batches, denoted by $D_1, D_2, \dots, D_N$ ^{39,40}. Traditional isolated learning frameworks, which develop models without incorporating self-adaptation mechanisms, often face significant limitations in prediction accuracy with unseen datasets and delayed process output measurements. The delayed feedback from measurements of viral particle concentrations in WWTPs is a classic example of this, and the rapid fluctuations in treatment conditions in different WWTPs further exacerbate the complexity. Thus, integrating mechanisms capable of retaining and leveraging prior knowledge is necessary to enable adaptive learning and to maintain robust model performance in the dynamic environments typical of WWTPs. This section focuses on the mathematical logic details of the adaptive framework.

Let us consider a local model f_t that is built based on batch data $\{(x_{jt}, y_{jt})\}_{j=1}^{n_t}$ for each data batch D_t , where n_t represents the number of samples in the batch, $x_{jt} \in R^d$ is the input vector and $y_{jt} \in R$ is the corresponding output value. In addition, assume that the system has processed $T-1$ complete input-output data batches and established the corresponding model set $f_1, f_2, \dots, f_{T-1}$. However, when the T -th batch of data starts to arrive, the system can only obtain the input data $\{x_{jT}\}_{j=1}^{n_T}$ immediately; therefore, the predictor should be able to accurately predict the actual process output $\{y_{jT}\}_{j=1}^{n_T}$ based only on the currently available input data x_{jT} and the previously accumulated knowledge $f_1, f_2, \dots, f_{T-1}$. Consequently, the system builds a complete model f_T only when the real output data for a new batch is accessible later and integrated into the existing knowledge base. This requires the prediction system to perform effective reasoning when the output data information is incomplete and continuously update and expand its knowledge base²⁵.

For each batch t , the lifelong learning framework constructs a local predictor defined as follows

$$f_t(x) = f(x, \theta_t), \quad (1)$$

where $\theta_t \in R^d$ is a batch-specific parameter vector. In general, the model parameter $\theta_t = Ls_t$, where $L \in R^{d \times k}$ and $s_t \in R^k$ are the sparse coefficients with $k < d$. Hence, L is a dictionary storing shared knowledge for all batches and s_t extracts essential knowledge features for a specific batch t ⁴¹. The factorization parameter θ_t enables knowledge transfer among batches. In a linear setting, one could introduce a linear model $f_t(x) = x^T \theta_t$ and an additional shared dictionary $K \in R^{d \times k}$ to the input feature vector, mapping the two spaces (i.e., L and K dictionaries) to the same sparse coefficient vectors S , and solve a least square optimization problem through a recursive updating scheme to reconstruct both the normalized input features and the local predictors²⁵.

Proposed lifelong learning-based LSTM predictor

In wastewater processes, the relationships between water quality input data and virus particle concentration output often exhibit highly nonlinear and time-dependent characteristics. The base model of the lifelong learning approach proposed in ref. 25 is linear and provides good predictions. One possible way to enhance the performance is to adopt nonlinear predictor models that can capture nonlinear relationships and long-term dependencies along with a nonlinear knowledge base model. In line with this, we implemented a streaming prediction of the virus particle concentration based on long short-term memory (LSTM) and tree-based predictors along with an LSTM knowledge base model to enhance prediction performance

while retaining the core advantages of the lifelong learning framework (see Text 2.3 and Algorithm 1, Supporting Information).

Let us define $X_t = [x_t^1; x_t^2; \dots; x_t^m] \in \mathbb{R}^{d \times 1}$ as the water quality input matrix, and $y_t = [y_t^1; y_t^2; \dots; y_t^n]^T \in \mathbb{R}^n$ as the corresponding virus particle concentrations. The first step of the streaming process prediction consists of constructing a linear local predictor based on a least square model or nonlinear local predictor based on LSTM and tree-based models from historical batches with water quality input and virus particle concentration output data (X_t, y_t) in an isolated learning manner. The second step lies in transforming the water quality input matrix X_t into the d -dimensional feature vector via an operator. The final step results in linking the encoded water quality input features with the local predictor via a coupled dictionary learning to completely exploit the water quality input and predict a new batch without the virus particle concentration output. In the following, we focus on constructing a base model with the LSTM algorithm. However, similar derivations also hold for GRU and the tree-based models such as XGB and CatBoost.

At time t , the output of the LSTM can be expressed as

$$h_t = \text{LSTM}(x_t, h_{t-1}, c_{t-1}),$$

where X_t is the input feature for batch t , h_{t-1} is the previous hidden state, and c_{t-1} is the previous cell state. For each new task or data batch, the task-specific parameter θ_t is first estimated.

The task-specific parameter at time t can be defined as

$$\theta_t = \text{LSTM}_{\text{params}}(X_t, y_t),$$

where X_t and y_t are the input and target values of the current task, respectively. To capture the local curvature of the loss surface, the Hessian matrix H_t is approximated as follows:

$$H_t = X_t^\top X_t + \frac{\beta I}{2n}, \quad (2)$$

where β is the regularization parameter and n is the number of samples. This approach is computationally efficient while retaining key curvature information, maximizing computational efficiency, adapting to the needs of online learning, and handling complexity in high-dimensional data environments.

To account for task-specific curvature, the feature space is transformed using Eq. (2) as follows:

$$\theta_{\text{transformed}} = H_t^{-\frac{1}{2}} \theta_t \text{ and } L_{\text{transformed}} = H_t^{-\frac{1}{2}} L \quad (3)$$

where L is the shared knowledge base matrix which is initially chosen as a random matrix of size $d \times k$. This transformation allows the model to operate in a task-specific geometric space, thereby improving its adaptability to new tasks. Subsequently, the sparse coefficients s_t are estimated to link the current task with the shared knowledge:

$$s_t = \text{LSTM}_{\text{predict}}(L_{\text{transformed}}) \quad (4)$$

This step enables the model to effectively exploit previously learned knowledge while maintaining specificity for new tasks.

The shared knowledge base is updated using Eqs. (2) and (4) as follows:

$$A = A + (s_t s_t^\top \otimes H_t) + \mu I, \quad (5)$$

$$b = b + [s_t \otimes (H_t \theta_t)], \quad (6)$$

where μ is the regularization parameter, $\otimes$ represents the Kronecker product, and s_t is the sparse coefficient. This allows the model to effectively

capture the interactions between tasks while maintaining computational efficiency¹².

Finally, the shared parameters L is updated using Eqs. (5) and (6) as follows:

$$\text{vec}(L) = (A^\top + \lambda I)^{-1} b^\top, \quad (7)$$

where λ is the regularization parameter. This update ensures that the model can balance new and old knowledge and avoid catastrophic forgetting. At this time, $\text{vec}(L)$ is the updated vectorized version of L after adaptation to the current batch information. We also need to reshape $\text{vec}(L)$ into a matrix form:

$$L = \text{reshape}(\text{vec}(L)). \quad (8)$$

At this time, L will be updated after batch adaptation and will be applied to the adaptation process of the next batch and entered into the loop. For a new input x_{new} , prediction is made through the LSTM model using the updated parameters:

$$y_{\text{new}} = \text{LSTM}_{\text{reconstruct}}(x_{\text{new}}, L). \quad (9)$$

Figure 6 illustrates the flowchart describing in detail the online prediction strategy for different batches of data, aiming to optimize the parameter adjustment and knowledge adaptation process of the local prediction model. The proposed nonlinear lifelong learning framework has significant advantages over linear models. The LSTM architecture allows for the modeling of complex nonlinear relationships, while the streaming process prediction mechanism enables the model to continuously update its knowledge base to adapt to changing conditions and new tasks. The combination of these two approaches solves many common practical problems in industrial processes. Here, by maintaining a shared knowledge base, the model transfers learning between tasks and improves performance on new related tasks. The inclusion of regularization terms λ , μ helps prevent overfitting and ensures the stability of the learning process.

We proposed a stopping criterion called the resurrection mechanism for components reaching a predefined threshold limit. The sum of each component in the matrices L is computed, and its limit is set to $\varepsilon = 10^{-8}$. This assessment step aims to identify components that have failed to contribute effectively to the current training data. For components reaching the limit ε , new random weights are assigned. This process is executed by generating random numbers from a standard normal distribution, ensuring diversity in the new weights. During the revival process, these components are reactivated, allowing them to participate in subsequent learning, thereby enhancing the model's ability to adapt to new data. To this end, this mechanism is reviewed and evaluated after the completion of each task, with the goal of ensuring that the model can continuously optimize the effectiveness of its components throughout long-term training. The threshold component revival mechanism provides an effective dynamic adjustment strategy for sparse learning models, significantly improving their adaptability and learning performance, and demonstrating the potential for addressing complex problems in changing treatment processes and environments.

Evaluation metrics

To evaluate the performance of the proposed streaming virus particle prediction, we utilized four standard metrics, including the coefficient of determination R^2 , RMSE, MAE, MSE, and symmetric mean absolute percentage error (SMAPE):

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2},$$

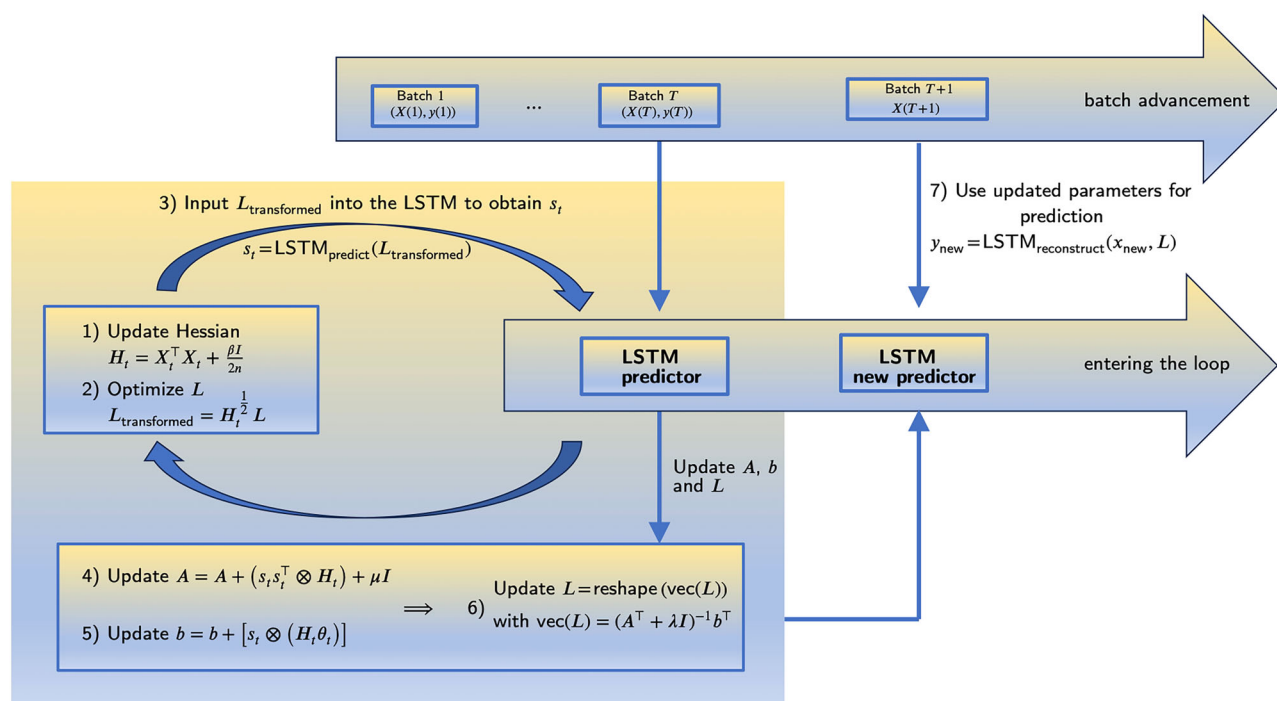


Fig. 6 | The flowchart describes in detail the online prediction strategy for different batches of data, aiming to optimize the parameter adjustment and knowledge adaptation process of the local prediction model. The flowchart includes the following steps: (1) update the Hessian matrix to adapt to the new data; (2) calculate the matrix L to prepare for the subsequent sparse coefficient calculation;

(3) use the matrix L to update the sparse coefficient s_t , which is the key to model adaptation; (4) update the matrix A and (5) vector b to reflect the new learning state; (6) Update the shared matrix L and reshape it, which is a manifestation of adaptability; (7) use the updated parameters to predict new data.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2},$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|,$$

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

$$\text{SMAPE} = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{\frac{|y_i| + |\hat{y}_i|}{2}} \times 100,$$

where y_i refers to the real output at sample i , $\hat{y}_i$ denotes the predicted output at sample i , $\bar{y}$ is the mean value, and n is the number of dataset samples. In all cases, a coefficient of determination R^2 close to 1 and a lower (RMSE, MAE, MSE and SMAPE) value indicate better prediction. For comparison purposes with the existing PLS predictor and the local ML predictors—tree-based ML models, GRU, and LSTM—these metrics are presented in Tables 1, 2, and 3, and the Supplementary material Tables S2, S3, and S4.

The evaluation metrics used a linear weighted averaging method in the adaptation phase. This method assigned higher weights to batches closer to the end, with the first batch receiving the lowest weight and the last batch receiving the highest weight. Each time the dataset is switched, the weighting scheme is reinitialized. This weighting approach effectively captures the evolution of evaluation metrics for the model in time series data while reducing the negative impact of poorly performing metrics during the early adaptation phase (when the model has not fully grasped the data patterns)

on the overall evaluation. This achieves an effective balance between temporal sensitivity and robustness.

Model setting for the three model adaptation cases

We implemented nonlinear predictors such as XGB, CatBoost, GRU and LSTM, and an existing linear predictor PLS. Genetic algorithm and Lasso were used to find the best feature combination. The detailed processing of the genetic algorithm and Lasso regression were given in the supplementary material (Text 2.1.7). In all the three model adaptation performance cases, the regularization parameters μ and β were appropriately chosen to increase the numerical stability of the matrix A and Hessian matrix, respectively, while λ was chosen for the stability of the inverse matrix calculation. The model architecture and hyperparameters of the five ML algorithms for the three developed model adaptation cases are given in the supplementary material (Tables S5, S6, and S7). For rapid training, the entire process utilized CUDA version 11.7 and PyTorch version 2.0.1+cu117.

Data availability

The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Received: 29 November 2024; Accepted: 24 March 2025;
Published online: 06 April 2025

References

1. UNESCO. The United Nations World Water Development Report 2018: Nature-Based Solutions for Water. New York, United States: UNESCO, 2018.
2. Grabow, W. O. K. The virology of wastewater treatment. *Water Res.* **2**, 675–701 (1968).

3. Corpuz, A. V. et al. Viruses in wastewater: Occurrence, abundance and detection methods. *Sci. Total Environ.* **745**, 140910 (2020).
4. Manti, A. et al. Bacterial cell monitoring in wastewater treatment plants by flow cytometry. *Water Environ. Res.* **80**, 346–354 (2008).
5. Alharbi, M., Hong, P.-Y. & Laleg-Kirati, T.-M. Sliding window neural network-based sensing of bacteria in wastewater treatment plants. *J. Process Control* **110**, 35–44 (2022).
6. Zambrano, J., Krustok, I., Nehrenheim, E. & Carlsson, B. A simple model for algae-bacteria interaction in photo-bioreactors. *Algal Res.* **19**, 155–161 (2016).
7. Yang, J. et al. Model-based evaluation of algal-bacterial systems for sewage treatment. *J. Water Process Eng.* **38**, 101568 (2020).
8. Ekundayo, T. C., Adewoyin, M. A., Ijabadeniyi, O. A., Igbinosa, E. O. & Okoh, A. I. Machine learning-guided determination of *Acinetobacter* density in waterbodies receiving municipal and hospital wastewater effluents. *Sci. Rep.* **13**, 7749 (2023).
9. Alharbi, M. S., Hong, P.-Y., & Laleg-Kirati, T.-M. Adaptive neural network-based monitoring of wastewater treatment plants. 3204–3211 (American Control Conference (ACC), 2022).
10. Aljehani, F., N'Doye, I., Hong, P.-Y., Monjed, M. K. & Laleg-Kirati, T.-M. Bacteria cells estimation in wastewater treatment plants using data-driven models. *IFAC-PapersOnLine* **58**, 718–723 (2024).
11. Kadoya, S. et al. A soft-sensor approach for predicting an indicator virus removal efficiency of a pilot-scale anaerobic membrane bioreactor (AnMBR). *J. Water Health* **22**, 967–977 (2024).
12. Farhi, N., Kohen, E., Mamane, H. & Shavitt, Y. Prediction of wastewater treatment quality using LSTM neural network. *Environ. Technol. Innov.* **23**, 101632 (2021).
13. Pisa, I., Santin, I., Morell, A., Vicario, J. L. & Vilanova, R. LSTM-based wastewater treatment plants operation strategies for effluent quality improvement. *IEEE Access* **7**, 159773–159786 (2019).
14. Mokhtari, H. A., Bagheri, M., Mirbagheri, S. A. & Akbari, A. Performance evaluation and modelling of an integrated municipal wastewater treatment system using neural networks. *Water Environ. J.* **34**, 622–634 (2020).
15. Wang, R. et al. Model construction and application for effluent prediction in wastewater treatment plant: Data processing method optimization and process parameters integration. *J. Environ. Manag.* **302**, 114020 (2022).
16. Haimi, H., Mulas, M., Corona, F. & Vahala, R. Data-derived soft-sensors for biological wastewater treatment plants: An overview. *Environ. Model. Softw.* **47**, 88–107 (2013).
17. Jordan, M. I. & Mitchell, T. M. Machine learning: Trends, perspectives, and prospects. *Science* **349**, 255–260 (2015).
18. Lee, J., Bagheri, B. & Kao, H.-A. A cyber-physical systems architecture for industry 4.0-based manufacturing systems. *Manuf. Lett.* **3**, 18–23 (2015).
19. Wang, J., Ma, Y., Zhang, L., Gao, R. X. & Wu, D. Deep learning for smart manufacturing: Methods and applications. *J. Manuf. Syst.* **48**, 144–156 (2018).
20. Ge, Z., Song, Z., Ding, S. X. & Huang, B. Data mining and analytics in the process industry: The role of machine learning. *IEEE Access* **5**, 20590–20616 (2017).
21. Alvi, M. et al. Deep learning in wastewater treatment: A critical review. *Water Res.* **245**, 120518 (2023).
22. Aljehani, F., N'Doye, I., Hong, P.-Y., Monjed, M. K. & Laleg-Kirati, T.-M. A calibration framework toward model generalization for bacteria concentration estimation in wastewater treatment plants. *Sci. Rep.* **14**, 31218 (2014).
23. Chen, Z., & Liu, B. Lifelong Machine Learning. (Springer Nature, 2022).
24. Hinton, G. E., Osindero, S. & Teh, Y.-W. A fast-learning algorithm for deep belief nets. *Neural Comput.* **18**, 1527–1554 (2006).
25. Liu, T. et al. Lifelong learning meets dynamic processes: An emerging streaming process prediction framework with delayed process output measurement. *IEEE Transactions Control System. Technology* **32**, 384–398 (2024).
26. Ruvolo, P., & Eaton, E. ELLA: An efficient lifelong learning algorithm. In: 30th International Conference on Machine Learning. 507–515 (Atlanta, GA, USA, 2013).
27. Ruvolo, P., & Eaton, E. Active task selection for lifelong machine learning. In: Proc. 27th AAAI Conference on Artificial Intelligence. 862–868 (Bellevue, WA, USA, 2013).
28. Ammar, et al. Safe policy search for lifelong reinforcement learning with sublinear regret. In: Proc. 32nd International Conference on Machine Learning. 2361–2369 (Lille, France, 2015).
29. Ammar, et al. Autonomous cross-domain knowledge transfer in lifelong policy gradient reinforcement learning. In: Proc. 24th International Joint Conference on Artificial Intelligence. 3345–3351 (Buenos Aires, Argentina, 2015).
30. Mendez, et al. Lifelong policy gradient learning of factored policies for faster training without forgetting. In: Proc. 34th Conference on Neural Information Processing Systems. 1–12 (Vancouver, BC, Canada, 2020).
31. Shao, W., Tian, X., Wang, P., Deng, X. & Chen, S. Online soft sensor design using local partial least squares models with adaptive process state partition. *Chemometrics Intell. Lab. Syst.* **144**, 108–121 (2015).
32. Jin, H., Chen, X., Wang, L., Yang, K. & Wu, L. Dual learning-based online ensemble regression approach for adaptive soft sensor modeling of nonlinear time-varying processes. *Chemometric Intell. Lab. Syst.* **151**, 228–244 (2016).
33. Chen, S., Cowan, C. F. N. & Grant, P. M. Orthogonal least squares learning algorithm for radial basis function networks. *IEEE Trans. Neural Netw.* **2**, 302–309 (1991).
34. Liu, T., Chen, S., Liang, S. & Harris, C. J. Selective ensemble of multiple local model learning for nonlinear and nonstationary systems. *Neurocomputing* **378**, 98–111 (2020).
35. Liu, T., Chen, S., Liang, S., Du, D. & Harris, C. J. Fast tunable gradient RBF networks for online modeling of nonlinear and nonstationary dynamic processes. *J. Process Control* **93**, 53–65 (2020).
36. Liu, T., Chen, S., Liang, S., Gan, S. & Harris, C. J. Multi-output selective ensemble identification of nonlinear and nonstationary industrial processes. *IEEE Trans. Neural Netw. Learn. Syst.* **33**, 1867–1880 (2022).
37. Ullah, S., Boullila, W., Koubaa, A. & Ahmad, J. MAGRU-IDS: A multi-head attention-based gated recurrent unit for intrusion detection in IoT networks. *IEEE Access* **11**, 114590–114601 (2023).
38. Cheng, H., Monjed, M. K., Myshkevych, Y., Wang, T. & Hong, P.-Y. Accounting for the microbial assembly of each process in wastewater treatment plants (WWTPs): study of four WWTPs receiving similar influent streams. *Appl. Environ. Microbiol.* **90**, e02253–23 (2024).
39. Lee, J., Lapira, E., Bagheri, B. & Kao, H.-A. Recent advances and trends in predictive manufacturing systems in big data environment. *Manuf. Lett.* **1**, 38–41 (2013).
40. Tsung, F. & Zhou, Y. Statistical monitoring of multistage processes: A review. *Qual. Reliab. Eng. Int.* **24**, 763–805 (2008).
41. Mairal, J., Bach, F., Ponce, J. & Sapiro, G. Online learning for matrix factorization and sparse coding. *J. Mach. Learn. Res.* **11**, 19–60 (2009).
42. Van Loan, C. F. The ubiquitous Kronecker product. *J. Computational Appl. Math.* **123**, 85–100 (2000).

Acknowledgements

This work is supported by Near-term Grand Challenge (AI) REI/1/5233-01-01 and KAUST-MEWA SPA (REP/1/6112-01-01) awarded to Peiying Hong. We thank the MODON WWTP operation team for granting us access to various wastewater samples, and Mehmet Mercangöz from Imperial College London for fruitful discussions on the lifelong learning framework.

Author contributions

J.C.: Conceptualization, methodology, investigation, software, validation, writing—original draft, writing—review and editing; I.N.: Conceptualization,

methodology, investigation, software, visualization, writing—original draft, writing—review and editing, supervision; Y.M.: Data curation, writing—review and editing; F.A.: Writing—review and editing; M.M.: Data curation, writing—review and editing; T.L.: Writing—review and editing; P.H.: Conceptualization, writing—review and editing, supervision, project administration, funding acquisition.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41545-025-00461-7>.

Correspondence and requests for materials should be addressed to Ibrahima N'Doye.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025