

RESEARCH ARTICLE

Particle Swarm Optimization for Hyperparameter Tuning of One-Dimensional Convolutional Neural Networks in Wastewater Classification: An Open Problem in Feature-Dependent Generalization

Mohammad Al-Ashoush^{1,*}

¹ Department of Computer Science, College of Information Technology, Amman Arab University, Amman, Jordan.

*Corresponding author: 202220861@aaau.edu.jo

Received: 25 June 2026 Accepted: 10 September 2026 Published: 6 October 2026

DOI: 10.15849/ijopcm.v19i4.101 License: Open access under CC BY 4.0. Authors retain copyright.

Publisher: Al-Zaytoonah University of Jordan (ZUJ), Amman, Jordan.

Cite this article: M. Al-Ashoush, "Particle Swarm Optimization for Hyperparameter Tuning of One-Dimensional Convolutional Neural Networks in Wastewater Classification: An Open Problem in Feature-Dependent Generalization," *Int. J. Open Problems Compt. Math.*, vol. 19, no. 4, 2026, doi: <https://doi.org/10.15849/ijopcm.v19i4.101>

ABSTRACT

Wastewater treatment plants generate large, heterogeneous, and often imbalanced operational records, and classifying the resulting environmental and process indicators remains a persistent obstacle to automating water-quality management. This paper condenses a broader investigation into a hybrid model that couples a one-dimensional Convolutional Neural Network (1D-CNN) with Particle Swarm Optimization (PSO) for hyperparameter and feature selection, evaluated on a six-year operational record (1,383 samples, 18 variables) from Melbourne's Eastern Wastewater Treatment Plant. Feature selection driven by PSO raised classification accuracy to 99% for Total Rainfall and 96% for Average Inflow, but the same procedure left Energy Consumption classification at only 62% accuracy, with several minority classes returning zero precision, recall, and F1-score, and a subset of ROC computations returning non-numeric (NaN) values due to insufficient class support. Rather than presenting these results as a settled solution, this paper argues that they expose an unresolved and label-dependent limitation of PSO-CNN pipelines: feature selection that is optimal for one operational target can be actively unsuitable for another drawn from the same dataset, and severe class imbalance undermines both the reliability of the reported metrics and the trustworthiness of the resulting model for operational deployment. We characterize the boundary conditions under which the PSO-CNN approach succeeds and fails, and we frame the imbalance-sensitive, target-dependent nature of feature selection as an open problem that current single-fold, single-objective PSO formulations do not resolve. The paper closes by outlining concrete directions (class-balancing prior to optimization, multi-objective PSO formulations, cross-validated fitness functions, and target-specific feature subsets) that future work is invited to pursue.

Keywords: Wastewater classification; convolutional neural networks; particle swarm optimization; feature selection; class imbalance; hyperparameter optimization; environmental monitoring.

1 Introduction

Water is central to public health, ecological stability, and economic activity, and the treatment of wastewater before discharge or reuse is a core safeguard against contamination [1]. A necessary step within this safeguard is the classification of pollutants and operational indicators so that the correct treatment pathway can be selected. Historically this classification relied on manual sampling and laboratory analysis, a process that is slow, labor-intensive, and susceptible to human error [2]. The maturation of machine learning and computer vision has opened an alternative: models that learn hierarchical representations directly from sensor and operational data [3].

CNNs are attractive for this task because they can automatically extract discriminative features from raw or lightly processed signals. Their effectiveness, however, is governed by a set of hyperparameters (learning rate, kernel size, number of filters, and depth) that are difficult to tune. Gradient-based tuning is prone to slow convergence, sensitivity to initialization, and entrapment in local optima [4]. PSO, a

population-based metaheuristic inspired by the collective movement of flocking organisms [5], offers a derivative-free alternative capable of exploring high-dimensional, non-convex hyperparameter spaces without the same sensitivity to initial conditions [6].

This paper reports a hybrid PSO-CNN pipeline applied to wastewater operational data and, critically, re-reads its own results as evidence of an unresolved methodological problem rather than as a finished solution. The pipeline performs well on some target variables and performs poorly (in a way that is not simply a matter of degree) on others drawn from the identical dataset and identical preprocessing pipeline. This asymmetry is the paper's central object of study: it suggests that PSO-driven feature selection, as currently formulated in a single-objective, single-split configuration, is target-dependent in a way that limits its reliability as a general-purpose wastewater classification tool. The paper does not propose a definitive fix; instead, it documents the boundary of what the approach can currently do and specifies the conditions a future solution would need to satisfy.

1.1 Research Questions

- RQ1: To what extent does PSO-based feature and hyperparameter selection improve 1D-CNN classification accuracy across multiple, heterogeneous wastewater indicators drawn from a single operational dataset?
- RQ2: Under what data conditions (class support, label distribution, variable type) does PSO-CNN feature selection fail to generalize, and what does that failure look like in the evaluation metrics?
- RQ3: What methodological changes would be required to make PSO-driven feature selection reliable across heterogeneous targets rather than favorable to only some of them?

1.2 Contribution

The contribution of this paper is not a finished, deployable classifier. It is (i) an empirical demonstration that a single PSO-CNN configuration produces highly divergent outcomes across labels from the same dataset (99% to 62% accuracy), (ii) a diagnosis of the class-imbalance and undefined ROC/AUC values due to insufficient class support issues (zero-support classes, NaN AUC values) that accompany the weaker outcomes, and (iii) an explicit problem statement intended to seed future work on imbalance-aware, multi-objective, target-agnostic feature selection for environmental time-series classification.

2 Related Work

Artificial intelligence has been applied to wastewater treatment across a range of tasks: predictive maintenance, anomaly detection, process control, and quality-index estimation [7]. Table 1 summarizes five representative studies alongside the present work, highlighting not only their reported performance but the specific generalization gap each leaves open.

Chawishborwornwong et al. [8] combined principal component analysis with artificial neural networks and a bootstrap procedure to estimate a water-quality index for the Lower Mun River basin, reporting a correlation coefficient of 0.993 and an incorporation rate above 95%, but the approach was validated on a single river basin and its transferability to other index definitions is untested. Maurya et al. [9] reviewed in-silico and machine-learning approaches to heavy-metal detection without proposing or validating a predictive model, leaving an explicit gap between literature synthesis and empirical validation. Aghdam et al. [10] compared Gene Expression Programming against several regression-based learners for predicting BOD5 and COD across seven Hong Kong plants, obtaining R^2 values of 0.784 and 0.861 respectively, but restricted their inflow data to one metropolitan sewer network. Nourani et al. [11], [12] used ensembles of feed-forward networks, support-vector regression, and adaptive neuro-fuzzy inference systems, improving BODeff/CODeff prediction accuracy by up to 20–24% through jittering and ensembling, though each study is confined to a single treatment plant and a narrow parameter set (BOD, COD, TN). Deepnarain et al. [13] paired artificial neural networks with fluorescent in-situ hybridization to relate filamentous bacterial composition to sludge settling behavior across seven South African plants, but the causal operating conditions behind excessive filamentous growth remained unresolved by their own account.

Across this body of work, two patterns recur: first, each study is validated on a narrow, single-source dataset, and second, none of the reviewed studies interrogate how differently their chosen method behaves across multiple, simultaneously available target variables from the same plant. This is precisely the gap the present study probes, not by adding another single-target success case, but by holding the dataset and pipeline constant and varying only the classification target, which surfaces a generalization problem that single-target studies cannot reveal.

Table 1. Comparative summary of related work and the present study.

Study	Method	Dataset / Domain	Key Result	Unresolved Issue
Chawishborwornwong et al. [8]	PCA + ANN + Bootstrap	Lower Mun River WQI	R = 0.993; RMSE = 0.86	Regional generalizability untested
Aghdam et al. [10]	GEP, multi-layer ANN, regression trees	7 WWTPs, Hong Kong (BOD5, COD)	$R^2 = 0.784$ (BOD5), 0.861 (COD)	Limited to one geographic inflow profile
Nourani et al. [11]	FFNN, SVR, ANFIS, ARIMA + jittering ensembles	Tabriz WWTP	Up to 20% accuracy gain via ensembling	Single-plant scope; no feature-selection layer
Deepnarain et al. [13]	ANN + multivariate statistics + FISH	7 South African WWTPs	SVI prediction from filamentous taxa	Mechanistic drivers of bulking left unresolved
Nourani et al. [12]	FFNN, ANFIS, SVM, MLR + ensembling	Nicosia WWTP	Up to 24% accuracy gain (BODeff)	Narrow parameter scope (BOD/COD/TN only)
Present study	1D-CNN + PSO-driven hyperparameter/feature selection	Melbourne Eastern WWTP, 6-yr multivariate record	99% (Total Rainfall), 96% (Average Inflow)	62% on Energy Consumption; unresolved class-imbalance and non-computable AUC cases

3 Materials and Methods

3.1 Dataset

The study uses the publicly available operational record “Energy consumption, climate and wastewater characteristics at Melbourne’s Eastern Wastewater Treatment Plant” [14], covering the period 2014–2019. The dataset¹ contains 1,383 records across 18 numeric (float64) columns spanning process variables (Average Outflow, Average Inflow, Energy Consumption, Ammonia, Biological Oxygen Demand, Chemical Oxygen Demand, Total Nitrogen) and meteorological covariates (temperature, humidity, rainfall, visibility, wind speed) together with temporal indices (Year, Month, Day). No missing values were present in the raw columns used for analysis.

3.2 Preprocessing

Preprocessing followed a standard pipeline: cleaning, decimal scaling of continuous features (e.g., rescaling Energy Consumption by a factor of 1,000 from 150,000–400,000 range to 150–400 for numerical stability), outlier inspection via box plots, and label encoding of the discretized target classes. Each candidate target variable (Total Rainfall, Average Inflow, Average Wind Speed, Ammonia, Energy Consumption) was independently discretized into class bins and used as a separate classification label, holding the same feature pool and preprocessing pipeline constant across targets. The data were split into training, validation, and test partitions, with the validation partition used inside the PSO fitness loop and the test partition withheld for final evaluation.

3.3 One-Dimensional CNN Architecture

The classifier is a 1D-CNN operating over the tabular feature vector treated as a sequence, using convolutional layers to learn local interactions among the environmental and process variables. Its

¹ <https://www.kaggle.com/code/ibrahimzaitoun/full-scale-waste-water-treatment-plant-data>

trainable hyperparameters, learning rate, number of filters, kernel size, and number of convolutional/dense layers — were treated as the search space for optimization rather than fixed a priori.

3.4 PSO-Driven Hyperparameter and Feature Selection

PSO maintains a swarm of candidate particles, each encoding one hyperparameter configuration together with a candidate feature subset. Each particle is evaluated by training the CNN on the training partition with its encoded configuration and scoring the resulting validation accuracy and loss; particle positions are then updated according to individual best and swarm-best experience, balancing exploration and exploitation of the search space [5]. Over successive iterations the swarm converges toward a hyperparameter/feature configuration that minimizes validation loss for the label under optimization. This is a single-objective formulation: the fitness function scores only the classification performance for one target label at a time, with no term penalizing feature-subset instability across labels or class-support imbalance.

3.5 Evaluation Metrics

Model performance was assessed using accuracy, precision, recall, F1-score, confusion matrices, and per-class ROC/AUC curves, together with macro- and support-weighted averages so that the effect of class imbalance on the aggregate accuracy figure could be examined rather than obscured by it.

4 Results

4.1 Baseline CNN without Feature Selection

With the full feature set and no PSO-based reduction, per-class AUC values were highly uneven within a single label. For Average Wind Speed, class-level AUC ranged from 0.933 to 0.983; for Average Inflow, from 0.60 to 0.89 with the strongest class reaching 0.891 and the weakest 0.599; for Energy Consumption, class-level AUC ranged as low as 0.483 and never exceeded 0.715; for Ammonia, the weakest class returned an AUC of 0.451, barely above chance. Several classes across labels returned non-numeric (NaN) AUC values because too few test instances existed to compute a valid ROC curve, an artifact of class-support imbalance rather than of model capacity.

4.2 PSO-CNN with Feature Selection

Introducing PSO-driven feature selection changed accuracy substantially, but unevenly across targets. Total Rainfall reached 99% accuracy, driven almost entirely by a single dominant class (Class 0, precision = recall = F1 = 1.00) while a near-zero-support class (Class 4, one instance) contributed nothing and was scored at 0.00 across all three metrics. Average Inflow reached 96% overall accuracy, again dominated by one large class (precision 0.96, recall 1.00), while a nine-instance minority class fell to an F1-score of 0.18 and a four-instance class scored 0.00 on every metric. Average Wind Speed reached 78% accuracy, with a four-instance class again scoring 0.25 across precision, recall, and F1. Ammonia reached 80% accuracy overall, yet three of its five classes (each with five or fewer supporting instances) scored 0.00 on precision, recall, and F1, meaning the model failed to identify a single true positive in those classes despite the seemingly strong headline accuracy. Energy Consumption, the target most resistant to the PSO-driven pipeline, reached only 62% accuracy, with one class (Class 0) scoring 0.00 across all metrics and a macro-averaged F1-score of only 0.33, indicating that the overall accuracy figure conceals substantial underlying failure. Table 2 summarizes these outcomes.

Table 2. Accuracy before and after PSO-based feature selection, by target label, with the associated class-support issue driving the residual error.

Target Label	Accuracy (no FS)	Accuracy (with PSO-FS)	Dominant class support issue
Total Rainfall	0.78	0.99	Class 0 dominates; Classes 1 and 4 near-zero support
Average Inflow	0.79	0.96	Class 2 support = 4; F1 = 0.00
Average Wind Speed	0.98	0.78	Class 3 support = 4; all metrics = 0.25
Ammonia	0.77	0.80	Classes 2–4 supports ≤ 5 ; F1 = 0.00
Energy Consumption	0.72	0.62	Class 0 F1 = 0.00; macro-F1 = 0.33

4.3 Correlational Structure

A Pearson correlation matrix over the feature set showed a strong positive association between maximum and minimum temperature ($r = 0.92$) and a moderate association between chemical and biological oxygen demand ($r = 0.52$), consistent with known physical and biochemical coupling among these variables. These correlations partly explain why feature selection benefits meteorologically driven labels such as Total Rainfall and Average Inflow more than it benefits Energy Consumption, whose class structure appears comparatively decoupled from the same predictor set.

5 Discussion: An Unresolved, Target-Dependent Limitation

The results in Section 4 do not describe a single model that classifies wastewater data well. They describe a pipeline whose apparent success is conditional on which target variable is being predicted, and whose headline accuracy figures (99%, 96%, 80%, 78%) mask near-total failure on minority classes within the very same labels. This is the paper’s central claim: PSO-based feature selection, in its present single-objective form, optimizes a feature subset and hyperparameter configuration that is locally well-suited to the dominant class structure of one target variable, but there is no mechanism in the fitness function that guarantees, or even encourages, that subset to remain informative for a differently structured target, or for the minority classes within it.

Three symptoms recur across the weaker outcomes and constitute, in our view, an open problem rather than a solved one. First, class-support scarcity: several classes across every label had fewer than ten supporting instances, and in the most extreme cases (Total Rainfall Class 4, Average Inflow Class 2) only a single instance was available, making any precision/recall estimate for that class statistically unstable regardless of the classifier used. Second, undefined ROC/AUC values due to insufficient class support in evaluation: a subset of ROC/AUC computations returned NaN values, meaning the standard evaluation protocol itself could not certify performance on those classes — the problem is not only that the model performs poorly, but that its poor performance cannot always be measured cleanly. Third, metric concealment: weighted-average accuracy and F1-scores, by construction, are dominated by majority classes, so a pipeline can report 80% or 96% accuracy while simultaneously failing completely ($F1 = 0.00$) on two or three of its constituent classes. Any deployment decision based on the headline accuracy alone would therefore be built on an incomplete picture of the model’s behavior.

We consider the Energy Consumption result the sharpest illustration of the underlying problem, precisely because it received the same preprocessing, the same architecture, and the same PSO search procedure as the successful labels, yet converged to a comparatively weak and unevenly distributed outcome (macro-F1 = 0.33 versus an overall accuracy of 0.62). This gap between macro- and weighted-average performance is, in effect, a symptom of the swarm optimizing for the metric it was given (overall validation accuracy) rather than for balanced classification performance, and no term in the fitness function currently penalizes that divergence.

It follows that the PSO-CNN approach, as formulated here and in the related work reviewed in Section 2, has not solved wastewater classification so much as it has solved a family of easier sub-problems within it, those where one class dominates the label distribution and the predictor set happens to correlate well with that class. The harder sub-problem, where classes are balanced, or where the target’s

relationship to the available predictors is weaker (as appears true for Energy Consumption), remains open.

5.1 Boundary Conditions Identified

- PSO-CNN feature selection performs strongly when the target label has one or two dominant classes with high support and when the feature set is physically or statistically well-correlated with the target (Total Rainfall, Average Inflow).
- Performance degrades sharply, in ways the aggregate accuracy metric under-reports, when three or more classes share the label space with fewer than ten instances each (Ammonia, Average Wind Speed).
- Performance degrades further, in both aggregate and per-class terms, when the predictor set's relationship to the target is comparatively weak or mediated by unobserved operational factors, as appears to be the case for Energy Consumption.
- A subset of per-class evaluations become mathematically undefined ROC/AUC values due to insufficient class support (NaN AUC) under extreme class scarcity, meaning the evaluation protocol itself requires revision before performance claims for those classes can be trusted.

6 Limitations

This study is limited to a single plant's six-year record and therefore inherits the same single-source constraint noted in the related work reviewed in Section 2. The PSO configuration used a single fitness objective (validation accuracy/loss) rather than a multi-objective formulation that could jointly weight macro-F1 or class-balanced accuracy. No resampling technique (e.g., SMOTE) or stratified k-fold cross-validation was applied prior to optimization, both of which are candidate mitigations for the class-imbalance symptoms documented above but were left outside the scope of this study. Finally, the discretization thresholds used to convert continuous targets into classes were fixed prior to optimization; how sensitive the reported accuracy gap is to that discretization choice is itself an open empirical question.

7 Conclusion and Open Problems

This paper condensed a broader investigation into PSO-optimized 1D-CNN classification of wastewater operational data and deliberately resisted framing its strongest results (99% and 96% accuracy) as evidence that the underlying problem has been solved. Held against the same dataset and pipeline, a materially weaker and internally uneven outcome for Energy Consumption (62% accuracy, macro-F1 = 0.33) demonstrates that PSO-driven feature selection, as presently formulated, is sensitive to target-specific class structure in ways its fitness function does not account for. We therefore treat the central finding of this paper not as a completed classifier but as a diagnosed, still-open problem: how should hyperparameter and feature-selection search be formulated so that its benefits transfer across heterogeneous, imbalanced targets drawn from the same operational dataset, rather than favoring whichever target happens to have a dominant, well-correlated majority class?

We invite future work along four concrete lines. First, imbalance-aware preprocessing — applying class-balancing techniques such as SMOTE, or stratified k-fold cross-validation within the PSO fitness loop, to test whether the Energy Consumption gap narrows once minority classes are adequately represented during optimization. Second, multi-objective PSO formulations that jointly optimize for overall accuracy and macro-F1 (or a class-balanced accuracy measure), so that the swarm cannot converge on configurations that silently sacrifice minority-class performance for aggregate accuracy. Third, target-agnostic or multi-task feature selection, in which a single PSO search is asked to find a feature subset that performs acceptably across several simultaneous targets rather than one at a time, directly testing whether the divergence documented in Section 4 is intrinsic to the dataset or an artifact of single-target optimization. Fourth, a revised evaluation protocol that treats undefined ROC/AUC values due to insufficient class support (NaN) per-class metrics as a first-class reporting requirement rather than an omission, so that headline accuracy figures cannot be presented without an accompanying account of

which classes could not be reliably evaluated at all. Addressing any one of these four directions would move the field from the label-dependent success reported here toward a wastewater classification pipeline whose reliability does not depend on which operational variable happens to be selected as the target.

Declarations

Funding. The author received no external funding for this research.

Conflicts of interest. The author declares that he has no competing interests.

Ethics approval. Not applicable.

Informed consent. Not applicable.

Data availability. The dataset analysed in this study is publicly available on Kaggle [14].

Author contributions. The author is solely responsible for the conceptualization, methodology, analysis, and writing of this manuscript.

Use of artificial intelligence. No AI tools were used beyond language editing.

References

- [1] S. F. Ahmed, N. Islam, N. Tasannum, A. Mehjabin, A. Momtahin, A. A. Chowdhury, et al., "Microplastic removal and management strategies for wastewater treatment plants," *Chemosphere*, vol. 347, Art. no. 140648, 2024. doi: <https://doi.org/10.1016/j.chemosphere.2023.140648>
- [2] N. A. Zulkifli et al., "Advancements in monitoring water quality based on various sensing methods: A review," *International Journal of Environmental Research and Public Health*, vol. 19, no. 21, p. 14352, Nov. 2022. doi: <https://doi.org/10.3390/ijerph192114080>
- [3] V. R. Allugunti, "A machine learning model for skin disease classification using convolution neural network," *International Journal of Computing, Programming and Database Management*, vol. 3, no. 1, pp. 141–147, 2022. doi: <https://doi.org/10.33545/27076636.2022.v3.i1b.53>
- [4] X. Xiao, M. Yan, S. Basodi, C. Ji, and Y. Pan, "Efficient hyperparameter optimization in deep learning using a variable length genetic algorithm," *arXiv preprint arXiv:2006.12703*, 2020. doi: <https://doi.org/10.48550/arXiv.2006.12703>
- [5] O. F. Aje and A. A. Josephat, "The particle swarm optimization (PSO) algorithm application – A review," *Global Journal of Engineering and Technology Advances*, vol. 3, no. 3, pp. 1–6, 2020. doi: <https://doi.org/10.30574/gjeta.2020.3.3.0033>
- [6] P. Singh, S. Chaudhury, and B. K. Panigrahi, "Hybrid MPSO-CNN: Multi-level particle swarm optimized hyperparameters of convolutional neural network," *Swarm and Evolutionary Computation*, vol. 63, Art. no. 100863, 2021. doi: <https://doi.org/10.1016/j.swevo.2021.100863>
- [7] J. Jin et al., "Recent concerns regarding artificial intelligent (AI) technologies have spurred studies into improving wastewater treatment efficiency and identifying low-carbon processes," *Separations*, vol. 12, no. 9, art. no. 237, 2025, doi: [10.3390/separations12090237](https://doi.org/10.3390/separations12090237).
- [8] C. Chawishborwornwong, S. Luanwuthi, C. Umpuch, and C. Puchongkawarin, "Bootstrap approach for quantifying the uncertainty in modeling of the water quality index using principal component analysis and artificial intelligence," *Journal of the Saudi Society of Agricultural Sciences*, vol. 23, no. 1, pp. 17–33, 2024. doi: <https://doi.org/10.1016/j.jssas.2023.08.004>
- [9] B. M. Maurya, N. Yadav, T. Amudha, J. Satheeskumar, A. Sangeetha, V. Parthasarathy, et al., "Artificial intelligence and machine learning algorithms in the detection of heavy metals in water and wastewater: Methodological and ethical challenges," *Chemosphere*, Art. no. 141474, 2024. doi: <https://doi.org/10.1016/j.chemosphere.2024.141474>
- [10] E. Aghdam, S. R. Mohandes, P. Manu, C. Cheung, A. Yunusa-Kaltungo, and T. Zayed, "Predicting quality parameters of wastewater treatment plants using artificial intelligence techniques," *Journal of Cleaner Production*, vol. 405, Art. no. 137019, 2023. doi: <https://doi.org/10.1016/j.jclepro.2023.137019>
- [11] V. Nourani, P. Asghari, and E. Sharghi, "Artificial intelligence based ensemble modeling of wastewater treatment plant using jittered data," *Journal of Cleaner Production*, vol. 291, Art. no. 125772, 2021. doi: <https://doi.org/10.1016/j.jclepro.2020.125772>
- [12] V. Nourani, G. Elkiran, and S. I. Abba, "Wastewater treatment plant performance analysis using artificial intelligence – an ensemble approach," *Water Science and Technology*, vol. 78, no. 10, pp. 2064–2076, 2018. doi: <https://doi.org/10.2166/wst.2018.477>

- [13] N. Deepnarain, M. Nasr, S. Kumari, T. A. Stenström, P. Reddy, K. Pillay, and F. Bux, “Artificial intelligence and multivariate statistics for comprehensive assessment of filamentous bacteria in wastewater treatment plants experiencing sludge bulking,” *Environmental Technology & Innovation*, vol. 19, Art. no. 100853, 2020. doi: <https://doi.org/10.1016/j.eti.2020.100853>
- [14] I. Zaitoun, “Full-scale wastewater treatment plant data,” Kaggle, 2024. [Online]. Available: <https://www.kaggle.com/code/ibrahimzaitoun/full-scale-waste-water-treatment-plant-data>