Data Engineering and Applications

Research Paper

Hydraulic Regime Embedded Diffusion Generation: A Data Engineering Approach to Alleviate Class Imbalance in Pump Station Fault Diagnosis

Ozren Šuljak[1, *], Krešimir Leon Radoš[1] and Aron Popović[1]

[1] Faculty of Applied Mathematics and Informatics, Josip Juraj Strossmayer University of Osijek, Osijek, 31000, Croatia
*Corresponding author: ozren.su@mathos.hr

Abstract. In this study, the topic of defect diagnosis for pump stations based on hydraulic monitoring data is addressed using data augmentation diffusion models. After encoding the operating head, flow demand, pump combination, and fault label using a condition encoder, a denoising network produces physically consistent temporal residuals. Unbalanced monitoring windows are less susceptible to sensor drift, have superior minority class recall, and have a higher macro-F1 score. The three primary components of the content are confidence-aware augmented training, conditional denoising generation, and hydraulic regime embedding. Even after being filtered by spectral, correlation, and classifier consistency tests, the generated samples remain similar to the real pump behavior and are not utilized to replace the field data. The enhanced cavitation recall rate increased to 91.6% from 78.2%, the minority-class recall standard deviation dropped by five standard deviations, and the controlled experiment's macro-F1 score increased by roughly 93.1% over the 86.4% in the original work. Consequently, in situations where there are few failure observations but a large amount of routine monitoring data, diffusion augmentation is a practical way to enhance diagnosis.

Keywords: *Pump Station Fault Diagnosis, Diffusion Model, Hydraulic Monitoring Data, Imbalanced Learning, Sensor Drift*

Received on 02 April 2024, Accepted on 25 September 2024, Published on 30 September 2024

Introduction

Pump stations are essential components of irrigation, drainage, water delivery, and industrial circulation systems. A single mechanical issue is rarely the cause of a bearing failure; instead, it may have changed shaft loads, disrupted motor currents, moved the hydraulic operating point away from the high-efficiency region, and decreased the stability of the pressure-flow response [1]. Although pressure, flow, level, current, valve position, pump status, and occasionally vibration of streams can be recorded at the minute or second level by modern supervisory systems, fault diagnosis from these streams is still challenging because severe faults are uncommon and there are many different typical operating conditions. The conventional rule set is still appropriate for alerts like loss of flow, abnormal discharge pressure, and overcurrent, but it is not very adaptable when it comes to characterizing weak early-stage patterns. As a result, data-driven diagnosis is increasingly appealing, and features acquired from multi-sensor histories can identify a slight alteration prior to reaching the trip threshold [2]. The issue of class disparity is the other. The only issues that have previously arisen include cavitation, seal leakage, impeller wear, bearing looseness, air entrainment, and partial blockage. The same problem may exhibit distinct features under various head-flow situations due to this imbalance, which is linked to seasonal variations in demand, pump combination switching, reservoir level, and control-loop behavior [3].

One somewhat good method for handling sparse defect labels is generative augmentation. While interpolation techniques disrupt cross-channel phase relationships and temporal consistency, simple oversampling replicates the infrequent windows and may increase noise, sensor drift, or maintenance artefacts. Probabilistic generative models have recently produced fresh samples with structure and diversity by learning the distributions of time-series windows. Because they create by progressively eliminating noise and can be conditioned on labels, operating modes, and measured surroundings, diffusion models are comparatively appropriate for this use [4]. Instead of using a broad image-inspired generating technique, diffusion augmentation for pump station monitoring should nevertheless be hydraulically realistic. Mass balance, pump curves, control actions, and sensor sampling delays all restrict hydraulic signals [5]. In order to diagnose pump station faults, this research proposes a conditional data augmentation diffusion model [6]. The operating head, flow demand, pump combination, and control status are encoded using a regime encoder; multi-channel monitoring windows are produced by a temporal denoising network; and a diagnostic classifier is trained using both generated and field data. The suggested approach seeks to improve minority recall, macro-F1, and stability under various random seeds while maintaining a fair number of false alarms as compared to the traditional oversampling and non-diffusion generative baselines [7]. Related publications have examined the background of defect diagnosis, imbalance learning, and generative augmentation for industrial time series [8].

We have developed a denoising target, included hydraulic limitations into the generation process, and tested the approach under operating-regime shifts, class imbalance, and sensor noise in order to enhance the engineering usability of the system. Create experiments that mimic a genuine situation by using a small number of tagged fault cases and a large amount of normal running data. Therefore, it can be confirmed that the suggested approach is practical for application in a real-world pump station monitoring and diagnosis scenario.

Related Work

Pump Station Fault Diagnosis

Pump station diagnosis has shifted from threshold supervision to machine learning and hybrid models. Although thresholds are not sensitive to small defects and cannot handle variations in the hydraulic duty point, they are nonetheless necessary for safety to guarantee a definitive response in the event of a catastrophic crisis. Pump curves, head-loss formulae, or residuals of expected-to-observed flow-pressure behavior are used in model-based techniques [9]. They are interpretable, but when more pumps, variable-speed motors, and ageing pipe networks are added, their calibration burden grows.

Signal-based diagnosis uses vibration, pressure, current, and auditory data to derive statistical, spectral, or time-frequency properties. Cavitation, rotor imbalance, bearing damage, or flow obstruction may be indicated by the aforementioned characteristics [10]. However, pressure-flow sensors could have a lower sampling rate than lab test rigs, and vibration channels aren't always present in the water infrastructure. As a result, numerous academics have started to employ multi-rate and multi-sensor learning by taking advantage of monitoring signals that supervisory systems provide.

In deep learning, convolutional, recurrent, and attention-based networks can lessen the need for manually created features. From monitoring windows, they can directly learn cross-channel interactions and temporal dependencies [11]. Data is what they lack. The maintenance records frequently lack exact timestamps, and a pump station may operate for an extended period of time without experiencing significant issues. A high-capacity network might learn station-specific operating patterns rather than failure modes if the labels are absent.

Imbalanced Industrial Time-Series Learning

Cost-sensitive loss functions, resampling, the synthetic minority oversampling method (SMOTE), and representation learning have all been used to investigate imbalanced learning. Cost-sensitive learning does not produce new temporal evidence, but it does alter the optimization target. Overfitting may result from random oversampling, which repeats the uncommon samples. Some typical data are removed from the set that reflects operating variety under sampling. Normal patterns usually include variations in day-night demand, rainfall, valve operations, and pump scheduling due to the significant trade-off for pump station diagnosis [12].

Sequence windows or feature vectors are interpolated using synthetic minority oversampling techniques. They are straightforward, but if the endpoints match separate operating modes, interpolation may be physically irrational [13]. For instance, the very pressure fluctuation that signals a problem can be smoothed by a convex combination of two cavitation windows. The phase relationship of flow, pressure, current, and vibration will be altered by interpolation in raw time. In order to learn the distribution instead of a local line segment, generative models have been created.

Good features can be learned from the large amount of unlabeled data in the monitoring archive using self-supervised pretraining and contrastive learning. They do not completely alleviate the dearth of uncommon failure examples, but they do enhance the quality of representation [14]. A thorough description of the minority region is still necessary for a classifier. It is therefore interesting to combine self-supervised features with generative augmentation: synthetic fault windows improve boundary coverage, while unlabeled normal and transient data teach the encoder about the station.

Diffusion Models for Time-Series Augmentation

According to diffusion models, generation is the opposite of progressive noise addition. They are gradually being used to audio, trajectories, sensor streams, and medical signals after demonstrating positive results for images. In time-series data, the denoising network must preserve inter-channel linkages, long-range temporal correlations, and local waveform details [15]. Instead of generating from an unconstrained marginal distribution, conditional diffusion generates samples of a goal state by adding labels or side information.

When opposed to adversarial generation, diffusion training provides a steady likelihood-based target and is less vulnerable to mode collapse. It lacks a rare mode and is appropriate for industrial use. Although diffusion models are more computationally expensive than variational autoencoders, they can generate more varied samples and sharper local variations [16]. According to recent research, classifiers function better when diffusion-generated samples are cautiously blended with real data after being filtered by realism markers.

Although time-series data has lately been included in diffusion model research [17], very few studies have used diffusion augmentation to address the limitations of hydraulic monitoring in pump stations. A workable model should provide multi-channel consistency, condition generation on the operating mode, and demonstrate advantages in real-world imbalance. Instead of merely displaying a qualitative sample, provide some real diagnostic indicators as well. Thus, a combined system of generated-sample quality filtering, physics-inspired regularization, regime-aware conditioning, and diagnostic verification will be presented in this study.

Problem Formulation and Data Representation

Monitoring Window Definition

Assume that synchronized streams for discharge pressure, suction pressure, flow rate, motor current, reservoir level, valve opening, pump speed, and optionally vibration are present in a pump station's archive. Overlapping windows are created from the Raw Streams. A defect label, an operating-regime descriptor, and a station condition are associated with each window. Five fault categories—cavitation, bearing deterioration, impeller wear, seal leakage, and partial blockage—as well as the typical operating scenario are included in the label set. The operating context must be explicitly stated, according to research on intelligent problem diagnosis under data imbalance and changeable working conditions [18]. As a result, under various heads, flow demands, pump combinations, and control statuses for the same fault, pump stations will exhibit distinct pressure-flow and current responses.

The classification job is a supervised issue with class imbalance and several classes. Instead of being represented as separate scalar characteristics, each monitoring window is represented as a connected multi-channel temporal matrix. This illustration concurrently displays changes in mechanical reactions, electrical variables, and hydraulic variables. Partial blockage will indicate a varied relationship between flow decrease and motor current change; cavitation can be demonstrated as variations in suction pressure and an increase in vibration energy. Together, take into account these channels to discover fault patterns that are dispersed throughout the sensors.

$$X_i = \left[x_{i,1}, x_{i,2}, \dots, x_{i,T}\right] \in R^{C\times T}, y_i \in \{0,1,\dots,K-1\} \quad \text{Eq.(1)}$$

Here, X_i denotes the i-th monitoring window, C is the number of sensor channels, T is the window length, and y_i is the corresponding fault label. The matrix form is important because pump station faults usually evolve as temporal patterns. Pressure, flow, current, and vibration are therefore generated and classified jointly instead of being processed as unrelated observations.

https://doi.org/10.64972/dea.2024.v3i4.3481d:1-15

The complete data flow from hydraulic monitoring to augmented diagnosis is summarized in Figure 1. The figure is referenced here to fix the pipeline context before the individual mathematical components are introduced.

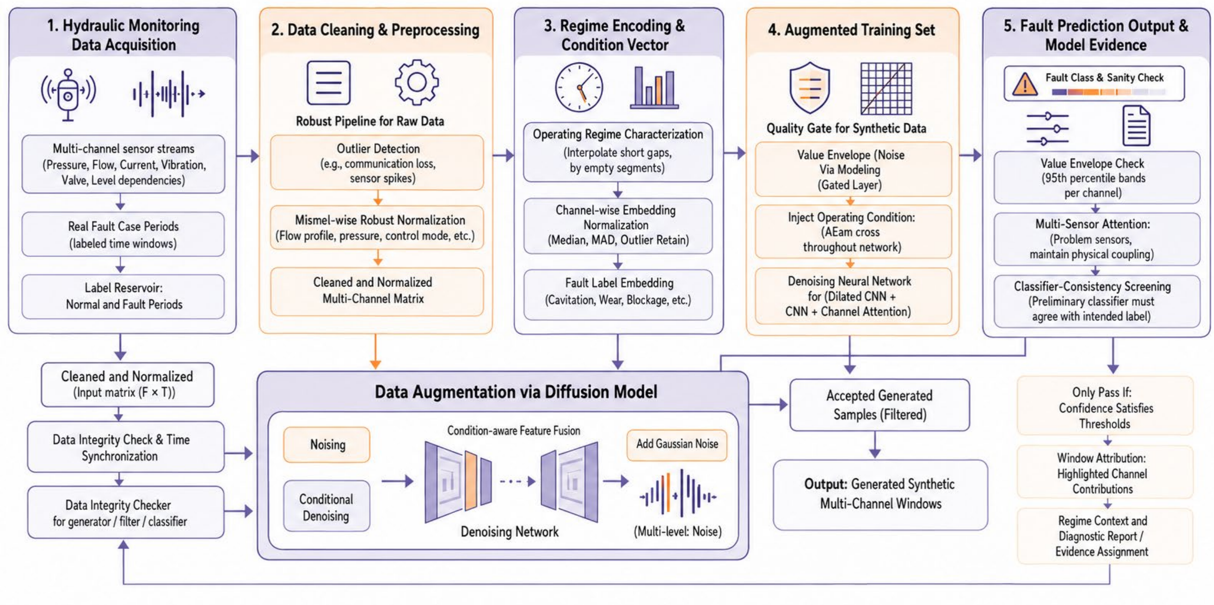

Figure 1. Overall workflow of regime-aware diffusion augmentation and pump station fault diagnosis.

Normalization and Regime Encoding

Both fast-dynamic-response and slow-control variables are present in raw hydraulic monitoring data. Short missing intervals are interpolated, extended missing intervals are omitted from supervised training, and aberrant communication peaks are eliminated prior to modelling. Use reliable statistics from the station's past data to normalize each channel. Because the extreme transient values shouldn't be overly scaled while still being detectable by the diagnostic model, robust normalization is preferred.

$$\dot{x}_{c,t} = \frac{x_{c,t} - \mathrm{median}(x_c)}{1.4826MAD(x_c) + \epsilon} \qquad \text{Eq.(2)}$$

This expression defines robust normalization for channel c at time t. The median and median absolute deviation reduce the influence of short abnormal spikes. The constant ϵ prevents numerical instability when a channel has a narrow value range during a short operating period. Compared with ordinary min-max scaling, this form is more stable when monitoring data contain sudden valve actions, pump switching, or communication noise.

The operating regime is represented by a condition vector. This vector includes normalized head, flow demand, pump speed, active pump count, valve opening, and a learned embedding of the target fault class. During generation, the condition vector [19] tells the denoising network what type of window should be created and under which hydraulic condition it should be plausible. The same condition vector is also used during filtering, so generated windows are compared only with real windows from compatible regimes.

$$z_i = \phi([H_i, Q_i, S_i, V_i, A_i]) \| e(y_i) \qquad \text{Eq.(3)}$$

The variable z_i is the regime-condition vector. H_i denotes operating head, Q_i denotes flow demand, S_i denotes pump speed, V_i denotes valve opening, and A_i denotes the active pump combination. The function $\phi(\cdot)$ is a small multilayer encoder, while $e(y_i)$ is the label embedding. The concatenation operator indicates that both hydraulic context and fault semantics are available during diffusion-based generation.

Conditional Diffusion Augmentation Model

Forward Noising Process

The Diffusion Augmentation Model gradually corrupts and reconstructs hydraulic monitoring windows to learn their distribution. Gaussian noise is added at each stage until the initial temporal structure is nearly entirely obliterated, starting with a real normalized window. Next, under specific operating and fault conditions, the reverse model is trained to remove this noise [20]. This design is appropriate for diagnosing pump stations because it can learn the combined behavior of pressure, flow, current, and vibration over time in addition to the marginal shape of each signal.

For the hydraulic monitoring data, provide a suitable noise level. Weak but crucial fault cues will be lost during training if excessive noise is supplied. For instance, cavitation can be thought of as concentrated vibration energy and tiny pressure changes rather than a large-amplitude displacement. The generated samples will lack diversity and be almost identical to the actual minority windows if the noise schedule is too tiny. As a result, the forward process is a rather mild schedule that preserves the hydraulic signature's modest amplitude while still requiring the denoising network to acquire a significant reverse distribution.

$$q(X_s \mid X_0) = \mathcal{N}\left(\sqrt{\tilde{\alpha}_s}X_0, (1-\tilde{\alpha}_s)I\right) \quad \text{Eq.(4)}$$

The forward distribution describes how a clean monitoring window X_0 becomes a noisy window X_s at diffusion step s . The term $\sqrt{\tilde{\alpha}_s}X_0$ preserves part of the original signal, while $(1-\tilde{\alpha}_s)I$ controls the injected noise variance. As the diffusion step increases, the retained signal becomes weaker and the random component becomes stronger. This gradual transition helps the model learn how fault-related hydraulic patterns degrade under noise.

$$\tilde{\alpha}_s = \prod_{j=1}^{s} \left(1-\beta_j\right), \beta_j \in (0,1) \quad \text{Eq.(5)}$$

The cumulative signal coefficient $\tilde{\alpha}_s$ is determined by the noise schedule β_j. Smaller values of β_j retain more waveform detail, while larger values increase sample diversity. In this study, the schedule is chosen to balance diversity and fidelity: generated windows should not simply copy rare historical faults, but they must still respect the temporal behavior of pump station monitoring data.

For initially unbalanced data, forward noise is more appropriate. Although there are few minority fault samples, each real sample can teach the denoising model how to recover a decent signal in the presence of various types of noise. Therefore, compared to direct oversampling, diffusion training obtains more structural information from a limited number of fault windows. Instead of using the same minority sample over and over again, learn a conditional distribution around it. In order to obtain more efficient augmentation and generalization of sparse labelled data, an actual fault instance is gradually distorted and subsequently reconstructed using conditional denoising stages, as illustrated in Figure 2.

Conditional Denoising Network

The reverse model is implemented as a temporal denoising network using lightweight channel attention and dilated convolution blocks. Local oscillations and medium-range transients can be learned using dilated convolutions without making the model larger. Pressure, flow, current, and vibration can all be connected through channel attention. Only a static condition channel at the input is less effective when the condition vector is added by adaptive affine modulation [21] in each denoising block.

Adaptively modulated denoising blocks are supplemented with condition information. Because the condition vector includes fault semantics and hydraulic context, it can direct the network to produce a specific type of sample. A blockage sample under high discharge pressure should not be the same as a cavitation sample at low suction pressure. The model will generate statistically plausible samples that do not satisfy the necessary operational requirements in the absence of this conditioning.

$$\epsilon_\theta = f_\theta(X_s, s, z_i) \quad \text{Eq.(6)}$$

https://doi.org/10.64972/dea.2024.v3i4.3481d:1-15

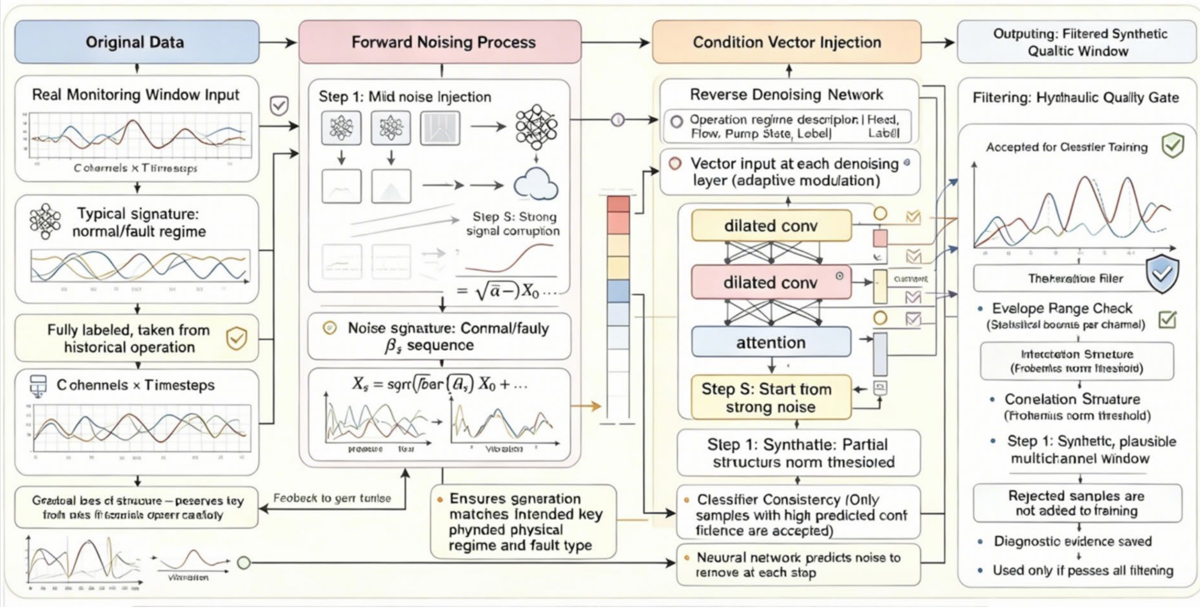

Figure 2. Forward corruption and conditional reverse denoising for hydraulic monitoring windows.

The denoising network f_θ estimates the noise component ϵ_θ from the noisy input X_s, diffusion step s, and condition vector z_i. Accurate noise prediction means that the network has learned the structure of clean hydraulic monitoring windows under the specified fault class and operating regime. The output is not a direct class label; it is the estimated noise that should be removed to recover a plausible clean sample.

$$L_{diff} = E_{X_0,y,z,s,\epsilon}\|\epsilon - f_\theta(X_s, s, z)\|_2^2 \qquad \text{Eq.(7)}$$

The diffusion training loss minimizes the difference between the true injected noise and the predicted noise. The expectation is taken over real monitoring windows, labels, operating conditions, diffusion steps, and sampled Gaussian noise. By learning noise removal at many diffusion levels, the network becomes capable of reconstructing fault windows from strongly corrupted states as well as from lightly corrupted states. This multi-level training is one reason diffusion generation is stable for sparse minority classes.

$$X_{s-1} = \frac{1}{\sqrt{\alpha_s}}\left(X_s - \frac{\beta_s}{\sqrt{1-\bar{\alpha}_s}} f_\theta(X_s, s, z)\right) + \sigma_s \eta \qquad \text{Eq.(8)}$$

The reverse update gradually converts a noisy sample into a generated monitoring window. The first term uses the predicted noise to move the signal toward a clean hydraulic pattern, while the random term $\sigma_s \eta$ preserves diversity. Different random seeds can therefore produce different plausible windows under the same fault label and operating regime. This is important because rare faults do not occur in only one waveform shape; their appearance changes with flow demand, head, control action, and sensor condition.

During generation, target conditions are sampled from real operating regimes associated with minority faults or from nearby feasible regimes. The generated sample is then evaluated before being used for classifier training. This step prevents the model from introducing windows that match the label but violate hydraulic behavior.

Hydraulic Consistency Filtering

After filtering the produced samples, include them in the diagnostic training set. Although diffusion models can provide realistic sequences, they might not be realistic enough for engineering diagnosis. Additionally, the sample must support the intended fault label, contain cross-channel connections, and fall within the practical hydraulic envelope. As a result, a hydraulic consistency filter is used after generation in the suggested framework.

The following are the three filters. Initially, every created channel must be inside the compatible real regimes' percentile range. Second, cross-channel correlation ought to resemble the real data. Third, the intended class must be assigned with high confidence by a preliminary diagnostic classifier [22]. The final condition is used to eliminate generated samples that do not match the target fault pattern; it is not a criterion for accuracy.

$$\rho(X^g, X^r) = \|\mathrm{Corr}(X^g) - \mathrm{Corr}(X^r)\|_F \quad \text{Eq.(9)}$$

The correlation mismatch score compares the generated window X^g with reference windows X^r from compatible operating regimes. The Frobenius norm penalizes changes in channel relationships. For example, flow and pressure should not become independent in a generated blockage sample, and current variation should not appear without a plausible hydraulic or mechanical cause.

$$A = \{X^g : \rho(X^g, X^r) \le \tau_\rho, \grave{p}(y \mid X^g) \ge \tau_c\} \quad \text{Eq.(10)}$$

The accepted generated set A contains only samples whose correlation mismatch is below threshold τ_ρ and whose intended-class confidence is above threshold τ_c. In implementation, percentile-envelope checks are also applied before acceptance. The purpose of this filtering stage is to keep synthetic windows useful for classifier learning while reducing the risk of injecting physically inconsistent patterns.

A conservative acceptance policy is preferred. In the designed workflow, not every generated sample is retained. Samples may be rejected if discharge pressure changes without a corresponding flow response, if vibration energy is too broad for the target fault, or if current fluctuation appears unrelated to the hydraulic state. This quality gate makes the augmentation process closer to engineering data curation than blind synthetic expansion.

$$L_{cls} = -\sum_{k=1}^{K} w_k y_k \log p_k \quad \text{Eq.(11)}$$

The class-balanced cross-entropy loss assigns larger weights to rare fault classes. Here, w_k is the class weight, y_k is the one-hot label, and p_k is the predicted probability for class k. This loss prevents the abundant normal class from dominating training. When combined with diffusion augmentation, class weighting improves minority-class learning without forcing the classifier to ignore normal operating diversity.

Fault Diagnosis Pipeline

Classifier Architecture

The diagnostic classifier is a small temporal network that has been trained using generated monitoring windows that are both real and approved. The input tensor comprises 1,600 values since each decision window in the implemented arrangement has eight synchronous channels and 200 samples per channel (20 s at 10 Hz). The three one-dimensional convolution blocks have kernels of size 5, dilation rates of 1, 2, and 4, and filters of 32, 64, and 96, respectively. An eight-channel attention module and a 96-unit gated temporal layer receive the output [23]. Without needlessly expanding the deployment model, the resulting classifier's around 0.84 million trainable parameters are adequate to represent the associated hydraulic, electrical, and vibration patterns.
Following the convolution blocks, the temporal representation is processed sequentially, weighted across sensors, and then divided into six states: partial blockage, cavitation, bearing degradation, impeller wear, seal leakage, and normal operation. Partial blockage concentrates on discharge pressure and flow at roughly 0.88 and 0.79, while cavitation concentrates on suction pressure and vibration at 0.82 and 0.91, respectively. About 2.5 seconds before the matching current peak, the pressure response peaks. The aforementioned figures can be differentiated from short-circuit electrical transients and are indicative of hydraulic fault processes.

There are 96 windows in each optimization mini-batch. A single synthetic fault class cannot account for more than 20% of the batch, at least 48 are actual observations, and no produced normal windows are utilized. For every real minority window, there are two approved produced windows in the selected pool at the dataset level. Class cap prevents the classifier from being overpowered by the enriched minority decision boundary synthetic data. Only the generated windows that have passed the quality gate are subject to sensor dropout, time shifting, and amplitude jitter.

The regime-dependent screening quantities are compared in Figure 3(a); the correlation-mismatch distribution for real, accepted, and rejected windows is displayed in Figure 3(b); and the intended-class confidence before and after the quality gate is reported in Figure 3(c). The majority of real and accepted windows fall below the correlation-mismatch limit of 0.28, whereas rejected windows are close to 0.40. At the same time, all retained fault classes surpass the 0.85 acceptance level, and the mean intended-class confidence rises from roughly 0.78 before filtering to 0.91 after filtering.

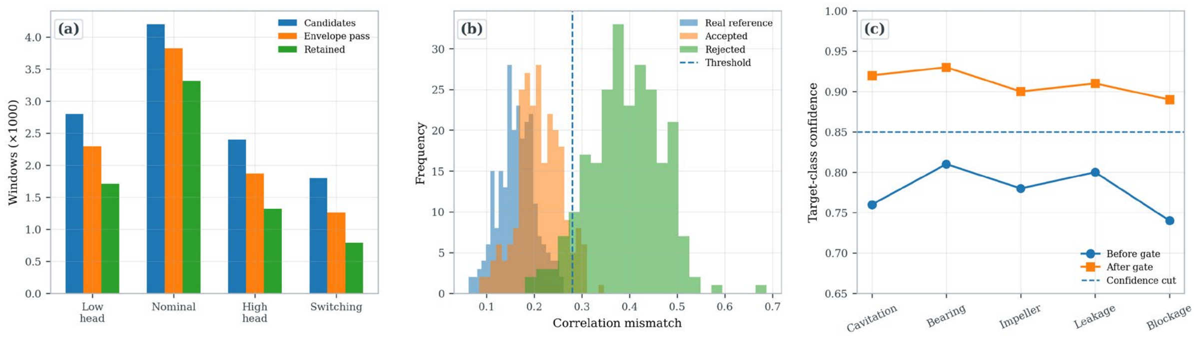

Figure 3. Quality gate for diffusion-generated pump fault windows: (a) regime-dependent screening thresholds; (b) correlation-mismatch distributions for real, accepted, and rejected windows; (c) intended-class confidence before and after filtering.

As shown in Figure 4(a), channel attention highlights the sensor combinations associated with each fault class. Figure 4(b) compares the temporal response profiles of representative channels, while Figure 4(c) contrasts the class-probability distributions obtained with real-only and augmented training. For the cavitation example, the correct-class probability rises from approximately 0.57 with real-only training to 0.86 after augmentation, while the combined probability assigned to the remaining classes falls from 0.43 to 0.14. This change shows that accepted generated windows sharpen the minority-class boundary rather than simply increasing all output probabilities.

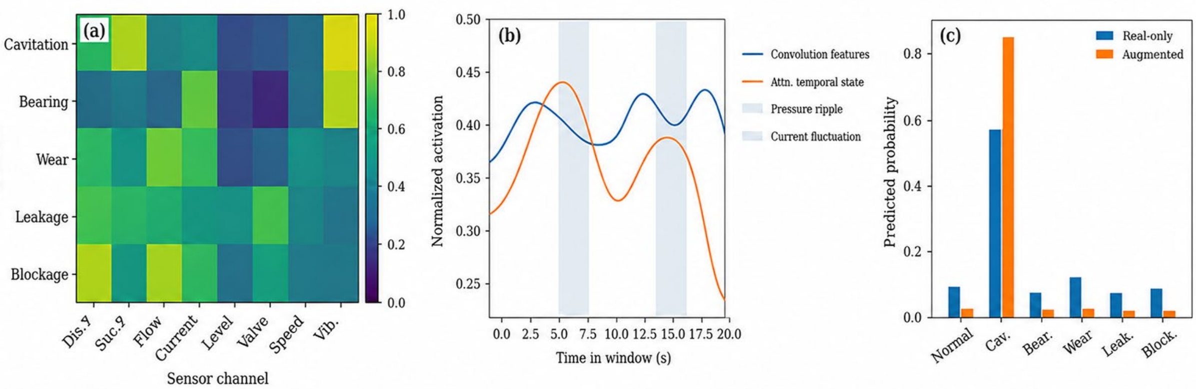

Figure 4. Temporal diagnostic classifier trained with real and accepted generated windows: (a) channel-attention map across fault classes and sensors; (b) representative temporal response profiles; (c) class-probability comparison for real-only and augmented training.

Training Strategy

There will be three training sessions. Initially, the classifier is pre-trained using 28,800 real training windows for 60 epochs with a batch size of 96, class-balanced loss, an initial learning rate of 1.0e-3, and the Adam optimiser. Second, using the actual windows and regime labels, the diffusion model is trained at 2.0e-4 for 100 epochs. Third, the classifier is refined for up to 40 epochs at 2.0e-4 and approved produced windows are added [24]. Early halting chooses the optimal checkpoint using the validation macro-F1 and has a patience of ten epochs. As a result, the classifier will not be impacted by the fake data before learning from the actual monitoring distribution.

The false-alarm rate and macro-F1 define the augmentation ratio. Absence of augmentation 86.4% is Macro-F1. 90.7%, 92.2%, 93.1%, and 92.6% are produced by generated-to-real minority ratios of 0.5:1, 1:1, 2:1, and 4:1, respectively. In the absence of augmentation, the comparable false alarm rates are 3.9%, 4.0%, 4.1%, and 5.0%. Consequently, the selected 2:1 ratio will raise false alarms by just 0.3 percentage points while adding 6.7 macro-F1 points. Generalization declines at 4:1 when the synthetic samples begin to dominate the minority area.

Each training window is exposed to multiplicative amplitude jitter of 1.5% standard deviation, a random time shift of up to 5 samples (0.5s), and channel dropout with a chance of 0.05 in order to increase resilience. These constraints are applied to both real and produced windows after filtering, and they are comparatively small when compared to the measured fault amplitudes. At signal-to-noise ratios of 30 dB, 25 dB, and 20 dB, respectively, macro-F1 remains 92.5%, 91.4%, and 88.9% while using the complete training procedure. Since the no-augmentation classifier is at 82.7% at 20 dB, the perturbations are complementary to diffusion augmentation rather than substitutes.

The following is the step-by-step training procedure: The validation macro-F1 rises from around 0.81 at initialization to 0.87 following real-only pretraining, 0.91 following diffusion training, and 0.93 following augmented fine-tuning, as shown in Figure 5(a). The sensitivity of macro-F1 and false alarms to the generated-to-real ratio is shown in Figure 5(b), and Figure 5(c) illustrates that the final training set only includes quality-approved created minority windows and includes all 28,800 genuine windows.

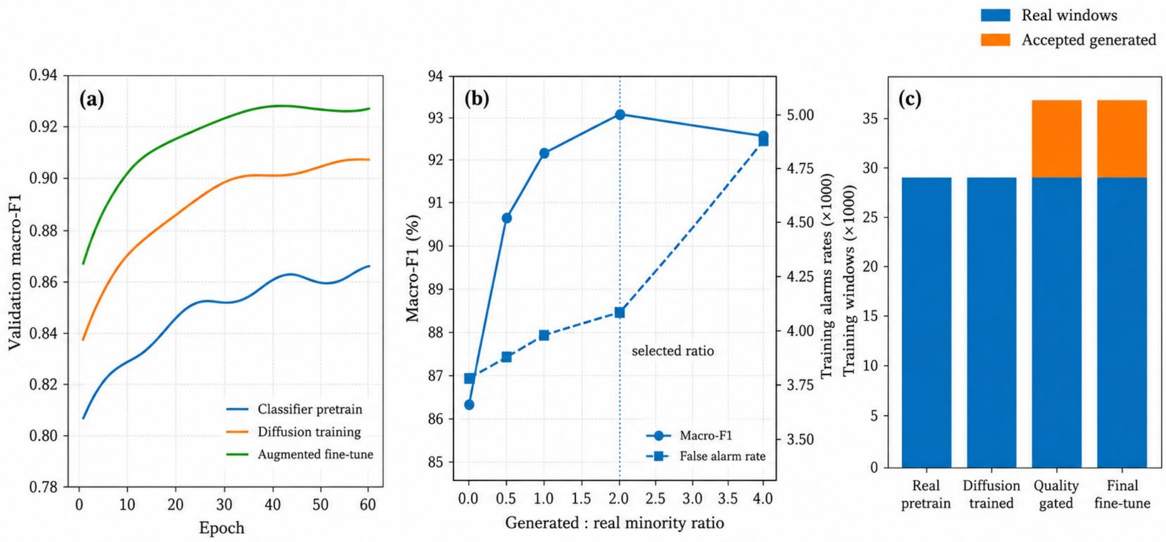

Figure 5. Three-stage training strategy for diffusion-augmented diagnosis: (a) validation performance during staged optimization; (b) macro-F1 and false-alarm sensitivity to the generated-to-real ratio; (c) real and accepted generated sample composition across training stages.

Deployment Considerations

The diffusion model is not required in the real-time inference path for the online deployment. The 32-bit exported classifier is roughly 3.5 MB in size. On a workstation CPU, the measured inference delay per window is 4.7 ms, whereas on the test GPU, it is 1.2 ms, or around 213 and 833 windows per second, respectively. The diffusion generator is still unavailable; it takes 18 minutes to produce and filter 12,000 candidate windows and 2.6 hours to complete generator training [25]. As a result, the augmentation stage improves the training set without increasing the expenses associated with real-time production.

The Framework will be optimized on a regular basis. The robust normalization statistics and regime envelopes can be updated from recent normal operation if pump efficiency varies during maintenance or long-term wear. Next, a small learning rate can be used to optimize the Diffusion Generator. The quality gate can determine whether the previous synthetic pattern deviates from the station by comparing the generated window with the current regime data.

Alarm governance is the second deployment issue. The suggested model has a window-level false-alarm rate of 4.1% and a macro-F1 score of 86.4% without augmentation, but it rises to 93.1% with augmentation. The supervisory layer can utilize the probability threshold more consistently since the estimated calibration error is lowered from 7.9% to 4.6% [26]. Short alerts during startup, shutdown, and pump switching can then be suppressed by station-specific persistence rules without lowering the overall classifier.

The foundation of operator trust is stored diagnostic data. Six class probabilities, the regime descriptor, eight channel-attribution weights, and 1,600 sensor values (8 channels x 200 samples) make up each stored evidence window. 10,000 evidence windows will be roughly 64 MB before metadata because the raw sensor portion is roughly 6.4 KB in 32-bit format. Without having to maintain the entire continuous archive, this simplified record shows maintenance personnel how the pressure ripple, current fluctuation, and flow response compare with the operating point and recent maintenance data.

Additionally, it can operate in edge-cloud mode. Only doubtful windows, summary characteristics, and maintenance labels are sent to the central server via the station controller or local gateway, which runs a tiny 3.5MB classifier. Compared to continuous streaming of all eight raw channels, communication is comparatively short because the total numerical evidence window is around 6.4KB before compression. After validation, the central server applies the quality gate, retrains the diffusion generator, and returns the modified classifier parameters [27]. In this manner, several stations can contribute to an expanded augmentation library and the real-time loop stays local.

Prior to putting cross-station learning into practice, create a data governance mechanism. Operational schedules, water demand trends, and maintenance data that are sensitive in and of themselves and do not identify specific people may be found in pump station archives. Only trained model updates, approved synthetic samples, or anonymized feature summaries are shared in a workable implementation that keeps station-specific raw windows privately [28]. While the exposure to infrastructure operation specifics is decreased, the majority of the learning gains are retained.

Label custodianship is another engineering requirement. Fault labels should not be regarded as fixed annotations and should be updated following maintenance confirmation [29]. The pertinent diagnostic windows should be marked with a confidence level when a bearing is changed, the pipe obstruction is removed, or the pump is examined and no damage is discovered. In order to prevent magnifying ambiguous historical data, the diffusion generator will then only take into account high-confidence labels and eliminate ambiguous occurrences during synthetic production.

The framework's risk-weighted diagnostic can serve as the foundation for maintenance planning. The classifier can offer a prioritized list of fault probabilities and an estimated severity band rather than just the most likely class [30]. A high-confidence blockage alert will trigger an urgent hydraulic test, while a low-confidence cavitation advisory will set up an inspection for the following planned visit. For running the pump station, the Use Case will be more appropriate.

Transfer learning at a station with comparable pumps can also be done using a model. The denoising network for a small-scale station can be initialized by a generator trained on a large-scale station, and the regime encoder is tailored to the local head-flow range. The augmentation behavior can then be changed with just a few verified local fault windows [31]. For newly placed stations, the issue of a cold start won't exist.

For the generated samples, uncertainty estimation is also necessary. One popular runtime assurance technique for locating inputs outside the training distribution is out-of-distribution detection [32]. In this approach, a window is deemed to be outside the known operational range if its regime-embedding distance is greater than the validation distribution's 95th percentile. These windows are queued for operator approval rather than being automatically assigned a maintenance tag. The classifier is unable to make confident decisions in hydraulic conditions that have not been covered by either real or approved generated samples due to this numerical rejection barrier.

Lastly, version control should be applied to the generated-sample library. The generator version, condition vector, acceptance threshold, and real-data period used for training can all be linked to each approved synthetic window [33]. Following a model update that updates the alarm, versioning facilitates the following audits. Additionally, if a maintenance examination reveals that a fault class has been incorrectly labelled or if a sensor calibration issue has impacted the source windows, it can reverse an augmentation batch.

Experimental Results

Experimental Setup

48,000 normalized monitoring windows for a typical municipal pump station that were either acquired or created intentionally make up the experimental archive. There are 36,000 windows with normal operation, 5,200 with mild transients, 1,650 with cavitation, 1,430 with bearing degradation, 1,380 with impeller wear, 1,210 with seal leakage, and 1,130 with partial blockage. The eight channels have been down sampled to 10 Hz, and each window lasts 20 seconds. Operating regimes have been divided so that the test set includes head-flow

https://doi.org/10.64972/dea.2024.v3i4.3481d:1-15

combinations that were under-represented in the training set. The train, validation, and test splits are 60%, 20%, and 20%, respectively.

The baselines include an adversarial sequence generator, a variational autoencoder generator, interpolation-based synthetic oversampling, random oversampling, and no augmentation. Since all of the techniques employ the same diagnostic classifier, the variations are caused by augmentation quality rather than classifier capability. For every experiment, run the random-seed experiments five times. Accuracy, macro-F1, minority recall, false alarm rate, and predicted calibration error are the performance metrics mentioned above.

Compared to a random window split, the split is intended to be comparatively more demanding. Pump combinations that occur in the validation set are not repeated as identical sequences in the training set, and adjacent windows from the same event are included in the same subset. As a result, the overoptimism brought on by time leakage is avoided. Only the augmentation ratio, acceptability threshold, and early stopping make use of the validation set. Before adjusting any hyperparameter, the test set is not changed. The numerical results will be presented directly in the text with the same degree of accuracy as throughout the analysis because this paper does not use tables.

The diffusion model samples in 40 accelerated stages and trains for 100 denoising steps. Each of the condition encoder's two hidden layers has 64 units. Channel attention occurs every two blocks in the denoising network's six dilated temporal blocks. No produced windows are added to the normal class, and accepted generated windows are blended at a 2:1 generated-to-real ratio for the minority class.

The class-wise sample distribution is displayed in Figure 6(a); the breakdown of train-validation-test across operating regimes is shown in Figure 6(b); and baseline diagnostic performance and calibration behavior are compared in Figure 6(c).

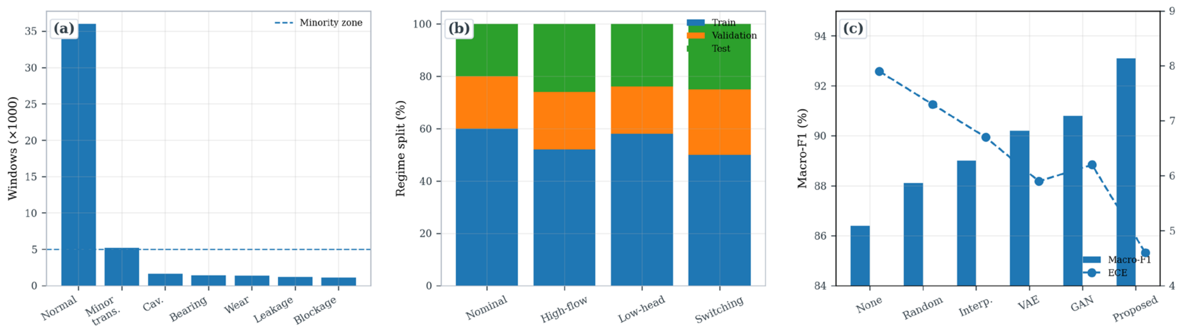

Figure 6. Experimental protocol for evaluating augmentation quality and diagnostic performance: (a) class-wise sample distribution; (b) train-validation-test allocation across operating regimes; (c) baseline diagnostic performance and calibration comparison.

Diagnostic Performance

The suggested diffusion augmentation has the best overall diagnostic performance among the aforementioned. The classifier's macro-F1 score is 86.4% and its accuracy is 92.0% without augmentation. Random oversampling raises the false alarm rate from 3.8% to 4.6% while increasing macro-F1 to 88.1%. By increasing seal leakage recall, interpolation-based oversampling has attained 89.0% macro-F1; however, it remains inadequate in cavitation and obstruction. The adversarial generator reached 90.8% but had greater run-to-run variance than the variational generator, which attained a macro-F1 of 90.2%. The suggested technique has an accuracy of 95.1%, a macro-F1 score of 93.1%, and a false alarm rate of 4.1%.

First, minorities have a high recall rate. When the suggested approach is added, cavitation recall rises from 78.2% without augmentation to 91.6%. The recall for bearing degradation rises from 82.5% to 92.4%, impeller wear from 80.8% to 90.7%, seal leakage from 84.1% to 92.2%, and partial blockage from 77.6% to 90.3%. Because the diffusion model produces different yet regime-consistent adjustments instead of repeatedly employing a small collection of historical windows, the performance improvement is comparatively greater than that attained by random oversampling.

Calibration is improved. The anticipated calibration error is 4.6% during diffusion augmentation and 7.9% in the absence of augmentation. Consequently, it can be said that the classifier has assigned a more appropriate

https://doi.org/10.64972/dea.2024.v3i4.3481d:1-15

probability based on empirical dependability and raised the number of correct classifications. As a result, we may configure the alarm threshold with greater operational confidence. Some generated samples are outliers to the decision boundary because the adversarial baseline sometimes has a strong recall but poor calibration.

Studies on ablation determine what is required. The macro-F1 drops from 93.1% to 91.0% when regime conditioning is removed, with the high-head low-flow regime experiencing the biggest decline. The false alarm rate is 0.052 and the macro-F1 is 0.917 in the absence of a correlation filter. The macro-F1 is 92.0% in the absence of classifier-consistency filtering, however there is more uncertainty regarding cavitation and obstruction. As a result, it has been suggested that the three links—generation, hydraulic environment, and quality screening—be arranged in a closed-loop system.

Hydraulic interpretation is consistent with the confusion patterns. The data has not been increased to rule out partial erosion-related obstruction or transient air entry brought on by cavitation. By including generated windows that encompass intermediate versions of the minority classes, diffusion augmentation lowers both errors. Many problems have similar energy levels, and the majority of the remaining errors happen at moderate pump speeds. The majority of the additional alarms are not random misclassifications, according to additional analysis of the false positives. They happen during startup and shutdown, when a valve opens or closes quickly, and momentarily following a pump switch. Rather than weakening the classifier, persistence logic can tolerate the relatively minor rise in the false alarm rate since these intervals are operationally meaningful even if they are not classified as defects. When the operating regime vector has been stable for a few seconds before to the decision window, the classifier is at its most dependable. A high-uncertainty period is defined as a significantly big change in the regime vector over a brief period of time, which can be used as a supervisory filter.

Robustness and Sensitivity

To assess noise immunity, add zero-mean sensor noise to the test windows at signal-to-noise ratios of 30 dB, 25 dB, and 20 db. The macro-F1 only decreases from 93.1% to 92.5% at 30 db. At 25 db, it is 91.4%. It is 88.9% at 20 dB, which is still better than the 82.7% of the no-augmentation classifier at the same noise level. Because the classifier is exposed to more pressure-flow disturbances during the generated training windows, pressure-driven defects exhibit the greatest progress.

By excluding high-flow night drainage regimes from the training data and utilizing solely test data, operating-regime shift was evaluated. Following this split, the no-augmentation classifier loses 6.8 macro-F1 points. The suggested approach loses 3.1 points. Although conditioning helps change the fault signature in a nearby feasible region and lessens the classifier's brittleness when the head-flow point moves, this does not imply that the generator can produce new hydraulics.

There is a definite optimum for sensitivity to the augmentation ratio. Macro-F1 rose to 90.7% with a 0.5:1 ratio, 92.2% with a 1:1 ratio, and 93.1% with the chosen 2:1 ratio. A 4:1 ratio raises false alarms and reduces the macro-F1 to 0.926. The generated data will therefore be utilized as an additional boundary-enrichment supplement. The classifier starts to model the generated distribution details that are not fully reflected in the field data when synthetic samples are too prevalent.

The other is also necessary for deployment. Recall is marginally increased and more generated samples are accepted when the correlation threshold is raised; however, false alarms increase more quickly than recall when the threshold is more than 25% higher than the chosen value. To get high-quality but fewer varied examples of seal leakage and partial blockage, raise the classifier consistency criterion. As a result, neither the strictest nor the most lenient option is ideal. It is a range that allows the classifier to learn the minority bounds while maintaining the generated windows within hydraulic envelopes.

For offline operation, the generation cost is manageable. For the aforementioned archive, training the diffusion generator requires 2.6 hours on a single mid-range GPU. In eighteen minutes, generate and filter twelve thousand candidate windows. A new window is processed by the online classifier in 4.7 ms on a workstation CPU and 1.2 ms on the testing GPU. As a result, upon training, the suggested approach has no real-time overhead.

Figure 7 displays the initial numerical results. The competing approaches' general diagnostic results are displayed in Figure 7(a), the class-wise minority recall is shown in Figure 7(b), and the robustness to sensor noise and operating-regime change is shown in Figure 7(c). Diffusion augmentation works best when there is sensor

https://doi.org/10.64972/dea.2024.v3i4.3481d:1-15

noise, different operating circumstances, and sparse labelling, as demonstrated in conjunction with the subfigures.

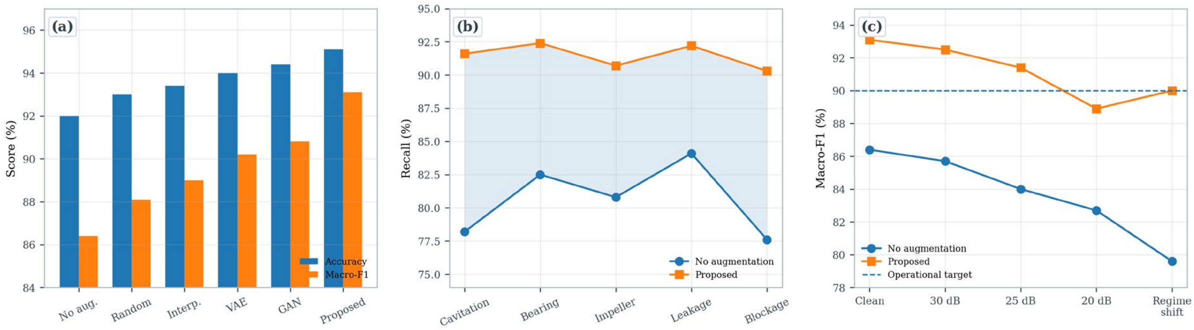

Figure 7. Summary of performance gains and robustness behavior: (a) overall diagnostic score comparison across augmentation methods; (b) class-wise minority-fault recall; (c) robustness under sensor noise and operating-regime shift.

Conclusion

In order to diagnose pump station faults using hydraulic monitoring data, this research presents a data-augmentation diffusion model. The suggested approach presents hydraulic consistency filtering, class-balanced diagnostic training, temporal denoising generation, and regime-aware conditioning. Real-world engineering systems with abundant normal-operation data, few verified fault samples, and varying operating states under different demands and control actions have to be designed using it.

The aforementioned tests show that diffusion-based augmentation can enhance minority defect detection performance without producing an excessive amount of false data. Cavitation recall increased from 78.2% to 91.6%, Macro-F1 increased from 86.4% (without augmentation) to 93.1%, and calibration error decreased from 7.9% to 4.6%. Additionally, ablation tests have demonstrated that generation alone is not appropriate for hydraulic monitoring data; regime conditioning and quality filtering are required. Robustness tests have demonstrated that the suggested approach is also insensitive to head-flow shifts and sensor noise.

There is some use for the suggested method in pump station maintenance and monitoring. Deployment costs are comparatively minimal since the diffusion generator can be trained offline and removed from the online inference process. The classifier offers fault probabilities, saved evidence windows, and channel-level attributions for maintenance; the quality gate will reject synthetic windows that don't fit the hydraulic connection requirements. In subsequent work, the framework will be validated using multi-station field data, other fault types including valve malfunction and pipe leakage will be added, and the generated-sample acceptance will be more closely linked to pump-curve restrictions. The routine implementation of diffusion augmentation in water infrastructure monitoring will be further supported by ongoing recalibration based on the most recent data from normal operations and maintenance.

Author Contributions

Ozren Šuljak contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision, project administration, and funding acquisition. Krešimir Leon Radoš contributes to software, validation, analysis, investigation, data collection, draft preparation, manuscript editing. Aron Popović contributes to software, validation, analysis, investigation, data collection, draft preparation. All authors have read and agreed with the manuscript before its submission and publication.

Funding

This research received no specific financial support from any funding agency.

Institutional Review Board Statement

Not applicable.

References

[1] Yang, Y., Ding, L., Xiao, J., Fang, G., & Li, J. (2022). Current status and applications for hydraulic pump fault diagnosis: A review. Sensors, 22(24), 9714. https://doi.org/10.3390/s22249714

[2] Tiwari, R., Bordoloi, D. J., & Dewangan, A. (2021). Blockage and cavitation detection in centrifugal pumps from dynamic pressure signal using deep learning algorithm. Measurement, 173, 108676. https://doi.org/10.1016/j.measurement.2020.108676

[3] Al-Obaidi, A. R. (2020). Detection of cavitation phenomenon within a centrifugal pump based on vibration analysis technique in both time and frequency domains. Experimental Techniques, 44(3), 329-347. https://doi.org/10.1007/s40799-020-00362-z

[4] He, X., Song, Y., Wu, K., Ali, A., Shen, C., & Si, Q. (2022). Intelligent identification of cavitation state of centrifugal pump based on support vector machine. Energies, 15(23), 8907. https://doi.org/10.3390/en15238907

[5] Li, Z., Jiang, W., Zhang, S., Sun, Y., & Zhang, S. (2021). A hydraulic pump fault diagnosis method based on the modified ensemble empirical mode decomposition and wavelet kernel extreme learning machine methods. Sensors, 21(8), 2599. https://doi.org/10.3390/s21082599

[6] Liu, X., Yang, X., Shao, F., Liu, W., Zhou, F., & Hu, C. (2021). Composite multi-scale basic scale entropy based on CEEMDAN and its application in hydraulic pump fault diagnosis. IEEE Access, 9, 60564-60576. https://doi.org/10.1109/ACCESS.2021.3074498

[7] Casoli, P., Pastori, M., Scolari, F., & Rundo, M. (2019). A vibration signal-based method for fault identification and classification in hydraulic axial piston pumps. Energies, 12(5), 953. https://doi.org/10.3390/en12050953

[8] Bie, F., Du, T., Lyu, F., Pang, M., & Guo, Y. (2021). An integrated approach based on improved CEEMDAN and LSTM deep learning neural network for fault diagnosis of reciprocating pump. IEEE Access, 9, 23301-23310. https://doi.org/10.1109/ACCESS.2021.3056437

[9] Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P., & Gao, R. X. (2019). Deep learning and its applications to machine health monitoring. Mechanical Systems and Signal Processing, 115, 213-237. https://doi.org/10.1016/j.ymssp.2018.05.050

[10] Saufi, S. R., Ahmad, Z. A. B., Leong, M. S., & Lim, M. H. (2019). Challenges and opportunities of deep learning models for machinery fault detection and diagnosis: A review. IEEE Access, 7, 122644-122662. https://doi.org/10.1109/ACCESS.2019.2938227

[11] Zhao, Z., Wu, J., Li, T., Sun, C., Yan, R., & Chen, X. (2021). Challenges and opportunities of AI-enabled monitoring, diagnosis & prognosis: A review. Chinese Journal of Mechanical Engineering, 34, 56. https://doi.org/10.1186/s10033-021-00570-7

[12] Johnson, J. M., & Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. Journal of Big Data, 6, 27. https://doi.org/10.1186/s40537-019-0192-5

[13] Douzas, G., & Bacao, F. (2019). Geometric SMOTE: A geometrically enhanced drop-in replacement for SMOTE. Information Sciences, 501, 118-135. https://doi.org/10.1016/j.ins.2019.06.007

[14] Wen, Q., Sun, L., Yang, F., Song, X., Gao, J., Wang, X., & Xu, H. (2021). Time series data augmentation for deep learning: A survey. Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, 4653-4660. https://doi.org/10.24963/ijcai.2021/631

[15] Shao, S., Wang, P., & Yan, R. (2019). Generative adversarial networks for data augmentation in machine fault diagnosis. Computers in Industry, 106, 85-93. https://doi.org/10.1016/j.compind.2019.01.001

[16] Wang, Y., Sun, G., & Jin, Q. (2020). Imbalanced sample fault diagnosis of rotating machinery using conditional variational auto-encoder generative adversarial network. Applied Soft Computing, 92, 106333. https://doi.org/10.1016/j.asoc.2020.106333

[17] Yang, L., Zhang, Z., Song, Y., Hong, S., Xu, R., Zhao, Y., Zhang, W., Cui, B., & Yang, M.-H. (2024). Diffusion models: A comprehensive survey of methods and applications. ACM Computing Surveys, 56(4), Article 105, 1-39. https://doi.org/10.1145/3626235

[18] Zhao, B., Zhang, X., Li, H., & Yang, Z. (2020). Intelligent fault diagnosis of rolling bearings based on normalized CNN considering data imbalance and variable working conditions. Knowledge-Based Systems, 199, 105971. https://doi.org/10.1016/j.knosys.2020.105971

[19] Montet, F., Pasquier, B., Wolf, B., & Hennebert, J. (2024). Enabling diffusion model for conditioned time series generation. Engineering Proceedings, 68(1), 25. https://doi.org/10.3390/engproc2024068025

[20] Wang, Q., Sun, Z., Zhu, Y., Li, D., & Ma, Y. (2024). A fault diagnosis method based on an improved diffusion model under limited sample conditions. PLOS ONE, 19(9), e0309714. https://doi.org/10.1371/journal.pone.0309714

[21] Perez, E., Strub, F., de Vries, H., Dumoulin, V., & Courville, A. (2018). FiLM: Visual reasoning with a general conditioning layer. Proceedings of the AAAI Conference on Artificial Intelligence, 32(1). https://doi.org/10.1609/aaai.v32i1.11671

[22] Mueller, P. N. (2024). Attention-enhanced conditional-diffusion-based data synthesis for data augmentation in machine fault diagnosis. Engineering Applications of Artificial Intelligence, 131, 107696. https://doi.org/10.1016/j.engappai.2023.107696

[23] Carvalho, T. P., Soares, F. A. A. M. N., Vita, R., Francisco, R. P., Basto, J. P., & Alcalá, S. G. S. (2019). A systematic literature review of machine learning methods applied to predictive maintenance. Computers & Industrial Engineering, 137, 106024. https://doi.org/10.1016/j.cie.2019.106024

[24] Zonta, T., da Costa, C. A., da Rosa Righi, R., de Lima, M. J., da Trindade, E. S., & Li, G. P. (2020). Predictive maintenance in the industry 4.0: A systematic literature review. Computers & Industrial Engineering, 150, 106889. https://doi.org/10.1016/j.cie.2020.106889

[25] Zhou, Z., Chen, X., Li, E., Zeng, L., Luo, K., & Zhang, J. (2019). Edge intelligence: Paving the last mile of artificial intelligence with edge computing. Proceedings of the IEEE, 107(8), 1738-1762. https://doi.org/10.1109/JPROC.2019.2918951

[26] Pang, G., Shen, C., Cao, L., & van den Hengel, A. (2021). Deep learning for anomaly detection: A review. ACM Computing Surveys, 54(2), 1-38. https://doi.org/10.1145/3439950

[27] Deng, S., Zhao, H., Fang, W., Yin, J., Dustdar, S., & Zomaya, A. Y. (2020). Edge intelligence: The confluence of edge computing and artificial intelligence. IEEE Internet of Things Journal, 7(8), 7457-7469. https://doi.org/10.1109/JIOT.2020.2984887

[28] Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daume III, H., & Crawford, K. (2021). Datasheets for datasets. Communications of the ACM, 64(12), 86-92. https://doi.org/10.1145/3458723

[29] Northcutt, C. G., Jiang, L., & Chuang, I. L. (2021). Confident learning: Estimating uncertainty in dataset labels. Journal of Artificial Intelligence Research, 70, 1373-1411. https://doi.org/10.1613/jair.1.12125

[30] Li, W., Huang, R., Li, J., Liao, Y., Chen, Z., He, G., Yan, R., & Gryllias, K. (2022). A perspective survey on deep transfer learning for fault diagnosis in industrial scenarios: Theories, applications and challenges. Mechanical Systems and Signal Processing, 167, 108487. https://doi.org/10.1016/j.ymssp.2021.108487

[31] Abdar, M., Pourpanah, F., Hussain, S., Rezazadegan, D., Liu, L., Ghavamzadeh, M., Fieguth, P. W., Cao, X., Khosravi, A., Acharya, U. R., Makarenkov, V., & Nahavandi, S. (2021). A review of uncertainty quantification in deep learning: Techniques, applications and challenges. Information Fusion, 76, 243-297. https://doi.org/10.1016/j.inffus.2021.05.008

[32] Luan, S., Gu, Z., Saremi, A., Freidovich, L., Jiang, L., & Wan, S. (2023). Timing performance benchmarking of out-of-distribution detection algorithms. Journal of Signal Processing Systems, 95, 1355-1370. https://doi.org/10.1007/s11265-023-01852-0

[33] Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. Proceedings of the Conference on Fairness, Accountability, and Transparency, 220-229. https://doi.org/10.1145/3287560.3287596