SPIDAR: System-level Physics-Informed Detection of Anomalies in Reactors

Ezgi Gursel [1], Bhavya Reddy[2], Benjamin Smith[1], Shahrbanoo Rezaei[1], Katy Daniels[1]
Jamie Baalis Coble[1], Mahboubeh Madadi[2], Vivek Agarwal[3]
Ronald Boring[3], Vaidav Yadav[3], Anahita Khojandi[1] *

[1]University of Tennessee, Knoxville, TN
[2]San Jose State University, San Jose, CA
[3]Idaho National Laboratory, Idaho Falls, ID

ABSTRACT

Nuclear power plants (NPPs) are home to many sensors that can be subject to various anomalies impacting the performance, safety, and reliability of NPPs. These sensor anomalies can arise due to degradation over time or as a result of various external factors. Anomaly detection models can be utilized to detect the presence of sensor errors in NPPs. Specifically, given the physical relationships present between sensed parameters in an NPP, physics-informed machine learning models can be developed to take advantage of the underlying physics of the system, described by physical equations, to ensure the model predictions remain physically consistent. As such, this study proposes SPIDAR: System-level Physics-Informed Detection of Anomalies in Reactors, a novel generative physics-informed system-level anomaly detection model for NPPs to detect sensor errors. Using data collected from a flow loop testbed, this study shows that SPIDAR can successfully detect anomalies present in an array of sensor data. Indeed, this study shows that SPIDAR outperforms state-of-the-art anomaly detection models, specifically a physics-uninformed GAN-based approach, highlighting the potential application of physics-informed machine learning for system-level anomaly detection in NPPs.

Keywords: physics-informed machine learning, anomaly detection, generative adversarial networks, nuclear power plants

1. INTRODUCTION

Nuclear power plants (NPPs) are home to a complex network of sensors that collect critical information regarding the operational status of the plant, from temperature to flow rate. NPPs contain on average over 10,000 sensors and detectors continuously monitoring the status of key plant parameters [1]. However, sensor faults that occur with degregation over time or due to external factors, can lead to faulty readings that may indicate abnormal operating conditions. Common sensor errors in an NPP setting include stuck data, outliers, and drifts [2]. As such, sensor fault detection in nuclear reactors has become the subject of extensive research. While obvious anomalies in the data can often be quickly detected by the trained operator and alarm systems, detecting subtler and gradual anomalies may be more challenging. Hence, real-time system-level anomaly detection is important to maintain the efficiency and the reliable operation of the plant.

Data-driven methods are popular approaches to detecting sensor errors or anomalies in the context of NPPs. These methods make use of available historical data to train models to identify possible faults or anomalies

*khojandi@utk.edu

in the data. Generally, data-driven methods for fault detection can be categorized as supervised learning methods, such as artificial neural networks (ANNs) and k-nearest neighbors, and unsupervised and reinforcement learning methods such as principal component analysis (PCA) [3].

Although data-driven models are common candidates and often suffice for anomaly detection tasks, a purely data-driven approach is hampered by several disadvantages, particularly when dealing with complex systems such as NPPs. Namely, data-driven methods lack generalization of the model outside of the given dataset [4, 5] and the predictions might be inconsistent with physical laws [6]. Furthermore, for complex systems, such as NPPs, it can be difficult, if not near impossible, to build large error-free training datasets that purely data-driven models would require to obtain sufficient performance. Additionally, data-driven methods are commonly regarded as 'black-boxes' due to their lack of interpretability, as it can be difficult to understand why the model has come to that prediction or output [3].

Physics-informed machine learning (PIML) can address these limitations by improving the interpretability of models. The combination of data-driven modeling with physical knowledge of the system can help in better understanding the physical process [7], given that the model is build upon the mathematical knowledge of the system. Additionally, the integration of physical laws into the model can help the model learn more robustly and efficiently [8]. Given the complicated nature of NPPs and the generally known physical relationship between sensed parameters, NPPs are suitable candidates for incorporating physics into ML models.

In this study, we develop a new system-level physics-informed approach for sensor anomaly detection in NPPs. Specifically, we develop SPIDAR or System-level Physics-Informed Detection of Anomalies in Reactors, which is a PIML model that detects the possible presence of anomalies in automatically collected sensor readings. We showcase the performance of SPIDAR using the data from a forced-flow loop testbed. We particularly develop SPIDAR to leverage generative adversarial networks (GANs), given their advantages in generating realistic samples from high-dimensional data [9] and learning unknown probability distributions. To the best of our knowledge, this is the first study that introduces a GAN-based physics-informed anomaly detection approach in the context of NPP sensor anomaly detection.

The rest of the manuscript is organized as follows. Section 2 provides a literature review on fault detection and diagnosis (FDD) in NPPs and PIML. Section 3 describes the data used in the study. Section 4 introduces SPIDAR. Section 5 discusses the results of the study. Finally, Section 6 provides insights from this study, presents possible future research directions, and concludes the manuscript.

2. Literature Review

FDD is a well-established field of research with respect to NPPs. Reviews on the applications and methods of FDD in NPPs can be found in [3, 10, 11]. Recent research has explored the use of different ML methods in FDD. For example, in [12], the authors propose the use of consistency index to detect sensor errors in NPPs using a long-short term memory (LSTM) network. Similarly, [13] uses ANN-based fault detection to detect sensor and actuator faults in nuclear reactors. In another study, [14] uses PCA for sensor FDD.

In the recent years, PIML has emerged as a promising research area with growing applications in reliability and safety [15]. [6] provides an overview of PIML and [8] provides a thorough survey of PIML methods and applications. Additionally, [4] provides a review of physics-informed neural networks (PINNs) in power systems, where the authors also discuss the application of PINNs for anomaly detection and location for high-impedance faults. In [16], the authors propose an information fusion method based on an ensemble of decision trees and SVMs that are trained on a testbed primary coolant system to detect drift. The authors highlight that sensors in an NPP are related through system laws, and the proposed model uses these relationships to create a function that provides estimates of sensor measurements under no drift conditions. Thus, any deviance between the expected and recorded measurements would be the estimate of drift. As an expansion of the aforementioned work, [17] uses PIML methods such as an ensemble of trees and SVM fusers and multi-sensor fusion to learn a regression function for sensor variables and provide an estimate of sensor errors in NPPs. In this study, data are collected from 20 primary coolant loop testbed scenarios (errors that can impact the primary coolant loop), such as calibration changes, minor sensor feed line leaks, air gaps, electromagnetic interference, as well as normal operations. Results of this study show that the root mean squared error (RMSE) of the sensor error estimate is less than 2% of the maximum measurement.

A relatively recent approach is to leverage physics information within GANs. For instance, [18] introduces physics-informed GANs (PI-GANs) that incorporate the physics of the system, in the form of stochastic different equations, into the architecture of a Wasserstein GAN, taking advantage of GAN's ability to learn the unknown terms from the data. Among one advantage of PI-GANs is that they do not suffer from the curse of dimensionality and can make use of multi-group data without any alignment between the data in the groups [18]. One common approach to incorporating physical-knowledge is through the modification of loss function of the ML model [15]. Likewise, in [19], the training of GANs is informed by a physics-informed loss function in the generator. PI-GANs have also been considered in different applications such as traffic flow (TrafficFlowGAN [20]) and fluid mechanics [21] and turbulence analysis [22]. Our study focuses on utilizing the GAN-based method with physics information for the purpose of anomaly detection for NPPs.

3. Data

The dataset used in this study is obtained from a physical testbed at the University of Tennessee, Knoxville, described in [9]. The forced-flow loop testbed includes a primary loop with a controllable 9-kW heater and variable speed pump. The secondary flow rate is controlled through a motor operated valve (MOV). Permanently mounted sensors on this system collect information from resistance temperature detectors (RTDs) about the temperatures (in °C) of the inlet and outlet of the heater region and inlet and outlet of the heat exchanger from the primary and secondary sides. Additionally, the coolant flow of the primary and secondary loops are captured (in milliamps) at a frequency of 0.25 Hz. In addition, three phases of the current and voltage to the variable speed pump and tri-axial acceleration values (in millivolts) of the pump can be recorded, as well as the control signals of the MOV, variable frequency drive, and heater. The readings are captured in intervals of approximately 3 seconds. The dataset used in our study contains approximately 1 hour of readings, and is assumed to come from steady state conditions. A diagram of the flow loop [9] is provided in Figure 5 in the Appendix.

4. Methodology

In this section, we first present the governing physics of the testbed. Next, we discuss our data preprocessing scheme, and introduce SPIDAR in detail. Finally, we present the evaluation metrics.

4.1. Physics of the Testbed

As discussed in Section 3, the sensor dataset collected by the testbed contains readings from resistance temperature detectors (expressed in °C) collected from the inlet and outlet of the heat exchanger on primary and secondary sides and two coolant flow values (expressed in milliamps) from the primary and secondary loops. Given the underlying physics of the testbed, the sensor variables are inherently related through a system of equations. The following calorimetric Equation (1), used to predict the primary coolant flow rate, governs the underlying physics of the testbed system [23]:

$$F_p = \frac{F_s(T_{sout} - T_{sin})}{T_{pin} - T_{pout}}, \tag{1}$$

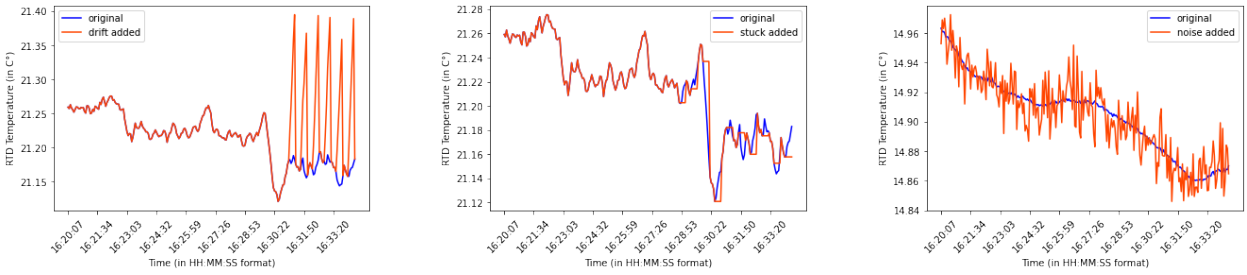
where F_p and F_s denote the primary and secondary flow rates and T_{pout}, T_{pin}, T_{sout} and T_{sin}, represent the outlet and inlet temperatures of the primary and secondary side of the heat exchanger, respectively. Differences between "expected" and measured readings in temperature or flow may be a sign of sensor error and may be a result of a potential issue with the loop system. Note that this equation can be rearranged to predict the values of any of the other sensors used in the equation, given the other temperature and flow readings are available and the system is in steady state.

4.2. Data Preprocessing

First, we reviewed the data and imputed the values of a sensor that was reporting erroneous readings, likely due to a loose sensor. Specifically, the existing readings from this faulty sensor were replaced with its expected values based on Equation (1) introduced in Section 4.1. Note that such approach is common for dealing with incomplete or inaccurate sensor readings [24].

Next, to test the ability of SPIDAR in identifying anomalies, and because our dataset is considered to lack anomalous instances, we artificially modified the normal dataset to introduce anomalies to a random subset of sensor readings. Such approach of adding external anomalies for the purpose of model testing is consistent with recent research works in nuclear engineering applications [12, 17] as well as other safety-critical applications, such as transportation [25]. Specifically, to simulate a real-world NPP environment, we introduced three commonly observed types of anomalies to at most 15% of the normal dataset: Drift, stuck data, and Gaussian noise. To introduce more variability in the anomalous dataset, drift and stuck data were introduced in "windows," where the anomalies start and stop briefly for a period of few time stamps until the duration of the anomaly. For our initial analysis, we test our model on 'single-sensor' anomalies, where one sensor - at most two sensors - is subject to an anomaly. We then consider multi-sensor anomalies, where multiple sensors are subject to random anomalies. We assume that each anomalous sensor are subject to only one type of anomaly, and that the anomalies are introduced to each sensor independently, but it is possible that multiple sensors are subject to different anomalies at the same time steps. Algorithm 1 provides a pseudocode of the anomaly generation process for single-sensor anomaly case. The algorithm for multi-sensor anomaly follows a similar process, except the probability threshold is higher to allow for more sensors to be anomalous. Figures 1a - 1b shows a sample of anomaly added sensors.

The normal dataset is then broken down into training-validation-testing sets, where 80% of the normal dataset is used for training, and 10% is used for validation and testing each. The anomalous dataset is used for the testing of the model. The time-series characteristic of the dataset is preserved in the splitting process. Figure 2 shows the overview of the train-validation-test split. The values are then normalized using min-max normalization and split into temporal windows of length, h = 11. Given that the sensor reading is captured in approximately 3 second intervals, as discussed in Section 3, a temporal window length of 11 allows us to capture any anomalies recorded with an 33-second interval. The length of the temporal window is chosen as an hyperparameter.



(a) Drift added **(b) Stuck added** **(c) Gaussian noise added**

Figure 1: Sample anomaly added temperature (RTD) sensors

4.3. SPIDAR

The model developed in this study incorporates the principles from the GAN-based anomaly detection model by [26] and the physics-based loss function approach from [19]. The architecture has an autoencoder network that acts as the model's generator, an encoder network, and a discriminator. The generator loss in the original GAN-based anomaly detection architecture considers three loss functions: adversarial, contextual, and encoder loss. The adversarial loss considers how successful the generator can distinguish between real and generated images, with the formula given as:

$$\mathcal{L}_{adv} = \mathop{\mathbb{E}}_{x \sim p_x} ||f(x) - \mathop{\mathbb{E}}_{x \sim p_x} f(G(x))||_2.$$

In this formula, p_x denotes the distribution of input data, $f(x)$ is the discriminator's output for the given instance x, and $||\cdot||_2$ denotes the $\mathcal{L}_2$ norm between the discriminator's output for the real instance and the generated instance, $G(x)$ [26].

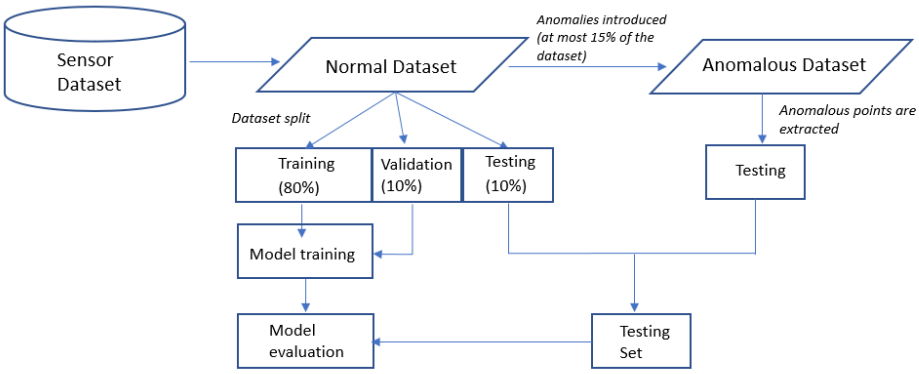


Figure 2: Overview for the train-validation-test split

The contextual loss considers how well the generator can learn contextual information and the formula is given as follows:

$$\mathcal{L}_{con} = \mathop{\mathbb{E}}_{x \sim p_x} ||x - G(x)||_1,$$

where $|| \cdot ||_1$ denotes the $\mathcal{L}_1$ norm.

The encoder loss considers the success of the generator in encoding the features of the generated image:

$$\mathcal{L}_{enc} = \mathop{\mathbb{E}}_{x \sim p_x} ||G_E(x) - E(G(x))||_2, \tag{2}$$

where $G_E(x)$ and $E(G(x))$ denote the feature of the generated instance and the real instance, respectively.

The overall objective for the generator thus becomes the weighted average of all losses, scaled by a corresponding $\lambda_{\text{[loss type]}}$. Thus, the objective for the generator can be described as:

$$\mathcal{L}_{gen} = \lambda_{adv}\mathcal{L}_{adv} + \lambda_{con}\mathcal{L}_{con} + \lambda_{enc}\mathcal{L}_{enc}, \tag{3}$$

where λ represents the hyperparameter of the respective weights for each loss, i.e., how heavily each loss contributes to the generator's objective.

In the SPIDAR approach, in addition to these losses, the generator's loss function is also informed by a fourth physics loss. The physics loss considers the output of the generator to the predicted output of generator for each of the sensor values using Equation (1) and encourages the model to provide physically-consistent results. For example, let $\hat{T}^G_{pin}$ denote the predicted T_{pin} value from the generator given the calorimetric equation and let T^G_{pin} denote the observed T_{pin} output from the generator, T^G_{pin} is compared with the $\hat{T}^G_{pin}$ for the physics loss calculation. For the physics loss, the output of the generator, T^G_{pin}, is penalized given its distance from $\hat{T}^G_{pin}$. The objective of the physics loss is to minimize the mean of the distance between the expected and observed generator's output for each sensor. Ideally, this mean should be zero, or as close to zero, indicating that the generator's output is producing physically-consistent results. The formula for the physics loss can be expressed as:

$$\mathcal{L}_{phy} = \frac{1}{n}\sum_{i=1}^{n}\left|T^G_i - \hat{T}^G_i\right|, \tag{4}$$

where n is the number of sensors, i is the sensor being considered, T_i^G and $\hat{T}_i^G$ denote the observed and predicted output of the generator, respectively.

Thus Equation (3) can be revised as follows:

$$\mathcal{L}_{gen} = \lambda_{adv}\mathcal{L}_{adv} + \lambda_{con}\mathcal{L}_{con} + \lambda_{enc}\mathcal{L}_{enc} + \lambda_{phy}\mathcal{L}_{phy}.$$

Since the physics loss captures the difference between the expected value from the Equation (1) and the measured value from the generator's output, the values inputted in the physics loss need to be in the original scale. Because our values are normalized in the preprocessing stage to optimize the training of the model, the values are rescaled back to original values for the physics loss calculation. The anomaly scores for each test window is found using the encoder loss, Equation (2).

Algorithm 1 Anomaly Generation Process for Sensors (single-sensor anomalous)

procedure ANOMALY GENERATION PROCESS
  for sensor s $\in$ S **do**
    assign probability $U(0, 1)$
    if probability ≤ 0.2 **then**
      generate random anomaly; each type of anomaly equally likely
      for anomaly type: gradual drift **do**
        begin drift at random $t \in T$, drift culminating to 1% of mean of signal for
        drift duration = 5 time stamps, pause drift for 5 time stamps, repeat until the end of drift
      end for
      for anomaly type: Gaussian noise **do**
        Gaussian noise with $\mu = 0, \sigma = \sigma$ of the signal / 2, duration for the length of the signal
      end for
      for anomaly type: stuck data **do**
        data is stuck at a random $t \in T$, stuck duration = 5 time stamps, normal values for 5
        time stamps, until end of anomaly, random duration
      end for
    end if
  end for
end procedure

4.4. Evaluation Metrics

The classification metrics used to evaluate the performance of our model are F1 score, accuracy, geometric mean (g-mean), sensitivity and specificity. These are common evaluation metrics for anomaly detection techniques.

F1 score is a measure of the performance of the binary classifier; accuracy is the total proportion of correctly identified instances. Note that in anomaly detection, accuracy can sometimes be a misleading metric given the imbalanced dataset [27]. G-mean is the geometric mean between between the sensitivity and specificity of a model. G-mean is often considered a better evaluation metric for imbalanced datasets and hence, it is more widely used in anomaly detection tasks [28]. For that reason, our main metric of interest is g-mean:

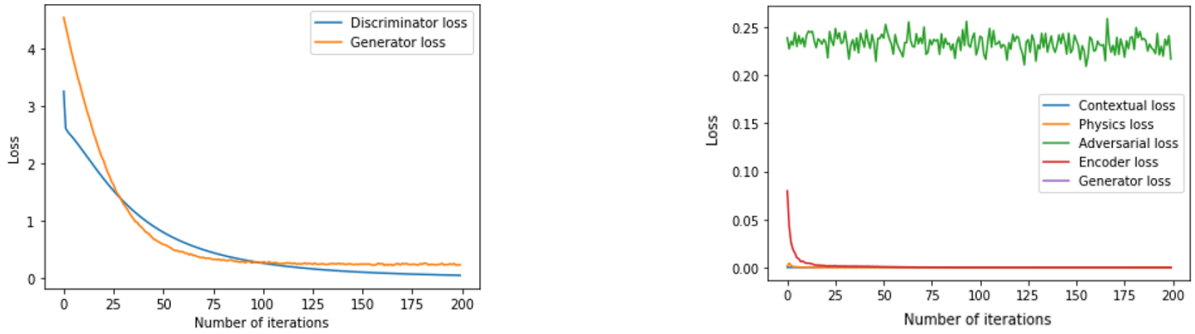
$$\text{G-Mean} = \sqrt{\text{Sensitivity} \cdot \text{Specificity}}.$$

Sensitivity, or recall, is a measure of true positives, or how well the model can correctly distinguishes anomalous instances and specificity is a measure of true negatives. Ideally, a model should have both high sensitivity and high specificity.

5. Results

We first consider the SPIDAR model with the weight of the physics loss, λ_{phy}, set to 20 with the weights of the other losses, $\lambda_{con}, \lambda_{adv}, \lambda_{enc}$, set to 1. We then compare our results to a GAN-based model without

physics losses, in which the weights of contextual, encoder, and adversarial losses are set to 1. The optimal anomaly score threshold is found by generating a normal distribution with the mean and standard deviation of the distribution equal to the median and standard deviation of the anomaly scores, respectively. Classification metrics are calculated for possible thresholds along this distribution, and the value that produces the best classification metrics is used as the anomaly score threshold, where if the anomaly score exceeds the threshold, the test window is marked as anomalous; otherwise it is considered 'normal.' To ensure fair comparison across models, we apply the same anomalies to both models, set the same random state, and verify that the models are sufficiently trained as seen in Figure 3. We repeat this experiment 10 times with different anomalous datasets under the single-sensor and multi-sensor anomaly cases and report the mean and the standard deviation of each classification metric in Table 1. Table 1 presents the results for the



(a) Convergence of generator and discriminator losses

(b) Generator loss and all the corresponding individual contributing losses

Figure 3: Plots of generator and discriminator losses after training with SPIDAR approach

SPIDAR and GAN-based approach when readings from a single sensor and multiple sensors (concurrently) are anomalous, across 10 different test datasets. As seen in the Table 1, we find that the inclusion of a highly-weighted physics loss benefits the performance of the model in correctly categorizing instances as normal and anomalous. For instance, Figure 4 presents an example histogram for the anomaly scores associated with SPIDAR and the GAN-based approach. As seen in this figure, for this example, the distinction between normal and anomalous instances is clear in the scores produced under SPIDAR, whereas those in the GAN-based approach are somewhat overlapping. This allows for a simple threshold to clearly distinguish normal and anomalous instances produced by SPIDAR, resulting in its improved performance over the GAN-based approach. Also, interestingly, under SPIDAR, model performance generally decreases in the single-sensor anomaly case, where anomalies are added to fewer sensors; in contrast, under GAN-based approach which lack a physics loss, the performance increases compared to a multi-sensor case, although SPIDAR still outperforms the GAN-based method under both cases. One possible explanation as to why the GAN-based method performs worse under the multi-sensor case is that it may not be able to handle the complexity of multiple sensors being anomalous and may have trouble learning the distributions of the normal data. On the other hand, SPIDAR can leverage the physical knowledge from the physics loss to better model the data distribution, and thus, is able to perform better, even when multiple sensors are anomalous.

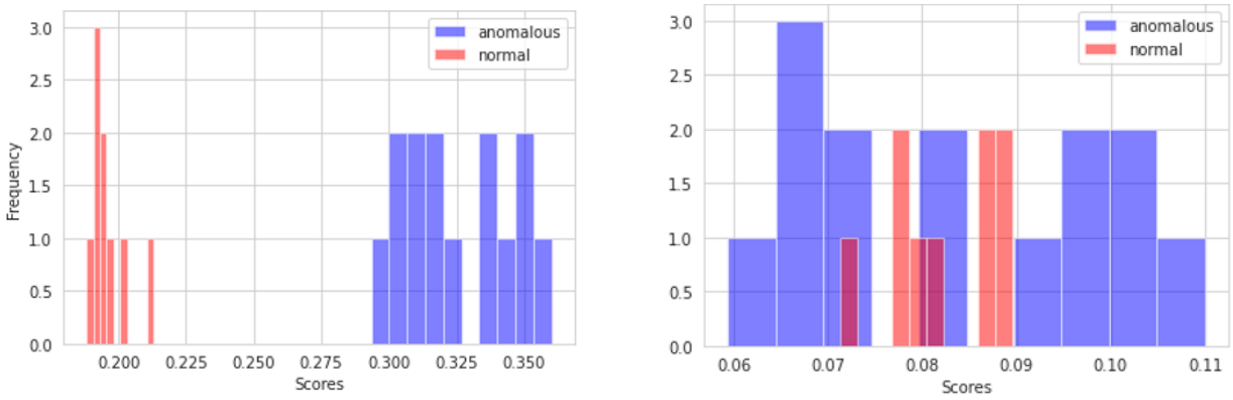
6. CONCLUSIONS

In this study, we propose SPIDAR, a system-level physics-informed anomaly detection approach for sensor anomalies in reactors. We consider the addition of a physics-informed loss to a GAN-based architecture to improve the model performance and encourage the model to provide physically sound results. Our results demonstrate that the inclusion of physics-informed loss provides improved model performance.

There are several future research steps to consider. As aforementioned in Section 4.2, the length h of the temporal window in our model is 11, which allows us to identify whether any sensor recorded an anomaly

Table 1: Performance comparisons (mean ± standard deviation) between SPIDAR and GAN-based approach when readings from single sensor vs. multiple sensors (concurrently) are anomalous, across 10 different test datasets.

	F1 Score	Accuracy	G-Mean	Sensitivity	Specificity
GAN-based (single-sensor)	0.845 ± 0.236	0.831 ± 0.234	0.809 ± 0.252	0.892 ± 0.232	0.762 ± 0.310
SPIDAR (λ_p = 20, single-sensor)	**0.922 ± 0.127**	**0.897 ± 0.167**	**0.883 ± 0.192**	**0.942 ± 0.125**	**0.840 ± 0.259**
GAN-based (multi-sensor)	0.660 ± 0.388	0.679 ± 0.347	0.637 ± 0.405	0.731 ± 0.413	0.657 ± 0.378
SPIDAR (λ_p = 20, multi-sensor)	**0.965 ± 0.082**	**0.959 ± 0.010**	**0.961 ± 0.094**	**0.967 ± 0.105**	**0.957 ± 0.092**



(a) SPIDAR **(b) GAN-based**

Figure 4: Sample histogram plots of scores for normal vs. anomalous instances with SPIDAR approach and GAN-based approach; anomaly scores calculated from $\mathcal{L}_{enc}$, Equation 2

within an approximately 33-second window. Decreasing the length of the window can allow us to consider a shorter time frame and may be more useful for real-time application purposes. Considering the contextual loss instead of the encoder loss in the anomaly score calculation and analyzing reconstruction plots of the generator's output can also allow us to gain insights as to how early the model detects each sensor's anomalies. Additionally, it would be interesting and useful analysis to further test the model across different anomaly probabilities, additional anomaly types, such as bias, and magnitudes and durations to understand how well the model is at classifying each type of anomaly and whether a certain type of anomaly is particularly challenging for anomaly detection purposes. Additional future work also includes expanding the model to include manually collected surveillance readings for a system-level anomaly detection model for NPPs that can detect the presence of both sensor anomalies and human error from the data collection process. This system-level approach could have the potential to facilitate error mitigation and promote safe and efficient NPP operations.

ACKNOWLEDGEMENTS

This work of authorship was prepared as an account of work sponsored by the U.S. Department of Energy, an agency of the U.S. Government, under Grant No. DE-NE0008978 to University of Tennessee-Knoxville. Neither the U.S. Government, nor any agency thereof, nor any of their employees makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights.

REFERENCES

[1] "Instrumentation and Control (I&C) Systems in Nuclear Power Plants: A Time of Transition - NTR2008 Supplement." (2008).

[2] J. W. Hines and R. Seibert. "Technical Review of On-Line Monitoring Techniques for Performance Assessment Volume 3: Limiting Case Studies." (2008).

[3] G. Hu, T. Zhou, and Q. Liu. "Data-driven machine learning for fault detection and diagnosis in nuclear power plants: A review." *Frontiers in Energy Research*, **volume 9**, p. 663296 (2021).

[4] B. Huang and J. Wang. "Applications of Physics-Informed Neural Networks in Power Systems-A Review." *IEEE Transactions on Power Systems* (2022).

[5] B. M. de Silva, J. Callaham, J. Jonker, N. Goebel, J. Klemisch, D. McDonald, N. Hicks, J. N. Kutz, S. L. Brunton, and A. Y. Aravkin. "Physics-informed machine learning for sensor fault detection with flight test data." *arXiv preprint arXiv:200613380* (2020).

[6] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang. "Physics-informed machine learning." *Nature Reviews Physics*, **volume 3**(6), pp. 422–440 (2021).

[7] J. A. Farber and D. G. Cole. "Detecting loss-of-coolant accidents without accident-specific data." *Progress in Nuclear Energy*, **volume 128**, p. 103469 (2020).

[8] Z. Hao, S. Liu, Y. Zhang, C. Ying, Y. Feng, H. Su, and J. Zhu. "Physics-Informed Machine Learning: A Survey on Problems, Methods and Applications." *arXiv preprint arXiv:221108064* (2022).

[9] E. Gursel, B. Reddy, A. Khojandi, M. Madadi, J. B. Coble, V. Agarwal, V. Yadav, and R. Boring. "Using artificial intelligence to detect human errors in nuclear power plants: A case in operation and maintenance." (2022).

[10] J. Ma and J. Jiang. "Applications of fault detection and diagnosis methods in nuclear power plants: A review." *Progress in nuclear energy*, **volume 53**(3), pp. 255–266 (2011).

[11] M. Sethu, B. Kotla, D. Russell, M. Madadi, N. A. Titu, J. B. Coble, R. L. Boring, K. Blache, V. Agarwal, V. Yadav, et al. "Application of Artificial Intelligence in Detection and Mitigation of Human Factor Errors in Nuclear Power Plants: A Review." *Nuclear Technology*, **volume 209**(3), pp. 276–294 (2023).

[12] J. Choi and S. J. Lee. "Consistency index-based sensor fault detection system for nuclear power plant emergency situations using an LSTM network." *Sensors*, **volume 20**(6), p. 1651 (2020).

[13] S. Banerjee, J. Deng, C. Gorse, V. Vajpayee, V. Becerra, and S. Shimjith. "ANN based sensor and actuator fault detection in nuclear reactors." In *2020 8th International Conference on Control, Mechatronics and Automation (ICCMA)*, pp. 88–94. IEEE (2020).

[14] W. Li, M. Peng, and Q. Wang. "Improved PCA method for sensor fault detection and isolation in a nuclear power plant." *Nuclear Engineering and Technology*, **volume 51**(1), pp. 146–154 (2019).

[15] Y. Xu, S. Kohtz, J. Boakye, P. Gardoni, and P. Wang. "Physics-informed machine learning for reliability and systems safety applications: State of the art and challenges." *Reliability Engineering & System Safety*, p. 108900 (2022).

[16] N. S. Rao, C. Greulich, P. Ramuhalli, S. M. Cetiner, and P. Devineni. "Sensor Drift Estimation for Reactor Systems by Fusing Multiple Sensor Measurements." In *2019 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, pp. 1–2. IEEE (2020).

[17] N. S. Rao, C. Greulich, P. Ramuhalli, A. Gurgen, F. Zhang, and S. M. Cetiner. "Estimation of sensor measurement errors in reactor coolant systems using multi-sensor fusion." *Nuclear Engineering and Design*, **volume 375**, p. 111024 (2021).

[18] L. Yang, D. Zhang, and G. E. Karniadakis. "Physics-informed generative adversarial networks for stochastic differential equations." *SIAM Journal on Scientific Computing*, **volume 42**(1), pp. A292–A317 (2020).

[19] Y. Yang and P. Perdikaris. "Adversarial uncertainty quantification in physics-informed neural networks." *Journal of Computational Physics*, **volume 394**, pp. 136–152 (2019).

[20] Z. Mo, Y. Fu, D. Xu, and X. Di. "Trafficflowgan: Physics-informed flow based generative adversarial network for uncertainty quantification." *arXiv preprint arXiv:220609319* (2022).

[21] P. Sharma, W. T. Chung, B. Akoush, and M. Ihme. "A Review of Physics-Informed Machine Learning in Fluid Mechanics." *Energies*, **volume 16**(5), p. 2343 (2023).

[22] M. Bode, M. Gauding, Z. Lian, D. Denker, M. Davidovic, K. Kleinheinz, J. Jitsev, and H. Pitsch. "Using physics-informed enhanced super-resolution generative adversarial networks for subfilter modeling in turbulent reactive flows." *Proceedings of the Combustion Institute*, **volume 38**(2), pp. 2617–2625 (2021).

[23] J. Coble, R. Tarver, and J. W. Hines. "Calorimetric analysis to infer primary circuit flow in integral and pool-type reactors." *IEEE Transactions on Nuclear Science*, **volume 64**(2), pp. 837–843 (2017).

[24] R. N. Faizin, M. Riasetiawan, and A. Ashari. "A review of missing sensor data imputation methods." In *2019 5th International Conference on Science and Technology (ICST)*, volume 1, pp. 1–6. IEEE (2019).

[25] F. Van Wyk, Y. Wang, A. Khojandi, and N. Masoud. "Real-time sensor anomaly detection and identification in automated vehicles." *IEEE Transactions on Intelligent Transportation Systems*, **volume 21**(3), pp. 1264–1276 (2019).

[26] S. Akcay, A. Atapour-Abarghouei, and T. P. Breckon. "Ganomaly: Semi-supervised anomaly detection via adversarial training." In *Asian conference on computer vision*, pp. 622–637. Springer (2018).

[27] C. K. Maurya, D. Toshniwal, and G. V. Venkoparao. "Online anomaly detection via class-imbalance learning." In *2015 Eighth International Conference on Contemporary Computing (IC3)*, pp. 30–35. IEEE (2015).

[28] J. Dahmen and D. J. Cook. "Indirectly supervised anomaly detection of clinically meaningful health events from smart home data." *ACM Transactions on Intelligent Systems and Technology (TIST)*, **volume 12**(2), pp. 1–18 (2021).

APPENDIX A

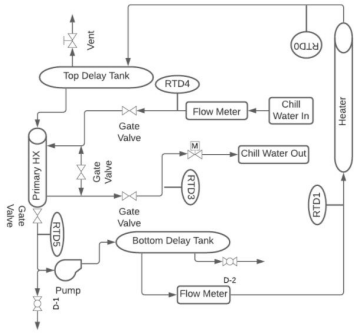


Figure 5: Flow Loop Diagram [9]