HVAC Fouling Detection with Root-Expansion Tiling-based Image Augmentation for Predictive Maintenance (RETINA-PdM)

Nicholas Strzelczyk [1] **Lucas Hartman** [1] **Ethan Pigou** [1] **Santiago Gomez-Rosero** [1] **Miriam A.M. Capretz** [1]

Abstract

Heating, ventilation, and air conditioning (HVAC) systems experience reduced energy efficiency due to fouling on heat exchangers, a problem traditional maintenance struggles to address cost-effectively. While computer vision offers a scalable solution, its effectiveness is limited by the scarcity of labeled data for such rare anomalies. This paper introduces the Root-Expansion Tiling-based Image Augmentation for Predictive Maintenance (RETINA-PdM) algorithm, a method for generating realistic and physics-inspired synthetic data that simulates natural fouling growth. By training convolutional neural networks on data produced by RETINA-PdM, we achieved a predictive F1 score of 0.9642, a significant improvement over previous methods that relied on geometrically simple patterns. This work provides building operators with a highly accurate and cost-efficient tool for predictive maintenance, paving the way for substantial energy savings and optimized building operations.

1. Introduction

The energy efficiency of buildings is heavily impacted by heating, ventilation, and air conditioning (HVAC) systems. A primary cause of inefficiency is fouling, where material accumulates on heat exchanges, which impedes thermal transfer and forces the system to consume more energy. Predictive maintenance (PdM) techniques using computer vision can be used to determine when cleaning is required, but the effectiveness of these techniques depends on large and labeled datasets which are difficult to acquire for rare anomalies like fouling.

Current solutions to reduce fouling on heat exchanges include preventative maintenance (El Marazgioui & El Fadar, 2022), where regular cleaning is done even when it may not be required, and sensor-based PdM which relies on monitoring the symptoms of fouling on HVAC systems using various sensor data. These symptoms include, but are not limited to, a decline in overall heat transfer efficiency (Ogbonnaya & Ajayi, 2017), reductions in fluid mass flow rates (Ogbonnaya & Ajayi, 2017), anomalies in commonly logged process variables (Sundar et al., 2020), and shifts in terminal temperature differences (Bobde et al., 2022). While preventative maintenance is effective at preventing large amount of fouling, it is inefficient as the maintenance may be done even when it is not strictly necessary. The sensor-based PdM is more efficient than preventative maintenance, but relies on expensive sensors to be installed and faces training data scarcity issues because fouling is rare.

A more robust solution to PdM is to use computer vision models, as they can operate on more common and easy-to-install hardware like cameras. While computer vision techniques do not require expensive sensors to operate, they still struggle with the data scarcity issue, necessitating the need for synthetic training data to represent the rare fouling patterns. The work of Strzelczyk et al. (Strzelczyk et al., 2024) demonstrated that a convolutional neural network (CNN) trained on synthetic data could effectively detect fouling, but this work was limited by its reliance on generating geometrically simple fouling spots. The uniform patterns used in this work did not capture the irregular and complex nature of real-world fouling, which limited the model's predictive accuracy. This work achieved a moderately high predictive F1 score of 0.8595, demonstrating the effectiveness of the approach despite the relatively simple synthetic data generation process.

This paper presents a new synthetic data generation algorithm called Root-Expansion Tiling-based Image Augmentation for Predictive Maintenance (RETINA-PdM) as its central contribution. It does this by simulating a more naturalistic foul-growth pattern within the synthetic data. This result represents a significant advancement in developing robust, data-driven PdM solutions for HVAC efficiency. By providing building operators a robust, highly accurate, and cost-efficient way to monitor fouling in their HVAC units,

[1] Department of Electrical and Computer Engineering, Western University, London, Ontario, N6A 5B9, Canada. Correspondence to: Nicholas Strzelczyk <nstrzelc@uwo.ca>.

Proceedings of the 42^{nd} International Conference on Machine Learning, Vancouver, Canada. PMLR 267, 2025.

this work has substantial and practical implications for the computational optimization of buildings.

The remainder of this paper is organized as follows: Section 2 examines the existing literature. Section 3 describes how RETINA-PdM operates. Section 4 presents the results, discusses the findings, and highlights the approach's limitations. Finally, Section 5 concludes the paper.

2. Related Works

Collecting and annotating real-world data is both costly and labor-intensive, which has encouraged a shift toward synthetic datasets. Existing research (Man & Chahl, 2022) highlights the advantage of data synthesis for automating data collection and labeling as it significantly reduces associated costs. They also warn about biases which may be induced by synthetic data due to limited variability. Previous research (Dankar & Ibrahim, 2021) reviews synthetic data generators primarily aimed at numerical and tabular data, while other work (Tian et al., 2023) extended these concepts specifically to predictive maintenance scenarios involving time-series data.

To support early fouling detection and preventative maintenance in industrial systems, recent work has leveraged synthetic data generation to overcome the chronic scarcity of labeled fault imagery. A GAN-based augmentation framework blends GAN-generated defect patches onto pristine images to dramatically expand training sets for visual fault classifiers (Jain et al., 2022). A digital-twin finite-element model simulates sub-surface defect heat signatures, interpolating them onto a pixel grid to generate synthetic thermograms for U-Net segmentation pre-training (Pareek et al., 2025). A deep convolutional GAN trained on limited borescope blade images fuses generated defects onto clean backgrounds to fine-tune object detection models for aircraft engine inspections (Schaller et al., 2025). Finally, a self-attention DC-GAN with a spatial–content attention discriminator synthesizes X-ray weld defect images, which, when combined with real data, enable semantic segmentation models to excel at pressure-vessel weld defect detection (Wang et al., 2024).

Physics-informed synthetic data generation approaches have also been developed to further enhance realism by integrating domain-specific physical models. Computational fluid dynamics simulations have generated datasets for defect detection and improved predictive maintenance pipelines (Lakshmanan et al., 2023). Integrating physics-driven models with generative adversarial networks has also shown potential for structural health monitoring applications (Luleci et al., 2022). Despite these advancements, a notable gap remains in applying these approaches to fouling detection in heat exchangers using image-based synthetic datasets. Unlike defect detection in industrial pipelines, fouling accumulation presents a dynamic challenge that requires specialized data generation techniques to reflect real-world conditions.

To the best of our knowledge, no other datasets based on images for heat exchangers exist. To address this gap, our work proposes a synthetic data generation algorithm tailored specifically for predictive maintenance in heat exchangers. Our approach uniquely integrates a random event-driven time-series simulation, realistic image synthesis based on physical models and real-world data, and automated annotation, which directly confronts the distinct complexities in fouling detection tasks.

3. Methodology

Figure 1 illustrates the RETINA-PdM methodology, comprising four core components: data collection, data generation, model training, and model evaluation.

3.1. Data Collection

Image data of the heat exchanger grid is collected using a 5G RTSP-enabled camera mounted at a fixed angle facing the cooling tower. The camera operates from 08:00 to 20:00, capturing one image per hour for 13 hours per day, seven days a week. The data collection period spans from July through August. Data from non-operational months and nighttime hours are excluded due to limited visibility and inactivity of the system. A Python service on an Ubuntu server manages image capture, naming, and storage. This structured raw dataset forms the foundation for the synthetic data generation process.

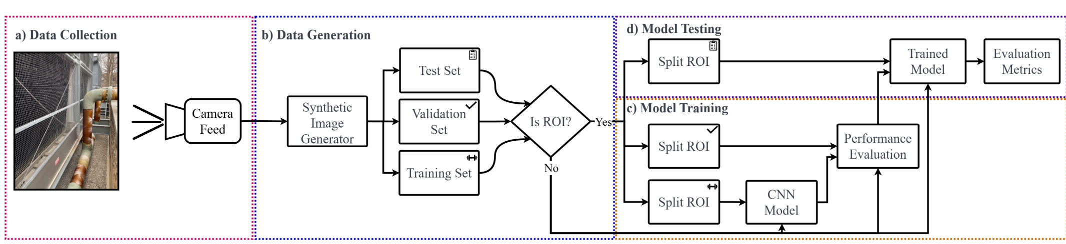

Figure 1. RETINA-PdM methodology for predictive maintenance.

3.2. Data Generation

We use root-expansion tiling (RET) to control fouling growth. The RET algorithm is shown in Algorithm 1. Starting from a small 20×36 pixel fouling centroid point, the generator expands the pattern outward in random directions, ensuring natural-looking spread and variation. Multiple fouling regions may overlap and grow into one another. Fouling intensity is modeled using a decreased intensity proportional to a tile's distance from the centroid, simulating higher central density and gradual fade. The generator pastes these fouling textures onto a black canvas of the same size as the target image (1920×1080), respecting the specified coordinates. To create the synthetic data, the generator blends these fouling textures into the real heat exchanger images so that the fouling realistically presents itself behind the metal mesh of the heat exchanger as shown in Figure 2.

To produce the label mask, all pixels representing tiles containing fouling are overwritten with the integer representing their class. Both binary and multiclass labels are produced separately. The multiclass labels take into account the various fouling intensities while binary labels simply label all fouling tiles as 1, else 0. To create the sample images, the generator blends the fouling textures into the real heat exchanger images using OTSU thresholding (Otsu, 1979). This results in the synthetic fouling realistically presenting itself behind the metal mesh of the heat exchanger. Each resulting (image, mask) pair represents a training sample. This process generates hundreds of labeled images per scenario while maintaining consistency with real-world texture and lighting.

Algorithm 1 Synthetic Fouling Generator

```
Input: numDays
image ← BlankImage()
centroids ← RandomTiles()
for d ← 1 to numDays do
    for tile ∈ centroids do
        for i ← 1 to RandomInt() do
            while tile ≠ ∅ do
                tile ← MoveRandomDirection()
            end while
            image[tile] ← ApplyFouling()
        end for
    end for
    for tile ∈ image do
        if tile has empty neighbor then
            image[tile] ← image[tile] · 0.25
        else if tile has empty neighbor 1 tile away then
            image[tile] ← image[tile] · 0.50
        end if
    end for
end for
```

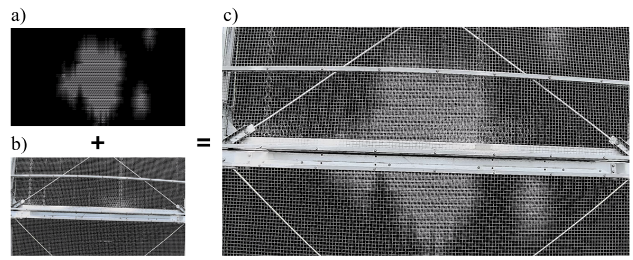

Figure 2. How the generated foul growth pattern (a) is combined with the raw heat exchange image (b) to create the synthetic data of fouling (c).

3.3. Model Training

We train two CNNs for the fouling detection task: U-Net (Ronneberger et al., 2015) and CGNet (Wu et al., 2021), both selected for their proven effectiveness in semantic segmentation. Each model is trained on the synthetic dataset generated by RETINA-PdM, where the input images contain simulated fouling and the corresponding masks serve as ground truth. Models are optimized using binary cross-entropy loss, and early stopping is applied based on validation performance. This setup enables a comparative evaluation of CNN architectures for pixel-wise fouling detection. Training was completed using two techniques: regions-of-interest (RoI) and full-image. The RoI technique splits each image into a 4x5 grid and passes each slice individually to the model. The full-image technique processes the entire image all at once. These techniques are compared for effectiveness.

3.4. Model Evaluation

We evaluate model performance on two test datasets not seen during training. The first test set mirrors training scenarios, while the second introduces scheduled maintenance patterns. Both test sets contain 756 images. These datasets assess both generalization and robustness to realistic variability.

Evaluation metrics include the F1 score, which balances precision and recall, quantifying the model's effectiveness in segmenting fouling under practical conditions, supporting its use in predictive maintenance systems.

4. Results

This section presents the experimental results obtained from the RETINA-PdM framework, including the configuration of fouling scenarios, quantitative evaluations from two experiments, a discussion of model performance, and an analysis of the technique's current limitations.

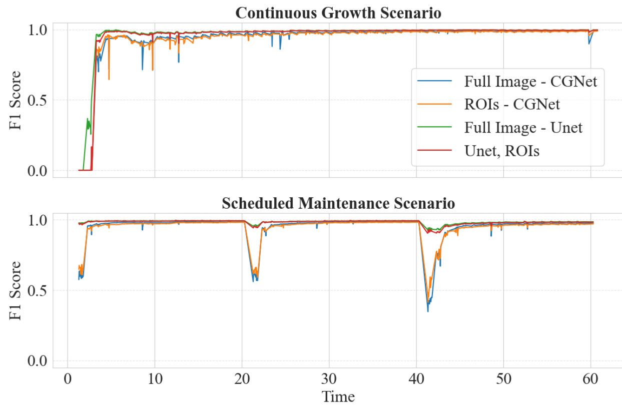

Figure 3. F_1 score over time across both scenario types.

The raw images were collected from a heat exchanger used in an HVAC system located at Western University. This project is part of the Campus as a Living Lab initiative. The model training was conducted using the binary cross-entropy loss function and the AdamW optimizer (Loshchilov & Hutter, 2017), with all training sessions running for 25 epochs. For reproducibility, each experiment was conducted using three different seeds and the results were averaged.

4.1. Scenario Configuration

To evaluate the different CNNs performance under different conditions, two types of scenarios were configured: a continuous growth scenario (CGS) and a scheduled maintenance scenario (SMS). The CGS simulates uninterrupted fouling accumulation over a 60-day period, producing 13 synthetic images per day corresponding to hourly intervals between 8:00 AM and 8:00 PM. Each image is paired with a corresponding label mask, and three CGS datasets were generated for training, validation, and testing. In contrast, the SMS incorporates a more realistic predictive maintenance (PdM) setting by introducing periodic maintenance every 20 days, as described in (Patil et al., 2022; Strzelczyk et al., 2024). On maintenance days, fouling growth is reset and resumes at newly randomized locations. Three SMS datasets were generated to form a second test set used exclusively for evaluating the model's ability to generalize to maintenance-influenced scenarios.

4.2. Experiment 1: Binary Classification of Fouling Presence

Experiment 1 evaluates the UNet's and CGnet's ability to perform binary classification on full heat exchanger images, determining whether any fouling is present or not. This experiment uses both CGS and SMS scenarios, providing insight into model performance under both idealized and more realistic predictive maintenance conditions. Each image is labeled with a binary class indicating the presence or absence of fouling. Figure 3 shows the performance of CNN models over both scenarios. This experiment resulted in a mean F1 score of 0.9458 and 0.9642 for the CGS and SMS test sets, respectively.

4.3. Experiment 2: Multiclass Fouling Severity Classification

Experiment 2 focuses on evaluating the model's ability to detect fouling severity using the SMS scenario, which better reflects realistic maintenance conditions. Unlike the binary classification in Experiment 1, this experiment uses a multiclass setting with four classes: clean metal surface, and fouling at 25%, 50%, and 100% obstruction levels. The goal is to determine whether incorporating varying degrees of fouling severity and explicitly modeling the metal structure improves the model's understanding and robustness. Figure 4 shows the F1 scores and behavior for the multi-classes on experiment 2. The average training time for CGNet was 23.08 minutes, while UNet required 88.35 minutes, highlighting CGNet's efficiency in training time and learning multiclass fouling patterns

4.4. Discussion

The effectiveness of the proposed approach was evaluated through two experiments. The first experiment focused on evaluating the ability of UNet and CGNet to perform binary classification on both CGS and SMS scenarios. Each model was trained and tested using both the RoI technique and the full-image approach, resulting in four total combinations. The results plotted in Figure 3 show that UNet generally yielded higher F1 scores across the entirety of both test sets. CGNet struggled noticeably more following maintenance occurrences in the SMS test set. This suggests that UNet is more consistent at performing binary classification of fouling, especially in more realistic contexts such as SMS. Both architectures performed well overall, with F1 scores that were consistently greater than 0.9.

The second experiment focused on evaluating the ability of both architectures to perform multiclass segmentation on the SMS test set. The four classes to be segmented were clean metal surface, and fouling at 25%, 50%, and 100% intensity. Figure 4 shows how each architectural combination performed on this task. Contrary to the results of the first experiment, CGNet yielded similar, and in some cases, better performance than UNet. Both variations of CGNet also show a clear reduction in F1 score standard deviation. Additionally, CGNet's significantly shorter training and inference time support the conclusion that it outperforms UNet in the task of multiclass fouling segmentation.

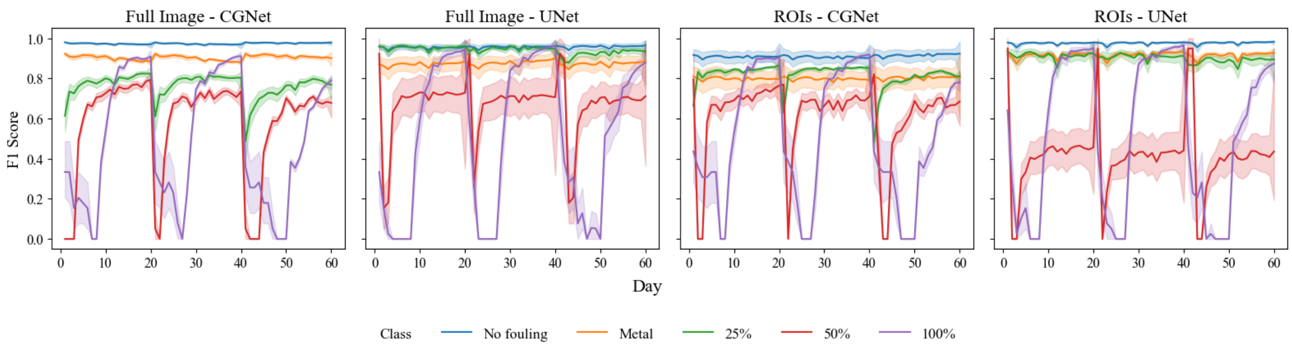

Figure 4. F_1 score across each level of fouling over time by model and input type for the scheduled maintenance scenario.

4.5. Limitations

While the RETINA-PdM framework provides a scalable and controllable approach for training fouling detection models, several limitations remain. First, the experimental validation currently relies entirely on synthetic data. The aim is to complement this with expert-labeled datasets and real IoT sensor data in future work to ensure practical relevance and reliability. Second, although RETINA-PdM performs robustly under a range of lighting and weather conditions, the current generator struggles to blend synthetic fouling effectively into rainy images, as well as those captured from oblique camera angles. Extending the framework to handle these scenarios would improve its realism and applicability.

5. Conclusion

This paper addressed the challenge of detecting fouling in HVAC heat exchangers, a task made more difficult due to the scarcity of labeled data. We introduced RETINA-PdM, a physics-inspired synthetic data generation algorithm that simulates realistic and irregular fouling growth patterns. By training CNNs on our synthetic data, we achieved a predictive F1 score of 0.9642, a significant leap from the 0.8595 F1 score of previous methods that used geometrically simple patterns. Our results demonstrate that the CGNet model, in particular, provides a highly accurate and computationally efficient solution. The success of RETINA-PdM offers building operators a powerful, data-driven, and cost-effective tool for predictive maintenance, paving the way for substantial energy savings and optimized building operations by enabling maintenance to be performed precisely when needed.

Future work will focus on expanding baseline comparisons to further validate the effectiveness of RETINA-PdM. In addition to synthetic image augmentation, alternative data generation strategies such as physics-based simulations, mathematical modeling, and generative adversarial networks could be explored to enhance realism and diversity. As no publicly available validation sets currently exist for this specific problem, establishing a shared benchmark would be a valuable contribution to the field. Lastly, while U-Net and CGNet provide strong baselines, both were originally designed for different tasks. Developing a task-specific architecture tailored for fouling detection could yield improvements in both computational efficiency and segmentation quality.

Acknowledgments

This work was supported by the Climate Action and Awareness Fund (CAAF) from Environment and Climate Change Canada, and the Ontario Research Fund – Research Excellence Grant - ORF-RE 10-045. This research has also been supported by the Vector Scholarship in Artificial Intelligence, provided through the Vector Institute, and supported by the Canadian Graduate Scholarship (Masters), awarded by the Government of Canada. S.G.-R. thanks the SENESCYT from Ecuador for the support received during his doctoral studies in Canada.

Impact Statement

This paper presents work whose goal is to advance the field of Machine Learning. There are many potential societal consequences of our work, specifically in the area of synthetic data generation for predictive maintenance. The introduced RETINA-PdM algorithm aims to address data scarcity issues in detecting HVAC fouling, promoting energy efficiency, sustainability, and reduced operational costs.

References

Bobde, P. B., Jadhav, T. V., Juwale, S. D., and Sawant, D. H. Prediction of Heat Exchanger Fouling Time. *International Journal of Advances in Engineering and Management (IJAEM)*, 4:845, 2022. doi: 10.35629/

5252-0405845849.

Dankar, F. K. and Ibrahim, M. Fake It Till You Make It: Guidelines for Effective Synthetic Data Generation. *Applied Sciences*, 11(5):2158, 2 2021. ISSN 2076-3417. doi: 10.3390/app11052158.

El Marazgioui, S. and El Fadar, A. Impact of cooling tower technology on performance and cost-effectiveness of CSP plants. *Energy Conversion and Management*, 258:115448, 4 2022. ISSN 01968904. doi: 10.1016/j.enconman.2022.115448.

Jain, S., Seth, G., Paruthi, A., Soni, U., and Kumar, G. Synthetic data augmentation for surface defect detection and classification using deep learning. *Journal of Intelligent Manufacturing*, 33(4):1007–1020, 4 2022. ISSN 0956-5515. doi: 10.1007/s10845-020-01710-x.

Lakshmanan, K., Tessicini, F., Gil, A. J., and Auricchio, F. A fault prognosis strategy for an external gear pump using Machine Learning algorithms and synthetic data generation methods. *Applied Mathematical Modelling*, 123:348–372, 11 2023. ISSN 0307904X. doi: 10.1016/j.apm.2023.07.001.

Loshchilov, I. and Hutter, F. Decoupled Weight Decay Regularization. *International Conference on Learning Representations*, 11 2017.

Luleci, F., Catbas, F. N., and Avci, O. Generative Adversarial Networks for Data Generation in Structural Health Monitoring. *Frontiers in Built Environment*, 8, 2 2022. ISSN 2297-3362. doi: 10.3389/fbuil.2022.816644.

Man, K. and Chahl, J. A Review of Synthetic Image Data and Its Use in Computer Vision. *Journal of Imaging*, 8(11):310, 11 2022. ISSN 2313-433X. doi: 10.3390/jimaging8110310.

Ogbonnaya, S. K. and Ajayi, O. O. Fouling phenomenon and its effect on heat exchanger: A review. *Frontiers in Heat and Mass Transfer*, 9, 2017. ISSN 21518629. doi: 10.5098/HMT.9.31.

Otsu, N. A threshold selection method from gray-level histograms. *IEEE Trans SMC*, 9:62–66, 1 1979.

Pareek, K. A., May, D., Meszmer, P., Ras, M. A., and Wunderle, B. Synthetic data generation using finite element method to pre-train an image segmentation model for defect detection using infrared thermography. *Journal of Intelligent Manufacturing*, 36(3):1879–1905, 3 2025. ISSN 0956-5515. doi: 10.1007/s10845-024-02326-1.

Patil, P., Srinivasan, B., and Srinivasan, R. Monitoring fouling in heat exchangers under temperature control based on excess thermal and hydraulic loads. *Chemical Engineering Research and Design*, 181:41–54, 5 2022. ISSN 02638762. doi: 10.1016/j.cherd.2022.02.032.

Ronneberger, O., Fischer, P., and Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Navab, N., Hornegger, J., Wells, W. M., and Frangi, A. F. (eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pp. 234–241, Cham, 2015. Springer International Publishing. ISBN 978-3-319-24574-4.

Schaller, T., Li, J., and Jenkins, K. W. A Data-Driven Approach for Automatic Aircraft Engine Borescope Inspection Defect Detection Using Computer Vision and Deep Learning. *Journal of Experimental and Theoretical Analyses*, 3(1):4, 2 2025. ISSN 2813-4648. doi: 10.3390/jeta3010004.

Strzelczyk, N., Gomez-Rosero, S., and Capretz, M. A. Predictive Maintenance Using Fouling Detection on Heat Exchangers. In *2024 34th International Conference on Collaborative Advances in Software and COmputiNg (CASCON)*, pp. 1–5. IEEE, 11 2024. ISBN 979-8-3315-0483-0. doi: 10.1109/CASCON62161.2024.10838159.

Sundar, S., Rajagopal, M. C., Zhao, H., Kuntumalla, G., Meng, Y., Chang, H. C., Shao, C., Ferreira, P., Miljkovic, N., Sinha, S., and Salapaka, S. Fouling modeling and prediction approach for heat exchangers using deep learning. *International Journal of Heat and Mass Transfer*, 159:120112, 10 2020. ISSN 0017-9310. doi: 10.1016/J.IJHEATMASSTRANSFER.2020.120112.

Tian, R., Gomez-Rosero, S., and Capretz, M. A. M. Health Prognostics Classification with Autoencoders for Predictive Maintenance of HVAC Systems. *Energies*, 16(20):7094, 10 2023. ISSN 1996-1073. doi: 10.3390/en16207094.

Wang, X., He, F., and Huang, X. A new method for deep learning detection of defects in X-ray images of pressure vessel welds. *Scientific Reports*, 14(1):6312, 3 2024. ISSN 2045-2322. doi: 10.1038/s41598-024-56794-9.

Wu, T., Tang, S., Zhang, R., Cao, J., and Zhang, Y. CGNet: A Light-Weight Context Guided Network for Semantic Segmentation. *IEEE Transactions on Image Processing*, 30:1169–1179, 2021. ISSN 1057-7149. doi: 10.1109/TIP.2020.3042065.