DETECTING INDUSTRIAL FOULING BY MONOTONICITY DURING ULTRASONIC CLEANING

Chang Rajani[1,3], *Arto Klami*[1], *Ari Salmi*[2], *Timo Rauhala*[3], *Edward Hæggström*[2], *Petri Myllymäki*[1]

[1]Dept. of Computer Science, Helsinki Institute for Information Technology HIIT, University of Helsinki
[2]Electronics Research Laboratory, Dept. of Physics, University of Helsinki, [3]Altum Technologies

ABSTRACT

High power ultrasound permits non-invasive cleaning of industrial equipment, but to make such cleaning systems energy efficient, one needs to recognize when the structure has been sufficiently cleaned without using invasive diagnostic tools. This can be done using ultrasound reflections generated inside the structure. This inverse modeling problem cannot be solved by forward modeling for irregular and complex structures, and it is difficult to tackle also with machine learning since human-annotated labels are hard get. We provide a deep learning solution that relies on the physical properties of the cleaning process. We rely on the fact that the amount of fouling is reduced as we clean more. Using this monotonicity property as indirect supervision we develop a semi-supervised model for detecting when the equipment has been cleaned.

Index Terms— ultrasound, convolutional neural network, fouling detection, industrial cleaning

1. INTRODUCTION

Biological, chemical, and mineral fouling of industrial production equipment costs billions of dollars every year and has a major negative impact on the environment. Fouling and clogging wastes energy, reduces production yield, and may break equipment. Heat exchanger fouling alone may contribute to 2.5% of global CO_2 emissions [1], while reducing the global GDP by 0.25%, i.e. circa $195 billion [2].

Traditional methods to clean industrial heat exchangers and pipes are mechanical (sand/soda/dry ice blasting, high pressure water cleaning, brushing) or chemical in nature. These methods, which require halting the production and dismantling the clogged equipment, can harm both production equipment and the environment. Non-invasive cleaning achieved by transmitting focused high-power ultrasound from transducers placed on the outer surface of the industrial equipment avoids these issues, and has been demonstrated in practice by Altum Technologies.

Naive application of such cleaning technology employs high sonic power for an overly long time to ensure that the equipment is properly cleaned. This approach wastes energy.
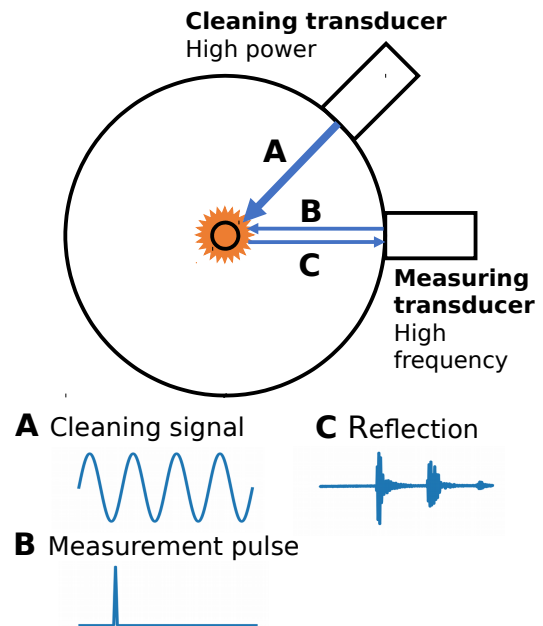
Fig. 1. Diagram visualizing the real-time cleaning and detection setup. The fouling (orange material) is on the external surface of an internal pipe within the closed equipment. We use high-power ultrasound (A) to clean the equipment while intermittently probing for the current status using high-frequency measurement pulse (B) and its echoes (C). A convolutional neural network trained in a semi-supervised fashion determines when the fouling has been removed.

An alternative is to non-invasively detect the residual amount of fouling, and to automatically stop cleaning when the structure is clean. Internal reflections of ultrasound detected outside of the equipment can in principle be used for this. Due to irregular and unknown properties of the fouling and due to the potentially complex internal structure inside the equipment, traditional techniques that use information about the specific properties of the setup are not applicable. We present a practical solution to the described problem, illustrated in Figure 1. The solution employs deep learning to detect, in real time, the fouling status based on reflections of high-frequency ul-

trasound diagnostic pulses.

From the perspective of machine learning the task is to detect the amount of fouling inside the structure based on ultrasonic signals measured from outside the structure. The input signals are 100 microseconds long reflections of high frequency ultrasound and the simplest predictive formulation is to predict, in real time, a binary class (fouled or clean) based on one individual signal. Unfortunately, there is never labeled training data available to teach such a classifier since the whole procedure is non-invasive.

We solve the problem with a classifier that learns to recognize when complete removal of fouling has been achieved without requiring individual annotations. It merely assumes that the cleaning process is *monotonous* – during cleaning the amount of fouling can only decrease – and that we have access to a collection of sufficiently long cleaning runs to guarantee that the structure is clean at the end of the process.

We formulate the solution as an extension of *structured pseudolabels* proposed by Longi *et al.* for semi-supervised classification of streaming data with temporally scarce labels [3]. The key idea of pseudo-labels [4] is to alternate between (a) learning the parameters of a model given current label estimates and (b) learning the label estimates given the current model parameters. We extend this framework by proposing two *monotonicity supervision* strategies that require the pseudo-labels to change monotonously during training.

To motivate the work we first present the cleaning apparatus and recapitulate the core properties of non-invasive ultrasonic monitoring in Section 2, before presenting the machine learning contributions of a convolutional neural network architecture in Section 3 and the monotonicity supervision principle in Section 4. Finally, in Section 5, we use the proposed technique on a laboratory equipment that allows detection of the ground truth for evaluation purposes. We demonstrate the solution in two separate tasks: (a) Recognize *post hoc* when the structure was cleaned based on measurement data from one cleaning run alone, (b) Detect when all fouling has been removed in real-time during a cleaning run, using model trained on earlier cleaning runs.

2. ULTRASONIC FOULING DETECTION

We perform fouling detection while we clean using technology provided by Altum Technologies. The setup (Figure 1) consists of an acrylic container with cleaning transducers mounted on the external surface. The setup features a single small acrylic pipe attached coaxially in the center of the external cylinder. The fouling is on the external surface of this inner pipe, and hence the setup represents a simplified version of a container with internal structure that accumulates fouling (e.g. a heat exchanger). The separate transducer on the outside of the container is used in pulse-echo mode to detect fouling. Each measurement pulse is interleaved with a cleaning pulse to avoid interference between the ultrasonic cleaning signal and the probing signal.

The standard approach for non-invasive imaging with ultrasound is based on direct physical modeling of the wave propagation under ideal conditions. For example, in our setup the two echoes detected in the signals (see Figure 3) arise from the two pipe-water interfaces of the internal structure. The acoustic impedance at the boundary of the interfaces changes when fouling is removed. Based on the properties of the measurement beam, the materials, and the physical dimensions of the equipment we could derive the expected reflections inside a clean structure and compare those to the measurement signal to determine the fouling state. Wallhäußer et al. [5] used such a technique to detect fouling. They used features such as the sonic time-of-flight calculated by hand for a specific setup and used a neural network to predict whether fouling is present. As this approach requires intimate knowledge about the container and type of fouling, it is impractical for use in a commercial setting.

For complex internal structures we cannot analyze the reflections on macroscopic level, but we can simulate the propagation of the measurement pulse using forward simulation techniques such as the Finite Element Method (FEM). FEM represents the acoustic waves using a fine discrete grid and simulates temporal progression explicitly, but the small wavelength of ultrasound requires an extremely fine numeric grid. Consequently, the simulation, especially when performed in 3D, quickly reaches hundreds of millions of degrees of freedom and becomes intractable. For complex structures the simulation time might reach thousands of hours, and naturally would require detailed knowledge of the internal structure.

In contrast to these approaches, we make no assumptions about the setup. Instead, our model simply relies on the fundamental differences between raw signals recorded in clean and fouled pipes (see Figure 3). It employs no characteristics of the structure or of the measurement equipment as input, is purely data-driven, and solves the problem solely based on measurement data collected during real cleaning runs.

3. CONVOLUTIONAL NEURAL NETWORK TO DETECT INDUSTRIAL FOULING

Convolutional neural networks (CNNs) have been successful in modeling data with spatial or temporal structure, such as images, video, and audio. Specifically in signal processing, where inputs are pressure-modulation signals represented by real vectors $x \in \mathbb{R}^D$, we can employ temporal (1-D) convolution to filter along the time-axis of the signal. Similarly the other main building block of CNNs, the pooling operation, is applied in the time dimension to encode temporal invariance. These kinds of models have successfully been applied in image recognition [6], to model text as one-dimensional series of characters or words [7], and in speech recognition [8].

Typical CNNs feature stacks of convolutional layers, pooling, and non-linearities. In principle any such structure

could be used to predict the fouling status based on the ultrasound probing signal, but it pays off to select the model structure to match the defining properties of the input signals. We can assume a few characteristics to hold, without making assumptions about the specific structure being cleaned:

1. The main signal corresponds to echoes of the measurement pulse from material interfaces inside the structure

2. The specific time coordinate of the echoes depends primarily on the structure, not on fouling

3. The fouling changes the reflection in unknown manner, possibly influencing both amplitude and the waveform

4. The measurement needs to be carried out with high sampling rate and is hence high-dimensional and sparse

For such signals we need models that base the decisions on temporally local events, attempting to detect the amplitude and waveform changes caused by fouling, while intentionally ignoring specific timing of the echoes that is primarily determined by the structure itself. No assumptions are made about the nature of the changes, which is captured by data-driven convolutional filters. The high dimensionality of the input together with a relatively small training data sets suggests using models of low complexity.

Building on these intuitions we designed the model illustrated in Figure 2. The input is a single $x \in \mathbb{R}^{2500}$ vector that is convolved using sixteen learned filters (amount of filters chosen by grid search cross-validation) of size $w = 5$. Each filter is followed by the rectified linear unit (ReLU). The pooling layer uses *1-max pooling*, which reduces the entire set of filter activations into a single scalar per filter that captures the highest possible activation. This pooling achieves perfect location invariance and was suggested originally for audio event recognition by Phan *et al.* [9]. Finally, the outputs of the 1-max pooling are converted into scalars $z_n \in [0, 1]$ that represent the probability of fouling at a certain time point by a dense layer with sigmoid activation.

Even though the proposed model borrows technical elements from [9], the full models are different. Their model was developed for audio event recognition based on spectrogram image features (SIF) computed for short overlapping windows of acoustic signals, whereas we use the 1-max pooling as part of a CNN operating on raw ultrasound signals.

This kind of model would be easy to train on labeled training data using gradient descent and backpropagation. We could then simply minimize the binary cross-entropy loss

$$\mathcal{L} = -\frac{1}{N_l} \sum_{n=1}^{N_l} \left[y_n \log(z_n) + (1 - y_n) \log(1 - z_n) \right], \quad (1)$$

where $y_n \in \{0, 1\}$ is the true label for sample x_n, and N_l is the amount of labeled training samples. However, since labeled training data is unavailable, we next present a training method that requires no direct supervision.
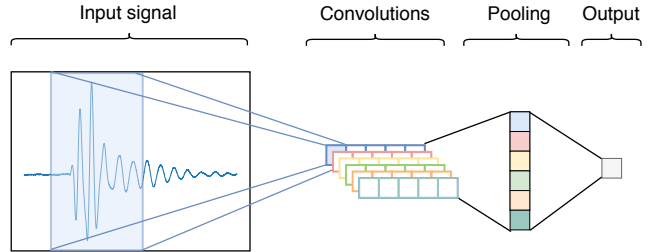
Fig. 2. CNN architecture employed to detect industrial fouling. The model features merely a few narrow-width temporal filters, along with 1-max pooling that makes it robust against temporal variance. A single real value is produced as output: the amount of fouling detected at a certain time point.

4. SUPERVISION BY MONOTONOCITY

Since we cannot monitor the amount of fouling during ongoing cleaning even in laboratory conditions, we cannot obtain annotated labels for individual samples. To overcome this challenge, we supervise the learning process by a global physical property: during the cleaning process the amount of fouling can only decrease.

Besides the monotonicity property we need to make two mild assumptions that can be guaranteed in laboratory conditions: (a) At the beginning of the process there is some fouling, and (b) the cleaning process is ran until the fouling is completely removed. Building on these three elements we create an essentially unsupervised solution for learning to detect when the clean state is reached.

The basic idea is to fix the labels of the first k and last k samples in the experiment to the fouled and clean states, respectively, treating the remaining samples as unsupervised examples[1]. For a single cleaning run we obtain $N_l = 2k$ labeled examples and $N_u \gg N_l$ unlabeled examples. We then adopt the *pseudo-label* approach [4] to solve this semi-supervised learning problem, by alternating between

1. Optimize the network parameters θ given current labels

2. Assign pseudo-labels for the unsupervised samples

Standard pseudo-labeling assigns the labels independently, providing for each unsupervised sample a label that matches the current network outputs according to a suitable criterion. To utilize the monotonicity property of our experimental structure we build on the *structured pseudo-label* concept [3], and replace the independent assignments with a global assignment of all labels with structural constraints.

We propose two novel instances of structured pseudo-labels to enforce the monotonicity constraint. The *label monotonicity* variant builds on the assumption that the real state of the system can only change once, from fouled to

---

[1]In practice $k = 1$ is enough when using monotonicity supervision, but for completeness we provide the equations for this more general case.

clean. It assigns binary pseudo-labels $\hat{y}_n$ for every unlabeled sample and enforces label monotonicity by requiring $\hat{y}_n = 1$ for all $n \leq t$ and $\hat{y}_n = 0$ for all $n > t$ for some threshold time t. The optimal threshold is obtained by maximizing the logarithmic likelihood of the pseudo-labels given the above constraint, using

$$\arg\max_t \left[\sum_{n=1}^{t} \log z_n + \sum_{n=t+1}^{N_u} \log(1 - z_n) \right], \quad (2)$$

where z_n is the current estimate for the probability of having a fouled class. The optimal solution is found by iterating over all possible thresholds in linear time.

The *probability monotonicity* variant assumes that the probability that the structure is fouled decreases monotonously with time. Instead of assigning binary pseudo-labels it uses continuous pseudo-labels $\hat{z}_n$ for every unlabeled sample. The labels are determined by solving the optimization problem

$$\arg\min_{\hat{\mathbf{z}}} \sum_{n=1}^{N_u} (z_n - \hat{z}_n)^2 \quad (3)$$
$$\text{s.t.} \quad \hat{z}_n - \hat{z}_{n-1} \leq 0 \quad \forall n \in [2, \dots, N_u],$$

where we use squared error to measure the deviance from the current model output z_n. The problem matches that of *isotonic regression* [10] and hence we use the pool-adjacent-violators algorithm [11] to find the optimal piece-wise monotonous decreasing function that minimizes the error. Isotonic regression is typically used for *post-hoc* probabilistic calibration of classifier outputs, but here it is used for iterative pseudo-labeling during optimization of the model.

A pseudo-code for training the model with monotonicity supervision is provided in Algorithm 1. Irrespective of the monotonicity supervision variant, the pseudo-label assignment step is guaranteed to find the globally optimal solution given the current network parameters. The network parameters θ are updated by gradient steps minimizing the loss that combines (1) for the $N_l = 2k$ labeled samples additively with a separate loss for the N_u pseudo-labeled samples. For label monotonicity the loss term is

$$-\frac{1}{N_u} \sum_{n=1}^{N_u} \left[\hat{y}_n \log(z_n) + (1 - \hat{y}_n) \log(1 - z_n) \right], \quad (4)$$

and for probability monotonicity it is

$$\frac{1}{N_u} \sum_{n=1}^{N_u} (z_n - \hat{z}_n)^2. \quad (5)$$

5. EXPERIMENTS

To demonstrate the proposed convolutional neural network for fouling detection, we conducted two computational experiments using the same empirical data. The first experiment

---

Algorithm 1 Monotonicity supervision for neural networks

1: Initialize network parameters θ for randomly
2: Train network on $N_l = 2k$ labeled instances
3: Predict labels $\hat{z}$ for all N_u unlabeled samples
4: **for** i in iters **do**
5:     **if** Label monotonicity **then**
6:        Set $\hat{\mathbf{y}}$ by solving (2)
7:        Update θ to improve (1) plus (4)
8:     **end if**
9:     **if** Probability monotonicity **then**
10:       Set $\hat{\mathbf{z}}$ by solving (3)
11:       Update θ to improve (1) plus (5)
12:   **end if**
13: **end for**

---

shows the model can be used to analyze a single cleaning run after completion to detect when the structure was clean. The second experiment shows how a model trained on previous cleaning runs provides real-time detection of cleaning.

The main goals of these experiments were to (a) show that the proposed model architecture can solve the fouling detection problem with high accuracy, and (b) that the monotonicity supervision techniques improve the accuracy. We do this by comparing three model variants: (1) Baseline CNN that uses our proposed architecture but only uses the labeled samples (first and last k samples), (2) Label monotonicity CNN that assigns binary pseudo-labels to all remaining samples, and (3) Probability monotonicity CNN that assigns continuous pseudo-labels to all remaining samples.

5.1. Experimental setup and data

We carried out seven cleaning runs using the setup illustrated in Figure 1. At the beginning of each run the pipe at the center of the structure was fouled and cleaning was carried out until it was clean. One run was discarded because the amount of fouling was insufficient to begin with – it did not conform to our constraint of going from fouled to clean. The total duration of each experiment varied, with average duration being close to 1.6 minutes corresponding to on average 950 measurements that were captured at roughly 0.1 sec intervals.

Each captured measurement signal contains 10,000 samples that spans 100 μs of the signal starting 150 μs after the measurement pulse. Within such a window we see both the reflection from the front surface of the internal pipe and the reflection from the back surface, see Figure 3. In this laboratory setup the second reflection is sufficiently dampened by the fouling that we can use it to construct ground truth labels for evaluation by simple thresholding, whereas the first reflection (2,500 time points starting from 175 μs after the measurement pulse) is used as input for our model. It is important to note that the second reflection is seen only because we use an acrylic inner structure – for real production equip-
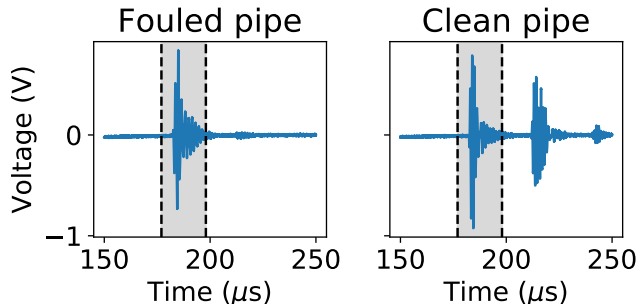
Fig. 3. Example measurements for fouled (left) and clean (right) structure. The shaded area indicates the part of the signal used as model input – the model does not see the second surface echo, which we use for evaluation.

ment the second echo would be absent.

5.2. Post-hoc fouling detection

The first scenario considers the problem of determining, *post hoc*, the time point at which the surface of the internal structure becomes clean. From a modeling perspective this is a challenging setup because of the scarcity of labeled data and the result is already useful *e.g.* in providing heuristics for the duration of future cleaning runs.

We solved the problem for all six cleaning runs independently, and report in Figure 4 (top) the average probability of the true class (fouled/clean), $\frac{1}{N_u}\left[y_n z_n + (1-y_n)(1-z_n)\right]$ for each cleaning run. All three methods solve the problem with relatively high accuracy already for $k = 1$, and having access to more labeled data helps but not dramatically.

Both pseudo-label variants outperform the baseline model, but not by a wide margin. The bottom subplot illustrates the behavior of the probability monotonicity supervision, and demonstrates that the pseudo-labels $\hat{z}$ smooth out the erroneous predictions of roughly 20% occurring after the structure has already been cleaned – even though this property has only a marginal effect on the overall accuracy, the monotonic predictions are qualitatively better.

5.3. Real-time fouling detection

A more valuable use-case is to train the fouling detector on historical data, so that it can be applied during future cleaning runs to determine the status in real time. We achieve this by training on five historical cleaning runs, allocating the pseudo-labels separately for each run during the training, and by evaluating the final model on the remaining run.

Figure 5 reports the average accuracy when leaving each of the six cleaning runs as evaluation data. Here the semi-supervised models trained using monotonic pseudo-labels are more accurate than the baseline model that only uses the end
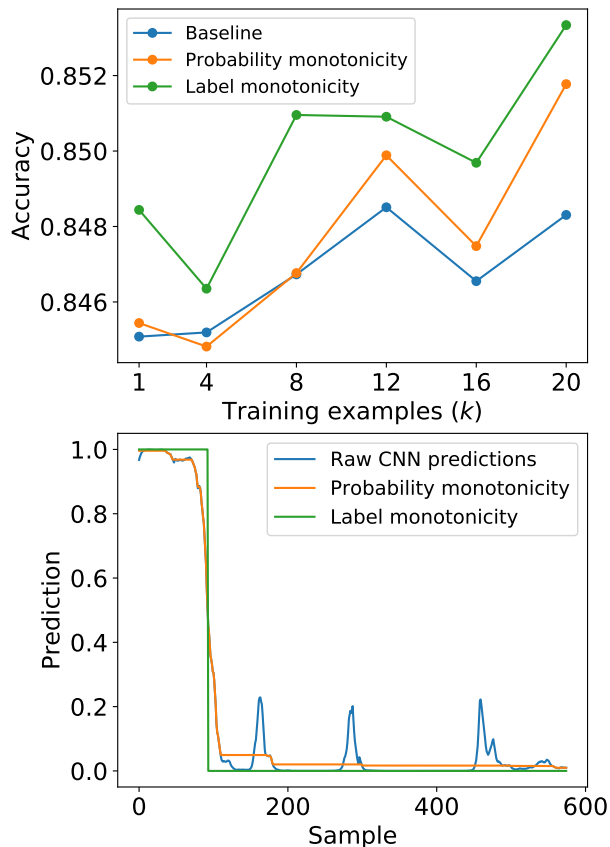

Fig. 4. **Top:** Accuracy (mean probability of correct label) of *post hoc* fouling detection during a cleaning run. **Bottom:** Illustration of the monotonicity enforcement step of Algorithm 1. This shows the effect of applying isotonic regression (orange) and binarization (green) to the raw predictions (blue).

points, with 15% relative error reduction in average probability, and already $k = 1$ is sufficient for optimal accuracy. The average deviation between the predicted and true time of cleaning was for the best variants roughly 10 seconds.

We also compare the methods on two example runs. For one of the runs the pseudo-label variants are better than the baseline that fails to detect the true clean state. The other example demonstrates an important property – when the baseline method already solves the problem, the monotonicity supervision does not hurt but merely returns the same solution.

6. DISCUSSION

We presented a machine learning solution for detecting when a closed structure being cleaned non-invasively has been fully cleaned. It learns to separate clean and fouled structures using a convolutional neural network applied to signals from non-invasive ultrasonic probing. It achieves this without us-
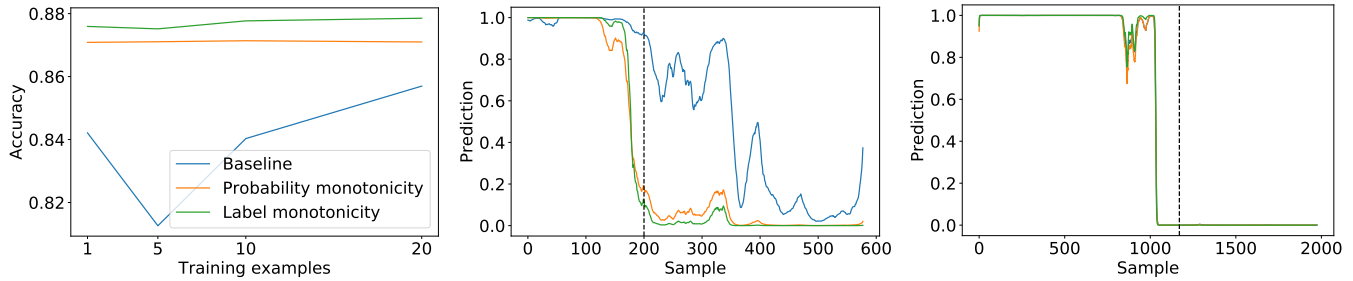
Fig. 5. **Left:** Both pseudolabel techniques outperform the baseline in the real-time detection scenario. **Middle and right:** Probability of fouling as function of time for two example cleaning runs, trained on 20 labeled examples. The horizontal axis denotes the signals sampled during cleaning, and the vertical line marks the ground truth based on the back surface echo.

ing explicit labels for the individual sensor signals, by only assuming that the cleaning proceeds monotonously – fouling can only be removed during the process. Our solution builds on existing audio event recognition models [9] and structured pseudo-labels [3], and extends the pseudo-label technique to work with a monotonously changing process in two ways: by providing pseudo-labels (1) directly for the cleaning state and (2) for the probability that the structure is clean.

We demonstrated the solution in practice on a laboratory equipment, detecting when simple internal structure has been fully cleaned within 10 seconds (on average) of the true transition. Even though the experimental setup here is one for which a physics-based solution would be possible, we showed that a labeling-free machine learning solution agnostic of the underlying physics can recognize the fouling status with an accuracy sufficient for practical use. This makes development of commercially applicable energy-efficient ultrasonic cleaning devices feasible, effectively minimizing energy consumed by overly long cleaning runs.

7. REFERENCES

[1] H Müller-Steinhagen, MR Malayeri, and AP Watkinson, "Heat exchanger fouling and cleaning," *Heat Transfer Engineering*, vol. 28, no. 3, pp. 173–176, 2011.

[2] H Müller-Steinhagen, MR Malayeri, and AP Watkinson, "Fouling of heat exchangers-new approaches to solve an old problem," *Heat transfer engineering*, vol. 26, no. 1, pp. 1–4, 2005.

[3] K Longi, T Pulkkinen, and A Klami, "Semi-supervised convolutional neural networks for identifying wi-fi interference sources," in *Proceedings of the 9th Asian Conference on Machine Learning*, 2017, vol. 77 of *Proceedings of Machine Learning Research*, pp. 391–406.

[4] "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks," .

[5] E Wallhäußer, Walid B H, M A Hussein, J Hinrichs, and T M Becker, "On the usage of acoustic properties combined with an artificial neural network–a new approach of determining presence of dairy fouling," *Journal of Food Engineering*, vol. 103, no. 4, pp. 449–456, 2011.

[6] A Krizhevsky, I Sutskever, and G E Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[7] N Kalchbrenner, E Grefenstette, and P Blunsom, "A convolutional neural network for modelling sentences," in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, 2014, vol. 1, pp. 655–665.

[8] O Abdel-Hamid, A-R Mohamed, H Jiang, and G Penn, "Applying convolutional neural networks concepts to hybrid NN-HMM model for speech recognition," in *Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2012, pp. 4277–4280.

[9] H Phan, L Hertel, M Maass, and A Mertins, "Robust audio event recognition with 1-max pooling convolutional neural networks," *Interspeech 2016*, pp. 3653–3657, 2016.

[10] B Zadrozny and C Elkan, "Transforming classifier scores into accurate multiclass probability estimates," in *Proceedings of the 8th ACM SIGKDD international conference on knowledge discovery and data mining*, 2002, pp. 694–699.

[11] M Ayer, H D Brunk, G M Ewing, W T Reid, and E Silverman, "An empirical distribution function for sampling with incomplete information," *Annals of Mathematical Statistics*, vol. 26, pp. 641–647, 1955.