

Unsupervised Machine Learning for Automated Corrosion Staging Using Optical Microscopy Images

Ashwin RajKumar^{1†}, Naveen Paluru^{1†}, Raji Susan Mathew²,
Prathamesh Shenai³, Dana Abdeen⁴, Nicholas Laycock⁴,
Phaneendra Yalavarthy^{1*}

^{1*}Department of Computational and Data Sciences, Indian Institute of Science, Bengaluru, 560012, India.

²School of Data Science, Indian Institute of Science Education and Research, Thiruvananthapuram, 695551, India.

³Shell Technology Centre, Shell India Markets Private Limited, KIADB Industrial Park, 562149, India.

⁴Qatar Shell Research and Technology Centre, Qatar Shell GTL Ltd, Qatar Science and Technology Park, Doha, 24750, Qatar.

*Corresponding author(s). E-mail(s): yalavarthy@iisc.ac.in;
Contributing authors: ashwinrajcum@iisc.ac.in; naveenp@iisc.ac.in;
rajisusanmathew@iisertvm.ac.in; prathamesh.shenai@shell.com;
dana.abdeen@shell.com; nick.laycock@shell.com;

[†]These authors contributed equally to this work.

Abstract

Corrosion poses a substantial economic burden, and machine learning is increasingly being explored for its potential in staging, predictive maintenance, and data-driven decision making. This study presents an unsupervised automated corrosion staging method based on image processing and machine learning using optical microscopy (OM) images. It detects and computes (i) the local porosity in a neighborhood of $5\mu\text{m} \times 5\mu\text{m}$ at pore locations, and (ii) the deposit thickness in (μm). The local porosity and deposit thickness were used to estimate the chloride concentration factor, associated pH, and the corrosion stage. The approach was tested on 48 ex-service OM images of under-deposit corrosion (UDC). A thickness-based approach yielded an accuracy of $\sim 73\%$ in classifying the corrosion stage in UDC compared to previous time-consuming approaches.

This is a significant step in automating the evaluation of the corrosion stage, enabling scalable data-driven corrosion assessment across critical infrastructures.

Keywords: Image Analysis, Machine Learning, Corrosion Staging, Porosity, Deposit Thickness, Optical Microscopy, Digitalization

1 Introduction

Corrosion is a widespread and interdisciplinary phenomenon that results in the degradation of metals, posing significant economic and safety challenges across various industries. It undermines the reliability, efficiency, and longevity of structural and functional components in critical sectors such as power generation, oil and gas, and infrastructure. Traditional methods for analyzing corrosion—such as visual inspection, potentiodynamic polarization, and electrochemical impedance spectroscopy—offer valuable insights into the overall corrosion process, particularly in materials like carbon steel [1–4]. However, they fall short in revealing detailed surface characteristics like porosity, which play a vital role in the initiation and progression of localized corrosion.

Pores on metal surfaces act as entry points for corrosive agents, significantly accelerating degradation through increased surface area exposure and the formation of micro-environments conducive to corrosion. Studies have shown that pore size, distribution, and inter-connectivity influence the kinetics of corrosion reactions [5]. Therefore, precise characterization of porosity is essential for predictive corrosion modeling. Advanced imaging techniques such as optical microscopy (OM), X-ray computed tomography, and scanning electron microscopy have made it possible to correlate surface microstructure with corrosion susceptibility [6]. In industrial applications, such insights facilitate more effective preventive maintenance strategies [7], especially in complex systems like steam generators and pipelines.

The advent of digital monitoring and machine learning (ML) offers new opportunities for automating corrosion analysis, reducing reliance on subjective manual inspection. ML enables the analysis of large, complex datasets and the discovery of hidden patterns, making it an ideal tool for predicting corrosion rates, characterizing degradation, and optimizing mitigation strategies [8, 9]. While supervised ML approaches have shown promise in predicting corrosion inhibitor performance and material degradation trends [10–12], they are often limited by the need for labeled datasets and the risk of overfitting [13]. Unsupervised ML, by contrast, circumvents these limitations by identifying intrinsic data patterns without requiring prior annotations. This makes it especially valuable in corrosion science, where labeled data are scarce and experimental conditions are variable [14].

Recent studies have begun to explore the role of unsupervised ML in materials science, such as in the classification of aluminum alloys [15] and quantification of nanopores in oxide films [16]. However, its application to corrosion characterization remains limited. In particular, under-deposit corrosion (UDC)—a localized corrosion form prevalent in high-pressure steam generators—has not been extensively studied

through unsupervised learning. UDC arises due to the accumulation of magnetite deposits that trap chlorides and other ionic contaminants beneath porous surface layers [1, 17–19]. These deposits alter local chemistry, such as pH and chloride concentration, and drive aggressive corrosion processes that are difficult to monitor using conventional methods [20–25].

Common industrial practices for mitigating UDC include reducing chloride concentration in steam generator water, monitoring deposit accumulation, and performing chemical cleaning before critical thresholds are reached [26–28]. However, controlling chlorine levels is often infeasible in large-scale systems, where trace chloride concentrations in the parts-per-billion range are intentionally maintained to promote uniform magnetite layer formation that protects steam generator tubes [29]. As a result, precise and continuous monitoring of deposit characteristics becomes critical to balance protective deposition with the risk of corrosion. Experimental systems for simulating magnetite deposits have been developed [30], yet their real-world durability and behavior remain uncertain. Imaging and quantification tools such as *ImageJ* [31] and *MATLAB* [32] have been applied to assess corrosion features at high temperatures. In this study, *MATLAB* was selected for image processing due to its compatibility with machine learning workflows, although future adaptations of the algorithm could also be developed as plugins for *ImageJ* to support wider community use.

Manual classification of UDC stages based on optical microscopy images has been proposed to address this issue, defining four corrosion stages based on deposit layering and porosity [33, 34]. However, this approach is labor-intensive, prone to interobserver variability, and challenging to scale across industrial operations. Automated image analysis, especially when combined with unsupervised ML techniques, offers a more robust, scalable, and objective method for staging corrosion and enabling predictive maintenance. Parameters such as surface texture, deposit thickness, and porosity—readily extracted from OM images—can serve as inputs for ML algorithms that estimate the extent and severity of corrosion [33, 35].

Recent studies on the progression of under-deposit corrosion (UDC) using optical microscopy have proposed a staging system based on the morphological features of corrosion deposits, particularly porosity and layer structure [33, 34]. This classification allows for more effective preventive maintenance and monitoring in pipeline systems. The UDC stages, as described in [33], are visually interpreted as follows:

- Stage 1 involves an unfractured scale with a thin protective magnetite layer and a nonporous barrier adjacent to the steel surface;
- Stage 2 is marked by a porous layer forming on the magnetite, resulting in a double-layer structure with pores concentrated on the outer surface;
- Stage 3 is characterized by the presence of multilayer corrosion product scales; and
- Stage 4 exhibits multilayer corrosion products with visible fractures between layers.

In this study, we present an unsupervised, image-based methodology for the automated staging of under-deposit corrosion. The proposed system computes two key surface parameters—local porosity and deposit thickness—from optical microscopy images of ex-service steam generator samples. These parameters are further used to estimate the chloride concentration factor and associated pH, both critical indicators

of corrosion severity. Our findings demonstrate that deposit thickness and porosity not only reflect the underlying chemical environment but also correlate well with expert-defined corrosion stages. Furthermore, we propose a fully automated staging algorithm based on these metrics, advancing the digitalization of corrosion diagnostics in industrial settings.

The main contributions of this study are as follows:

- Development and validation of an automated algorithm for thickness measurement and porosity computation, which shows good agreement with manual measurements and domain expertise calculations.
- Analysis of chloride concentration factor and pH as a function of deposit thickness, revealing a correlation between increasing deposit thickness, higher chloride concentration, and decreasing pH.
- Demonstration of the relationship between the progression of UDC and the increase in the chloride concentration factor along with the decrease in pH values.
- Quantification of pH values for different stages of under-deposit corrosion (UDC), particularly for stages 3 and 4, with stage 3 showing minimum pH values between 2.8 and 3.5, and stage 4 between 1 and 1.5.
- Identification of a potential threshold for transition from stage 3 to stage 4 UDC, based on pH values between 2.8 and 3.
- Finally, a fully automated algorithm for automated corrosion staging based on OM images was developed using the computed deposit thickness and local porosity from the unsupervised machine learning method.

These findings contribute to a better understanding of the UDC process and provide quantitative data for predicting and monitoring corrosion stages in industrial applications.

2 Results

All implementations were performed using *MATLAB* on a Linux workstation with an Intel i9-9900X, 3.3GHz, 10 cores processor and 128 GB of RAM. Fig. 1 shows the input OM image with the calibrated thickness obtained using Eq.(8), the local porosity map obtained using Eqs. (6), the local porosity as a function of thickness obtained using Eq.(7) and the attributes of the sample OM images were considered. As expected, the porosity values were lower on the metal side (x) than on the oxide interface (y). The local porosity map estimated using the proposed method is directly correlated with the visual assessment of porosity throughout the width of the tube deposit. Further, representative images from each stage is shown in Figure 2 along with the calibrated thickness obtained using the proposed algorithm.

Figure 3 shows the confusion matrices corresponding to (a) porosity, (b) local porosity, (c) thickness, and (d) UDC stage classifications based on the voting method. These results corresponded to the 48 sample OM images used by Abitha et al. [33]. The true stage of the UDC corresponds to the actual stage of the UDC determined by domain expertise [33]. The diagonal elements in each matrix represent the number of correctly classified samples and the nondiagonal elements represent the number

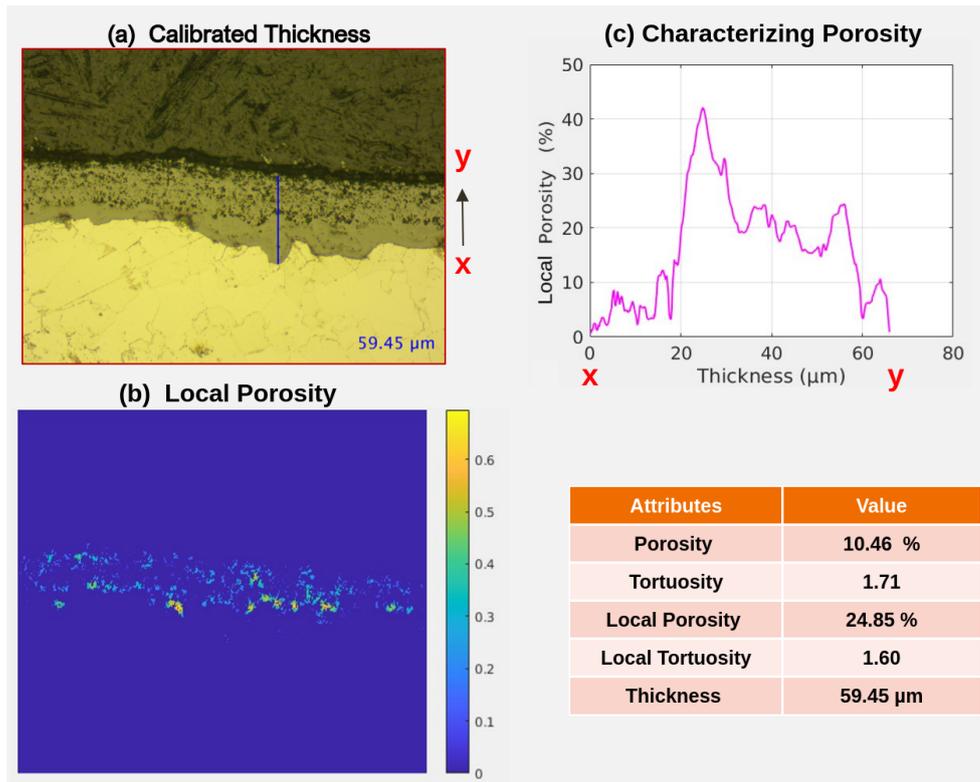


Fig. 1 Representative results obtained using the proposed automated pore characterization on sample OM image. (a) Input image with calibrated thickness, (b) local porosity map, and (c) local porosity computed as a function of thickness and attributes obtained.

of misclassified samples. Because the thickness-based criterion yielded the maximum number along the diagonal with 72.9% accuracy, it was used for classification in the results presented in this study. The classification methods based on porosity and thickness exhibit relatively lower accuracy, and the voting-based approach tends to produce conservative predictions—often assigning a corrosion stage that is equal to or greater than the actual stage. Given the critical implications of missing a true case of corrosion, the proposed approach was intentionally designed to be conservative, favoring false positives over false negatives. This aligns with our broader goal of ensuring robust detection, even at the cost of overestimation, to enhance reliability in early-stage corrosion monitoring.

Further in the proposed approach, the initial segmentation is performed using an unsupervised deep learning method—k-means clustering—which does not rely on predefined class labels. This segmentation is then used to compute local porosity and deposit thickness. The final classification into UDC stages is carried out based on quantitative thresholds of deposit thickness, rather than through a supervised learning model trained directly on labeled class data. As such, the class distribution

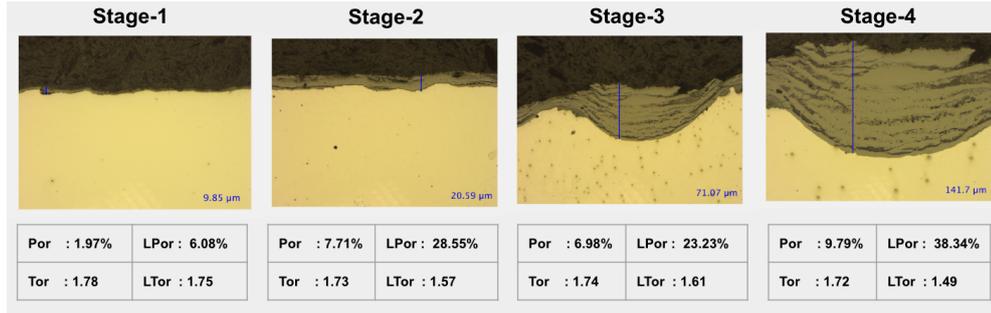


Fig. 2 Representative results obtained using the proposed automated pore characterization on sample OM images from each of the four stages. In each column, the top row corresponds to the input image with calibrated thickness shown in the insets. The attributes such as porosity (Por), tortuosity (Tor), local porosity (LPor) and local tortuosity (LTor) corresponding to the respective images are also shown for reference.

arises naturally from the underlying physical measurements, and no supervised loss function or class-based model training was applied. Therefore, class imbalance in the traditional supervised learning sense does not influence the classification performance.

To understand how the thickness and local porosity of the deposit vary with the UDC stage, the deposit thickness and local porosity versus the UDC stage are plotted in Figs. 4(a) and 4(b), respectively. As expected, the thickness of the deposit and local porosity increased with increasing stages of corrosion. Furthermore, the plots pinpoint the thresholds at which the transition occurred from Stage 3 to Stage 4 UDC. The thickness of the deposit was measured as the maximum value reading for each sample and was free from variation due to spallation. The highest reading in Stage 3 was approximately 90 μm and the lowest reading in Stage 4 was approximately 95 μm . Similarly, for Stage 3, the median porosity reported was $\sim 30\%$, which is consistent with the expert analysis of the subject matter reported by Abitha et al. [33]. Figure 5 shows a 2D histogram detailing the distribution of OM images with respect to the thickness of the deposit and local porosity. As shown in Fig. 5, both the thickness of the deposit and the local porosity increase as the UDC progresses. In Stage 1, the OM images exhibited the lowest porosity and thickness, corresponding to the green peak. Similarly, the peaks corresponding to the subsequent stages shifted toward higher values of the deposit thickness and porosity as the UDC progressed.

Figure 6 shows the analysis of the Cl^- concentration factor (Fig. 6 (a)) and pH (at 286°C) (Fig. 6(b)) as a function of the thickness of the deposit. For the thickness of the deposit $< 30\mu\text{m}$, the chloride concentration factor under the deposits was not high, and the local pH remained within the range of 5.5 to 5.8, which is close to the pH of the bulk water (5.8). However, as the thickness of the deposit increased, the concentration factor continued to increase and the pH continued to decrease (Fig. 6). Figure 7 top row shows the Cl^- Concentration Factor (Fig. 7(a)) and pH (at 286°C) (Fig. 7(b)) variation across the stages obtained using the manual thickness measurement and porosity computation with domain expertise [33] and bottom row shows the

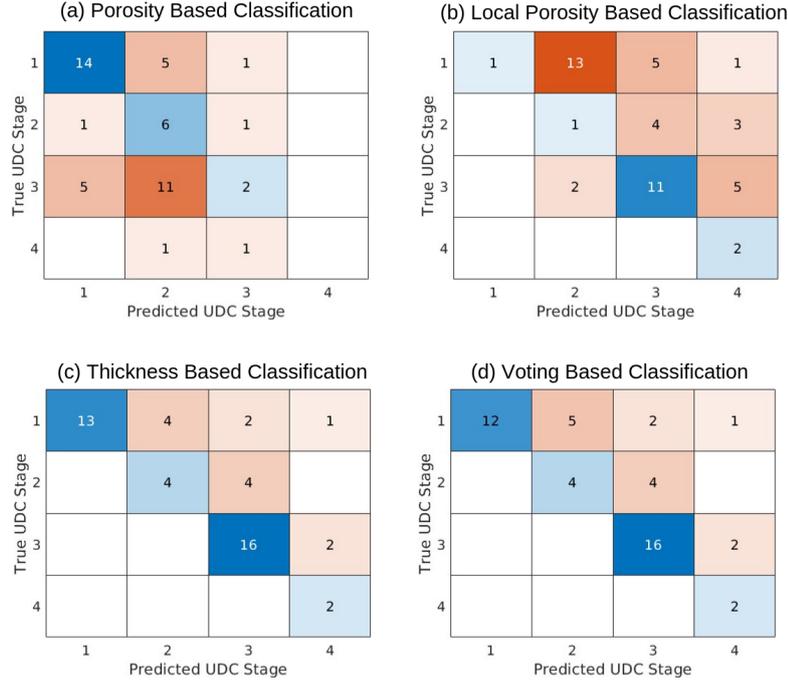


Fig. 3 Confusion matrices corresponding to (a) porosity, (b) local porosity, (c) thickness and (d) voting method based UDC stage classifications. These results correspond to the 48 sample OM images used in [33]. The true UDC stage corresponds to UDC stage determined by [33]

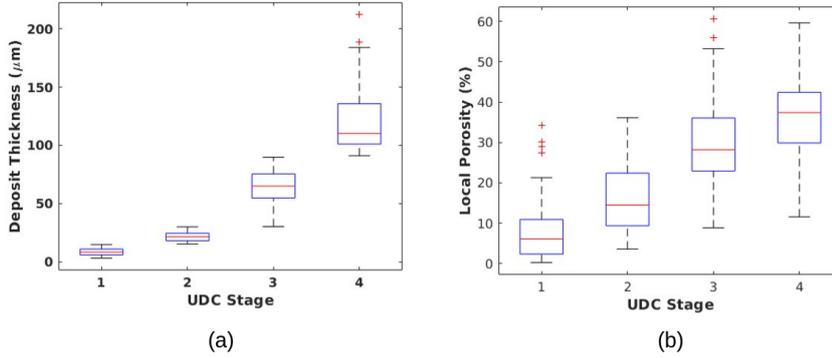


Fig. 4 Deposit thickness (a) and local porosity (b) versus UDC stage. As expected, the deposit thickness and local porosity increased during the later corrosion stages

Cl^- Concentration Factor (Fig. 7(c)) and pH (at 286°C) (Fig. 7(d)) variation across the stages computed using the proposed automated algorithm. This shows that the automated algorithm agrees well with manual thickness measurement and porosity computation with domain expertise. The minimum pH value of stage-3 was approximately between 2.8 and 3.5 (Fig. 7 (d)). Similarly, for Stage 4, the minimum pH value

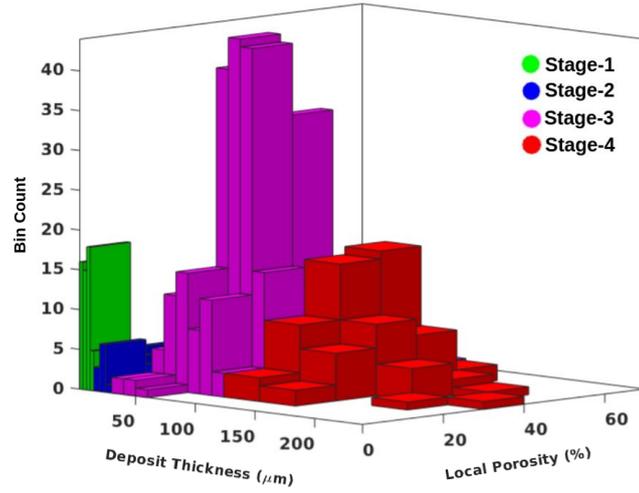


Fig. 5 2D histogram detailing the distribution of OM images with respect to deposit thickness and local porosity

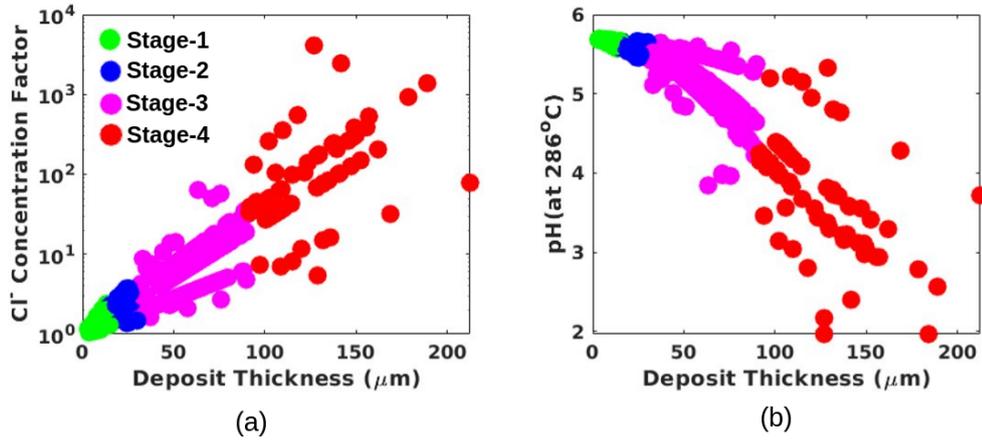


Fig. 6 The chloride concentration factor and corresponding pH variation as a function of the deposit thickness.

was between 1 and 1.5 (Fig. 7(d)). As shown in Fig. 7, Cl^- Concentration Factor increases and pH decreases as the UDC stage progresses. Although there is a spread for both measures across stages, a stable threshold to decide the transition from stage 3 to stage 4 is approximately between 2.8 and 3 for the pH values.

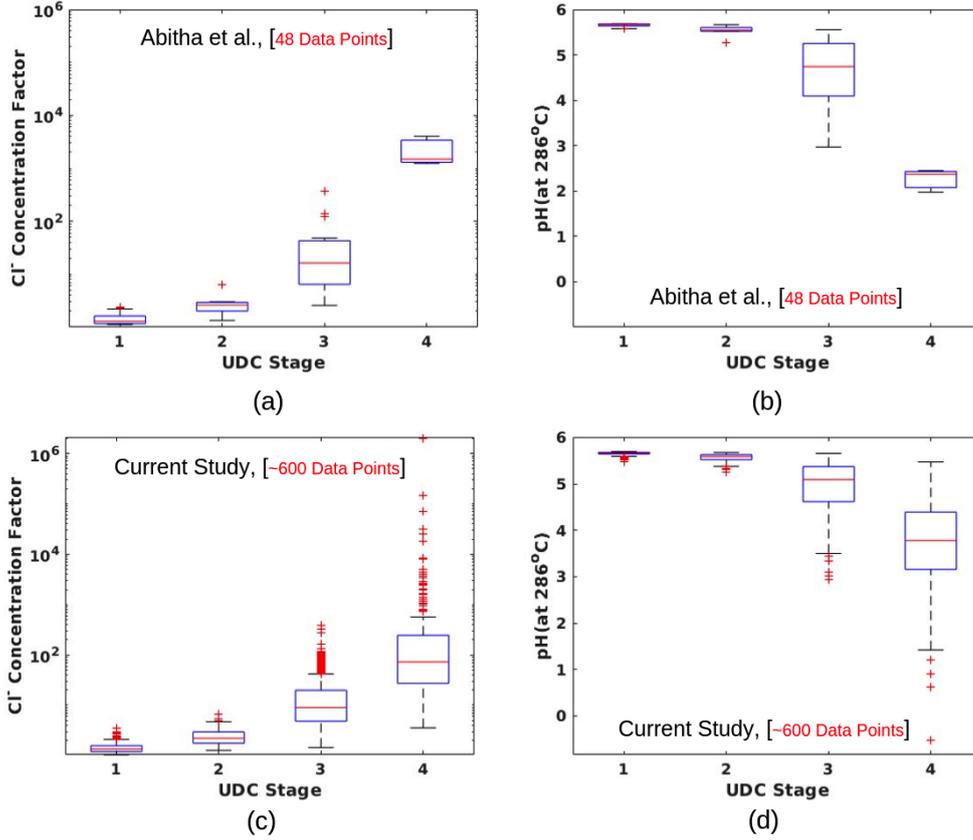


Fig. 7 The top row shows (a) the Chloride concentration factor versus UDC stage and (b) pH versus UDC stage using manual thickness measurement and porosity computation with domain expertise, as reported by Abitha et al.[33]. The bottom row shows (c) Chloride concentration factor versus UDC stage and (d) pH versus UDC stage obtained using the proposed automated algorithm.

3 Discussion

This study demonstrated an unsupervised machine-learning-based automated image processing algorithm for corrosion staging in optical microscopy images. The algorithm automatically computed the deposit thickness and local porosity directly, which facilitated the characterization of the corrosion stage. Furthermore, the calculated porosity facilitates the quantitative analysis of the chloride concentration factor and associated pH value. In particular, the observed deposit thickness transition between Stages 3 and 4 aligned well with the estimated pH threshold. This alignment is significant because it suggests a potential link between the deposit morphology, local chemistry, and the progression of the corrosion stage, offering valuable insight into the mechanisms of UDC.

This work further demonstrates $\sim 73\%$ accuracy in the characterization of the UDC stage in ex-service samples. Although this level of accuracy is promising, future work will focus on improving the robustness and performance of the algorithm for more complex corrosion morphologies. In contrast to traditional manual analysis, the proposed unsupervised algorithm offers an important advantage in eliminating human subjectivity associated with repetitive tasks associated with labeled data generation in supervised machine-learning approaches [9], paving the way for a broader application of unsupervised techniques in image analysis. Specifically, it demonstrated the feasibility of extracting meaningful quantitative information from complex corrosion images without relying on manual annotation. Using unsupervised machine learning and automated characterization techniques, this study presents a valuable tool for the assessment of UDC, which will ultimately contribute to improved maintenance strategies in large-scale settings.

4 Methods

4.1 Computation of Local Porosity

The method for computing the local porosity is divided into three parts: (i) pore detection, (ii) estimation of the local porosity, and (iii) characterization of the porosity. The process of characterizing the porosity throughout the thickness of the boiler deposit is illustrated in Fig. 8.

K-means [36] is an unsupervised clustering algorithm that clusters (or groups) data points according to their similarity score (often the mean squared error) is used for the pores detection. In contrast to other machine learning methods, unsupervised learning reduces dependency on labeled data, which is very rare in corrosion datasets. In a given image, a random cluster of centroids is initially selected, and the cluster centroids are refined over multiple iterations until the convergence criterion is satisfied. As shown in Fig. 9 (a), there are three clusters: (i) the waterside region, (ii) the deposit, and (iii) the metal pipe onto which the data points can be grouped. Each pixel (data point) has three intensity values corresponding to the Red, Green, and Blue (RGB) color model. An in-depth breakdown of the detection of pores from the optical microscopy (OM) images is shown in Fig. 9.

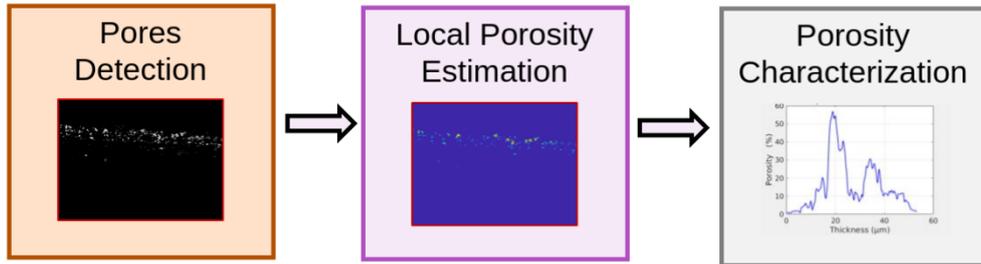


Fig. 8 Cartoon picture showcasing the key steps involved in the characterization of local porosity across the thickness of boiler deposit

Consider an OM image consisting of N pixels, where each pixel (x_1, \dots, x_N) is represented by vector 3×1 . The objective was to divide the data into three clusters such that the distance within each cluster was minimized and the distance between each cluster was maximized. For a given data point x_n and centroid μ_k of the k^{th} cluster, the binary variable $g_{nk} \in 0, 1$ indicates whether x_n belongs to cluster k . In other words, if x_n is assigned to group k , g_{nk} takes a value of 1; otherwise, it takes a value of 0. The clustering process aims to minimize the following:

$$J = \sum_{n=1}^N \sum_{k=1}^3 g_{nk} \|x_n - \mu_k\|^2 \quad (1)$$

The algorithm starts by assigning random values to each cluster centroid, μ_k . Initially, objective function J in Eq. (1) is minimized with respect to g_{nk} while keeping μ_k fixed. Then, J is minimized with respect to μ_k while keeping the value of g_{nk} fixed. This two-stage process is repeated until convergence is achieved. Subsequently, blob removal [37] was performed to fill the small black/white regions in the segmented image $S(r, c)$ (Fig. 9 (b)) to obtain the tube deposit image $T(r, c)$ (Fig. 9(c)). Here, (r, c) corresponds to the r^{th} row and c^{th} column of the OM image. This is followed by a logical AND operation [37] to obtain tube deposit with pores $T_p(r, c)$ (Eq. (2)) as shown in Fig. 9(d),

$$T_p(r, c) = \begin{cases} 1 & \text{if } S(r, c) = 1, \text{ and } T(r, c) = 1 \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

Finally, pores $P(r, c)$ were extracted using an XOR operation with $T(r, c)$ (Fig. 9(c)) and $T_p(r, c)$ (Fig. 9(d)) as shown in Eq. 3. Equation (3) produces an output consisting of the pores present within the deposit region, as shown in Fig. 9(e).

$$P(r, c) = \begin{cases} 1 & \text{if } T(r, c) = 1, \text{ and } T_p(r, c) = 0 \\ 1 & \text{if } T(r, c) = 0, \text{ and } T_p(r, c) = 1 \\ 0 & \text{otherwise.} \end{cases} \quad (3)$$

In the local porosity estimation process, a localized porosity map is initially generated by aggregating the porosity values within a predefined neighborhood relative to each pore. To compute the porosity within this neighborhood, a square matrix denoted by M was employed, with all its entries set to 1 and a size of 55×55 , equivalent to a physical area of $5\mu m \times 5\mu m$. This matrix M is then convolved with the pore image $P(r, c)$ and tube deposit image $T(r, c)$, as indicated in Eq. (4), and Eq. (5), respectively.

$$V_v^{local}(r, c) = \sum_{dr=-a}^a \sum_{dc=-b}^b M(dr, dc)P(r - dr, c - dc) \quad (4)$$

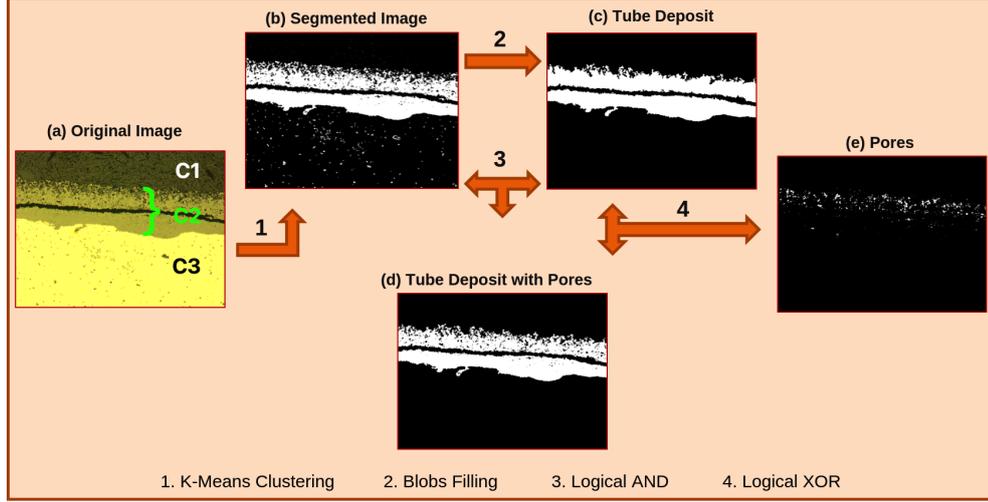


Fig. 9 Illustration of the important steps involved in pore detection (a) Original OM image (clusters C1: sample support, C2: deposit region and C3: metal tube), (b) Segmented Image obtained after K-means clustering, (C) blob-filled tube deposit, (d) tube deposit with pores, and (e) detected pores. The operations performed at each step were provided correspondingly at the bottom of the image.

$$V_T^{local}(r, c) = \sum_{dr=-a}^a \sum_{dc=-b}^b M(dr, dc)T(r - dr, c - dc), \quad (5)$$

The edges of P and T are appropriately zero padded. However, $V_v^{local}(r, c)$ and $V_T^{local}(r, c)$ are cropped to maintain the same size as P . This step is performed using the `conv2d` function in MATLAB [32]. A sample of accumulated pores ($V_v^{local}(r, c)$) and accumulated tube deposits ($V_T^{local}(r, c)$) in the considered local neighborhood is shown in Figs. 10 (a) and 10 (b), respectively. Then the local porosity map/image (Fig. 10(c)) is computed as

$$\phi^{local}(r, c) = \begin{cases} \frac{V_v^{local}(r, c)}{V_T^{local}(r, c)} & \text{if } P(r, c) = 1 \\ 0 & \text{otherwise.} \end{cases} \quad (6)$$

Finally, the averaged local porosity is computed as the mean of the local porosity map given in the equation. (6). Tortuosity was computed using the empirical relation given in [33]. Similarly, the local tortuosity was computed using the same empirical relation for tortuosity, but with porosity replaced with the averaged local porosity.

To characterize the local porosity along the thickness $\phi^{th}(r)$ of the deposit, the $\phi^{local}(r, c)$ in Eq.(6) is averaged across each row r of the local porosity image,

$$\phi^{th}(r) = \frac{1}{n_C} \sum_{c=1}^C \phi^{local}(r, c), \quad (7)$$

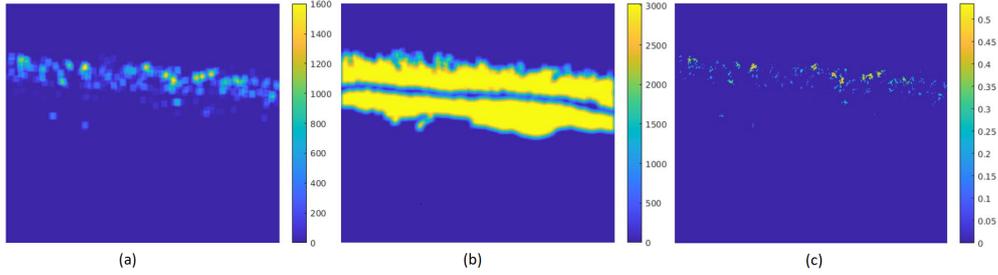


Fig. 10 Example case (from left to right) showing (a) accumulated pores, (b) accumulated tube deposits, and (c) local porosity map obtained using Eq. (6).

Table 1 Criteria for UDC stage classification based on percentage of porosity and thickness

UDC Stage	Porosity (%)	Thickness (μm)
1	0 - 2	0 - 15
2	2 - 10	15 - 30
3	10 - 30	30 - 90
4	> 30	> 90

where n_C is the number of pores in the r^{th} row. The thickest portion in the deposit on each sample image is identified and the deposit width at that location is estimated as

$$\text{Thickness} = n_p \times \text{resolution}, \quad (8)$$

where n_p is the number of pixels and *resolution* is the length of the pixels expressed in μm . The resolution of the pixels in this case is $0.11\mu\text{m}$.

4.2 Criteria for Corrosion Stage Classification

This study further evaluated the importance of the thickness of the calculated deposit and the local porosity by using them to classify the UDC stages in [33]. The classification was performed based on the porosity, which was first computed as the fraction of total pores to the total tube deposit. By analyzing the experimental studies reported by Abitha et al. [33], each OM image is assigned a UDC stage based on porosity value as listed in Table 1. The same criteria based on the porosity values were utilized for the averaged local-porosity-based classification. In the thickness-based classification, the UDC stage is assigned according to the thickness values reported in Table 1 [33]. Next, for the voting method, if two of the aforementioned criteria agree, the corresponding stage is assigned to the OM image. If none agreed with each other, a thickness-based criterion was used.

Four boilers operating with all-volatile treatment (AVT) implementation produced 12 coil pieces, each measuring approximately 1 m in length. The OM images used

in this study were acquired from a group of boilers operating simultaneously and theoretically identically, all suffering from UDC. Typically, boilers have a helical coil design [38, 39] with boiling water on the exterior and hot process gas inside. The details of the image acquisition and corrosion stage characterization for UDC are provided in Abitha et al. [33].

5 Data and Code Availability

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request. The software implementation can be accessed from the GitHub repository provided in <https://github.com/NaveenPaluru/UDC-Analysis>.

6 Author Contributions

ARK: Original draft writing, software, methodology, investigation, formal analysis, and validation; NP: Original draft writing, software, methodology, investigation, formal analysis, and validation; RSM: Review and editing, validation, methodology, formal analysis, and investigation; DA: supervision, resources, and formal analysis; PS: supervision, resources, and formal analysis; NL: Review and editing of the manuscript, supervision, resources, acquisition of funding, formal analysis, and conceptualization; PY: Review and editing of the manuscript, supervision, resources, acquisition of funding, formal analysis, and conceptualization.

7 Competing Interests

The authors declare that they have no competing interests.

References

- [1] Obot, I. B. Under-deposit corrosion on steel pipeline surfaces: mechanism, mitigation and current challenges. *Journal of Bio-and Tribo-Corrosion* **7**, 49 (2021).
- [2] Hladky, K., Callow, L. & Dawson, J. Corrosion rates from impedance measurements: an introduction. *British Corrosion Journal* **15**, 20–25 (1980).
- [3] Jones, D. A. *Principles and prevention of corrosion* (Prentice Hall, 1996).
- [4] Bard, A. J., Faulkner, L. R. & White, H. S. *Electrochemical methods: Fundamentals and applications* (John Wiley & Sons, Ltd, 2022).
- [5] Trethewey, K. R. & Chamberlain, J. *Corrosion for Science and Engineering* (Longman, 1998).

- [6] Cipolloni, G., Menapace, C., Cristofolini, I. & Molinari, A. A quantitative characterisation of porosity in a cr–mo sintered steel using image analysis. *Materials Characterization* **94**, 58–68 (2014). URL <https://www.sciencedirect.com/science/article/pii/S1044580314001399>.
- [7] Singh, S., Goyal, K. & Bhatia, R. A review on protection of boiler tube steels with thermal spray coatings from hot corrosion. *Materials Today: inproceedings* **56**, 379–383 (2022). URL <https://www.sciencedirect.com/science/article/pii/S2214785322002437>. International Conference on Materials, Machines and Information Technology-2022.
- [8] Shivadhanush, B. R. *et al.* Corrosion and risk modelling analytics. *Proceedings of the AMPP CORROSION Conference* (2024). Paper ID: AMPP-2024-20833.
- [9] Coelho, L. B. *et al.* Reviewing machine learning of corrosion prediction in a data-oriented perspective. *npj Materials Degradation* **6**, 8 (2022).
- [10] Diao, Y., Yan, L. & Gao, K. Improvement of the machine learning-based corrosion rate prediction model through the optimization of input features. *Materials & Design* **198**, 109326 (2021).
- [11] Yan, L., Diao, Y., Lang, Z. & Gao, K. Corrosion rate prediction and influencing factors evaluation of low-alloy steels in marine atmosphere using machine learning approach. *Science and technology of Advanced Materials* **21**, 359–370 (2020).
- [12] Roy, A. *et al.* Machine-learning-guided descriptor selection for predicting corrosion resistance in multi-principal element alloys. *npj Materials Degradation* **6**, 9 (2022).
- [13] Sutojo, T. *et al.* A machine learning approach for corrosion small datasets. *npj Materials Degradation* **7**, 18 (2023).
- [14] Bhat, N., Birbilis, N. & Barnard, A. S. Unsupervised learning and pattern recognition in alloy design. *Digital Discovery* **3**, 2396–2416 (2024).
- [15] Tiwari, T., Jalalian, S., Mendis, C. & Eskin, D. Classification of t6 tempered 6xxx series aluminum alloys based on machine learning principles. *JOM* **75**, 4526–4537 (2023).
- [16] Zhang, H. *et al.* Nano-porosity effects on corrosion rate of zr alloys using nanoscale microscopy coupled to machine learning. *Corrosion Science* **208**, 110660 (2022).
- [17] Brown, B. & Moloney, J. Under-deposit corrosion. *Trends in oil and gas corrosion research and technologies* 363–383 (2017).

- [18] Ali, M. *et al.* Corrosion-related failures in heat exchangers. *Corrosion Reviews* **39**, 519–546 (2021).
- [19] Shokri, A. & Sanavi Fard, M. Under deposit corrosion failure: mitigation strategies and future roadmap. *Chemical Papers* **77**, 1773–1790 (2023).
- [20] Dooley, R. B. & Bursik, A. Caustic gouging. *PowerPlant Chemistry* **12**, 188–192 (2010).
- [21] Dooley, R. B. & Bursik, A. Acid phosphate corrosion. *PowerPlant Chemistry* **12**, 368 (2010).
- [22] Dooley, R. B. & Bursik, A. Hydrogen damage. *PowerPlant Chemistry* **12**, 122 (2010).
- [23] Dooley, R. Deposition in boilers—review of soviet and russian literature. *EPRI, Paolo Alto, CA* **1004193** (2003).
- [24] Dooley, R. Deposition on drum boiler tube surfaces. Tech. Rep., final technical report, EPRI-Russia (2005).
- [25] Dooley, B. in *15 - developing the optimum cycle chemistry provides the key to reliability for combined cycle/hrsg plants* (ed. Eriksen, V. L.) *Heat Recovery Steam Generator Technology* 321–353 (Woodhead Publishing, 2017). URL <https://www.sciencedirect.com/science/article/pii/B9780081019405000154>.
- [26] Friend, D. & Dooley, R. Volatile treatments for the steam-water circuits of fossil and combined cycle/hrsg power plants. *IAPWS Report, Niagara Falls, Canada* (2010).
- [27] Garcia, A., Place, T. & Wood, S. Development of key performance indicator for management of under deposit corrosion (udc). *Proceedings of the International Pipeline Conference* **50251** (2016). Paper No: V001T03A029.
- [28] Dooley, B. & Anderson, B. Trends in hrsg reliability—a 10-year review. *Power-Plant Chemistry* **21**, 158 (2019).
- [29] Dooley, R. Fossil plant cycle chemistry and steam. *Steam, Water, and Hydrothermal Systems: Physics and Chemistry, Meeting the Needs of Industry* 35–49 (2000).
- [30] Jeon, S.-H., Shim, H.-S., Lee, J.-M., Han, J. & Hur, D. H. Simulation of porous magnetite deposits on steam generator tubes in circulating water at 270 °c. *Crystals* **10** (2020).
- [31] Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. Nih image to imagej: 25 years of image analysis. *Nature methods* **9**, 671–675 (2012).

- [32] Inc., T. M. Matlab (2022).
- [33] Ramesh, A. *et al.* Critical deposit loading thresholds for under deposit corrosion in steam generators. *Corrosion* **78**, 584–598 (2022).
- [34] Thyagarajan, A. *et al.* The virtual corrosion engineer. *Corrosion and Materials Degradation* **2**, 762–769 (2021).
- [35] Dwivedi, D., Lepková, K. & Becker, T. Carbon steel corrosion: a review of key surface properties and characterization methods. *RSC advances* **7**, 4580–4610 (2017).
- [36] MacQueen, J. Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics* **5**, 281–298 (1967).
- [37] Soille, P. *et al.* *Morphological image analysis: principles and applications* Vol. 2 (Springer, 1999).
- [38] Bozzoli, F., Cattani, L., Mocerino, A. & Rainieri, S. Experimental estimation of the local heat-transfer coefficient in coiled tubes in turbulent flow regime. *Journal of Physics: Conference Series* **745**, 032034 (2016).
- [39] Mitrovic, J. *Heat Exchangers: Basics Design Applications* (BoD–Books on Demand, 2012).

Figure Legends

Figure 1. Representative results obtained using the proposed automated pore characterization on sample OM image. (a) Input image with calibrated thickness, (b) local porosity map, and (c) local porosity computed as a function of thickness and attributes obtained.

Figure 2. Representative results obtained using the proposed automated pore characterization on sample OM images from each of the four stages. In each column, the top row corresponds to the input image with calibrated thickness shown in the insets. The attributes such as porosity (Por), tortuosity (Tor), local porosity (LPor) and local tortuosity (LTor) corresponding to the respective images are also shown for reference.

Figure 3. Confusion matrices corresponding to (a) porosity, (b) local porosity, (c) thickness and (d) voting method based UDC stage classifications. These results correspond to the 48 sample OM images used in [33]. The true UDC stage corresponds to UDC stage determined by [33]

Figure 4. Deposit thickness (a) and local porosity (b) versus UDC stage. As expected, the deposit thickness and local porosity increased during the later corrosion stages

Figure 5. 2D histogram detailing the distribution of OM images with respect to deposit thickness and local porosity

Figure 6. The chloride concentration factor and corresponding pH variation as a function of the deposit thickness.

Figure 7. The top row shows (a) the Chloride concentration factor versus UDC stage and (b) pH versus UDC stage using manual thickness measurement and porosity computation with domain expertise, as reported by Abitha et al.[33]. The bottom row shows (c) Chloride concentration factor versus UDC stage and (d) pH versus UDC stage obtained using the proposed automated algorithm.

Figure 8. Cartoon picture showcasing the key steps involved in the characterization of local porosity across the thickness of boiler deposit

Figure 9. Illustration of the important steps involved in pore detection (a) Original OM image (clusters C1: sample support, C2: deposit region and C3: metal tube), (b) Segmented Image obtained after K-means clustering, (C) blob-filled tube deposit, (d) tube deposit with pores, and (e) detected pores. The operations performed at each step were provided correspondingly at the bottom of the image.

Figure 10. Example case (from left to right) showing (a) accumulated pores, (b) accumulated tube deposits, and (c) local porosity map obtained using Eq. (6).