1. **Narayanganj Dockyard Biofouling Dataset (NDBD):**

Our dataset, comprising 35 high-resolution images with 92 annotated biofouling instances from Dockyard & Engineering Works LTD, Narayanganj, Bangladesh. Details are noted in Table 1, and sample images are shown in Figure 1. While collecting this dataset, we maintained diversity and human presence. However, ethical standards as detailed in section 2 were ensured.

|  |  |
| --- | --- |
| **Table 1: Characteristics of Narayanganj Dockyard Biofouling Dataset (NDBD)** | |
| **Characteristic** | **Description** |
| Total Original Images and number of Instances | **Raw 35 high-resolution** Marine Biofouling images collected from **Dockyard & Engineering Works LTD, Narayanganj, Bangladesh**. The total count of annotated biofouling regions (e.g., barnacle clusters) across all 35 images is **92** |
| Key Variabilities | **Angles:** varied viewing angles relative to hull surfaces |
| **Lighting:** diverse illumination conditions (e.g., different times of day, shadows, and highlights) |
| **Background:** heterogeneous backgrounds (e.g., water, dock structures, and equipment) |
| **Instance regions:** biofouling regions varying in size from small patches to extensive coverage |
| **Human presence:** human presence in approximately half the images to simulate real dockyard operations and occlusions (handled per ethics guidelines in Section 4.2) |
| **Image scale and coverage:** Collected at 3 different coverage thresholds (mask pixels / total image pixels). categorized 4 as Full-coverage (>80% frame coverage), 20 as Medium-scale (80% to 30%), and 11 as Contextual (<30%). |
| Annotation Tool | **CVAT:** Computer Vision Annotation Tool (CVAT.ai 2025), an open-source, web-based platform widely adopted in computer vision research. All 35 images were manually annotated with precise polygons around regions containing biofouling. Quality was cross-checked to minimize errors and corresponding mask labels were then generated for instance segmentation. |



Raw Images

Annotated Images

(a)

(b)

(c)



Raw Images

Annotated Images

(d)

(e)

(f)

**Figure 1.** Representative raw images from the marine biofouling dataset with corresponding instance-level annotations. (a–c) Scenes without human presence: (a) contextual daylight view, (b) submerged ship-hull biofouling, and (c) heavy (>80%) fouling coverage captured from varying angles and illumination conditions. (d–f) Scenes with human presence: (d) under-hull inspection in low-light conditions, (e) side-view cleaning operation, and (f) dockyard environment with visible infrastructure.

1. **Ethics and Data Privacy Considerations**

Ethical data curation is critical for computer vision datasets. Specially, in marine applications to protect privacy, ensure consent, and mitigate biases. This dataset adheres to frameworks like "Datasheets for Datasets," focusing on purpose, privacy, consent, and diversity. (Timnit Gebru 2018)

The dataset, collected via collaboration with **Dockyard & Engineering Works LTD, Bangladesh**. Images depict vessel hull surfaces in a dockyard setting and contain no personally identifiable information (PII), such as human faces, worker identities, or vessel registration details, which were anonymized during preprocessing (Robinson 2024, Kaur 2024). Institutional consent was obtained, complying with local regulations, GDPR for data sharing (Complete guide to GDPR compliance 2025, University n.d.), and ethical guidelines from the International Maritime Organization (IMO) on marine worker safety (IMO 2006, ILO 2006). We conducted a preliminary bias audit during preprocessing, ensuring balanced representation across coverage types.

# We are releasing the anonymized dataset publicly under Creative Commons Attribution-NonCommercial 4.0, with usage guidelines prohibiting re-identification.

**References**

2025. "Complete guide to GDPR compliance." *GDPR.eu.* https://gdpr.eu/.

CVAT.ai, Corporation. 2025. *Computer Vision Annotation Tool (CVAT).* Accessed February 2025. https://www.cvat.ai/.

ILO, International Labour Organization. 2006. *Maritime Labour Convention.* Geneva, Switzerland, February 23. https://www.ilo.org/international-labour-standards/maritime-labour-convention-2006.

IMO, International Maritime Organization. 2006. "Guidelines on the Basic Elements of a Shipboard Occupational Health and Safety Programme." *MSC-MEPC.2/Circ.3.* London, June 5. https://www.register-iri.com/wp-content/uploads/MSC-MEPC.2-Circ.3.pdf.

Kaur, Dr. Jagreet. 2024. "blog." *xenonstack.* October 22. https://www.xenonstack.com/blog/ethical-considerations-in-computer-vision.

Robinson, Ryan. 2024. "blog page." *Alation.* july 16. https://www.alation.com/blog/data-ethics-in-ai-6-key-principles-for-responsible-machine-learning/.

Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, Kate Crawford. 2018. "Datasheets for Datasets." *ArXiv.* doi:https://doi.org/10.48550/arXiv.1803.09010.

University, Florida State. n.d. "GDPR and Research." *FSU Research — Investigator Resources.* https://www.research.fsu.edu/research-offices/ohsp/investigator-resources/confidentiality-privacy-and-information-security-in-human-research/gdpr-and-research/.