An IEEE Societies and Technical Councils Publication
IEEE Data Descriptions

Received 6 September 2024; revised 13 October 2024 and 13 October 2024; accepted 13 October 2024; Date of publication 16 October, 2024; date of current version 15 November, 2024.

Digital Object Identifier 10.1109/IEEEDATA.2024.3481394

Collection: *Wound Healing Assay Dataset (WHAD) and Cell Adhesion and Motility Assay Dataset (CAMAD)*

LEONARDO O. IHEME [1], SEVGI ONAL [2,3], YUSUF S. ERDEM [4], MAHMUT UCAR [4], OZDEN YALCIN-OZUYSAL [5], DEVRIM PESEN-OKVUR [5], BEHCET U. TOREYIN [6], AND DEVRIM UNAY [4]

[1]Epam Systems, Istanbul 34396, Türkiye
[2]Department of Biotechnology and Bioengineering, Izmir Institute of Technology Graduate School of Science and Engineering, Izmir 35430, Türkiye
[3]Center for Advanced Biomaterials for Health Care, Istituto Italiano di Tecnologia, 80125 Napoli, Italy
[4]Department of Electrical and Electronics Engineering, Faculty of Engineering, Izmir Democracy University, Izmir 35140, Türkiye
[5]Department of Molecular Biology and Genetics, Faculty of Science, Izmir Institute of Technology, Izmir 35430, Türkiye
[6]Department of Artificial Intelligence and Data Engineering, Istanbul Technical University Informatics Institute, Istanbul 34469, Türkiye

CORRESPONDING AUTHORS: Ozden Yalcin-Ozuysal (e-mail: ozdenyalcin@iyte.edu.tr); Devrim Pesen-Okvur (e-mail: devrimpesen@iyte.edu.tr); Devrim Unay (e-mail: unaydevrim@gmail.com).

This work was supported in part by Marie Curie IRG under Grant FP7 PIRG08-GA-2010-27697; in part by the Scientific and Technological Research Council of Turkey (TUBITAK) under Grant 113Z088 and Grant 119E578; and in part by Izmir Institute of Technology under Grant 2016IYTE65.

This article has supplementary downloadable material available at https://doi.org/10.1109/IEEEDATA.2024.3481394, provided by the authors.

ABSTRACT This article introduces two novel datasets for cell motility and wound healing research: the wound healing assay dataset (WHAD) and the cell adhesion and motility assay dataset (CAMAD). WHAD comprises time-lapse phase-contrast images of wound healing assays using genetically modified MCF10A and MCF7 cells, while CAMAD includes MDA-MB-231 and RAW 264.7 cells cultured on various substrates. These datasets offer diverse experimental conditions, comprehensive annotations, and high-quality imaging data, addressing gaps in existing resources. The collection methods, experimental designs, and annotation processes are detailed, emphasizing the reliability and utility of our datasets. WHAD and CAMAD enable researchers to develop robust algorithms for cell tracking, segmentation, and behavior analysis in phase-contrast microscopy images. The applications of these datasets span cancer research, image processing technique development, and automated analysis of wound healing and cell motility assays. By providing these resources, we aim to accelerate progress in biomedical image analysis and cell biology research.

IEEE SOCIETY/COUNCIL Engineering in Medicine and Biology Society (EMBS)

DATA TYPE/LOCATION Images; Izmir, Türkiye

DATA DOI/PID 10.21227/f77j-j355

INDEX TERMS Cancer research, cell motility, cell tracking, dataset, image analysis, phase-contrast microscopy, segmentation, wound healing.

BACKGROUND

Cell motility and wound healing are fundamental biological processes crucial for embryonic development, tissue homeostasis, and disease progression [17]. Understanding these processes is essential for advancing our knowledge in related fields such as developmental biology, regenerative medicine, and cancer research. The ability to accurately observe, quantify, and analyze cell behavior in various contexts has become increasingly important in biomedical research.

Phase-contrast microscopy (PCM) has emerged as a powerful tool for studying cell dynamics, allowing researchers to observe living cells without the need for staining or

TABLE I. Summary Cell Lines and Genetic Modifications for WHAD

Cell Line	Condition	Modification	Culture Medium	Assay Medium
MCF10A	CC	Control (No modification)	DMEM-F12 + 5% Horse Serum + 20 ng/mL EGF + 0.5 μg/mL Hydrocortisone + 100 ng/mL Cholera Toxin + 10 μg/mL Insulin + 1% Penicillin/Streptomycin	DMEM-F12 + 1% Penicillin/Streptomycin
	NC	NICD overexpression		
	C61	CYR61 shRNA expression		
	N61	NICD + CYR61 shRNA		
MCF7	Control	LacZ overexpression	DMEM + 10% fetal bovine serum + 1% penicillin/streptomycin	DMEM + 1% FBS + 1% Penicillin/Streptomycin
	SEMA6D	SEMA6D overexpression		

fluorescent labeling. However, the analysis of PCM data presents unique challenges, including variability in image quality, cell segmentation difficulties, and the need for robust tracking algorithms [18].

To address these challenges and accelerate progress in the field, there is a growing need for comprehensive, well-annotated datasets that capture diverse cellular behaviors under various experimental conditions. Such datasets are essential for developing and validating advanced image analysis algorithms, machine learning models, and automated cell tracking systems.

Overview of Datasets

This article introduces two novel datasets designed to advance research in cell motility and wound healing: the wound healing assay dataset (WHAD) and the cell adhesion and motility assay dataset (CAMAD). These datasets aim to provide researchers with high-quality, diverse, and well-annotated data for developing and validating algorithms in cell biology and biomedical image analysis.

WHAD

WHAD consists of time-lapse PCM images capturing the process of wound healing in cell monolayers. The dataset features MCF10A (normal breast epithelial) and MCF7 (breast cancer) cell lines with various genetic modifications as explained in Table I. Images are captured at regular intervals over the course of wound closure, with manual annotations providing precise delineation of wound areas and detached cell boundaries.

CAMAD

CAMAD focuses on individual cell behavior in different microenvironments, providing time-lapse videos of MDA-MB-231 (breast cancer) and RAW 264.7 (macrophage) cells on various substrates. The dataset includes 16 distinct experimental setups, each consisting of 600 frames captured over 5 h. Detailed manual annotations include cell boundary delineation and measurements of cell area, circularity, and aspect ratio.

Both datasets are designed to complement existing resources in the field by offering experimental diversity, high-quality imaging, comprehensive annotations, and broad applicability. They support a wide range of applications, from basic research in cell biology to the development of advanced image processing techniques and automated analysis systems for wound healing and cell motility assays.

By providing these datasets, we aim to facilitate the development of more robust and generalizable algorithms for cell tracking, segmentation, and behavior analysis, ultimately advancing our understanding of cellular behavior in health and disease.

Related Datasets

Several datasets focusing on cell motility and tracking are available in the scientific community [25]. However, our dataset addresses some unique aspects and fills certain gaps left by the existing data. Here, we review some of the most relevant datasets and highlight how our contribution complements the field.

Cell Tracking Datasets

The cell tracking challenge [19] provides two relevant datasets: Glioblastoma-astrocytoma U373 cells on a polyacrylamide substrate and pancreatic stem cells on a polystyrene substrate. Both datasets use PCM and are openly accessible. The glioblastoma dataset consists of 230 frames captured every 15 min, while the pancreatic stem cell dataset comprises 600 frames with a 10-min interval. These datasets offer valuable resources for algorithm development in cell tracking. However, they are limited to single experiments, specific cell types, and substrates.

Another significant contribution is [8]. This extensive dataset includes 48 experiments with a total of 49 919 frames, providing a rich resource for cell-tracking algorithm development and validation.

Mitosis Detection Datasets

Liu et al. [3] presented 7180 frames from five experiments, Huh et al. [2] used 1013 frames from a single experiment, and Su et al. [15] compiled 16 208 frames from 16 experiments, all with a 5-min interval.

These datasets provide valuable resources for developing and bench-marking mitosis detection algorithms. However, they are often limited in their scope, focusing on specific cell types or experimental conditions.

Wound Healing Assay Datasets

In the context of wound healing assays, Zaritsky et al. [6] introduced a dataset with raw live time-lapse *in vitro* data, investigating the impact of growth factors on wound healing. This dataset encompasses two cell lines under various

conditions, with six unique experimental setups repeated four to six times. It consists of 1024×1024 pixel grayscale images and corresponding binary mask files.

Kauanova et al. [13] showcased a pipeline for analyzing cell motility metrics using their wound healing assay dataset, acquired through transmitted light microscopy. The open wound regions are manually annotated, and they plan to release the raw data.

Bobadilla et al. [9] developed a novel wound front migration quantification metric using a publicly available dataset of 24 wound healing assays, each containing 48 frames of 1000×1000 pixel grayscale images.

Suarez-Arnedo et al. [12] created a dataset of 30 manually annotated images from a 28-h wound healing assay using phase-contrast microscopy to validate their analysis tool.

Our Dataset Contribution

WHAD and CAMAD complement these existing resources in several ways.

1) *Experimental Diversity*: We provide datasets that encompass 11 experiments with a total of 5490 frames for WHAD and 16 experiments with a total of 9600 frames for CAMAD, under various experimental conditions.
2) *Temporal Resolution*: Our datasets have consistent time intervals between frames adjusted during data acquisition. This feature allows for the study of cell behavior under selected temporal resolutions during data processing. This is particularly useful for developing adaptive algorithms that can handle diverse time-lapse conditions.
3) *Open Access*: Like some of the mentioned datasets, ours is freely available to the research community [22], [23], promoting reproducibility and collaborative advancement in the field.
4) *Comprehensive Ground Truth*: We provide detailed annotations for cell segmentation, which is often a crucial preprocessing step for both cell tracking and mitosis detection.
5) *Broader Applicability*: While many existing datasets focus on specific cell types or experimental conditions, our dataset aims to capture a wider range of cellular behaviors, making it more generally applicable for algorithm development and testing.

By providing this dataset, we aim to fill the gap between highly specific, single-condition datasets and the need for more diverse, multiexperiment data in the field of cell motility analysis. Our dataset can serve as a valuable resource for researchers developing robust, adaptable algorithms for cell segmentation, tracking, and behavior analysis in phase-contrast microscopy images.

Applications and Impact of WHAD & CAMAD

Our newly introduced datasets have been instrumental in advancing various areas of biomedical research. Here, we highlight some significant studies that demonstrate the broad impact and applications of these datasets in the fields of cell motility and wound healing.

Cell–Cell Interactions and Cancer Research

Onal et al. [14] utilized our dataset to investigate the interaction between breast cancer cells (BCCs) and macrophages. Their study, employing cell-on-a-chip devices, revealed that macrophages exhibit chemotaxis toward BCC due to the secretion of colony-stimulating factor-1 (CSF-1) by BCC. Conversely, BCC requires direct contact with macrophages for interaction with macrophage-derived epidermal growth factor (EGF). This work elucidates the paracrine-juxtacrine loop between BCC and macrophages, highlighting the role of EGF and CSF-1 in cancer cell motility and adhesion.

Part of MCF10A WHAD was utilized to investigate whether CYR61 gene functions downstream of oncogenic Notch1 signaling. The data demonstrated that CYR61 is a mediator of Notch1-induced pro-metastatic migration phenotype in normal breast epithelial cells [10]. The MCF7 WHAD was utilized to investigate the role of the SEMA6D gene in noninvasive BCC lines. WHAD analysis showed that SEMA6D overexpression induced the migration of MCF7 cells while shifting the migratory behavior toward a more detached phenotype [16].

Image Processing and Analysis Techniques

Advancements in image processing have been crucial for maximizing the utility of cell motility and wound healing datasets. Ucar et al. [24] introduced a preprocessing pipeline to detect and correct distortions in phase-contrast microscopy time-lapse images. In addition to our dataset, this study included 15 395 frames from 27 experiments and 830 frames from the cell tracking challenge. The proposed pipeline effectively addresses blank frames and intensity variations, significantly enhancing the accuracy of subsequent image analysis processes. In another innovative approach, Erdem et al. [20] proposed a data synthesis method using neural style transfer to generate realistic wound healing assay images. This method supports the training of deep learning models by providing synthetic data that closely resemble real PCM images, which is particularly beneficial when labeled data are scarce. The effectiveness of this approach was validated using images from the WHAD.

Automated Analysis of Wound Healing and Cell Motility

The complexity and time demands of manual analysis have driven the development of automated techniques for analyzing PCM images. Erdem et al. [18] discussed a comprehensive workflow for the automated analysis of wound healing and cell motility assays. Their work, using time-lapse images of MDA-MB-231 BCCs and MCF10A normal epithelial cells, covered image acquisition, preprocessing, segmentation, tracking, and quantification. This automated process is essential for managing large-scale analyses efficiently. Building on this, Mayali et al. [11] presented an automated method

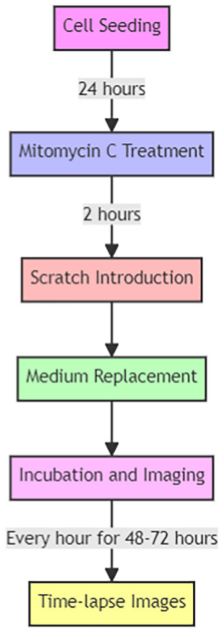


FIG. 1. Wound healing assay process.

specifically for analyzing wound healing microscopy image series. Our dataset enabled the development and comparison of traditional image processing and deep learning-based approaches for automated segmentation of the wound area. These methods facilitate the rapid acquisition of reliable and reproducible results, which is crucial for wound healing research.

The studies highlighted here collectively demonstrate the significant impact of WHAD and CAMAD in advancing biomedical image analysis.

COLLECTION METHODS AND DESIGN

WHAD

Cell Lines and Culture Conditions

MCF10A (CRL-10317) and MCF7 (HTB-22) cell lines were obtained from the American Type Culture Collection (ATCC). These cells were genetically modified through lentiviral infection, as described in previous studies [10], [16]. The culture conditions and genetic modifications are summarized in Table I. Further details about experiments are recorded in the supplementary material `whad_experiments_summary.xlsx`.

Wound Healing Assay Protocols

The main steps of the wound healing assay are illustrated in Fig. 1. Briefly, cells were seeded on 12-well plates at the density of 1×10^6 cells/well and 7.5×10^5 cells/well, for MCF10A and MCF7 cells, respectively. A 2-h treatment with $10\ \mu\text{g/mL}$ mitomycin C was applied to block proliferation. The phase-contrast images of defined positions were recorded every hour for 48 or 72 h with Leica DMI8 confocal microscope supplemented with 5% CO_2 at 37 °C.

CAMAD

Cell Lines and Culture Conditions

CAMAD includes two cell lines: MDA-MB-231 and RAW 264.7. Table II summarizes the culture conditions for each cell line. Further details are reported in the supplementary material, `camad_experiments_summary.xlsx`. Twenty-four thousand cells of either MDA-MB-231 or RAW 264.7 were seeded in serum-free Leibowitz's L15 medium supplemented with 0.35% BSA (L15-BSA) on each substrate.

Substrate Preparation

All substrates were based on 15×15 mm glass surfaces with polydimethylsiloxane (PDMS) frames to contain cells in culture medium. Various coatings were applied to these substrates as follows.

1) *Matrigel Coating*: $112.5\ \mu\text{L}$ of $100\ \mu\text{g/mL}$ matrigel per substrate, incubated at room temperature for 60–90 min.
2) *Collagen Coating*:
   a) Glass treated with 1N HCl, rinsed, and coated with 0.1 mg/mL polylysine (PLL).
   b) $112.5\ \mu\text{L}$ of $47.5\ \mu\text{g/mL}$ collagen I per substrate, incubated for 60 min.
3) *Cell-Derived Matrix Coating*:
   a) *Disperse macrophage matrix:* 6000 RAW 264.7 cells cultured for seven days.
   b) *Confluent macrophage matrix:* 48 000 RAW 264.7 cells cultured for seven days.
   c) *Confluent breast cancer matrix:* 118 000 MDA-MB-231 cells cultured for seven days.
   d) *Decellularization:* 2M urea treatment for 80–85 min in total.

Fig. 2 illustrates the substrate preparation process. Further details can be found in Section A of the Appendix (available online). This diverse range of substrates allows for the study of cell motility under various extracellular matrix conditions, providing a rich dataset for analyzing cell behavior in different microenvironments.

VALIDATION AND QUALITY

To ensure quality, the data provided here were selected based on the criteria that they can be manually annotated. Manual annotation required a visually satisfactory high signal-to-noise ratio. Spatial resolution was satisfied by using 10× and 40× objectives to give pixel size of $0.454 \times 0.454\ \mu\text{m}$ and of $0.117 \times 0.117\ \mu\text{m}$ for WHAD and CAMAD, respectively. Any artifacts were screened out by the two-tier annotation approach.

Furthermore, our datasets provide higher image resolution than other publicly available datasets in these domains to our knowledge. WHAD and CAMAD have pixel resolutions of 1920×1440 and 2568×1912, respectively. Other datasets published by [3], [5], [6], and [9] have 1392×1040,

TABLE II. Cell Lines and Culture Conditions for CAMAD

Cell Line	Culture Conditions	Collection Method	Culture Medium
MDA-MB-231	37 °C, 5% CO_2	Trypsinization	DMEM + 10% FBS + 1X penicillin-streptomycin + 1X L-glutamine
RAW 264.7		Mechanical collection	RPMI + 5% FBS + 1X penicillin-streptomycin + 1X L-glutamine

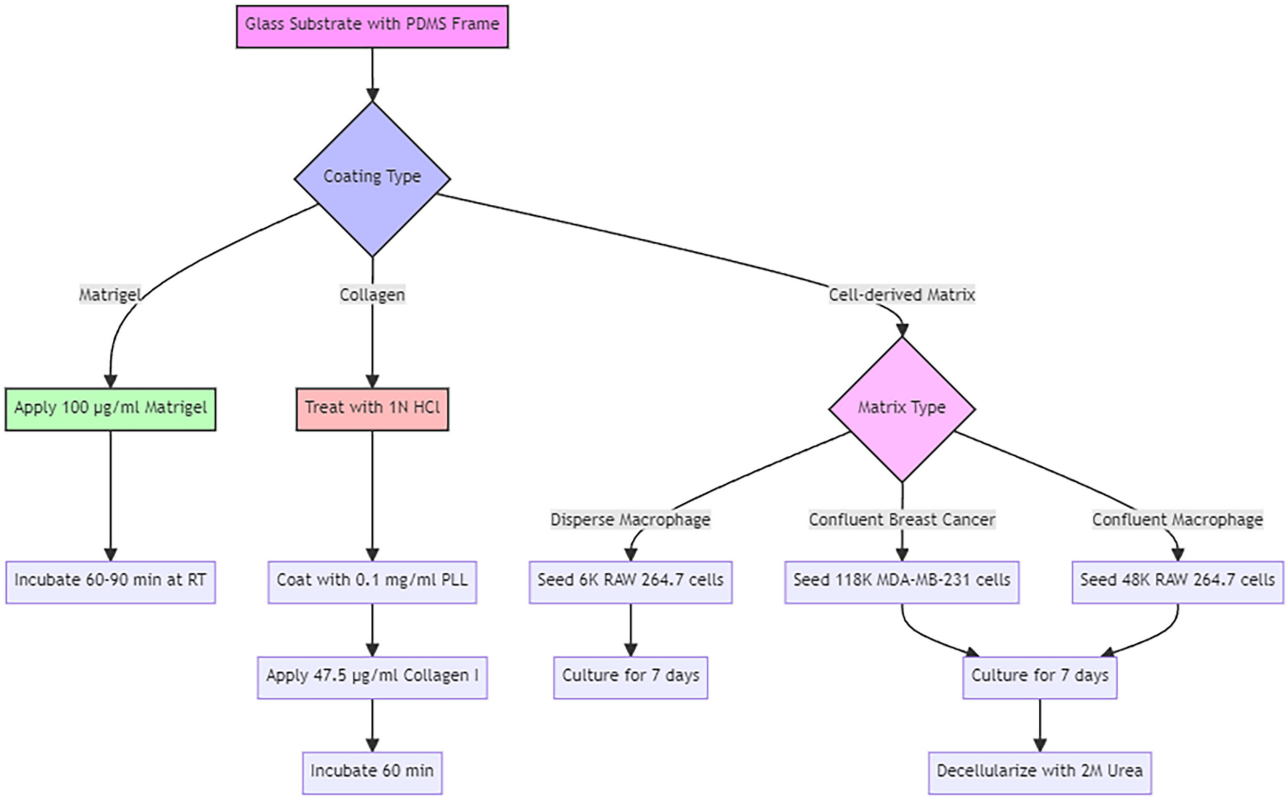


FIG. 2. Substrate preparation process for CAMAD.

1024×1024, 1300×1030, and 1000×1000 pixel resolutions, respectively.

WHAD Annotation

The wound healing assay images were manually annotated using two publicly available tools: ImageJ [4] and Supervisely (https://supervise.ly/). The annotation process involved the following steps.

Open Wound Region Delineation: Each open wound region was annotated as a polygon as demonstrated in Fig. 3.

Expert Supervision: All annotations were performed by two expert annotators who were supervised by field experts for two years while conducting the annotation process and further approved by an expert biologist with over 15 years of experience. Special attention was given to potential confusion between debris and living cells.

Temporal Analysis: The number of annotated frames per experiment varied depending on the healing speed of the wounds.

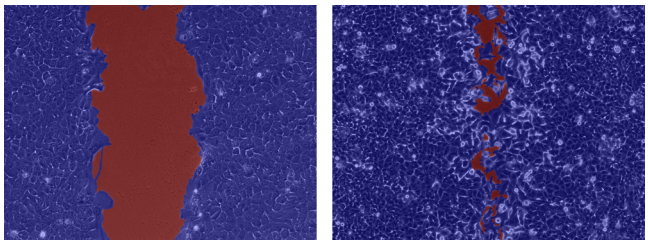
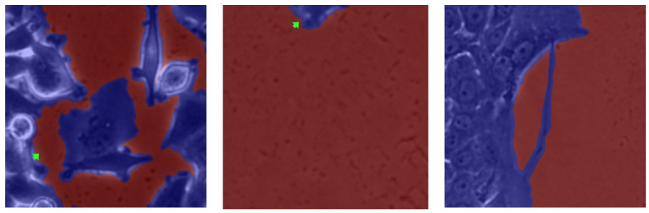


FIG. 3. Examples of WHAD annotation showing wound region delineation. Cell regions are highlighted in blue, and empty wound regions are highlighted in red.

To maintain consistency in the annotations, the following rules, as illustrated in Fig. 4, were established:

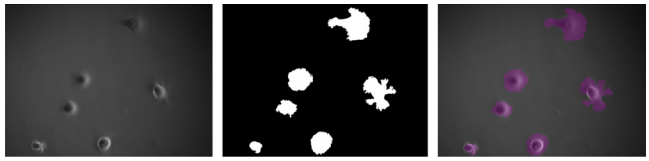
Detached Cell Groups: Isolated cell groups not connected to the main cell mass were annotated and labeled as detached cells as shown in Fig. 4(a).

Incomplete Cells: Cells not fully within the frame as shown in Fig. 4(b) were excluded from annotation to avoid



(a) (b) (c)

FIG. 4. Illustration of annotation rules for WHAD. Cell regions are highlighted in blue, and empty wound regions are highlighted in red. (a) Detached cell groups. (b) Incomplete cells. (c) Empty regions.



(a) (b) (c)

FIG. 5. Example of cell annotation in CAMAD. (a) Original image. (b) Binary mask image. (c) Cells delineated.

introducing bias into the analysis. These cells were also not included in the annotated empty wound region.

Empty Regions: Separate empty regions near the main wound area were annotated as empty wound regions as demonstrated in Fig. 4(c). However, isolated empty regions distant from the main wound were not annotated.

CAMAD Annotation

Manual annotation of cells was carried out using Fiji/ImageJ [4]. Only cells that were not in contact with other cells were included in the analysis. The images were preprocessed twice using the contrast limited adaptive histogram equalization (CLAHE) tool to enhance local contrast.

Cell boundaries were then delineated using the free-hand tool and recorded with the ROI manager. Subsequently, cell area, circularity, and aspect ratio were quantified using the analyze particles tool. Fig. 5 demonstrates the outputs of each process. To ensure accuracy, all manual annotations were performed by at least two independent experts. Annotations were then compared and, if necessary, discrepancies were resolved through reannotation.

RECORDS AND STORAGE

WHAD Records

The WHAD comprises 2-D grayscale PCM time-series images captured during wound healing assays under diverse conditions, along with corresponding manual annotations. The dataset is organized into two primary directories.

1) `images`: It contains the raw assay images in TIFF format.
2) `masks`: It contains the corresponding binary mask images in PNG format, where white pixels delineate cell regions and black pixels represent empty space.

File Naming Conventions

1) *Assay Images*: `assay\_name\_frame\_xxx.tif` where: `assay\_name` reflects the specific assay conditions (e.g., cell line, treatment, and time point) and `frame\_xxx` indicates the frame number in the time series.
2) *Mask Images*:
   a) *Main cell group masks*: `assay\_name\_frame\_xxx\_mask.png`. These masks cover the two primary cell groups involved in wound closure.
   b) *Detached cell masks*: `assay\_name\_frame\_xxx\_mask\_Cyy.png`. These masks individually delineate detached cell regions within the same frame, with `Cyy` being a unique identifier for each detached cell region (e.g., C01 and C02) which is consistent throughout the time series.

It is important to note that each assay image may have multiple corresponding mask images if detached cells are present. To create a single mask for all cells in a frame, the individual mask images can be combined using a logical OR operation. Furthermore, assays are grouped into subfolders within the main "MCF7" and "MCF10A" directories based on cell line and experimental conditions, enhancing organization and searchability.

CAMAD Records

The CAMAD dataset comprises time-lapse videos, corresponding annotated image frames, and the associated editable region of interest (ROI) files.

Video Records

The `videos` directory houses time-lapse videos in AVI format, capturing 16 experiments. Each video is reconstructed at 20 frames per second (FPS) and consists of 600 frames, spanning 5 h of real-time observation. Video filenames adhere to the convention `exp[experiment number]\_[additional details].avi`, e.g., `exp1\_glassmatrigel231liveimaging5aug.20fps.avi`.

Image and Mask Records

The `images` directory stores the annotated frames extracted from the videos. Within the subdirectory of each experiment (e.g., `exp1`), two further subdirectories exist.

1) `images`: It contains the annotated image frames in TIFF format (`exp[experiment number]\_[frame number].tif`).
2) `masks`: It contains binary masks corresponding to the annotated regions in PNG format (`exp[experiment number]\_[frame number].png`).

ROI Records

The `rois` directory contains zipped archives (`exp[experiment number]\_roi.zip`) of the editable ImageJ ROI files for each annotated frame.

Annotation Details

Annotations commence at the beginning of each video and are made at 10-min intervals for the initial 50 min. If needed, the interval can be varied as in experiment 2. The resultant order of annotated frames per experiment may differ due to the selected annotation intervals. The annotated frames for each video can be found in the supplementary material, `camad_experiments_summary.xlsx`.

Storage Details and Access Information

Both datasets are freely available to use for research purposes and are stored in [22] and [23].

INSIGHTS AND NOTES

Advanced Research Opportunities

The WHAD and CAMAD datasets are crucial for advancing cell motility and wound healing research. They enable the development and validation of image processing algorithms for cell segmentation and tracking in phase-contrast microscopy. Researchers can use these datasets to analyze the effects of genetic modifications on wound healing dynamics, particularly in breast cancer [10], [16], and investigate the impact of different extracellular matrix compositions on cell adhesion and motility [14].

Detached Cell Analysis in WHAD

The WHAD dataset offers unique insights into detached cell behavior during wound healing assays. Epithelial cells usually migrate collectively due to cell-cell adhesion, but migration patterns can change with cell type and extracellular environment [1]. For example, epithelial-mesenchymal transition (EMT) alters both cell morphology and migration behavior [7].

In wound healing assays, most epithelial cells migrate collectively, making it easy to analyze cell migration by calculating changes in the open area. However, cells undergoing EMT may detach and migrate individually. The detailed annotations of the WHAD dataset facilitate the analysis of this behavior, revealing important information about migration dynamics and EMT. This is particularly valuable in cancer metastasis studies, where changes in migration patterns are associated with disease phenotype. In addition, analysis of changes in the migration behavior could be utilized in developmental biology research where the transition between epithelial and mesenchymal phenotypes is frequently observed [7]. Thus, including detached cell masks in WHAD enables comprehensive analyses, potentially leading to new insights in cancer and developmental biology research.

Applications in Drug Testing and Cancer Research

These datasets are valuable for drug testing and cancer research, serving as baseline controls for high-throughput screening of compounds affecting cell motility or wound healing. In cancer research, they help develop predictive models for cancer cell invasiveness based on motility patterns [14], [17]. Similar to other datasets [2], [3], [5], [6], [8], [9], [12], [15], [19], they also provide a foundation for automated systems to assess cellular health and behavior, aiding diagnostics, early stage toxicology screening and personalized medicine approaches [18].

Limitations and Future Directions

While WHAD and CAMAD offer extensive data, they could be further expanded with future work using a wider variety of cell types and experimental conditions. The latter can be achieved by the high applicability of the sample preparation and data collection methods described in this work. Integrating other imaging modalities, such as fluorescence microscopy, could provide complementary molecular-level insights. Long-term studies beyond the current time frames could offer valuable information on cell behavior over extended periods.

Overall, WHAD and CAMAD significantly contribute to cell biology and biomedical imaging research. Their applications extend beyond their original scope, offering opportunities for groundbreaking cellular and molecular research in developmental biology and cancer. Developing standardized benchmarks and challenges based on these datasets could drive innovation in biomedical image analysis to advance our understanding of cellular behavior in health and disease.

SOURCE CODE AND SCRIPTS

The source code and scripts for visualization and video processing are freely available at [21].

ACKNOWLEDGMENT

The authors thank Zeynep Ulger and Simge Yucel for their help with manual cell tracking, and Aydin Ayanzadeh and Berkay Mayali for organizing mask and image directories. The authors also thank Muhammed Balikci for contributing to data annotation.

L.O.I. and S.O. led the conceptualization of the study. L.O.I. and Y.S.E. contributed to software development and visualization. Y.S.E. and M.U. were involved in data preparation. S.O., O.Y.-O., and D.P.-O. handled data curation, with O.Y.-O. and D.P.-O. also conducting investigation and validation. B.U.T. and D.U. conducted formal analysis and provided supervision. L.O.I., S.O., Y.S.E., M.U., O.Y.-O., D.P.-O., and D.U. contributed to the writing process. All authors reviewed the final manuscript.

The authors have declared no conflict of interest.

REFERENCES

[1] P. Friedl and K. Wolf, "Plasticity of cell migration: A multiscale tuning model," *J. Cell Biol.*, vol. 188, no. 1, pp. 11–19, 2010, doi: 10.1083/jcb.200909003.

[2] S. Huh, D. F. E. Ker, R. Bise, M. Chen, and T. Kanade, "Automated mitosis detection of stem cell populations in phase-contrast microscopy images," *IEEE Trans. Med. Imag.*, vol. 30, no. 3, pp. 586–596, Mar. 2011, doi: 10.1109/TMI.2010.2089384.

[3] A. A. Liu, K. Li, and T. Kanade, "A semi-Markov model for mitosis segmentation in time-lapse phase contrast microscopy image

sequences of stem cell populations," *IEEE Trans. Med. Imag.*, vol. 31, no. 2, pp. 359–369, Feb. 2012, doi: 10.1109/TMI.2011.2169495.

[4] C. A. Schneider, W. S. Rasband, and K. W. Eliceiri, "NIH Image to ImageJ: 25 years of image analysis," *Nat. Methods*, vol. 9, no. 7, pp. 671–675, 2012, doi: 10.1038/nmeth.2089.

[5] D. Gurari et al., "How to collect segmentations for biomedical images? A benchmark evaluating the performance of experts, crowdsourced non-experts, and algorithms," in *Proc. IEEE Wint. Conf. Appl.*, 2015, pp. 1169–1176.

[6] A. Zaritsky, S. Natan, D. Kaplan, E. Ben-Jacob, and I. Tsarfaty, "Live time-lapse dataset of in vitro wound healing experiments," *GigaScience*, vol. 4, no. 1, 2015, Art. no. s13742-015, doi: 10.1186/s13742-015-0049-6.

[7] M. A. Nieto, R. Y. J. Huang, R. A. Jackson, and J. P. Thiery, "EMT: 2016," *Cell*, vol. 166, no. 1, pp. 21–45, 2016, doi: 10.1016/j.cell.2016.06.028.

[8] D. F. E. Ker et al., "Phase contrast time-lapse microscopy datasets with automated and manual cell tracking annotations," *Sci. Data*, vol. 5, no. 1, pp. 1–12, 2018, doi: 10.1038/sdata.2018.237.

[9] A. V. P. Bobadilla et al., "In vitro cell migration quantification method for scratch assays," *J. R. Soc. Interface*, vol. 16, no. 151, 2019, Art. no. 20180709, doi: 10.1098/rsif.2018.0709.

[10] M. Ilhan et al., "Pro-metastatic functions of Notch signaling is mediated by CYR61 in breast cells," *Eur. J. Cell Biol.*, vol. 99, nos. 2–3, 2020, Art. no. 151070, doi: 10.1016/j.ejcb.2020.151070.

[11] B. Mayali, O. Şayliğ, Ö. Yalçin-Özuysal, D. Pesen-Okvur, B. U. Töreyin, and D. Ünay, "Automated analysis of wound healing microscopy image series—A preliminary study," in *Proc. Med. Technol. Congr. (TIPTEKNO)*, Piscataway, NJ, USA: IEEE, 2020, pp. 1–4.

[12] A. Suarez-Arnedo, F. T. Figueroa, C. Clavijo, P. Arbeláez, J. C. Cruz, and C. Muñoz-Camargo, "An image J plugin for the high throughput image analysis of in vitro scratch wound healing assays," *PLoS One*, vol. 15, no. 7, 2020, Art. no. e0232565, doi: 10.1371/journal.pone.0232565.

[13] S. Kauanova, A. Urazbayev, and I. Vorobjev, "The frequent sampling of wound scratch assay reveals the 'opportunity' window for quantitative evaluation of cell motility-impeding drugs," *Front. Cell Dev. Biol.*, vol. 9, 2021, Art. no. 640972, doi: 10.3389/fcell.2021. 640972.

[14] S. Onal, M. Turker-Burhan, G. Bati-Ayaz, H. Yanik, and D. Pesen-Okvur, "Breast cancer cells and macrophages in a paracrine-juxtacrine loop," *Biomaterials*, vol. 267, 2021, Art. no. 120412, doi: 10.1016/j.biomaterials.2020.120412.

[15] Y. T. Su, Y. Lu, J. Liu, M. Chen, and A. A. Liu, "Spatio-temporal mitosis detection in time-lapse phase-contrast microscopy image sequences: A benchmark," *IEEE Trans. Med. Imag.*, vol. 40, no. 5, pp. 1319–1328, May 2021, doi: 10.1109/TMI.2021.3052854.

[16] Z. E. Gunyuz et al., "SEMA6D differentially regulates proliferation, migration, and invasion of breast cell lines," *ACS Omega*, vol. 7, no. 18, pp. 15769–15778, 2022, doi: 10.1021/acsomega.2c00840.

[17] Y.-N. Lin and A. Wellstein, "Cell migration," in *Cell Movement in Health and Disease*, M. Schnoor, L.-M. Yin, and S. X. Sun, Eds. Academic Press, 2022, ch. 5, pp. 67–82. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780323901956000115

[18] Y. S. Erdem et al., "Automated analysis of phase-contrast optical microscopy time-lapse images: Application to wound healing and cell motility assays of breast cancer," in *Diagnostic Biomedical Signal and Image Processing Applications With Deep Learning Methods (Intelligent Data-Centric Systems)*, K. Polat and S. Öztürk, Eds., Cambridge, MA, USA: Academic Press, 2023, ch. 7, pp. 137–154. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780323961295000135

[19] M. Maška et al., "The cell tracking challenge: 10 years of objective benchmarking," *Nat. Methods*, vol. 20, no. 7, pp. 1010–1020, 2023, doi: 10.1038/s41592-023-01879-y.

[20] Y. S. Erdem et al., "Novel neural style transfer based data synthesis method for phase-contrast wound healing assay images," *Biomed. Signal Process. Control*, vol. 96, 2024, Art. no. 106514, doi: 10.1016/j.bspc.2024.106514.

[21] L. O. Iheme et al., "leonardo-iheme/whad_camad_datasets: Initial whad_camad code release," Jul. 2024. [Online]. Available: https://doi.org/10.5281/zenodo.12805891

[22] L. O. Iheme et al., "Wound healing assay dataset (WHAD) and cell adhesion and motility assay dataset (CAMAD)," Jul. 2024. [Online]. Available: https://doi.org/10.5281/zenodo.12806149

[23] L. O. Iheme et al., "Wound healing assay dataset (WHAD) and cell adhesion and motility assay dataset (CAMAD)," 2024. [Online]. Available: https://dx.doi.org/10.21227/f77j-j355

[24] M. Ucar et al., "Blank frame and intensity variation distortion detection and restoration pipeline for phase-contrast microscopy time-lapse images." *Electrica*, vol. 24, no. 1, pp. 60–66, 2024, doi: 10.5152/electrica.2024.23030.

[25] R. Yazdi and H. Khotanlou, "A survey on automated cell tracking: Challenges and solutions," *Multimedia Tools Appl.*, vol. 83, no. 34, pp. 81511–81547, 2024, doi: 10.1007/s11042-024-18697-9.