Strengthening Scientific Integrity: Digital Forensics for Biomedical Research Imaging

João Phillipe Cardenuto*
Universidade Estadual de Campinas
Campinas-SP, Brasil
Email: phillipe.cardenuto@ic.unicamp.br

Daniel Moreira
Loyola University Chicago
Chicago-IL, USA

Anderson Rocha
Universidade Estadual de Campinas
Campinas-SP, Brasil

***Abstract*—To combat the increasing number of misconduct cases in science, this Ph.D. research addressed the challenge of scientific integrity with a pioneering investigation into digital forensic analysis specifically tailored for biomedical research imaging. This work conducted extensive research into key manipulation types—copy-move forgery, image reuse, and AI-generated content—developing novel, fully explainable, and auditable computational detection methods for each. In a commitment to transparency and to promote research in the area, these techniques are provided as open-source resources. Besides the isolated techniques for each type of image forged, a central contribution is the development of an end-to-end system, created through collaboration with international forensic experts and the U.S. Office of Research Integrity (ORI). This system automates the analysis of scientific publications, starting from PDF documents and ending by identifying figures with potential integrity concerns.**

I. Introduction

Scientific image misconduct has been an increasing concern to science. After manually screening over 20,000 biomedical papers, Bik *et al.* [1] found inappropriate image duplication in approximately 4%, with half showing clear signs of deliberate manipulation. Bucci [2] analyzed the PubMed Central repository[1] [3], concluding that about 6% of its articles contained manipulated images. Furthermore, Acuna *et al.* [4], employing a duplication detection framework on over 760,000 articles from PubMed Open Access, identified inappropriate duplications in 9% of the biomedical publications of this repository.

While image editing software (e.g., Photoshop) contributes to individual cases [5], the landscape of scientific misconduct has grown more complex with the emergence of systematic, fraudulent paper production, referred to as "paper mills." Christopher [6], a scientific journal editor, identified multiple submissions across different topics featuring suspiciously similar figures, suggesting systematic fabrication. Subsequently, Dr. Bik and others confirmed the prevalence of paper mills, compiling a list of hundreds of suspect articles [7]–[9]. These articles were often identified by figure reuse, which was subjected to post-processing (e.g., color changes, cropping, rotation). Their efforts identified over 600 potentially fraudulent papers and led to numerous retractions. More recently, Nature News shared an estimate that at least 400,000 articles have been produced by paper mills [10], highlighting the scale and organized nature of the problem.

Furthermore, the advancement of generative artificial intelligence (AI) could make the scenario even worse. Models capable of creating highly realistic fake images from simple text prompts pose a significant future challenge to integrity. Recently, Qi *et al.* [11] demonstrated that Generative Adversarial Networks (GANs) can synthesize high-quality Western blot images—a widely used biological image used for protein analysis—that are indistinguishable from authentic ones, even to experienced biomedical researchers. Going further, this capability could potentially be exploited to fabricate entire scientific figures.

Despite this challenging landscape, **the primary line of defense in many journals and research integrity offices still relies heavily on the manual screening of images –** a laborious process that is susceptible to human error and severely limited in scalability. While a few researchers attempt to explore automated tools (*e.g, [12]*), the development and widespread adoption of validated, accessible solutions have lagged. Proprietary tools have emerged, offering potential assistance, but often lack rigorous, systematic evaluation, and might pose a false sense of security [13], [14]. This absence of validated, transparent tools underscores the urgent need for reliable, automated methods to safeguard scientific integrity.

Research Contributions

Addressing this critical challenge, our Ph.D. thesis presents a pioneering investigation into biomedical image integrity through the lens of digital forensics. This research organized the types of scientific manipulation and developed several novel forensic detectors to tackle scientific misconduct. These detectors employ diverse approaches, including **image manipulation localization, image provenance analysis, and the identification of AI-generated images.**

A key contribution is an integrated system, named SILA [15], incorporating multiple image analysis methods, developed through international partnerships between institutions in the USA (including collaboration with the Office of Research Integrity - ORI), Italy, and UNICAMP (Brazil). Furthermore, the thesis assesses Western blot image quality, a type of image frequently associated with integrity concerns. Our analysis revealed that approximately 87% of 90,000

*This work is related to the first author's Ph.D. Thesis.

[1]PubMed Central: An extensive repository of biomedical papers.

Western blots–claimed as "raw data"–extracted from publications exhibit significant quality issues, often stemming from compression or post-processing artifacts, which can render subsequent forensic analysis inconclusive.

In a commitment to advancing the field and fostering collaborative solutions to this critical challenge, the developed software tools and datasets generated during this research have been made freely available as open-source resources.

In the next three sections, we will detail three core solutions (one per section) developed during our Ph.D. research, all published in leading journals or conferences.

II. Identifying systematic scientific frauds by provenance analysis

Integrity researchers and whistleblowers have been increasingly concerned about fraudulent organizations, known as paper mills, which illegally mass-produce scientific articles for profit [9]. Although multiple instances of paper mills have been exposed through manual screening, no digital forensic solutions have been developed to track articles from such entities.

Our key idea for detecting such systematic fraud is that these organizations repeatedly reuse and manipulate the same scientific images across different articles [17]. Exploiting such a pattern, we designed and developed an end-to-end image provenance analysis solution to track images and documents produced by paper mills.

The solution begins with a large collection of suspect PDF articles and ends by identifying reused and manipulated images across large article sets. We evaluated our solution using a dataset reported by Dr. Elisabeth Bik and other investigators [8]. To test our method, we added thousands of distractor documents to Dr. Bik's collection, recognizing that in real-world scenarios, paper mill articles represent only a small fraction of the total investigated collection. **As a result, our method effectively identified all reported paper mill articles.**

The core of our provenance analysis solution, summarized in Figure 1, operates after the initial preprocessing stages (also developed by our research). During the preprocessing, figures are automatically extracted from PDF documents. Then, these extracted figures undergo parsing and content filtering, identifying biomedical images and discarding graphs, drawings, and related content. Subsequently, a Convolutional Neural Network (CNN) is employed to generate descriptive embeddings for each figure's content. These resulting embeddings are then indexed and stored within a central database, as depicted at the top of Figure 1.

After pre-processing, a parallel analysis is initiated for each figure indexed in the database (see Figure 1). This involves performing a similarity search to identify the top-k most similar figures within the database based on their embeddings. The content of the query figure is then matched against these retrieved top-k similar images. A content matching score, quantifying the percentage of overlapping or similar area between the query and retrieved images, is calculated for each pair.

Based on these pairwise content matching scores, an adjacency matrix is constructed, representing the relationships between figures. Edges in this matrix connect figures with significant content similarity, weighted by their matching score. Groups of figures exhibiting content reuse are identified as connected components within this graph structure. Finally, by computing the Maximum Spanning Tree (MST) for each identified group, we generate the resulting provenance graphs. These graphs explicitly track the reuse and potential manipulation of figures across different publications.

This solution and dataset have been published in PlosONE:

- **Cardenuto, J.P.**, Moreira, D., and Rocha, A. (2024). *"Unveiling Scientific Articles from Paper Mills with Provenance Analysis"*, PlosONE, 19(10): e0312666, https://doi.org/10.1371/journal.pone.0312666

III. Artificial intelligence-generated scientific image detection and source attribution

Integrity experts expect that, sooner or later, paper mills will use artificial intelligence technology to scale their production with never-before-seen images capable of fooling even specialists in the biomedical field. In this context, our Ph.D. thesis performed an in-depth analysis of artificially generated images to investigate how to detect synthetic scientific images. As a result, we have found that **current generative models include unique and identifiable artifacts that could be used for AI image identification.** By exploiting such artifacts, we proposed two new AI image detection methods that are fully explainable based on Fourier and texture-based analyses. The proposed methods outperformed the current state-of-the-art methods based on handcrafted or deep-learning techniques on a dataset of synthetic Western blot images generated by Generative Adversarial Networks (GANs) and Diffusion-based models.

Figure 4 illustrates the exploited artifacts. These patterns, often referred to as checkerboard artifacts due to their visual appearance, are typically introduced during the upsampling stages (frequently involving transposed convolutions, i.e., deconvolutions) within generative models. This stage, which is responsible for transforming latent representations (embeddings) into the pixel-based RGB image space, is a common component in all image generation architectures. The specific implementation of this upsampling process can leave distinct, often periodic, fingerprints in the resulting image, which serve as detectable indicators of synthetic content.

The checkerboard artifacts are particularly pronounced in the Fourier domain, where they often appear as unique energy peaks in the magnitude spectrum (check Figure 5). Based on this observation, we propose a detection approach centered on Fourier-based analysis, specifically by extracting and analyzing these energy peaks from an input image.

Furthermore, our investigation revealed that texture features, quantified by Gray Level Co-occurrence Matrix (GLCM) analysis, are also consistently altered in AI-generated images,

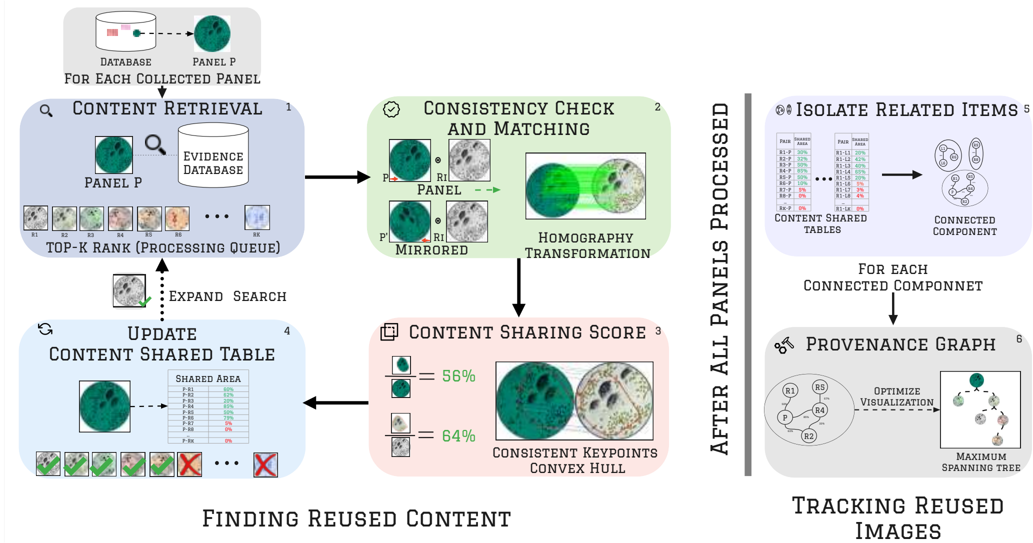

Fig. 1. Overview of the proposed provenance analysis pipeline. For each query figure, a similarity search retrieves the top-k most similar figures from the database. Pairwise content matching then quantifies the similarity between the query and retrieved figures. An adjacency matrix, weighted by these similarity scores, captures the relationships between figures. Groups of related figures are identified as connected components within this graph. Finally, Maximum Spanning Trees (MSTs) are computed for each group to construct the provenance graphs, indicating the reuse and potential manipulation across publications. Figure reproduced from [16] under Creative Commons license.

likely due to artificial patterns (e.g., the checkerboard) introduced during the image generation process.

The proposed workflow, summarized in Figure 5, outlines our detection methodology. It initiates with the extraction of residual noise from the input image, a step designed to enhance the subtle, low-level artifacts of synthetic content. The image residual is then subjected to both Fourier domain analysis (examining energy peak characteristics) and texture analysis (extracting informative GLCM features).

Our evaluation demonstrates that both fourier and texture-based features can be used for a one-class classification approach. By training a classifier, specifically Probabilistic Principal Component Analysis (PPCA) [18], using only features extracted from pristine (non-AI-generated) images, we were able to effectively distinguish AI-generated images from authentic ones based on deviations from the learned pristine distribution.

The findings, method, and dataset from this front were published in the paper:

- **Cardenuto, J.P.**, Mandeli, S., Moreira, D., Bestagini, P., Delp, E., and Rocha, A. (2024). *"Explainable Artifacts for Synthetic Western Blot Source Attribution"*, IEEE International Workshop on Information Forensics and Security (WIFS), Rome, Italy. doi: 10.1109/WIFS61860.2024.10810680.

IV. A System for Scientific Image Analysis - SILA

Another key outcome of this research, developed through collaboration with the Office of Research Integrity (ORI-USA) and an international team of digital forensic researchers, is a system for image analysis (SILA) – the first open-source system specifically designed for scientific image examination within the context of research integrity.

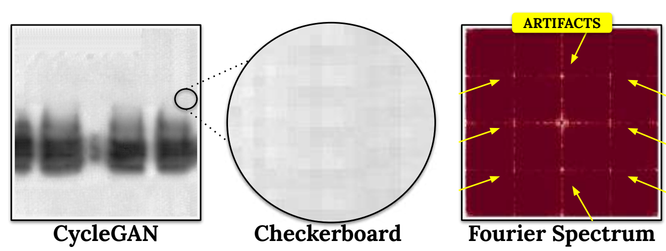

Fig. 2. Checkerboard Artifacts

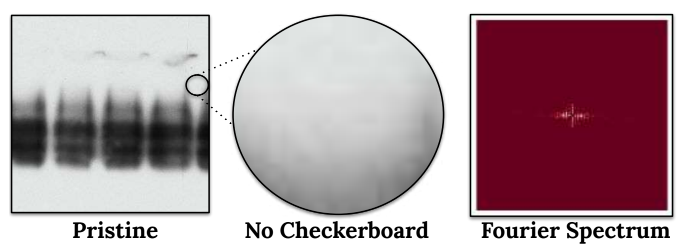

Fig. 3. No Artifacts

Fig. 4. Comparison between a CycleGAN (a) and a pristine (b) Western blot image. The CycleGAN image contains checkerboard artifacts visible when zooming into the image. The highlighted Fourier spectrum peaks (see the yellow arrows) also indicate the presence of those artifacts. Image reproduced from © 2024 IEEE International Workshop on Information Forensics and Security (WIFS) [19] article.

SILA implements an end-to-end pipeline employing fully explainable and auditable methods, a critical requirement for investigations in this sensitive domain. Starting with a collection of articles in PDF format, SILA automatically extracts figures, identifies visually similar image panels across the

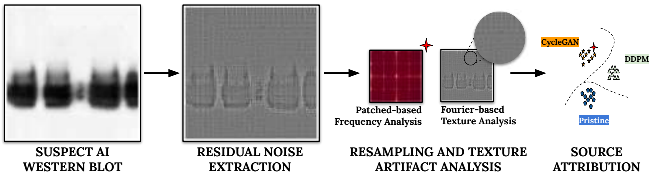

Fig. 5. Proposed workflow. Given a questioned Western blot, we perform residual noise extraction, fourier artifacts, and texture analysis to perform synthetic image detection. Image reproduced from © 2024 IEEE International Workshop on Information Forensics and Security (WIFS) [19] article.

collection, and performs both copy-move forgery detection and image provenance analysis to trace image reuse and manipulation.

The individual modules were rigorously evaluated using a custom dataset curated by our research team. This dataset contains actual documented cases of image manipulation within the biomedical literature. Ground truth annotations were established through a consensus process involving multiple collaborators, guided by the information provided in the official retraction notices for the manipulated articles. Due to copyright restrictions, the original figures and articles used to construct this dataset cannot be publicly released. However, to promote reproducibility and further research, the dataset's comprehensive metadata and annotations are publicly available at github.com/danielmoreira/sciint/tree/dataset (Last accessed: March 31, 2025).

Both the dataset and SILA system have been published in Scientific Reports - Nature:

- Moreira, D., **Cardenuto, J.P.**, et al. *SILA: a system for scientific image analysis.* Scientific Reports - Nature, 12, 18306 (2022). https://doi.org/10.1038/s41598-022-21535-3

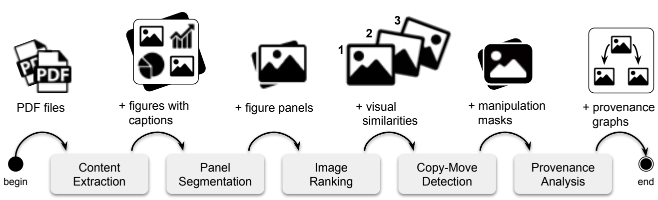

Fig. 6. SILA's workflow starts with PDF figure extraction and ends with provenance analysis. Figure reproduced from [15] under Creative Commons license.

V. Conclusions and Future Work

In conclusion, this work investigated the core problems of scientific image integrity from a forensic perspective. The proposed solutions open numerous research opportunities for future work facilitated by freely available datasets and methods developed by the herein research, including:

- Image quality assessment;
- Image and document provenance analysis;
- Fully-explainable AI-generated image detection;
- Copy-move forgery detection;
- A system for image analysis;

Besides promoting a new generation of integrity methods, we hope our research stimulates new discussions to understand the current quality of scientific images, a comprehensive understanding of the limitations of existing integrity methods, and potential guidelines based on forensic and integrity knowledge that publishers and integrity offices can feasibly implement.

VI. Research Accomplishments

The main accomplishments of this research are:

- Four journal publications directly related to digital forensics and scientific integrity [20], [21], [15], [22];
- One journal publication related to digital forensics and synthetic realities [23];
- One conference paper related to digital forensics and scientific integrity [19];
- Three new datasets/benchmarks for digital forensics applied to scientific integrity [24], [16], and the SILA dataset available at https://github.com/danielmoreira/sciint/tree/dataset;

The full list of publications is summarized in Table I.

In recognition of the excellence of this Ph.D. thesis, our research has earned the Google Latin America Award **(LARA)** [25]. The Google LARA Award recognizes innovative and cutting-edge computer science research projects that impact science and society, created in Latin America. Additionally, this Ph.D. thesis has been selected for presentation in the *Congresso da Sociedade Brasileira de Computação* (CSBC) 2025 thesis award, which means it is among the top 10 best computer science theses defended in Brazil in 2025.

Besides that, our research has also been featured in multiple media articles, as enumerated in Table II.

Finally, we thank Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) for funding this research, which was fundamental to accomplishing all the listed achievements.

References

[1] E. M. Bik, A. Casadevall, and F. C. Fang, "The prevalence of inappropriate image duplication in biomedical research publications," *mBio*, vol. 7, no. 3, Jul. 2016. [Online]. Available: https://doi.org/10.1128/mbio.00809-16

[2] E. M. Bucci, "Automatic detection of image manipulations in the biomedical literature," *Cell Death & Disease*, vol. 9, no. 3, Mar. 2018. [Online]. Available: https://doi.org/10.1038/s41419-018-0430-3

[3] NCBI Resource Coordinators, "Pubmed: the database," National Center for Biotechnology Information [Internet], 2005, available from: https://www.ncbi.nlm.nih.gov/books/NBK153385/. Accessed on June 2024.

[4] D. E. Acuna, P. S. Brookes, and K. P. Kording, "Bioscience-scale automated detection of figure element reuse," *bioRxiv*, Feb. 2018, available at https://doi.org/10.1101/269415 (Access March 2025).

[5] M. Rossner and K. M. Yamada, "What's in a picture? the temptation of image manipulation," *Journal of Cell Biology*, vol. 166, no. 1, pp. 11–15, Jul. 2004. [Online]. Available: https://doi.org/10.1083/jcb.200406019

[6] J. Christopher, "Systematic fabrication of scientific images revealed," *FEBS Letters*, vol. 592, no. 18, pp. 3027–3029, Sep. 2018. [Online]. Available: https://doi.org/10.1002/1873-3468.13201

[7] D. Chawla, "A single 'paper mill' appears to have churned out 400 papers, sleuths find," *Science*, Feb. 2020. [Online]. Available: https://doi.org/10.1126/science.abb4930

TABLE I
SUMMARY OF THE PH.D. PUBLICATIONS

Title	Venue	DOI
Benchmarking Scientific Image Forgery Detectors	Science and Engineering Ethics (2022)	10.1007/s11948-022-00391-4
SILA: a system for scientific image analysis	Nature - Scientific Reports (2022)	10.1038/s41598-022-21535-3
Forensic Analysis of Synthetically Generated Western blot Images	IEEE Access (2022)	10.1109/ACCESS.2022.3179116
The age of synthetic realities	APSIPA Trans. on Signal and Info. Proc. (2023)	10.1561/116.00000138
Unveiling Scientific Articles from Paper Mills	Plos ONE (2024)	10.1371/journal.pone.0312666
Explainable Artifacts for Synthetic Western blot Source Attribution	IEEE WIFS (2024)	10.1109/WIFS61860.2024.10810680

TABLE II
PH.D. MEDIA COVERAGE

Title	Venue
Sistema promete detectar adulterações em imagens de artigos científicos	Revista FAPESP (2023)
Cientista da computação da Unicamp cria método que aponta papers falsos	Revista FAPESP (2023)
Grupo desenvolve ferramentas contra plágio e conteúdo falso	Jornal da UNICAMP (2025)
Pesquisa desenvolve ferramentas contra plágio e conteúdo falso	Programa Trocando em Miúdos - UFU (2025)
Pesquisador da Unicamp desenvolve ferramenta que detecta imagens fraudulentas	G1 - Globo e EPTV (2025)

[8] E. Bik, "The stock photo paper mill," Science Integrity Digest [Internet], 2020, available at https://scienceintegritydigest.com/2020/07/05/the-stock-photo-paper-mill. (Accessed March 2025).

[9] H. Else and R. V. Noorden, "The fight against fake-paper factories that churn out sham science," *Nature*, vol. 591, no. 7851, pp. 516–519, Mar. 2021. [Online]. Available: https://doi.org/10.1038/d41586-021-00733-5

[10] R. Van Noorden, "How big is science's fake-paper problem?" *Nature*, vol. 623, no. 7987, p. 466–467, Nov. 2023. [Online]. Available: http://dx.doi.org/10.1038/d41586-023-03464-x

[11] C. Qi, J. Zhang, and P. Luo, "Emerging concern of scientific fraud: Deep learning and image manipulation," BioRxiv [Preprint], Nov. 2020, available from https://doi.org/10.1101/2020.11.24.395319.

[12] H. Farid, "Exposing digital forgeries in scientific images," in *Proceeding of the 8th workshop on Multimedia and security - MM&Sec '06*. ACM Press, 2006. [Online]. Available: https://doi.org/10.1145/1161366.1161374

[13] M. Rossner, "A false sense of security," *Journal of Cell Biology*, vol. 183, no. 4, pp. 573–574, Nov. 2008. [Online]. Available: https://doi.org/10.1083/jcb.200810172

[14] P. S. Brookes, "Misconduct detection — evolving methods & lessons from 15 years of scientific image sleuthing," *Journal of Law, Medicine & Ethics*, p. 1–11, 2025.

[15] D. Moreira, J. P. Cardenuto, R. Shao *et al.*, "Sila: a system for scientific image analysis," *Scientific Reports*, vol. 12, no. 1, Oct. 2022. [Online]. Available: http://dx.doi.org/10.1038/s41598-022-21535-3

[16] J. P. Cardenuto, D. Moreira, and A. Rocha, "UPM - DATASET," Jul. 2024, available at https://zenodo.org/records/12806479 (Accessed March 2025). [Online]. Available: https://zenodo.org/records/12806479

[17] J. A. Byrne and J. Christopher, "Digital magic, or the dark arts of the 21stcentury—how can journals and peer reviewers detect manuscripts and publications from paper mills?" *FEBS Letters*, vol. 594, no. 4, p. 583–589, Feb. 2020. [Online]. Available: http://dx.doi.org/10.1002/1873-3468.13747

[18] M. E. Tipping and C. M. Bishop, "Probabilistic principal component analysis," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 61, no. 3, pp. 611–622, 01 2002. [Online]. Available: https://doi.org/10.1111/1467-9868.00196

[19] J. P. Cardenuto, S. Mandelli, D. Moreira, P. Bestagini, E. Delp, and A. Rocha, "Explainable artifacts for synthetic western blot source attribution," in *2024 IEEE International Workshop on Information Forensics and Security (WIFS)*, 2024, pp. 1–6. [Online]. Available: https://doi.org/10.1109/WIFS61860.2024.10810680

[20] J. P. Cardenuto and A. Rocha, "Benchmarking scientific image forgery detectors," *Science and Engineering Ethics*, vol. 28, no. 4, Aug. 2022. [Online]. Available: http://dx.doi.org/10.1007/s11948-022-00391-4

[21] S. Mandelli, D. Cozzolino, E. D. Cannas *et al.*, "Forensic analysis of synthetically generated western blot images," *IEEE Access*, vol. 10, pp. 59 919–59 932, 2022. [Online]. Available: https://doi.org/10.1109/ACCESS.2022.3179116

[22] J. P. Cardenuto, D. Moreira, and A. Rocha, "Unveiling scientific articles from paper mills with provenance analysis," *PLOS ONE*, vol. 19, no. 10, p. e0312666, Oct. 2024. [Online]. Available: http://dx.doi.org/10.1371/journal.pone.0312666

[23] J. P. Cardenuto, J. Yang, R. Padilha *et al.*, "The age of synthetic realities: Challenges and opportunities," *APSIPA Transactions on Signal and Information Processing*, vol. 12, no. 1, 2023. [Online]. Available: http://dx.doi.org/10.1561/116.00000138

[24] J. P. Cardenuto and A. Rocha, "Recod.ai scientific image integrity dataset (rsiid)," Aug. 2022, available at https://zenodo.org/records/15095089 (Accessed March 2025). [Online]. Available: https://zenodo.org/records/15095089

[25] Google, "Conheça os vencedores do prêmio lara 2021, o programa de bolsas de pesquisa do google," 2021, available at https://blog.google/intl/pt-br/novidades/iniciativas/conheca-os-vencedores-do-premio-lara-2021-o-programa-de-bolsas-de-pesquisa-do-google (Accessed March 2025).