Strengthening Scientific Integrity: Digital Forensics for Biomedical Research Imaging

João Phillipe Cardenuto[1], **Daniel Moreira**[2], **Anderson Rocha**[1]

[1]Instituto de Computação – Universidade Estadual de Campinas (UNICAMP)
Campinas, SP, Brasil

[2]Department of Computer Science – Loyola University Chicago
Chicago, IL, U.S.A.

{phillipe.cardenuto, anderson.rocha}@ic.unicamp.br dmoreira1@luc.edu

Abstract. *To fight against the increasing misconduct cases in science, this Ph.D. research confronted the challenge of scientific integrity with a pioneering investigation into digital forensic analysis specifically tailored for biomedical images. 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 to 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.*

1. Introduction

Scientific image misconduct has been an increasing concern to science. After manually screening over 20,000 papers, Bik *et al.* [Bik et al. 2016] found inappropriate image duplication in approximately 4%, with half showing clear signs of deliberate manipulation. Bucci [Bucci 2018] analyzed the PubMed Central repository[1] [NCBI Resource Coordinators 2005], concluding that about 6% of its articles contained manipulated images. Furthermore, Acuna *et al.* [Acuna et al. 2018], employing a duplication detection framework on over 760,000 articles from PubMed Open Access, identified inappropriate duplications in 9% of the publications of this repository.

While image editing software (e.g., Photoshop) contributes to individual cases [Rossner and Yamada 2004], the landscape of scientific misconduct has grown more complex with the emergence of systematic, fraudulent paper production, referred to as "paper mills." Christopher [Christopher 2018], 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 [Chawla 2020, Bik 2020, Else and Noorden 2021]. These articles were often identified by figure reuse, which was

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

subjected to simple post-processing (e.g., color changes, cropping, rotation). Their efforts, which identified over 600 potentially fraudulent papers and led to numerous retractions, highlighted the scale and organized nature of the paper mill 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.* [Qi et al. 2020] 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,* [Farid 2006]), 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 [Rossner 2008]. This absence of validated, transparent tools underscores the urgent need for reliable, automated methods to assist in the critical task of safeguarding scientific integrity.

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 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 critically assesses Western blot image quality, a type of image frequently associated with integrity concerns. Our analysis revealed that approximately 87% from 90,000 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. Next, we will detail three core solutions developed during our Ph.D. research, all published in leading journals or conferences.

2. 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. 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 [Byrne and Christopher 2020]. 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 [Bik 2020]. 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 (illustrated in 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

3. 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

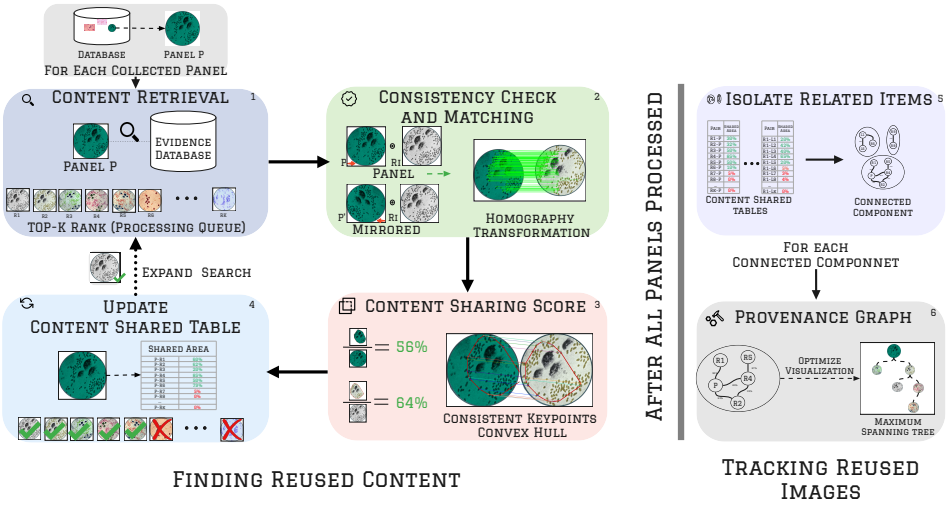

Figura 1. Overview of the proposed figure provenance analysis pipeline. For each query figure, a similarity search retrieves the top-k most similar figures from the database. Pairwise content matching then quantifies the similarity between the query and retrieved figures. An adjacency matrix, weighted 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 [Cardenuto et al. 2024b] under Creative Commons license.

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 using handcrafted or deep-learning techniques on a dataset of synthetic Western blot images generated by Generative Adversarial Networks (GANs) and Diffusion-based models.

Figure 2 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 3). 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.

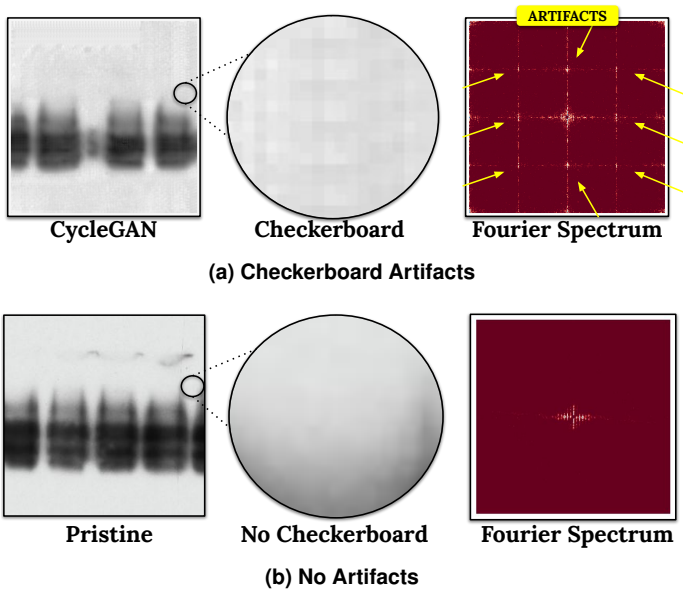

Figura 2. Comparison between a CycleGAN (a) and a pristine (b) Western blot image. The CycleGAN image contains checkerboard artifacts visible when zooming into the image. The highlighted Fourier spectrum peaks (see the yellow arrows) also indicate the presence of those artifacts. Image reproduced from © 2024 IEEE International Workshop on Information Forensics and Security (WIFS) [Cardenuto et al. 2024a] article.

Furthermore, our investigation revealed that texture features, quantified by Gray Level Co-occurrence Matrix (GLCM) analysis, are also consistently altered in AI-generated images, likely due to artifacts introduced during the image generation process (such as deconvolution). These features derived from the GLCM provide an additional, complementary method for distinguishing AI-generated images from pristine ones.

The proposed workflow, summarized in Figure 3, 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) [Tipping and Bishop 2002], using only features extracted from pristine (non-AI-generated) images, we were able to effectively distinguish novel 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.

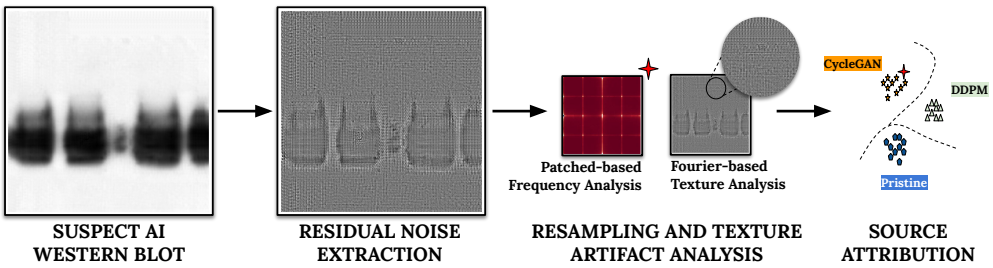

Figura 3. Proposed workflow. Given a questioned Western blot, we perform residual noise extraction, fourier artifacts, and texture analysis to perform synthetic image detection. Image reproduced from © 2024 IEEE International Workshop on Information Forensics and Security (WIFS) [Cardenuto et al. 2024a] article.

4. 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.

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 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.* Sci Rep 12, 18306 (2022). https://doi.org/10.1038/s41598-022-21535-3

5. Research Accomplishments

In summary, the main accomplishments of this research are:

- Four journal publications directly related to digital forensics and scientific integrity [Cardenuto and Rocha 2022a], [Mandelli et al. 2022], [Moreira et al. 2022], [Cardenuto et al. 2024b];
- One journal publication related to digital forensics and synthetic realities [Cardenuto et al. 2023];

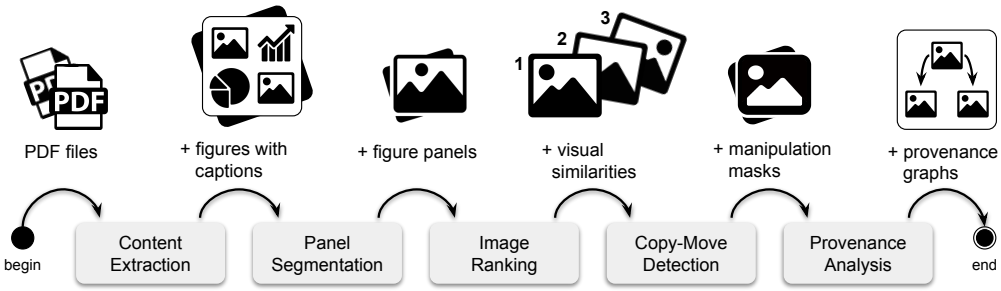

Figura 4. SILA's workflow starts with PDF figure extraction and ends with provenance analysis. Figure reproduced from [Moreira et al. 2022] under Creative Commons Attribution 4.0 International License.

- One conference paper related to digital forensics and scientific integrity [Cardenuto et al. 2024a];
- Three new datasets/benchmarks for digital forensics applied to scientific integrity [Cardenuto and Rocha 2022b], [Cardenuto et al. 2024c], and the SILA dataset available at https://github.com/danielmoreira/sciint/tree/dataset;
- The Google Latin America Award (LARA) [Google 2021].

6. 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, the limitations of existent integrity methods, and potential guidelines based on forensic and integrity knowledge that publishers and integrity offices can feasibly implement.

7. References

Referências

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

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

Bik, E. M., Casadevall, A., and Fang, F. C. (2016). The prevalence of inappropriate image duplication in biomedical research publications. *mBio*, 7(3).

Bucci, E. M. (2018). Automatic detection of image manipulations in the biomedical literature. *Cell Death & Disease*, 9(3).

Byrne, J. A. and Christopher, J. (2020). Digital magic, or the dark arts of the 21stcentury—how can journals and peer reviewers detect manuscripts and publications from paper mills? *FEBS Letters*, 594(4):583–589.

Cardenuto, J. P., Mandelli, S., Moreira, D., Bestagini, P., Delp, E., and Rocha, A. (2024a). Explainable artifacts for synthetic western blot source attribution. In *2024 IEEE International Workshop on Information Forensics and Security (WIFS)*, pages 1–6.

Cardenuto, J. P., Moreira, D., and Rocha, A. (2024b). Unveiling scientific articles from paper mills with provenance analysis. *PLOS ONE*, 19(10):e0312666.

Cardenuto, J. P., Moreira, D., and Rocha, A. (2024c). UPM - DATASET. Available at `https://zenodo.org/records/12806479` (Accessed March 2025).

Cardenuto, J. P. and Rocha, A. (2022a). Benchmarking scientific image forgery detectors. *Science and Engineering Ethics*, 28(4).

Cardenuto, J. P. and Rocha, A. (2022b). Recod.ai scientific image integrity dataset (rsiid). Available at `https://zenodo.org/records/15095089` (Accessed March 2025).

Cardenuto, J. P., Yang, J., Padilha, R., et al. (2023). The age of synthetic realities: Challenges and opportunities. *APSIPA Transactions on Signal and Information Processing*, 12(1).

Chawla, D. (2020). A single 'paper mill' appears to have churned out 400 papers, sleuths find. *Science*.

Christopher, J. (2018). Systematic fabrication of scientific images revealed. *FEBS Letters*, 592(18):3027–3029.

Else, H. and Noorden, R. V. (2021). The fight against fake-paper factories that churn out sham science. *Nature*, 591(7851):516–519.

Farid, H. (2006). Exposing digital forgeries in scientific images. In *Proceeding of the 8th workshop on Multimedia and security - MM&Sec '06*. ACM Press.

Google (2021). Conheça os vencedores do prêmio lara 2021, o programa de bolsas de pesquisa do google. 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).

Mandelli, S., Cozzolino, D., Cannas, E. D., et al. (2022). Forensic analysis of synthetically generated western blot images. *IEEE Access*, 10:59919–59932.

Moreira, D., Cardenuto, J. P., Shao, R., et al. (2022). Sila: a system for scientific image analysis. *Scientific Reports*, 12(1).

NCBI Resource Coordinators (2005). Pubmed: the database. National Center for Biotechnology Information [Internet]. Available from: `https://www.ncbi.nlm.nih.gov/books/NBK153385/`. Accessed on June 2024.

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

Rossner, M. (2008). A false sense of security. *Journal of Cell Biology*, 183(4):573–574.

Rossner, M. and Yamada, K. M. (2004). What's in a picture? the temptation of image manipulation. *Journal of Cell Biology*, 166(1):11–15.

Tipping, M. E. and Bishop, C. M. (2002). Probabilistic principal component analysis. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 61(3):611–622.