MediFor Scientific Integrity Project
Purdue University and USC Information Sciences Institute
System User Manual
Jan 22, 2021
Version 5.0

Ruiting Shao and Edward J. Delp
# Table of Contents

*Introduction* ................................................................................................................................. 2  
*Contact Information and Acknowledgment* .................................................................................. 3  
*Access to Prototype System* ......................................................................................................... 4  
*System Homepage* ......................................................................................................................... 5  
*System Header* ............................................................................................................................... 6  
*User Creation* ................................................................................................................................. 7  
*Case Creation* ................................................................................................................................. 8  
*Upload PDFs/Images* ....................................................................................................................... 10  
*Image and Text Analysis Functions* ............................................................................................. 15  
  *Similarity Ranking* ....................................................................................................................... 17  
  *Panel Extraction* ......................................................................................................................... 18  
  *Simple Image Manipulation Operation* ...................................................................................... 20  
  *Copy/Move Detection* ............................................................................................................... 21  
  *Provenance Analysis* ................................................................................................................... 24  
*Case Reports* ................................................................................................................................ 26  
*Conclusion* .................................................................................................................................... 27
Introduction

Since the earliest days of science, illustrations have been widely used in scientific publications to describe the proposed processes, aid in theoretical interpretation, and present experimental results. Later, with the advent and popularization of photography, photographs (images) were accepted for use in scientific publications, especially for recording experimental results.

Recently, with the transition from classical photography to digital imaging, it is easy to modify and synthesize images with digital image editing tools. Some editing is acceptable in scientific publications, such as intensity calibrations for visual enhancement or synthesis aimed at making comparison of different results easier. Other modifications may be either mistakes (for example, the author inadvertently provides incorrect representations of experimental results) or misconduct (author intends to deceive).

The use of modern image forensics might be useful for examining images in scientific publications to determine if they have been altered. The goal of this project is to use tools and methods developed in the DARPA MediFor\(^1\) program to develop a prototype system for analysts to conduct an investigation into the integrity of a publication or document. A cloud-based scientific integrity system was designed to help analysts examine images extracted from pdf files of scientific publications. The system has a web-based interface that allows an analyst to upload documents, use the image forensics tools, and download results and reports.

Currently, the system incorporates various elements including image extraction, image ranking, panel extraction, manipulation detection, and provenance analysis.

This document is intended to provide a brief introduction on how to use the system.

Please Note: The current system is not a production or commercial product. This is a prototype system to be used as a testbed for developing methods for the analysis of scientific papers and to exploit methods developed by the DARPA MediFor Program. Several aspects of the user interface were suggested by HHS.

We ask users to understand that this is a testbed and as such it is not “finished”. We ask all users to please make suggestions and provide us with your feedback and comments. You can email us at good-science@ecn.purdue.edu with any comments.

\(^1\) https://www.darpa.mil/program/media-forensics
Contact Information and Acknowledgment

This work was performed by the Purdue University and USC ISI teams and was funded by the DARPA MediFor Program (MediFor Program Manager: Dr. Matt Turek)

Purdue Principal Investigator: Edward Delp

USC Principal Investigator: Wael Abd-Almageed

Research Team:

Luisa Verdoliva, UNINA
Walter Scheirer, University of Notre Dame
Patrick Flynn, University of Notre Dame
Anderson Rocha, University of Campinas
Kevin Bowyer, University of Notre Dame
Stefano Tubaro, Politecnico di Milano
Paolo Bestagini, Politecnico di Milano

To contact the team, please send an email to: good-science@ecn.purdue.edu

Acknowledgement

This material is based on research sponsored by the Defense Advanced Research Projects Agency (DARPA) and the Air Force Research Laboratory (AFRL) under agreement number FA8750-16-2-0173. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of DARPA, AFRL, the Department of Defense, the Department of Health and Human Services or the U.S. Government.
**Access to Prototype System**

The current system runs on a server located at Purdue University. A user can access the system using a web browser. *We recommend you use the Firefox browser and turn off any VPN you are using.* A user would then use their login credentials to access their workspace on the system. The login credentials can be requested for access to the system.

To access the current system, go to [https://rose.ecn.purdue.edu/](https://rose.ecn.purdue.edu/)

For a better user experience, we recommend that you clear the browser cache every time you start using the system.

The opening page looks as below:

![Opening page](image)

For first-time users, you can watch the demo video to see how the system works.


The paper set we used for the demo video is at: [https://lorenz.ecn.purdue.edu/~shao72/public/demo-paper-set.zip](https://lorenz.ecn.purdue.edu/~shao72/public/demo-paper-set.zip)

This user manual is at: [https://lorenz.ecn.purdue.edu/~shao72/public/user-manual-scientific-integrity-v5.pdf](https://lorenz.ecn.purdue.edu/~shao72/public/user-manual-scientific-integrity-v5.pdf)
System Homepage

After you log into the system, it will look like this:

The interface of an “ADMIN”

You can check the current users of the system in the “Users” section as seen in “Main Navigation” sidebar.

You can edit your profile in “Edit Profile” section as seen in “Main Navigation” sidebar.
System Header

The system header contains 5 different functions to help a user better use the system.

Func 1: clicking on this area will lead the user back to the homepage

Func 2: clicking this icon will allow the user to minimize or expand the “Main Navigation” sidebar shown on the left

Func 3: clicking this icon will hide or expand the user’s avatar and information in the upper left corner of the homepage (this function is only valid in the homepage)

Func 4: clicking this icon will display the system interface in full screen mode

Func 5: clicking this icon will let the user set their visual preference for the system, such as the display theme and layout
User Creation

Important: Only an “ADMIN” has the right to create a new “USER”.

Step 1: Click the “Create User” button on the homepage

Step 2: Fill in the “User Email” and submit

When the submission is complete, the system will display the message “User created! Please contact sbairedd@purdue.edu to receive the information”.

![Image of User Creation Process]

Step 1

Step 2
Case Creation

Step 1: Click the “Create Case” button

Step 2: Fill the required fields and submit
Once the case is submitted, it will appear on the homepage.
Upload PDFs/Images

Click on the case that you want to investigate. There are two ways to upload data into the system, “Upload Image” and “Upload PDF”.

The acceptable image formats are PNG and JPG.

The acceptable PDFs are ones created by Word or LaTeX.

Note: the uploaded images or PDFs should not share the same file names!

If the user selects “Upload Image”:

Step 1: click the “Pick Image” button in “Upload Image” section

Step 2: select the wanted images from local computer

Step 3: if the selected images are correct, click the “Upload Image” button at the bottom of the page. Otherwise, please click the “Pick Image” button at the bottom of the page to reselect the desired images
If a user wants to upload images from different folders on the local computer, they need to do the same process for each folder.
If the user selects “Upload PDF”:

Step 1: click the “Pick PDF” button in “Upload PDF” section

Step 2: select the wanted PDFs from local computer

Step 3: if the selected PDFs are correct, click the “Upload PDF” button at the top of the page. Otherwise, please click the “Pick PDF” button at the top of the page to reselect the desired PDFs.

If a user wants to upload PDFs from different folders on the local computer, they need to do the same process for each folder.

The system will take a few minutes to complete image and caption extraction from the PDF documents.
Once the upload process is completed, the uploaded images and extracted images along with captions will be displayed in the “Search” section with a check box displayed in the upper left corner of the image. During the upload process, the system also classifies images into different categories. Each image will be classified into one or several classes, and these class labels will be displayed below it. Currently, the system can classify images into 7 different categories, namely graphs, westernblot, flowcytometry, biological, panels, letters and others.

There is an image category selection option at the top of the “Search” section. It is a drop-down list of the 7 image categories listed above. The system will display images according to the classes selected by the user and display all images by default (that is, no image class is selected).
The user can also view the uploaded PDF documents in the “View PDFs” section.
Image and Text Analysis Functions

In the “Search” section, all the images, either extracted from the PDFs or uploaded directly, can be viewed. Each image carries a check box and, if relevant, a link tracing back to the source PDF file, a caption from the source document and image labels.

The user can use the “Delete Selected Images” function to delete the selected unwanted images.

When the cursor is placed on the red text, it will display the corresponding caption extracted from the PDF.

When the user clicks on the blue link, or the source PDF file name, the system will display the source document in another tab.

The user can also modify the image class labels by clicking on the labels.
For each image, the user can click the ellipsis in the upper right corner to perform image similarity ranking, panel extraction, or image deletion.
Similarity Ranking

The system is capable of finding images that are similar to a particular image of interest. Select “Show Similar Images” for an image. The system will rank the remaining images based on a similarity score from high to low.

Then user can go back to the “Search” section by clicking the “Back” button on the similarity result page.
The image similarity ranking process will be performed within the images under the classes selected by the user. By default, the query image is compared with all uploaded and extracted images.

**Panel Extraction**

On selecting “Panel Extraction” for the query image, the system will automatically extract the sub-images from the selected panel. After the panel extraction process is completed, the system will prompt with the message “Panel Extraction Done!”. The extracted images along with the caption, source link and image labels will be displayed in the “Search” section.
The images in the red box are the extracted sub-images from the query image.
Simple Image Manipulation Operation

In current system, users can perform some simple image manipulation operations by clicking on the image to meet their goals.

The system allows the user to make simple adjustments on color, brightness, contrast, exposure, gamma and saturation value. It also allows the user to clip color value or apply an adjustable amount of noise to the image.

The user can also perform histogram equalization and auto contrast on the images. They are one-click functions based on the original image.

Note: Currently, these operations are only helpful for image visualization. Later we will add a function to save the modified image into the audit trail.
Copy/Move Detection

The system is capable of determining if part of an image has been copied and moved to another part of the image.

Step 1: Select one or multiple images to analyze

Step 2: Click the “Image Analysis” button

Step 3: Select the “Copy Move Detection”

Step 4: Click the “Submit” button

Note: in the current version of the system, copy/move manipulations within a single image can be detected.
The system will display the copy/move detection results in two columns. The left column shows the selected images and right column shows the masks indicating the copy move area.
Note: the system will also display the past copy/move detection results in the results page, the results of the latest query images are usually displayed at the bottom of the results page. However, if the query image has already performed the copy/move detection before, the result of this query image will display right after the past results.

For convenience, we recommend users check the copy/move detection results in the “History” section. The history results are arranged by processing time and also indicate the analyst.

Once the copy/move detection is completed, the user needs to click the “Search” section to go back to the image page.
Provenance Analysis

The system can determine the relationship between images from different papers. This is known as provenance analysis.

Step 1: Select the image to be analyzed, only one image can be selected

Step 2: Click the “Provenance Graph” button

The provenance analysis process will be performed within the images under the classes selected by the user. By default, the query image is compared with all uploaded and extracted images.

It will take several minutes for the provenance analysis to be executed by the server. For this analysis, the goal is to find and match feature points from one image to another image. This analysis will eventually generate a graph that indicates the potential similar/identical areas between images.
Case Reports

Once the users complete their analysis, a case report can be generated and exported using the “Case Report” section.

In this section, users can report their findings based on the analysis they have done. They can edit this section as needed and export the report as a PDF (we will also be able to export spreadsheets in the very near future, as well as the various image analysis results).
Conclusion

This document is a brief introduction as to how the system works and will be continuously updated as we develop new tools and user interface enhancements.

We welcome questions, comments, and suggestions.

Please contact us at: good-science@ecn.purdue.edu