PharmaSUG 2023 - Paper SD-311

A Computer Vision Approach to Automate Figure Validation

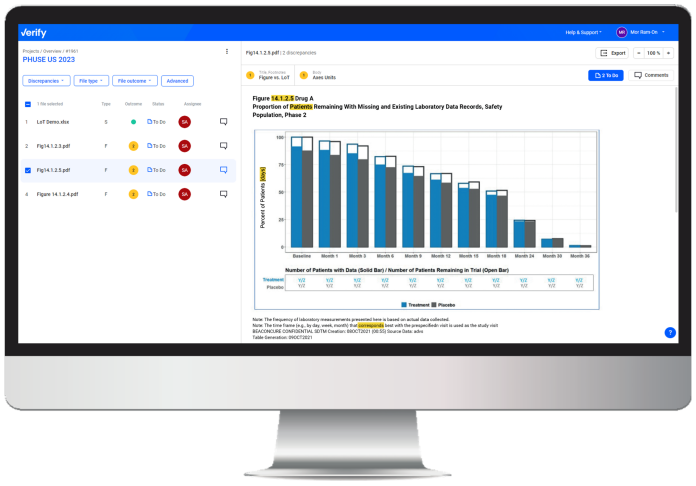
Ilan Carmeli, Beaconcure

Abstract

The validation of tables, listings, and figures (TLFs) is a burdensome and lengthy process. Especially the checking of figures needs a lot of visual comparisons. In order to review and validate the figures, the statistical programmer needs to open the table on one screen and the figures on a second screen and compare it manually. The statistical programmer looks for consistency of the figures data, consistency of the figure format including legend, title, and footnotes, consistency with the mock shells, and accuracy of the X and Y axes. By doing this manually the chance of overlooking discrepancies is high. In this presentation we will present how using AI-based image recognition by using machine learning (ML) and deep learning (DL) methods can help in automating such complex manual tasks and align to the clinical standards.

Introduction

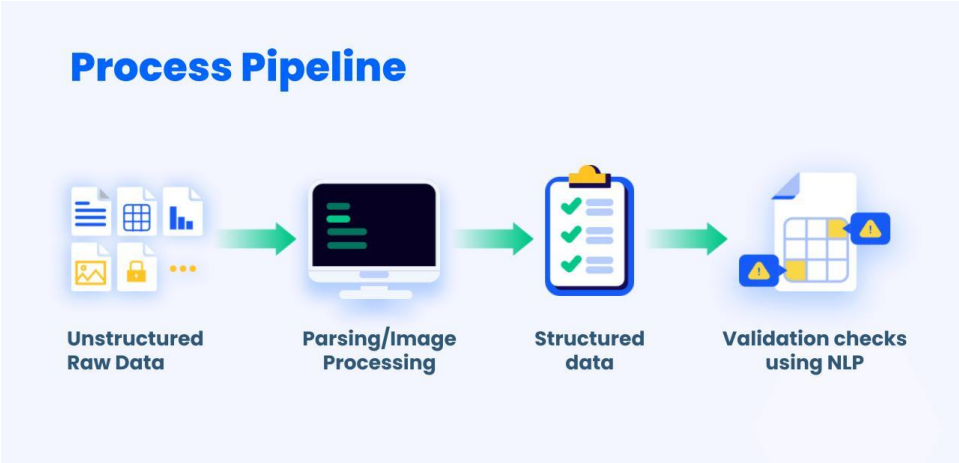
Figures play a critical role in conveying information in scientific and technical papers. In the pharmaceutical industry, the validation of figures is particularly crucial, as it can impact the efficacy and safety of drugs. However, the current process of validating figures is mainly manual and prone to errors. In this paper, we showcase a product called Verify, which based on a computer vision approach, automates the validation of figures. The solution utilizes both format and logical checks to ensure accuracy and reliability. In addition, Verify provides a comprehensive solution to manage the validation process, allowing for collaboration and tracking of changes and issues.



The Importance of Accurate Figures in the Pharmaceutical Industry

Figures consist of a title section, a footer section, and the plot itself, and they are subject to discrepancies that can lead to misinterpretation and erroneous conclusions. Possible discrepancies include mismatches between figures and lists of tables (LoT), mismatches in figure number, title, and source data between figures and mock shells, lack of axes units, and inconsistency in series type, colors, and names in the legend.

To ensure accuracy and reliability, Verify utilizes a computer vision approach to automate the validation process for figures in the pharmaceutical industry. The solution utilizes advanced machine learning and computer vision algorithms to perform format and logical checks. Verify provides an easy-to-use interface for users to upload the outputs they wish to validate. The system then checks the format and logical content of the figure, ensuring that all the data and information presented are accurate and reliable.

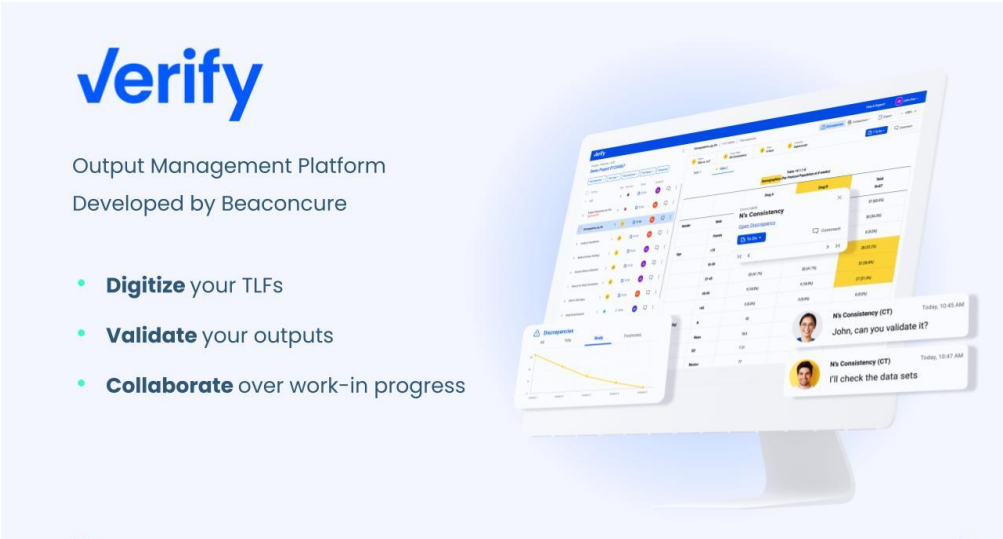


Solution - Verify

Verify utilizes a computer vision solution approach to automate the pharmaceutical industry's validation process for figures. The system employs advanced machine learning and computer vision algorithms to perform format and logical checks.

The solution provides an easy-to-use interface for users to upload outputs they wish to validate. The system checks the logical content of the figure, ensuring that all the data and information presented are accurate and reliable (This means, without Possible discrepancies include mismatches between figures and lists of tables (LoT), mismatches in figure number, title, and source data between figures and mock shells, lack of axes units, and inconsistency in series type, colors, and names in the legend)

In addition, Verify allows users to collaborate on the validation process. The system provides a collaborative workspace where users can comment on the output, manage issues, track changes, and ensure transparency.

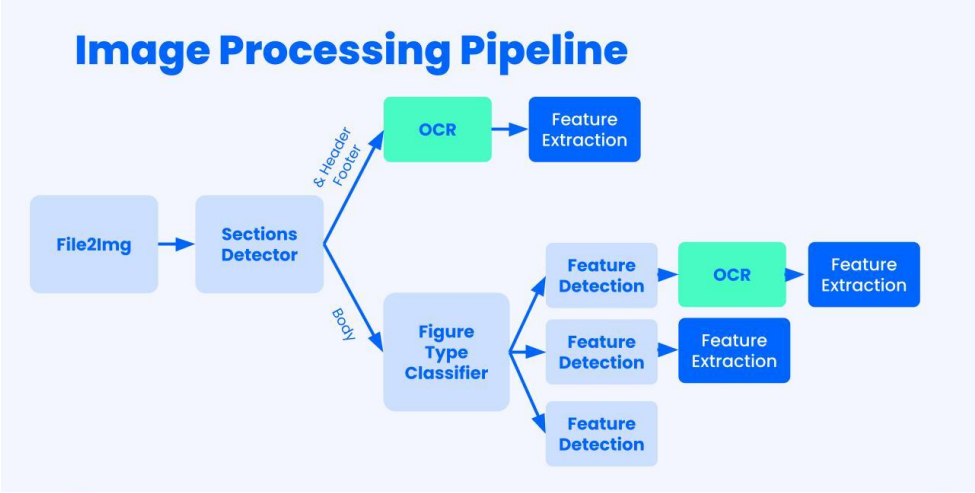


Within-Table, table checks are performed to validate the numeric logic within tables, such as the sum of numeric values and hierarchy checks. An instance of a discrepancy detected in the sum of a mutually exclusive sub-group in the table is demonstrated in the example presented below.

Feature Extraction Pipeline

To automate the validation of figures, we developed a Feature extraction pipeline that includes, detecting and classifying the relevant sections on the page, extracting important features,, classifying the figure type, and extracting text using the OCR technique. Feature extraction techniques are employed to extract specific features of the figure, such as axes titles, legends, and data points. Figure type classification is used to identify the type of figure.

The Verify solution utilizes advanced computer vision approaches to ensure that all aspects of the figure are extracted correctly. The system employs the OCR technique to extract text data from the figure image. The extracted text data is then processed using feature extraction techniques to extract specific features such as axes titles, legends, and data points. The figure type classification technique is used to identify the type of figure and validate it accordingly.



Conclusion

In conclusion, Verify provides an efficient and accurate solution for validating figures, tables, listings, and more in the pharmaceutical industry. Its computer vision approach and advanced parsing pipeline ensure that all aspects of the output are accurately validated, providing a more efficient and accurate validation process. With Verify, the validation process can be completed quickly and accurately, freeing up valuable resources for other critical aspects of drug development.

Additionally, Verify's collaborative workspace allows for greater transparency in the validation process, ensuring that all stakeholders are aware of the status of the validation. This promotes better communication and collaboration between team members, leading to a more efficient and accurate validation process.

In summary, Verify provides an innovative and reliable solution for the validation of figures, tables, listings, and more in the pharmaceutical industry. Its computer vision approach and advanced parsing pipeline ensure that all discrepancies are caught and corrected.

In addition, Verify utilizes feature extraction techniques to extract specific features of the figure, such as axes titles, legends, and data points. This allows Verify to check the content of the figures and their format, ensuring that all information presented is accurate and reliable.

The Verify solution also includes a collaborative workspace that enables users to manage the validation process. This workspace allows users to upload outputs they wish to validate, comment on them, manage issues, track changes, and more. This promotes better communication and collaboration between team members, leading to a more efficient and accurate validation process.

To automate the validation of figures, Verify employs a feature extraction pipeline that includes OCR, feature extraction, and figure type classification. The OCR technique is used to extract text data from the figure image, while feature extraction techniques are employed to extract specific features of the figure. Figure type classification is used to identify the type of figure, such as a scatter plot or a bar graph. This pipeline ensures that all aspects of the figure are accurately validated, providing an added layer of reliability to the validation process.

Information

—

Beaconcure Technologies, Inc.

OUR OFFICES

Israel

Israel Beck st 34

Tel Aviv

USA

117 Kendrick St, Needhm, MA

Email: ilan@beaconcure.com

Web: www.beaconcure.com