

Validação da Extração de Dados de Gráficos de Artigos Científicos Utilizando Inteligência Artificial

Suzan Roberta Tombini Venturella^a, Fernanda Batista de Souza^{a,b}, Natan Padoin^a, Cíntia Soares^{a*}

^aUniversidade Federal de Santa Catarina (UFSC), Centro Tecnológico (CTC), Departamento de Engenharia Química e Engenharia de Alimentos (EQA), Programa de Pós-Graduação em Engenharia Química (PósENQ), Florianópolis-SC, Brasil

^bUniversidade Tecnológica Federal do Paraná (UTFPR), Departamento Acadêmico de Engenharias (DAENG), Francisco Beltrão-PR, Brasil

*cintia.soares@ufsc.br

RESUMO

A extração de dados a partir de gráficos científicos é prática recorrente, especialmente quando tabelas numéricas não são disponibilizadas nos artigos originais. Neste contexto, este trabalho avaliou a confiabilidade da extração automatizada baseada em inteligência artificial (IA) em comparação com a extração manual, em que os eixos e pontos são delimitados por um operador humano. Dados pareados de coordenadas obtidos de gráficos selecionados segundo critérios de resolução e distribuição dos pontos foram analisados. A comparação incluiu teste de Wilcoxon, Coeficiente de Correlação Intraclass (ICC), métricas de erro e análise de Bland-Altman. Embora o teste inferencial tenha indicado diferença estatisticamente significativa, as discrepâncias apresentaram baixa magnitude, com erros absolutos reduzidos e concordância excelente ($ICC \geq 0,999$). Os resultados indicam que a abordagem automatizada apresenta elevada consistência e precisão, constituindo alternativa viável para digitalização e reutilização de dados experimentais.

Palavras-chave: Análise estatística; automação computacional; ferramentas computacionais; tratamento de dados.

1. Introdução

A disponibilidade de dados numéricos confiáveis é fundamental para o desenvolvimento de modelos matemáticos, validação experimental e análises de sensibilidade. No entanto, muitos resultados experimentais são publicados apenas em formato gráfico, sem a disponibilização das tabelas correspondentes, o que compromete a reprodutibilidade e dificulta a reutilização dos dados para modelagem e validação cruzada.

Isotermas de adsorção, por exemplo, são amplamente utilizados na caracterização de materiais e na análise de equilíbrio em sistemas de separação. A ausência de dados tabulados nesses casos limita a reprodutibilidade e a reutilização das informações para modelagem e validação de processos.

A extração de dados pode ser realizada por ferramentas de digitalização manual, como o *WebPlotDigitizer*, que permitem a calibração dos eixos e a marcação dos pontos por um operador humano. Apesar de eficaz, esse processo é demorado e pode estar sujeito à variabilidade do operador, sobretudo em regiões com sobreposição de dados ou baixa resolução.

Avanços em visão computacional e inteligência artificial têm viabilizado métodos automatizados de extração gráfica, baseados em detecção visual, segmentação e reconhecimento de padrões. Estudos recentes indicam desempenho consistente dessas abordagens na interpretação de gráficos científicos (Decatur e Krishnan, 2021; Polak e Morgan, 2025). Contudo, sua adoção exige validação sistemática quanto à concordância e ao viés em relação a métodos consolidados, empregando métricas, como análise de Bland-Altman (Bland e Altman, 1986) e Coeficiente de Correlação Intraclass (Shrout e Fleiss, 1979).

Nesse contexto, este trabalho tem como objetivo validar a extração automatizada baseada em IA, comparando-a à digitalização manual quanto à concordância estatística, quantidade de dados extraídos e eficiência operacional.

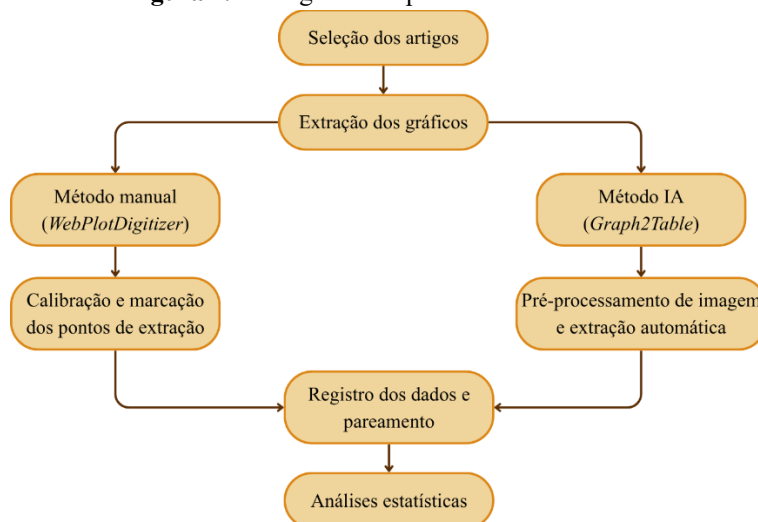
2. Metodologia

O procedimento metodológico adotado neste estudo foi estruturado de forma sequencial, contemplando desde a seleção do gráfico até a etapa de validação estatística dos dados extraídos. Inicialmente, realizou-se a obtenção da imagem do gráfico a partir de artigos científicos que não disponibilizavam tabelas numéricas

correspondentes. Em seguida, a extração de dados foi conduzida por dois métodos distintos: manual e automatizado, permitindo posterior comparação entre as abordagens.

A Figura 1 apresenta o fluxograma geral do procedimento adotado, destacando as etapas de extração, totalização, pareamento e análise estatística. As subseções seguintes descrevem detalhadamente cada fase do processo.

Figura 1. Fluxograma do procedimento adotado.



Fonte: elaborado pela autora (2026)

2.1. Seleção dos artigos e gráficos

Cinco artigos científicos contendo dados experimentais apresentados exclusivamente em formato gráfico foram selecionados. Em um dos artigos, dois gráficos distintos foram analisados, totalizando seis gráficos avaliados. Os gráficos utilizados são do tipo dispersão (*scatter plot*), com eixos X e Y em escala linear. Os dados extraídos correspondem a isotermas de adsorção, expressando a relação entre pressão (eixo X) e capacidade de adsorção (eixo Y). Esses dados são utilizados na caracterização de materiais adsorventes e na análise de equilíbrio em sistemas de separação.

A seleção foi baseada em dois critérios: o primeiro considerou a distribuição dos dados, buscando incluir gráficos com regiões de pontos esparsos e regiões com alta concentração, especialmente em baixas pressões, a fim de avaliar o desempenho da ferramenta de Inteligência Artificial (IA) em diferentes níveis de complexidade visual. O segundo critério envolveu a qualidade e a resolução das imagens, uma vez que nitidez, contraste e definição influenciam diretamente a precisão da extração, impactando tanto o processo manual quanto o automatizado.

2.2. Processo de extração de dados

A extração foi conduzida por dois métodos independentes. A extração manual foi realizada utilizando o software *WebPlotDigitizer*, na qual os eixos são previamente calibrados e os pontos são marcados manualmente pelo pesquisador. A extração automatizada, por sua vez, foi conduzida utilizando o software *Graph2Table*, ferramenta que realiza a identificação automática de curvas e conversão de elementos gráficos em dados numéricos.

Para a validação comparativa, realizou-se o pareamento dos pontos correspondentes entre os dois métodos. Pontos cujas coordenadas extraídas manualmente não apresentaram correspondência direta com aquelas obtidas pela IA foram desconsiderados na análise de concordância.

2.3. Pré-processamento das imagens

Na etapa inicial da extração automatizada, observou-se limitação na identificação de pontos, especialmente em regiões com alta densidade de dados. Para mitigar esse efeito, a extração com o *Graph2Table* foi realizada em três etapas sucessivas, associadas ao aprimoramento progressivo da qualidade das imagens.

As figuras foram convertidas para o formato .PNG (*Portable Network Graphics*) compressão sem perda), com resolução de 300 dpi e dimensão de 1829 pixels × 1524 pixels. A adoção do .PNG buscou preservar bordas e

evitar artefatos típicos de compressão com perda (como em arquivos .JPEG - *Joint Photographic Experts Group*), que podem comprometer a definição de linhas e pontos. O aumento da resolução teve como objetivo reduzir erros de quantização, proporcionando representação geométrica mais fiel e maior precisão na detecção automática dos dados.

2.4. *Análise estatística*

Após a extração, os dados foram organizados em planilhas e estruturados em colunas X_manual, X_IA, Y_manual e Y_IA, permitindo o pareamento direto das observações.

A análise foi conduzida em ambiente Python para avaliação da concordância entre as abordagens. Aplicou-se o teste de Wilcoxon para amostras pareadas, diante da ausência de garantia de normalidade, seguido do cálculo do Coeficiente de Correlação Intraclassa (ICC) para quantificação da concordância. A análise de Bland-Altman foi utilizada para avaliação do viés médio e dos limites de concordância ao longo da faixa de valores.

2.5. *Avaliação de desempenho temporal*

Além da análise estatística de concordância, uma avaliação comparativa do desempenho temporal dos métodos de extração foi realizada. O tempo foi mensurado desde o início do processo de calibração dos eixos do gráfico até a obtenção da planilha final contendo os dados coletados.

Para o método manual, o tempo incluiu as etapas de definição de escala e marcação individual dos pontos. Para o método baseado em IA, o tempo considerado incluiu a execução do algoritmo de detecção automática e a exportação dos dados obtidos.

A medição foi realizada para cada gráfico individualmente, de modo a avaliar a eficiência operacional das abordagens analisadas.

3. **Resultados e discussão**

3.1. *Pré-processamento da imagem na extração automatizada*

A extração automatizada utilizando o *Graph2Table* foi realizada em três condições distintas: (i) imagem original sem processamento; (ii) imagem convertida para o formato .PNG; e, (iii) imagem em formato .PNG com aprimoramento de resolução (300 dpi, 1829 pixels $\times$ 1524 pixels).

A extração manual realizada com o *WebPlotDigitizer* resultou na obtenção de 324 dados (162 pares) de pressão e capacidade de adsorção. Por sua vez, a extração automatizada conduzida por meio do *Graph2Table* totalizou 304 dados (152 pares).

Considerando ambos os métodos, 628 dados no total foram coletados. Para viabilizar a validação comparativa, procedeu-se ao pareamento dos pontos correspondentes entre as duas abordagens, resultando em 564 dados efetivamente comparáveis. Esse valor representa aproximadamente 89,8 % do total de dados coletados, indicando elevado grau de correspondência entre os métodos.

A diferença observada entre o número total de pontos extraídos manualmente e por IA pode estar associada à maior capacidade interpretativa do operador humano em regiões de sobreposição, baixa definição gráfica ou densidade elevada de pontos. Por outro lado, o alto percentual de dados pareados evidencia que a ferramenta automatizada foi capaz de reproduzir a maior parte das informações obtidas manualmente, demonstrando consistência na identificação dos dados experimentais representados nos gráficos.

Os pontos não pareados não foram considerados nas análises estatísticas subsequentes, uma vez que a ausência de correspondência direta inviabiliza a avaliação de concordância entre os métodos.

3.2. *Análise de eficiência operacional*

Observou-se redução significativa no tempo de extração ao empregar a abordagem automatizada. Para um dos artigos, o tempo de coleta manual foi de 5 min e 52 s e 7 min e 23 s para diferentes temperaturas, enquanto a extração por IA foi realizada em 35 s e 52 s, respectivamente. Nos demais casos, os tempos de extração manual variaram entre 1 min e 55 s e 5 min e 3 s, enquanto a abordagem automatizada apresentou tempos entre 10 s e 57 s. De forma geral, a utilização da IA resultou em reduções expressivas no tempo de coleta, variando aproximadamente entre 80 % e 95 % em relação ao método manual, evidenciando ganho substancial de eficiência operacional.

3.3. *Análise de concordância entre os métodos*

A comparação estatística foi realizada com base em pares de dados correspondentes entre os métodos manual e automatizado. Na condição inicial, utilizando a imagem sem processamento, 104 pares de dados foram

identificados. Após a conversão para o formato .PNG e posterior aprimoramento da resolução, o número de pontos extraídos aumentou progressivamente até atingir 152 pares de dados na condição final.

Esse crescimento corresponde a um incremento aproximado de 46 % no número de pontos identificados pela ferramenta automatizada ao longo das etapas de aprimoramento da imagem. O resultado evidencia que o desempenho do algoritmo de detecção é significativamente influenciado pela qualidade visual da figura analisada.

É importante destacar que esse aumento se refere exclusivamente à evolução do desempenho da ferramenta de IA entre as diferentes condições de processamento da imagem, não implicando superioridade em relação ao método manual.

3.3.1 Teste de Wilcoxon

O teste foi aplicado com o objetivo de verificar a existência de diferenças sistemáticas entre os valores obtidos pelos métodos manual e automatizado.

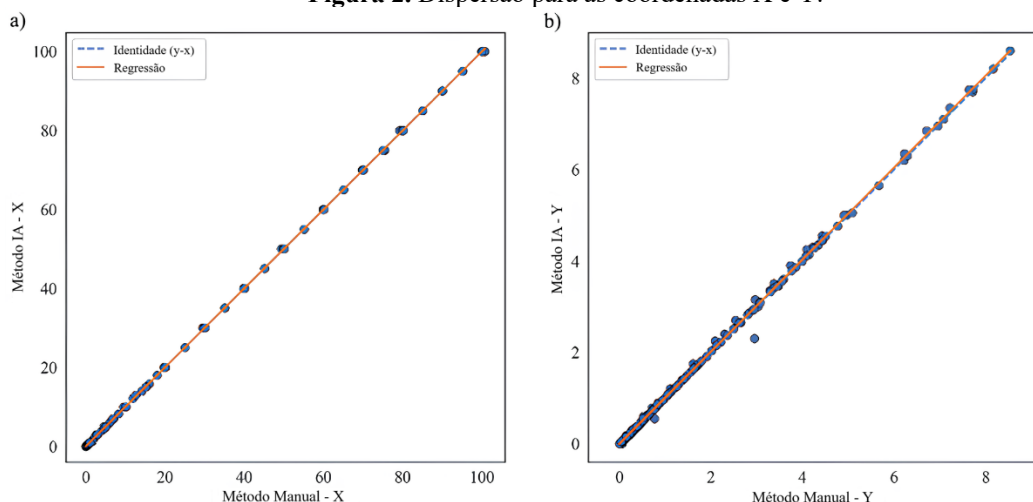
Para a coordenada X, o teste resultou em $p = 0,0068$, enquanto para a coordenada Y obteve-se $p = 0,0003$, indicando diferença estatisticamente significativa entre os métodos ao nível de 5 %. Entretanto, a interpretação desses resultados deve considerar a magnitude das diferenças observadas e o tamanho amostral ($n = 141$). Amostras de maior dimensão aumentam a potência estatística do teste, tornando-o capaz de detectar diferenças muito pequenas. De fato, as diferenças médias foram de 0,002 para X e $-0,013$ para Y, valores substancialmente reduzidos quando comparados à amplitude total das respectivas variáveis.

Embora o teste indique diferença estatisticamente significativa, a magnitude das discrepâncias observadas é reduzida. Conforme discutido por Sullivan e Feinn (2012) e Wasserstein e Lazar (2016), significância estatística não implica necessariamente relevância prática, especialmente em amostras de maior dimensão, nas quais pequenas diferenças podem tornar-se detectáveis pelo aumento da potência do teste.

3.3.2 Coeficiente de Correlação Intraclass (ICC)

A análise de concordância por meio do Coeficiente de Correlação Intraclass (ICC) revelou valores extremamente elevados: para a coordenada X = 1,000 e para coordenada Y = 0,999.

Figura 2. Dispersão para as coordenadas X e Y.



Fonte: elaborado pela autora (2026)

De acordo com a classificação proposta por Koo e Li (2016), valores de ICC superiores a 0,90 indicam concordância excelente. Assim, os resultados obtidos evidenciam consistência praticamente perfeita entre os métodos avaliados. A elevada concordância observada também pode ser visualizada por meio dos gráficos de dispersão apresentados na Figura 2. Em ambas as coordenadas, verifica-se forte alinhamento dos pontos em torno da linha de identidade, indicando correspondência praticamente perfeita entre os métodos manual e automatizado. No caso da coordenada X (Figura 2a), o coeficiente de determinação foi $R^2 = 1,000$, com equação de regressão próxima da identidade, evidenciando ausência de viés sistemático relevante. Para a coordenada Y (Figura 2b), observou-se $R^2 = 0,999$, igualmente indicando elevada linearidade e mínima divergência entre os métodos.

3.3.3 Métricas de erro

As métricas de erro confirmam a precisão da extração automatizada e são apresentadas na Tabela 1.

Tabela 1. Métricas de erro.

Coordenadas	MAE	RMSE	Erro Máximo
X	0,091	0,171	0,800
Y	0,037	0,077	0,650

Fonte: elaborado pela autora (2026)

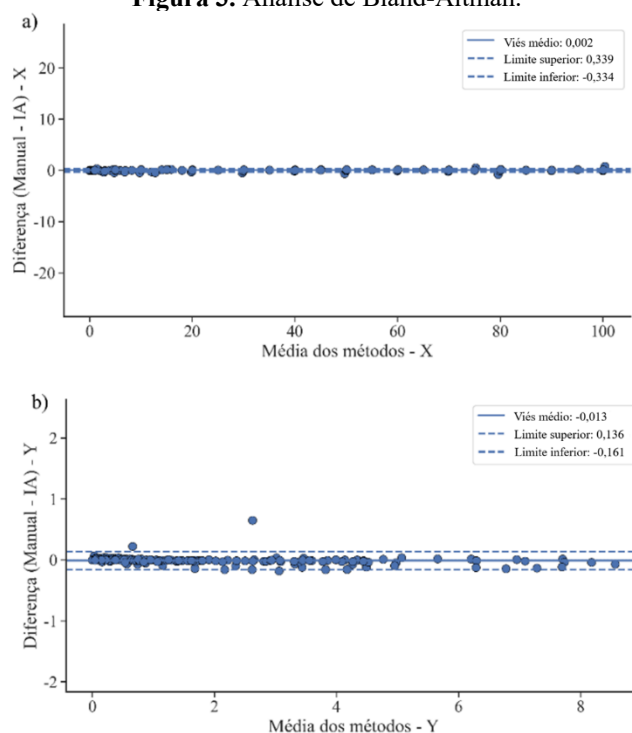
Considerando a faixa de variação das variáveis analisadas, esses valores representam erros absolutos reduzidos. O MAE indica que, em média, as diferenças entre os métodos permanecem inferiores a uma unidade na escala original, enquanto o RMSE evidencia baixa dispersão das discrepâncias.

De forma integrada, os resultados indicam que o método baseado em inteligência artificial apresenta elevada precisão e estabilidade, reproduzindo os valores extraídos manualmente com variações de pequena magnitude e sem evidência de erro sistemático expressivo.

3.3.4 Análise de Bland-Altman

A análise de Bland-Altman permitiu avaliar o viés médio e os limites de concordância entre os métodos, conforme ilustrado na Figura 3.

Figura 3. Análise de Bland-Altman.



Fonte: elaborado pela autora (2026)

Para a coordenada X (Figura 3a), o viés médio foi de 0,002, com limites de concordância de $-0,334$ a $0,339$. Observa-se que a maior parte dos pontos se distribui dentro desses limites, sem evidência clara de tendência proporcional ao aumento da magnitude da variável. Para a coordenada Y (Figura 3b), o viés médio foi de $-0,013$, com limites de concordância de $-0,161$ a $0,136$. De maneira semelhante, as diferenças permanecem majoritariamente concentradas em torno da média, indicando ausência de viés sistemático relevante.

Em ambas as coordenadas, a dispersão dos pontos demonstra estabilidade da diferença ao longo da faixa de valores analisada, reforçando a elevada concordância entre os métodos.

4. Conclusão

Este estudo avaliou a confiabilidade da extração automatizada de dados gráficos baseada em inteligência artificial em comparação ao método manual. Embora o teste de Wilcoxon tenha indicado diferença estatisticamente significativa entre os métodos, as diferenças médias observadas foram de baixa magnitude, com métricas de erro reduzidas e concordância excelente segundo o Coeficiente de Correlação Intraclass ($ICC \geq 0,999$).

A análise integrada demonstra que o método automatizado apresenta elevada consistência e precisão em relação à extração manual, configurando alternativa tecnicamente viável para digitalização de dados gráficos.

Como limitação, destaca-se a dependência da qualidade da imagem para otimização do desempenho da ferramenta. Recomenda-se, como trabalhos futuros, a ampliação da amostra de gráficos e a avaliação do impacto de diferentes padrões de resolução e complexidade visual sobre a acurácia da extração.

Agradecimentos: Os autores agradecem à CAPES e ao CNPq (Processos 305472/2025-9 – C.S. e 312247/2022-2 – N.P.) pelo apoio financeiro.

Referências

- J. M. Bland e D. Altman: Statistical Methods for Assessing Agreement Between Two Methods of Clinical Measurement, *The Lancet* (327), 307–310, 1986.
- D. Decatur e S. Krishnan: VizExtract: Automatic Relation Extraction from Data Visualizations, arXiv: 2021.
- T. K. Koo e M. Y. Li: A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research, *Journal of Chiropractic Medicine* (15), 155–163, 2016.
- M. P. Polak e D. Morgan: Leveraging Vision Capabilities of Multimodal LLMs for Automated Data Extraction from Plots, arXiv: 2025.
- P. E. Shrout e J. L. Fleiss: Intraclass Correlations: Uses in Assessing Rater Reliability, *Psychological Bulletin* (86), 420–428, 1979.
- G. M. Sullivan e R. Feinn: Using Effect Size - or Why the P Value Is Not Enough, *Journal of Graduate Medical Education* (4), 279–282, 2012.
- R. L. Wasserstein e N. A. Lazar: The ASA Statement on p-Values: Context, Process, and Purpose, *The American Statistician* (70), 129–133, 2016.