

Universidade Federal de Uberlândia - UFU

Bacharelado em Ciência da Computação

GBC213 - Multimídia - 2025

Douglas Ferreira Martins - 12311BCC058

Gustavo Cruz Biazi - 12111BCC025

Lázaro Joabe - 12211BCC022

Nathan Augusto Coelho dos Santos - 11811BCC046

Relatório

Analizador de imagens para quantificação bioquímica

Uberlândia, MG
Setembro de 2025

1 Introdução

Este projeto implementa um sistema para análise quantitativa de ensaios de Bradford, permitindo a extração de dados quantitativos de imagens digitalizadas de reações colorimétricas entre proteínas e reagente de Bradford. A aplicação possibilita ao usuário carregar imagens, definir regiões de interesse (ROI), calibrar um modelo com base em imagens padrão e analisar amostras desconhecidas para prever concentrações de proteína. O sistema foi desenvolvido em Python, utilizando uma interface gráfica interativa para facilitar o uso por especialistas da área bioquímica, atendendo aos requisitos da disciplina GBC213 - Multimídia.

2 Fluxo de trabalho

A aplicação se divide em duas fases de execução, uma fase de calibração e uma fase de análise, garantindo um fluxo de trabalho completo desde a configuração do modelo até a predição em novas amostras.

2.1 Fase de Calibração

A fase de calibração é o coração do sistema, onde o modelo de predição é treinado com base em amostras de concentração conhecida. As etapas são:

- Carregamento de imagens de calibração.
- Detecção automática de círculos (poços) usando a Transformada de Hough.
- Definição de múltiplas ROIs de fundo para correção de iluminação não uniforme.
- Aplicação de um sistema avançado de filtragem de poços (fotométrico e cromático) para aumentar a precisão da detecção.
- Redetecção guiada por ROIs de fundo para refinar a localização dos poços.
- Associação de uma concentração conhecida a um poço selecionado e armazenamento do par (intensidade corrigida, concentração).
- Calibração de um modelo de regressão linear a partir dos pontos coletados.

2.2 Fase de Análise

Uma vez que o modelo está calibrado, a fase de análise permite a quantificação de amostras desconhecidas:

- Análise individual de imagens desconhecidas.
- Aplicação automática dos filtros e das configurações de ROI de fundo definidas na calibração, garantindo consistência.
- Processamento em lote de pastas inteiras de imagens.
- Predição de concentrações para cada poço detectado utilizando o modelo calibrado.
- Geração de um arquivo CSV com os resultados consolidados do processamento em lote.

3 Processamento e Modelo de cor

O processamento de imagem foi projetado para detectar e quantificar com precisão os poços de reação. A imagem, carregada no formato BGR, é primeiramente convertida para escala de cinza para a maioria das operações de detecção e cálculo de intensidade.

A detecção inicial dos poços é realizada pela função ‘cv2.HoughCircles’, que identifica formas circulares na imagem. Para refinar essa detecção, o sistema implementa uma série de filtros avançados:

1. **Correção de Fundo:** O usuário pode selecionar múltiplas regiões de interesse (ROIs) que representem o fundo da imagem. O sistema calcula a intensidade média de pixels nessas áreas (I_{fundo}) e a utiliza para corrigir a intensidade média de cada poço detectado (I_{poco}). A intensidade corrigida, que serve como variável para o modelo de regressão, é calculada como:

$$I_{corrigida} = I_{fundo} - I_{poco}$$

2. **Redetecção Guiada pelo Fundo:** Após a seleção de ROIs de fundo, o sistema pode realizar uma nova detecção. Ele cria uma máscara binária, suprimindo pixels com intensidade similar à média do fundo, e aplica a Transformada de Hough novamente sobre a imagem mascarada. Isso melhora significativamente a detecção em imagens com iluminação irregular.
3. **Filtro Fotométrico:** Este filtro descarta poços detectados que são fotometricamente semelhantes ao fundo. Um poço é removido se a diferença absoluta entre sua intensidade média e a do fundo for menor que um limiar (`abs_diff_thresh`), e, simultaneamente, o contraste entre o disco interno do poço e seu anel externo for baixo.
4. **Filtro Cromático (Modelo de Cor):** Para distinguir poços contendo reagente de poços vazios ou com artefatos, o sistema emprega um filtro híbrido que opera nos espaços de cor CIE Lab e HSV. Um poço é considerado válido se atender a pelo menos uma das seguintes condições, que indicam uma assinatura de cor distinta do fundo:

- A diferença de cor perceptual (Delta E CIE 1976) entre o poço e o fundo no espaço Lab é superior a um limiar ($\Delta E_{ab}^* \geq \text{deltaE_min}$). A fórmula é:

$$\Delta E_{ab}^* = \sqrt{(L_2^* - L_1^*)^2 + (a_2^* - a_1^*)^2 + (b_2^* - b_1^*)^2}$$

- **OU** a saturação do poço no espaço HSV é maior que um limiar ($S_{poco} \geq \text{s_min}$) **E** a diferença absoluta de brilho entre o poço e o fundo também é superior a um limiar ($|V_{poco} - V_{fundo}| \geq \text{v_diff_min}$).

Este pipeline multifacetado garante que apenas os poços de reação mais relevantes sejam selecionados para a calibração e análise, aumentando a confiabilidade dos resultados.

4 Modelo matemático e resultados da predição

Para quantificar a concentração de proteína, o projeto utiliza um modelo de regressão linear simples, que estabelece uma relação linear entre a intensidade corrigida da cor do poço e a concentração da amostra.

O modelo matemático é definido pela equação de uma reta:

$$C = a \cdot I + b$$

Onde:

- C é a concentração de proteína prevista.
- I é a intensidade corrigida ($I_{fundo} - I_{poco}$) calculada na etapa de processamento.
- a (alfa) é o coeficiente angular da reta, representando a sensibilidade da cor à concentração.
- b (beta) é o intercepto linear, representando a concentração teórica para uma intensidade nula.

Durante a fase de calibração, o sistema armazena múltiplos pares de $(I_{corrigida}, C_{conhecida})$. Ao final, os coeficientes a e b são calculados utilizando o método dos mínimos quadrados, implementado pela função ‘numpy.polyfit’ com grau 1.

A qualidade do ajuste do modelo aos dados de calibração é avaliada pelo Coeficiente de Determinação, R^2 , que é calculado pela fórmula:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} = 1 - \frac{\sum_i (C_i - \hat{C}_i)^2}{\sum_i (C_i - \bar{C})^2}$$

Onde C_i são as concentrações reais, \hat{C}_i são as concentrações previstas pelo modelo e \bar{C} é a média das concentrações reais. Um valor de R^2 próximo de 1 indica que o modelo se ajusta bem aos dados de calibração. Após a calibração, o sistema exibe a equação final e o valor de R^2 para o usuário.

5 Modelo de Cores e Compressão

Para aumentar a flexibilidade da análise, o sistema possui a capacidade de o usuário selecionar o modelo de cor e os componentes específicos a serem utilizados como características (features) para o treinamento do modelo de regressão. Através de uma janela de configurações inicial, é possível escolher entre os espaços de cor RGB, HSV e CIE L*a*b*, bem como selecionar se a análise será baseada em todos os componentes do modelo (ex: H, S e V) ou em um único componente de interesse, como o Matiz ('H') ou o eixo cromático 'b*'. A extração desses dados é realizada pela função `extract.features`, que calcula o valor médio do(s) componente(s) selecionado(s) dentro da região de interesse circular, convertendo a cor média da ROI para o espaço de cor escolhido antes de retornar os valores numéricos que alimentarão o modelo. Essa funcionalidade permite ao usuário investigar qual dimensão da cor possui a correlação mais linear e robusta com a concentração de proteína, otimizando a precisão das predições para diferentes condições de iluminação e reagentes.

Para atender ao requisito de análise do impacto da compressão de imagem, o sistema implementa uma funcionalidade dedicada à geração de um conjunto de dados para este experimento, conforme as diretrizes do projeto. Através da janela de configurações inicial, o usuário pode especificar uma pasta contendo as imagens originais e um diretório de destino. A aplicação então executa a função `save_compressed_versions`, que cria automaticamente subpastas nomeadas de acordo com os níveis de qualidade JPEG definidos (por padrão, 95, 75, 50 e 25). Cada imagem original é lida e salva como um novo arquivo .jpg em cada uma dessas subpastas, utilizando a biblioteca OpenCV. O nível de compressão é controlado precisamente pelo parâmetro `cv2.IMWRITE_JPEG_QUALITY` da função `cv2.imwrite`, garantindo que cada conjunto de imagens corresponda ao nível de qualidade desejado. Este processo gera a base de dados necessária para a condução do protocolo experimental, permitindo a posterior análise da relação entre o tamanho do arquivo e o erro de medição da concentração.

6 Resultados

A imagem utilizada nos testes foi a imagem com nome "165556.jpg", pois continha uma grande variedade de cores, implicando em diversas concentrações com diferenças de iluminação, o que permitirá uma calibração melhor do modelo do que outras imagens que possuem pouca variedade. Para o cálculo do erro com referência à fórmula original, foi utilizada a fórmula do Erro Percentual Absoluto. No caso, o valor real foi definido como sendo as concentrações previstas no dataset original, e o valor previsto o valor obtido nas versões comprimidas.

$$\text{EPA} = \left| \frac{\text{Valor Previsto} - \text{Valor Real}}{\text{Valor Real}} \right| \times 100\%$$

Os resultados para cada quantidade de compressão podem ser vistos na Tabela 1 com a média entre os 3 modelos. A diferença entre cada modelo era menor que 10^{-3} , para todas as linhas da coluna. Essa tabela apresenta os erros para cada imagem para cada porcentagem de compressão realizada nela.

Tabela 1: Tabela de Erro Percentual por Nível de Compressão - Média 3 Modelos

Imagem	Erro @ 95%	Erro @ 75%	Erro @ 50%	Erro @ 25%
0bf3bcff.jpg	12.61%	10.59%	4.98%	5.48%
5e11a902.jpg	8.54%	0.57%	1.05%	4.57%
07fdf52a.jpg	83.49%	86.27%	89.65%	87.34%
8ff57f96.jpg	20.41%	5.77%	21.09%	10.30%
023fa417.jpg	-0.25%	-22.25%	0.00%	-8.45%
78a7e629.jpg	3.18%	5.80%	6.59%	4.02%
518de58b.jpg	0.32%	0.16%	0.40%	4.77%
587f3bf7.jpg	5.14%	12.61%	6.41%	73.31%
165528.jpg	225.92%	46.33%	211.57%	102.62%
165535.jpg	0.00%	9.97%	6.66%	22.12%
165550.jpg	0.01%	12.25%	3.55%	0.34%
165554.jpg	0.01%	9.64%	0.74%	0.60%
165556.jpg	108.82%	55.17%	86.99%	107.46%
37383707.jpg	14.46%	0.33%	7.57%	31.49%
b49ac8c4.jpg	0.01%	9.35%	12.86%	12.21%
ba31611e.jpg	6.72%	8.10%	21.38%	4.34%
bb21875a.jpg	28.94%	7.63%	33.23%	0.17%
c5eeceb9.jpg	2.99%	2.63%	2.29%	2.87%
c3551634.jpg	63.51%	66.51%	64.71%	0.71%
ef351254.jpg	0.03%	16.46%	0.05%	3.82%
f1027d6d.jpg	0.00%	0.03%	0.28%	3.21%

Em seguida, a Tabela 2 mostra o tamanho de cada imagem após as diversas etapas de compressão, demonstrando a eficiência das diversas etapas de compressão.

Tabela 2: Tabela de Tamanho de Arquivo (KB) por Nível de Compressão

Imagem	Original (KB)	95% (KB)	75% (KB)	50% (KB)	25% (KB)
0bf3bcff.jpg	149	277	148	116	75
5e11a902.jpg	138	259	139	110	71
07fdf52a.jpg	138	257	137	107	68
8ff57f96.jpg	154	287	153	120	78
023fa417.jpg	112	215	115	92	59
78a7e629.jpg	138	257	137	109	71
518de58b.jpg	141	264	141	112	72
587f3bf7.jpg	161	299	159	125	81
165528.jpg	269	283	109	68	47
165535.jpg	160	168	67	43	29
165550.jpg	356	373	143	88	59
165554.jpg	2436	2438	704	435	320
165556.jpg	2343	2336	644	394	297
37383707.jpg	152	290	154	119	75
b49ac8c4.jpg	139	261	139	111	72
ba31611e.jpg	133	247	133	106	69
bb21875a.jpg	145	270	145	115	74
c5eeceb9.jpg	113	215	115	93	60
c3551634.jpg	123	232	124	99	64
ef351254.jpg	139	261	139	110	71
f1027d6d.jpg	142	266	143	113	73

Com os dados obtidos nas 3 tabelas de erro, podemos efetuar o cálculo da média dos erros para cada compressão, e com ela, gerar o gráfico necessário. Os cálculos finais podem ser vistos na Tabela 3, e o gráfico na Imagem 1.

Tabela 3: Erro x Compressão

Nível de Qualidade JPEG (%)	Erro Médio Absoluto (%)
95	27.87
75	18.50
50	27.72
25	23.34

O gráfico representando essa tabela pode ser visto na figura abaixo.

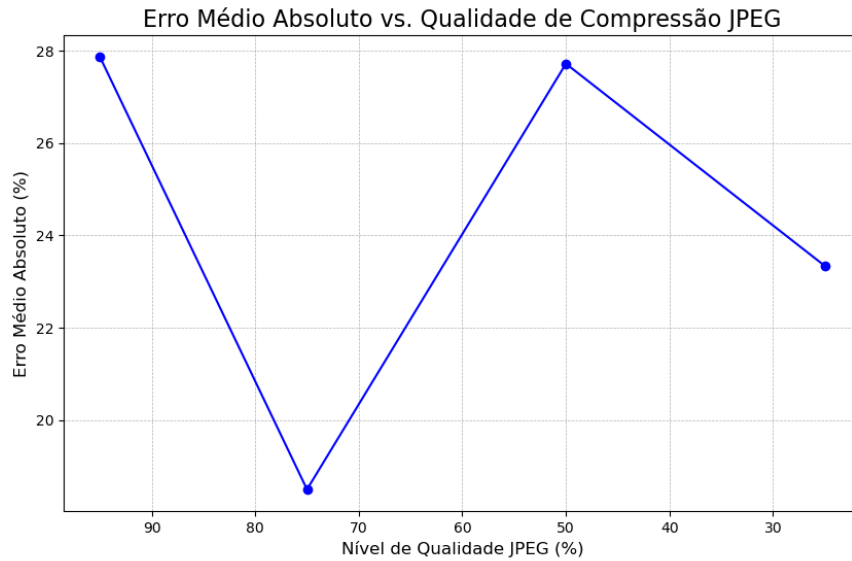


Figura 1: Gráfico da relação entre o erro médio absoluto e a qualidade da compressão JPEG.

7 Uso de ferramentas de IA Generativa

Para o código, as LLMs ChatGPT e Gemini foram utilizadas na prototipagem, criação do arcabouço geral do projeto, que foi então preenchido com o código próprio. No desenvolvimento do relatório, o Gemini foi utilizado para a correção tanto de erros gramaticais ou ortográficos quanto do código \LaTeX e também para melhorar a fluidez do texto (separação em etapas no segmento referente ao processamento de imagem, por exemplo).

Prompts utilizados (resumo)

1. Arquitetura

“Proponha a estrutura de pastas e módulos para um app de análise de imagens com PySimpleGUI e OpenCV, separando responsabilidades (GUI, detecção de poços, seleção de ROI, regressão linear, utilidades). Inclua comentários e docstrings.”

2. GUI básica (calibração e análise)

“Crie uma interface PySimpleGUI com duas janelas: Calibração e Análise. Permita trocar entre elas e reaproveitar um modelo de regressão calibrado.”

3. Detecção por HoughCircles

“Implemente detecção de poços em placas via `cv2.HoughCircles`, com `dp`, `minDist`, `param1/param2` e faixa de raios ajustável. Retorne círculos como (x, y, r) .”

4. Pontuação de poços (score)

“Dado um círculo, calcule o score como $\text{média}(\text{anel}) - \text{média}(\text{disco})$ em tons de cinza, para ranquear poços e destacar o melhor.”

5. ROI de fundo + correção de iluminação

“Permita selecionar uma ROI de fundo (“branco”) e calcule intensidade corrigida: $\text{fundo} - \text{poço}$. Preveja concentração a partir dessa intensidade.”

6. Múltiplas ROIs de fundo

“Suporte múltiplas ROIs de fundo e use a média das ROIs como referência. Reaplique a detecção e filtre círculos parecidos com o fundo.”

7. Redetecção guiada pelo fundo

“A partir das ROIs de fundo, crie uma máscara anti-fundo com limiar adaptativo e rode Hough novamente numa faixa de raio plausível.”

8. Filtro fotométrico (gray/contraste)

“Após a detecção, remova poços cujo disco seja parecido com o fundo ($|disc - bg| < T$) ou cujo contraste anel–disco seja baixo.”

9. Filtro de cor (Lab/HSV)

“Implemente um filtro que mantenha poços com $\Delta E(\text{Lab}) \geq \tau$ ou saturação/valor (HSV) suficientemente diferentes do fundo.”

10. “Selecionar poço manual” + expansão por similares

*“Ao selecionar manualmente um poço (semente), procure outros poços similares (faixa de raio $\pm 20\%$, ΔE ou cinza próximo) e **adicione** ao conjunto já detectado.”*

11. Propagação calibração \rightarrow análise

“Carregue ROIs de fundo e referências de poços da calibração na janela de análise; reexecute filtros com sliders próprios da análise.”

12. Painel Avançado (sliders/checkboxes)

*“Adicione um painel com: **param2** (Hough), Δ_{abs} (disco–fundo), $\Delta_{contraste}$ (anel–disco), ΔE_{\min} , S_{\min} , $|V - V_{\text{fundo}}|$, e toggles para ‘Redetectar guiado’, ‘Filtro fotométrico’ e ‘Filtro de cor’.”*

13. Modelo de regressão com R^2

*“Implemente uma classe de calibração linear (mínimos quadrados), exponha **ajustar**, **prever**, **coeficientes** e R^2 , valide que há ≥ 2 pontos.”*

8 Conclusão

O requisito de implementar uma interface gráfica foi realizado com sucesso. Durante o uso, todas as informações relevantes são exibidas diretamente na janela da aplicação. Na fase de calibração, o usuário pode visualizar os poços detectados, as ROIs de fundo e os dados de cada ponto de calibração adicionado. Após calibrar, a equação do modelo linear ($C = a \cdot I + b$) e o coeficiente R^2 são exibidos em uma janela, exibindo a qualidade do modelo. Na fase de análise, a concentração prevista para o poço selecionado é exibida em tempo real.

A funcionalidade mais poderosa para análise de um grande volume de amostras é o processamento em lote. O usuário pode selecionar uma pasta contendo múltiplas imagens de ensaios. O sistema itera sobre cada imagem, aplica o processo de detecção e filtragem, e utiliza o modelo calibrado para prever a concentração. Ao final do processo, um arquivo chamado ‘resultados_analise.csv’ é gerado no mesmo diretório. Este arquivo contém duas colunas: ‘Imagem’ e ‘Concentracao_Prevista’, listando o nome de cada arquivo de imagem e o valor de concentração correspondente, ou “N/A” em caso de erro na análise. Isso permite uma análise fácil dos dados e a exportação dos valores para outras ferramentas de software.