

Coffee bean quality analysis using convolutional neural networks

Bernardo Silva Ribeiro Duarte¹, Alan Diego Aurelio Carneiro², Pedro Moises de Sousa³

¹Instituto de Ciências Exatas e Tecnológicas - Universidade Federal de Viçosa - Campus Rio Paranaíba (UFV-CRP)

Rodovia MG230, Km7, Caixa Postal 22 - 38.810-000 - Rio Paranaíba - MG - Brasil

bernardo.duarte@ufv.br, alan.carneiro@ufv.br, pedromois@ufv.br

Abstract—Coffee grading is the main procedure in its production. One of the processes to grade coffee is done manually, requiring a lot of training and experience from experts. The main objective of this work is to use technologies based on Artificial Intelligence, with Convolutional Neural Network models, and together with the application of image processing techniques, to improve the quality analysis of coffee beans. We performed a comparison of pre-trained models, namely AlexNet and DenseNet, using a coffee bean dataset. A total of 4272 coffee bean images from the USK-COFFEE database were used in this work. The model training, testing and validation processes were performed with an 80/10/10 division of the obtained images. Classification metrics such as Recall, F-1 Score, were used for the detailed analysis of the performance models. ROC curves were used to analyze their distinction.

Index Terms—Coffee, Neural Network, classification, analysis.

I. INTRODUÇÃO

O Brasil é o maior produtor e exportador, e o segundo maior consumidor de café no mundo, atrás somente dos Estados Unidos [1]. Com uma área destinada à cafeicultura, totalizando 2,26 milhões de hectares em 2023, a produção brasileira de café pode atingir 54,94 milhões de sacas do grão beneficiado, segundo a Conab [1].

A classificação do café é o principal procedimento em sua produção. Um dos processos para se classificar o café é feito manualmente, exigindo muito treinamento e experiência dos especialistas, porém mesmo com especializações, os classificadores necessitam de pausas, devido as exaustões físicas e mentais, e podem cometer erros na classificação devido a essas condições [2].

Com a evolução das tecnologias de inteligência artificial, foi possível se desenvolver diversas ferramentas que impulsionaram indústrias, como por exemplo a indústria FOSS, que é a fornecedora líder global de análise para indústrias alimentícias e agrícolas [4]. Como também é demonstrado no trabalho do Ping Wang [7], utilizando deep learning, onde se foi criada uma rede neural com um compactador, a fim de aprimorar a identificação de grãos de café, permitindo assim, ter um método com menos custo e mais eficiente.

Portanto, esse projeto tem como o objetivo utilizar de tecnologias baseadas em Inteligência artificial, com modelos de Redes Neurais Convolucionais, como Densenet e AlexNet [3], e juntamente com a aplicação de técnicas de processamento de imagens, como a filtragem CLAHE [5], para melhorar a

análise de qualidade de grãos de café. Esperando assim, que os resultados forneçam as ferramentas de apoio, que agilizem o processo de análise dos grãos.

II. REFERENCIAL TEÓRICO

A. Café

Os grãos de café podem ter várias características, como tamanhos, cores, dentre outras, podendo ser de diversas maneiras. Neste trabalho, dois tipos de avaliação serão utilizados para a classificação de cada grão, como as exemplificadas na fig.1 abaixo.

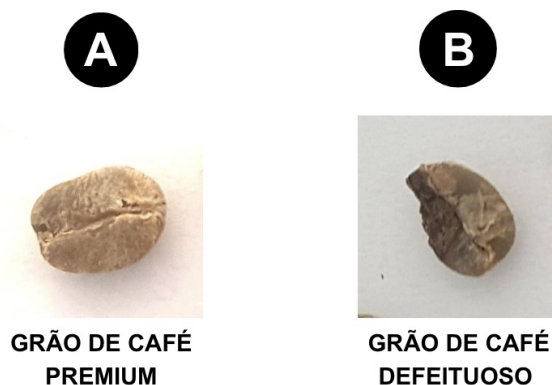


Fig. 1. Comparação de grãos do dataset.(imagem feita pelo autor)

B. Redes Neurais

Uma rede neural é um método de inteligência artificial que ensina computadores a processar dados de uma forma inspirada pelo cérebro humano. É um tipo de processo de aprendizado de máquina, chamado aprendizado profundo, que usa neurônios interconectados em uma estrutura em camadas, semelhante ao cérebro humano. A rede neural cria um sistema adaptativo que os computadores usam para aprender com os erros e se aprimorar continuamente. As redes neurais artificiais tentam solucionar problemas complicados, como resumir documentos ou reconhecer rostos com grande precisão [5].

C. Redes Neurais Convolucionais (CNN)

A operação convolucional é uma função matemática executada por uma técnica para extrair características dos dados de entrada, permitindo que o modelo aprenda representações complexas das informações por meio de uma função matemática. A ideia principal da CNN é obter os dados de entrada nas camadas superiores e combiná-los de forma complexa em camadas inferiores, passando por diversas camadas ocultas, onde o modelo aprende características complexas. Elas são usadas principalmente para identificação de objetos e classificação de imagens [5].

D. Arquitetura de uma CNN

A arquitetura da rede neural convolucional é feita de 3 principais recursos: campo receptivo local, onde cada neurônio das camadas entre a entrada e saída, é conectado em um campo receptivo, pois assim, os neurônios artificiais conseguem extrair elementos como cantos, arestas e pontos da imagem; Compartilhamento de pesos, uma técnica para reduzir a quantidade de parâmetros treináveis, com o objetivo de tornar os modelos mais eficientes, com o filtro convolucional usando os mesmos pesos para todas as suas aplicações na imagem, tornando o modelo mais leve e rápido, reduzindo o número de parâmetros usados; Subsampling (pooling), reduz o tamanho de entrada de dados, reduzindo assim os parâmetros do modelo. Uma das técnicas mais usadas é o pooling, que reduz a dimensionalidade dos mapas de características, onde é preservado aqueles que são considerados os mais importantes [5], como exemplificado o funcionamento na fig. 2.

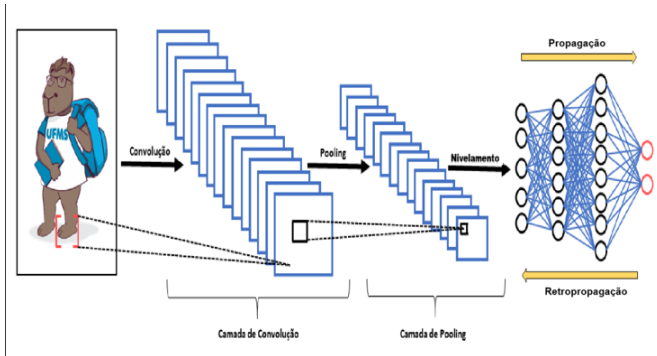


Fig. 2. Arquitetura de redes neurais convolucionais.(imagem feita pelo autor)

E. Camadas Convolucionais

Uma camada convolucional tem uma grande importância principalmente na etapa de extração de características na entrada de dados. A camada é formada por uma série de filtros, onde cada filtro é projetado para detectar padrões em regiões da imagem de entrada, extraindo informações significativas para solução efetiva de tarefas acerca de visão computacional [5].

F. Camadas de Pooling

Logo após uma camada convolucional, há uma camada de *pooling*, com o objetivo de diminuir a quantidade de parâmetros na rede. A camada de Pooling pega cada saída da camada convolucional e reduz a quantidade de amostras. Geralmente o uso da camada de *Pooling* são para situações em que a identificação de informação é mais importante do que sua classificação, tornando o processo rápido e preciso [5], como é exemplificado pela fig. 3.

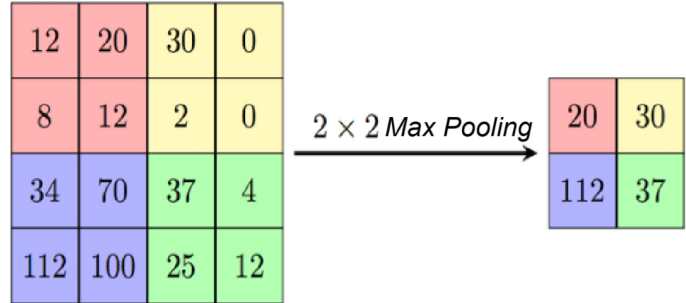


Fig. 3. Exemplo de *Max Pooling*.

G. Camadas Totalmente Conectadas

As camadas totalmente conectadas são um tipo de camada em redes neurais que cada neurônio da camada anterior é ligado a cada neurônio da camada totalmente conectada. Diferente das camadas de pooling e convolucionais, as camadas totalmente conectadas são usadas no processo de classificação de informação [5]. O objetivo é permitir que a rede neural aprenda representações mais complexas das entradas de dados.

H. Filtragem CLAHE

A *Contrast Limited Adaptive Histogram Equalization*, mais conhecida como CLAHE [6], é uma técnica de processamento de imagens que melhora o contraste da imagem, limitando a amplificação do contraste para evitar o excesso de ruído. Recebendo as imagens do dataset original transformadas em escala cinza, e assim, melhorando o contraste para que a análise dos grãos seja mais eficaz.

III. TRABALHOS RELACIONADOS

No trabalho [7] é proposto um modelo com um sistema inteligente de inspeção de qualidade de grãos de café que utiliza do método de KD, um método que faz a transferência de aprendizado de uma máquina mais complexa para uma mais simples. Foram usadas imagens de uma base privada, que continham um total de 4626 imagens, com 4000 para testes e 626 para validação. O modelo proposto atingiu 95% de acurácia enquanto a ResNet-18 chegou a 93%.

No trabalho [8], se tem como objetivo desenvolver um sistema automático de classificação para os produtores de grãos de café em Timor-Leste, que foi baseado na rede neural chamada de “Neocognitron”. No trabalho se utilizou imagens da base de dados ImageNet, onde se utilizou um total de 13.000 imagens de 6.500 grãos de café, com tamanhos de

256 x 256 pixels. Como resultado, se separou com sucesso os grãos, em algumas classes, com mais de 90% de acurácia, sendo maior em comparação a rede em que foi baseada.

O trabalho [9] tem como objetivo principal, classificar 3 grãos de café diferentes utilizando suas imagens, através do método de transfer learning, utilizando 4 modelos diferentes baseados em Redes Neurais Convolucionais, que são SqueezeNet, Inception V3, VGG16 e VGG19. O conjunto de dados utilizado no treinamento dos modelos foi criado especialmente para este estudo. Um total de 1.554 imagens de grãos de café dos tipos de café Espresso, Quênia e Starbucks Pike Place foram coletadas com o mecanismo criado, sendo elas 530 de café Espresso, 502 do tipo Quênia e 522 do tipo Starbucks Pike Place. Processos de treinamento e teste de modelo foram realizados com as imagens obtidas. Foram utilizadas métricas de sucesso de classificação, precisão, recall e pontuação F-1 para a análise detalhada dos modelos de performances. Curvas ROC foram utilizadas para analisar sua distinção. Como resultado dos testes, o sucesso médio de classificação dos modelos foi determinado como 87,3% para SqueezeNet, 81,4% para Inception V3, 78,2% para VGG16 e 72,5% para VGG19. Estes resultados demonstram que o SqueezeNet é o modelo de maior sucesso. Acredita-se que este estudo possa contribuir com o tema da separação dos grãos de café na indústria.

O trabalho [12] tem como objetivo principal é avaliar quão efetivamente vários modelos pré-treinados podem prever tipos de café usando técnicas avançadas de aprendizado profundo, com uma comparação abrangente de vários modelos pré-treinados, incluindo AlexNet, LeNet, HRNet, Google Net, Mobile V2 Net, ResNet (50), VGG, Efficient, Darknet e DenseNet, utilizando um conjunto de dados do tipo café. As imagens foram coletadas e salvas automaticamente em formato PNG, com 4800 imagens no total, classificadas em 4 graus diferentes, com 1200 para cada grau. O trabalho utilizou de várias avaliações de métricas de resultados, incluindo sensibilidade (1,0000), especificidade (0,9917), precisão (0,9924), valor preditivo negativo (1,0000), precisão (1,0000) e pontuação F1-score (0,9962), a análise forneceu diferenciados resultados.

IV. METODOLOGIA

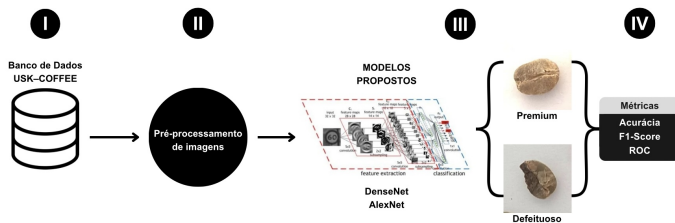


Fig. 4. Diagrama da metodologia (imagem feita pelo autor).

A. Banco de dados

Neste estudo, utilizamos o banco de dados público *USK-COFFEE*, que disponibiliza imagens com resolução de 400x400 *pixels*. O banco de dados propõe duas classes de grãos de café: *defect*, *premium*. O banco de dados contém no total 4.272 imagens de grãos de café em formato PNG e codificação RGB. Separados em duas classes, *premium* e *defeituosos*. Destes 2.136 seriam grãos *premium* e 2.136 seriam grãos *defeituosos*.

B. Pré-processamento de Imagens

A etapa de pré-processamento das imagens envolveu a aplicação da filtragem CLAHE [6] na base de dados, como demonstrado na fig. 5., com o objetivo de melhorar o desempenho do modelo. Para entender o impacto da filtragem no comportamento do modelo, foram comparadas imagens com a filtragem e as imagens originais, sem nenhum tipo de filtragem. Essa análise comparativa permitiu avaliar como cada técnica de pré-processamento influencia os resultados, contribuindo para a otimização do desempenho do modelo.

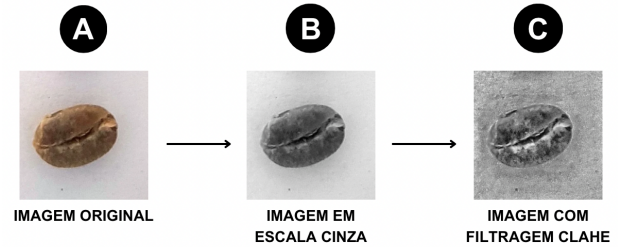


Fig. 5. Imagem do uso da filtragem em um grão de café (imagem feita pelo autor).

C. Distribuição

Para a realização dos experimentos, as imagens foram divididas da seguinte forma: 80% do banco, 3.424 imagens, foram utilizadas para treinamento, 10% do banco, 424 imagens, para teste e 10% do banco, 424 imagens, para validação, como a fig.6. está representando, com intuito de validar a generalização do modelo da melhor maneira.

D. Métricas de avaliação

A acurácia [11] na validação foi usada para escolher o melhor conjunto de hiperparâmetros durante a otimização. Para avaliar o modelo final, consideramos a precisão [11] e o F1-score [11] nos conjuntos de validação e teste. Comparar

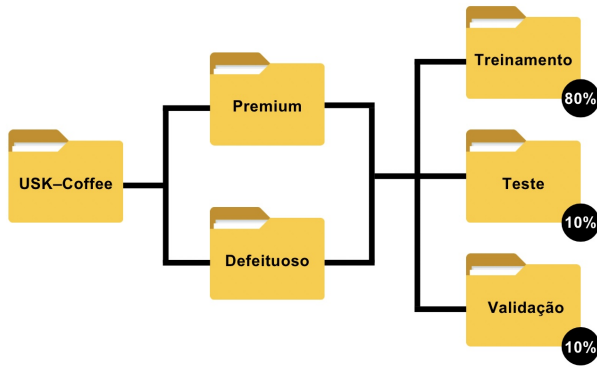


Fig. 6. Imagem da distribuição do banco de dados (imagem feita pelo autor).

esses índices entre a validação e o teste nos ajuda a entender a capacidade do modelo de aplicar o que aprendeu durante o treinamento em novos dados. Curvas ROC [11] foram usadas para analisar sua distinção, que é calculada a pela relação entre a taxa de verdadeiros positivos (TPR, também conhecida como sensibilidade) e a taxa de falsos positivos (FPR) em diferentes limiares de classificação

Acurácia: Medida que reflete a proporção de previsões corretas em relação ao total de previsões feitas. É calculada considerando tanto verdadeiros positivos (TP) quanto verdadeiros negativos (TN) [11].

$$\text{Acurácia} = \frac{TP + TN}{TP + TN + FP + FN}$$

F1-Score: É a média da precisão e da sensibilidade. Essa métrica é valiosa em casos de classes desbalanceadas, pois oferece um único valor que considera tanto falsos positivos (FP) quanto falsos negativos (FN) [11].

$$\text{F1-Score} = \frac{2 \cdot TP}{2 \cdot TP + FP + FN}$$

V. RESULTADOS

Os experimentos foram realizados utilizando das redes *Densenet* e *AlexNet*, e treinaram a CNN por 50 e 100 épocas utilizando duas abordagens para o pré-processamento do banco de dados: sem filtro e com o filtro CLAHE [6]. A seguir, são apresentados os resultados obtidos pelas métricas de avaliação como Loss, Acurácia, F1-Score e a Área Sob a Curva (ROC-AUC) para cada configuração.

A. Resultados sem Filtro em 100 épocas (Densenet)

Na rede *Densenet*, já é possível identificar resultados satisfatórios, com uma acurácia de 0.76 e F1-Score de 0.80, mas ainda denota-se possibilidade de melhoria no modelo apresentado, como demonstrado pela matriz confusão da fig. 7.

- **Val Loss:** 0.6601
- **Val Accuracy:** 0.7618
- **Melhor F1-Score:** 0.80

- **ROC-AUC:** 0.763

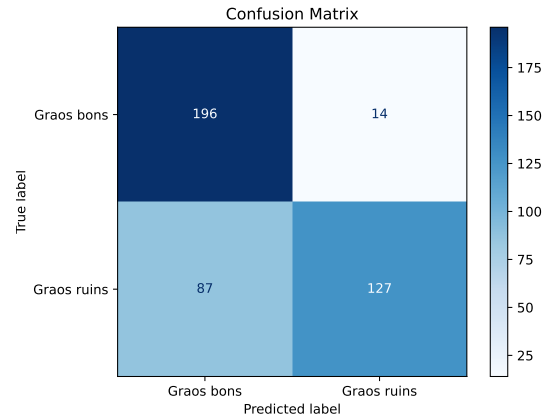


Fig. 7. Matriz de Confusão Sem Filtro (Feito pelo autor).

B. Resultados com Filtro em 100 épocas (Densenet)

Já com a utilização da filtragem CLAHE, é possível notar aumento significativo na acurácia, chegando a 0.82, como demonstrado pela matriz confusão da fig. 8.

- **Val Loss:** 1.249
- **Val Accuracy:** 0.8208
- **Melhor F1-Score:** 0.83
- **ROC-AUC:** 0.821

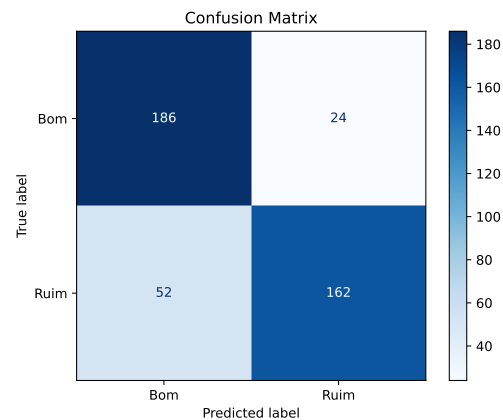


Fig. 8. Matriz de Confusão Com Filtro (Feito pelo autor).

C. Resultados sem Filtro em 100 épocas (Alexnet)

Na utilização da rede *Alexnet*, os resultados brutos já se iniciaram de forma muito satisfatória, chegando a 0.85 de acurácia, , como demonstrado pela matriz confusão da fig. 9.

- **Val Loss:** 0.5261
- **Val Accuracy:** 0.8255
- **Melhor F1-Score:** 0.87
- **ROC-AUC:** 0.857

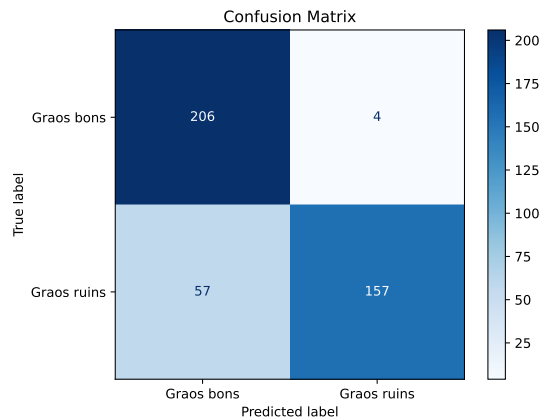


Fig. 9. Matriz de Confusão Sem Filtro (Feito pelo autor).

D. Resultados com Filtro em 100 épocas (Alexnet)

Já com o uso da filtragem CLAHE, também foi obtido um aumento nos valores da acurácia, chegando a 0.87, como demonstrado pela matriz confusão da fig. 10.

- **Val Loss:** 2.1515
- **Val Accuracy:** 0.8703
- **Melhor F1-Score:** 0.88
- **ROC-AUC:** 0.871

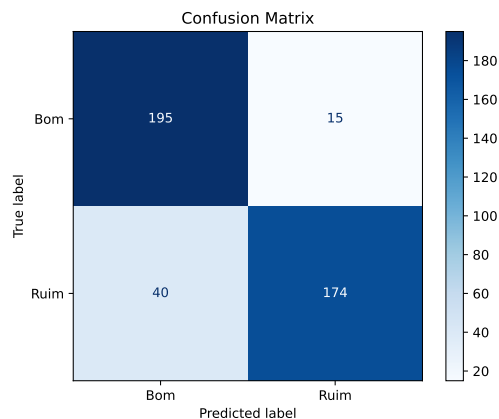


Fig. 10. Matriz de Confusão Com Filtro (Feito pelo autor).

E. Comparação Geral dos Resultados

Os resultados indicam que as abordagens sem filtro apresentaram uma performance pior do que as abordagens utilizando da filtragem CLAHE [6] em geral. A tabela I demonstra uma comparação de cada rede neural, com e sem o uso da filtragem CLAHE, comparando as métricas de acurácia, loss e

curva ROC. Foram implementados testes com maiores épocas, porém não houve melhora significativa nos trabalhos.

TABLE I

Métricas	Acurácia	Loss	Curva ROC	Acurácia	Loss	Curva ROC
Densenet	0.7618	0.6601	0.763	0.8208	1.249	0.821
Alexnet	0.8255	0.5261	0.857	0.8703	2.1515	0.871
Filtragem	Sem filtragem CLAHE			Com filtragem CLAHE		

F. Ambiente de Hardware

Os experimentos foram realizados em um PC equipado com uma GPU NVIDIA RTX 4060. Os experimentos foram desenvolvidos utilizando Python 3.12.4, PyTorch 2.4, torchvision 0.19 com CUDA Toolkit 12.1.

VI. CONCLUSÃO

Neste estudo, avaliamos o desempenho das redes Densenet e Alexnet aplicada à classificação da qualidade de grãos de café, utilizando dois tipos de imagens da base de dados, uma normal e outro com a filtragem CLAHE [6].

Os resultados obtidos mostraram que a aplicação de técnicas de filtragem melhora significativamente o desempenho das duas redes, especialmente em termos de precisão e F1-score. Especificamente, o filtro CLAHE [6] demonstrou ser mais eficaz em realçar características relevantes dos grãos de café, resultando em uma melhor taxa de acertos em comparação com os dados sem filtragem. A utilização de filtragem não apenas melhorou a capacidade do modelo de generalizar para dados desconhecidos, mas também reduziu a ocorrência de erros críticos, como falsos negativos.

Esses achados reforçam a importância de pré-processamento de imagens em tarefas de classificação, sugerindo que com uma escolha adequada de técnicas de filtragem pode ser decisiva para melhorar a performance de redes neurais profundas em aplicações específicas. Além disso, o estudo abre caminho para futuras pesquisas que possam explorar o uso combinado de múltiplas técnicas de filtragem ou a aplicação dessas metodologias em outras arquiteturas de redes neurais e conjuntos de dados.

VII. ACKNOWLEDGMENTS

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001. Bernardo Silva Ribeiro Duarte received a scholarship from PIBIC/CNPq

REFERENCES

- [1] GOV (14 de abril de 2023). Brasil é o maior produtor mundial e o segundo maior consumidor de café. Acesso em: <https://www.gov.br/agricultura/pt-br/assuntos/noticias/brasil-e-o-maior-produtor-mundial-e-o-segundo-maior-consumidor-de-cafe>
- [2] ARUNKUMAR, J. R.; BERIHUN MENGIST, Tagele. Developing Ethiopian Yirgacheffe Coffee Grading Model using a Deep Learning Classifier. International Journal of Innovative Technology and Exploring Engineering, v. 9, n. 4, p. 3303-3309, 2020

- [3] RONNEBERGER, Olaf; FISCHER, Philipp; BROX, Thomas. U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer International Publishing, 2015. p. 234–241. Disponível em: < https://doi.org/10.1007/978-3-319-24574-4_28 >
- [4] FossAnalytics. Disponível em: <https://www.fossanalytics.com/pt-br/about-foss/who-we-are>
- [5] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” in Advances in Neural Information Processing Systems, 2012, pp. 1097–1105.
- [6] Karel Zuiderveld. Graphics gems iv. chapter Contrast Limited Adaptive Histogram Equalization, pages 474–485. Academic Press Professional, Inc., San Diego, CA, USA, 1994. Disponível em: <https://doi.org/10.1016/B978-0-12-336156-1.50061-6>
- [7] WANG, Ping et al. Deep convolutional neural network for coffee bean inspection. Sensors and Materials, v. 33, n. 7, p. 2299-2310, 2021. Disponível em: <https://doi.org/10.18494/SAM.2021.3277>
- [8] PINTO, Carlito et al. Classification of Green coffee bean images based on defect types using convolutional neural network (CNN). In: 2017 international conference on advanced informatics, concepts, theory, and applications (ICAICTA). IEEE, 2017. p. 1-5.
- [9] UNAL, Yavuz et al. Application of pre-trained deep convolutional neural networks for coffee beans species detection. Food Analytical Methods, v. 15, n. 12, p. 3232-3243, 2022.
- [10] K. Eves and J. Valasek, “Adaptive control for singularly perturbed systems examples,” Code Ocean, Aug. 2023. [Online]. Available: <https://codeocean.com/capsule/4989235/tree>
- [11] BELCIUG, S. Artificial intelligence in cancer: diagnostic to tailored treatment. Cambridge: Academic, 2020
- [12] Hassan, E. Enhancing coffee bean classification: a comparative analysis of pre-trained deep learning models. Neural Comput Applic 36, 9023–9052 (2024). <https://doi.org/10.1007/s00521-024-09623-z>