

Classificação Inteligente do Teste de Corte do Cacau com Redes Neurais Convolucionais Profundas

Daniel Benoliel Malcher, Elloá B. Guedes

¹Grupo de Pesquisa em Sistemas Inteligentes
Universidade do Estado do Amazonas (UEA)
Av. Darcy Vargas, 1200 – Manaus – Amazonas
{dbm.eng18, ebgcosta}@uea.edu.br

Resumo. *Este trabalho considera a classificação automática de amêndoas de cacau no Teste de Corte como uma tarefa de Visão Computacional abordada com Redes Neurais Convolucionais. Com uma base de dados com exemplos de 14 classes, foram realizados experimentos com validação cruzada holdout os quais mostraram as redes Inception, VGG-16 e EfficientNetB0 como tendo melhor desempenho para a tarefa (acurácia média igual a $68,28\% \pm 3,21$, $65,54\% \pm 3,05$ e $55,04\% \pm 2,03$, respectivamente). Tais modelos foram combinados em um comitê com votação suave cuja acurácia média foi igual a $89,79\% \pm 0,92$, o que superou os modelos individuais avaliados. Este é um resultado promissor que pode apoiar objetivos estratégicos na melhoria da qualidade da produção do cacau no Brasil.*

Abstract. *This work aims at addressing the problem of classification of cocoa beans at the Cut-Test as a Computer Vision task with Convolutional Neural Networks. By using a realistic balanced dataset with images labeled by experts into fourteen classes, we carried out an experimental evaluation with holdout cross-validation and repetitions that showed Inception, VGG-16 and EfficientNetB0 as Top-3 architectures with best performance regarding accuracy ($68,28\% \pm 3,21$, $65,54\% \pm 3,05$ and $55,04\% \pm 2,03$, respectively). We combined such models into a soft voting ensemble with resulting accuracy of $89,79\% \pm 0,92$ that outperforms all individual solutions evaluated. This is a promising result that may support strategical goals of improving the quality of cocoa beans production in Brazil.*

1. Introdução

O cacau (*Theobroma cacao*) é um fruto originário da América Central que serve de matéria prima para a produção de manteiga de cacau, para a extração de flavonóides, para fabricação de chocolate, dentre outros [Ferreira et al. 2013]. Em 2020 o Brasil produziu 264.586 t de cacau, gerando R\$ 3,182 M em lucros, o que representa apenas 4,6 % da participação global no mercado de cacau. Em termos geográficos, a região Norte do País lidera a produção nacional (53,2 %); a Bahia é o único estado produtor na região Nordeste; e nas regiões Norte do Espírito Santo e de Minas Gerais encontram-se 2,8 % da área colhida nacionalmente [Brainer 2021].

No Brasil, o Ministério da Agricultura, Pecuária e Abastecimento (MAPA), por meio da Comissão Executiva do Plano da Lavoura Cacaueira (CEPLAC), tem como um de seus objetivos tornar o País autossuficiente na produção de cacau para o mercado interno até 2025, fortalecendo a cadeia produtiva para produção de amêndoas de cacau com qualidade e responsabilidade sócio-ambiental, além do estímulo ao consumo consciente de seus derivados [MAPA 2022]. Exemplos de desafios endógenos a serem enfrentados para alcançar esta meta são lidar com as infestações por fungos, as estiagens e a falta de modernização da produção; enquanto que alguns dos principais desafios exógenos identificados são a instabilidade macroeconômica,

a superprodução de cacau em outros Países e a queda dos preços do produto [Bahia de Aguiar and De Moura Pires 2019].

Um dos aspectos cruciais para elevar o valor agregado da produção de cacau é a qualidade do beneficiamento, processo do qual se extraem os subprodutos do cacau. O beneficiamento inicia-se com a colheita dos frutos maduros, separação (maduros, sobremaduros e aprodrecidos) e segue-se com a etapa de corte, na qual cada amêndoa (semente) deve ser retirada manualmente da casca e a polpa deve ser inspecionada. Uma vez que as sementes foram separadas, devem ser armazenadas em sacos plásticos sem contato com o ar e transportadas para a fermentação [Ferreira et al. 2013]. Uma vez que a fermentação é concluída, segue-se para uma inspeção da qualidade final das sementes, denominada *teste de corte*.

O teste de corte inicia com procedimentos para garantir a seleção de uma boa amostra de uma dada produção: todas as amêndoas produzidas são separadas em quatro partes, das quais uma é eliminada. Misturam-se as restantes e reparte-se novamente de igual modo até que reste 1 kg de amêndoas, caracterizando a amostra a ser analisada [Ferreira et al. 2013]. Todas as sementes desta amostra são cortadas longitudinalmente e dispostas em uma tábua e devem ser inspecionadas. Conforme instrução normativa do MAPA, a inspeção deve ser conduzida por agrônomos ou técnicos em agropecuária devidamente habilitados com vistas a estabelecer classificações para as amêndoas da amostra no que diz respeito à cor e diversos tipos de defeitos [MAPA 2008, Saito 2019]. É importante ressaltar que a realização do teste de corte demanda mão de obra altamente especializada; reside na aferição dos achados visuais das amêndoas, o que pode estar sujeito à subjetividade; e pode ter seu desempenho e abrangência comprometidos em função do número de amêndoas a serem inspecionadas em um dado intervalo de tempo. Em face dos desafios práticos evidenciados na realização do referido teste, há a oportunidade para propor e avaliar soluções de Visão Computacional para este contexto, especialmente considerando que esta área da Computação é uma pedra angular da Agricultura Digital [Tang et al. 2002] e um eixo estratégico a ser desenvolvido e adotado nas próximas duas décadas na Agricultura brasileira [EMBRAPA 2014].

Diante do exposto, o objetivo deste trabalho consistiu em apresentar os resultados da utilização de Redes Neurais Convolucionais Profundas (CNNs, do inglês *Convolutional Neural Networks*) aplicadas ao problema de classificação de imagens de amêndoas na ocasião do teste de corte. Para tanto, fez-se uso de uma base de dados pública, gratuita e rotulada contendo mais de mil exemplos de amêndoas para fornecer experiência à diferentes arquiteturas canônicas de CNNs. Os resultados obtidos mostram que a tarefa é desafiadora, mas que há potencial de uso das CNNs segundo um comitê com votação majoritária suave para classificação automática e inteligente, com acurácia média de $89,79\% \pm 0,92$ e F_1 -Score de $89,75\% \pm 0,93$ em um cenário de validação cruzada *holdout* com repetições.

Com vistas a apresentar o que se propõe, o presente artigo está estruturado como segue. Os trabalhos relacionados, incluindo uma replicação de experimentos neles propostos, encontram-se descritos na Seção 2. A metodologia utilizada para condução das atividades é apresentada na Seção 3. Os resultados obtidos e a discussão são mostrados na Seção 4. Por fim, as considerações finais e perspectivas de trabalhos futuros são apresentados na Seção 5.

2. Trabalhos Relacionados

Trabalhos previamente encontrados na literatura que abordam o problema do teste de corte como uma tarefa de Visão Computacional efetuam a extração de características das imagens para obtenção de atributos preditores que serão utilizados para o treinamento de algoritmos clássicos de Aprendizado de Máquina. Nestes, destacam-se as características texturais de Haralick [Haralick et al. 1973], obtidas a partir da matriz de co-ocorrência de níveis de cinza,

a média dos componentes de cor de cada canal da imagem e ainda o perímetro, uma característica estrutural das sementes [Santos et al. 2018, Santos et al. 2019a].

Em ambos os trabalhos, as redes neurais artificiais *multilayer perceptron* destacaram-se dentre as técnicas investigadas para classificação, obtendo os melhores resultados experimentais: 99,73 % e 92,97 % de acurácia, respectivamente. Conforme análise de importância dos atributos para classificação realizada em um destes trabalhos [Santos et al. 2019a], dentre as dez características mais relevantes, 70 % são de natureza textural e 20 % são relacionadas à cor. A técnica de validação cruzada utilizada por ambos os autores é do tipo *holdout* com 80 % dos exemplos disponíveis para treino e 20 % para teste. Embora os resultados sejam animadores, um dos trabalhos indaga a possibilidade de ocorrência de *overfitting* para as configurações das redes neurais abordadas [Santos et al. 2018].

Para investigar a possibilidade de ocorrência de *overfitting* indicada pelos autores, no escopo deste trabalho foi feita uma replicação experimental da classificação automática proposta no trabalho relacionado mais recente, em que foi considerada a mesma base de dados, a mesma grade de busca de parâmetros das redes neurais artificiais e 16 atributos preditores, sendo 13 características texturais de Haralick (81,25 % do total de atributos preditores) e 3 componentes de cor. Não foi possível extrair a característica estrutural pela omissão da menção do algoritmo utilizado originalmente.

Com a utilização de *scripts* na linguagem Python, disponíveis em <https://github.com/elloa/wcama2022>, os resultados experimentais da busca em grade apontaram a mesma arquitetura identificada pelos autores mencionados como sendo a melhor adequada para a tarefa, mas com algumas diferenças nos hiperparâmetros. Entretanto, ao observar a taxa de acerto deste modelo sob as mesmas condições de validação, houve uma grande diferença nas métricas de desempenho: a acurácia e o F_1 -Score obtidos no conjunto de testes pelo melhor modelo foram iguais a 39,36 % e 35,62 %, respectivamente. A matriz de confusão da Fig. 1 ilustra as previsões do modelo no conjunto de testes desse cenário experimental. Uma posterior validação cruzada *10-fold* com 5 repetições resultou em acurácia igual a 44,63 % \pm 3,77 e F_1 -Score igual a 44,49 % \pm 3,76, corroborando para a suposição de *overfitting*.

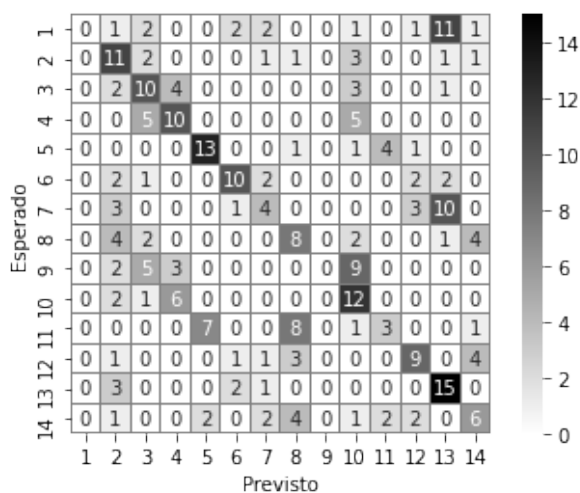


Figura 1: Matriz de confusão obtida com uma rede neural artificial com 28 neurônios na camada oculta, taxa de aprendizado igual a 10^{-1} e 10^4 épocas de treinamento.

Em face dos achados durante a replicação dos experimentos, observa-se a necessi-

dade de investigação de métodos mais robustos de classificação para o problema em questão, os quais possam simultaneamente aumentar a taxa de acerto perante o problema e também incorporar técnicas de regularização e validação durante o treinamento que minimizem *overfitting* nos classificadores. A busca por soluções para estes aspectos foram motivadores para a realização do presente trabalho, cuja metodologia é descrita a seguir.

3. Materiais e Métodos

O problema da aferição da qualidade das amêndoas de cacau foi abordado no escopo deste trabalho como uma tarefa de classificação multi-classe mediante Aprendizado Supervisionado com CNNs, modelos do estado da arte de *Deep Learning* para este tipo de tarefa de Visão Computacional [Khan et al. 2018]. Os dados experimentais, modelos, a avaliação de desempenho e a elaboração de comitês são descritos detalhadamente nas subseções a seguir.

3.1. Dados Experimentais

Como fonte de experiência para aprendizado das CNNs, a base de dados *Cut-Test-Classified Cocoa Beans* [Santos et al. 2019b] foi utilizada no escopo desse trabalho. Esta base foi escolhida em virtude de ser balanceada, conter exemplos coletados no Brasil e também por ser a mesma utilizada nos trabalhos relacionados na literatura, o que propicia uma comparação equânime dos resultados. Nessa base de dados há 1.400 exemplos de imagens de sementes de cacau no teste de corte, rotuladas por especialistas segundo 14 classes, representadas por números inteiros segundo a ordem lexicográfica, conforme segue: (1) Aglutinada (*Agglutinated*); (2) Quebradiça (*Brittle*); Compartmentalizada (*Compartmentalized*) com cores (3) marrom (*brown*), (4) parcialmente violeta (*partially purple*), (5) violeta (*purple*), (6) ardósia (*slaty*) e (7) branca (*white*); (8) Achatada (*Flattened*); (9) Embolorada (*Moldered*); e Chocha (*Plated*) com as mesmas cores do caso compartmentalizada (Classes 10 a 15). A base de dados encontra-se balanceada, com 100 exemplos de cada classe, alguns dos quais encontram-se ilustrados na Figura 2. As catorze classes dizem respeito à cor, formato e segmentação interna da semente, o que se relaciona com a qualidade final da amêndoa e, por conseguinte, com o seu valor comercial.

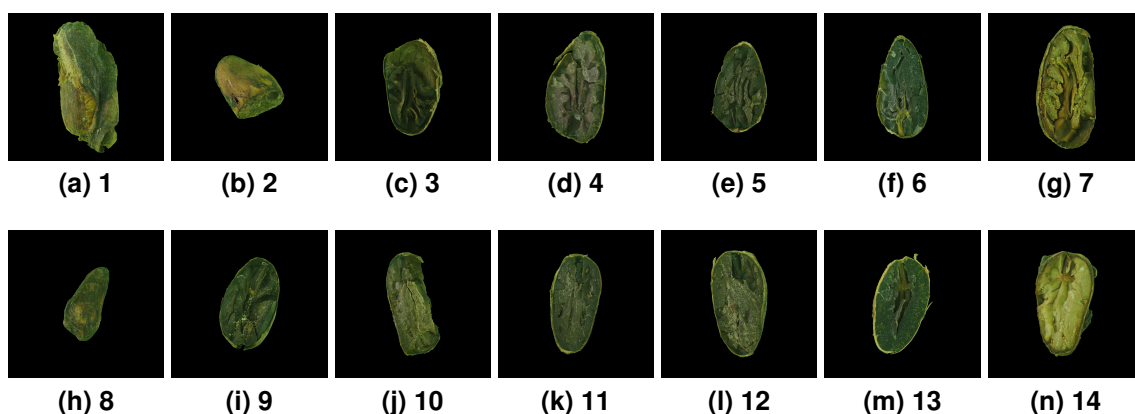


Figura 2: Exemplos de sementes da base de dados e suas respectivas classes.
Fonte: [Santos et al. 2019b].

3.2. Arquiteturas, Parâmetros e Hiperparâmetros

Levando em consideração a adoção de CNNs como modelo de referência na tarefa considerada, foi efetuado um levantamento na literatura visando elencar redes com comprovada

eficiência em problemas de classificação de imagens [Russakovsky et al. 2015]. Como resultado, as seguintes CNNs foram elencadas:

1. **LeNet**. Arquitetura originalmente proposta para a identificação de dígitos manuscritos, é composta por duas camadas convolucionais com camadas de *max pooling*. Após esse processo de extração de características, duas camadas densas produzem a classificação resultante [LeCun et al. 1998];
2. **AlexNet** Apresenta 5 camadas convolucionais acompanhadas de 4 camadas de *max pooling*, tendo sido desenvolvida para expandir o que foi proposto originalmente pela *LeNet* para contemplar *datasets* maiores e com mais características. Para produzir a saída, possui 3 camadas densas. Quando comparada à sua predecessora, possui melhor desempenho em diversos problemas de classificação de imagens [Krizhevsky et al. 2012];
3. **MobileNet**. Proposta com o objetivo de ser uma arquitetura embarcável em dispositivos móveis, sua principal característica é possuir apenas 2 camadas convolucionais e não fazer uso de camadas de *pooling*, o que demanda menos recursos computacionais quando comparada às alternativas existentes na literatura [Howard et al. 2017];
4. **VGG-16**. Possui 16 camadas de profundidade em conjunto com um grande número de convoluções com filtros pequenos, de tamanho 3×3 . Além disso, utiliza *dropout* de pesos com vistas a evitar *overfitting* [Simonyan and Zisserman 2014]. A utilização desta CNN neste trabalho foi feita mediante transferência de aprendizado com pesos pré-treinados da base de dados ImageNet [Deng et al. 2009];
5. **Inception**. Em sua terceira versão, é uma topologia que considera ramificações internas e que balanceia o aumento da profundidade sem aumentar significativamente o número de parâmetros, o que geralmente resulta em alta acurácia em problemas de Visão Computacional [Szegedy et al. 2015a, Szegedy et al. 2015b];
6. **EfficientNet**. Uma família de arquiteturas de CNNs projetadas para balancear os parâmetros de profundidade, largura e resolução das redes neurais convolucionais com vistas a aumentar a performance e diminuir o número de parâmetros e operações requeridas no treinamento [Tan and Le 2019]. Foram consideradas as topologias B0, por conter o menor número de parâmetros, e a B3, por ter número de parâmetros próximo à Inception.

Os hiperparâmetros foram utilizados conforme os seguintes critérios: (i) considerou-se um número máximo de 300 épocas; (ii) para evitar *overfitting*, adotou-se a técnica de *Early Stopping* com paciência de 100 épocas; (iii) as imagens de entrada eram coloridas, foram normalizadas e, exceto para as EfficientNets, possuíam dimensões 299×299 ; (iv) adotou-se a técnica de *Model Checkpoint* para monitorar a acurácia no conjunto de validação e persistir em disco o conjunto de pesos que provia melhor generalização; (v) a taxa de aprendizado foi igual a 10^{-4} ; (vi) adotou-se o otimizador Nadam por utilizar o momentum de Nesterov para lidar com platôs no algoritmo do gradiente descendente [Dozat 2016]; e (vii) a escolha do hiperparâmetro relativo ao tamanho dos *batches* deu-se de forma empírica para cada arquitetura, balanceando o número de parâmetros e a utilização dos recursos computacionais disponíveis, a exemplo de memória principal. Conforme sugestões de ajuste fino disponíveis na literatura [Fu 2020], a rede EfficientNetB0 utilizou as imagens normalizadas com resolução de 224 pixels e a EfficientNetB3 considerou resolução de 300 pixels.

No escopo desse trabalho, as camadas densas e finais das redes VGG-16, Inception e EfficientNets foram redimensionadas para conter um número de neurônios compatível com as classes do problema abordado e também foram inseridas camadas de *dropout* entre elas com o objetivo de colaborarem na regularização do treinamento, visto que essas arquiteturas

são as mais complexas dentre as mencionadas, o que colabora para mitigar problemas como *overfitting*, *underfitting*, desvanecimento do gradiente durante o treinamento, dentre outros.

3.3. Avaliação de Desempenho

Para avaliar os modelos selecionados perante o problema proposto, considerou-se 5 repetições distintas de validações cruzadas do tipo *holdout*, com 70 % dos dados disponíveis para treino, 10 % para validação e 20 % para testes, em que esta última partição foi utilizada para aferir o desempenho. Os valores médios das seguintes métricas de desempenho nas 5 repetições sintetizaram tal avaliação, em que $C = \{1, \dots, 15\}$ denota o rótulos das classes do problema:

$$\text{Acurácia} = \frac{1}{|C|} \sum_{c \in C} \left(\frac{\text{TP}_c + \text{FP}_c}{\text{TP}_c + \text{FP}_c + \text{FN}_c + \text{FP}_c} \right) \quad \text{Precisão} = \frac{1}{|C|} \sum_{c \in C} \left(\frac{\text{TP}_c}{\text{TP}_c + \text{FP}_c} \right) \quad (2)$$

$$\text{Revocação} = \frac{1}{|C|} \sum_{c \in C} \left(\frac{\text{TP}_c}{\text{TP}_c + \text{FN}_c} \right) \quad (3) \quad F_1\text{-Score} = 2 \left(\frac{\text{Precisão} \cdot \text{Revocação}}{\text{Precisão} + \text{Revocação}} \right) \quad (4)$$

3.4. Comitês de Modelos

Uma vez que os melhores pesos das CNNs foram persistidos em disco após o treinamento, além do uso individual de tais arquiteturas também foram consideradas estratégias de Aprendizado por Comitê (*Ensemble Learning*). Há diversos relatos na literatura em diferentes domínios de aplicação que mostram que se há bons aspectos de diversidade no comitê, os vieses dos classificadores individuais costumam ser mitigados, o que favorece um melhor desempenho coletivo [Polikar 2006, Zhang and Ma 2012]. Levando isto em consideração, foram tomadas as três melhores CNNs com desempenho individual no tocante à acurácia para compor comitês com as seguintes estratégias de votação:

1. **Votação Majoritária Simples.** Nesta estratégia a mesma entrada é fornecida às CNNs integrantes do comitê e obtém-se a classificação produzida por cada uma delas. A classificação produzida pelo comitê será então o resultado majoritário das classificações individuais;
2. **Votação Majoritária Suave.** Considera os vetores de distribuição de probabilidades perante as classes do problema produzidos pelas CNNs a partir de uma mesma entrada. Tais vetores são somados e a classificação resultante é atribuído à classe com maior valor acumulado. Essa estratégia reflete a confiança majoritária nas classificações das CNNs individuais.

4. Resultados e Discussão

Para a execução dos *scripts* de treinamento das CNNs foi utilizado um servidor com a seguinte configuração: processador Intel(R) Core(TM) i7-8700 CPU @ 3.20GHz, 32 GB de memória principal, 960 GB de memória secundária e 2 placas de vídeo NVIDIA GTX 1080 Ti com 11 GB de VRAM para aceleração em *hardware* do treinamento. Os resultados obtidos dos testes para as diferentes arquiteturas de CNNs consideradas encontram-se dispostos na Tabela 1 sob a forma de média e desvio padrão das 5 repetições.

Conforme os resultados observados e considerando a métrica de acurácia, verifica-se que a CNN Inception obteve desempenho superior às demais, seguida pela VGG-16 e então

Tabela 1: Resultado experimental das arquiteturas de CNNs perante 5 repetições da validação cruzada *holdout*.

	Acurácia	F ₁ -Score	Precisão	Revocação	Parâmetros
LeNet	43,00 % ± 7,33	40,93 % ± 8,26	42,19 % ± 10,88	43,00 % ± 7,33	8.537.784
AlexNet	39,14 % ± 4,93	38,71 % ± 4,75	41,34 % ± 4,84	39,14 % ± 4,93	83.186.798
MobileNet	34,92 % ± 5,06	33,34 % ± 5,71	36,82 % ± 8,78	34,92 % ± 5,06	36.136.470
VGG-16	65,54 % ± 3,05	60,38 % ± 3,29	71,93 % ± 9,19	57,57 % ± 5,74	31.978.654
Inception	68,28 % ± 3,21	67,49 % ± 2,51	70,74 % ± 1,54	68,28 % ± 3,21	22.026.582
EfficientNetB0	55,04 % ± 2,03	53,90 % ± 2,90	58,01 % ± 2,06	53,07 % ± 2,83	4.194.305
EfficientNetB3	26,13 % ± 16,52	29,99 % ± 12,86	51,16 % ± 13,10	25,64 % ± 13,38	10.954.893

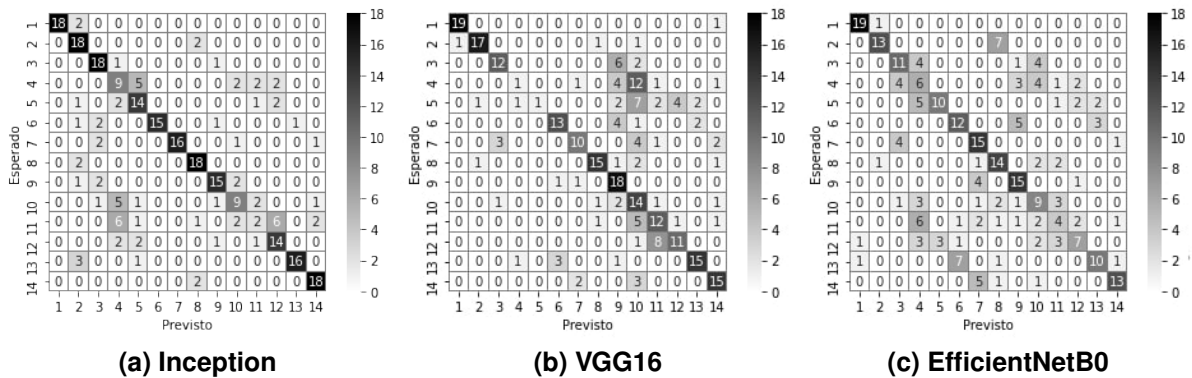


Figura 3: Matrizes de confusão das três CNNs com maior acurácia para uma das repetições da validação cruzada.

pela EfficientNetB0. As matrizes de confusão para previsões feitas por estas redes encontram-se ilustradas na Fig. 3, em que é possível observar que houve consistência na classificação correta para algumas classes em todos os modelos (Classes 1 e 14, por exemplo) e que algumas classes mostraram-se particularmente mais sujeitas à erros de classificação (Classes 4 e 5, por exemplo). Para as demais redes consideradas, a acurácia ficou abaixo de 50 % independentemente do número de parâmetros nelas presentes, o que ressalta a importância de efetuar experimentos com diversas arquiteturas distintas de CNNs.

Uma vez identificados os três melhores modelos individuais para a tarefa em questão, partiu-se para a elaboração dos comitês conforme as estratégias previamente mencionadas. Os resultados obtidos encontram-se descritos em termos de média e desvio padrão na Tabela 2 e as matrizes de confusão encontram-se ilustradas na Figura 4. Nota-se que a diagonal principal da matriz de confusão do comitê com votação majoritária suave é densa, o que demonstra a robustez na correteza das previsões.

Tabela 2: Desempenho dos comitês por estratégia de votação.

	Acurácia	F ₁ -Score	Precisão	Revocação
Votação Majoritária Simples	70,93% ± 2,28	70,71% ± 2,30	72,86% ± 1,56	70,93% ± 2,28
Votação Majoritária Suave	89,79% ± 0,92	89,75% ± 0,93	90,30% ± 0,98	89,79% ± 0,92

A elaboração dos comitês mostrou-se uma estratégia bem sucedida para aumentar a eficiência na tarefa considerada, sem um ônus adicional de treinamento ou inclusão significativa de novos parâmetros. Em particular, conforme Fig. 5, é possível verificar um desempenho

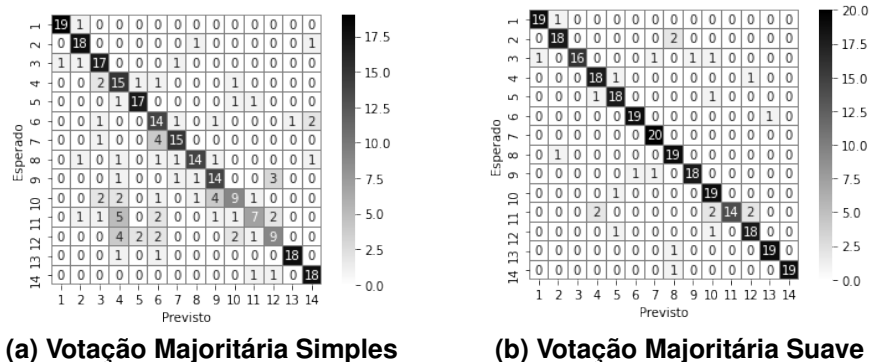


Figura 4: Matrizes de confusão dos comitês para uma das repetições da validação cruzada.

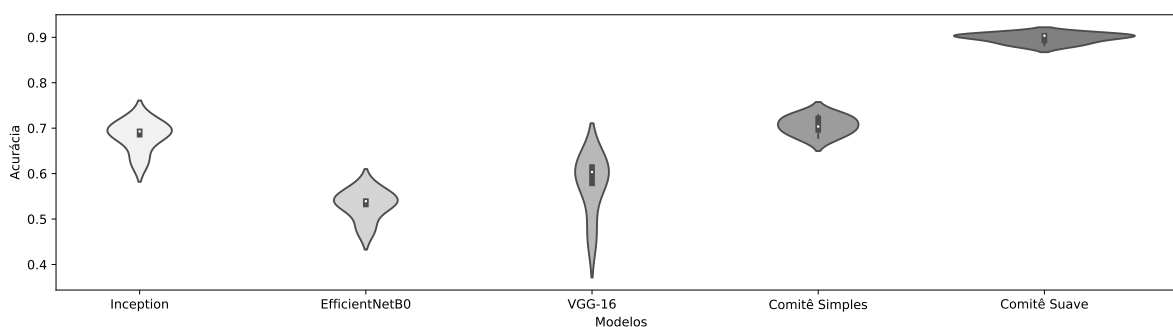


Figura 5: Distribuição da acurácia das três melhores CNNs e dos comitês.

experimental superior do comitê com votação majoritária suave, com ganho percentual de acurácia de 21,51 % quando comparado à Inception e de 18,86 % quando comparado ao outro comitê elaborado. Em face de tais achados experimentais, considera-se essa a solução de referência para a tarefa de classificação considerada no escopo deste trabalho.

Ao levar em conta os resultados experimentais nesta mesma base de dados relatados em outros trabalhos, verifica-se um decréscimo percentual de 9,94 % da acurácia da solução de referência. Porém, quando a solução de referência aqui proposta é comparada com os valores obtidos da reprodução dos experimentos da literatura, conforme Seção 2, observa-se um ganho de 50,43 %, o que verifica-se compatível com o estado da arte em que tarefas de classificação de imagens quando abordadas por CNNs consistentemente registram desempenho superior às abordagens tradicionais de Visão Computacional [Russakovsky et al. 2015].

5. Considerações Finais

Neste trabalho foi proposta uma solução inteligente para o teste de corte do cacau baseada no uso de comitês de CNNs. Para tanto, considerou-se dados experimentais oriundos de uma base de dados realística para esta tarefa, modelos de CNNs do estado da arte de *Deep Learning* e uma estratégia de validação cruzada com repetições para mitigar os vieses decorrentes de eventuais flutuações estatísticas. Os resultados obtidos ressaltaram um comitê com votação majoritária suave das arquiteturas Inception, VGG-16 e EfficientNetB0 como tendo melhor desempenho para a tarefa, com acurácia média igual a $89,79\% \pm 0,92$, superando um outro comitê proposto e também todas as CNNs individuais avaliadas.

A solução obtida apresenta um grande potencial de uso prático no teste de corte do cacau, mas ressalta-se a importância de estender a sua avaliação considerando o uso de ima-

gens em condições menos controladas de iluminação e com resoluções menores como, por exemplo, quando adquiridas por *smartphones*. Experimentos adicionais que promovam um comparativo com especialistas humanos também são essenciais antes da sugestão de adoção da solução proposta em ampla escala.

Em trabalhos futuros, almeja-se ampliar a avaliação da solução na direção mencionada, realizar novos experimentos com outras estratégias de aprendizado por comitê e também verificar se o ajuste fino de parâmetros e hiperparâmetros das CNNs pode colaborar para um aumento nas métricas de desempenho.

Agradecimentos

Os autores agradecem o apoio financeiro da Fundação de Amparo à Pesquisa do Estado do Amazonas (FAPEAM) por meio do programa PAIC 2020-2021 e 2021-2022. Agradecem também o apoio material do Laboratório de Sistemas Inteligentes (LSI) da Universidade do Estado do Amazonas.

Referências

- Bahia de Aguiar, P. C. and De Moura Pires, M. (2019). A região cacauzeira do sul do estado da Bahia (Brasil): crise e transformação. *Cuad. Geogr. Rev. Colomb. Geogr.*, 28(1):192–208.
- Brainer, M. S. D. C. P. (2021). Produção de cacau. *Caderno Setorial do Escritório Técnico de Estudos Econômicos do Nordeste*, (149):1–23.
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE.
- Dozat, T. (2016). Incorporating Nesterov Momentum into Adam. In *Proceedings of the 4th International Conference on Learning Representations*, pages 1–4, San Juan, Puerto Rico.
- EMBRAPA (2014). Visão 2014–2034 – O Futuro do Desenvolvimento Tecnológico da Agricultura Brasileira (Síntese). Distrito Federal, Brasil, 55p.
- Ferreira, A. C. R., Ahnert, D., de Melo Neto, B. A., and Mello, D. L. N. (2013). *Guia de Beneficiamento de Cacau de Qualidade*. Instituto Cabruca, Ilhéus, Bahia.
- Fu, Y. (2020). Image classification via fine-tuning with Efficientnet. Keras. Disponível em https://keras.io/examples/vision/image_classification_efficientnet_fine_tuning/. Acesso em 11 de maio de 2022.
- Haralick, R. M., Shanmugam, K., and Dinstein, I. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-3(6):610–621.
- Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., and Adam, H. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. Disponível em <http://arxiv.org/abs/1704.04861>. Acesso em 11 de maio de 2022.
- Khan, S., Rahmani, H., Shah, S., and Bennamoun, M. (2018). *A Guide to Convolutional Neural Networks for Computer Vision*. Number 1 in Synthesis Lectures on Computer Vision. Morgan & Claypool Publishers.
- Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems*, pages 1097–1105, USA. ACM.

- LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324.
- MAPA (2008). Instrução normativa n. 38, de 23 de junho de 2008. Disponível em https://members.wto.org/crnattachments/2008/sps/BRA/08_2038_00_x.pdf. Acesso em 11 de maio de 2022.
- MAPA (2022). Comissão executiva do plano da lavoura cacauera. Disponível em <https://www.gov.br/agricultura/pt-br/assuntos/ceplac>. Acesso em 11 de maio de 2022.
- Polikar, R. (2006). Ensemble based systems in decision making. *IEEE Circuits and Systems Magazine*, 6(3):21–45.
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., and Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252.
- Saito, S. (2019). Entendendo o relatório de classificação de amêndoas. Centro de Inovação do Cacau. Disponível em https://pctsb.org/cic/wp-content/uploads/2019/11/Nota_tecnica_1_Como_e_feita_a_classificacao_no_CIC.pdf. Acesso em 11 de maio de 2022.
- Santos, F., Canuto, A., Bedregal, B., Palmeira, E., and Silva, I. (2019a). Supervised methods applied to the construction of a vision system for the classification of cocoa beans in the cut-test. In *Anais do XVI Encontro Nacional de Inteligência Artificial e Computacional*, pages 72–83, Porto Alegre, RS, Brasil. SBC.
- Santos, F., Palmeira, E., and Jesus, G. (2019b). An Image Dataset of Cut-Test-Classified Cocoa Beans. *Data in Brief*, 24:103916.
- Santos, F. A., Palmeira, E. S., and Jesus, G. Q. (2018). Color, structural and textural features for the classification of a Cocoa beans image dataset using Artificial Neural Network. In *Anais do XIV Workshop de Visão Computacional*, pages 80–84, Ilhéus, Bahia.
- Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. Disponível em <https://arxiv.org/abs/1409.1556>. Acesso em 11 de maio de 2022.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015a). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, volume 1, pages 1–9, Boston, Massachusetts, USA. IEEE.
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2015b). Rethinking the inception architecture for computer vision. Disponível em <https://arxiv.org/abs/1512.00567>. Acesso em 11 de maio de 2022.
- Tan, M. and Le, Q. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97, pages 6105–6114.
- Tang, S., Zhu, Q., Zhou, X., Liu, S., and Wu, M. (2002). A conception of digital agriculture. In *IEEE International Geoscience and Remote Sensing Symposium*, pages 3026–3028, Canada. IEEE.
- Zhang, C. and Ma, Y. (2012). *Ensemble Machine Learning: Methods and Applications*. Springer Science+Business Media, USA.