

Application of Convolutional Neural Network in Coffee Capsule Count Aiming Collection System for Recycling

1st Henrique Wippel Parucker da Silva
Bel. em Ciência da Computação (UDESC)
Joinville, Brasil
henriquewps@gmail.com

2nd Gilmário Barbosa dos Santos
Dep. Ciência da Computação
Universidade do Estado de Santa Catarina (UDESC)
Joinville, Brasil
gilmario.santos@udesc.br

Resumo—The coffee capsules brought practicality and speed in the preparation of the drink. However, with its popularization came a major environmental problem, the generation of a large amount of garbage, which for 2021 has an estimated 14 thousand tons of garbage, only coming from the capsules. To avoid this disposal it is necessary to recycle them, however it is not a trivial job, since they are composed of various materials, as well as the collection of these capsules presents challenges. Therefore, a collection system is of great value, which, in addition to being automated, generates bonuses proportional to the quantity of discarded capsules. This work is dedicated preliminary tests on the development of such a system using a convolutional neural network for the detection of coffee capsules. This algorithm was trained with two image sets, one containing images with reflection and the other without, which presented an accuracy of approximately 97%.

Index Terms—Image Processing, Image classification, Coffee Capsules, Convolutional Neural Networks.

I. INTRODUÇÃO

A planta do café teve origem na África, mais especificamente na Etiópia, Quênia e Sudão [1]. Entretanto, foi no oeste da Arábia onde começou o seu cultivo. Logo se percebeu que as sementes de café poderiam ser moídas e utilizadas para a produção de uma bebida estimulante, a qual começou a mostrar seu potencial comercial nessa região, pois era de alto consumo entre monges em seus rituais religiosos, visto que os mantinham acordados. Com essa popularização, em 1475 foi criada a primeira cafeteria, a Kiva Han, na Turquia, a qual ajudou ainda mais a difusão da bebida, chegando em Veneza no ano de 1615 onde começou a se expandir pela Europa [2].

Já no século XIX, o aumento do consumo do café fez crescer a necessidade de técnicas mais rápidas para a produção desta, sendo assim começaram os estudos para produzir café em máquinas a vapor. Em 1884 o italiano Angelo Moriondo criou um dos primeiros protótipos, mas foi Luigi Bezerra em 1901 que aperfeiçoou o dispositivo de produção de café introduzindo um porta-filtro, possibilitando o preparo do café em segundos diretamente em uma xícara café.

Com o avanço da tecnologia e a popularização do consumo de café, os métodos de preparos da bebida foram evoluindo.

Até que em 1970, no departamento de Pesquisa e Desenvolvimento da Nestlé, foi desenvolvido uma máquina de café expresso que utilizava cápsulas do produto moído, para o preparo da bebida. Esta tecnologia foi sofrendo modificações e ganhando espaço entre os consumidores.

Infelizmente, as cápsulas de café, ao mesmo tempo que trouxeram praticidade e rapidez no preparo da bebida, também acarretaram em um grande problema ambiental, segundo afirma Silva et al. [3]. No Brasil, por exemplo, a previsão de descarte desse material será de aproximadamente 14 mil toneladas em 2021.

Uma forma de mitigar o problema do grande volume de descarte de cápsulas é a reciclagem, de acordo com Bolton [4], a coleta de cápsulas vem sendo uma tarefa que exige cuidados e é frequentemente realizada por pessoas treinadas para correta separação entre as cápsulas e outros materiais similares descartados, portanto, o desenvolvimento de dispositivo automático que permita a coleta correta de cápsulas de café seria de grande valia especialmente para a disseminação de pontos de coleta em variados locais inclusive agregando um sistema e bonificação. A intensão é que a população deposite as cápsulas de café, nas máquinas coletoras, as quais contabilizariam as cápsulas por meio de um sistema de visão artificial e, a título de incentivo, o usuário receberia algum bônus proporcional à quantidade de cápsulas de café depositadas. Logo, o sistema deve garantir a contagem apenas das cápsulas de café, evitando fraudes. Sendo assim, um requisito importante desse sistema é a capacidade de rapidamente realizar a contagem correta das cápsulas depositadas.

Silveira [5] relata iniciativas do setor da indústria de cápsulas de café visando implementar postos de entrega voluntária nos estados de São Paulo e do Ceará, tendo conseguido 7% de aumento no recolhimento de cápsulas para a reciclagem. O tipo de dispositivo utilizado para a coleta não era automatizado, demonstrando haver espaço para desenvolvimento tecnológico nesse campo.

Ao longo do tempo é comum que produtos comerciais sofram modificações nos seus rótulos de identificação. Para um sistema capaz de identificar determinada marca de cápsula de

café, é conveniente que se escolha um modelo de classificador que seja robusto em relação a eventuais mudanças no rótulo do produto.

A escolha do modelo Rede Neural Convulucional (RNC) é adequada para tal finalidade, visto que quaisquer variações de rótulos podem ser assimiladas pelo “aprendizado” da rede, por meio de um novo treinamento contendo exemplos de visualização deste novo rótulo. Inclusive, o sistema classificador pode até permitir a coexistência de dois ou mais rótulos distintos enquanto restarem produtos com rótulos antigos nos estoques dos estabelecimentos de venda.

Este trabalho apresenta estudos experimentais de um sistema baseado em uma RNC e técnicas de processamento de imagens na coleta e contagem de cápsulas de café em um sistema que poderá ser adaptado para lidar com bonificação.

II. REDE NEURAL ARTIFICIAL - RNA

Um modelo de Inteligência Artificial muito difundido é a Rede Neural Artificial (RNA), cuja unidade básica é o neurônio artificial, o qual é descrito brevemente a seguir. Maiores detalhes podem ser encontrados nas referências citadas, particularmente nos textos de Geron [6] e Buduma [7].

O modelo de neurônio artificial foi concebido nos anos 1940 por McCulloch e Pitts, correspondendo basicamente à representação gráfica exibida na Fig. 1. Um neurônio recebe estímulos de entrada (via dendritos), que determinam um estado de ativação e envia estímulos de saída (via terminais do axônio), esses estímulos dependem do estado assumido pelo núcleo do neurônio.

A RNA emula a rede neural real utilizando uma rede de neurônios artificiais em multicamadas (Fig. 2). O aprendizado da RNA decorre da capacidade de cada neurônio capturar parte de um conhecimento complexo.

A Eq. $y = X.W + b_k$ descreve uma função de ativação (ϕ) de entradas (w_i, x_i), onde $X = \{x_1, x_2, \dots, x_n\}$ é o vetor de estímulos de entrada, $W = \{w_1, w_2, \dots, w_n\}$ o vetor de pesos representando a respectiva ponderação de cada estímulo de entrada e b_k é um vetor de viés (*bias*) utilizado para ajustes dos estímulos de entrada.

O ajuste (*tuning*) dos chamados hiper parâmetros (número de camadas e neurônios, tipo de função de ativação, etc.) é fator importante no processo de aprendizado da MLP. Maiores detalhes podem ser vistos nos trabalhos de Buduma e Locascio [7] e Géron [6].

A RNA necessita de treinamento supervisionado, em um processo que busca ajustar o modelo ao seu conjunto de dados de treinamento. É preciso que o modelo seja capaz de generalizar, apresentado uma alta acurácia tanto na classificação de dados de treinamento, quanto na classificação de dados de validação e de teste. De fato, o treinamento pode gerar certas “anomalias de aprendizado” conhecidas como super ajuste (*overfitting*) ou subajuste (*underfitting*). Maiores detalhes podem ser encontrados em Géron [6] e Buduma [7].

Assim como as RNAs, as Redes Neurais Convolucionais (RNCs) contam com neurônios, pesos e funções não lineares. Porém, ao invés de ter como entrada vetores de características

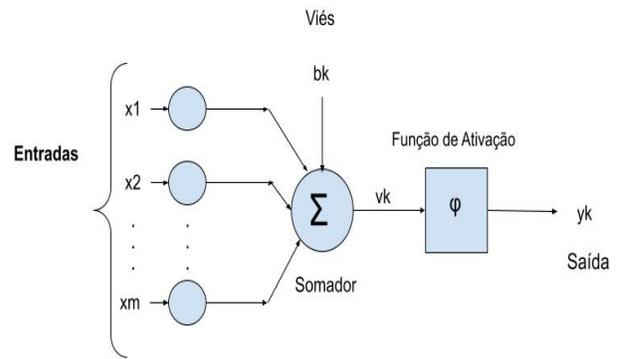


Figura 1. Representação de um neurônio artificial. Fonte: [8]

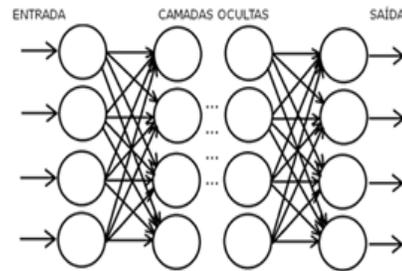


Figura 2. Um exemplo de RNA multicamadas de neurônios.

previamente extraídos das imagens, as próprias RNCs extraem as feições importantes presentes nas imagens por meio de camadas de filtros convolucionais extratores de características importantes para a classificação da imagem de entrada.

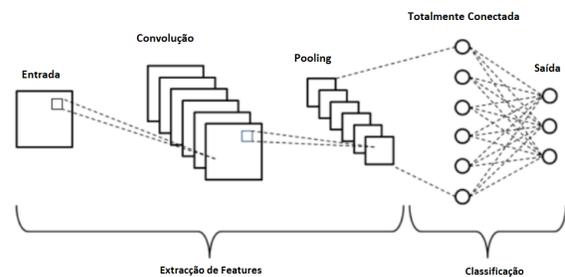


Figura 3. Exemplo de arquitetura de uma RNC. Fonte: [9].

Um exemplo de arquitetura RNC pode ser visto na Fig. 3 e basicamente conta com as camadas convolucional+retificação, *pooling*, softmax e totalmente conectadas:

- Camada Convolucional: implementa a aplicação de convoluções com filtros detectores de características de baixo nível (bordas, contornos etc);
- Camada de Retificação: advém do inglês *Rectified Linear Unit*, é utilizada após a camada de convolução com uma função de ativação não linear com o objetivo de zerar valores negativos (Eq. 1);

- Camada de *Pooling* tem o objetivo de reduzir o tamanho espacial de uma imagem, largura e altura. Sendo assim, permitindo diminuir o custo computacional de uma rede, além de ajudar a controlar o *overfitting*;
- Camada Softmax: tem como objetivo obter a maior probabilidade de uma imagem de entrada pertencer a uma determinada classe;
- Camada Totalmente Conectada: basicamente é uma RNA acoplada à saída da camada de *pooling* para executar a classificação propriamente dita.

$$f(x) = \begin{cases} x & \text{if } n \geq 0 \\ 0 & \text{if } n < 0 \end{cases} \quad (1)$$

III. MÉTRICAS DE AVALIAÇÃO

Para a avaliação dos resultados nos experimentos são utilizadas as seguintes métricas clássicas: matriz de confusão, *Precision* e *Recall* e acurácia, conforme Eq.2, Eq.3 e Eq.4. Nestas equações, TP, FP e FN correspondem aos totais de casos classificados como verdadeiro-positivos (*True Positives*), falso-positivos (*False Positives*) e falso-negativos (*False Negatives*). Os valores dessas métricas variam entre zero e a unidade, sendo esta última resultado ideal obtido de um classificador sem erros.

A matriz de confusão resulta dos testes finais do modelo utilizando o conjunto-teste (conjunto de casos não utilizados no treinamento da rede). Quanto a *Precision* (Eq. 2 e *Recall* (Eq. 3), essas métricas são representações concisas do conteúdo da matriz de confusão.

Precision indica quanto das predições positivas para certa classe foram corretamente detectadas pelo classificador (baseia-se nas colunas da matriz de confusão). Em complemento, *Recall*, também conhecida como TPR (*True Positive Rate*), informa quanto dos casos verdadeiramente positivos para certa classe foram corretamente classificados/detectados pelo classificador (baseia-se nas linhas da matriz de confusão).

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$Acurácia = \frac{TP + TN}{TP + FN + TN + FP} \quad (4)$$

IV. EXPERIMENTOS

A seguir são descritas as etapas do desenvolvimento prático do trabalho:

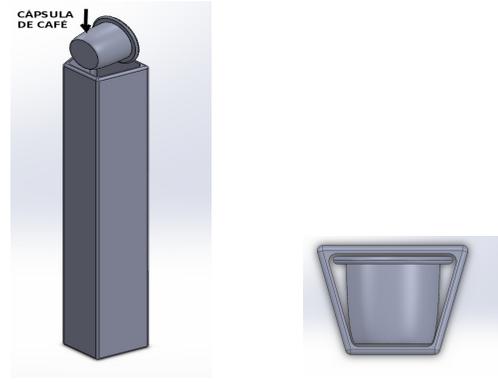


Figura 4. Esquerda: vista oblíqua do bocal de coleta exibindo uma cápsula sendo depositada na entrada do bocal coletor. Direita: vista superior do bocal de coleta também exibindo a cápsula.

A. Aquisição de Imagens

A aquisição de imagens para a formação das bases de dados foi realizada por meio de aparato específico provendo iluminação controlada e homogênea sobre o objeto em um ambiente confinado.

O dispositivo de iluminação foi um dispositivo *Ring Light* (iluminação circular), o qual proporciona uma luz uniforme, difusa, no alvo e ao longo de suas bordas, como também destaca características das superfícies. Esse tipo de iluminação atende às especificações usuais para um sistema de captura de imagens [10].

As cápsulas de café utilizadas são reais e da marca comercial Nescafé^(TM) *Dolce Gusto* Espresso Intenso. Visto que o rótulo da cápsula de café contém informações únicas e distintas, foi esquematizado um bocal de coleta com o objetivo garantir que o rótulo fique voltado para o campo de visão da câmera. Esse sistema pode ser visto na Fig. 4.

As cápsulas depositadas no bocal coletor caem pelo corpo do bocal e são exibidas com o rótulo voltado para a câmera. Sem a garantia de orientação ótima, O rótulo pode ser capturado de “ponta-cabeça” (Fig. 5), por exemplo, ou pode apresentar reflexo luminoso (a reflexão no rótulo pode ocorrer devido à posição relativa do plano do rótulo da cápsula, em queda pelo bocal, o que pode refletir a iluminação em uma direção coincidente à direção do eixo óptico da câmera).

O aparato utilizado para a captura de imagens pode ser visto na Fig. 6 com os seguintes itens:

- A) Câmera utilizada para realizar a captura dos rótulos das capsulas de café: fabricante Cognex^(TM) - CAM-CIC-5000-24-GC;
- B) Iluminação em forma de anel, a fim de manter uma luz uniforme em volta do objeto de interesse;
- C) Sensor para acionar a câmera no instante que a cápsula estiver no campo de visão. Trata-se de um sensor óptico do fabricante Balluff^(TM), modelo BOS023M, que apresenta conexão direta (cabo) com a câmera. Esse tipo de sensor detecta diferenças de contraste, reconhecendo objetos dependendo da superfície, cor e material. Caracteriza-se pela instalação, montagem e alinhamento simples;

D) Bocal para introduzir a cápsula de café e guiá-la até o ponto correto da captura de imagem, com o rótulo voltado para o campo de visão da câmera.



Figura 5. Exemplos de imagens de cápsulas capturadas.

- 200 imagens de uma palheta de música.
- 216 imagens de um pendrive.
- 200 imagens de um pendrive preto.
- 204 imagens de um cartão sd.

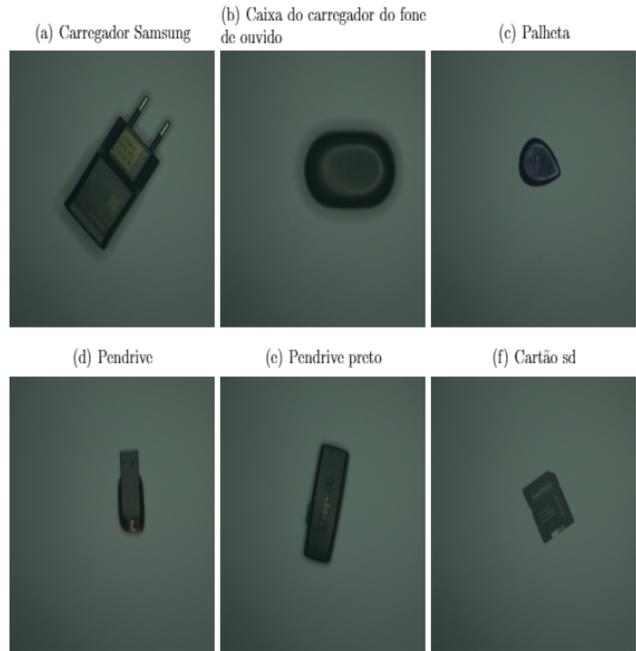


Figura 7. Exemplo de não cápsulas de café

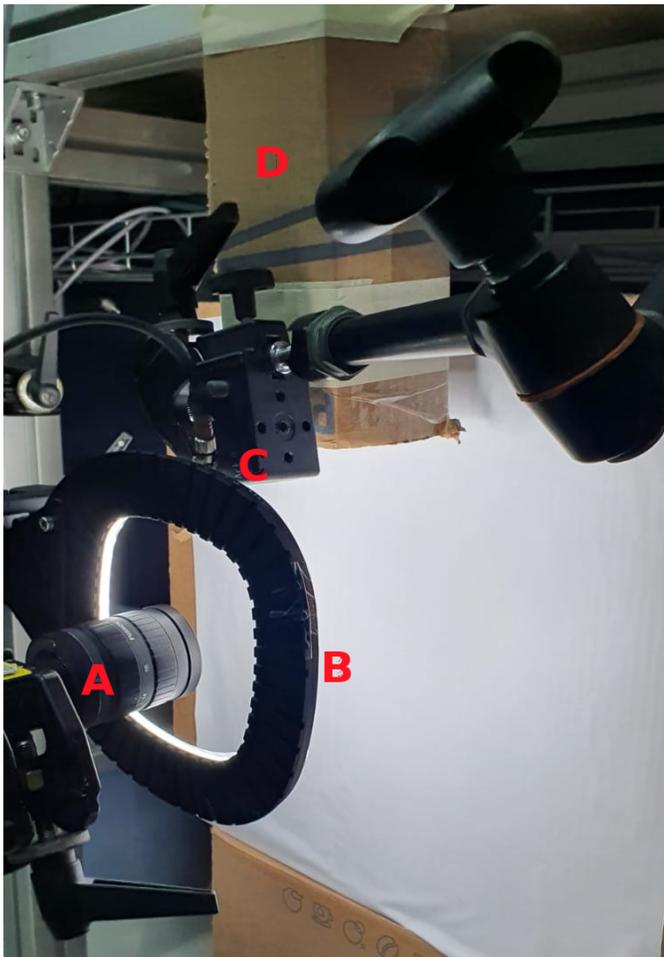


Figura 6. Protótipo do ambiente de captura: Vistas do ambiente de captura de imagem: A) Câmera, B) Iluminação em anel (*ring light*), C) Gatilho de captura (*trigger*), D) Bocal coletor.

Ao total, foram capturadas 613 imagens de cápsulas de café e 1252 imagens de objetos que não são cápsulas de café, como os da Figura 7 com a seguinte distribuição:

- 200 imagens de um carregador de celular marca-1.
- 132 imagens da caixa do carregador de um fone marca-2.

B. Pré Processamento

A preparação das imagens passa por uma operação de subtração de fundo visando eliminar qualquer ruído remanescente no fundo da imagem. Para tanto foi utilizada uma imagem sem o objeto de interesse. Ilustrando essa etapa do pré processamento realizado, temos a Fig. 8 representando a imagem original, apenas o fundo da foto, e o resultado da operação.

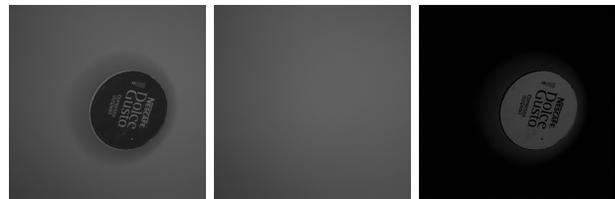


Figura 8. Cápsula capturada, fundo da imagem, cápsula processada para remoção do fundo.

Como um pré processamento para a rede neural artificial também foi realizado a transformação da imagem RGB para a escala de cinza, e feito um redimensionamento de tamanho dela para 180x180 pixels para melhor eficiência computacional pois as imagens originais apresentavam resolução de 1800 x 1800 pixels.

V. REDE NEURAL UTILIZADA

A RNC utilizada para os experimentos é baseada no trabalho de Huang et al.[11] para a classificação de grãos de café. No melhor dos esforços, não foi localizado nenhum trabalho especificamente voltado à detecção de cápsulas de café. Por outro lado, o problema da classificação de grãos de café envolve um alto nível de complexidade, a separação de grãos identificando produto de alta qualidade de produto de qualidade menor, onde, detalhes de relativos ao tamanho, forma e cor do grão são sutis diferenças que precisam ser aprendidos pela rede. Nesse contexto Huang et al. [11] atingiram uma acurácia expressiva da ordem de 93,34%, esses aspectos (alta complexidade do problema versus alta acurácia na classificação) levaram à escolha do modelo RNC usado por Huang et al. [11] para enfrentamento da problemática tratada neste artigo. A seguir é descrita a arquitetura da RNA de Huang et al. [11]:

- 1) Camada de convolução com entrada de 180x180, com *stride* de 1 e função de ativação ReLu, seguida de uma camada de *pooling* com *stride* de 4.
- 2) Camada de convolução, com *stride* de 1 e função de ativação ReLu, seguida de uma camada de *pooling* com *stride* de 3.
- 3) Camada totalmente conectada, também com a função de ativação ReLu
- 4) Camada totalmente conectada com a função de ativação Softmax

Esse algoritmo foi construído na linguagem Python, versão 3.6, com o Tensorflow na versão 2.3 e biblioteca CuDa da Nvidia na versão 10.1. Em relação ao computador, foi utilizada a CPU Ryzen 5600 e uma placa de vídeo RTX 2060 Super. O modelo de câmera utilizada para a captura foi o CAM-CIC-5000-20-GC da Cognex^(TM).

Para treinar a rede neural convolucional foram realizados dois testes com bases de dados diferentes: $dataset_1$ e $dataset_2$. O $dataset_1$ contém imagens de cápsulas de café sem reflexo de iluminação no rótulo, ao passo que o $dataset_2$ corresponde ao $dataset_1$ acrescido de imagens de cápsulas com reflexo de luz em seus rótulos, vide Figura 9. Os dois testes contam com imagens de objetos que não são cápsulas.

No primeiro teste foram usadas 415 imagens para cada classe e no segundo teste foram usadas 613 imagens. Essa diferença de quantidades entre os experimentos ocorreu devido ao fato do primeiro teste não conter imagens com reflexos.

Durante os experimentos foram utilizadas as seguintes configurações para a rede:

- *Learning Rate*: 0,0001
- *Bach size*: 16
- Função *Loss*: *Binary Crossentropy*
- Otimizador: Adam
- Base de dados dividida manualmente entre 80% para treinamento e 20% para testes



Figura 9. Esquerda: imagem sem reflexo. Direita: imagem com reflexo no rótulo.

VI. RESULTADOS E ANÁLISES

O primeiro teste da rede neural artificial, ou seja, o que não contém cápsulas de café com reflexo no banco de imagens, apresentou a matriz de confusão representada na Figura 10.

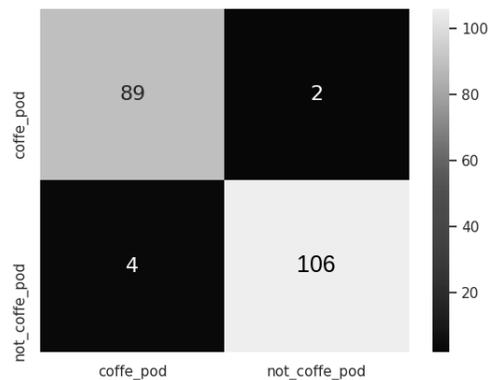


Figura 10. Matriz de confusão do teste 1

Com esses dados podemos calcular o *Recall*, *Precision* e a *Acurácia*, visto que temos a quantidade de 89 verdadeiros positivos, 106 verdadeiros negativos, 2 falsos positivos e 4 falsos negativos.

- *Recall* de 0,956, o que indica que várias imagens foram classificadas corretamente.
- *Precision* da rede neural de 0,978, indicando poucos falsos positivos.
- *Acurácia* total do sistema de 97,01%.

Já para o segundo teste da rede neural artificial, incluindo imagens contendo reflexo, resultou na seguinte matriz de confusão exibida na Fig11, com 118 verdadeiros positivos, 102 verdadeiros negativos, 1 falso positivo e 5 falsos negativos.

Sendo assim temos as seguintes métricas para esse teste:

- *Recall* de 0,959, indicando uma probabilidade alta de detectar um objeto.
- *Precision* da rede neural artificial é de 0,991, ou seja, 99,1%, indicando uma quantidade baixa de falsos positivos.
- *Acurácia* total do sistema de 97,35%

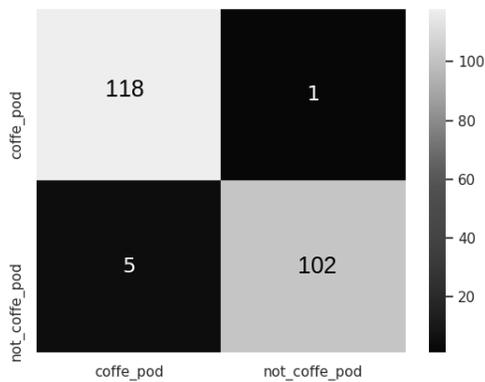


Figura 11. Matriz de confusão do teste 2

Nota-se que em ambos experimentos a acurácia alcançada é excelente, da ordem de 97% de acerto. A ocorrência de reflexo no rótulo não deteriorou o desempenho da rede, o que indica o seu potencial para aplicações reais, nas quais a posição relativa entre o rótulo e a fonte de iluminação podem gerar reflexões e distorções nas imagens capturadas. Nesse sentido, um sistema de coleta automática poderá dispensar tratamento de imagens para a remoção dos efeitos do reflexo, acarretando um menor tempo de resposta. Característica importante para a aplicação desejada, na qual o tempo de espera deve ser minimizado para que sem demora (tempo real) o usuário (dispensador das cápsulas) tenha conhecimento do nível de bonificação a ser contemplado.

VII. CONCLUSÃO

Motivado pela problemática do acúmulo de resíduos provenientes do uso de cápsulas de café e a dificuldade de sua reciclagem, o presente trabalho descreve resultados preliminares de experimentos de identificação de cápsulas de café por meio do seu rótulo visando contribuir com o desenvolvimento de um sistema de coleta e bonificação desse tipo de material.

Uma RNC foi treinada com duas bases de dados contendo imagens de cápsulas de café da marca Nescafé (TM), mais especificamente do sabor *Dolce Gusto* Espresso Intenso, com e sem reflexo de iluminação. Os testes produziram resultados bastante satisfatórios com uma acurácia geral de aproximadamente 97%. Entretanto, não foram realizados testes quanto a outras marcas, nem levando em conta possíveis tentativas de fraude mais sofisticadas (visto que o sistema ideal prevê bonificação pelo descarte correto do produto utilizado).

Dentre os trabalhos futuros, sugere-se realizar experimentos com outras marcas de cápsulas de café, buscando RNC que possam rodar em *hardware* mais compacto e de custo mais baixo. A justificativa para tanto reside na perspectiva de embarque do sistema computacional em um aparato de coleta de cápsulas a ser produzido em larga escala, sendo importante a dimensão física desse aparato bem como questões relativas à dissipação de calor e custo final desse conjunto (aparato

coletor e sistema computacional). Em resumo, o desafio para trabalhos futuros consiste em se buscar um modelo de classificador ótimo rodando em um dispositivo computacional de baixo custo, alto desempenho computacional, baixa emissão de calor e pequenas dimensões físicas.

AGRADECIMENTOS

Os autores gostariam de agradecer ao apoio da Universidade do Estado de Santa Catarina (UDESC), particularmente ao fomento do EDITAL N° 01/2021 PIPES 2021-2022.

REFERÊNCIAS

- [1] J. Morris, *Coffe a Global History*. Reaktion Books, 2019.
- [2] S. Talita, “História do café - a origem e a trajetória da bebida no mundo.” 2016. [Online]. Available: <https://www.graogourmet.com/blog/historia-do-cafe/>
- [3] E. C. e Silva et al., “Governança privada e sustentabilidade na indústria do café,” *Revista agroalimentaria*, vol. 25, no. 48, pp. 35–51, 2015.
- [4] D. Bolton, “Precision manufacturing is essential to capsule success,” 2015.
- [5] K. P. Silveira, “Iniciativas sustentáveis: Nespresso - transparência pela sustentabilidade.” 2018. [Online]. Available: <https://www.fiesp.com.br/indices-pesquisas-e-publicacoes/case-nespresso/>
- [6] A. Geron, *Hands-On Machine Learning with Scikit-Learn and TensorFlow, Concepts, Tools, and Techniques to Build Intelligent Systems*. O’Reilly, 2017.
- [7] N. Buduma and N. Locascio, *Fundamentals of deep learning: designing next-generation machine intelligence algorithms*. O’Reilly, 2017.
- [8] S. Haykin, *Neural networks and learning machines*. McMaster University, Canada: Pearson, 2009.
- [9] V. Phung and E. Rhee, “A deep learning approach for classification of cloud image patches on small datasets,” *Journal of Information and Communication Convergence Engineering*, vol. 16, pp. 173–178, 01 2018.
- [10] COGNEX, “Iluminação de visão industrial,” 2016. [Online]. Available: <https://www.cognex.com/pt-br/what-is/machine-vision/components/lighting>
- [11] N.-F. Huang, D.-L. Chou, and C.-A. Lee, “Real-time classification of green coffee beans by using a convolutional neural network,” in *2019 3rd International Conference on Imaging, Signal Processing and Communication (ICISPC)*, 2019, pp. 107–111.