

Classifying Glycemic Index from Food Images with Convolutional Neural Networks

Jailson P. Januário, Elloá B. Guedes, Fábio S. de Silva

¹Grupo de Pesquisas em Sistemas Inteligentes
Laboratório de Sistemas Inteligentes
Escola Superior de Tecnologia
Universidade do Estado do Amazonas
Av. Darcy Vargas, 1200 – Manaus – Amazonas

{jppj.ads, ebgcosta, fssilva}@uea.edu.br

Abstract. *The identification of food glycemic index is fundamental for dietary planning and glycemic control, especially in patients with diabetes. In order to determine the glycemic index from food images, this paper considers the use of a convolutional neural network architecture to address two classification tasks, one with only fruits and the other with foods in general. Upon evaluating the proposed model, we verified an accuracy of 98.22% for the first task and of 84.30% for the second task, results that are competitive with the state of the art and that motivate the development of computational solutions that promote the improvement of quality of life through healthy eating.*

Resumo. *A identificação do índice glicêmico de alimentos é fundamental para o planejamento dietético e controle glicêmico, especialmente de pacientes com diabetes. Tendo em vista a determinação do índice glicêmico a partir de imagens de alimentos, este trabalho considera o uso de uma arquitetura de rede neural convolucional para abordar duas tarefas de classificação, sendo uma de frutas e outra de alimentos em geral. Ao avaliar o modelo proposto, obteve-se acurácia de 98,22% para a primeira tarefa e de 84,30% para a segunda, resultados que se mostram competitivos com o estado da arte e que motivam o desenvolvimento de soluções computacionais que promovem a melhoria da qualidade de vida por meio da alimentação saudável.*

1. Introdução

O índice glicêmico é um método de classificação de alimentos baseado na quantidade de carboidratos nele disponíveis e no potencial destes carboidratos em promoverem um aumento na glicose sanguínea [Marsh et al. 2011]. Promover uma dieta rica em alimentos de baixo índice glicêmico promove melhorias na manutenção dos índices de glicose sanguínea, na redução da lipídemia, na diminuição da demanda por insulina e na mitigação de fatores diretamente ligados ao desenvolvimento de diabetes mellitus [Silva and de Mello 2006].

A incorporação do índice glicêmico no planejamento dietético de pacientes com diabetes contribui para a melhora do controle glicêmico [Marsh et al. 2011], o que também é corroborado por outros autores da literatura [Greenwood et al. 2013]. A sugestão de dietas neste perfil compõe diretrizes nutricionais disseminadas por diversas associações em todo o mundo para prevenção e controle de diabetes [SBD 2017, Imran et al. 2018]. Os valores de referência do índice glicêmico dos alimentos são comumente compilados

em tabelas como forma de instrumento de orientação nutricional ao público em geral [Philippi et al. 1999]. Porém, em tabelas amplamente difundidas, o número de alimentos contemplados supera os 2.400 itens, vide [Atkinson et al. 2008], o que pode tornar a consulta a tais valores uma tarefa exaustiva para o público em geral em contextos cotidianos.

Para contornar essa dificuldade prática são necessários esforços na concepção de soluções que colaborem com a rápida e fácil identificação do índice glicêmico associado a um alimento. Com o advento da Visão Computacional, e o recente progresso promovido nesta área por meio das técnicas de *Deep Learning*, este trabalho tem por objetivo apresentar uma abordagem baseada no uso de Redes Neurais Convolucionais (CNNs, do inglês, *Convolutional Neural Networks*) para esta tarefa, na qual a descoberta do índice glicêmico é apoiada pela classificação de alimentos a partir de imagens. Os resultados obtidos apresentados neste trabalho mostram-se promissores, pois obteve-se uma acurácia de 84,30% nesta tarefa, equiparável com as mais recentes contribuições da literatura, o que pode ensejar o desenvolvimento de soluções para orientação nutricional, monitoramento e controle alimentar.

Para apresentar o que se propõe, este trabalho está organizado como segue, a Seção 2 apresenta uma visão geral do estado da arte para a classificação de imagens de alimentos com técnicas de *Deep Learning*. Em seguida, na Seção 3, são apresentados os materiais e métodos considerados na elaboração da solução proposta. Os resultados obtidos e a discussão são apresentados na Seção 4. Por fim, as considerações finais e perspectivas de trabalhos futuros são apresentados na Seção 5.

2. Trabalhos Relacionados

A identificação de alimentos a partir de imagens já foi abordada em outros trabalhos da literatura com vistas à classificação. Acompanhando o progresso recente das técnicas de *Deep Learning* para classificação de imagens, nota-se que muitos dos resultados recentes para classificação de alimentos também acompanham essa tendência, em especial, considerando o uso de Redes Neurais Convolucionais. Alguns destes trabalhos do estado da arte são apresentados a seguir.

Considerando o reconhecimento de alimentos e bebidas com vistas a fornecer orientações nutricionais para pacientes acometidos pela doença de Parkinson, Mezgec & Seljak [2017] elaboraram uma arquitetura própria de CNN para esta tarefa, a qual foi preparada para classificar 520 tipos de alimentos e bebidas e treinada com imagens oriundas da internet. Os resultados obtidos denotaram uma acurácia total de 86,72% do modelo proposto, com 55% de acerto nas 5 classes mais frequentes. Segundo os autores, o modelo obtido é promissor e vem sendo usado na prática como parte de uma aplicação móvel para o público-alvo considerado.

Muresan & Oltean [2018], por sua vez, restringiram-se ao problema de classificação de frutas, considerando 60 classes distintas e propondo uma base de dados para esta tarefa, a qual passou a ser disponibilizada de maneira pública e gratuita. Para tal, os autores apresentaram um modelo de CNN desenvolvido especificamente para o projeto, adotando algumas abordagens com uso de imagens nos modos grayscale, RGB e HSV. O resultado com o melhor desempenho na tarefa apresentou uma acurácia total de 97,04% nos dados reservados para testes, utilizando aumento de dados e combinando imagens no espaço de cores HSV com representações em grayscale. Porém o trabalho se restringe apenas a um tipo de alimento, onde o conjunto de dados utilizado na tarefa foi subme-

tido a um pré-processamento que preencheu o fundo das imagens com uma cor branca de modo que a fruta a ser detectada estivesse evidenciada.

Os métodos de reconhecimento de imagens voltados para classificação de alimentos mostram-se uma área promissora, quando o foco é estimar valores nutricionais e monitorar os hábitos alimentares das pessoas para os cuidados em uma alimentação saudável. Freitas et al [2018] apresentaram uma aplicação móvel para o auxílio no monitoramento e controle alimentar utilizando métodos de reconhecimento de imagens. Para tal, os autores constituíram um conjunto de dados com mais 1.000 imagens e 5 classes de alimentos, a fim de definir um modelo de aprendizagem para classificação de alimentos os autores utilizaram em sua abordagem modelos canônicos de CNNs. O resultado com o melhor desempenho obtido demonstrou uma precisão de classificação 86,94% no conjunto reservado para validação, que, para os autores, mostra-se promissora diante do estado da arte.

A solução proposta neste escopo se difere dos trabalhos previamente mencionados, pois visa a proposição e uso de uma arquitetura própria de rede neural convolucional com vistas promover a associação da classificação da imagem do alimento com o respectivo índice glicêmico em uma tabela de referência [Atkinson et al. 2008]. A tarefa de Aprendizado de Máquina considerada e a metodologia adotada para obtenção dos resultados propostos encontram-se descritas na seção a seguir.

3. Materiais e Métodos

Para endereçar o que se propõe e considerando o amplo escopo da classificação de alimentos, considerou-se abordar duas tarefas de classificação mediante Aprendizado Supervisionado, a primeira delas com vistas a abranger apenas frutas, e a segunda, mais geral, compreendendo também outros tipos de alimentos. Em ambos os casos, todas as classes consideradas possuem índice glicêmico estabelecido. Assim, após a classificação, tem-se então o mapeamento para o índice correspondente. Diante do exposto, cada uma das tarefas a ser abordada encontra-se descrita detalhadamente a seguir.

3.1. Tarefa de Classificação de Frutas

Para a classificação de frutas foi considerada uma tarefa de classificação multi-rótulo com dados experimentais oriundos da base de dados *Fruits-360 dataset* [Muresan and Oltean 2018]. Esta base é composta por 65.429 exemplos de frutas organizadas segundo 95 classes com imagens no modo RGB com dimensões 100×100 pixels, conforme exemplos ilustrados na Figura 1. Observa-se que os exemplos de tal base são dispostos de maneira controlada, encontrando-se segmentados do *background* e centralizados.

Os exemplos disponíveis na base Fruits-360 encontram-se distribuídos de maneira heterogênea, conforme ilustra o histograma de exemplos por classe da Figura 2. As classes minoritárias, a citar *Peach* e *Lemon Meyer*, por exemplo, representam cerca de 0,61% do total de amostras disponíveis cada, o que ressalta a natureza desbalanceada desta base de dados.

Para conduzir a tarefa considerada, a base de dados foi particionada segundo o método *holdout* de validação cruzada, em que 70% dos dados foram utilizados para treinamento do modelo e o restante para testes. Após o treinamento, o modelo foi aferido



Figura 1: Amostras de imagens do Fruits-360 dataset.

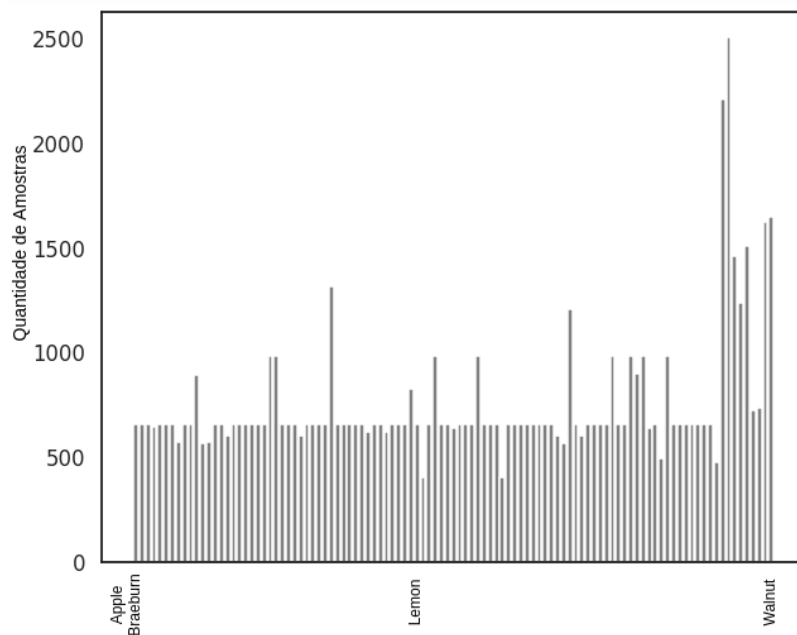


Figura 2: Histograma de distribuição de exemplos por classe na base de dados Fruits-360, considerando a ordem lexicográfica das classes.

perante o desempenho no conjunto de testes segundo as métricas de acurácia e *Micro F-Score*, em que esta última foi preferida por contemplar a média harmônica entre a precisão e a revocação por classe, considerando as diferentes frequências nas classes do problema.

Visando prover um melhor ajuste de pesos pelo otimizador na etapa de treinamento do modelo, os valores dos pixels das imagens foram normalizados para o intervalo $[0,1]$ [Chollet 2018]. Além disso, visando explorar possíveis comportamentos distintos do modelo mediante o fornecimento dos exemplos em diferentes representações, conforme metodologia adotada por Muresan & Oltean [2018], foram considerados os espaços de cores *Grayscale*, *RGB* e *HSV*. Ademais, optou-se por não utilizar técnicas para promover o aumento artificial de dados.

3.2. Tarefa de Classificação de Alimentos

Visando aumentar o escopo da tarefa de classificação anteriormente proposta, foi feito um acréscimo à base de dados dos exemplos presentes na base Food-11 [MMSPG 2017]. Esta última é composta de 16.643 imagens de alimentos agrupadas segundo 11 classes, conforme ilustra a Figura 3. Percebe-se que os exemplos foram adquiridos em

cenários mais realísticos, sem padronização no processo de aquisição nem tampouco prévia segmentação de fundo. Nota-se ainda que o número de exemplos por classe é heterogêneo, o que também sugere desbalanceamento na base mencionada.

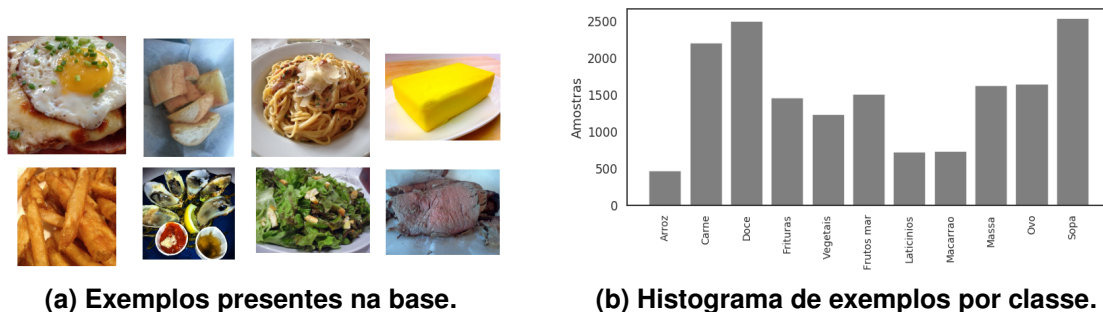


Figura 3: Exemplos e distribuição de amostras por classe na base de dados Food-11.

Os exemplos disponíveis na Food-11 foram então redimensionados para 100×100 pixels e, após a unificação, obteve-se um total de 82.072 exemplos distribuídos conforme 105 classes a serem utilizados em uma nova tarefa de classificação multi-rótulo. Os mesmos percentuais de partição *holdout* foram adotados, bem como de normalização e representação do espaço de cores. Nesta segunda tarefa, as métricas de desempenho acurácia e Micro *F-Score* também foram utilizadas como referência para avaliação do modelo.

3.3. Proposição de Modelo: Arquitetura, Parâmetros e Hiperparâmetros

Encontrar uma rede neural convolucional adequada para uma determinada tarefa de Visão Computacional não é uma tarefa fácil. Para contornar esta dificuldade, é comum utilizar-se de arquiteturas canônicas com bom desempenho em tarefas consagradas, quer seja preservando sua estrutura mas efetuando um novo treinamento, quer seja transferindo o aprendizado para o novo domínio [Khan et al. 2018]. Uma outra alternativa possível diz respeito à proposição de novas arquiteturas a partir daquelas já existentes, por meio da adição ou remoção de camadas.

A arquitetura de rede utilizada neste projeto é uma rede neural convolucional inspirada no modelo canônico VGG-16, que foi desenvolvida com o propósito de classificar e reconhecer objetos em imagens [Simonyan and Zisserman 2015]. Assim, foram feitas adaptações propostas nesta arquitetura que incluem a adição das seguintes camadas: *Batch Normalization*, com vistas a reduzir a mudança de covariância nas camadas internas da rede e acelerar o treinamento [Buduma 2017]; *Global Average Pooling*, com vistas a sumarizar massivamente as características relevantes da entrada para a camada seguinte [Brownlee 2016]; e *Dropout*, cujo o emprego visa reduzir a ocorrência de *overfitting*. Além destas adições, o número de neurônios nas camadas convolucionais foi modificado, com vistas a produzir mais mapas de características. Ao final, o número resultante de parâmetros mostrou-se inferior à arquitetura original, o que pode ser especialmente útil nos quesitos de armazenamento do modelo após o treinamento. A Figura 4 apresenta a visão geral da arquitetura proposta e os parâmetros utilizados. Ressalta-se que o número de neurônios na camada de saída será estimado levando em consideração o quantitativo de classes da tarefa para o qual será preparado.

O modelo proposto apresenta blocos de convolução com números crescentes de neurônios, onde os números entre parênteses indicam o número de filtros de saída da camada *Conv2D*, como pode ser observado na Figura 4. Nas ativações internas, seguiu-se o padrão utilizado na VGGNet, cuja vantagem é a não ativação de todos os neurônios de cada camada. Além disso, a camada de saída do modelo proposto é constituída por uma quantidade de neurônios compatível com cada tarefa, sendo sujeitos à função de ativação *softmax*.

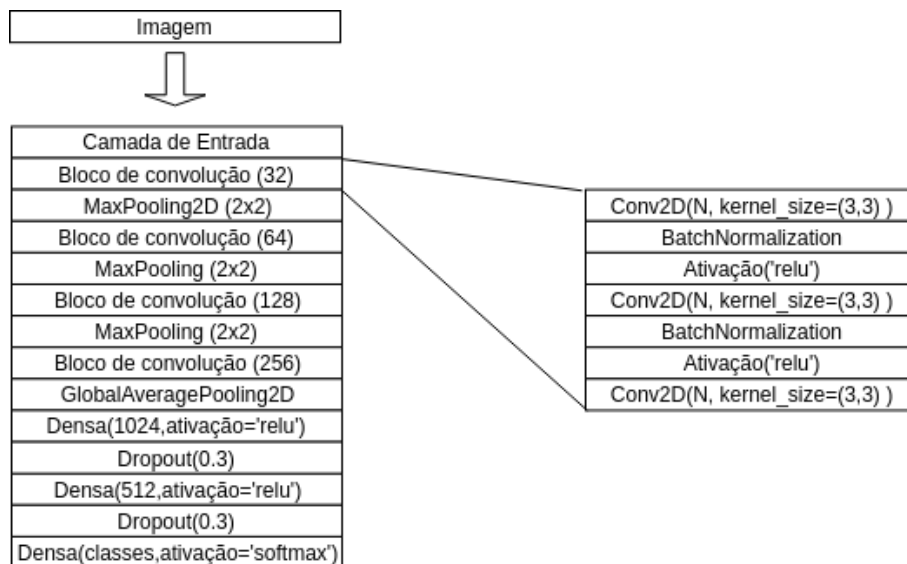


Figura 4: Modelo de rede neural convolucional elaborado a partir da VGG-16 para as tarefas de classificação propostas.

Para efetuar o treinamento deste modelo, os seguintes hiperparâmetros foram considerados: otimizador SGD para estimação do gradiente descendente do algoritmo de *backpropagation*; taxa de aprendizado igual a 0,01; *momentum* de 0,9; *batch* de tamanho 32; e, número máximo de épocas igual a 150. A escolha destes hiperparâmetros foi efetuada de maneira empírica. Visando evitar *overfitting*, adotou-se o treinamento com *early stopping* e *patience* igual a 30, em que a mudança na perda no conjunto de validação foi a métrica monitorada. Na ocasião da interrupção do treinamento, foram preservados os pesos na época em que a menor perda foi registrada.

Para a implementação do modelo proposto, foi adotado um conjunto de bibliotecas na plataforma Python, a citar: Numpy ¹, Keras ² e Tensorflow ³. Além disso, a execução do treinamento e teste do modelo proposto fez uso de um desktop com processador Intel Core i7 8a geração, 16 GB de memória principal, 2 TB de memória secundária e duas placas de vídeo NVIDIA GeForce 1080 com 11 GB de memória cada.

4. Resultados e Discussões

Conforme descrito na metodologia, o modelo proposto foi treinado e testado para cada uma das tarefas consideradas, tendo as métricas de desempenho sido coletadas em cada um desses cenários individualmente. A Tabela 1 sintetiza o número de parâmetros do

¹<http://www.numpy.org/>

²<https://keras.io/>

³<https://www.tensorflow.org/>

modelo perante cada cenário de cada tarefa e o número de épocas demandado para o treinamento. Nota-se que o número de parâmetros variou em função da dimensão da entrada e do número de classes da tarefa associada e que o número de épocas demandado por cada modelo também foi distinto.

Tabela 1: Síntese dos parâmetros e épocas do modelo proposto perante cada tarefa e cenário considerados

Tarefa	Espaço de Cores	Parâmetros Treináveis	Épocas
Classificação de Frutas	<i>Grayscale</i>	2.794.143	41
	RGB	2.794.719	51
	HSV	2.794.719	80
Classificação de Alimentos	<i>Grayscale</i>	2.799.273	74
	RGB	2.799.849	72
	HSV	2.799.849	71

Após a etapa de treinamento, foi conduzida a avaliação dos modelos perante os exemplos disponíveis no conjunto de testes, juntamente com a coleta das métricas de desempenho previamente especificadas, resultados estes que encontram-se dispostos na Tabela 2. Os resultados obtidos do desempenho do modelo na tarefa de classificação, em especial, mostram-se muito animadores, pois foram aferidos como tendo melhor Micro *F-Score* igual a 0,98, um alto índice de precisão e revocação para a tarefa considerada. Levando em conta que os espaços de cores com diferentes componentes resultaram em um melhor desempenho do modelo que usando apenas tons de cinza, há evidências de que a coloração contribuiu para uma melhor classificação das frutas, além de suas formas.

Tabela 2: Performance das métricas no conjunto de teste das tarefas de reconhecimento.

Tarefa	Espaço de Cores	Acurácia	Micro F-Score
Classificação de Frutas	<i>Grayscale</i>	93,70%	0,94
	RGB	96,63%	0,97
	HSV	98,22%	0,98
Classificação de Alimentos	<i>Grayscale</i>	79,28%	0,79
	RGB	84,30%	0,84
	HSV	83,99%	0,84

Levando em conta o desempenho dos modelos, nota-se que houve melhor performance na primeira tarefa, o que enseja que as características disponíveis e a preparação deste conjunto de dados foram favoráveis ao aprendizado. Na tarefa de classificação de alimentos, além de haver um número maior de classes, uma porcentagem significativa das imagens ali contidas foram coletadas em contextos realísticos, o que aumentou a complexidade da tarefa de aprendizado. De maneira geral, é possível observar também que o desempenho foi melhor quando havia componentes de cor nos exemplos.

Em relação à tarefa de classificação de alimentos, a Figura 5a sintetiza a matriz de confusão do modelo proposto treinado com imagens do espaço de cores RGB para 10 classes escolhidas aleatoriamente da base de dados utilizada. É possível observar que há

alta densidade na diagonal principal desta matriz, evidenciando a classificação correta dos respectivos exemplos. Os F-Scores associados à estas classes encontram-se dispostos na Figura 5b, em que nota-se que não há uniformidade nesta métrica em relação às diferentes classes, sugerindo que há algumas delas mais facilmente distinguíveis do que outras.

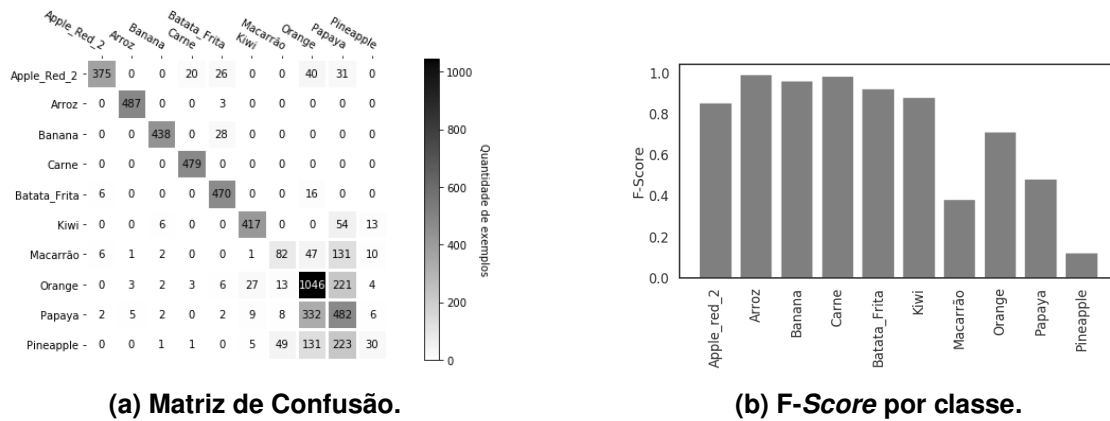


Figura 5: Desempenho do modelo proposto a tarefa de classificação de alimentos para 10 classes selecionadas aleatoriamente.

De maneira geral, o modelo adaptado a partir da arquitetura canônica VGG-16 mostrou-se efetivo para as tarefas propostas, pois, mesmo com um número de épocas relativamente baixo, houve bom desempenho, aliado ao fato de que não foi efetuado um ajuste fino de parâmetros, o que viria a requerer mais recursos computacionais.

Considerando uma relação dos resultados propostos com o estado da arte, tem-se a síntese apresentada na Tabela 3. Para a tarefa de classificação de frutas, o modelo proposto superou o desempenho relatado em um trabalho relacionado do estado da arte. No tocante à tarefa de classificação de alimentos, o melhor desempenho verificado é comparável, mas não supera as métricas já consolidadas. Apesar disso, ressalta-se que no trabalho aqui proposto há pouco esforço de pré-processamento, sem uso de técnica de aumento de dados, e que o número de camadas convolucionais utilizados é menor em comparação com a literatura.

Tabela 3: Comparação do desempenho da solução proposta com modelos análogos propostos na literatura.

Tarefa	Modelo	Acurácia
Classificação de Frutas	Proposto	98,22%
	[Muresan and Oltean 2018]	97,04%
Classificação de Alimentos	[Freitas et al. 2018]	86,94%
	[Mezgec and Seljak 2017]	86,72%
	Proposto	84,30%

5. Conclusões e Trabalhos Futuros

No escopo deste trabalho foram abordadas duas tarefas de Aprendizado de Máquina com vistas a estimar o índice glicêmico em alimentos a partir de imagens, o que pode vir a

corroborar o desenvolvimento de soluções para orientação nutricional, monitoramento e controle alimentar. Mediante Aprendizado Supervisionado, as tarefas consideradas foram endereçadas com uma arquitetura de rede neural convolucional obtida a partir de adaptações na arquitetura canônica VGG-16. Sem considerar o uso de técnicas de aumento artificial de dados e explorando apenas variações nos espaços de cores das imagens, o desempenho do melhor modelo para a tarefa de classificação de frutas registrou uma acurácia de 98,22%, superando o estado da arte para a mesma base de dados, enquanto que para a tarefa de classificação de alimentos foi de 84,30%, um resultado comparável ao da literatura, mas ligeiramente inferior.

Considerando a busca de melhorias nas métricas verificadas para colaborar com o problema de estimação do índice glicêmico de alimentos segundo uma tarefa de Visão Computacional, os próximos passos a serem adotados consideram o emprego de técnicas de transferência de aprendizado, a exploração de outras arquiteturas de redes neurais convolucionais, a realização de ajuste fino de parâmetros e hiperparâmetros dos modelos e a exploração de técnicas de pré-processamento das imagens. Além disto, com vistas a favorecer a coleta de novos exemplos e aproximar os resultados propostos do público-alvo, tem-se em mente o desenvolvimento de uma aplicação móvel para a predição do índice glicêmico a partir de imagens coletadas com o auxílio de *smartphones*.

Agradecimentos

Os autores agradecem o apoio financeiro e material provido pela Fundação de Amparo à Pesquisa do Estado do Amazonas (FAPEAM) e pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) por meio do Programa PPP Edital 04/2017.

Referências

- Atkinson, F. S., Foster-Powell, K., and Brand-Miller, J. C. (2008). International tables of glycemic index and glycemic load values: 2008. *Diabetes Care*, 31(12):2281–2283.
- Brownlee, J. (2016). *Deep Learning With Python*. Machine Learning Mastery, Estados Unidos.
- Buduma, N. (2017). *Fundamentals of Deep Learning*. O'Reilly Media, Estados Unidos.
- Chollet, F. (2018). *Deep Learning with Python*. Manning Shelter Island.
- Freitas, C. N. C., Cordeiro, F. R., and da Silva, A. J. (2018). Food recognition system for nutrition monitoring. In *Anais do XV Encontro Nacional de Inteligência Artificial e Computacional*, pages 186–197, Porto Alegre, RS, Brasil. SBC.
- Greenwood, D. C., Threapleton, D. E., Evans, C. E., Cleghorn, C. L., Nykjaer, C., Woodhead, C., and Burley, V. J. (2013). Glycemic index, glycemic load, carbohydrates, and type 2 diabetes. 36(12):4166–4171.
- Imran, S. A., Agarwal, G., Bajaj, H. S., and Ross, S. (2018). Targets for glycemic control. *Canadian Journal of Diabetes*, 42:S42 – S46. Diabetes Canada 2018 Clinical Practice Guidelines for the Prevention and Management of Diabetes in Canada.
- Khan, S., Rahmani, H., Shah, S. A. A., and Bennamoun, M. (2018). *A Guide to Convolutional Neural Networks for Computer Vision*. Morgan & Claypool, Austrália, 1 edition.

- Marsh, K., Barclay, A., Colagiuri, S., and Brand-Miller, J. (2011). Glycemic index and glycemic load of carbohydrates in the diabetes diet. *Current Diabetes Reports*, 11(2):120–127.
- Mezgec, S. and Seljak, B. K. (2017). Nutrinet: A deep learning food and drink image recognition system for dietary assessment. *Nutrients*, 9(657):1–19.
- MMSPG (2017). Multimedia Signal Processing Group (MMSPG), Ecole Polytechnique Fédérale de Lausanne. Lausanne, Suíça. Disponível em <https://mmspg.epfl.ch/downloads/food-image-datasets/>. Acesso em 27 de agosto de 2019.
- Muresan, H. and Oltean, M. (2018). Fruit recognition from images using deep learning. *Acta Univ. Sapientiae, Informatica*, 10(1):26–42.
- Philippi, S. T., Latterza, A. R., Cruz, A. T. R., and Ribeiro, L. C. (1999). Pirâmide alimentar adaptada: Guia para escolha dos alimentos. *Revista de Nutrição Campinas*, 12(1):65–80.
- SBD (2017). Diretrizes da Sociedade Brasileira de Diabetes 2017-2018. Sociedade Brasileira de Diabetes. Clannad Editora Científica.
- Silva, F. M. and de Mello, V. D. F. (2006). Índice glicêmico no manejo do diabetes melito. *Revista HCPA*, 26(2):73–81.
- Simonyan, K. and Zisserman, A. (2015). Very deep convolution networks for large-scale image recognition.