

USO DE APRENDIZADO DE MÁQUINA COMBINADO COM PROCESSAMENTO DE IMAGEM DIGITAL: UMA ESTRATÉGIA PARA CLASSIFICAÇÃO DE AMOSTRAS DE CHÁS

Pedro Raimann Gonçalves¹; Patrick Guilherme Roza¹; Vanderlei Aparecido de Lima¹

¹Universidade Tecnológica Federal do Paraná - Campus Pato Branco

Contato: valima@utfpr.edu.br



O uso do aprendizado de máquina aliado ao processamento digital de imagens permite a classificação de chás, garantindo o controle de qualidade e permite a detecção de adulterações nessa matriz.

INTRODUÇÃO

O uso de aprendizado de máquina, ou *Machine Learning* (ML), em combinação com o processamento digital de imagens tem se apresentado como uma tecnologia promissora para realização de análises voltadas ao controle de qualidade e monitoramento de alimentos. Essa abordagem permite a detecção de compostos em matrizes alimentares e a identificação de possíveis adulterações ou fraudes por meio da interpretação de variações na coloração dos alimentos, sejam elas perceptíveis ou não ao olho humano (MOUSAVIZADEGAN et al., 2024).

Um ponto chave na investigação de matrizes alimentares é a falta de padronização durante sua produção em escala. Alterações na coloração de um produto alimentício podem sinalizar inconsistências em processos industriais, resultando em variações na composição do produto e em suas características organolépticas. Essas diferenças, por sua vez, podem impactar negativamente a percepção e aceitação do consumidor.

Alterações na coloração de matrizes alimentares também podem ser indicativas de adulterações ou fraudes. No Brasil, por exemplo, chás são frequentemente comercializados em pó ou em pequenos fragmentos, o que facilita a substituição da matéria-prima vegetal por outras de qualidade inferior (PASSOS; RAYMUNDO, 2022). Essa prática compromete tanto a autenticidade quanto a qualidade do produto, prejudicando o consumidor.

Este artigo propõe um modelo de classificação baseado em aprendizado de máquina e processamento digital de imagem para distinguir amostras de chás de camomila, erva-doce e capim-cidreira, utilizando como critério principal a coloração dessas matrizes.

PROCESSO DE CLASSIFICAÇÃO

Aquisição de imagens e processamento

Inicialmente foi feito o preparo das amostras dos três tipos de chás: chá de camomila, erva-doce e capim-cidreira, totalizando 312 amostras, sendo 104 para cada tipo de chá. Em seguida foram realizados os registros fotográficos de cada amostra, utilizando um smartphone sob luz ambiente. As imagens foram recortadas para um tamanho de 40 x 40 pixels e processadas digitalmente por meio do software GIMP e no software ChemoStat. O software é utilizado para separar e quantificar a intensidade dos canais de cor RGB e dos espaços de cor HSV associados a imagem de cada amostra. O processamento das imagens digitais no ChemoStat gera como saída, os valores dos atributos R (vermelho), G (verde), B (azul), H (tonalidade), S (saturação), V (valor de iluminância), L (brilho) e I (intensidade).

Os dados obtidos foram submetidos a uma seleção e filtragem por meio do VIF (Fator de Inflação de Variância), no RStudio, com objetivo de evitar a multicolinearidade e remover variáveis altamente correlacionadas. Foram selecionados apenas os atributos com VIF inferior a 5,0, sendo esses: R, H, S e V.

Aplicação do modelo para classificação

A classificação dos três tipos de amostras de chás foi realizada por meio do algoritmo *Support Vector Machine* (Máquina de vetor de suporte, SVM). Esse algoritmo separa dados em diferentes classes por meio da criação de hiperplanos, buscando o hiperplano ótimo que maximiza a distância entre os pontos de diferentes categorias.

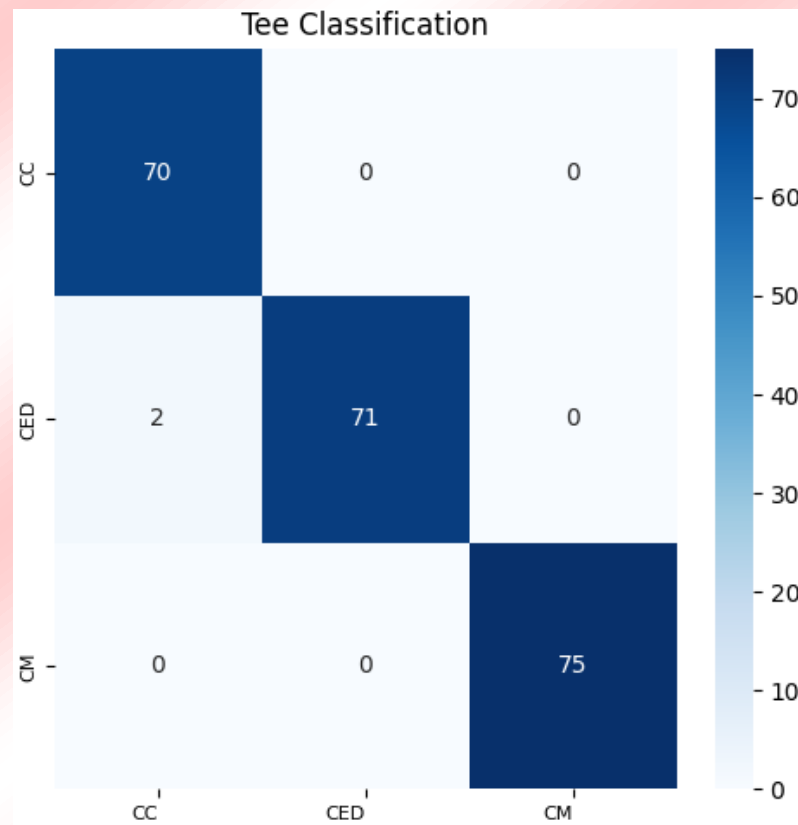
O algoritmo foi implementado utilizando a linguagem de programação *Python*, com o suporte da biblioteca *Scikit-learn*. Os dados foram divididos em dois subconjuntos: 70% para o conjunto de treino, utilizado para aprendizado do modelo, e 30% para o conjunto de teste, utilizado para validação do modelo de classificação.

Avaliação dos resultados

Os resultados gerados pelo modelo construído são apresentados na forma de uma Matriz de Confusão (Figura 1), onde em um eixo são representadas as classes esperadas e no outro as classes previstas pelo modelo. A matriz de confusão para o conjunto teste denotou alta performance para classificação de chás de tonalidades de cores semelhantes. Somente houve dois erros (duas amostras) de chá de erva-doce que foram classificadas como chá de camomila.



Figura 1. Matriz de Confusão para os dados de treino com o modelo SVM.



Com base no número de acertos e erros, são calculadas métricas de avaliação da performance do modelo (Tabela 1).

Tabela 1. Resultado das métricas da performance do modelo SVM para classificação de três tipos de chás.

	Precisão	Recall	F1-score	Suporte
0 - CC	97,00%	100,00%	99,00%	70
1 - CED	100,00%	97,00%	99,00%	73
2 - CM	100,00%	100,00%	100,00%	75
Acurácia global			99,00%	218

CC = Chá de capim-cidreira, CED = Chá de erva-doce, CM = chá de camomila

Os resultados para as classificações dos chás apresentaram alta performance, onde a precisão mínima de acertos foi superior a 95%, porém a acurácia global foi de 99,00% (Tabela 1).

CONSIDERAÇÕES FINAIS

Os resultados das métricas avaliadas mostram que o modelo SVM se ajustou muito bem aos dados, alcançando 99% de acurácia global na classificação das amostras dos três tipos de chá com base nos atributos de cor R, H, S e V.

Essa abordagem utilizando processamento digital de imagens e algoritmos de aprendizado de máquina demonstra grande potencial como ferramenta para padronização de matrizes alimentícias no



setor de produção de alimentos e/ou de bebidas. Com essas abordagens também é possível monitorar o controle de qualidade e detecção de possíveis adulterações em matrizes alimentícias e de bebidas, oferecendo vantagens significativas para empresas produtoras de chás.

Como método de implementação, existe ainda, a possibilidade do desenvolvimento de softwares que automatizem o processo de captura de imagens e análise de dados, tornando a tecnologia mais rápida e acessível. Futuras pesquisas podem combinar a análise de imagem com dados da composição de cada amostra, aumentando a robustez do modelo na identificação de amostras para seu controle de qualidade em indústrias de chás.

REFERÊNCIAS

MOUSAVIZADEGAN, M. et al. Machine learning-assisted image-based optical devices for health monitoring and food safety. *TrAC. Trends in Analytical Chemistry*, [s.l.], v. 177, p. 117794, 1 jun. 2024. Disponível em: <https://doi.org/10.1016/j.trac.2024.117794>. Acesso em: 18 nov. 2024.

PASSOS, J. DE L.; RAYMUNDO, C. E. V. Uso da anatomia foliar no controle de qualidade do chá verde *Camellia sinensis* L. Kuntze (*Theaceae*). *Revista Fitos*, [s.l.], v. 16, n. 1, p. 18-28, 31 mar. 2022. Disponível em: <https://revistafitos.far.fiocruz.br/index.php/revista-fitos/article/view/1124>. Acesso em: 8 nov. 2024.

