

# **Comparação entre Classificadores Tradicionais e Deep Learning na Identificação de Imagens de Alimentos do Dataset Food-101**

**Luany RoseLima de Mélo, Zion Henrique Parcelino Cordeiro Germano, Raquel Mariana Silva Nascimento, Lucas Gabriel de Souza Cabral, Isaac Pessoa Dias Gomes, Lisandro Luan Silva Chagas, Luciano Souza Cabral**

Curso Integrado de Desenvolvimento de Sistemas – Instituto Federal de Educação, Ciência e Tecnologia de Pernambuco (IFPE - Campus jaboatão dos Guararapes)  
Caixa Postal 54080-000 – Jaboatão dos Guararapes – PE – Brazil

lrlm@discente.ifpe.edu.br,  
zhpcg@discente.ifpe.edu.br,  
llsc@discente.ifpe.edu.br,  
ipdg@discente.ifpe.edu.br,  
rmsn1@discente.ifpe.edu.br,  
lgsc5@discente.ifpe.edu.br  
luciano.cabral@jaboatao.ifpe.edu.br

## **Resumo:**

O reconhecimento automático de imagens de alimentos é uma área fundamental para o avanço de aplicações em monitoramento dietético e saúde. Este artigo relata a implementação e avaliação de um sistema de classificação de imagens utilizando modelos de Machine Learning (ML) tradicionais, nomeadamente Support Vector Machine (SVM) e Random Forest (RF), em um subconjunto de dez classes do dataset Food-101. O objetivo principal foi contrastar a eficácia de uma abordagem baseada em características de baixo nível (valores de pixel achados) com o desempenho superior alcançado por arquiteturas de Deep Learning (DL), conforme documentado na literatura. A solução implementada alcançou uma acurácia de validação de aproximadamente 31.00% com o modelo SVM. Essa baixa performance é discutida em detalhe, demonstrando a ineficácia da extração de características simples e ressaltando a imperatividade do uso de recursos hierárquicos de redes neurais convolucionais (CNNs) para lidar com a complexidade e a variabilidade das imagens de alimentos.

## 1. Introdução\*

A classificação precisa de alimentos por meio de suas imagens é uma tarefa desafiadora, mas com vasto potencial de aplicação em saúde pública e nutrição. Projetos que visam a identificação automática de refeições dependem fundamentalmente da robustez dos algoritmos de visão computacional. Enquanto o campo tem sido revolucionado pelo Deep Learning, a avaliação de métodos de Machine Learning tradicional serve como um ponto de referência crucial, especialmente para sistemas com restrições de hardware ou de interpretabilidade.

O presente projeto focou na construção de um classificador para o popular dataset Food-101. Especificamente, investigamos o desempenho de modelos clássicos que utilizam características extraídas diretamente dos pixels redimensionados, atingindo no desenvolvimento de uma aplicação web demonstrativa (via Streamlit ou Flask) com persistência de dados em SQLite e MySQL. O núcleo desta análise reside na comparação direta entre a acurácia obtida pelo nosso classificador tradicional e os resultados de ponta atingidos por modelos de DL, como AlexNet, VGG16, e EfficientNetV2, reportados na literatura recente.

## 2. Contextualizando

Os estudos científicos sobre o dataset Food-101, como os conduzidos por Şengür et al. (2020), estabeleceram que a chave para a alta performance reside na capacidade de extrair "deep features". Nesses trabalhos, a combinação de características extraídas das camadas convolucionais de modelos pré-treinados, como a VGG16 e a AlexNet, em conjunto com um classificador SVM, demonstrou alcançar uma acurácia em torno de 62.44% no dataset completo. Ademais, o ajuste fino (fine-tuning) de uma CNN pré-treinada elevou esse índice para 79.86%, sublinhando a importância da representação de alto nível.

Outros pesquisadores, como Rasyidi et al. (2024), ao trabalharem com um dataset mais diversificado e de maior complexidade (alimentos indonésios), confirmaram a superioridade das arquiteturas modernas de DL, como a EfficientNetV2, reportando acurárias acima de 85.00%. Essa consistência na literatura estabelece um parâmetro de excelência que os modelos tradicionais, por sua natureza, lutam para replicar. A tabela comparativa desses resultados com o nosso experimento é essencial para contextualizar a lacuna de desempenho.

### 3. Solução Proposta

Para este experimento, foi selecionado um subconjunto de dez classes do Food-101. A metodologia de extração de características seguiu o princípio mais simples do ML: as imagens foram redimensionadas para 64x64 pixels e achadas em um vetor de 12.288 features. Essa representação, embora direta, falha em preservar a vizinhança espacial e as estruturas hierárquicas das texturas, formas e cores, que são vitais para a distinção entre classes visualmente semelhantes de alimentos.

Os modelos SVM com kernel RBF e Random Forest foram treinados sobre essas características. Os resultados obtidos em nosso conjunto de validação revelaram a principal limitação da abordagem. O SVM, apesar de ter demonstrado o melhor desempenho de generalização entre os dois, atingiu apenas 31.00% de acurácia de validação. É notável que ambos os modelos exibiram um elevado overfitting, com a acurácia de treino atingindo 82.50% (SVM), evidenciando a incapacidade das características simples de generalizar para o dataset de teste. A matriz de confusão dos experimentos ilustra a falha do modelo em distinguir consistentemente entre as classes.

### 4. A Lacuna de Desempenho

A acurácia de 31.00% obtida no classificador baseado em pixel contrasta de forma dramática com os 62.44% a 79.86% alcançados pelos trabalhos de referência que utilizaram Deep Features no mesmo dataset. Essa diferença não é uma mera questão de ajuste de hiperparâmetros, mas sim uma limitação fundamental da própria representação dos dados.

As imagens de alimentos possuem complexidade visual inerente: as variações de iluminação, ângulos de câmera, texturas sutis e a alta variabilidade intra-classe, exigem características que capturem abstrações complexas. Enquanto os modelos tradicionais dependem de características extraídas manualmente ou de pixels brutos – que são sensíveis ao deslocamento e à escala – as CNNs aprendem a construir representações hierárquicas. As primeiras camadas da CNN detectam bordas e texturas, enquanto as camadas mais profundas combinam essas informações para formar representações conceituais, como a forma de uma massa ou a textura de um vegetal.

Portanto, o resultado de 31.00% não invalida o uso do SVM ou do Random Forest, mas sim a qualidade das features de entrada utilizadas. Demonstra-se que, para o domínio de classificação de imagens de alimentos, a extração automática de características profundas via Deep Learning é um requisito essencial para se alcançar um desempenho comparável ao estado da arte.

Apesar da limitação de acurácia, a solução final foi empacotada em uma aplicação web, demonstrando a integração de Machine Learning com deployment em nuvem e a funcionalidade de persistência de dados, que armazena cada interação do usuário no banco de dados, estabelecendo um framework completo para futuras iterações com modelos de DL mais avançados.

## 5. Conclusão

O presente estudo confirmou que abordagens de Machine Learning tradicional, quando alimentadas apenas com características de pixel de baixo nível, são insuficientes para lidar com a complexidade do dataset Food-101, resultando em uma acurácia de aproximadamente 31.00% e alto overfitting. Este resultado reforça a literatura, que aponta o Deep Learning e a Transferência de Aprendizado como as metodologias ideais para este domínio.

Como trabalho futuro, é imperativo migrar a solução para a utilização de Deep Features extraídas de arquiteturas consagradas (como VGG16 ou EfficientNet) ou, idealmente, realizar o ajuste fino (fine-tuning) de uma CNN pré-treinada. Essa mudança metodológica é a próxima etapa crucial para elevar a acurácia do sistema de classificação para patamares competitivos, aproximando-se ou superando os 79.86% reportados na literatura.

Referências e artigos lidos:

- [https://www.researchgate.net/publication/336730300\\_Food\\_Image\\_Classification\\_with\\_Deep\\_Features](https://www.researchgate.net/publication/336730300_Food_Image_Classification_with_Deep_Features)
- <https://www.beei.org/index.php/EEI/article/view/7996>