

**CLASSIFICAÇÃO AUTOMATIZADA DE CASTAS DE ABELHAS (APIS
MELLIFERA) VIA VISÃO COMPUTACIONAL E APRENDIZADO DE
MÁQUINA**

André Silva (andre_ss1@outlook.com)

Nilsa Duarte Da Silva Lima (nilsa.lima@ufrr.br)

A diferenciação precisa entre abelhas operárias e zangões é fundamental para monitorar a saúde e a dinâmica populacional das colmeias, o equilíbrio entre essas duas castas indica o estado de desenvolvimento e a saúde do enxame. Identificar essas castas automaticamente ajuda a monitorar a produtividade e a detectar anomalias na colônia sem a necessidade de intervenções manuais lentas. Este estudo utilizou o Orange Data Mining, uma ferramenta de programação visual que permite construir fluxos complexos de análise de dados de forma intuitiva e sem a necessidade de codificação. O objetivo central foi avaliar a eficácia de algoritmos de aprendizado de máquina na identificação automatizada dessas castas a partir de imagens digitais. A metodologia seguiu o ciclo de vida de projetos de IA, iniciando com a importação das imagens e a extração de características complexas. Foram comparados quatro modelos de classificação: Neural Network, SVM, kNN e Logistic Regression.

O experimento foi estruturado em três fases distintas. Inicialmente, utilizou-se um conjunto de 452 instâncias (244 operárias e 208 zangões) para treinamento. Na segunda fase, os modelos foram desafiados com um teste

cego contendo 418 imagens inéditas (204 operárias e 214 zangões) para avaliar a capacidade de generalização em condições reais. Por fim, realizou-se uma validação final com um lote de 20 novas imagens, sendo 10 de cada casta, para confirmar a robustez do sistema. No treinamento inicial, todos os modelos atingiram métricas perfeitas de 1.000 em AUC e acurácia (CA). Entretanto, o teste cego com as 418 instâncias trouxe resultados mais realistas, onde a Logistic Regression destacou-se com a maior precisão, alcançando uma acurácia de 0.990. Os outros modelos, como SVM (0.986) e Neural Network (0.974), também apresentaram bom desempenho, enquanto o kNN obteve o resultado inferior com 0.967.

Na fase de validação final, o modelo de Logistic Regression processou as 20 instâncias inéditas, cada uma representada por 1000 variáveis numéricas, e obteve 100% de acerto. As análises visuais por meio de gráficos de dispersão demonstraram uma separação linear absoluta entre os grupos de operárias e zangões, provando que o modelo capturou de forma eficaz as distinções visuais entre as castas. Conclui-se que o Orange é uma ferramenta robusta para a fenotipagem digital na apicultura. A Logistic Regression revelou-se a arquitetura mais consistente e eficiente entre os testes, sendo uma solução para implementação em sistemas de monitoramento automatizado devido à sua alta capacidade de generalização e precisão em dados novos.

Palavras-chave: visão computacional; apicultura; aprendizado de máquina;.