

Classificação de Imagens no Conjunto de Dados MPEG7 Modificado

1st Elisa Ribeiro Gonçalves
UFV Campus Rio Paranaíba
Rio Paranaíba, Brasil
elisa.r.goncalves@ufv.br

2nd Pedro Henrique Campos Moreira
UFV Campus Rio Paranaíba
Rio Paranaíba, Brasil
pedro.henrique.moreira@ufv.br

3th Bernardo Silva Ribeiro Duarte
UFV Campus Rio Paranaíba
Rio Paranaíba, Brasil
bernardo.duarte@ufv.br

Abstract—The K-Nearest Neighbors (KNN) algorithm is widely used for pattern recognition due to its simplicity and effectiveness. This study applied KNN to classify images from the modified MPEG7 dataset, aiming to evaluate the impact of feature normalization and different K values on classification performance. The dataset underwent preprocessing, including standardization of features using mean and standard deviation normalization, followed by training and testing with multiple K values (3, 5, 7, 9, 11, 13, 15). Performance was evaluated using metrics such as accuracy, precision, recall, and F1-Score. The results showed that $K = 3$ achieved the highest accuracy of 84.68%, with F1-Scores per class ranging from 0.59 to 0.94, indicating varying levels of classification difficulty. In addition to KNN, the Random Forest Classifier was used, a decision tree-based model that combines multiple trees to improve accuracy and reduce overfitting. The Random Forest achieved an accuracy of 86.12%, outperforming KNN for this dataset. These results highlight the potential of both KNN and Random Forest for image classification tasks and open avenues for future model improvements.

Index Terms—Image Classification, MPEG7 Dataset, Random Forest, Principal Component Analysis (PCA), Decision Tree, Computer Vision, Feature Extraction.

I. INTRODUCTION

A Visão Computacional é uma área da inteligência artificial que utiliza algoritmos para processar e extrair informações de imagens e vídeos. Com os avanços em aprendizado de máquina e deep learning, essa tecnologia tem sido amplamente aplicada em diversos setores. A segmentação de imagens, uma técnica essencial nesse campo, divide uma imagem em regiões distintas para identificar áreas de interesse, facilitando tarefas como reconhecimento facial, análise de movimento e diagnóstico médico automatizado.

O objetivo deste trabalho foi aplicar uma técnica de segmentação de imagens para melhorar o desempenho de algoritmos de aprendizado de máquina, utilizando ferramentas de visão computacional em *Python*. A segmentação foi implementada para identificar e isolar regiões de interesse nas imagens, o que contribuiu significativamente para o processo de reconhecimento de padrões. A partir desse pré-processamento, técnicas de aprendizado supervisionado, como o uso de KNN, foram aplicadas para classificar e analisar os dados, com resultados expressivos, comprovando a eficácia do pipeline metodológico adotado para o reconhecimento e análise de padrões em imagens do conjunto MPEG7 Modificado.

II. METODOLOGIA

O desenvolvimento do projeto foi estruturado em diversas etapas sequenciais para alcançar um classificador e realizar a segmentação, extração de características morfológicas, divisão do conjunto de dados, treinamento e avaliação do modelo, utilizando métricas como acurácia, precisão, recall e F1-score.

A. Base de dados

O conjunto de dados, é composto por imagens em formato digital, adequadas para segmentação e extração de características morfológicas.

B. Segmentação e extração de características morfológicas das imagens

No desenvolvimento deste trabalho, a segmentação das imagens, que geralmente é a primeira etapa do processo para isolar os objetos de interesse, foi simplificada, uma vez que o dataset disponibilizado já incluía imagens previamente segmentadas. Esse tratamento inicial garantiu que os objetos estivessem claramente delineados, eliminando a necessidade de aplicar técnicas de pré-processamento, como a conversão para escala de cinza ou a binarização, que são comumente utilizadas para uniformizar as informações visuais e destacar as regiões de interesse.

Com as imagens segmentadas, foi possível avançar diretamente para a extração das características morfológicas, etapa essencial para representar os padrões presentes no dataset e preparar os dados para o treinamento do modelo de classificação.

Após isso, foi realizada a extração de características morfológicas dos objetos isolados. Essas características incluíram métricas como área, perímetro, compactidade, momentos estatísticos, razão de aspecto (altura/largura), entre outras descrições geométricas que sintetizam propriedades de forma. Esses atributos constituíram a base de dados que seria utilizada pelo modelo de classificação.

C. Divisão do conjunto de dados

Em seguida, foi feita a divisão do conjunto de dados em subconjuntos de treinamento e teste. A divisão seguiu proporções de 70% para treinamento, e 30% para teste. A função *train_test_split* da biblioteca *Scikit-learn* foi utilizada

para essa etapa, com a configuração do parâmetro *random_state=42*, assegurando a reprodutibilidade dos resultados. Com o objetivo de aumentar a variabilidade do conjunto de treinamento, foram aplicadas técnicas de *data augmentation*. Esse processo incluiu rotações aleatórias(horizantal e vertical), ajustes de brilho e contraste, e aplicação de desfoque gaussiano, simulando condições reais de captura de imagens. Essa etapa foi fundamental para garantir que o modelo fosse robusto a variações, podendo reconhecendo os padrões independentemente de alterações em ângulo, iluminação ou qualidade da imagem. Após o *data augmentation*, as imagens foram verificadas para identificar e corrigir arquivos corrompidos, convertendo-os para RGB e salvando-os novamente. Em seguida, as classes do *dataset* foram identificadas e organizadas, com listagens das imagens correspondentes a cada classe, garantindo uma estrutura clara e consistente para o treinamento e avaliação.

D. Treinamento e avaliação

Após a preparação do conjunto de dados, iniciou-se o treinamento do modelo utilizando o classificador *Random Forest*. Este modelo foi escolhido por sua versatilidade em lidar com diferentes tipos de dados e por sua eficácia contra o *overfitting*, devido ao uso de múltiplas árvores de decisão. Durante o treinamento, foram ajustados os hiperparâmetros do modelo para maximizar seu desempenho.

Durante o processo de treinamento e avaliação, foi utilizado o classificador *K-Nearest Neighbors* (KNN), devido à sua simplicidade e eficácia em cenários de classificação baseada em distância. Após a divisão dos dados em conjuntos de treinamento e teste, o modelo foi treinado utilizando as características extraídas das imagens segmentadas e pré-processadas. A avaliação do desempenho do modelo foi feita com base em métricas como acurácia, precisão, recall e F1-score. A acurácia, especificamente, foi calculada também utilizando o KNN para prever as classes das imagens de teste, comparando as previsões do modelo com os rótulos reais. O desempenho do KNN foi analisado e comparado com outros classificadores, possibilitando uma análise mais robusta da eficiência do modelo para a tarefa proposta.

Adicionalmente, foi gerada uma matriz de confusão para identificar possíveis confusões entre as classes. A análise dos resultados revelou um desempenho geral satisfatório, com uma acurácia aproximada de 84%. Essa taxa de acerto indica a capacidade do modelo em identificar padrões complexos com precisão.

Para ilustrar os resultados foram feitas diversas visualizações, como a plotagem de previsões versus rótulos verdadeiros, a demonstração de uma das árvores de decisão que compõe o *Random Forest* e a visualização em 2D das classificações por meio da Análise de Componentes Principais (PCA). Esses gráficos mostraram claramente as previsões corretas e incorretas, oferecendo uma visão detalhada do comportamento do classificador.

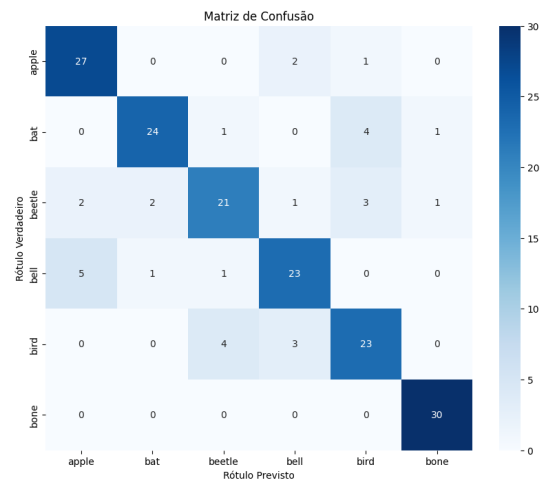


Fig. 1. Caption

III. RESULTADOS

Os resultados obtidos indicaram que a abordagem proposta é eficiente para o reconhecimento de padrões no conjunto de dados. A metodologia incluiu normalização das características, segmentação, extração de atributos e aprendizado de máquina utilizando o algoritmo *K-Nearest Neighbors* (KNN). A normalização permitiu centralizar os dados, reduzindo a média das características do conjunto de treinamento para valores próximos de zero e normalizando o desvio padrão.

O melhor desempenho foi alcançado com $K = 3$, obtendo uma acurácia de 84,68%, enquanto outras escolhas de K apresentaram ligeira redução no desempenho. Além disso, o classificador demonstrou eficácia variável por classe, com métricas de F1-Score alcançando valores como 0,93 para "apple" e 0,94 para "bone", mas menores para "bell" (0,67) e "bird" (0,59), evidenciando diferenças no nível de dificuldade entre as classes. A precisão geral média foi de 81%, refletindo a robustez do modelo para lidar com classes distintas.

A análise das métricas de precisão, recall e F1-Score por classe revelou que o modelo apresentou alta precisão para classes como "bone" (1,00) e "apple" (0,93), enquanto o recall foi mais desafiador para classes como "bird" (0,50). Esses resultados mostram que a performance do modelo pode ser afetada por fatores como a similaridade entre as características das classes.

Conclui-se que o modelo proposto é capaz de classificar eficientemente as imagens do conjunto MPEG7 Modificado, atingindo bons índices de desempenho.

REFERENCES

- [1] "O Que é Visão Computacional? - Data Science Academy". Data Science Academy. Consult. 2024-11-18. [Em linha].
- [2] G. Câmara. "CLASSIFICAÇÃO COM RANDOM FOREST". Medium. Consult. 2024-11-19. [Em linha]. Disponível: <https://medium.com/@gustavocamara/classificação-com-random-forest-6c508a046537>.