Abstract

This project constructs and evaluates three neural network architectures for food image classification: CNN, RNN, and LSTM. Labeled food photographs were preprocessed with data augmentation techniques including scaling to 224×224 pixels, rotation, shear, zoom, horizontal flip, and brightness modifications. The dataset was split into training, validation, and test sets for robust evaluation.

The CNN model utilized a pre-trained VGG16 backbone with GlobalAveragePooling2D, a 256-unit dense layer, and softmax classification, achieving 82% accuracy. In contrast, both RNN and LSTM models achieved only 58% accuracy despite incorporating TimeDistributed layers and sequential processing capabilities. The sequential models struggled with static image data lacking temporal dependencies and performed poorly on underrepresented classes.

Class imbalance significantly affected performance across all models, particularly for categories like 'Macarons-Choco' and 'Macarons-Mint'. The confusion matrix analysis revealed misclassifications between visually similar food categories. CNN with VGG16 transfer learning demonstrated superior hierarchical feature extraction capabilities compared to sequential architectures, proving convolutional networks are most effective for food image classification tasks.

Commented [YB1]: *Project scope and methodology
*Performance results and technical specifications
*Analysis of challenges and limitations
*Main conclusions about model effectiveness