Abstract

Indonesia is a country with a rich cultural heritage and a wide variety of traditions, which highlights the need to preserve various aspects of its heritage, including its scripts. However, over time, these scripts have become less commonly studied, partly due to their complex forms, such as Balinese and Javanese scripts. Moreover, the Sundanese script is also at risk of extinction. This study conducted seven different class combinations: individual scripts, pairwise combinations, and all three scripts together. The results were 91.64% for Balinese, 93.19% for Javanese, 98.44% for Sundanese, 93.42% for Balinese-Javanese, 95.52% for Balinese-Sundanese, 96.43% for Javanese-Sundanese, and 95.72% for all three scripts combined. Through comprehensive error analysis and characters-likeness rate indices, the study identified specific character pairs with high likeness rate that posed classification challenges, notably "la" and "ha" (Javanese), "nga" and "nya" (Balinese), "sa" and "da" (Javanese), and "ha" and "la" (Balinese). This study focuses on the likeness rate between paired characters, both within a single language (intra-language) and across languages (inter-language). The findings demonstrate that the CNN model effectively distinguishes between characters from different scripts, without misclassifications occurring either within or across languages. This study contributes to the digital preservation of Indonesia's linguistic heritage.

Keywords: balinese script, javanese script, sundanese script, cnn, vgg-16