ABSTRACT

Aksara is an orthographic system resulting from regional communities which includes characters and pronunciation systems for writing regional languages, one of which is Sundanese script. Various studies as an effort to digitize in the preservation of Sundanese script with various methods in machine learning with a focus on Optical Recognition Character (OCR) especially Convolutional Neural Network (CNN) have been carried out. However, the current data is limited to characters that are part of swara, ngalagena, rarangkén and numbers. There is only a few of OCR using data in the form of a combination of ngalagena and swara with rarangkén into a word dataset. With the addition of the word dataset, it can produce two models, namely the ensemble model and the Connectionist Temporal Classification (CTC) model. The ensemble model with an ensemble weight of 0.5 achieved an accuracy of 0.912, precision of 0.889, sensitivity of 0.701, and F1-score of 0.755. While, the CTC model successfully achieved a Character Error Rate (CER) value of 0.069 and a Word Error Rate (WER) value of 0.087.

Keywords: *Optical Character Recognition*, Aksara Sunda, *Convolutional Neural Network*, *Recurrent Neural Network*, *Connectionist Temporal Classification*, EfficientNet.