

KLASIFIKASI *ISOLATED CHARACTER* AKSARA BALI PADA CITRA LONTAR MENGGUNAKAN METODE CNN DAN SVM

Oleh

Gusti Ayu Devani Zelvia, 2215101061

Jurusan Teknik Informatika

Program Studi S1 Ilmu Komputer

ABSTRAK

Penelitian ini mengembangkan dan membandingkan model *Convolutional Neural Network* (CNN) dan *hybrid CNN-SVM* untuk klasifikasi *isolated character* aksara Bali pada citra manuskrip lontar. Pelestarian aksara Bali menjadi tantangan karena semakin sedikitnya generasi muda yang mampu membacanya. Dataset yang digunakan berasal dari ICFHR 2016 Challenge 3: *Isolated Character Recognition of Balinese Script in Palm Leaf Manuscript Images* (AMADI_LontarSet), terdiri dari 133 kelas karakter dengan 10.540 citra latih dan 7.673 citra uji. Teknik augmentasi data diterapkan untuk memperluas variasi data latih. Model CNN kustom dilatih secara *end-to-end*, kemudian pada pendekatan *hybrid*, lapisan klasifier CNN diganti dengan *Support Vector Machine* (SVM). Metrik evaluasi meliputi akurasi, presisi, *recall*, dan *F1-score*. Hasil menunjukkan model CNN mandiri mencapai akurasi 92,38%, melampaui *state-of-the-art* sebelumnya dari tim G20 ICFHR 2018 (92,17%). Pengujian pada model CNN mendapatkan akurasi sebesar 92,38% dengan *F1-Score Macro* 0,8476 dan *F1-Score Weighted* 92,41%. Model CNN+SVM berhasil meningkatkan performa pada seluruh metrik utama, dengan akurasi menjadi 92,60% (selisih 0,22%), *F1-Score Macro* menjadi 0,8537 (selisih 0,61%), dan *F1-Score Weighted* menjadi 92,67% (selisih 0,26%), menjadikannya model terbaik pada dataset yang sama. Jumlah kelas berperforma rendah berkurang dari 16 menjadi 14 kelas setelah integrasi SVM. Temuan ini mendukung pengembangan sistem digitalisasi aksara Bali pada manuskrip lontar.

Kata kunci: Aksara Bali, CNN, CNN-SVM, Klasifikasi, Manuskrip Lontar, SVM.

CLASSIFICATION OF *ISOLATED CHARACTER* OF BALINESE SCRIPT IN LONTAR IMAGES USING CNN AND SVM METHODS

By

Gusti Ayu Devani Zelvia, 2215101061

Department of Informatics Engineering

S1 Computer Science Study Program

ABSTRACT

This study develops and compares a Convolutional Neural Network (CNN) model and a hybrid CNN–SVM approach for classifying isolated Balinese script characters in lontar manuscript images. The preservation of Balinese script has become a challenge as fewer young generations are able to read it. The dataset was sourced from ICFHR 2016 Challenge 3: Isolated Character Recognition of Balinese Script in Palm Leaf Manuscript Images (AMADI_LontarSet), consisting of 133 character classes with 10,540 training images and 7,673 test images. Data augmentation techniques were applied to expand training data variation. A custom CNN architecture was trained end-to-end; in the hybrid approach, the CNN classifier layer was replaced with a Support Vector Machine (SVM). Evaluation metrics include accuracy, precision, recall, and F1-score. Results show that the standalone CNN model achieved 92.38% accuracy, surpassing the previous state-of-the-art by ICFHR 2018 Team G20 (92.17%). Testing on the CNN model achieved an accuracy of 92.38%, with a Macro F1-Score of 0.8476 and a Weighted F1-Score of 92.41%. The CNN+SVM model successfully improved performance across all key metrics, with accuracy increasing to 92.60% (a difference of 0.22%), Macro F1-Score increasing to 0.8537 (a difference of 0.61%), and Weighted F1-Score increasing to 92.67% (a difference of 0.26%), making it the best-performing model on the same dataset. The number of low-performing classes decreased from 16 to 14 after SVM integration. These findings support the development of a Balinese script digitalization system for lontar manuscripts.

Keywords: Balinese Script, Classification, CNN, CNN-SVM, Lontar Manuscript, SVM.