

INTISARI

Pengenalan Gestur Bahasa Isyarat Indonesia Menggunakan Model Hybrid Deep Learning

by

Muhammad Yusuf Daffa Izzalhaqqi

19/444312/PA/19374

Bahasa isyarat menggunakan gerakan tangan yang spesifik bagi setiap negara sebagai alat komunikasi antara komunitas yang dapat mendengar dan komunitas tuna rungu. Dengan prevalensi global gangguan pendengaran yang diperkirakan akan meningkat dua kali lipat pada tahun 2050, studi mengenai bahasa isyarat menjadi sangat penting. Berbagai pendekatan menggunakan *deep learning* telah diusulkan oleh beberapa peneliti untuk mengenali gerakan dalam bahasa isyarat yang terkenal. Dalam penelitian ini, sebuah model *hybrid deep learning* dikembangkan untuk mengenali gerakan alfabet dalam Bahasa Isyarat Indonesia (BISINDO), dengan memanfaatkan keunggulan CNN dalam menangkap fitur spasial dan LSTM dalam menangkap fitur temporal.

Dataset utama yang terdiri dari 23.000 gambar yang mewakili 23 gerakan dalam Bahasa Isyarat Indonesia dikumpulkan untuk melatih dan mengevaluasi model *deep learning* mengingat keterbatasan ketersediaan dataset publik. Gambar-gambar ini melalui tahap pemisahan, augmentasi, dan pra-pemrosesan sebelum pelatihan model. Kinerja model *hybrid CNN-LSTM* yang diusulkan dibandingkan dengan model komponennya menggunakan beberapa metrik evaluasi termasuk matriks kebingungan, akurasi, presisi, recall, dan skor F1.

Hasil penelitian menunjukkan bahwa model *hybrid* menunjukkan akurasi yang tinggi yaitu 99,60%, 84,87%, dan 98,00% pada set pelatihan, validasi, dan pengujian. Namun, CNN mencapai akurasi yang sedikit lebih tinggi yaitu 99,74%, 98,04%, dan 99,90% pada set yang sama. Kedua model juga memberikan rata-rata makro yang mendekati sempurna, menunjukkan tingkat presisi, recall, dan skor F1 yang tinggi di semua kelas. Sebaliknya, LSTM menunjukkan *overfitting* dengan akurasi yang jauh lebih rendah yaitu 97,23%, 51,17%, dan 65,10% pada set pelatihan, validasi, dan pengujian. Model ini juga menunjukkan kinerja klasifikasi yang buruk dengan skor rata-rata makro 0,665, 0,650, 0,581 untuk presisi, recall, dan skor F1.

Kata Kunci: *Pengenalan Bahasa Isyarat, BISINDO, Deep Learning, CNN, LSTM, Hybrid Model*

ABSTRACT

GESTURE RECOGNITION IN INDONESIAN SIGN LANGUAGE USING HYBRID DEEP LEARNING MODELS

by

Muhammad Yusuf Daffa Izzalhaqqi

19/444312/PA/19374

Sign language employs country-specific hand gestures as a means of communication between the hearing and hearing-impaired communities. With the global prevalence of hearing loss reaching projected to double by 2050, studies on sign language are of great importance. Various deep learning approaches have been proposed by researchers to recognize gestures in well-known sign languages. In this study, a hybrid CNN-LSTM model is developed to recognize alphabet gestures in Indonesian Sign Language (BISINDO), leveraging the strengths of CNN and LSTM in capturing spatial and temporal features, respectively.

A primary dataset of 23,000 images representing 23 gestures in Indonesian Sign Language are collected to train and evaluate the deep learning model, given the limited availability of public dataset. These images undergo stages of splitting, augmentation, and preprocessing prior to the model training. The performance of the proposed hybrid CNN-LSTM model is compared to that of its constituent models using several evaluation metrics including confusion matrix, accuracy, precision, recall, and F1 score.

The results suggest that the hybrid model exhibits remarkable accuracies of 99.60%, 84.87%, and 98.00 in training, validation, and testing sets respectively. However, the CNN achieves slightly higher accuracies of 99.74%, 98.04%, and 99.90% in same sets. Both models also yield near-perfect macro averages, indicating high degree of precision, recall, and F1-score across all the classes. In contrast, the LSTM exhibits overfitting with significantly lower accuracies of 97.23%, 51.17%, and 65.10% on the training, validation, and testing sets, respectively. It also demonstrates poor classification performance with macro average scores of 0.665, 0.650, 0.581 for precision, recall, and F1-score, respectively.

Keywords: *Sign Language Recognition, BISINDO, Deep Learning, CNN, LSTM, Hybrid Model*