

ABSTRAK

Pengenalan gestur tangan dinamis kontinu bisa menjembatani interaksi manusia dengan komputer melalui informasi visual. Aspek kontinu diperuntukkan supaya pengenalan gestur lebih *robust* karena sifat alamiah dari gestur manusia adalah berlangsung secara berkelanjutan. Pengenalan gestur dinamis kontinu menghasilkan fitur spasial dan temporal yang diproses dengan algoritme yang bisa memahami aspek waktu untuk melakukan prediksi. *Long Short-Term Memory* (LSTM) merupakan algoritme yang dirancang untuk mempelajari informasi yang memiliki aspek urutan waktu, namun dalam riset gestur dinamis menunjukkan keterbatasan dalam akurasi dibandingkan arsitektur yang lebih baru, contohnya *attention mechanism*. Penelitian ini mengusulkan pendekatan baru pada algoritme yang bertugas untuk melakukan prediksi gestur tangan dinamis. *Extended Long Short Term Memory* (xLSTM) hadir sebagai pembaharuan dari LSTM untuk mengatasi berbagai kekurangannya dari segi struktural untuk mengolah data dalam bentuk sekuens. xLSTM hadir dengan penerapan fungsi gerbang eksponensial yang disertai sel sLSTM dan mLSTM untuk mendukung kemampuan revisi informasi dan kapasitas memori yang lebih mumpuni untuk menangkap dependensi temporal. Penelitian ini mengusung *MediaPipe Hands* sebagai ekstraktor fitur tangan dalam bentuk informasi *skeleton* yang kemudian dipreproses ke bentuk sekuens. Dalam konteks gestur tangan dinamis kontinu, *IPN Hands Dataset* digunakan sebagai bahan untuk menunjukkan performa model dalam melakukan klasifikasi gestur secara berkelanjutan. Evaluasi performa dilakukan mencakup gestur navigasi antarmuka pengguna seperti *Pointing1F*, *Pointing2F*, *Click1F*, *Click2F*, *Throw Up*, *Throw Down*, *Throw Left*, *Throw Right*, *Open Twice*, *Double Click1F*, *Double Click2F*, *Zoom In*, dan *Zoom Out*, ditambah *Non-Gesture* atau *Idle*. Hasil pengujian menunjukkan bahwa dalam kondisi *benchmarking* yang setara, model xLSTM mencapai *test accuracy* sebesar 92,3% dengan F1-Score 0,88, mengungguli LSTM standar yang hanya mencapai akurasi 84,5% dan F1-Score 0,7130. Performa xLSTM ini kemudian ditingkatkan lebih lanjut melalui proses *tuning* manual, mencapai *test accuracy* optimal sebesar 93,27% dengan F1-Score 0,8989 pada konfigurasi terbaik. Model ini diuji secara lokal inferensinya dengan sampel yang menunjukkan rata-rata *confidence score* 99,3%, FPS 29,87, latensi 33,08 ms, dan waktu inferensi model 10,63 ms, berjalan di perangkat CPU sebagai demonstrasi dari pengenalan gestur secara *real-time*.

Kata kunci *MediaPipe*, xLSTM, Gestur

ABSTRACT

Continuous dynamic hand gesture recognition can bridge human-computer interaction through visual information. The continuous aspect is intended to make gesture recognition more robust, as the natural characteristic of human gestures is that they occur continuously. Continuous dynamic gesture recognition produces spatial and temporal features that must be processed using algorithms capable of understanding the temporal aspect to perform predictions. Long Short-Term Memory (LSTM) is an algorithm designed to learn information with temporal sequence aspects; however, in dynamic gesture research, it has shown limited accuracy compared to more recent architectures such as attention mechanisms. This research proposes a new approach to the algorithm responsible for dynamic hand gesture prediction. Extended Long Short-Term Memory (xLSTM) emerges as an update to LSTM to overcome its structural weaknesses in processing sequential data. xLSTM applies exponential gating functions accompanied by sLSTM and mLSTM cells to support information revision capability and a more robust memory capacity for capturing temporal dependencies. This research utilizes MediaPipe Hands as a hand feature extractor in the form of skeleton information, which is subsequently preprocessed into sequences. In the context of continuous dynamic hand gestures, the IPN Hands Dataset is used to demonstrate the model's performance in continuous gesture classification. Performance evaluation covers user interface navigation gestures such as Pointing1F, Pointing2F, Click1F, Click2F, Throw Up, Throw Down, Throw Left, Throw Right, Open Twice, Double Click1F, Double Click2F, Zoom In, and Zoom Out, along with Non-Gesture or Idle. The test results show that under equal benchmarking conditions, the xLSTM model achieves a test accuracy of 92.3% with an F1-Score of 0.88, outperforming standard LSTM, which only achieves an accuracy of 84.5% and an F1-Score of 0.7130. The performance of xLSTM was further improved through a separate manual tuning process, reaching an optimal test accuracy of 93.27% with an F1-Score of 0.8989 at the best configuration. The system was also tested locally for its inference performance, showing an average confidence score of 99.3%, 29.87 FPS, an average latency of 33.08 ms, and an average inference time of 10.63 ms when running on a CPU as a demonstration of real-time gesture recognition.

Keywords MediaPipe, xLSTM, Gesture