

BAB I

PENDAHULUAN

1.1 Latar Belakang

Human Activity Recognition (HAR) berbasis video merupakan bidang penelitian yang menarik dan menantang pada era ini terutama dalam mengidentifikasi dan mengenali tindakan manusia dibidang pengawasan keamanan berbasis video, bantuan medis serta realitas virtual (Żelawski & Hachaj, 2021). HAR merupakan kunci untuk lebih memahami berbagai semantik yang terkandung dalam video (Zhang dkk., 2020). Hingga saat ini, petugas keamanan dalam menganalisis suatu kejadian masih dilakukan secara manual, yaitu dengan mengamati video tindakan manusia yang diperoleh dari cctv secara berulang-ulang, yang mana sangat menyita waktu dan tenaga. Sedangkan pengidentifikasian secara otomatis dapat dilakukan dengan bantuan program *machine learning* yang disematkan pada sistem sehingga dapat mengklasifikasi tindakan manusia. Pengenalan tindakan pada video memiliki banyak aplikasi, seperti pengambilan video berbasis konten, sistem pengawasan untuk tujuan keamanan dan privasi, dan interaksi manusia-komputer, telah menarik perhatian besar dalam visi komputer baru-baru ini (Mokari dkk., 2020).

Deep learning menjadi teknik yang paling banyak digunakan untuk proses pembelajaran mesin dengan berbagai fitur tingkat tinggi dalam pengenalan tindakan dan perilaku berbasis video. Pendekatan *deep learning* pada HAR menggunakan arsitektur sederhana *Convolutional Neural Networks* (CNN) dalam operasi

konvolusi untuk mempelajari kerangka pada video setiap melakukan proses pelatihan. Lapisan konvolusi ini mengekstrak dan mempelajari fitur spasial untuk melatih model klasifikasi. Pendekatan lain pada HAR menggunakan arsitektur klasik *Recurrent Neural Network* (RNN), yaitu *Long Short-Term Memory* (LSTM). Pada arsitektur LSTM memiliki mekanisme internal yang disebut gerbang dalam mengatur aliran informasi, sehingga pada arsitektur LSTM pandai dalam memodelkan urutan video dengan kerangka yang panjang (Donahue dkk., 2016).

Penelitian yang dilakukan (Muhammad dkk., 2021) dengan judul *Human action recognition using attention based LSTM network with dilated CNN features*. Penelitian ini berfokus pada teknik pengenalan tindakan arsitektur pembelajaran mesin untuk representasi visual *frame* video pada tahap pelatihan yang dapat memengaruhi penentuan perbedaan fitur, seperti perbedaan visual dengan temporal. Pada penelitian ini mengusulkan arsitektur pembelajaran mesin pada HAR menggunakan algoritma *Bi-Directional Long Short-Term Memory* (BiLSTM) dan *Dilated Convolutional Neural Network* (DCNN) yang secara selektif berfokus pada fitur dalam masukan *frame* yang efektif untuk mengenali berbagai tindakan manusia dalam video.

Penelitian lain yang terkait yaitu *Human action recognition using convolutional LSTM and fully-connected LSTM with different attentions* (Zhang dkk., 2020). Penelitian ini bertujuan untuk memfasilitasi pengenalan tindakan dengan arsitektur jaringan perhatian ganda STDAN, yang menggabungkan *Convolutional LSTM* dan *Fully-Connected LSTM*.

Penelitian lain yang dilakukan oleh (Donahue dkk., 2016) dengan judul *Long-term Recurrent Convolutional Networks for Visual Recognition and Description*. Penelitian ini menjelaskan tentang kelas arsitektur konvolusi berulang yang dapat dilatih secara *end to end* dan sesuai pada tugas pemahaman visual skala besar untuk pengenalan aktivitas, teks gambar, dan deskripsi video.

(Park dkk., 2016) menggunakan metode RNN berbasis kamera kedalaman pada HAR untuk layanan kesehatan dan perawatan sosial. Metode yang diusulkan mencapai akurasi pengenalan rata-rata 99,55% dan mampu mengenali dua belas aktivitas manusia secara andal.

(Arif dkk., 2019) mengusulkan kombinasi metode 3D-CNN dengan LSTM. Pada prosesnya jaringan konvolusi 3 dimensi akan menggabungkan informasi diskriminatif dari video ke dalam peta yang disebut *motion map*. Penggabungan *motion map* dan *frame video* dapat dijadikan teknik untuk meningkatkan panjang video pelatihan secara iteratif. Penerapan skema fusi berbobot linier berfungsi untuk menggabungkan peta fitur jaringan menjadi fitur *spatio-temporal* dan penerapan *encoder-decoder* pada LSTM untuk melakukan proses prediksi akhir.

Penelitian lain mengenai HAR menggunakan teknik *skeleton joints* yang dilakukan oleh (Surayahani dkk., 2018) dengan judul *Human Activity Recognition Based On Optimal Skeleton Joints Using Convolutional Neural Network*. Penelitian ini berfokus pada *noise* yang menyebabkan titik *joint* yang tidak relevan dan tidak bergerak, sehingga menjadi penyebab utama penurunan kinerja HAR.

(Deep & Zheng, 2019) melakukan penelitian menggunakan metode *transfer learning* VGG-16 untuk mendapatkan fitur gambar yang dalam dan proses

klasifikasi *machine learning* yang terlatih pada CNN dalam pengenalan aktivitas manusia. Akurasi metode yang diusulkan pada penelitian ini yaitu mencapai 96.95%.

Penelitian yang dilakukan oleh (Zhao dkk., 2020) dengan judul *Improved two-stream model for humanaction recognition*, berfokus pada struktur *two-stream* menggunakan model berbasis LSTM dalam aliran spasialnya untuk mengekstrak fitur spasial dan temporal pada *RGB frame*.

Terdapat penelitian mengenai HAR yang berfokus pada proses multi-kelas klasifikasi dalam peningkatan akurasi dengan biaya komputasi rendah (Xu dkk., 2018) dan pengurangan kompleksitas model dengan proses eliminasi yang dibutuhkan untuk teknik fitur tingkat lanjut (Mutegeki & Han, 2020).

(Liu dkk., 2017) melakukan penelitian dalam perancangan arsitektur model menggunakan kombinasi metode 3DCNN dan LSTM. Pada penelitian ini model yang diusulkan mampu melakukan *stack* pada *frame* video, mengekstrak fitur waktu dan spasial, dan melakukan proses pelatihan pada kumpulan data masukan untuk mencapai kinerja pengenalan yang baik. Model LSTM yang didasarkan pada hubungan antar *frame* dalam waktu yang berbeda diterapkan untuk mempertimbangkan informasi yang dihasilkan pada *frame* sebelumnya.

Berdasarkan penelitian terkait setiap metode memiliki kelebihan dan kekurangannya tersendiri pada arsitektur model *machine learning* dalam identifikasi aktivitas manusia. Maka dari itu, penelitian ini dilakukan dengan menggunakan kombinasi metode CNN dan LSTM. Pada prosesnya, lapisan konvolusi yang digunakan untuk mengekstraksi fitur spasial dari *frame*, akan

diumpangkan ke lapisan LSTM pada setiap jaringan untuk pemodelan urutan temporal. Dengan cara ini jaringan pada model akan mempelajari fitur *spatiotemporal* secara langsung dalam pelatihan *end-to-end*, untuk menghasilkan model yang kuat.

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas maka diperoleh rumusan masalah sebagai berikut:

1. Bagaimana proses klasifikasi aktivitas manusia menggunakan kombinasi algoritma CNN dan LSTM berbasis pengolahan visual pada video?
2. Bagaimana parameter kualitas model *machine learning* menggunakan kombinasi algoritma CNN dan LSTM untuk mengklasifikasi aktivitas manusia berbasis pengolahan visual pada video?

1.3 Batasan Masalah

Berdasarkan latar belakang diatas terdapat batasan masalah dalam penelitian ini sebagai berikut:

1. Perancangan model *machine learning* yang dilakukan berfokus pada kombinasi algoritma CNN dan LSTM pada HAR.
2. Data pada penelitian ini menggunakan dataset UCF50.
3. Aktivitas manusia yang menjadi data dalam proses penelitian terdiri dari 10 aktivitas olahraga.

4. Ukuran *frame* video masukan adalah 64×64 piksel dengan 3 *channel* warna dan maksimal panjang sekuen adalah 20 *frame*.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang diperoleh maka tujuan penelitian ini sebagai berikut:

1. Melakukan proses klasifikasi aktivitas manusia menggunakan kombinasi algoritma CNN dan LSTM berbasis pengolahan visual pada video.
2. Menganalisa performa variasi pada parameter *batch size* dan *dropout layer* menggunakan kombinasi algoritma CNN dan LSTM untuk mengklasifikasi aktivitas manusia berbasis pengolahan visual pada video.

1.5 Manfaat Penelitian

Melalui penelitian pada tugas akhir ini diharapkan dapat memberi manfaat sebagai berikut:

1. Memberikan pengetahuan terkait proses klasifikasi aktivitas manusia pada *deep learning* dengan menggunakan kombinasi algoritma CNN dan LSTM.
2. Memberikan model *machine learning* yang dapat digunakan untuk mengidentifikasi dan mengenali tindakan manusia dibidang olahraga berbasis video sebagai referensi dalam penelitian selanjutnya.
3. Kontribusi terhadap area penelitian mengenai bidang kajian *Computer Vision* dengan berfokus pada bidang *deep learning* berupa optimalisasi parameter kualitas pada HAR.

1.6 Metodologi Penelitian

Berdasarkan latar belakang diatas disusunlah metodologi penelitian untuk melakukan penelitian sebagai berikut:

a. Studi Literatur

Melakukan pendalaman materi mengenai model *machine learning* pada HAR yaitu dengan metode CNN dan LSTM. Literatur yang dilakukan yaitu melalui buku, jurnal, artikel, dan website yang membahas tentang *activity recognition*.

b. Definisi Persoalan Penelitian

Mendefinisikan persoalan penelitian yang akan diteliti berdasarkan kajian dari tahap studi literatur.

c. Persiapan Data

Melakukan persiapan data untuk kebutuhan eksperimen pengujian model penelitian yang akan dilakukan.

d. Pemodelan

Pembuatan model penelitian dengan merancang dan mengimplementasikan arsitektur yang menjadi usulan model penelitian.

e. Evaluasi Hasil

Evaluasi dilakukan dengan beberapa data *testing* untuk mengetahui performa pada model yang telah dibangun menggunakan metode *confusion matrix*.

1.7 Sistematika Penulisan

Sistematikan penulisan Tugas Akhir disusun agar penulisan laporan penelitian dapat terarah sesuai tahapan penelitian. Adapun sistematika yang akan diuraikan secara garis besar, yaitu:

BAB I PENDAHULUAN

Bab ini menguraikan tentang garis besar keseluruhan penelitian. Terdiri dari latar belakang, rumusan masalah, batasan masalah, tujuan, manfaat, metodologi dan sistematika penulisan penelitian.

BAB II TINJAUAN PUSTAKA

Bab ini berisi mengenai beberapa penelitian yang terkait dengan topik penelitian yang akan dilakukan. Selain itu, bagian ini juga berisi mengenai penjelasan teori-teori yang akan digunakan dalam penelitian.

BAB III METODOLOGI PENELITIAN

Bab ini berisi tentang metodologi atau tahapan-tahapan penelitian yang akan dilakukan. Selain itu, bagian ini juga memaparkan mengenai jenis penelitian, waktu dan tempat penelitian.

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisi tentang hasil eksperimen yang dilakukan berdasarkan tahapan-tahapan yang dimuat dalam metodologi penelitian.

BAB V KESIMPULAN DAN SARAN

Bab ini berisi tentang kesimpulan hasil eksperimen yang merupakan rangkuman pembahasan pada penelitian dan saran untuk penelitian selanjutnya yang membahas mengenai topik sejenis atau terkait.