



UNIVERSITAS DIPONEGORO

**KLASIFIKASI *HUMAN ACTION RECOGNITION* (HAR)
BERBASIS MEDIAPIPE DAN *LONG SHORT-TERM MEMORY*
(LSTM)**

TUGAS AKHIR

**Diajukan sebagai salah satu syarat untuk memperoleh gelar
Sarjana Teknik**

ICHSAN ARSYI PUTRA

21120118120029

**FAKULTAS TEKNIK
DEPARTEMEN TEKNIK KOMPUTER
SEMARANG**

2022

HALAMAN PENGESAHAN

Tugas Akhir ini diajukan oleh

Nama : Ichsan Arsyi Putra
NIM : 21120118120029
Jurusan/Program Studi : Teknik Komputer
Judul Tugas Akhir : Klasifikasi *Human Action Recognition* (HAR)
Berbasis MediaPipe dan *Long Short-Term Memory*
(LSTM)

Telah berhasil dipertahankan di hadapan Tim Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Teknik pada Jurusan/Program Studi Teknik Komputer, Fakultas Teknik, Universitas Diponegoro.

TIM PENGUJI

Pembimbing I : Dr. Oky Dwi Nurhayati, S.T., M.T. ()
Pembimbing II : Dania Eridani, S.T., M.Eng. ()
Ketua Penguji : Dr. Adian Fatchur Rochim, S.T., M.T. ()
Anggota Penguji : Risma Septiana, S.T., M.Eng. ()

Semarang, 4 Juli 2022

Ketua Departemen Teknik Komputer



Dr. Adian Fatchur Rochim, S.T., M.T.


NIP. 197302261998021001

HALAMAN PERNYATAAN ORISINALITAS

Tugas Akhir ini adalah hasil karya saya sendiri, dan semua sumber baik yang dikutip maupun yang dirujuk telah saya nyatakan dengan benar

Nama : Ichsan Arsyi Putra

NIM : 21120118120029

Tanda Tangan : 

Tanggal : 8 Juni 2022

**HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI
TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS**

Sebagai civitas akademika Universitas Diponegoro, saya yang bertanda tangan di bawah ini:

Nama : Ichsan Arsyi Putra
NIM : 21120118120029
Jurusan/Program Studi : Teknik Komputer
Fakultas : Teknik
Jenis Karya : Tugas Akhir

demikian pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Diponegoro **Hak Bebas Royalti Noneksklusif** (*Non-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul:

Klasifikasi *Human Action Recognition* (HAR) Berbasis MediaPipe dan Long Short-Term Memory (LSTM)

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti/Noneksklusif ini, Universitas Diponegoro berhak menyimpan, mengalihmedia/formatkan, mengelola, dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik hak cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Semarang
Pada Tanggal : 8 Juni 2022

Yang menyatakan



Ichsan Arsyi Putra

KATA PENGANTAR

Puji syukur kehadiran Allah SWT yang telah melimpahkan rahmat-Nya sehingga penulis dapat menyelesaikan laporan tugas akhir yang berjudul “Klasifikasi *Human Action Recognition* (HAR) Berbasis MediaPipe dan *Long Short-Term Memory* (LSTM)”.

Laporan tugas akhir ini merupakan salah satu syarat untuk menyelesaikan studi di Departemen Teknik Komputer Fakultas Teknik Universitas Diponegoro. Selain itu, penyusunan laporan tugas akhir ini juga diharapkan dapat memberikan manfaat dalam bidang pendidikan.

Dalam penyusunan laporan tugas akhir ini, penulis banyak mendapatkan dukungan, doa, bimbingan, dan dorongan dari berbagai pihak. Oleh karena itu, penulis menyampaikan ucapan terima kasih kepada:

1. Bapak Dr. Adian Fatchur Rochim, S.T., M.T. selaku Ketua Departemen Teknik Komputer yang telah menyediakan lingkungan akademis yang kondusif bagi mahasiswa yang menulis tugas akhir.
2. Ike Pertiwi Windasari, S.T., M.T. selaku Koordinator Tugas Akhir yang telah memberikan persetujuan dan bimbingan dalam pelaksanaan kegiatan tugas akhir.
3. Ibu Dr. Oky Dwi Nurhayati S.T., M.T. selaku dosen pembimbing I yang telah memberikan bimbingan dan petunjuk dalam pembuatan tugas akhir.
4. Ibu Dania Eridani S.T., M.Eng. selaku dosen pembimbing II yang telah memberikan bimbingan dan petunjuk dalam pembuatan tugas akhir.
5. Bapak dan Ibu dosen Departemen Teknik Komputer yang telah memberikan materi perkuliahan berharga dalam kehidupan akademis penulis.
6. Bapak Samiran dan Ibu Ngatni selaku kedua orang tua yang telah memberikan dukungan mental, moral, spiritual, material, dan finansial kepada penulis.
7. Diana Nur Auliasari selaku kakak penulis yang telah memberikan saran berharga dalam penulisan tugas akhir.
8. Teman-teman angkatan 2018 selaku rekan seperjuangan yang telah menyediakan sarana diskusi serta memberikan saran yang berharga.

9. Kakak-kakak tingkat angkatan 2017, 2016, dan 2015 yang telah memberikan nasihat dalam perkuliahan dan kegiatan kampus lainnya.
10. Adik-adik tingkat angkatan 2019, 2020, dan 2021 yang selalu memberikan inspirasi dan semangat dalam aktif berkarya.
11. Serta semua pihak yang tidak dapat Penulis sebutkan satu-persatu yang telah membantu penyelesaian Tugas Akhir ini.

Penulis menyadari bahwa laporan tugas akhir ini masih memerlukan perbaikan di berbagai aspek. Oleh sebab itu, penulis mengharapkan kritik, saran, dan masukan dari pembaca. Akhir kata, penulis mengucapkan terima kasih dan semoga laporan tugas akhir ini dapat bermanfaat bagi orang banyak.

Semarang, 8 Juni 2022

A handwritten signature in black ink, appearing to read 'Ichsan Arsyi Putra', enclosed within a hand-drawn, irregular triangular or arrow-like shape pointing to the right.

Ichsan Arsyi Putra

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN.....	ii
HALAMAN PERNYATAAN ORISINALITAS.....	iii
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI.....	iv
KATA PENGANTAR	v
DAFTAR ISI.....	vii
DAFTAR GAMBAR	ix
DAFTAR TABEL.....	xi
ABSTRAK	xii
ABSTRACT.....	xiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan Tugas Akhir.....	2
1.4 Batasan Masalah.....	2
1.5 Manfaat Penelitian.....	3
1.6 Sistematika Penulisan.....	3
BAB II KAJIAN PUSTAKA	4
2.1 Penelitian Terdahulu.....	4
2.2 <i>Human Action Recognition (HAR)</i>	7
2.2.1 Representasi Aksi.....	7
2.2.2 Klasifikasi Aksi.....	7
2.2.3 <i>Deep Network</i>	8
2.3 Klasifikasi.....	9
2.4 <i>Deep Learning</i>	9
2.5 Pustaka MediaPipe	11
2.6 <i>Long Short-Term Memory</i>	12
2.7 <i>Dataset Weizmann</i>	14
2.8 <i>Data Preprocessing</i>	14

2.8.1	<i>Feature Selection</i>	14
2.8.2	<i>Data Labeling</i>	14
2.9	<i>Data Augmentation</i>	15
2.10	Model OSEMN.....	15
2.11	<i>Confusion Matrix</i>	16
BAB III METODE PENELITIAN.....		19
3.1	Spesifikasi Perangkat Penelitian	19
3.1.1	Spesifikasi Perangkat Keras	19
3.1.2	Spesifikasi Perangkat Lunak	19
3.2	Model OSEMN.....	19
3.2.1	<i>Obtain-1</i>	20
3.2.2	<i>Scrub-1</i>	22
3.2.3	<i>Explore-1</i>	24
3.2.4	<i>Obtain-2</i>	24
3.2.5	<i>Scrub-2</i>	25
3.2.6	<i>Explore-2</i>	25
3.2.7	<i>Model</i>	26
3.2.8	<i>Interpret</i>	31
BAB IV HASIL DAN PEMBAHASAN		33
4.1	Analisis dan Pembahasan	33
4.1.1	Analisis <i>Training</i> dan <i>Validation</i>	33
4.1.2	Evaluasi <i>Testing</i>	42
4.2	Implementasi Model pada Video Demo.....	50
BAB V PENUTUP.....		55
5.1	Kesimpulan.....	55
5.2	Saran	56
DAFTAR PUSTAKA		57
BIODATA MAHASISWA		63

DAFTAR GAMBAR

Gambar 2. 1 Ilustrasi hubungan antara <i>Artificial Intelligence</i> , <i>Machine Learning</i> , <i>Deep Learning</i> , dan <i>Deep Neural Network</i> [26]	10
Gambar 2. 2 Contoh struktur <i>Deep Neural Network</i> sederhana [27].....	10
Gambar 2. 3 Deteksi objek menggunakan MediaPipe [28]	11
Gambar 2. 4 <i>Pose Landmarks</i> [30]	12
Gambar 2. 5 Ilustrasi algoritma LSTM [33]	12
Gambar 2. 6 <i>Dataset Weizmann</i>	14
Gambar 2. 7 Perbandingan ekspektasi dan realitas penggunaan model OSEMN [39]	15
Gambar 3. 1 Alur kerja penelitian tugas akhir	20
Gambar 3. 2 Jumlah video per kelas pada <i>dataset Weizmann</i>	22
Gambar 3. 3 Satu detik awal (25 <i>frames</i>) pada video “ira_bend.avi” sebelum (kiri) dan sesudah pemotongan durasi (kanan)	22
Gambar 3. 4 Jumlah video per kelas pada <i>dataset Weizmann</i> hasil <i>data augmentation</i>	24
Gambar 3. 5 Ilustrasi ekstraksi ciri video menggunakan pustaka MediaPipe: <i>Pose</i>	25
Gambar 3. 6 Arsitektur model VA-LSTM-SYSU	28
Gambar 3. 7 Arsitektur model VA-LSTM-SBU.....	28
Gambar 3. 8 Arsitektur model LSTM-PASL.....	28
Gambar 3. 9 Arsitektur model LSTM-AHM	29
Gambar 3. 10 Arsitektur Model 1	30
Gambar 3. 11 Arsitektur Model 2	30
Gambar 3. 12 Arsitektur Model 3	31
Gambar 3. 13 Beberapa peragaan aksi pada video demo.....	31
Gambar 4. 1 Grafik <i>loss</i> (a) dan <i>accuracy</i> (b) pada proses <i>training</i> dan <i>validation</i> terhadap model VA-LSTM-SYSU.....	33
Gambar 4. 2 Grafik <i>loss</i> (a) dan <i>accuracy</i> (b) pada proses <i>training</i> dan <i>validation</i> terhadap model VA-LSTM-SBU	34

Gambar 4. 3 Grafik <i>loss</i> (a) dan <i>accuracy</i> (b) pada proses <i>training</i> dan <i>validation</i> terhadap model LSTM-PASL	36
Gambar 4. 4 Grafik <i>loss</i> (a) dan <i>accuracy</i> (b) pada proses <i>training</i> dan <i>validation</i> terhadap model LSTM-AHM.....	37
Gambar 4. 5 Grafik <i>loss</i> (a) dan <i>accuracy</i> (b) pada proses <i>training</i> dan <i>validation</i> terhadap Model 1.....	38
Gambar 4. 6 Grafik <i>loss</i> (a) dan <i>accuracy</i> (b) pada proses <i>training</i> dan <i>validation</i> terhadap Model 2.....	40
Gambar 4. 7 Grafik <i>loss</i> (a) dan <i>accuracy</i> (b) pada proses <i>training</i> dan <i>validation</i> terhadap Model 3.....	41
Gambar 4. 8 <i>Confusion matrix</i> hasil proses <i>testing</i> terhadap model VA-LSTM-SYSU	43
Gambar 4. 9 <i>Confusion matrix</i> hasil proses <i>testing</i> terhadap model VA-LSTM-SBU	44
Gambar 4. 10 <i>Confusion matrix</i> hasil proses <i>testing</i> terhadap model LSTM-PASL	45
Gambar 4. 11 <i>Confusion matrix</i> hasil proses <i>testing</i> terhadap model LSTM-AHM	46
Gambar 4. 12 <i>Confusion matrix</i> hasil proses <i>testing</i> terhadap Model 1	47
Gambar 4. 13 <i>Confusion matrix</i> hasil proses <i>testing</i> terhadap Model 2	48
Gambar 4. 14 <i>Confusion matrix</i> hasil proses <i>testing</i> terhadap Model 3	49
Gambar 4. 15 Deteksi <i>jack</i> oleh VA-LSTM-SBU (a) dan deteksi <i>wave1</i> oleh Model 2 (b).....	53
Gambar 4. 16 Deteksi <i>run</i> oleh Model 1 (a) dan <i>wave2</i> oleh Model 2 (b)	53
Gambar 4. 17 Hasil deteksi aksi <i>side</i> oleh LSTM-PASL (a) dan <i>jump</i> oleh Model 3 (b)	54

DAFTAR TABEL

Tabel 2. 1 Perbandingan metode dan hasil penelitian terdahulu.....	4
Tabel 2. 2 Jenis-jenis klasifikasi aksi.....	7
Tabel 2. 3 <i>Confusion matrix</i>	16
Tabel 2. 4 Rumus-rumus <i>confusion matrix</i> pada masalah klasifikasi banyak kelas	17
Tabel 3. 1 Jumlah video tiap kelas pada <i>dataset</i> Weizmann	20
Tabel 3. 2 Contoh hasil proses <i>data augmentation</i>	23
Tabel 3. 3 Nama kelas dan kode angka.....	25
Tabel 3. 4 Dimensi <i>array</i> pada data X	26
Tabel 3. 5 Dimensi <i>array</i> pada data y	26
Tabel 3. 6 Inisialisasi nilai <i>hyperparameter</i> yang sama pada ketiga model	29
Tabel 3. 7 Perubahan <i>hyperparameter</i> pada Model 1	30
Tabel 3. 8 Perubahan <i>hyperparameter</i> pada Model 2.....	30
Tabel 4. 1 Hasil nilai TP, FP, TN, dan FN dari proses <i>testing</i> model VA-LSTM- SYSU	43
Tabel 4. 2 Hasil nilai TP, FP, TN, dan FN dari proses <i>testing</i> model VA-LSTM- SBU.....	44
Tabel 4. 3 Hasil nilai TP, FP, TN, dan FN dari proses <i>testing</i> model LSTM-PASL	45
Tabel 4. 4 Hasil nilai TP, FP, TN, dan FN dari proses <i>testing</i> model LSTM-AHM	46
Tabel 4. 5 Hasil nilai TP, FP, TN, dan FN dari proses <i>testing</i> Model 1	47
Tabel 4. 6 Hasil nilai TP, FP, TN, dan FN dari proses <i>testing</i> Model 2	48
Tabel 4. 7 Hasil nilai TP, FP, TN, dan FN dari proses <i>testing</i> Model 3	49
Tabel 4. 8 Hasil pengukuran kualitas klasifikasi tiap model	50
Tabel 4. 9 Urutan perubahan deteksi aksi oleh tiap model	50

ABSTRAK

Human Action Recognition (HAR) merupakan topik yang sangat penting di bidang Machine Learning dan Computer Vision. Salah satu metode yang pernah diajukan adalah kombinasi penggunaan antara pustaka MediaPipe sebagai pengekstrak ciri citra dengan Long Short-Term Memory (LSTM) sebagai pengklasifikasi. Penelitian HAR dapat menjadikan akurasi deteksi dan waktu training sebagai indikator dalam menentukan model terbaik. Berdasarkan kondisi tersebut, peneliti mencoba mengadaptasi model LSTM yang pernah diajukan untuk pengaplikasian HAR dengan data hasil ekstraksi landmarks citra melalui pustaka MediaPipe. Peneliti juga membandingkan hasil akurasi dan waktu training pada model-model LSTM yang digunakan.

Penelitian dilaksanakan berdasarkan kerangka kerja OSEMN yang terdiri atas tahap Obtain, Scrub, Explore, Model, dan iNterpret. Dataset yang digunakan dalam penelitian adalah dataset Weizmann sehingga terdapat 10 kelas aksi dalam bentuk video. Sebelum digunakan, dataset Weizmann dikenakan proses data preprocessing dan data augmentation. Ekstraksi ciri dilakukan terhadap video menggunakan pustaka MediaPipe: Pose dan menghasilkan data landmarks. Proses training dan validation dilakukan dengan tiga variasi arsitektur model neural network yang berfokus pada layer Long Short-Term Memory (LSTM). Evaluasi performa model dilakukan melalui interpretasi confusion matrix dan perhitungan nilai accuracy, loss, precision, dan F₁score.

Hasil penelitian yang diperoleh berupa tujuh model LSTM dengan nilai akurasi testing mencapai 82% dengan waktu training selama 10 menit 50 detik. Dari ketujuh model, waktu training model tercepat diketahui memiliki durasi selama 1 menit 4 detik namun akurasi testing-nya hanya mencapai 77%.

Kata kunci: Klasifikasi, Deep Learning, Human Action Recognition, MediaPipe, Long Short-Term Memory

ABSTRACT

Human Action Recognition is an important research topic in Machine Learning and Computer Vision domains. One of the proposed methods is a combination of MediaPipe library as a features extractor along with Long Short-Term Memory as the classifier. Research in HAR could refer to the testing accuracy and training duration as indicators to evaluate the model performance. Based on that circumstance, this research tried to adapt proposed LSTM models to implement HAR with image features extracted by MediaPipe library. Furthermore, there would be a comparison between LSTM models based on their testing accuracy and training duration.

This research was conducted under OSEMN methods consisting of five main processes namely Obtain, Scrub, Explore, Model, and iNterpret. The dataset used in this research was preprocessed Weizmann dataset with data preprocessing and data augmentation implementations. The video feature extracting process was done by MediaPipe: Pose library. Training and validation processes were conducted on three variants of the neural network model focusing on LSTM layers. The processes were finished by model performance evaluation based on confusion matrices interpretation and calculations of accuracy, error rate, precision, recall, and F₁score.

This research yielded seven LSTM model variants with the highest testing accuracy at 82% taking 10 minutes and 50 seconds of training duration. We also found that the fastest training duration was 1 minute and 4 seconds yet its testing accuracy only reached 77%.

Keywords: *Classification, Deep Learning, Human Action Recognition, MediaPipe, Long Short-Term Memory*