

**DETEKSI TINGKAT *ENGAGEMENT* SISWA PADA
PEMBELAJARAN DARING BERDASARKAN
EMOSI WAJAH DAN SIKAP SISWA**

TESIS

**Karya tulis sebagai salah satu syarat
untuk memperoleh gelar Magister dari
Institut Teknologi Bandung**

**Oleh
ARI APRIANSYAH
NIM: 23523301
(Program Studi Magister Informatika)**



**INSTITUT TEKNOLOGI BANDUNG
Maret 2025**

**DETEKSI TINGKAT ENGAGEMENT SISWA PADA
PEMBELAJARAN DARING BERDASARKAN
EMOSI WAJAH DAN SIKAP SISWA**

Oleh
Ari Apriansyah
NIM: 23523301
(Program Studi Magister Informatika)

Institut Teknologi Bandung

Menyetujui
Tim Pembimbing

Tanggal 20 November 2025

Ketua

Anggota

Dr. Ir. Kusprasapta Mutijarsa, S.T., M.T.

Dr. Fadhil Hidayat, S.Kom., M.T.

DAFTAR ISI

DAFTAR ISI.....	ii
DAFTAR LAMPIRAN.....	iv
DAFTAR GAMBAR DAN ILUSTRASI.....	v
DAFTAR TABEL.....	vii
DAFTAR SINGKATAN DAN LAMBANG	viii
BAB I Pendahuluan.....	1
I.1 Latar Belakang	1
I.2 Rumusan Masalah.....	5
I.3 Tujuan Penelitian	6
I.4 Hipotesis	6
I.5 Batasan Penelitian.....	7
I.6 Kontribusi Penelitian	7
I.7 Sistematika Penulisan	8
BAB II Tinjauan Pustaka.....	10
II.1 <i>Learning Management System</i>	10
II.2 <i>Affective Computing</i>	11
II.3 <i>Automated Emotion Recognition</i> di Lingkungan Pendidikan.....	11
II.3.1 Modalitas Teks.....	13
II.3.2 Modalitas Audio.....	16
II.3.3 Modalitas Sinyal Tubuh.....	19
II.3.4 Modalitas Visual	22
II.3.5 Multimodal.....	26
II.4 <i>Facial Expressions Recognition</i> di Lingkungan Pendidikan Daring.....	27
II.5 Oklusi pada Pengenalan Ekspresi Wajah.....	29
II.6 Klasifikasi Emosi dalam Deteksi <i>Engagement</i> Siswa	31
II.6.1 <i>Machine Learning</i> (ML)	32
II.6.2 <i>Deep Learning</i> (DL)	32
II.6.3 Kombinasi <i>Machine Learning</i> Klasik dan <i>Deep Learning</i>	34
II.7 Peta Literatur.....	35
BAB III Metodologi Penelitian.....	36
III.1 Identifikasi Masalah dan Motivasi.....	36
III.2 Definisi Tujuan dan Solusi	37
III.3 Perancangan dan Pengembangan.....	38
III.4 Demonstrasi	39
III.5 Evaluasi	39
III.6 Komunikasi.....	40
BAB IV Analisis dan Desain Awal.....	41
IV.1 Lingkungan Penerapan	41
IV.2 Analisis Kebutuhan Sistem.....	41
IV.2.1 Kebutuhan Fungsional	42
IV.2.2 Kebutuhan Non-Fungsional	43
IV.2.3 Kebutuhan Pengguna	43
IV.2.4 Kebutuhan Perangkat Keras dan Perangkat Lunak.....	44
IV.3 Analisis Desain	44
IV.4 Analisis Wajah.....	45

IV.5	Desain Solusi	46
IV.5.1	Desain Solusi Umum	46
IV.5.2	Desain Algoritma Program	47
IV.6	Desain Pengujian	50
IV.7	Pengembangan	52
IV.7.1	Perangkat Keras	52
IV.7.2	Perangkat Lunak	52
IV.7.2.1	Desain Halaman <i>Login</i>	53
IV.7.2.2	Desain Halaman <i>Asynchronous Learning</i>	53
IV.7.2.3	Desain Halaman Analisis Emosi.....	54
IV.7.3	Pengembangan Metode Usulan	54
BAB V	Hasil dan Evaluasi	56
V.1	Hasil Metode Usulan	56
V.1.1	Hasil pada Wajah yang Tertutup Oklusi Tangan (Vertikal)	57
V.1.2	Hasil pada Wajah dengan Kondisi Miring (<i>Non-Frontal</i>).....	62
V.1.3	Hasil pada Wajah yang Tertutup Oklusi Tangan (Horizontal)	66
V.1.4	Hasil Komparasi terhadap Metode Frontalisasi Lain	68
V.2	Pembahasan	69
V.2.1	Performa <i>Face Detector</i> pada Model.....	69
V.2.2	Wajah yang Teroklusi oleh Tangan Sebagian	70
V.2.3	Wajah dengan Kondisi Miring (<i>Non-Frontal</i>).....	73
V.2.4	Oklusi Parsial pada Area Mulut.....	75
V.3	Tampilan Antarmuka	76
V.4	Implementasi FER pada LMS	79
BAB VI	Kesimpulan	81
DAFTAR PUSTAKA	83
LAMPIRAN	91

DAFTAR LAMPIRAN

Lampiran A *Source code* fungsi utama metode usulan (*Half-Flip 2D*)..... 91

DAFTAR GAMBAR DAN ILUSTRASI

Gambar I.1	Contoh tangkapan layar wajah mahasiswa saat pembelajaran daring dengan berbagai kondisi dan wajah yang teroklusi oleh tangan	5
Gambar II.1	Siklus penerapan AER pada lingkungan pendidikan (Yu dkk., 2024)	12
Gambar II.2	Deteksi <i>engagement</i> siswa menggunakan FER, <i>eye blink</i> , dan <i>head pose</i> (Gupta dkk., 2023a)	24
Gambar II.3	Alur pengenalan emosi dengan FER digabungkan dengan analisis sentimen (Cui dkk., 2021)	26
Gambar II.5	Proses utama <i>facial expressions recognition</i>	28
Gambar II.6	(a) Keadaan wajah dengan oklusi <i>partial</i> oleh tangan, (b) keadaan wajah dengan oklusi total (Jain & Learned-Miller, 2010)	30
Gambar II.7	Gambaran umum alur <i>machine learning</i>	32
Gambar II.8	Distribusi penggunaan metode <i>machine learning</i> dan <i>deep learning</i> dalam rentang tahun 2010 – 2024 (Yu dkk., 2024)	34
Gambar II.9	Peta literatur deteksi emosi siswa pada media pembelajaran daring	35
Gambar II.10	Peta literatur tentang cara mengatasi masalah oklusi pada FER ...	35
Gambar III.1	<i>Design Science Research Methodology</i>	36
Gambar IV.1	Alur penerapan mekanisme umpan balik pada LMS	41
Gambar IV.2	Contoh hasil analisis wajah dengan paramater arah pandangan mata, arah kepala, dan emosi wajah.....	45
Gambar IV.3	Contoh kondisi wajah yang arah kepala dan mata tidak menghadap monitor atau menghadap kesamping (gambar dibuat dengan AI)	46
Gambar IV.4	Gambaran umum ajuan solusi sistem.....	46
Gambar IV.5	Detail teknis dan fokus penelitian pada bagian <i>pre-processing</i>	47
Gambar IV.6	Diagram alir algoritma kondisi penggunaan FER untuk menentukan <i>engagement</i> siswa.....	47
Gambar IV.7	Skema metode usulan dalam penanganan masalah oklusi dan wajah <i>non-frontal</i> pada wajah.....	49
Gambar IV.8	Proses penambahan tangan sintesis, (a) gambar asli, (b) tangan sintesis, (c) hasil augmentasi	51
Gambar IV.9	Data wajah untuk pengujian, (a) wajah dengan oklusi tangan, (b) wajah non-frontal	52
Gambar IV.10	Rancangan desain halaman <i>login</i>	53
Gambar IV.11	Rancangan halaman <i>asynchronous learning</i>	53
Gambar IV.12	Rancangan halaman analisis emosi.....	54
Gambar IV.13	Langkah pada metode usulan.....	54
Gambar V.1	Hasil percobaan metode usulan, (a) data wajah dengan oklusi tangan, (b) wajah dalam keadaan non-frontal	56
Gambar V.2	Perbandingan probabilitas emosi pada gambar sebelum dan sesudah diterapkan metode usulan	57
Gambar V.3	Visualisasi perbandingan akurasi data sebelum dan sesudah penerapan metode usulan pada data wajah dengan oklusi tangan menggunakan model CNN	58

Gambar V.4	Visualisasi perbandingan akurasi data sebelum dan sesudah penerapan metode usulan pada data wajah dengan oklusi tangan menggunakan model DeepFace.....	59
Gambar V.5	Visualisasi perbandingan akurasi data sebelum dan sesudah penerapan metode usulan pada data wajah dengan oklusi tangan menggunakan model FaceLib	60
Gambar V.6	Visualisasi perbandingan akurasi data sebelum dan sesudah penerapan metode usulan pada data wajah dengan oklusi tangan menggunakan model SVM.....	61
Gambar V.7	Hasil akurasi data wajah <i>non-frontal</i> dengan CNN.....	63
Gambar V.8	Hasil akurasi data wajah <i>non-frontal</i> dengan DeepFace	64
Gambar V.9	Hasil akurasi data wajah <i>non-frontal</i> dengan FaceLib.....	65
Gambar V.10	Hasil akurasi data wajah <i>non-frontal</i> dengan SVM	66
Gambar V.11	Kondisi wajah tertutup tangan pada bagian bawah (mulut)	66
Gambar V.12	Hasil penerapan metode usulan terhadap wajah dengan oklusi tangan pada bagian bawah wajah.....	67
Gambar V.13	Metode frontalisasi wajah 2D oleh Tsai dkk., (2021)	68
Gambar V.14	Hasil komparasi, (A) gambar <i>input</i> , (B) metode usulan, (C) metode oleh Tsai dkk., (2021).....	68
Gambar V.15	Hasil metode usulan, (A) kondisi berhasil, (B) kondisi gagal	72
Gambar V.16	Arsitektur DeepFace (Taigman dkk., 2014).....	73
Gambar V.17	Halaman <i>login</i>	76
Gambar V.18	Halaman <i>dashboard</i>	77
Gambar V.19	Halaman akses materi pembelajaran <i>asynchronous</i>	77
Gambar V.20	Halaman hasil ringkasan <i>engagement</i> siswa	78
Gambar V.21	Keterangan hasil analisis wajah siswa per <i>frame</i> , (a) kondisi <i>highly engaged</i> , wajah lurus ke layar tanpa halangan, (b) kondisi <i>highly engaged</i> , wajah lurus ke layar dengan halangan tangan, (c) kondisi <i>boredom</i> , wajah tidak menghadap ke layar, (d) kondisi <i>sleepy</i> , mata tertutup, (e) kondisi <i>boredom</i> kepala ditopang tangan	79

DAFTAR TABEL

Tabel II.1	Studi terkait yang menggunakan metode <i>deep learning</i> pada deteksi emosi dengan modalitas visual pada lingkungan pendidikan	33
Tabel III.1	Tahap identifikasi masalah dan motivasi	36
Tabel III.2	Tahap tujuan dan solusi	38
Tabel III.3	Tahap perancangan dan pengembangan	38
Tabel III.5	Tahapan demonstrasi	39
Tabel III.6	Tahap evaluasi	39
Tabel III.7	Tahap komunikasi	40
Tabel IV.1	Alternatif solusi untuk deteksi emosi siswa	44
Tabel V.1	Hasil akurasi data wajah dengan oklusi tangan dengan CNN	58
Tabel V.2	Hasil akurasi data wajah dengan oklusi tangan dengan DeepFace	59
Tabel V.3	Hasil akurasi data wajah dengan oklusi tangan dengan FaceLib ..	60
Tabel V.4	Hasil akurasi data wajah dengan oklusi tangan dengan SVM.....	61
Tabel V.5	Hasil akurasi data wajah miring (<i>non-frontal</i>) dengan CNN	62
Tabel V.6	Hasil akurasi data wajah miring (<i>non-frontal</i>) dengan DeepFace.	63
Tabel V.7	Hasil akurasi data wajah miring (<i>non-frontal</i>) dengan FaceLib....	64
Tabel V.8	Hasil akurasi data wajah miring (<i>non-frontal</i>) dengan SVM	65
Tabel V.9	Jumlah data uji wajah dengan oklusi tangan (horizontal)	66
Tabel V.10	Performa model dalam mendeteksi wajah pada wajah yang teroklusi tangan pada bagian mulut	67
Tabel V.11	Hasil pengujian metode usulan pada wajah yang teroklusi tangan pada bagian mulut	68
Tabel V.12	Hasil komparasi metode frontalisasi	69

DAFTAR SINGKATAN DAN LAMBANG

SINGKATAN	Nama	Pemakaian pertama kali pada halaman
AER	<i>Automated Emotion Recognition</i>	11
BVP	<i>Blood Volume Pressure</i>	20
CMU	<i>Carnegie Mellon University</i>	17
CNN	<i>Convolutional Neural Network</i>	3
CNN-RF	<i>Convolutional Neural Network - Random Forest</i>	34
COVID-19	<i>Corona Virus Disease 2019</i>	1
DL	<i>Deep Learning</i>	4
DSRM	<i>Design Science Researcr Methodology</i>	8
ECG	<i>Elektrokardiography</i>	20
EDA	<i>Electrodermal Activity</i>	19
EEG	<i>Electroencephalography</i>	16
EKG	<i>Electrocardiography</i>	19
EMG	<i>Elektromiography</i>	20
FER	<i>Facial Expressions Recognition</i>	3
GSR	<i>Galvanic Skin Response</i>	20
IB-BiLSTM	<i>Information Block Bidirectional LSTM</i>	14
IEMOCAP	<i>Interactive Emotional Dyadic Motion Capture</i>	18
IoT	<i>Internet of Things</i>	3
KDD	<i>Knowledge Discovery in Databases</i>	35
LBP	<i>Local Binary Pattern</i>	22
LMS	<i>Learning Management System</i>	1
LSTM	<i>Long Short Term Memory</i>	13
MER	<i>Multimodal Emotion Recognition</i>	26
MFCC	<i>Mel-Frequency Cepstral Coefficients</i>	17
ML	<i>Machine Learning</i>	15
MOOC	<i>Massive Online Open Course</i>	2
NLP	<i>Natural Language Processing</i>	27
PANAS	<i>Positive and Negative Affect Schedule</i>	15
PPG	<i>Photoplethysmography</i>	20
RSP	<i>Respiration</i>	20
SAM	<i>Self-Assessment Manikin</i>	15
SER	<i>Speech Emotion Recognition</i>	16
SKT	<i>Skin Temperature</i>	20
SVM	<i>Support Vector Machine</i>	14
TTS	<i>Text-To-Speech</i>	17
WHO	<i>World Health Organization</i>	1

LAMBANG

BAB I Pendahuluan

Bab ini membahas berbagai dasar dan pedoman yang menjadi fokus utama penelitian. Tujuannya adalah untuk memberikan pemahaman menyeluruh mengenai penelitian serta urgensinya dalam mencapai tujuan yang telah ditetapkan.

I.1 Latar Belakang

Transformasi digital dalam dunia pendidikan menjadi langkah penting bagi institusi pendidikan tinggi untuk beradaptasi dengan era revolusi industri 4.0. Salah satu arah pengembangan yang banyak diambil oleh berbagai universitas dan lembaga pendidikan adalah penerapan konsep *smart campus*, dimana teknologi informasi digunakan untuk meningkatkan efisiensi, mutu, dan fleksibilitas pembelajaran. Pandemi COVID-19 menjadi pemicu utama bagi percepatan transformasi ini, dimana kejadian tersebut memaksa sistem pendidikan di seluruh dunia beralih secara mendadak ke metode pembelajaran daring. Perubahan ini menghadirkan berbagai tantangan, terutama dalam menjaga kualitas interaksi antara pengajar dan peserta didik.

Meskipun pandemi telah berakhir dan kegiatan belajar mengajar tatap muka kembali dilakukan, pembelajaran daring tetap menjadi bagian dari sistem pendidikan modern, khususnya dalam bentuk *asynchronous learning* yang merupakan metode pembelajaran daring dimana peserta dan pengajar tidak harus hadir atau berinteraksi pada waktu yang sama dan biasa menggunakan suatu sistem *Learning Management System* (LMS). Melalui LMS, materi pembelajaran dapat disampaikan dalam bentuk dokumen maupun video sehingga peserta didik memiliki fleksibilitas dalam menentukan waktu belajar. Namun demikian, sistem *asynchronous learning* juga menimbulkan permasalahan lain yaitu kurangnya interaksi langsung antara pengajar dan peserta didik. Akibatnya, pengajar sulit menilai tingkat fokus, keseriusan, dan *engagement* peserta didik selama mengakses materi pembelajaran (Peterson, 2023).

Untuk mengatasi keterbatasan tersebut, diperlukan mekanisme umpan balik berbasis teknologi yang mampu menyimpulkan tingkat *engagement* peserta didik dengan memanfaatkan data emosi dalam berbagai bentuk seperti teks, suara, atau

visual (Yu dkk., 2024). Salah satu pendekatan yang banyak digunakan adalah analisis sentimen berbasis teks, di mana emosionalitas peserta didik dapat diukur melalui komentar atau tanggapan dalam forum pembelajaran daring (Xu, 2023). Beberapa penelitian tentang penerapan analisis sentimen untuk mengetahui emosi dari siswa telah dilakukan oleh beberapa peneliti. Peneliti X. Zhang & Qin, (2022) membuat sistem yang menerapkan analisis sentimen untuk mengevaluasi emosi siswa pada pelajaran Ideologi dan Politik dalam *platform* daring. Pendekatan analisis sentimen juga telah dilakukan oleh Mao dkk., (2023) untuk mengevaluasi tingkat emosional siswa dalam pelajaran *Internet of Things* (IoT). Walaupun demikian, pendekatan analisis sentimen dinilai kurang tepat untuk dijadikan aspek penilaian utama untuk mendapatkan data emosi siswa, karena dimungkinkan untuk terjadinya emosi palsu dari siswa dalam tulisannya (Khoshnam & Baraani-Dastjerdi, 2022).

Bentuk mekanisme umpan balik yang lainnya adalah berbasis *audio*. Berdasarkan referensi penelitian lampau, Bao dkk., (2024) menggabungkan fitur *audio* dan gerakan mata untuk mengevaluasi emosi siswa pada *massive daring open course* (MOOC). Mereka merancang sistem yang dapat mengklasifikasikan emosi seperti minat, kebahagiaan, kebingungan, dan kebosanan dengan hasil akurasi sebesar 81,90%. Penelitian lain yang menggunakan fitur *audio* dalam mendeteksi emosi siswa pada pembelajaran daring adalah Bekmanova dkk., (2022), didasari oleh pembelajaran jarak jauh karena COVID-19, mereka membangun sistem yang dapat mendeteksi emosi positif dan negatif dari ucapan lisan siswa dengan bahasa Kazakh dengan hasil akurasi sebesar 79,7%. Berdasarkan hasil penelitian – penelitian di atas, membuktikan bahwa audio dapat digunakan sebagai basis penilaian untuk memberikan masukan dari siswa kepada pengajar berupa data emosi siswa. Walaupun demikian, penggunaan *audio* dalam deteksi emosi siswa pada pembelajaran daring memiliki sejumlah masalah, seperti penggunaan fitur *audio* yang menyebabkan penurunan akurasi jika digabungkan dengan fitur lain (Pan dkk., 2024), lalu adanya batasan kuantitas terhadap *database* yang akan digunakan sebagai data latih (Zhu-Zhou dkk., 2022), hingga mahalnya biaya komputasi pengolahan sinyal *audio* (Z. T. Liu dkk., 2021).

Pendekatan lain yang semakin populer adalah berbasis visual. Memanfaatkan teknologi *computer vision*, banyak penelitian yang menggunakan teknologi *facial expressions recognition* (FER) untuk mendapatkan data emosi dari pengamatan ekspresi wajah siswa pada lingkungan pembelajaran daring. Bhardwaj dkk., (2021) merancang sistem untuk menghitung tingkat *engagement* siswa dengan menerapkan FER selama kelas daring dan mendapatkan hasil akurasi sebesar 93,6% dalam klasifikasi emosinya menggunakan algoritma *Convolutional Neural Network* (CNN). Penerapan FER dalam lingkungan pendidikan daring lainnya juga pernah dilakukan oleh Gupta dkk., (2023). Mereka membangun sistem FER yang dapat mendeteksi tingkat *engagement real-time* dengan menganalisis ekspresi wajah, lalu mengklasifikasikan siswa menjadi “Terlibat” atau “Tidak Terlibat” berdasarkan keadaan emosinya. Algoritma yang mereka gunakan untuk klasifikasi emosi siswa adalah algoritma ResNet-50 dan mendapatkan hasil mencapai akurasi 92,3% dalam skenario *real-time*. Berdasarkan hal tersebut, dibandingkan dengan solusi alternatif seperti analisis sentimen dan pengenalan suara, penerapan FER pada pembelajaran daring baik *synchronous* maupun *asynchronous* bisa dijadikan sebagai solusi yang paling logis untuk menyelesaikan masalah kebutuhan mekanisme umpan balik dari siswa pada lingkungan pembelajaran daring baik dari sisi batasan-batasan tiap solusi yang ada maupun dari biaya dan beban komputasi.

Walaupun demikian, penerapan FER pada pembelajaran daring juga mempunyai beberapa kekurangan, terutama dalam keadaan oklusi, seperti wajah yang tertutup oleh tangan, kacamata, atau objek lain (Li dan Deng, 2022). Kondisi ini dapat mengurangi akurasi dalam mendeteksi ekspresi secara tepat, sehingga sistem mungkin menghasilkan prediksi yang kurang akurat atau bahkan keliru dalam menginterpretasikan emosi pengguna (Mensah dkk., 2024). Selain itu, variasi pencahayaan dan sudut pandang kamera yang tidak seragam juga dapat memengaruhi performa model dalam mengenali ekspresi wajah secara konsisten.

Beberapa penelitian yang telah lampau mencoba untuk mengatasi masalah oklusi pada penerapan FER pada lingkungan nyata. Ada beberapa cara untuk mengatasi oklusi pada penerapan FER yang telah dilakukan peneliti, antara lain dengan menerapkan teknik pemrosesan citra seperti merekonstruksi bagian wajah yang

terhalang (Poux dkk., 2022), menggunakan model *deep learning* (DL) yang dilatih dengan data augmentasi oklusi agar lebih *robust* terhadap data oklusi (X. Zhang dkk., 2022), serta memanfaatkan multi-modalitas, seperti kombinasi informasi dari gambar dan data lainnya, misalnya gabungan data wajah dan suara (Freire-obregón dan Castrillón-santana, 2024 dan Mamieva dkk., 2023). Meskipun berbagai metode telah dikembangkan untuk mengatasi oklusi pada penerapan FER, solusi-solusi di atas merupakan teknik yang memerlukan komputasi yang kompleks, sehingga mengakibatkan kenaikan latensi pada prosesnya.

Alternatif lain yang lebih ringan adalah perbaikan pada tahap pra-pemrosesan data. Cao dkk., (2021) mengusulkan metode berbasis penggabungan setengah wajah dalam bentuk 3D untuk memperbaiki pengenalan emosi pada wajah yang tertutup sebagian. Metode ini terbukti meningkatkan akurasi, tetapi teknik berbasis 3D memerlukan sumber daya komputasi tinggi (Morar dkk., 2017). Untuk mengatasi keterbatasan ini, penelitian ini akan mengajukan metode pra-pemrosesan baru yang disebut Half-Flip 2D, yang memanfaatkan setengah wajah yang tidak teroklusi dengan cara *mirroring* untuk merekonstruksi wajah penuh dalam format dua dimensi. Pendekatan ini diharapkan mampu mempertahankan ketepatan deteksi emosi seperti pengolahan 3D, namun dengan kebutuhan komputasi yang jauh lebih rendah

Berdasarkan uraian di atas, pada penelitian ini akan dikembangkan sebuah aplikasi media pembelajaran daring dengan fitur yang dapat merepresentasikan LMS sesungguhnya dengan mengimplementasikan metode usulan dalam suatu sistem yang dapat menilai tingkat *engagement* siswa berdasarkan hasil dari FER dan sikap siswa. Sikap siswa yang akan dinilai adalah berdasarkan arah kepala dan arah mata siswa. Adapun bentuk materi yang digunakan dalam aplikasi media pembelajaran yang akan dibuat, terbatas pada video rekaman paparan perkuliahan daring yang akan diakses oleh siswa dan dimanfaatkan oleh sistem untuk mendeteksi emosi yang tampak pada wajah siswa guna menilai tingkat *engagement* mereka saat melakukan pembelajaran daring secara *asynchronous*. Data penilaian tingkat *engagement* siswa yang dihasilkan oleh aplikasi diharapkan bisa menjadi data yang

bisa digunakan sebagai umpan balik sekaligus solusi atas kelemahan pada suatu media pembelajaran daring khususnya pada *asynchronous learning*.



Gambar I.1 Contoh tangkapan layar wajah mahasiswa saat pembelajaran daring dengan berbagai kondisi dan wajah yang teroklusi oleh tangan

Fokus penelitian ini adalah pada pengembangan metode *pre-processing* data wajah yang mengalami oklusi, khususnya ketika wajah tertutup oleh tangan sebagaimana ditunjukkan pada Gambar I.1. Kondisi ini sering kali terjadi dalam konteks pembelajaran daring, dimana siswa secara alami menutupi sebagian wajah mereka karena berbagai alasan, seperti keluhan kesehatan pada leher atau kepala (Beekman dkk., 2002), maupun akibat faktor psikologis seperti upaya untuk lebih fokus atau merasa bosan. Dengan demikian, metode *pre-processing* yang diusulkan bertujuan untuk meningkatkan probabilitas emosi pada wajah yang tidak menghadap lurus ke depan maupun teroklusi oleh tangan agar tetap dapat diklasifikasikan secara optimal dalam sistem FER selama pembelajaran daring berlangsung.

I.2 Rumusan Masalah

Berdasarkan latar belakang di atas, bahwa penerapan sistem FER pada LMS untuk mengetahui tingkat kefokusan siswa melalui modalitas visual masih memiliki kekurangan terutama pada pembelajaran *asynchronous*, dan juga adanya masalah penurunan akurasi FER pada kondisi wajah yang teroklusi tangan, maka dirumuskan suatu masalah pada penelitian ini, yaitu :

1. Bagaimana meyimpulkan tingkat *engagement* siswa berdasarkan gabungan modalitas FER, arah pandangan mata, dan arah kepala?
2. Apakah *pre-processing* gambar 2D yang memanfaatkan data setengah wajah dapat meningkatkan probabilitas emosi wajah siswa yang teroklusi ?

I.3 Tujuan Penelitian

Tujuan penelitian ini dibagi menjadi tujuan umum dan tujuan khusus. Tujuan umum penelitian ini adalah membuat aplikasi media pembelajaran daring dengan menerapkan fitur umpan balik menggunakan modalitas FER, arah pandangan, dan arah kepala untuk mendapatkan data tingkat *engagement* dari siswa saat melakukan pembelajaran daring. Adapun tujuan khusus dari penelitian ini adalah sebagai berikut:

1. Aplikasi media pembelajaran daring dapat meyimpulkan tingkat *engagement* siswa menggunakan *multi-modal* yaitu FER, arah pandangan mata, dan arah kepala.
2. Melakukan pengukuran efek *pre-processing* gambar 2D yang memanfaatkan data setengah wajah terhadap probabilitas emosi wajah siswa yang teroklusi.

I.4 Hipotesis

Wajah yang tertutup sebagian oleh tangan dan keadaan wajah miring dapat menurunkan akurasi sistem FER. Hal ini terjadi karena area-area penting pada wajah, seperti mata, mulut, dan alis menjadi tertutup sehingga informasi visual yang dibutuhkan untuk mengenali emosi tidak dapat terekstraksi secara optimal (Ruan dkk., 2022a). Selain itu, pose kepala yang miring (tidak menghadap ke depan) juga mengakibatkan menurunnya akurasi FER karena terjadi *self-occlusion* dan perubahan tampilan geometri fitur wajah pada gambar, sehingga sistem sulit mengenali emosi secara akurat (Engin dkk., 2018).

Penghapusan objek oklusi seperti tangan dan penerapan *frontalization* pada wajah miring dapat memperbaiki akurasi FER karena fitur-fitur wajah penting yang awalnya tertutup atau hilang bisa dikembalikan sehingga model FER dapat mengekstraksi fitur wajah dengan lengkap dan menghasilkan kesimpulan emosi yang tepat.

I.5 Batasan Penelitian

Adanya keterbatasan dalam hal waktu, tenaga dan biaya, dirasa perlu untuk menetapkan batasan dalam penelitian ini, Hal ini penting untuk dilakukan agar penelitian tetap fokus pada permasalahan yang diangkat. Berikut batasan-batasan yang ditetapkan untuk penelitian ini.

1. Media pembelajaran daring yang akan dibuat mengacu pada LMS dengan fitur yang berfokus pada pembelajaran *asynchronous*.
2. Penerapan sistem FER hanya pada pembelajaran *asynchronous* yang berbentuk video rekaman paparan perkuliahan.
3. Bentuk oklusi yang diteliti hanya pada kondisi sebagian wajah tertutup tangan dan wajah dalam keadaan miring (tidak menghadap ke depan), sebagaimana penjelasan di latar belakang yang menyebutkan bahwa keadaan sebagian wajah tertutup tangan lazim dilakukan oleh siswa dalam menjalani pembelajaran secara daring baik karena masalah kesehatan leher atau kepala (Beekman dkk., 2002).

I.6 Kontribusi Penelitian

Penelitian ini diharapkan dapat memberikan kontribusi baik terhadap penyelesaian masalah maupun dalam bidang ilmu pengetahuan. Adapun kontribusi yang diharapkan adalah:

1. Menghasilkan sebuah konsep mekanisme umpan balik pada pembelajaran daring berdasarkan tingkat *engagement* siswa menggunakan data emosi wajah dan sikap siswa.
2. Menghasilkan temuan ilmiah baru terkait penerapan metode pre-processing pada gambar wajah yang teroklusi menggunakan data setengah wajah dalam format 2D untuk mempertahankan akurasi dalam pengenalan emosi wajah pada kondisi teroklusi tangan dan keadaan wajah miring, dengan fokus pada pendekatan yang dapat dijalankan pada sistem dengan sumber daya komputasi terbatas.

I.7 Sistematika Penulisan

Sistematika penulisan tesis akan dibagi menjadi lima bab yang diawali dengan penjabaran umum di setiap rinciannya.

Bab I Pendahuluan

Memberikan penjelasan mengenai latar belakang dilakukannya penelitian. Dari latar belakang yang telah dijabarkan, selanjutnya disusun suatu rumusan masalah, tujuan penelitian, batasan penelitian, kebaruan dan kontribusi penelitian serta sistematika penulisan.

Bab II Tinjauan Pustaka

Menjelaskan mengenai gambaran tinjauan literatur dan kajian ilmu yang digunakan untuk mendukung suatu penelitian. Tinjauan pustaka ini selanjutnya dimanfaatkan dalam mengidentifikasi suatu permasalahan, konsep, dan ilmu pengetahuan untuk mendapatkan solusi dari penelitian yang akan dilakukan.

Bab III Metodologi Penelitian

Pembahasan terkait dengan metodologi yang akan digunakan di dalam penelitian. *Design Science Research Methodology* (DSRM) dipilih sebagai metode dalam penulisan tesis ini. Terdapat enam tahapan utama di dalam DSRM, yaitu identifikasi masalah & motivasi, menentukan tujuan solusi, perancangan & pengembangan, demonstrasi, evaluasi, dan komunikasi.

Bab IV Analisis dan Desain

Menjelaskan hasil analisis terhadap beberapa alternatif solusi beserta alasan pemilihan solusi yang dianggap paling sesuai dengan kebutuhan sistem. Selain itu, bab ini juga memaparkan rancangan dari solusi terpilih, meliputi arsitektur, alur proses, dan komponen utama yang akan digunakan dalam pengembangan sistem.

Bab V Hasil

menyajikan hasil implementasi dari desain yang telah dibuat pada bab sebelumnya. Pada bagian ini dijelaskan kinerja sistem yang telah dibangun, hasil pengujian yang dilakukan, serta evaluasi terhadap pencapaian tujuan yang telah ditetapkan. Bab ini juga menampilkan tanggapan atau analisis terhadap hasil yang diperoleh untuk

memastikan bahwa sistem berfungsi sesuai dengan kebutuhan dan spesifikasi yang direncanakan.

Bab VI Kesimpulan dan Saran

Memberikan penjelasan mengenai kesimpulan dari penelitian yang telah dilakukan. Selain itu, juga akan disampaikan mengenai beberapa saran yang bisa menjadi pertimbangan untuk dilakukannya penelitian lanjutan.

BAB II Tinjauan Pustaka

Bab ini mencakup tinjauan pustaka yang berperan sebagai acuan penelitian dalam mencapai tujuan penelitian yang telah didefinisikan. Tinjauan pustaka dilakukan terhadap referensi eksternal yang kredibel dan sesuai dengan kebutuhan penggerjaan tesis.

II.1 *Learning Management System*

Learning Management System (LMS) adalah sebuah platform perangkat lunak yang dirancang untuk mengelola, mendistribusikan, dan melacak aktivitas pembelajaran secara daring. LMS banyak digunakan di institusi pendidikan, perusahaan, dan organisasi untuk menyelenggarakan pelatihan serta kursus secara efisien. Dengan LMS, pengguna dapat mengakses materi pembelajaran, mengikuti kuis, mengerjakan tugas, dan berinteraksi dengan pemateri atau sesama peserta secara fleksibel dari mana saja dan kapan saja.

Manfaat dari *Learning Management System* (LMS) sangat beragam dan memberikan dampak positif bagi proses pembelajaran. Pertama, LMS dapat meningkatkan *engagement* siswa serta hasil pembelajaran, karena memungkinkan pengalaman belajar yang lebih interaktif dan terstruktur (Alotaibi, 2024). Selain itu, sistem ini juga berperan dalam meningkatkan aksesibilitas dan kesetaraan pendidikan, memungkinkan siswa dari berbagai latar belakang untuk memperoleh materi pembelajaran dengan lebih mudah (Alotaibi, 2024).

Selain manfaat tersebut, LMS juga mendukung manajemen dan penilaian kursus yang lebih efisien, membantu pendidik dalam mengorganisir materi, memberikan tugas, serta mengevaluasi kemajuan siswa secara sistematis (Tran dan Meacheam, 2020). Lebih jauh, penggunaan LMS memiliki dampak positif terhadap kepuasan siswa, karena menyediakan pengalaman belajar yang fleksibel, mudah diakses, dan sesuai dengan kebutuhan individu (Zaareer dkk., 2024). Terakhir, LMS dapat memfasilitasi interaksi yang lebih baik antar peserta, baik antara siswa dengan pemateri maupun dengan sesama siswa, melalui fitur diskusi, kolaborasi, dan komunikasi yang tersedia dalam platform (Costley dkk., 2022).

Dengan semakin berkembangnya teknologi, LMS terus mengalami inovasi, termasuk dengan penerapan kecerdasan buatan untuk personalisasi pembelajaran dan analisis data yang lebih mendalam (Alotaibi, 2024). Penggunaan LMS di berbagai sektor menunjukkan bahwa sistem ini menjadi bagian penting dalam transformasi digital pendidikan dan pelatihan, memberikan kemudahan dalam akses ilmu pengetahuan serta meningkatkan efisiensi proses pembelajaran.

II.2 *Affective Computing*

Affective computing adalah bidang *interdisipliner* dalam kecerdasan buatan yang berfokus pada pengenalan, interpretasi, dan simulasi emosi manusia oleh komputer. Konsep ini pertama kali diperkenalkan oleh Rosalind Picard pada tahun 1995, yang menekankan pentingnya emosi dalam interaksi manusia dan teknologi. Dengan kemajuan teknologi, *affective computing* kini diterapkan dalam berbagai domain, seperti interaksi manusia-mesin, perawatan kesehatan, pendidikan, dan hiburan. Kemampuan sistem untuk mengenali dan merespons emosi pengguna dapat meningkatkan pengalaman pengguna secara signifikan (Afzal dkk., 2024).

Dalam dunia pendidikan *affective computing* memiliki peran, terutama dalam meningkatkan pengalaman belajar yang lebih personal dan adaptif (Villegas-Ch dkk., 2024). Teknologi ini memungkinkan sistem pembelajaran digital untuk mengenali emosi siswa melalui ekspresi wajah, nada suara, atau bahkan respons fisiologis (Yu dkk., 2024). Dengan memanfaatkan *affective computing*, sistem dapat menyesuaikan materi dan metode pembelajaran sesuai dengan kondisi emosional siswa, misalnya dengan memperlambat penjelasan jika siswa terlihat kebingungan atau memberikan motivasi saat mereka merasa frustasi (Shi, 2024). Hal ini dapat meningkatkan efektivitas pembelajaran dan membantu siswa lebih fokus serta termotivasi dalam proses belajar (Lin dkk., 2022).

II.3 *Automated Emotion Recognition* di Lingkungan Pendidikan

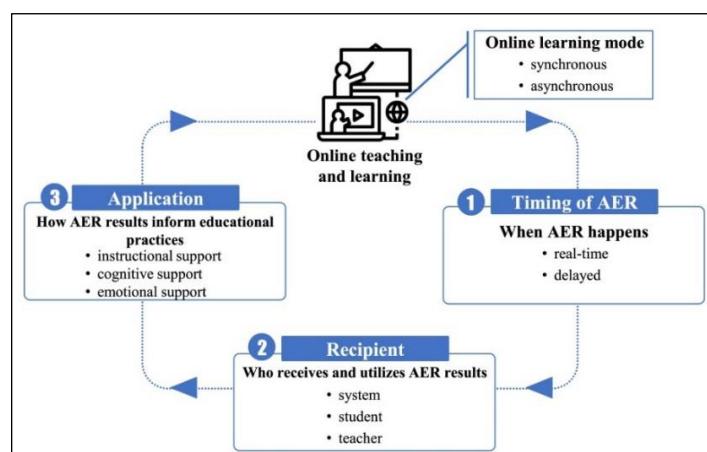
Automated emotion recognition (AER) dalam lingkungan pendidikan merujuk pada teknologi berbasis kecerdasan buatan yang secara otomatis mendeteksi dan menganalisis emosi siswa selama proses pembelajaran. Sistem ini menggunakan berbagai teknik seperti pengenalan wajah, analisis suara, pemantauan fisiologis,

dan pemrosesan bahasa alami untuk memahami kondisi emosional siswa secara *real-time* (Yu dkk., 2024). Dengan memanfaatkan teknologi ini, pendidik dan sistem pembelajaran adaptif dapat menyesuaikan strategi pengajaran agar lebih sesuai dengan keadaan emosional siswa, sehingga meningkatkan *engagement* mereka dalam proses belajar .

Dalam praktiknya, AER dapat diterapkan baik dalam kelas fisik maupun pembelajaran daring. Di kelas tradisional, kamera dan sensor dapat digunakan untuk menganalisis ekspresi wajah dan bahasa tubuh siswa, sementara dalam pembelajaran daring, sistem dapat mengevaluasi ekspresi wajah dan intonasi suara dari video konferensi. Informasi yang diperoleh dari AER memungkinkan pengajar untuk memberikan intervensi yang lebih tepat, seperti mengubah metode penyampaian materi atau memberikan dukungan tambahan bagi siswa yang mengalami kesulitan. Dengan demikian, teknologi ini berpotensi meningkatkan efektivitas pembelajaran serta menciptakan pengalaman belajar yang lebih adaptif dan responsif terhadap kebutuhan emosional siswa.

Dalam konteks pendidikan, AER bertujuan untuk:

1. Membantu guru dalam memahami kebutuhan siswa dengan memberikan wawasan tentang kondisi emosional mereka (Llurba dkk., 2024).
2. Meningkatkan *engagement* siswa dengan menyesuaikan strategi pembelajaran berdasarkan emosi mereka (Yu dkk., 2024).
3. Mempersonalisasi pengalaman belajar dan meningkatkan penyesuaian diri siswa dalam tugas-tugas berbasis kelompok (Ngo dkk., 2024).



Gambar II.1 Siklus penerapan AER pada lingkungan pendidikan (Yu dkk., 2024)

Peniliti Yu dkk., (2024) dalam ulasannya mencoba untuk memetakan bagaimana siklus proses pengaplikasian AER dalam lingkungan pendidikan yang dapat dilihat pada gambar II.1. Yu dkk., (2024) menjelaskan bahwa penerapan AER pada pembelajaran daring bisa dalam pembelajaran *synchronous* maupun pembelajaran *asynchronous*. Penerapannya juga bisa dalam keadaan *real-time* maupun *delayed*. Kemudian entitas yang terkait adalah aplikasi sistem pembelajaran daring, guru atau pemateri, dan siswa. Hasil dari penerapan AER bisa berbentuk instruksional, sebagai contoh jika diterapkan secara real-time, sistem bisa memperingatkan siswa jika tidak fokus, lalu bisa juga berbentuk penilaian kognitif, dan juga berbentuk data emosi dari siswa.

Atas dasar tersebut, penulis memilih untuk melakukan penelitian pada pembelajaran *asynchronous* dengan keadaan *delayed* dan menghasilkan data emosi dari siswa dengan tambahan modalitas arah pandangan dan arah kepala yang kemudian disimpulkan untuk menentukan *engagement* dari siswa pada pembelajaran daring dengan alasan bahwa pengenalan emosi pada keadaan *delayed* memiliki beban *resource* yang lebih ringan dibandingkan dengan sistem *real-time* (Pascual dkk., 2022).

II.3.1 Modalitas Teks

Penggunaan modalitas teks dalam lingkungan pendidikan bisa digunakan untuk penilaian pedagogik. Penilaian pedagogik menggunakan teks tanpa menerapkan AER telah biasa dilakukan seperti menggunakan kuisioner (Herrero-Alvarez dkk., 2024), melakukan survey (Benabbes dkk., 2023), dan penilaian komulatif berbasis teks pada platform *daring education* (Xu, 2023). Namun, pendekatan tersebut tidak akurat untuk menilai siswa dari segi emosionalnya. Analisis sentimen adalah salah satu pendekatan yang dapat diterapkan untuk membangun sebuah sistem penilaian pedagogik dengan menerapkan AER melalui data modalitas teks.

Peneliti Zhang & Qin, (2022) membuat sistem yang menerapkan analisis sentimen untuk mengevaluasi emosi siswa pada pelajaran Ideologi dan Politik dalam platform daring. Mereka mengombinasikan *Convolutional Neural Network* (CNN) dan *Long Short-Term Memory* (LSTM) untuk menganalisis komentar mengenai kursus ideologis dan politik secara daring. Pendekatan ini bertujuan untuk

meningkatkan akurasi dalam mengenali sentimen pengguna dibandingkan dengan metode tradisional. Hasil penelitiannya menunjukkan bahwa algoritma CNN+LSTM berhasil mengungguli metode konvensional seperti SVM dan LSTM. Kinerja terbaik dicapai dengan menggunakan metode vektorisasi teks Word2vec, yang mampu menangkap representasi semantik kata dengan lebih baik. Dalam analisis terhadap 12.000 komentar, ditemukan bahwa 86,36% dari komentar tersebut mengandung sentimen positif, sementara 13,64% lainnya bersifat negatif. Temuan ini memberikan wawasan berharga bagi pengelola kursus untuk melakukan perbaikan berdasarkan umpan balik negatif yang diberikan oleh peserta.

Pendekatan analisis sentimen juga dilakukan oleh Mao dkk., (2023) dalam evaluasi terhadap tingkat emosional siswa dalam pelajaran *Internet of Things* (IoT). Mao dkk., (2023) menggunakan algoritma *Information Block Bidirectional Long-Short Term Memory* (IB-BiLSTM) untuk meningkatkan analisis sentimen dalam teks pada pendidikan animasi secara *daring*. Algoritma ini memanfaatkan teknologi IoT untuk mengumpulkan data multimodal. Model yang dikembangkan mencapai akurasi 93,92% dan skor F1 90,34%, menunjukkan keefektifannya dalam mendeteksi perubahan emosi selama proses pembelajaran. Penelitian ini berkontribusi pada pengembangan pendidikan yang lebih personal dan meningkatkan *engagement* emosional siswa.

Walaupun demikian, analisis sentimen masih memiliki kelemahan hanya mendeteksi emosi positif dan negatif dan memiliki keterbatasan dalam menangkap nuansa emosi yang lebih kompleks. Pendekatan biner ini tidak dapat membedakan emosi spesifik seperti marah, sedih, takut, atau senang, yang dapat memberikan wawasan lebih mendalam tentang perasaan siswa (Gamage dkk., 2024). Berdasarkan hal tersebut, penulis menyimpulkan bahwa modalitas teks kurang cocok untuk digunakan dalam mendeteksi emosi siswa pada penelitian ini.

Data modalitas teks dapat berasal dari dokumen tertulis, transkrip percakapan, media sosial, artikel berita, hingga korpus linguistik yang dikurasi. Metode koleksi dapat dilakukan melalui *web scraping*, ekstraksi dari basis data publik, penggunaan API dari *platform* pembelajaran daring, serta melalui anotasi manual oleh pakar bahasa. Format penyimpanan data teks dapat berupa TXT, JSON, CSV, atau

database berbasis teks seperti Elasticsearch. Data yang telah dikumpulkan kemudian digunakan untuk berbagai keperluan, seperti analisis sentimen, pemrosesan bahasa alami (*Natural Language Processing*), dan pengembangan model kecerdasan buatan berbasis teks.

Peneliti Benabbes dkk., (2023) mengumpulkan data siswa langsung dari *platform e-learning*. Pengumpulan data tersebut dilakukan dengan mendapatkan umpan balik melalui survei dari siswa dan guru untuk menilai fungsionalitas dan kemudahan dalam penggunaan sistem pembelajaran. Data umpan balik yang didapatkan selanjutnya dianalisis untuk menilai *engagement* siswa dalam pembelajaran.

Metode serupa juga dilakukan oleh Herrero-Alvarez dkk., (2024), mereka mengumpulkan data tes menggunakan dua kuesioner yang diberikan kepada siswa di akhir kegiatan. Proses ini menghasilkan total 10 tanggapan per siswa. Tanggapan dikumpulkan menggunakan *platform Google Forms*, yang memfasilitasi pengumpulan data emosional dari peserta. Selain itu, keadaan emosional siswa dinilai melalui instrumen *Self-Assessment Manikin* (SAM) dan *Positive and Negative Affect Schedule* (PANAS), yang dirancang khusus untuk mengukur berbagai dimensi emosi.

Jenis data modalitas teks lainnya yang bisa digunakan untuk menilai *engagement* siswa adalah melalui data kuantitatif. Peneliti Tao dkk., (2022) menganalisis data yang diambil dari data seberapa banyak jumlah *post* yang dibuat siswa dalam media pembelajaran daring menggunakan metode *machine learning* (ML) dengan tambahan parameter lainnya. Walaupun tidak dapat menilai emosi dari siswa, namun metode tersebut cukup efisien jika hanya untuk menilai *engagement* atau keaktifan siswa dalam pembelajaran daring.

Dalam analisis sentimen, Zhang dan Qin, (2022) memanfaatkan komentar siswa pada platform *Massive Open Online Course* (MOOC). Dari data komentar tersebut dikumpulkan dan dilakukan analisis dengan menggunakan *deep learning* untuk menilai emosi dari siswa berdasarkan teks komentar yang ditulisnya.

Berbeda dengan peneliti di atas, Mao dkk., (2023) mendapatkan data modalitas teks dari konversi data sensor yaitu *touchscreen* dan *keyborad* untuk menilai feed back

berupa teks, lalu untuk menangkap emosi menggunakan beberapa perangkat seperti sensor *body motion*, perangkat *eye tracker*, dan sensor EEG. Mereka juga menggunakan *smart pen* atau *handwriting tablet* untuk mendapatkan data tulisan siswa, dan yang terakhir menggunakan *speech recognition* untuk mendapatkan ekspresi verbal siswa. Semua data yang dikumpulkan berupa teks yang kemudian diproses dengan metode IB-BiLSTM yang hasil keluarannya adalah klasifikasi emosi. Apa yang dilakukan oleh Mao dkk., (2023) menunjukkan bahwa penggunaan modalitas teks dapat diambil dari bentuk apapun, baik visual, *audio*, ataupun sinyal walaupun hasilnya akan lebih spesifik dalam mengenali emosi, namun prosesnya menjadi tidak efisien untuk diterapkan dalam lingkungan pendidikan khususnya dalam pembelajaran daring.

Berdasarkan uraian di atas, pendekatan deteksi emosi dengan modalitas teks dinilai kurang tepat untuk diterapkan pada media pembelajaran daring, selain karena hanya bisa mendeteksi emosi biner (*positif* atau *negatif*) dan tidak dapat membedakan emosi spesifik seperti marah, sedih, takut, atau senang, yang dapat memberikan wawasan lebih mendalam tentang perasaan siswa (Gamage dkk., 2024), alasan lainnya adalah dimungkinkan untuk terjadinya emosi palsu dari siswa dalam tulisannya (Khoshnam & Baraani-Dastjerdi, 2022), seperti kata “tidak” jika diungkapkan dengan nada rendah bukan berarti emosinya negatif, sehingga diperlukan modalitas pendukung lain, seperti *audio* ataupun *visual* (Das dkk., 2024).

II.3.2 Modalitas *Audio*

Dalam lingkungan pendidikan, modalitas audio dapat digunakan untuk menentukan emosi siswa dengan teknologi *speech emotion recognition* (SER). Beberapa literatur penelitian terdahulu, ada beberapa peneliti yang menggunakan modalitas audio sebagai parameter penentu emosi di bidang pendidikan. Tanko dkk., (2022) pada penelitian membuat sebuah pendekatan SER yang bertujuan untuk mendeteksi emosi pemateri saat mengajar berdasarkan ucapan mereka. Untuk mencapai tujuan ini, dua kumpulan data pidato dalam bahasa Turki dan Inggris dikumpulkan. Model pembelajaran baru tersebut dinamai *ShoePat23* dan berhasil mencapai akurasi klasifikasi sebesar 94,97% untuk bahasa Inggris dan 96,41% untuk bahasa Turki.

Hasil penelitian menunjukkan bahwa *ShoePat23* merupakan teknik yang efektif dan efisien untuk pengenalan emosi dalam ucapan, terutama dalam aplikasi dunia nyata. Selain itu, studi ini mengungkapkan bahwa model ini dapat digunakan untuk mengembangkan alat pengendalian diri baru yang dapat diintegrasikan ke dalam platform pendidikan jarak jauh guna mengevaluasi kinerja pemateri.

Pendekatan lain bernama *Esernet* yang merupakan improvisasi dari sistem SER juga dapat digunakan dalam deteksi emosi siswa (T. Liu dkk., 2025). Ada dua temuan utama dalam penelitian yang dilakukan oleh T. Liu dkk, yaitu adanya isyarat emosional yang signifikan serta perbedaan kecil antara emosi yang berbeda, yang keduanya berperan penting dalam meningkatkan akurasi SER. Model yang diperkenalkan menggunakan strategi dua jalur. Jalur pertama berfokus pada ekstraksi isyarat ucapan penting dengan teknik masking, sementara jalur kedua menggabungkan informasi multiskala dari sinyal suara. Arsitektur *Transformer* digunakan untuk menangkap hubungan semantik jarak jauh dalam sinyal ucapan. Hasil eksperimen menunjukkan bahwa *Esernet* mampu mengenali isyarat penting dan hubungan jarak jauh dengan baik. Namun, penelitian ini juga mengakui bahwa model ini memiliki keterbatasan, terutama dalam hal kebutuhan komputasi yang tinggi akibat strategi multi-tugas yang digunakan.

Dalam modalitas audio khususnya dalam pendekatan SER, beberapa literatur terdahulu menunjukkan bahwa mereka menggunakan dua jenis data, yaitu data publik, dan data *private*. Data modalitas *audio* dapat diperoleh dari berbagai sumber, termasuk menggunakan mikrofon secara langsung, sensor ultrasonik, atau perangkat rekaman lainnya, dengan format yang beragam seperti sinyal mentah, spektrogram, atau fitur yang diekstraksi seperti *Mel-Frequency Cepstral Coefficients* (MFCC). Dari berbagai literatur disebutkan bahwa beberapa peneliti menggunakan dua jenis *dataset* modalitas audio, yaitu *dataset publik* dan *private*.

Beberapa peneliti yang menggunakan *dataset* publik adalah Guan, (2024). Dia menggunakan *dataset* CMU ARCTIC, dimana CMU ARCTIC adalah kumpulan *dataset* ucapan (*speech corpus*) yang dikembangkan oleh Carnegie Mellon University (CMU). *Dataset* ini dirancang untuk penelitian dan pengembangan dalam bidang *speech synthesis* atau *text-to-speech* (TTS). CMU ARCTIC berisi

rekaman suara dari berbagai penutur dengan aksen yang berbeda, masing-masing membaca sekitar 1.132 kalimat yang dipilih secara khusus untuk mencakup berbagai fonema dalam bahasa Inggris. Selain dari situs resminya, *dataset* tersebut juga tersedia pada situs *Kaggle* maupun *Hugging Face*.

Peneliti T. Liu dkk., (2025) juga menggunakan *dataset* publik pada penelitiannya. Mereka menggunakan gabungan dari *dataset* IEMOCAP dan EmoDB. IEMOCAP adalah kumpulan data yang terdiri dari rekaman video aktor yang melakukan dialog naskah dengan berbagai ekspresi emosional. *Dataset* mencakup data multimodal, seperti bingkai video, sinyal audio, dan subtitle teks, sehingga cocok untuk mengevaluasi metode pengenalan emosi multimodal (Le dkk., 2023). Sedangkan Berlin Database of Emotional Speech atau yang lebih dikenal dengan nama EmoDB adalah kumpulan data yang terdiri dari rekaman suara aktor yang mengekspresikan emosi yang berbeda. Rekaman ini dirancang untuk menangkap berbagai keadaan emosional, termasuk kemarahan, ketakutan, kebahagiaan, kesedihan, jijik, kebosanan, dan keadaan netral. Basis data disusun untuk menyediakan serangkaian ekspresi emosional yang komprehensif untuk analisis dan pelatihan model (Kanwal & Asghar, 2021). *Dataset* IEMOCAP dan EmoDB bisa didapatkan pada situs *Kaggle* maupun *Hugging Face*.

Peneliti Tanko dkk., (2022), menggunakan *dataset private* dengan cara mengumpulkan *dataset English Speech Emotion Recognition* dari total 45 pembicara, yang mencakup 15 wanita dan 30 pria, berusia antara 24 hingga 42 tahun. File *audio* direkam ketika pembicara menyampaikan kuliah tentang forensik digital menggunakan slide kuliah yang disiapkan. Rekaman dibuat dengan ponsel, dan audio ditangkap dalam lima format berbeda, yaitu MP4, MPEG, AAC, OGG, dan M4A. *Dataset* yang terkumpul terdiri dari berbagai kelas emosional, dengan 44 suara dengan label “menarik”, 47 dengan label “netral”, dan 43 dengan label “membosankan”.

Berdasarkan uraian di atas, meskipun modalitas audio dapat mendeteksi emosi dengan baik, ada beberapa masalah saat menggunakannya saat mengidentifikasi emosi siswa dalam pembelajaran daring. Jika fitur *audio* digabungkan dengan fitur lain, ini dapat menurunkan akurasi (Pan dkk., 2024). Selain itu, pemrosesan sinyal

audio membutuhkan sumber daya komputasi yang besar (Z. T. Liu dkk., 2021), yang membuatnya tidak efisien dalam penerapan skala luas. Kemudian masalah lainnya yaitu jumlah data yang tersedia untuk pelatihan sering kali terbatas (Zhu-Zhou dkk., 2022). Atas pertimbangan tersebut, penulis tidak menggunakan modalitas *audio* dalam menentukan *engagement* siswa berdasarkan emosi, mengingat *resource* yang dimiliki pada lokus penelitian tidak terlalu besar.

II.3.3 Modalitas Sinyal Tubuh

Dalam lingkungan yang terkontrol, deteksi emosi siswa kadang juga dilakukan dengan memanfaatkan signal dari tubuh yang ditangkap melalui *tools* atau *sensor* yang terpasang pada bagian tubuh siswa. Penggunaan perangkat IoT juga menjadi pilihan *tools* untuk mendapatkan data signal tubuh siswa (Awais dkk., 2021). Peneliti Awais dkk., (2021) mengusulkan kerangka kerja berbasis IoT untuk menganalisis emosi manusia dengan menggunakan sinyal fisiologis yang diproses melalui model pembelajaran mendalam berbasis LSTM. Model ini mampu mengenali berbagai emosi, seperti lucu, membosankan, relaksasi, dan menakutkan, dengan akurasi lebih dari 95%. Kerangka IoT yang dikembangkan memastikan komunikasi data yang cepat dan andal antara perangkat sensor dan hub IoT, dengan latensi kurang dari 1 milidetik. Sistem ini diharapkan dapat mendukung siswa, institusi pendidikan, dan layanan kesehatan, terutama dalam pembelajaran jarak jauh selama pandemi COVID-19 dan potensi wabah di masa depan.

Pendekatan lainnya adalah dengan memanfaatkan sinyal EEG (Fernandez dkk., 2024). Studi yang dilakukan Fernandez dkk., (2024) menyoroti meningkatnya minat dalam deteksi stres menggunakan sinyal fisiologis, terutama data EEG, untuk meningkatkan pemantauan kesehatan mental dan kesejahteraan. Data dikumpulkan melalui protokol khusus yang dirancang untuk menginduksi stres pada peserta, dengan biosinyal yang direkam dari EEG, EDA, dan EKG. Berbagai teknik ekstraksi fitur dan model pembelajaran mesin diuji untuk mengidentifikasi stres secara akurat, dengan hasil yang menunjukkan bahwa pemilihan fitur, model, dan ukuran jendela yang tepat sangat berpengaruh terhadap kinerja sistem. LightGBM terbukti sebagai metode paling efektif, mencapai akurasi tertinggi di berbagai kondisi, sementara data EEG memberikan hasil yang lebih baik dibandingkan EKG

dan EDA. Integrasi semua sinyal fisiologis menghasilkan performa terbaik, meskipun efektivitasnya tetap bergantung pada konteks pengukuran stres.

Penggunaan sinyal EEG dalam deteksi emosi siswa juga dilakukan oleh peneliti Moise dkk., (2024). Studi yang mereka lakukan bertujuan untuk mengembangkan model pengenalan emosi yang berkinerja tinggi dengan memanfaatkan data psikologis yang relevan dalam proses pendidikan. Tujuh emosi yang dianalisis meliputi kebosanan, kebingungan, frustrasi, rasa ingin tahu, kegembiraan, konsentrasi, dan kecemasan. Model ini menggunakan arsitektur 1D-CNN yang dioptimalkan untuk lima saluran EEG, dengan hasil akurasi tinggi, di mana pengenalan emosi kegembiraan mencapai 99,91%. Untuk menjelaskan prediksi model, studi ini menerapkan LIME, yang menunjukkan bahwa analisis frekuensi fitur dapat membantu memperjelas hasil klasifikasi. Namun, terdapat beberapa keterbatasan, seperti data yang tidak dikumpulkan langsung dalam skenario pembelajaran serta tantangan dalam mengumpulkan data EEG dari peserta pendidikan.

Data modalitas sinyal dapat berasal dari berbagai sumber, seperti EEG, ECG, EMG, atau sinyal dari sensor IoT. Proses koleksi data umumnya dimulai dengan pemilihan perangkat akuisisi yang sesuai, seperti biosensor, perangkat medis, atau sensor industri, yang mampu menangkap sinyal dengan resolusi dan kecepatan sampling yang memadai. Setelah itu, data dikumpulkan dalam kondisi yang telah ditentukan, seperti lingkungan laboratorium yang terkendali atau melalui perangkat wearable yang digunakan dalam aktivitas sehari-hari. Selama proses akuisisi, penting untuk memastikan kualitas sinyal dengan mengurangi *noise* melalui teknik penyaringan (*filtering*) atau *artifact removal*. Data yang telah dikumpulkan kemudian disimpan dalam format yang sesuai, seperti CSV, EDF, atau HDF5, untuk selanjutnya diproses dan dianalisis dalam penelitian atau pengembangan sistem berbasis sinyal.

Peneliti Awais dkk., (2021) mengumpulkan data uji menggunakan sensor fisiologis, yang mencakup berbagai modalitas, seperti EEG, EKG, EMG, GSR, RSP, SKT, BVP, dan PPG. Sebanyak tiga puluh peserta terlibat dalam eksperimen akuisisi data, di mana mereka diperlihatkan delapan video yang dirancang untuk membangkitkan empat kategori emosi yang berbeda yaitu, relaksasi, kebosanan,

kesenangan, dan ketakutan. Sinyal fisiologis direkam secara kontinu dengan frekuensi pengambilan sampel 1.000 Hz untuk menangkap respons emosional peserta secara objektif.

Lalu peneliti Fernandez dkk., (2024) mengumpulkan data menggunakan pengaturan komprehensif yang mencakup akuisisi sinyal fisiologis khususnya EEG, EKG, dan EDA dari subjek yang sama. Proses ini dilakukan dengan mengikuti protokol standar untuk memastikan konsistensi dalam seluruh data yang dikumpulkan. Konfigurasi elektroda yang digunakan untuk menangkap sinyal ini dirancang agar memungkinkan akuisisi yang akurat dan simultan, sehingga memfasilitasi analisis holistik terhadap respons stres subjek. Selain itu, wawancara pasca-eksperimen dan kuesioner dilakukan untuk mengumpulkan data kualitatif, yang kemudian dibandingkan dengan data fisiologis yang dikumpulkan selama pengujian.

Moise dkk., (2024) juga melakukan metode serupa dengan menggunakan alat sensor dalam pengumpulan datanya. Mereka mengumpulkan data uji menggunakan pendekatan multimodal, dengan fokus pada sinyal EEG dan berbagai sinyal fisiologis perifer. *Dataset DEAP*, yang digunakan dalam penelitian ini, mencakup sinyal EEG yang direkam dari 32 subjek saat mereka menonton 40 video musik. Sinyal EEG ditangkap menggunakan elektroda AgCl, yang diposisikan sesuai dengan sistem 10/20 internasional, dengan laju pengambilan sampel 512 Hz. Selain itu, sinyal fisiologis *perifer* seperti *galvanic skin response* (GSR), amplitudo respirasi, suhu kulit, EKG, dan EMG, juga direkam menggunakan sensor yang ditempatkan di wajah, leher, dan tangan kiri, serta sabuk respirasi untuk mengukur pola pernapasan. Metode pengumpulan data yang komprehensif ini bertujuan untuk meningkatkan keakuratan pengenalan emosi dalam proses pembelajaran.

Walaupun pengenalan emosi menggunakan signal asli tubuh manusia seperti EEG lebih efektif dibanding metode tradisional (Li dan Deng, 2022a), namun penggunaan sensor atau tools dalam menangkap sinyal tubuh manusia pasti memerlukan perangkat yang lebih banyak daripada pegenalan emosi lainnya seperti FER, SER, dan SA (Li & Deng, 2022a). Atas dasar tersebut, penulis juga tidak

memilih pendekatan menggunakan modalitas sinyal tubuh untuk diterapkan dalam penentuan *engagement* siswa berdasarkan emosinya.

II.3.4 Modalitas Visual

Dalam beberapa tahun ke belakang, modalitas visual merupakan modalitas terbanyak yang digunakan oleh beberapa peneliti dalam mendeteksi emosi pada lingkungan pendidikan (Yu dkk., 2024), terutama dengan pendekatan FER. Ketersediaan *dataset* publik yang relatif banyak dan masih banyaknya peluang penelitian tentang FER, menjadikan FER sebagai topik penelitian yang masih menarik untuk terus dikembangkan terutama di bidang pendidikan seperti yang dilakukan oleh Wang, (2021). Wang menyimpulkan bahwa perilaku belajar daring dapat dianalisis secara efektif menggunakan teknik pengenalan emosi berbasis gambar. Proses ini mencakup ekstraksi *keyframe* dari ekspresi wajah dengan menerapkan *enhanced local binary pattern* (LBP) dan *wavelet transform*. *Keyframe* yang diperoleh digunakan untuk mengekstrak fitur ekspresi rata-rata untuk mengoptimalkan akurasi dalam *emotion recognition*. Sistem analisis perilaku pembelajaran daring dikembangkan oleh Wang, (2021) adalah dengan mengintegrasikan *face-based emotion recognition methods*. Model klasifikasi emosi pada *image-based daring learning* dirancang menggunakan *attention mechanism*, yang terbukti meningkatkan performa klasifikasi. Hasil penelitiannya menunjukkan bahwa model ini dapat diimplementasikan untuk *daring learning behavior analysis*.

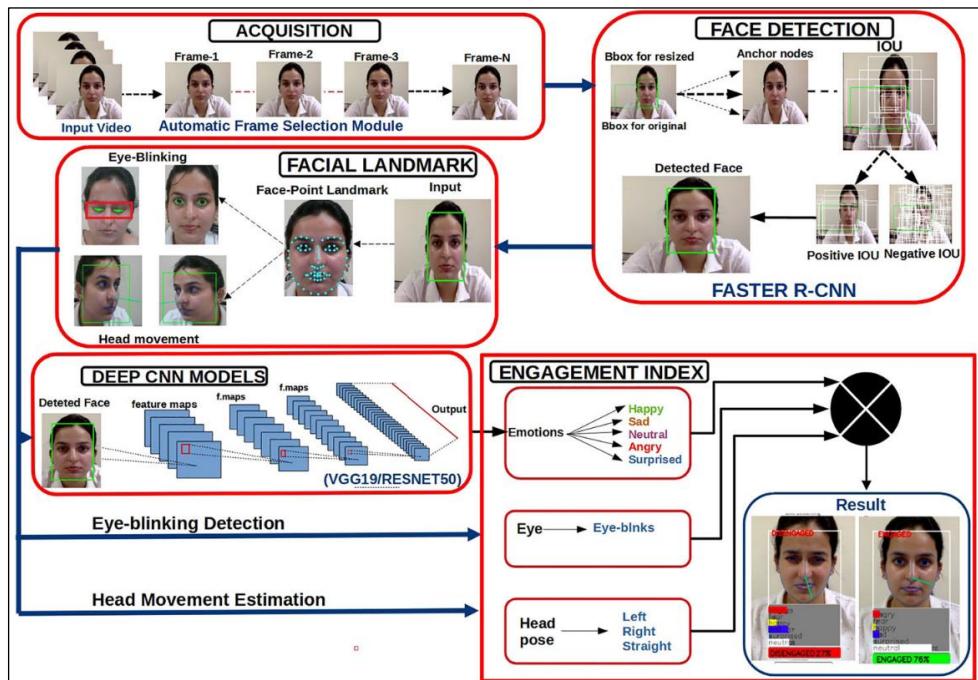
Pendekatan menggunakan modalitas *visual* juga bisa digunakan dalam mem membuat sebuah *visual analytics system* (Zeng dkk., 2021). Peneliti Zeng dkk., (2021) memperkenalkan EmotionCues, yaitu sebuah *visual analytics system* yang dirancang untuk merangkum dan menganalisis emosi siswa dalam suatu video kelas secara luring. Dengan fitur visualisasi interaktif, EmotionCues memungkinkan *end-users*, seperti guru dan para orang tua untuk mendapatkan *insights* mengenai *engagement* siswa dan emosinya.

Penerapan FER memiliki tantangan tersendiri terhadap emosi anak (Rathod dkk., 2022). Peneliti Rathod dkk., (2022) menyimpulkan bahwa model *deep learning* efektif untuk deteksi emosi pada anak-anak dalam media pembelajaran daring

dengan menggunakan *dataset* LIRIS serta *dataset* khusus yang dikembangkan untuk anak usia 7 hingga 10 tahun. Rathod dkk., (2022) menyatakan bahwa *facial expressions* pada anak-anak berbeda dari orang dewasa, sehingga digunakan 3D 468 *key geometric points* untuk meningkatkan *emotion recognition accuracy* untuk mendeteksi emosi pada anak-anak. Hasil akurasinya mencapai 90.98% melalui *comprehensive comparative analysis* dengan tujuh model arsitektur CNN.

Penerapan FER juga bisa dalam keadaan real-time pada video saat pembelajaran daring (Mehta dkk., 2022). Peneliti Mehta dkk., (2022) menyimpulkan bahwa level konsentrasi siswa dapat dideteksi secara efektif melalui *real-time facial expression analysis*. Tiga tingkat konsentrasi yang diidentifikasi dan diperoleh dari *facial emotions* adalah “*highly concentrated*”, “*nominally concentrated*”, dan “*not concentrated at all*”. Hasil penelitian ini menunjukkan efektivitas model *deep learning*, khususnya VGG16, dalam *facial emotion classification*, dengan akurasi tinggi dalam mendeteksi tingkat konsentrasi. Secara keseluruhan, hasil penelitian mengindikasikan bahwa mengintegrasikan emotion recognition ke dalam *e-learning systems* dapat meningkatkan pengalaman belajar dengan menyediakan umpan balik secara langsung terkait *engagement* siswa.

Selain dengan FER, deteksi *engagement* siswa juga bisa menggunakan *head pose* (Gupta dkk., 2023b) dan *eye gaze* (Tian dkk., 2024). Dengan teknik tersebut, sistem bisa menyimpulkan keadaan siswa apakah fokus terhadap pelajaran atau tidak. Peneliti Gupta dkk., (2023a), menggabungkan *head pose* dan *eye blink* (gambar II.2) dalam mendeteksi *engagement* siswa. Dilatarbelakangi atas meningkatnya permintaan pembelajaran daring selama pandemi COVID-19, Gupta dkk., (2023a) mengusulkan sistem *student engagement detection* untuk memberikan umpan balik. Sistem ini dirancang untuk mendukung *e-learning platforms* dengan mendeteksi tingkat *engagement* siswa secara *real-time*. Pendekatan multimodal diterapkan dengan mengombinasikan data dari *facial expressions*, *eye movements*, dan *head movements* untuk mengklasifikasikan siswa sebagai *engaged* atau *disengaged*. Model *deep learning* digunakan untuk *facial emotion recognition*, *blink detection*, dan *head movement tracking*, yang kemudian dihitung menjadi *Engagement Index* (EI). Hasil eksperimen menunjukkan bahwa sistem ini mencapai akurasi sebesar 92.58%.



Gambar II.2 Deteksi *engagement* siswa menggunakan FER, *eye blink*, dan *head pose* (Gupta dkk., 2023a)

Data modalitas visual dapat berasal dari berbagai sumber, seperti kamera RGB, sensor, kamera termal, atau kombinasi dari beberapa perangkat. Beberapa literatur menunjukkan banyak peneliti yang menggunakan *dataset* publik untuk membuat model latih, seperti Affectnet (Trabelsi dkk., 2023), FER-2013 (Pordoy dkk., 2024), CK+ (Harb dkk., 2023). Selain itu, ada pula yang menggabungkan beberapa *dataset* publik, seperti yang dilakukan oleh Gupta dkk., (2023c), Meriem dkk., (2022), dan Aly dkk., (2023).

Untuk mendapatkan *dataset* publik pada modalitas visual, langkah pertama yang dapat dilakukan adalah mencari sumber *dataset* yang sesuai dengan kebutuhan penelitian, seperti Kaggle, ImageNet, AffectNet, atau Google *Dataset* Search. Setelah menemukan *dataset* yang relevan, periksa lisensi dan syarat penggunaannya untuk memastikan kepatuhan terhadap hak cipta. Selanjutnya, *dataset* dapat diunduh melalui API atau tautan resmi yang disediakan oleh penyedia *dataset*. Selain itu, jika *dataset* yang tersedia belum mencukupi, data tambahan dapat dikumpulkan dari sumber terbuka lainnya atau melalui teknik *web scraping* dengan tetap memperhatikan etika pengambilan data.

Selain *dataset* publik, beberapa peneliti juga menggunakan data private yang terdiri dari kumpulan wajah siswa yang diambil langsung dalam pengamatan selama beberapa waktu. Pada modalitas visual, peneliti lebih cendrung menggunakan kamera, baik webcam yang tertanam pada perangkat komputer maupun kamera profesional seperti yang dilakukan oleh Rathod dkk., (2022) yang menggunakan kamera DSLR untuk mengambil data wajah siswa yang merupakan anak-anak untuk dijadikan model latih.

Pengumpulan data menggunakan *webcam* pada lingkungan pembelajaran daring lazim dilakukan oleh beberapa peniliti. Gupta dkk., (2023a) pada penelitiannya mengumpulkan data uji dengan menangkap video *streaming* secara *real-time* dari *webcam* yang digunakan oleh siswa di lingkungan *E-learning*. Video yang ditangkap kemudian diekstrak menjadi *frame* wajah setiap 20 detik. Kumpulan *frame* wajah tersebut dianalisis untuk mendeteksi titik-titik kunci pada wajah, kedipan mata, dan gerakan kepala. Cara yang sama juga dilakukan oleh Gupta dkk., (2023b) pada penelitiannya yang serupa.

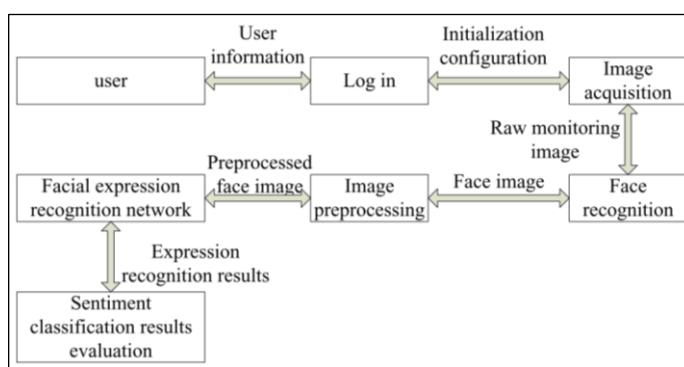
Penggunaan *webcam* juga dilakukan oleh Shobana dan Kumar, (2021), yaitu dengan cara *webcam* diaktifkan segera setelah pelajar masuk ke sistem kuis, merekam video penilaian berdasarkan pertanyaan, dan mengubah video menjadi *frame* gambar secara berkala untuk analisis lebih lanjut. Peneliti seperti Aly dkk., (2023), Trabelsi dkk., (2023), Sassi dkk., (2023), dan Li, (2025), juga menggunakan *webcam* dalam pengumpulan data baik untuk data latih, maupun data uji.

Cara pengumpulan data yang berbeda dilakukan oleh Bao dkk., (2024b). Mereka mengumpulkan data uji menggunakan perangkat pelacak mata, yaitu TobiiTX300 yang mencakup modul gerakan mata dan modul tampilan 23 inci. Perangkat beroperasi pada frekuensi sampling 60 Hz/s dan terhubung ke komputer desktop yang mengontrol prosedur eksperimental dan merekam data. Pengumpulan data melibatkan 68 mahasiswa berusia 20-23 tahun, dengan rasio pria dan wanita seimbang 1:1. Percobaan dilakukan pada lingkungan laboratorium terkontrol dengan kecerahan konstan. Apa yang dilakukan oleh Bao dkk., (2024b) menunjukkan bahwa ada cara lain untuk mengumpulkan data dari modalitas visual selain menggunakan kamera.

Berdasarkan uraian di atas, modalitas visual merupakan pilihan yang sangat logis untuk diterapkan pada sebuah sistem pembelajaran daring. Hal ini didasari pada banyaknya data publik yang bisa digunakan dalam membuat model FER sehingga proses pengumpulan data penelitian bisa lebih cepat dan efisien. Namun, penggunaan modalitas visual pada pembelajaran daring juga mempunyai beberapa kekurangan, terutama pada penerapan FER dalam keadaan oklusi, seperti wajah yang tertutup oleh tangan, kacamata, atau objek lain (Li dan Deng, 2022). Atas dasar tersebut, penelitian tentang cara mengatasi keadaan oklusi dalam penerapan FER masih dipandang perlu untuk dilakukan untuk menyempurnakan penerapan FER di berbagai bidang.

II.3.5 Multimodal

Beberapa peneliti menggunakan lebih dari satu elemen dalam pendekripsi emosi siswa atau yang lebih dikenal dengan sebutan *multimodal emotion recognition* (MER). Cui dkk., (2021) pada penelitiannya menggabungkan FER (visual) dan analisis sentimen (teks) untuk mendekripsi emosi siswa pada *Business English Class* dengan alur pada gambar II.3.



Gambar II.3 Alur pengenalan emosi dengan FER digabungkan dengan analisis sentimen (Cui dkk., 2021)

Penggunaan multimodal dalam deteksi di lingkungan pendidikan pernah dilakukan peneliti. Orosoo dkk., (2024) menggunakan dua kumpulan data utama dalam penelitiannya yaitu *Multimodal Emotion Lines Dataset* (MELD) dan *EnterFace'05*. MELD terdiri dari percakapan dalam serial televisi *Friends*, yang mencakup modalitas audio, visual, dan teks, memungkinkan analisis emosi yang komprehensif. *Dataset EnterFace'05* berisi urutan video yang menyediakan

informasi penting untuk pengenalan emosi, menunjukkan bahwa data video dikumpulkan untuk analisis emosi. Studi ini juga menekankan penggunaan *Natural Language Processing* (NLP) untuk mengonversi suara menjadi teks, yang menunjukkan bahwa data audio dikumpulkan dan diproses untuk meningkatkan akurasi pengenalan emosi. Secara keseluruhan, pengumpulan data melibatkan modalitas audio dan visual, kemungkinan melalui rekaman video dan input audio.

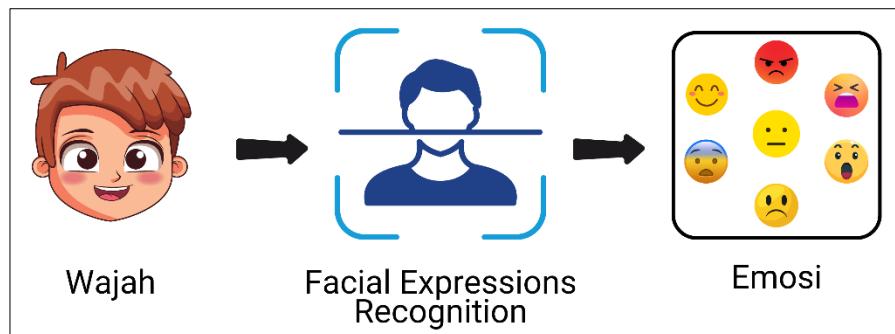
Hal berbeda dilakukan oleh peneliti Roy dkk., (2023). Mereka mengekstrak video menjadi text yang merupakan hasil konversi dari audio, lalu audio itu sendiri diekstrak terpisah dan dijadikan fitur MFCC kemudian diproses dengan *deep learning*, lalu *frame* dari video diproses dengan model *yolo*, yang pada akhirnya menggabungkan semua skor dari sisi teks, audio, dan video menjadi nilai akhir untuk menetukan *engagement* dalam media pembelajaran daring. Roy dkk., (2023) mengumpulkan data uji dengan menerapkan model di server *cloud*, di mana pengguna diperlihatkan video dari kuliah video sumber terbuka berdasarkan minat mereka. Setelah menyelesaikan video, pengguna diminta untuk memberikan peringkat untuk *engagement* mereka pada skala 1 hingga 10. Metode ini memungkinkan pengumpulan data *engagement* pengguna yang terkait langsung dengan konten video yang dilihat.

Walaupun pengenalan emosi dengan multimodal berpotensi meningkatkan akurasi pengenalan emosi (Hunukumbure dkk., 2021), namun pengenalan emosi dengan multimodal memiliki beberapa tantangan, salah satunya adalah masalah sinkronisasi data dari berbagai sumber dan pengembangan model yang harus mampu menangani data yang heterogen (Das dkk., 2024), dan juga pasti memiliki beban komputasi yang lebih tinggi daripada pengenalan emosi dengan *unimodal* (Khan dkk., 2024).

II.4 *Facial Expressions Recognition* di Lingkungan Pendidikan Daring

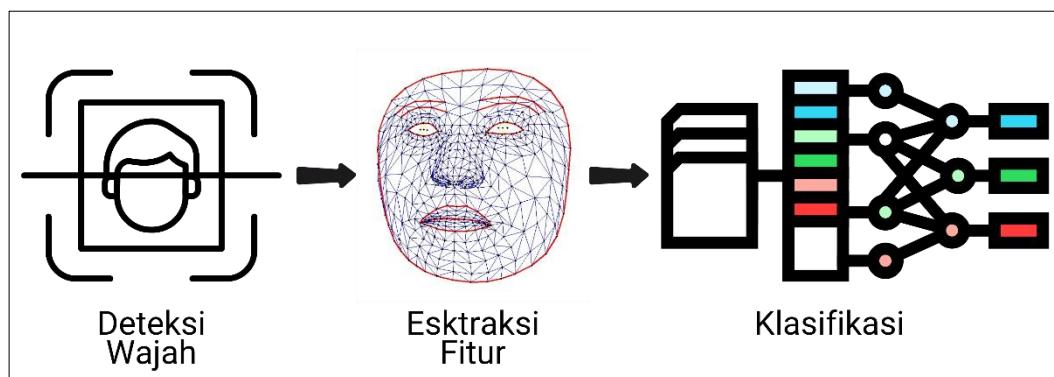
Facial Expression Recognition (FER) adalah bidang penelitian dalam computer vision dan kecerdasan buatan yang berfokus pada identifikasi dan klasifikasi ekspresi wajah manusia berdasarkan fitur visual. Teknologi ini memanfaatkan algoritma *machine learning* dan *deep learning* untuk mengenali berbagai emosi seperti bahagia, sedih, marah, terkejut, takut, jijik, dan netral. Proses utama dalam

sistem FER mencakup beberapa tahap, yaitu deteksi wajah, ekstraksi fitur, dan klasifikasi ekspresi (gambar II.5).



Gambar II.4 Sistem pengenalan emosi wajah

Deteksi wajah dilakukan dengan metode seperti *Haarcascade* (Sukumaran dan Manoharan, 2024) atau MTCNN (Zeng dkk., 2021) untuk menemukan posisi wajah dalam gambar. Setelah wajah terdeteksi, fitur-fitur penting seperti bentuk mata, mulut, dan alis diekstraksi menggunakan teknik seperti *Local Binary Pattern* (LBP) dan *Convolutional Neural Networks* (Sowjanya & Krithiga, 2024). Tahap akhir adalah klasifikasi ekspresi, yang biasanya dilakukan menggunakan model seperti *Random Forest* (Shobana & Kumar, 2021), atau jaringan saraf dalam seperti ResNet50 (Aly dkk., 2023) dan VGG19 (Gupta dkk., 2023c).



Gambar II.5 Proses utama *facial expressions recognition*

Dalam lingkungan pendidikan daring, FER dapat digunakan untuk menganalisis ekspresi wajah siswa guna memahami tingkat *engagement*, kebingungan, atau kejemuhan mereka selama pembelajaran. Dengan menggunakan kamera dan algoritma berbasis *deep learning*, sistem FER dapat mendeteksi emosi siswa secara *real-time*, memungkinkan pengajar untuk menyesuaikan metode pengajaran agar lebih interaktif dan efektif. Teknologi ini juga dapat diterapkan dalam sistem *e-*

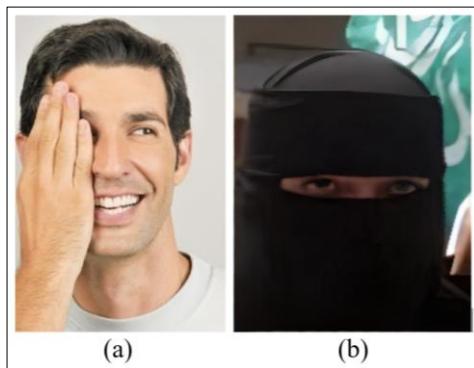
learning adaptif, di mana materi pelajaran secara otomatis disesuaikan berdasarkan respons emosional siswa. Selain itu, FER dapat membantu dalam evaluasi pengalaman belajar, dan mendukung pengajar dalam memberikan umpan balik yang lebih personal, serta meningkatkan kualitas interaksi dalam pembelajaran jarak jauh.

Meskipun banyak kemajuan telah dicapai, FER masih menghadapi beberapa tantangan, seperti perbedaan budaya dalam ekspresi emosi (Mattioli & Cabitza, 2024), variasi pencahayaan, serta oklusi atau halangan pada wajah. Salah satu tantangan terbesar adalah menangani wajah yang terhalang oleh tangan, kacamata, atau masker, yang dapat mengurangi akurasi pengenalan emosi wajah (Li dan Deng, 2022).

II.5 Oklusi pada Pengenalan Ekspresi Wajah

Oklusi adalah kondisi di mana suatu objek atau bagian tertentu tertutupi oleh objek lain sehingga tidak terlihat sepenuhnya. Dalam konteks pengenalan ekspresi wajah, oklusi terjadi ketika bagian wajah tertutup oleh elemen seperti kacamata, masker, tangan, rambut, atau benda lainnya. Dalam banyak kasus, metode pengenalan ekspresi wajah berbasis *deep learning* mengalami penurunan akurasi ketika menghadapi kondisi oklusi yang signifikan (Wickline dkk., 2025).

Oklusi dapat bersifat parsial atau total, tergantung pada seberapa besar bagian wajah yang tertutup. Oklusi parsial terjadi ketika hanya sebagian kecil wajah yang tertutup, seperti mata yang tertutup kacamata hitam atau mulut yang tertutup masker (Wickline dkk., 2025). Sementara itu, oklusi total dapat terjadi ketika hampir seluruh wajah tertutupi (C. Zhu dkk., 2022). Untuk mengatasi tantangan ini, berbagai teknik seperti augmentasi data, deteksi oklusi, dan rekonstruksi fitur wajah digunakan dalam sistem pengenalan ekspresi berbasis kecerdasan buatan.



Gambar II.6 (a) Keadaan wajah dengan oklusi *partial* oleh tangan, (b) keadaan wajah dengan oklusi total (Jain & Learned-Miller, 2010)

Untuk mengatasi permasalahan ini, berbagai pendekatan telah dikembangkan, seperti augmentasi data (Malakar dkk., 2024), penggunaan model yang lebih robust terhadap oklusi (Kim & Lee, 2023), serta teknik deteksi dan rekonstruksi wajah (N. Zhang dkk., 2023). Penelitian tentang rekonstruksi wajah lainnya juga pernah dilakukan, seperti yang dilakukan oleh K. Zhu dkk., (2023). Mereka melakukan rekonstruksi wajah dalam bentuk 3D. Menghapus objek oklusi di wajah dan menggunakan *template* (database bagian-bagian wajah) dalam bentuk 3D, berbagai jenis path wajah sebagai media untuk mengenerate (mengisi) bagian oklusi yang dihapus. Penelitian ini berhasil menghapus dan menggenerate wajah oklusi menjadi wajah tanpa oklusi, namun sayangnya tidak ada pengujian lebih lanjut terhadap nilai akurasi klasifikasi emosi wajah. Walaupun demikian metode yang dilakukan Kaifeng Zhu dkk, bisa dijadikan referensi dalam hal pendekatan yang tepat dalam menghapus oklusi pada wajah.

Ada cara lain untuk mengatasi oklusi pada wajah yaitu pada bagian *pre-processing* Peneliti Cao dkk., (2021) mencoba untuk mengenali emosi dari wajah yang *non-frontal* dengan berbasis gambar 3D. Mereka melakukannya dengan mengambil setengah bagian wajah dan menggabungkannya dalam bentuk 3D kemudian melakukan penguatan bentuk otot disekitar hidung untuk dapat meningkatkan akurasi pengenalan emosi dari wajah 3D. Hasilnya ada kenaikan akurasi terhadap beberapa emosi dasar setelah melakukan proses penguatan otot muka disekitar hidung.

Walaupun bentuk data wajah dalam 3D dapat mengandung lebih banyak informasi dan meningkatkan akurasi pengenalan emosi wajah (Cao dkk., 2021), namun mengubah gambar 2D ke bentuk 3D dapat merubah informasi asli dari data wajahnya seperti bentuk otot muka yang tidak sesuai dengan data aslinya. Data gambar dalam bentuk 3D juga memiliki kompleksitas komputasi yang lebih tinggi gambar 2D sehingga memerlukan waktu proses yang lebih lama dari gambar 2D (Morar dkk., 2017). Apa yang dilakukan oleh Cao dkk., (2021) tersebut menginspirasi penulis untuk mengusulkan sebuah metode yang sama namun dalam bentuk gambar 2D dengan harapan hasil akurasinya bisa menyamai dengan bentuk gambar 3D namun dengan kompleksitas pemrosesan yang lebih sederhana sehingga bisa diterapkan pada resource dengan spesifikasi rendah.

II.6 Klasifikasi Emosi dalam Deteksi *Engagement* Siswa

Berdasarkan beberapa literatur, dalam modalitas visual para peneliti melakukan berbagai macam cara untuk menilai *engagement* siswa dalam pembelajaran, ada yang langsung dengan menilai terhadap *eye gaze* dan *head pose* (Gupta dkk., 2023a), namun tidak sedikit juga yang menggunakan data emosi siswa yang didapat dari proses *emotion recognition*. *Emotion Recognition* adalah proses dimana sistem menganalisis ekspresi wajah untuk mengklasifikasikan emosi seseorang. Beberapa peneliti pada studi ini menggunakan nilai probabilitas setiap emosi pada siswa untuk menentukan tingkat engagement dari siswa dalam pembelajaran yang dilakukan. Untuk mendapatkan nilai emosi tersebut beberapa peneliti menggunakan teknik *machine learning* konvensional, seperti *Random Forest* (Shobana dan Kumar, 2021), atau bahkan metode *deep learning*, seperti CNN (Bao dkk., 2024b) dan ResNet-50 (Aly dkk., 2023). Model ini dilatih pada *dataset* ekspresi wajah seperti FER2013, AffectNet, atau CK+ yang mengandung berbagai kategori emosi seperti marah, bahagia, sedih, netral, takut, jijik, dan terkejut. Setelah wajah terdeteksi, fitur-fitur penting diekstraksi dan digunakan untuk melakukan klasifikasi emosi. Keberhasilan Emotion Detection sangat bergantung pada kualitas data, teknik augmentasi, serta arsitektur model yang digunakan untuk menangkap perbedaan ekspresi wajah secara akurat.

II.6.1 Machine Learning (ML)

Kelebihan ML dalam *automatic emotion recognition* (AER) terletak pada efisiensi komputasi, interpretabilitas, dan fleksibilitas dalam berbagai jenis data. Dibandingkan dengan *deep learning* (DL) yang membutuhkan jumlah data besar dan daya komputasi tinggi, ML dapat bekerja dengan baik bahkan pada *dataset* terbatas dan dengan *hardware* yang lebih sederhana (Ülgen Sönmez dan Varol, 2024). Adapun gambaran umum pada tugas klasifikasi menggunakan *machine learning* dapat dilihat pada gambar II.7.



Gambar II.7 Gambaran umum alur *machine learning*

Pendekatan ML Konvensional pernah dilakukan peneliti untuk mengklasifikasikan tingkat pemahaman siswa dalam sebuah tes dalam platform digital. Salah satunya adalah peneliti Shobana dan Kumar, (2021) yang memanfaatkan ML untuk menganalisis perilaku non-verbal dalam menilai tingkat akuisisi pengetahuan peserta didik selama proses pengajaran kuis dalam platform digital (*I-Quiz*). Model yang digunakan adalah *Random Forest Classifier* yang dilatih menggunakan *dataset* berisi perilaku non-verbal serta skor pertanyaan-jawaban untuk memprediksi tingkat pemahaman peserta secara akurat. Sistem *I-Quiz* dirancang untuk menangkap perilaku non-verbal melalui kamera menghadap ke depan yang merekam ekspresi wajah, pergerakan mata, gestur, serta postur tubuh. Dari hasil rekaman tersebut, sistem menghasilkan fitur yang mencakup berbagai parameter seperti emosi, tingkat kedipan, skor kepercayaan diri, serta waktu yang dibutuhkan dalam menjawab pertanyaan. Dengan menggunakan fitur tersebut Shobana dan Kumar, (2021) melakukan klasifikasi nilai tingkat akuisisi pengetahuan peserta didik.

II.6.2 Deep Learning (DL)

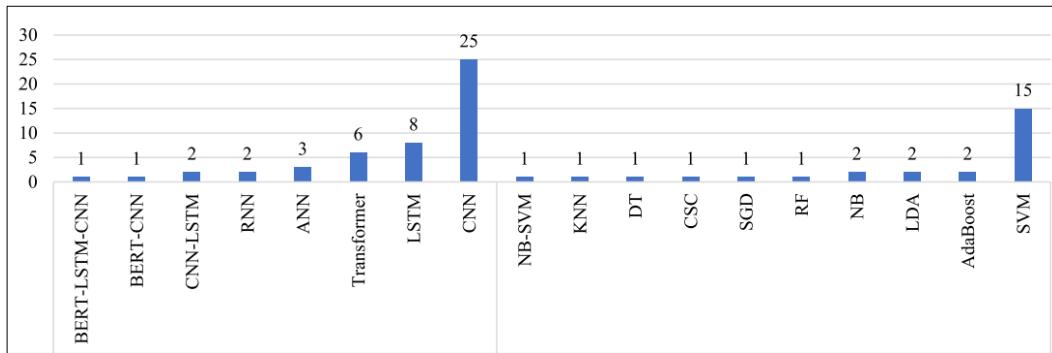
Kelebihan *Deep Learning* (DL) dalam *automatic emotion recognition* (AER) terletak pada kemampuannya mengekstrak fitur secara otomatis, menangkap pola kompleks dalam data, dan menghasilkan akurasi tinggi dibandingkan metode ML

klasik. Dalam penerapan FER, DL dapat menganalisis ekspresi wajah secara lebih mendalam, tanpa memerlukan proses ekstraksi fitur manual (S. Zhang dkk., 2024). Meskipun membutuhkan daya komputasi yang lebih besar daripada ML klasik, DL menawarkan fleksibilitas, skalabilitas, dan kemampuan generalisasi yang lebih baik, menjadikannya pilihan utama dalam banyak aplikasi AER berbasis gambar, audio, dan multimodal (Trigka & Dritsas, 2025).

Tabel II.1 Studi terkait yang menggunakan metode *deep learning* pada deteksi emosi dengan modalitas visual pada lingkungan pendidikan

Metode	Referensi	Matrik Evaluasi
CNN	(Bao dkk., 2024b)	Accuracy (81.9 %)
	(Bhardwaj dkk., 2021)	Accuracy (93.6 %)
	(Sowjanya dan Krithiga, 2024)	Accuracy (94.8%)
	(Rathod dkk., 2022)	Accuracy (90.98%)
	(Sukumaran dan Manoharan, 2024)	Accuracy (>80%)
DenseAttNet	(Mehta dkk., 2022)	Accuracy (>80%)
ResNet50	(Aly dkk., 2023)	Accuracy (97.72%)
VGG19	(Gupta dkk., 2023c)	Accuracy (92.58%)
YoloV5	(Trabelsi dkk., 2023)	Accuracy (76%)
Neural Architecture Search (NAS)	(Harb dkk., 2023)	Accuracy (86.29%)
Dual Path Stacked Attention Network (DPSAN)	(H. Zhu dkk., 2022)	Accuracy (93.26%)
Multi-Frame Transfer Learning (MFTL)	(Pordoy dkk., 2024)	Accuracy (87.8%)

Hasil rangkuman studi pada tabel II.1 menunjukkan bahwa metode CNN adalah metode yang paling banyak digunakan dalam klasifikasi emosi khususnya dalam penerapannya pada bidang pendidikan daring. Hal ini juga di dukung oleh data hasil penelitian yang dilakukan oleh Yu dkk., (2024) yang dapat dilihat pada gambar II.8.



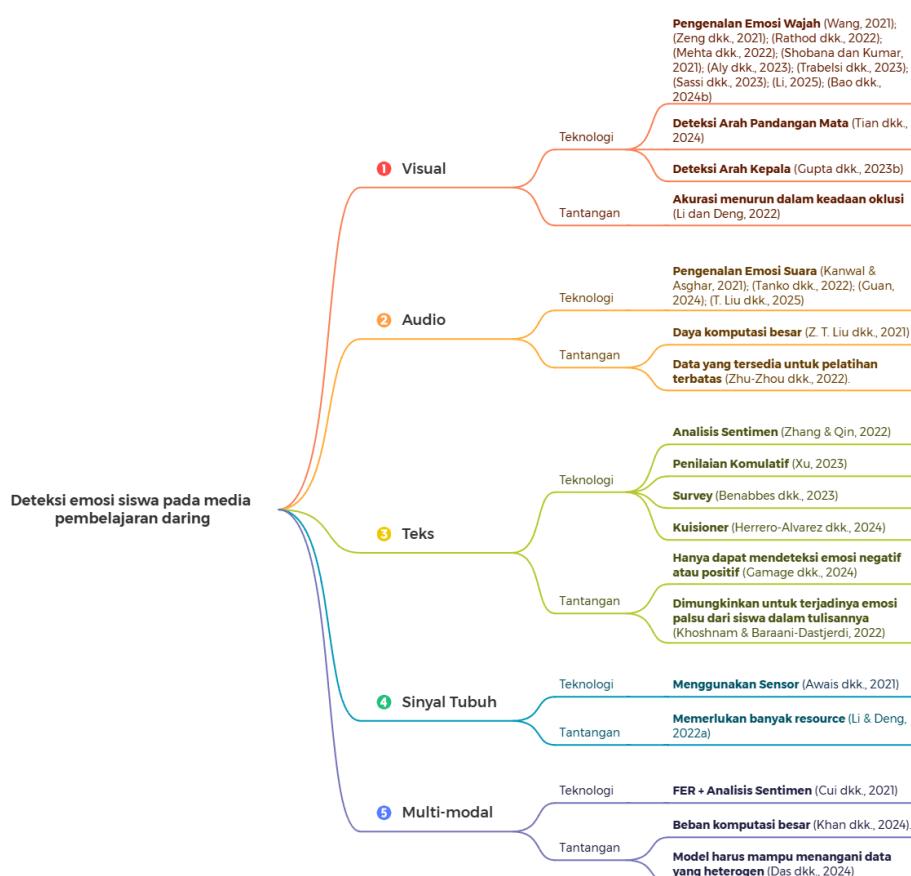
Gambar II.8 Distribusi penggunaan metode *machine learning* dan *deep learning* dalam rentang tahun 2010 – 2024 (Yu dkk., 2024)

II.6.3 Kombinasi *Machine Learning* Klasik dan *Deep Learning*

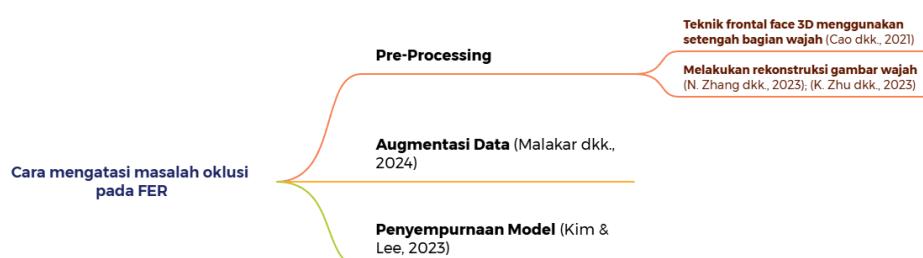
Penggabungan metode ML Klasik dan DL juga mungkin dilakukan untuk mendeteksi emosi wajah siswa. Sassi dkk., (2023) menggabungkan CNN dan Random Forest untuk menganalisis secara *real-time* emosi siswa saat pembelajaran daring. Dalam implementasinya, Sassi dkk., (2023) menggunakan model CNN-RF untuk pengenalan emosi, yang menunjukkan kinerja lebih tinggi serta kompleksitas yang lebih rendah dibandingkan dengan metode konvensional. Model CNN-RF yang diusulkan menunjukkan kinerja yang unggul dengan mencapai akurasi sebesar 71,86%, presisi 70,56%, *recall* 74,35%, dan skor F1 sebesar 72,09%. Hasil ini lebih tinggi dibandingkan dengan tolok ukur yang ada, menegaskan efektivitas pendekatan yang digunakan. Akurasi yang tinggi menunjukkan bahwa model mampu mengklasifikasikan sampel dengan baik, sementara presisi yang tinggi mengindikasikan rendahnya tingkat prediksi positif palsu. Keunggulan lain dari model ini terletak pada arsitektur yang efisien, yang memungkinkan peningkatan akurasi tanpa mengorbankan kinerja komputasi. Dengan hanya 5,17 juta parameter, model ini tetap ringan dan optimal untuk aplikasi dunia nyata, terutama dalam lingkungan dengan sumber daya yang terbatas. Evaluasi model dilakukan menggunakan *dataset* FER2013, yang berisi gambar wajah dengan berbagai ekspresi emosi. Selama proses pelatihan, teknik augmentasi data diterapkan untuk memperkaya kumpulan data, sehingga meningkatkan generalisasi model dalam mengenali emosi dari berbagai variasi gambar.

II.7 Peta Literatur

Berdasarkan beberapa literatur yang telah dibahas, penulis melakukan pemetaan tentang referensi penelitian terdahulu beserta tantangannya yang berkaitan dengan pilihan-pilihan yang ada untuk mendeteksi emosi siswa pada pembelajaran daring, hasilnya dapat dilihat pada gambar II.10. Selain itu, penulis juga membuat peta literatur tentang penelitian yang sudah dilakukan untuk mengatasi masalah wajah yang teroklusi dalam penerapan FER yang dapat dilihat pada gambar II.11.



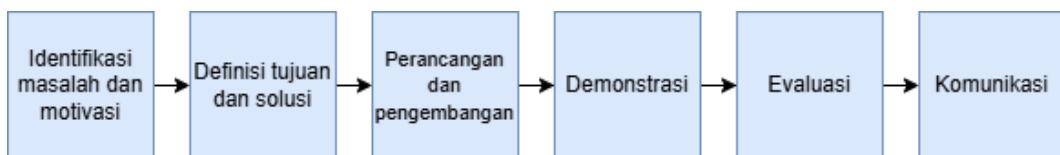
Gambar II.9 Peta literatur deteksi emosi siswa pada media pembelajaran daring



Gambar II.10 Peta literatur tentang cara mengatasi masalah oklusi pada FER

BAB III Metodologi Penelitian

Semua metodologi yang akan digunakan dalam penelitian dibahas dalam bab ini. Gambar III.1 menunjukkan tahapan penelitian *Design Science Research Methodology* (DSRM) yang akan diterapkan dalam penelitian ini.



Gambar III.1 *Design Science Research Methodology*

Gambar III.1 menunjukkan bahwa DSRM terdiri dari enam tahapan penelitian: identifikasi masalah, definisi tujuan dan solusi, perancangan dan pengembangan, demonstrasi, evaluasi, dan komunikasi. Subbab berikutnya memberikan penjelasan tentang setiap langkah yang harus dilakukan.

III.1 Identifikasi Masalah dan Motivasi

Pada tahap identifikasi masalah dan motivasi, fokus akan berada pada identifikasi masalah dan motivasi untuk melakukan penelitian. Tabel III.1 menunjukkan langkah-langkah *input*, proses, dan *output* pada tahap identifikasi masalah dan motivasi.

Tabel III.1 Tahap identifikasi masalah dan motivasi

<i>Input</i>	- Riset penelitian - Dokumen ilmiah
Proses	Kajian literatur
<i>Output</i>	- Informasi mekanisme umpan balik pada pembelajaran daring - Alasan pemilihan FER sebagai pendekatan umpan balik pada pembelajaran daring

Berdasarkan tabel III.1, *input* pada tahap ini diperoleh dengan melakukan pengkajian literatur terkait topik yang akan dibahas, yaitu penerapan mekanisme umpan balik pada LMS. Masalah dalam penelitian ini diperoleh dari telaahan dokumen strategi pengembangan institusi pendidikan yang menekankan pentingnya penerapan konsep smart campus, salah satunya melalui penggunaan *Learning Management System* (LMS) sebagai pendukung pembelajaran daring.

Berdasarkan wawancara langsung dengan dosen dan dipadukan dengan studi literatur, ditemukan bahwa LMS yang digunakan masih memiliki kekurangan yaitu tidak tersedianya mekanisme umpan balik yang dapat menilai tingkat *engagement* (*engagement*) serta emosi siswa selama pembelajaran daring, khususnya pada pembelajaran asynchronous. Kekurangan seperti minimnya interaksi, keterlambatan respon, dan kurangnya fitur pendukung monitoring *engagement* siswa menjadi tantangan yang perlu diatasi dalam pengembangan media pembelajaran berbasis LMS.

Berdasarkan beberapa literatur menyebutkan bahwa penerapan deteksi emosi siswa pada pembelajaran daring bisa melalui pendekatan *facial expression recognition* (FER), *sound emotion recognition* (SER), *analysis sentiment*, atau gabungan beberapa pendekatan tersebut. Memperhitungkan kelebihan dan kelemahan masing-masing pendekatan tersebut, disimpulkan bahwa pendekatan facial expression recognition adalah solusi yang paling logis untuk diterapkan pada LMS. Namun penerapan FER pada lingkungan pendidikan khususnya pada pembelajaran daring masing memiliki kekurangan pada kondisi wajah yang teroklusi, dimana kondisi tersebut lazim terjadi saat pembelajaran daring. Contohnya adalah saat siswa menyangga kepalaanya dengan tangan karena faktor kelelahan atau terlalu fokus, maupun faktor lainnya. Kondisi wajah yang teroklusi tersebut dapat menurunkan tingkat probabilitas dari suatu emosi, sehingga diperlukan penelitian lebih lanjut untuk dapat mengatasi masalah wajah teroklusi pada penerapan FER di LMS.

Berdasarkan pengkajian data pada tahapan *input*, didapatkan *output* berupa permasalahan yang dapat diidentifikasi, yaitu diperlukannya pengembangan LMS yang memiliki fitur untuk dapat menilai *engagement* dan emosi siswa yang *robust* terhadap wajah yang teroklusi, khususnya wajah yang tertutupi oleh tangan sebagian, sehingga didapatkan emosi siswa yang lebih akurat, dan hasilnya dapat membantu pengajar dalam mengevaluasi materi maupun siswa yang bersangkutan dengan lebih tepat.

III.2 Definisi Tujuan dan Solusi

Fokus tahap kedua DSRM adalah mendefinisikan tujuan dan solusi berdasarkan masalah yang telah diidentifikasi sebelumnya. Identifikasi tujuan penelitian kemudian memberikan solusi untuk mengatasi masalah yang ada. Pada tahap tujuan dan solusi, Tabel III.2 menunjukkan tahapan *input*, proses, dan *output*.

Tabel III.2 Tahap tujuan dan solusi

<i>Input</i>	<ul style="list-style-type: none"> - Informasi mekanisme umpan balik pada pembelajaran daring - Alasan pemilihan FER sebagai pendekatan umpan balik pada pembelajaran daring
Proses	<ul style="list-style-type: none"> - Eksplorasi cara untuk menangani keadaan oklusi pada FER - Eksplorasi model FER yang tersedia dan sesuai
<i>Output</i>	Rancangan solusi

Berdasarkan hasil pada tahap identifikasi masalah dan motivasi, hal selanjutnya yang dilakukan adalah melakukan eksplorasi cara untuk menangani keadaan oklusi pada FER, dan eksplorasi model FER yang tersedia dan sesuai. Dari studi literatur yang telah dilakukan dan menghasilkan beberapa kemungkinan cara yang bisa diterapkan dalam menangani oklusi pada wajah, penulis memilih cara pada bagian *pre-processing* dengan harapan improvisasi pada bagian ini tidak akan menambah beban komputasi yang lebih banyak jika dibandingkan dengan melakukan improvisasi pada bagian pembuatan atau fine-tuning pada model klasifikasinya.

III.3 Perancangan dan Pengembangan

Tahap ketiga dalam DSRM adalah mendefinisikan perancangan dan pengembangan berdasarkan rancangan solusi yang telah didefinisikan sebelumnya. Pada tahap ini akan berfokus pada pengembangan metode *pre-processing* untuk mengatasi masalah oklusi wajah oleh tangan pada penerapan FER di LMS. Tabel III.3 merupakan tahapan *input*, process, dan *output* pada tahap perancangan dan pengembangan.

Tabel III.3 Tahap perancangan dan pengembangan

<i>Input</i>	Rancangan solusi
Proses	Pengembangan metode <i>pre-processing</i> untuk mengatasi masalah oklusi wajah oleh tangan pada penerapan FER di LMS
<i>Output</i>	Metode <i>pre-processing</i> dapat mengatasi masalah oklusi wajah oleh tangan dalam penerapan FER

Berdasarkan tabel III.3, *input* pada tahap ini adalah berasal dari rancangan solusi pada bagian sebelumnya. Selanjutnya akan dilakukan pengembangan metode *pre-processing* untuk mengatasi masalah oklusi wajah oleh tangan pada penerapan FER di LMS. Pengembangan metode tersebut akan dilakukan dengan metode *design thinking*.

III.4 Demonstrasi

Tahap keempat dalam metodologi DSRM adalah demonstrasi. Pada tahap ini akan berfokus pada uji coba terhadap model yang telah dikembangkan. Pada Tabel III.5 merupakan *input*, proses, dan *output* pada tahap demonstrasi.

Tabel III.4 Tahapan demonstrasi

<i>Input</i>	Metode <i>pre-processing</i> dapat mengatasi masalah oklusi wajah oleh tangan dalam penerapan FER
Proses	Uji coba metode yang telah dikembangkan
<i>Output</i>	Data emosi wajah siswa

Berdasarkan Tabel III.5, input pada tahapan ini yaitu metode *pre-processing* yang telah dikembangkan pada hasil tahapan perancangan dan pengembangan. Berdasarkan input tersebut, dilakukan uji coba terhadap metode yang telah dibangun. Hasil dari tahapan demonstrasi adalah mendapatkan data *engagement* siswa yang dapat digunakan sebagai bahan evaluasi oleh pengajar ketika melakukan pembelajaran daring menggunakan LMS.

III.5 Evaluasi

Tahap kelima dari DSRM yaitu tahap evaluasi. Pada tahap ini akan berfokus pada pengukuran performa dari model yang telah dikembangkan. Pada Tabel III.6 merupakan tahapan *input*, proses, dan *output* pada tahap evaluasi.

Tabel III.5 Tahap evaluasi

<i>Input</i>	Data emosi wajah siswa
Proses	Evaluasi probabilitas emosi wajah siswa
<i>Output</i>	Persentase pengaruh terhadap probabilitas emosi wajah siswa setelah menerapkan metode usulan

Proses pada tabel III.6 adalah mengevaluasi data emosi wajah siswa hasil proses demonstrasi untuk mengetahui perubahan probabilitas emosi setelah menerapkan

metode usulan. Evaluasi akan dilakukan dengan membandingkan probabilitas emosi pada data wajah dengan keadaan *baseline* dan data wajah setelah diterapkan proses *pre-processing* dengan metode usulan.

III.6 Komunikasi

Tahap terakhir dalam DSRM adalah komunikasi. Pada tahap ini akan berfokus pada dokumentasi hasil yang kemudian dikomunikasikan dalam bentuk tulisan ilmiah. Tabel III.7 merupakan tahapan *input*, proses, dan *output* pada tahap komunikasi.

Tabel III.6 Tahap komunikasi

<i>Input</i>	- Rancangan solusi - Metode preprocessing data wajah dengan memanfaatkan bagian setengah wajah - Hasil evaluasi perubahan probabilitas emosi sebelum dan sesudah menerapkan metode usulan
Proses	Penyusunan dokumen ilmiah
<i>Output</i>	- Tesis - Paper

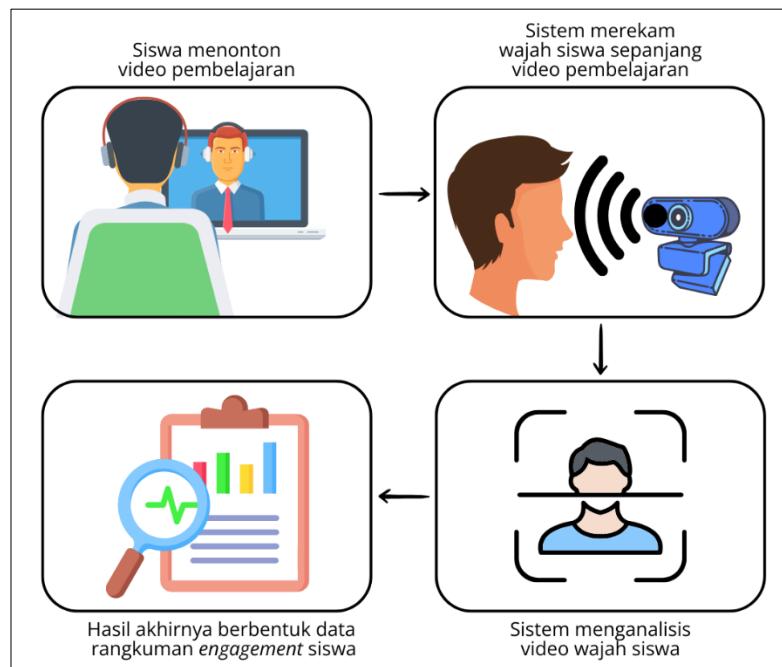
Berdasarkan Tabel III.7, *input* pada tahap evaluasi diperoleh dari seluruh tahapan sebelumnya yang kemudian akan dilakukan penyusunan dokumen ilmiah yang menghasilkan dokumen tesis dan publikasi *paper*.

BAB IV Analisis dan Desain Awal

Bagian ini akan menjelaskan gambaran penelitian yang akan dilakukan. Pada bagian ini juga akan dijelaskan lingkungan yang akan di analisis dan kondisi-kondisi kapan FER akan digunakan untuk menentukan *engagement* dari siswa.

IV.1 Lingkungan Penerapan

Lingkungan penerapan penelitian ini adalah lingkungan pembelajaran daring dengan memanfaatkan *platform learning management system*. Lalu, kondisi penerapan *facial expressions recognition* pada *learning management system* adalah khusus pada kondisi pembelajaran *asynchronous* yang mewajibkan siswa untuk menonton video rekaman pembelajaran yang diberikan oleh pemateri seperti apa yang tertera pada gambar IV.1



Gambar IV.1 Alur penerapan mekanisme umpan balik pada LMS

IV.2 Analisis Kebutuhan Sistem

Dalam analisis kebutuhan sistem ini, akan dibahas secara mendalam mengenai kebutuhan fungsional, non-fungsional, serta kebutuhan pengguna yang mendasari integrasi metode usulan dengan *Learning Management System* (LMS), guna

memastikan pemantauan emosi siswa yang efektif dan memberikan umpan balik yang konstruktif bagi pengajar dalam proses pembelajaran *asynchronous*.

IV.2.1 Kebutuhan Fungsional

Kebutuhan fungsional sistem ini mencakup serangkaian fitur dan fungsi yang harus ada untuk memastikan sistem dapat berjalan sesuai dengan tujuan yang diinginkan, mendukung proses operasional, dan memberikan hasil yang diharapkan. Adapun kebutuhan fungsional dari sistem yang akan dibuat adalah:

1. Pengenalan ekspresi wajah (FER), dimana sistem harus dapat mengenali ekspresi wajah siswa bahkan ketika wajah tertutup sebagian oleh tangan atau miring dengan penerapan metode yang memanfaatkan setengah bagian wajah yang kemudian digabungkan untuk membentuk wajah yang seolah-olah menghadap ke depan.
2. Kesimpulan *Engagement*, berdasarkan hasil dari FER, sistem harus dapat memberikan analisis *engagement* siswa, yang dapat digunakan oleh pengajar untuk menilai seberapa baik tingkat *engagement* siswa dalam pembelajaran daring.
3. Integrasi dengan *Learning Management System* (LMS), yaitu Sistem FER harus terintegrasi dengan platform LMS yang digunakan untuk pembelajaran daring, memungkinkan pemantauan emosi tepat setelah sesi pembelajaran berakhir dan memberikan umpan balik kepada pengajar.
4. Penyimpanan dan pengelolaan data, dimana data emosi yang terkumpul dari setiap sesi pembelajaran harus disimpan dan dikelola dengan baik, memungkinkan untuk analisis jangka panjang mengenai pola *engagement* siswa.
5. Umpan balik, dimana sistem harus mampu memberikan umpan balik kepada pengajar berdasarkan data siswa tentang tingkat *engagement* dari data hasil analisis emosi yang dilakukan.

IV.2.2 Kebutuhan Non-Fungsional

Kebutuhan non-fungsional sistem ini mencakup aspek-aspek yang mendukung kinerja, dan kualitas pengalaman pengguna, seperti tingkat akurasi yang harus lebih tinggi saat menerapkan metode usulan atau minimal menyamai akurasi model yang digunakan, serta kecepatan pemrosesan untuk menjaga kualitas pengalaman pengguna saat menggunakan sistemnya. Adapun kebutuhan non-fungsional dari sistem yang akan dibuat adalah akurasi dan keandalan, dimana sistem harus memiliki tingkat akurasi yang tinggi dalam mendeteksi ekspresi wajah dan mengidentifikasi emosi, bahkan dalam kondisi wajah yang tertutup sebagian atau miring. Kinerja sistem harus konsisten dan dapat diandalkan dalam berbagai posisi wajah.

IV.2.3 Kebutuhan Pengguna

Kebutuhan pengguna sistem ini mencakup kemampuan bagi siswa untuk menggunakan aplikasi dengan mudah dalam memantau ekspresi wajah mereka selama pembelajaran daring, bagi pengajar untuk mengakses analisis emosi dan tingkat *engagement* siswa secara *real-time*, serta bagi administrator untuk mengelola dan memelihara sistem secara efisien, memastikan integrasi yang lancar dengan platform pembelajaran yang ada. Adapun kebutuhan pengguna pada sistem yang dibuat antara lain :

1. Siswa harus dapat menggunakan aplikasi LMS dengan mudah, dimana aplikasi secara otomatis memonitor ekspresi wajah mereka selama sesi pembelajaran dan memberikan umpan balik mengenai tingkat *engagement* mereka.
2. Pengajar harus dapat melihat analisis emosi dan *engagement* siswa dalam bentuk yang mudah dipahami, untuk dapat membuat penyesuaian pada pendekatan pengajaran mereka. Mereka juga harus dapat mengakses statistik yang menunjukkan pola *engagement* siswa selama sesi pembelajaran.

IV.2.4 Kebutuhan Perangkat Keras dan Perangkat Lunak

Sistem harus dapat berjalan dengan baik pada perangkat keras umum yang digunakan siswa dan pengajar, seperti komputer dengan spesifikasi kamera yang cukup untuk mendeteksi ekspresi wajah.

Dari sisi perangkat lunak, sistem harus berbasis pada perangkat lunak yang mendukung pengenalan ekspresi wajah menggunakan teknik *computer vision* dan *machine learning* dan juga harus *cross platform*, sehingga tidak perlu melakukan pengaturan yang memberatkan pengguna.

IV.3 Analisis Desain

Berdasarkan tinjauan literatur yang sudah dilakukan, ada beberapa alternatif solusi yang bisa dilakukan dalam menyelesaikan permasalahan penelitian. Adapun analisisnya dijabarkan dalam tabel IV.1.

Tabel IV.1 Alternatif solusi untuk deteksi emosi siswa

Modalitas	Metode	Kelebihan	Kekurangan
Visual	Pengenalan Emosi Wajah	Model dan <i>dataset</i> publik sudah banyak, mudah untuk didapatkan dan dimodifikasi, pemrosesan relatif lebih ringan	Akurasi turun pada data wajah yang teroklusi
Audio	Pengenalan Emosi Suara	Akurasi pengenalan emosi baik	Beban komputasi sedikit lebih berat
Teks	Analisis Sentimen	Beban komputasi ringan karena hanya memproses teks	Hanya bisa menentukan emosi positif dan negatif, juga seringnya muncul emosi palsu dalam tulisan
Sinyal Tubuh	Jam Tangan Pintar atau perangkat Sensor	Akurasi lebih akurat	Memerlukan <i>resources</i> yang banyak dan tidak murah
Multi-Modal	FER + Analisis Sentimen	Akurasi lebih akurat	Beban komputasi bisa bertambah sehingga bisa menambah waktu pemrosesan

Dihubungkan dengan analisis kebutuhan sistem pada sub bab IV.2, pemilihan modalitas yang tepat untuk dijadikan solusi penelitian adalah modalitas visual

dengan metode pengenalan emosi wajah yang mana pemrosesannya bisa lebih ringan dan murah dibandingkan dengan modalitas lain dengan tetap menjaga akurasi pengenalan emosinya.

Untuk memenuhi kebutuhan minimal perangkat keras dan perangkat lunak sistem sesuai dengan penjelasan pada sub bab IV.2.4, sistem akan dibuat dengan berbasis web menggunakan *flask* yaitu *python* sebagai *back-end* dan *html+css+js* sebagai *front-end*. Penggunaan *flask* akan memudahkan integrasi antara *front-end* dan *back-end* dan bisa diakses melalui sebuah aplikasi *browser* sehingga pengguna tidak perlu melakukan pengaturan apapun dalam menggunakannya.

IV.4 Analisis Wajah

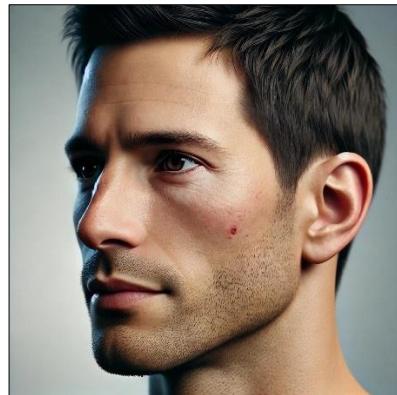
Ada beberapa parameter untuk menentukan *engagement* siswa, pertama adalah arah kepala (*head pose*), lalu arah pandangan mata (*eye gaze*), dan emosi wajah. Sebagai contoh, gambar IV.2 memperlihatkan hasil analisis wajah menggunakan tiga parameter yang digunakan (mata, kepala, emosi) dari sebuah proses ekstraksi wajah baik yang bisa berasal dari gambar maupun video.



Gambar IV.2 Contoh hasil analisis wajah dengan paramater arah pandangan mata, arah kepala, dan emosi wajah

Adapun penggunaan pendekatan FER tidak selalu pada setiap gambar yang di eksekusi, contoh: jika kondisi kepala dan mata sedang menatap ke samping (tidak melihat layar minitor) seperti pada gambar IV.3, maka sistem akan langsung

menyimpulkan bahwa siswa tersebut dalam keadaan tidak fokus tanpa harus dilakukan proses pengenalan emosi wajah.



Gambar IV.3 Contoh kondisi wajah yang arah kepala dan mata tidak menghadap monitor atau menghadap kesamping (gambar dibuat dengan AI)

IV.5 Desain Solusi

Bagian ini akan membahas tentang desain solusi yang diperlukan untuk membuat keseluruhan sistem mulai dari desain solusi umum, desain metode usulan, desain tampilan antarmuka, hingga desain pengujian dalam penelitian ini.

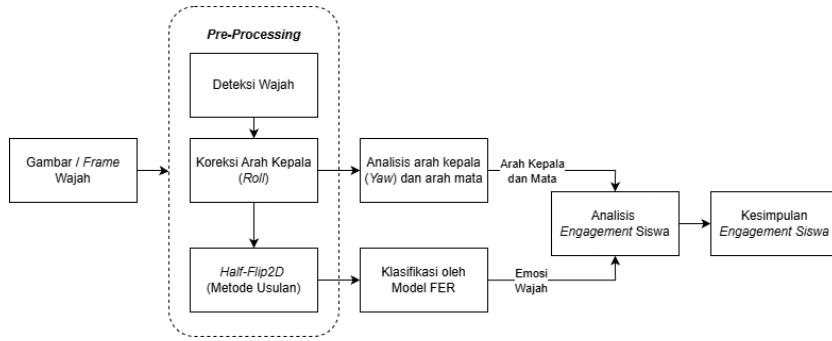
IV.5.1 Desain Solusi Umum

Adapun gambaran umum desain keseluruhan sistem yang dibuat dapat dilihat pada gambar IV.4.



Gambar IV.4 Gambaran umum ajuan solusi sistem

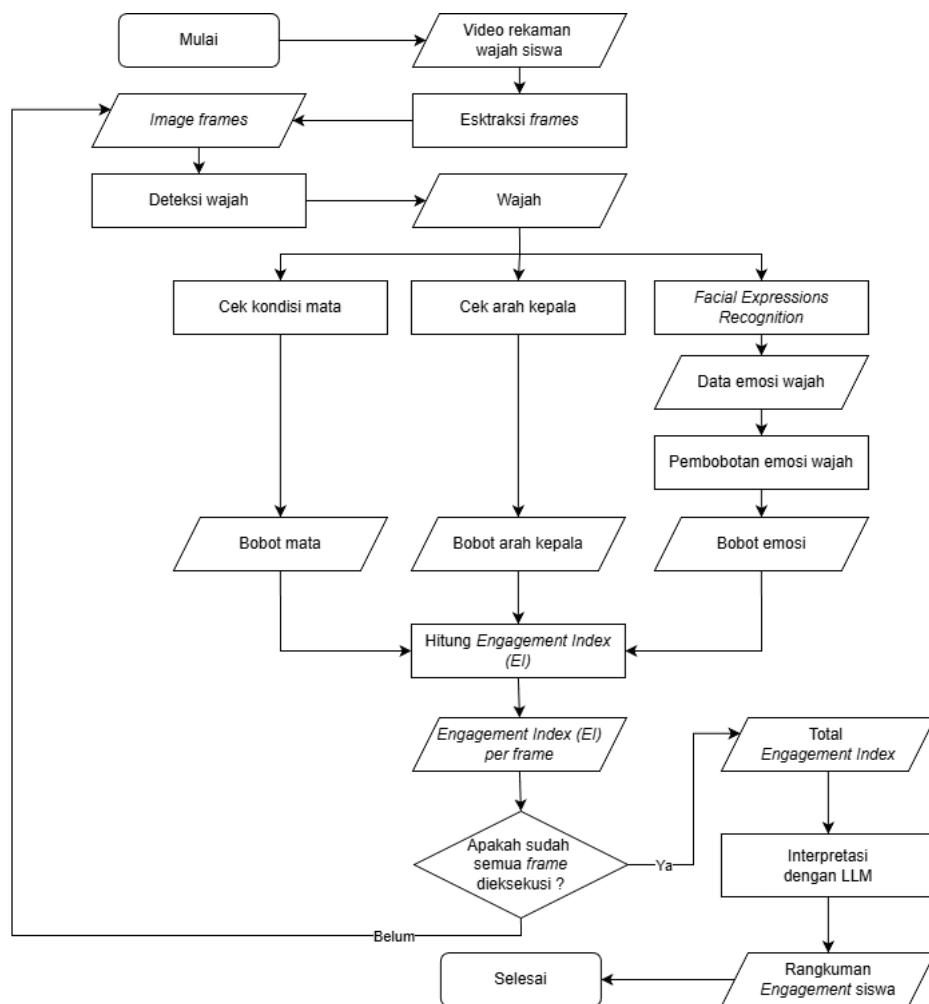
Proses teknis dari ajuan penelitian ini mulai dari proses deteksi wajah, pre-processing, klasifikasi emosi oleh model FER, analisis *engagement*, dan keluaran akhirnya berupa kesimpulan *engagement* siswa. Walaupun keseluruhan sistem akan dibuat sesuai dengan gambar IV.4, namun fokus penelitian ini hanya pada bagian *pre-processing* saja, sehingga pengujinya juga akan berfokus pada implikasi yang dihasilkan dari metode usulan pada bagian *pre-processing*. Adapun detail teknis yang menjadi fokus penelitian ini dijelaskan pada gambar IV.5.



Gambar IV.5 Detail teknis dan fokus penelitian pada bagian *pre-processing*

IV.5.2 Desain Algoritma Program

Pengenalan emosi wajah hanya diterapkan ketika mata dan kepala masih terdeteksi dalam keadaan melihat monitor. Lebih lengkapnya desain algoritma untuk menentukan *engagement* siswa dapat dilihat pada diagram alir gambar IV.6.

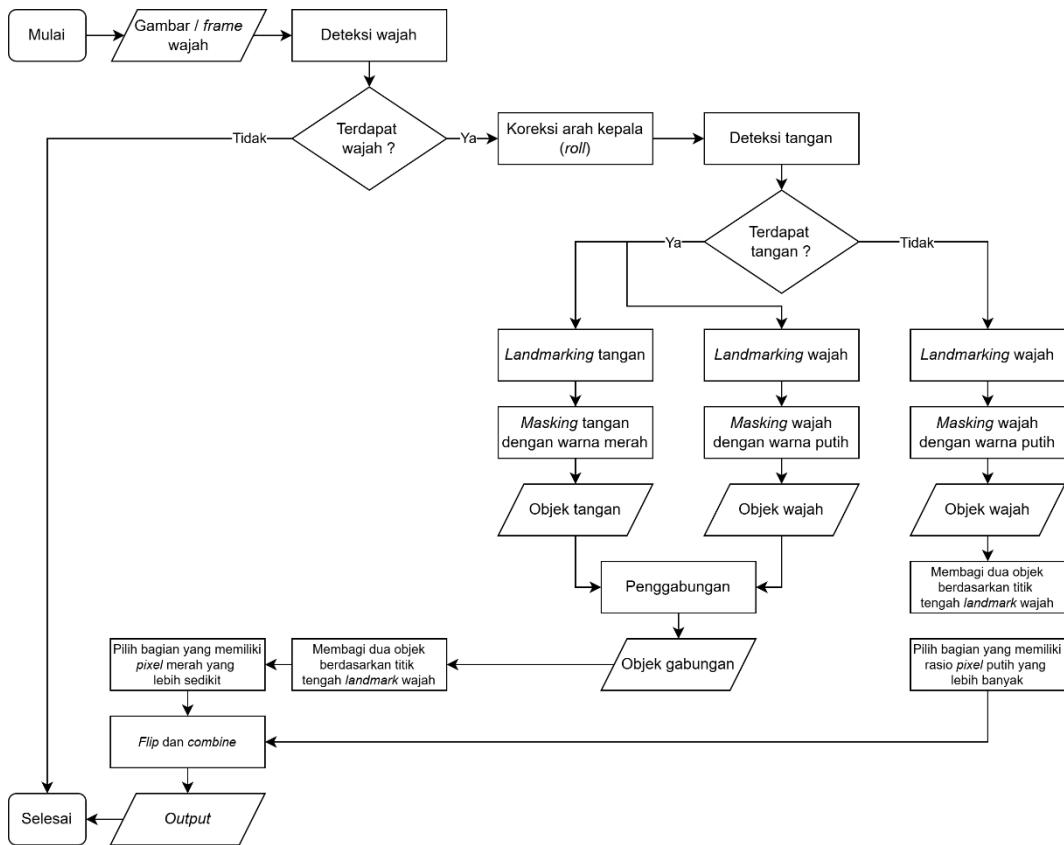


Gambar IV.6 Diagram alir algoritma kondisi penggunaan FER untuk menentukan *engagement* siswa

Penjelasan diagram alir pada gambar IV.6 adalah sebagai berikut:

1. Video rekaman wajah siswa diekstrak dan diproses satu *frame* per detik.
2. Kumpulan *frame* tersebut diproses satu per satu selama panjang video untuk dianalisis.
3. Proses pertama pada satu *frame* adalah dengan melakukan deteksi wajah.
4. Data wajah yang selanjutnya dilakukan proses deteksi arah mata, arah kepala melihat, dan deteksi emosi dengan FER, yang masing-masing hasilnya akan diberikan bobot sesuai dengan literatur (Sukumaran & Manoharan, 2024).
5. Bobot ketika kepala atau arah pandangan melihat lurus ke layar adalah 1, dan ketika tidak melihat lurus ke layar maka bobotnya 0.
6. Bobot ketika mata terbuka adalah 2.5, dan ketika tertutup adalah 0.
7. Bobot emosinya antara lain *angry* 0.1, *disgust* 0.9, *fear* 0.5, *happy* 1.1, *neutral* 1.4, *sad* 0.3, *surprised* 0.7.
8. Bobot pada poin 5, 6, dan 7 dijumlahkan, total hasilnya (EI) akan diklasifikasikan menjadi *engagement status* dari siswa dengan aturan jika EI lebih besar sama dengan 4.5 maka statusnya adalah ***highly engagement***, jika EI lebih kecil dari 4.5 dan lebih besar sama dengan 4 maka statusnya adalah ***confused***, jika EI lebih kecil dari 4 dan lebih besar sama dengan 2.5 maka statusnya adalah ***boredom***, jika EI lebih kecil dari 2.5 dan lebih besar dari 0 maka statusnya ***sleepy***, dan jika EI sama dengan 0 maka statusnya ***very not engaged***.
9. Proses akan diiterasi sebanyak frame yang tersedia hasil dari proses pada poin 1.
10. Ketika semua *frame* sudah diproses maka proses terakhir adalah menganalisis seluruh data *engagement index* menggunakan LLM untuk mendapatkan kesimpulan akhir dan *insight* yang berupa analisis *engagement* siswa juga sedikit saran perbaikan atas materi yang diberikan oleh pengajar, dan proses selesai.

Hasil akhir dari seluruh proses tersebut akan disajikan dalam bentuk grafik, sehingga memudahkan untuk melakukan analisis mendalam terhadap masing-masing hasil dari siswa.



Gambar IV.7 Skema metode usulan dalam penanganan masalah oklusi dan wajah *non-frontal* pada wajah

Penjelasan pada gambar IV.7 adalah sebagai berikut:

1. Proses dimulai dari gambar *input* yang akan dilakukan proses deteksi wajah, jika wajah tidak terdeteksi maka proses langsung akan selesai, dan jika wajah terdeteksi maka akan dilanjutkan pada proses selanjutnya.
2. Selanjutnya akan dilakukan penyesuaian arah *roll* kepala, sehingga kepala yang miring akan dibuat tegak lurus.
3. Proses selanjutnya adalah deteksi tangan, jika ada tangan terdeteksi maka akan dilakukan *landmarking* pada objek tangan dan wajah, namun jika tidak ada tangan yang terdeteksi maka *landmarking* hanya dilakukan pada objek wajah saja.
4. Hasil *landmarking* selanjutnya akan dilakukan *masking* untuk membentuk objek, wajah diakan dibuat objek dengan warna putih, dan tangan akan dibuat objek berwarna merah.

5. Gambar objek yang hanya wajah saja maupun yang merupakan gabungan objek tangan dan wajah selanjutnya akan dibagi dua berdasarkan titik tengah wajah, sehingga menghasilkan bagian wajah kiri dan wajah kanan.
6. Selanjutnya adalah proses memilih bagian gambar wajah kiri ataupun kanan. Jika objeknya hanya wajah, maka proses pemilihannya adalah dengan memilih luas *pixel* putih yang lebih besar, namun jika objeknya merupakan gabungan wajah dan tangan, maka proses pemilihannya adalah bagian yang memiliki *pixel* merah yang lebih sedikit ataupun bagian yang tidak memiliki *pixel* merah sama sekali.
7. Bagian gambar wajah yang terpilih dilakukan proses *flip* dan *combine* sehingga hasilnya akan membentuk wajah yang seolah menghadap ke depan dan tanpa adanya oklusi tangan.
8. Proses selesai.

Wajah dari hasil metode usulan pada gambar IV.8 akan dilakukan proses FER sesuai dengan gambar IV.5 dengan memanfaatkan model FER yang sudah ada untuk mendapatkan data emosi wajah dari siswa. Hasil data emosi wajah siswa akan digabungkan dengan data arah pandangan mata dan data arah kepala menghadap untuk disimpulkan *engagement* dari siswa yang melakukan pembelajaran daring dengan materi *asynchronous video*.

IV.6 Desain Pengujian

Pengujian akan dilakukan dengan membandingkan akurasi pengenalan emosi wajah pada wajah yang tanpa dilakukan pre-processing dengan metode usulan dan data wajah yang dilakukan pre-processing dengan metode usulan. Hal tersebut untuk mendapatkan seberapa besar pengaruh dari metode yang diusulkan dalam meningkatkan akurasi pada wajah yang teroklusi. Adapun langkah-langkah pengujianya adalah sebagai berikut :

1. Mengumpulkan *dataset* publik FER yang tersedia.
2. Membersihkan *dataset* yang terkumpul seperti menghapus gambar yang tidak sesuai dengan label emosinya dengan menggunakan tools *py-feat* (Cheong dkk., 2023).

3. Memisahkan data wajah yang lurus dan kondisi *non-frontal*.
4. Data wajah lurus yang sudah bersih kemudian dilakukan augmentasi berupa penambahan objek tangan sintesis pada sebagian wajah dan hasilnya dilabeli dengan “*Dataset Before*”.
5. “*Dataset Before*” lalu dilakukan proses dengan metode usulan, dan hasilnya akan dilabeli dengan “*Dataset After*”.
6. *Dataset before* dan after pada poin 4 masing-masing akan dilakukan pengujian akurasi dengan 4 model yaitu CNN, DeepFace, Facelib, dan SVM.
7. Hasil akurasi pada *dataset* before dan after akan dibandingkan dan dihitung selisihnya untuk melihat pengaruhnya yang bisa jadi naik dan bisa jadi turun.
8. Hasil akurasi juga akan diuji per kelas emosi, sehingga tampak emosi mana yang paling berdampak setelah dilakukan pemrosesan dengan metode usulan.

Dataset yang akan digunakan dalam pengujian adalah *dataset* FER publik gabungan dari Affectnet dan RafDb. Untuk memenuhi kebutuhan data wajah yang tertutup tangan, maka data wajah lurus yang diseleksi dari gabungan *dataset* tersebut akan dilakukan proses penambahan tangan sintesis, hasilnya merupakan representasi kemungkinan terburuk pada kondisi wajah tertutup sebagian oleh tangan (vertikal). Adapun prosesnya dapat dilihat pada gambar IV.8.



Gambar IV.8 Proses penambahan tangan sintesis, (a) gambar asli, (b) tangan sintesis, (c) hasil augmentasi

Pada akhirnya data wajah yang akan digunakan dalam pengujian ini adalah data wajah yang teroklusi dengan tangan sebagian, dan data wajah yang miring (*non-frontal*) sesuai dengan gambar IV.9.



Gambar IV.9 Data wajah untuk pengujian, (a) wajah dengan oklusi tangan, (b) wajah non-frontal

IV.7 Pengembangan

Sistem keseluruhan akan dibuat dengan berbasis aplikasi web. Menggunakan *flask* sebagai *back-end* dan *html css js* sebagai *front-end* untuk mendukung sistem yang bisa diakses dengan *cross-platform*.

IV.7.1 Perangkat Keras

Perangkat keras yang digunakan dalam penelitian ini menggunakan sebuah *personal computer* dengan spesifikasi CPU Ryzen 5 5600X dengan *6 cores 12 threads 3.7Ghz Up to 4.6Ghz*, GPU Nvidia 3060 12Gb yang digunakan untuk melakukan pelatihan model untuk pengujian data, dan RAM DDR4 16Gb. Spesifikasi tersebut tergolong pada kategori *mid-range*, dan cukup untuk melakukan sebuah pelatihan data ataupun tugas *computer vision* lainnya.

IV.7.2 Perangkat Lunak

Perangkat lunak yang digunakan dalam pengembangan sistem antara lain, sistem operasi Windows 11, aplikasi VSCode untuk melakukan programming, dan Jupyter Notebook untuk melakukan analisis hasil. Python dengan library Flask digunakan sebagai *back-end* karena kemampuannya dalam mengintegrasikan *python* dengan tugas pemrosesan *computer vision* dan *html-css-js* sebagai media yang bertugas sebagai *interface* sistem. *Database* yang digunakan adalah *mysql*, karena data pada sistem ini masih terstruktur sehingga *mysql* dinilai cukup untuk melakukan tugasnya dengan baik.

IV.7.2.1 Desain Halaman Login

Halaman login dibuat dengan tampilan yang sederhana dan mudah digunakan, sehingga pengguna dapat mengakses sistem tanpa kesulitan. Penempatan elemen seperti kolom username, password, serta tombol masuk diatur secara jelas untuk memudahkan proses autentikasi. Adapun hasil rancangan halaman *login* pada penelitian ini dapat dilihat pada gambar IV.10.



Rancangan desain halaman *login* yang menampilkan form input untuk Username dan Password, serta dua tombol yang belum dilabelkan.

Gambar IV.10 Rancangan desain halaman *login*

IV.7.2.2 Desain Halaman Asynchronous Learning

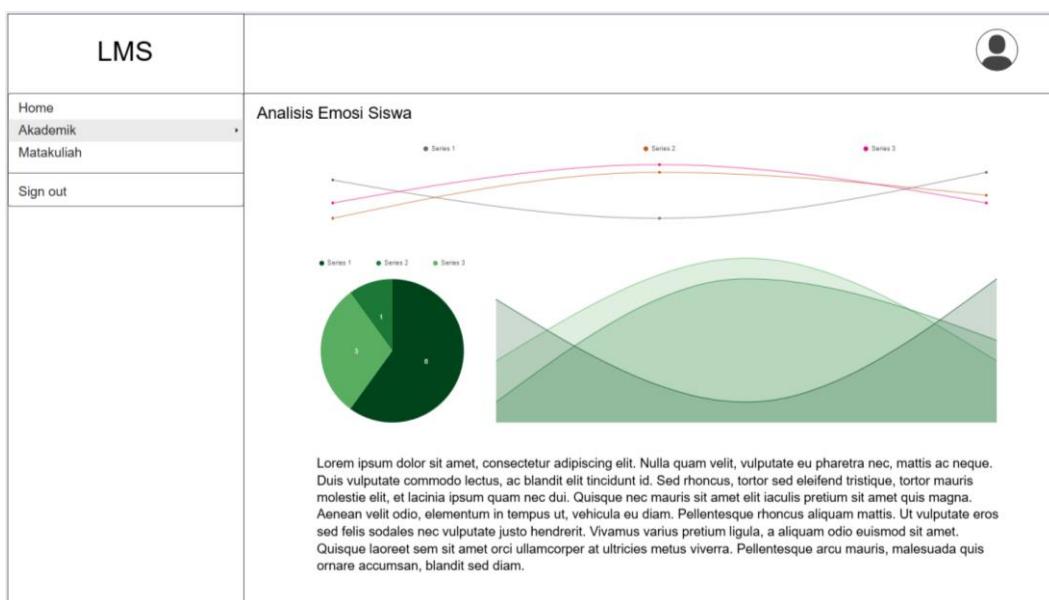
Halaman *asynchronous learning* akan berisi beberapa elemen, yaitu *frame* video pembelajaran, *frame* tampilan dari *webcam*, dan tombol navigasi untuk memulai dan menyelesaikan video pembelajaran. Adapun hasil rancangan halaman *asynchronous learning* pada penelitian ini dapat dilihat pada gambar IV.11.



Gambar IV.11 Rancangan halaman *asynchronous learning*

IV.7.2.3 Desain Halaman Analisis Emosi

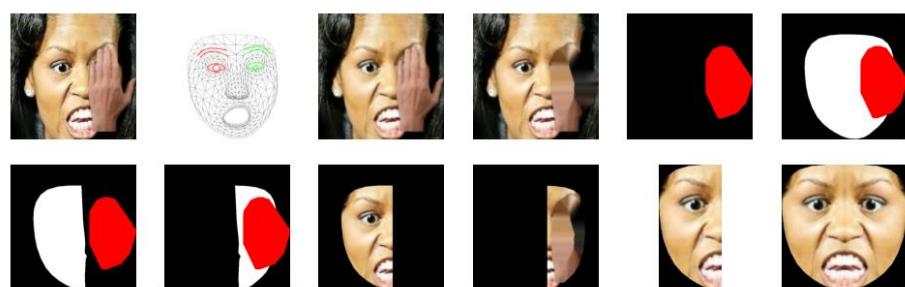
Halaman ini akan berisi data berupa grafik tentang analisis emosi siswa berdasarkan hasil dari penilaian sistem saat siswa sedang mengakses materi pada *asynchronous learning*. Pada bagian ini akan ditampilkan analisis kefokusan, tingkat *engagement*, emosi dominan, hingga sedikit masukan umpan balik kepada pengajar terhadap materi yang diberikan. Adapun hasil rancangan halaman analisis emosi pada penelitian ini dapat dilihat pada gambar IV.12.



Gambar IV.12 Rancangan halaman analisis emosi

IV.7.3 Pengembangan Metode Usulan

Gambar yang akan diproses adalah hasil tangkapan kamera yang dilakukan pada pihak *front-end*. Gambar tersebut akan dikirim ke *back-end* melalui *socket io*, lalu dilakukan *pre-processing* terlebih dahulu dengan metode usulan. Adapun langkah pada metode usulan ini dapat dilihat pada gambar IV.13.



Gambar IV.13 Langkah pada metode usulan

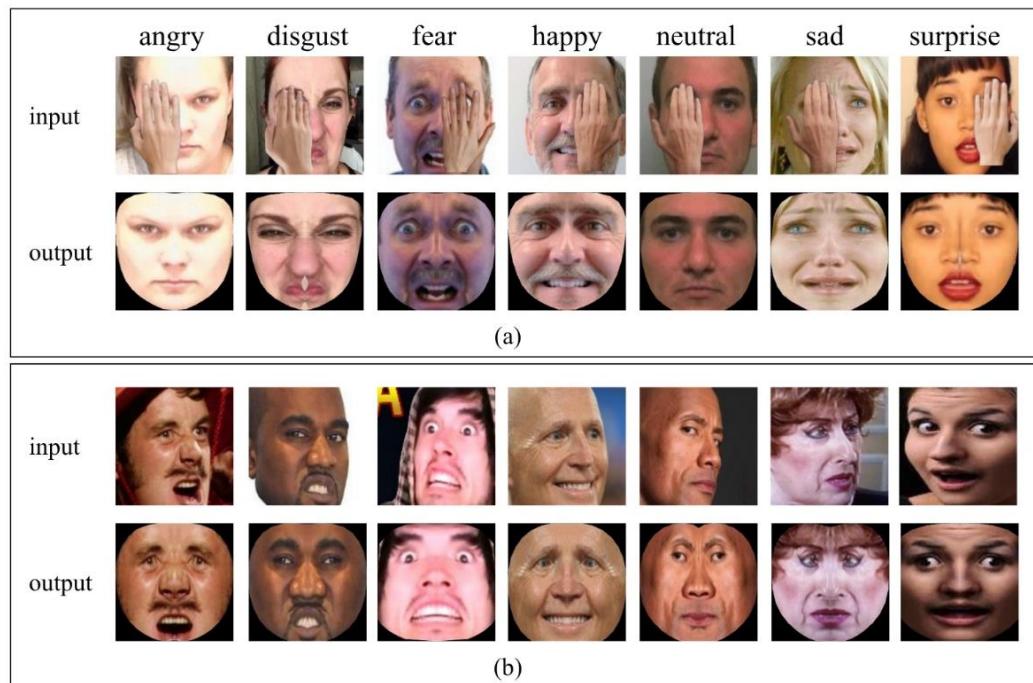
Pada gambar IV.8 dapat dijelaskan bahwa gambar yang berupa wajah yang teroklusi objek tangan akan dilakukan konversi ke bentuk *mesh* dengan memanfaatkan *library mediapipe*. Lalu gambar tersebut akan dikoreksi arah kepalanya (*roll*) yang dimana jika wajahnya miring (*roll*) akan diluruskan terlebih dahulu. Selanjutnya wajah yang sudah diluruskan akan dideteksi objek tangannya, jika ada tangan terdeteksi maka tangan tersebut akan dibuat menjadi *blur*. Wajah dan tangan tersebut dilakukan masking menjadi objek berbentuk objek putih untuk bagian wajah, dan objek merah untuk bagian tangan. Objek hasil *masking* merah dan putih tersebut akan dibagi dua sesuai dengan titik tengah wajah. Hasil baginya akan dibandingkan dengan menghitung masing-masing luas pixel putihnya. Bagian yang diambil adalah bagian yang memiliki pixel putih lebih banyak dari bagian lainnya yang dimana hal tersebut berarti tidak ada atau objek oklusiya lebih sedikit. Berdasarkan objek *masking* yang dipilih, kemudian wajah asli akan *di-stretching* ke bentuk masking putih yang telah diluruskan bagian tengahnya. Lalu bagian wajah asli yang telah *di-stretching* tersebut diduplikasi dan dikombinasikan menjadi wajah yang seolah menghadap ke depan tanpa adanya tangan yang menutupi. Hasil wajah tersebutlah yang akan dilakukan proses FER dengan model-model FER yang telah ada.

BAB V Hasil dan Evaluasi

Bab ini membahas hasil dari penerapan solusi yang telah dirancang. Selain itu, pada bagian ini juga akan dijelaskan evaluasi sesuai dengan rancangan rencana evaluasi pada bab sebelumnya. Setiap tahapan pengembangan dijelaskan secara detail dalam subbab masing-masing.

V.1 Hasil Metode Usulan

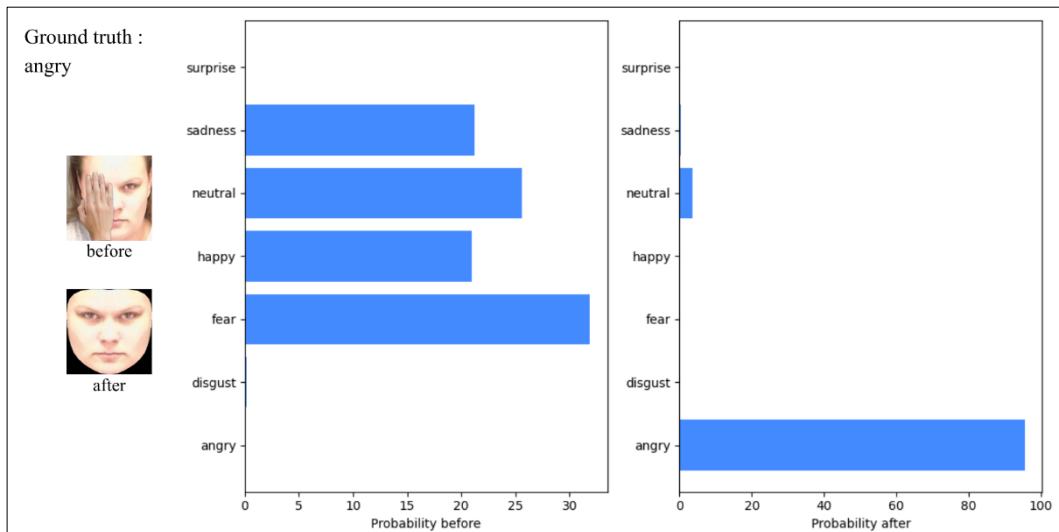
Bagian ini menyajikan hasil penerapan metode usulan terhadap data uji yang berupa wajah dengan oklusi tangan dan wajah yang tidak menghadap ke depan. Gambar V.1 menunjukkan hasil dimana gambar *input* adalah data yang diujikan, dan gambar *output* adalah hasil dari penerapan metode usulan.



Gambar V.1 Hasil percobaan metode usulan, (a) data wajah dengan oklusi tangan, (b) wajah dalam keadaan non-frontal

Gambar V.1 berisi dua subbagian (a) dan (b) yang masing-masing memperlihatkan contoh data wajah dari tujuh kategori emosi: *angry*, *disgust*, *fear*, *happy*, *neutral*, *sad*, dan *surprise*. Hasilnya menunjukkan keberhasilan metode usulan melakukan frontalisasi atau transformasi terhadap wajah sehingga orientasi dan struktur wajah menjadi lebih seragam dan frontal tanpa menghilangkan fitur yang penting untuk FER pada citra asli.

Penilaian keberhasilan metode usulan adalah dengan membandingkan emosi pada gambar sebelum diterapkan metode usulan dan gambar setelah diterapkan metode usulan.



Gambar V.2 Perbandingan probabilitas emosi pada gambar sebelum dan sesudah diterapkan metode usulan

Gambar V.2 menunjukkan pada gambar sebelum diterapkan metode usulan, hasil komposisi probabilitas emosinya menjadi ambigu sehingga hasil emosinya menjadi salah atau berbeda dengan emosi yang seharusnya. Namun, setelah gambar diterapkan metode usulan, maka emosinya menjadi benar atau sesuai dengan label emosi yang seharusnya disebabkan oleh probabilitas yang awalnya ambigu menjadi *confidence* terhadap label emosi yang seharusnya.

V.1.1 Hasil pada Wajah yang Tertutup Oklusi Tangan (Vertikal)

Pada bagian ini akan disajikan hasil pengukuran akurasi *before* dan *after* wajah dengan oklusi tangan sebagian menggunakan 4 model yaitu CNN, DeepFace, FaceLib, dan SVM sebagai validasi.

1. Hasil Evaluasi pada Model CNN

Hasil pengujian dengan model CNN pada tabel V.1 menunjukkan bahwa metode usulan berhasil menaikkan akurasi seluruh label emosi pada wajah yang teroklusi tangan sebagian (vertikal) dengan rata-rata kenaikan keseluruhan sebesar 33.79%.

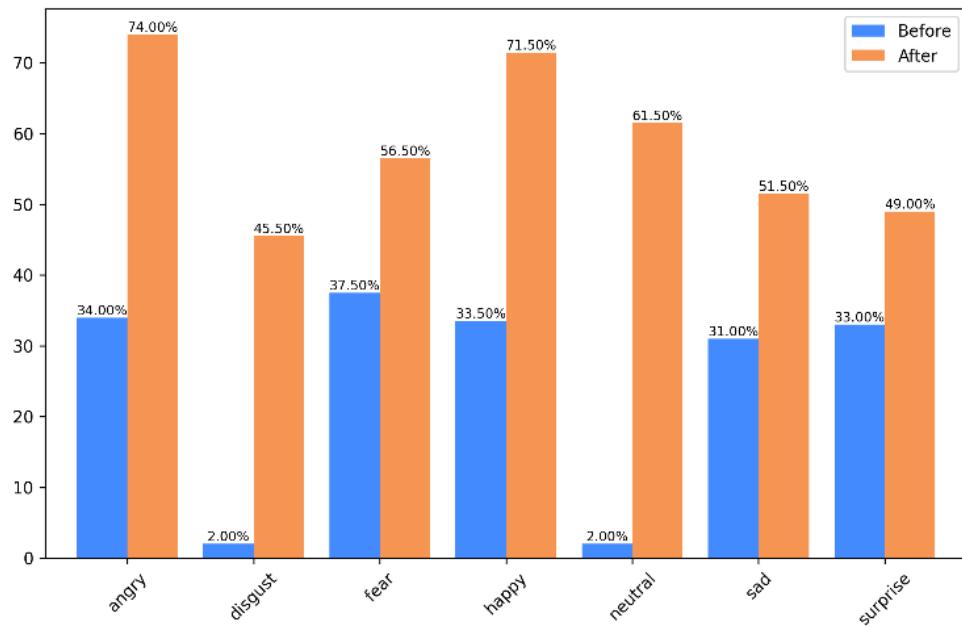
Tabel V.1 Hasil akurasi data wajah dengan oklusi tangan dengan CNN

Emosi	Jumlah Wajah terdeteksi (%)	Sebelum ^a (%)	Setelah ^b (%)	Selisih (b - a)	Kesimpulan
Angry	100	34.00	74.00	40.00	Naik
Disgust	100	2.00	45.50	43.50	Naik
Fear	100	37.50	56.50	19.00	Naik
Happy	100	33.50	71.50	38.00	Naik
Neutral	100	2.00	61.50	59.50	Naik
Sad	100	31.00	51.50	20.50	Naik
Surprise	100	33.00	49.00	16.00	Naik
Rata-rata				33.79	Naik

a. Akurasi sebelum penerapan metode usulan

b. Akurasi setelah penerapan metode usulan

Adapun visualisasi hasil pengujian data sebelum dan sesudah penerapan metode usulan dengan model CNN dapat dilihat pada gambar V.3.



Gambar V.3 Visualisasi perbandingan akurasi data sebelum dan sesudah penerapan metode usulan pada data wajah dengan oklusi tangan menggunakan model CNN

2. Hasil Akurasi pada Model DeepFace

Hasil pengujian dengan model DeepFace pada tabel V.2 menunjukkan bahwa metode usulan berhasil menaikkan akurasi pada beberapa label emosi pada wajah yang teroklusi tangan sebagian (vertikal). Walaupun terdapat penurunan dan tetap pada emosi *surprise* dan *fear*, namun secara keseluruhan hasilnya mengalami kenaikan rata-rata sebesar 14.29%.

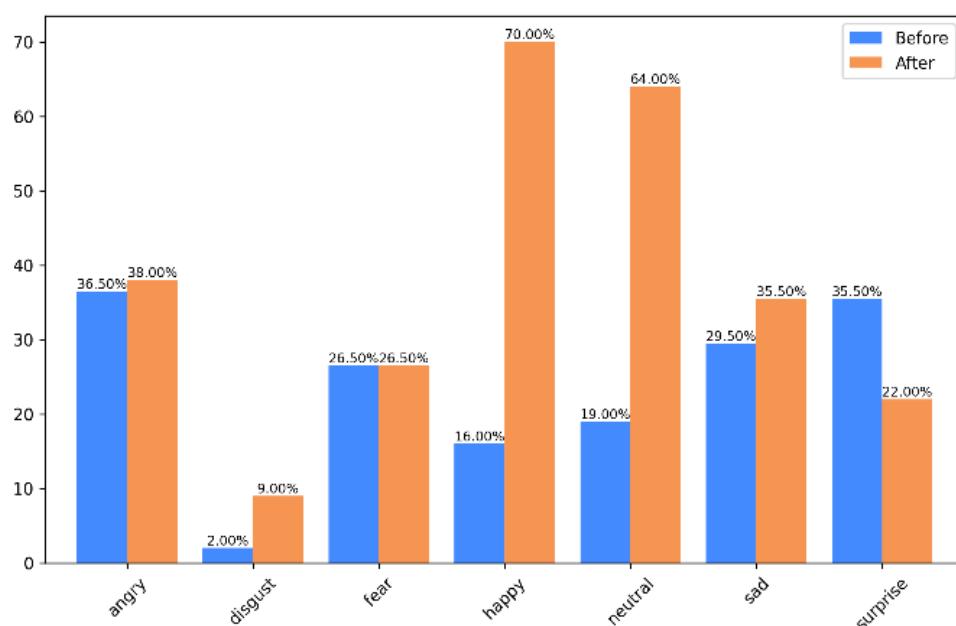
Tabel V.2 Hasil akurasi data wajah dengan oklusi tangan dengan DeepFace

Emosi	Jumlah Wajah terdeteksi (%)	Sebelum ^a (%)	Setelah ^b (%)	Selisih (b - a)	Kesimpulan
Angry	100	36.50	38.00	1.50	Naik
Disgust	100	2.00	9.00	7.00	Naik
Fear	100	26.50	26.50	0.00	Tetap
Happy	100	16.00	70.00	54.00	Naik
Neutral	100	19.00	64.00	45.00	Naik
Sad	100	29.50	35.50	6.00	Naik
Surprise	100	35.50	22.00	-13.50	Turun
Rata-rata				14.29	Naik

a. Akurasi sebelum penerapan metode usulan

b. Akurasi setelah penerapan metode usulan

Adapun visualisasi hasil pengujian data sebelum dan sesudah penerapan metode usulan dengan model DeepFace dapat dilihat pada gambar V.4.



Gambar V.4 Visualisasi perbandingan akurasi data sebelum dan sesudah penerapan metode usulan pada data wajah dengan oklusi tangan menggunakan model DeepFace

3. Hasil Akurasi pada Model FaceLib

Hasil tidak biasa ditunjukkan oleh FaceLib, dimana model ini tidak berhasil mendeteksi wajah pada kondisi tertutup tangan sebagian. Sehingga jumlah data yang diukur hanyalah sejumlah wajah yang terdeteksi oleh model FaceLib. Hasil pengujian dengan model FaceLib pada tabel V.3

menunjukkan bahwa metode usulan berhasil menaikkan akurasi pada beberapa label emosi pada wajah yang teroklusi tangan sebagian (vertikal). Terdapat penurunan pada emosi *Surprise*, walaupun demikian, secara keseluruhan hasilnya mengalami kenaikan rata-rata sebesar 10.66%.

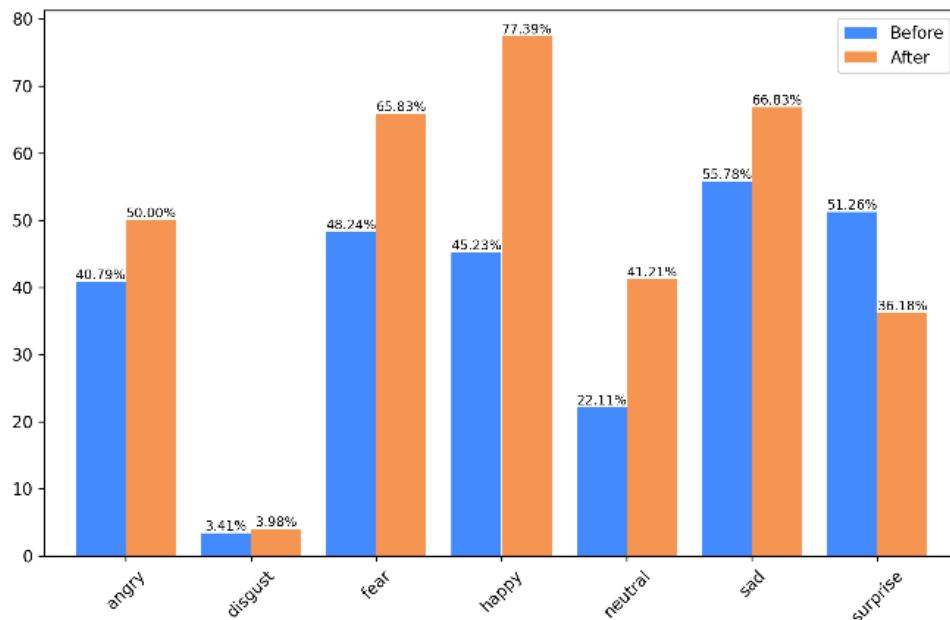
Tabel V.3 Hasil akurasi data wajah dengan oklusi tangan dengan FaceLib

Emosi	Jumlah Wajah terdeteksi (%)	Sebelum ^a (%)	Setelah ^b (%)	Selisih (b - a)	Kesimpulan
Angry	76.00	40.79	50.00	9.21	Naik
Disgust	88.00	3.41	3.98	0.57	Naik
Fear	99.50	48.24	65.83	17.59	Naik
Happy	99.50	45.23	77.39	32.16	Naik
Neutral	99.50	22.11	41.21	19.10	Naik
Sad	99.50	55.78	66.83	11.05	Naik
Surprise	99.50	51.26	36.18	-15.08	Turun
Rata-rata				10.66	Naik

a. Akurasi sebelum penerapan metode usulan

b. Akurasi setelah penerapan metode usulan

Adapun visualisasi hasil pengujian data sebelum dan sesudah penerapan metode usulan dengan model FaceLib dapat dilihat pada gambar V.5.



Gambar V.5 Visualisasi perbandingan akurasi data sebelum dan sesudah penerapan metode usulan pada data wajah dengan oklusi tangan menggunakan model FaceLib

4. Hasil Akurasi pada Model SVM

Hasil pengujian dengan model SVM pada tabel V.4 menunjukkan bahwa keberhasilan metode usulan hanya pada emosi angry, happy, neutral, dan sad. Terdapat penurunan pada emosi *surprise* dan *fear*, dan tidak ada perubahan terhadap emosi *disgust*. Walaupun demikian, secara keseluruhan hasilnya mengalami kenaikan rata-rata sebesar 6.49%.

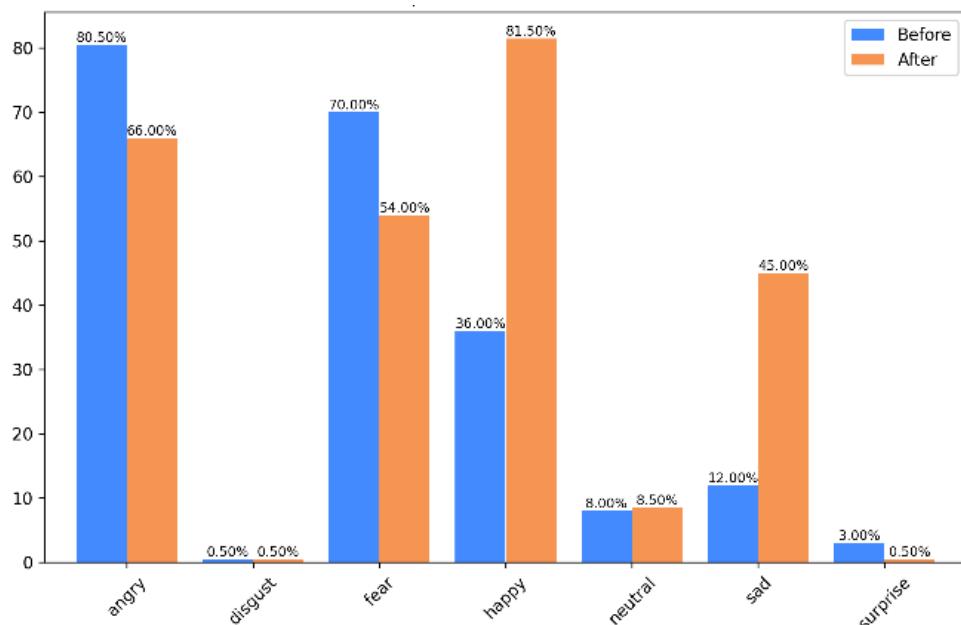
Tabel V.4 Hasil akurasi data wajah dengan oklusi tangan dengan SVM

Emosi	Jumlah Wajah terdeteksi (%)	Sebelum ^a (%)	Setelah ^b (%)	Selisih (b - a)	Kesimpulan
Angry	100	80.50	66.00	14.50	Naik
Disgust	100	0.50	0.50	0.00	Tetap
Fear	100	70.00	54.00	-16.54	Turun
Happy	100	36.00	81.50	45.50	Naik
Neutral	100	8.00	8.50	0.50	Naik
Sad	100	12.00	45.00	33.00	Naik
Surprise	100	3.00	0.50	-2.50	Turun
Rata-rata				6.49	Naik

a. Akurasi sebelum penerapan metode usulan

b. Akurasi setelah penerapan metode usulan

Adapun visualisasi hasil pengujian data sebelum dan sesudah penerapan metode usulan dengan model SVM dapat dilihat pada gambar V.6.



Gambar V.6 Visualisasi perbandingan akurasi data sebelum dan sesudah penerapan metode usulan pada data wajah dengan oklusi tangan menggunakan model SVM

V.1.2 Hasil pada Wajah dengan Kondisi Miring (*Non-Frontal*)

Pada bagian ini akan disajikan hasil pengukuran akurasi *before* dan *after* wajah dalam kondisi *non-frontal* menggunakan 4 model yaitu CNN, DeepFace, FaceLib, dan SVM sebagai validasi.

1. Hasil Akurasi pada Model CNN

Hasil pengujian pada wajah miring dengan model CNN pada tabel V.5 menunjukkan bahwa metode usulan berhasil menaikkan akurasi emosi pada wajah dalam kondisi *non-frontal* dengan rata-rata kenaikan keseluruhan sebesar 16.53%. Walaupun demikian, terdapat akurasi yang turun pada label emosi *angry* sebesar -26.39%.

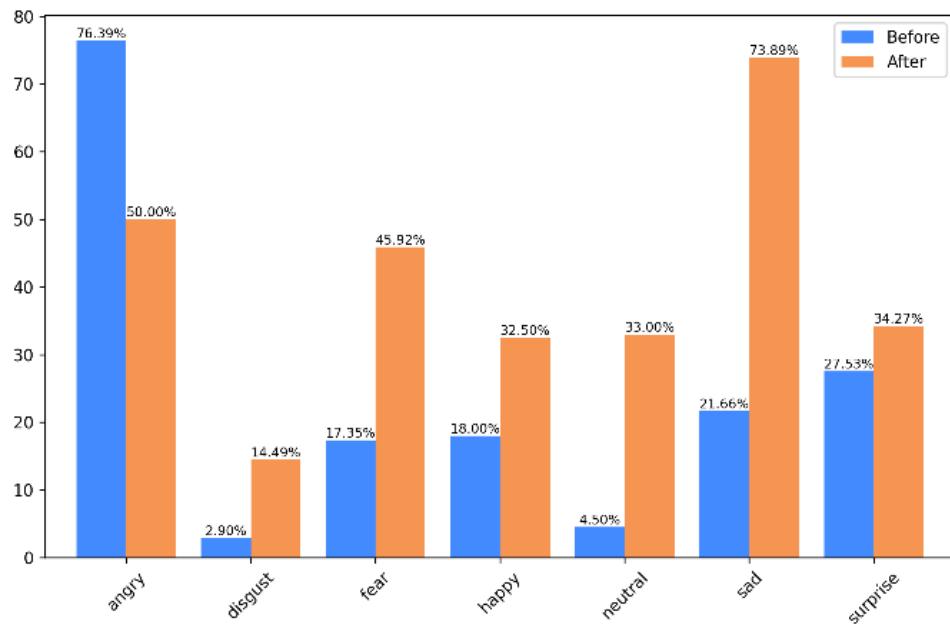
Tabel V.5 Hasil akurasi data wajah miring (*non-frontal*) dengan CNN

Emosi	Jumlah Wajah terdeteksi (%)	Sebelum ^a (%)	Setelah ^b (%)	Selisih (b - a)	Kesimpulan
Angry	100	76.39	50.00	-26.39	Turun
Disgust	100	2.90	14.49	11.59	Naik
Fear	100	17.35	45.92	28.57	Naik
Happy	100	18.00	32.50	14.50	Naik
Neutral	100	4.50	33.00	28.50	Naik
Sad	100	21.66	73.89	52.23	Naik
Surprise	100	27.53	34.27	6.74	Naik
Rata-rata				16.53	Naik

a. Akurasi sebelum penerapan metode usulan

b. Akurasi setelah penerapan metode usulan

Adapun visualisasi hasil pengujian data sebelum dan sesudah penerapan metode usulan dengan model CNN dapat dilihat pada gambar V.7.



Gambar V.7 Hasil akurasi data wajah *non-frontal* dengan CNN

2. Hasil Akurasi pada Model DeepFace

Hasil pengujian pada wajah miring dengan model DeepFace pada tabel V.6 menunjukkan bahwa metode usulan hanya berhasil menaikkan akurasi emosi wajah dalam kondisi *non-frontal* pada emosi *angry* dan *neutral* saja masing-masing sebesar 1.39% dan 14.00%, dengan sisanya mengalami penurunan. Walaupun demikian, jumlah penurunan rata-rata hanya sebesar -1.03%.

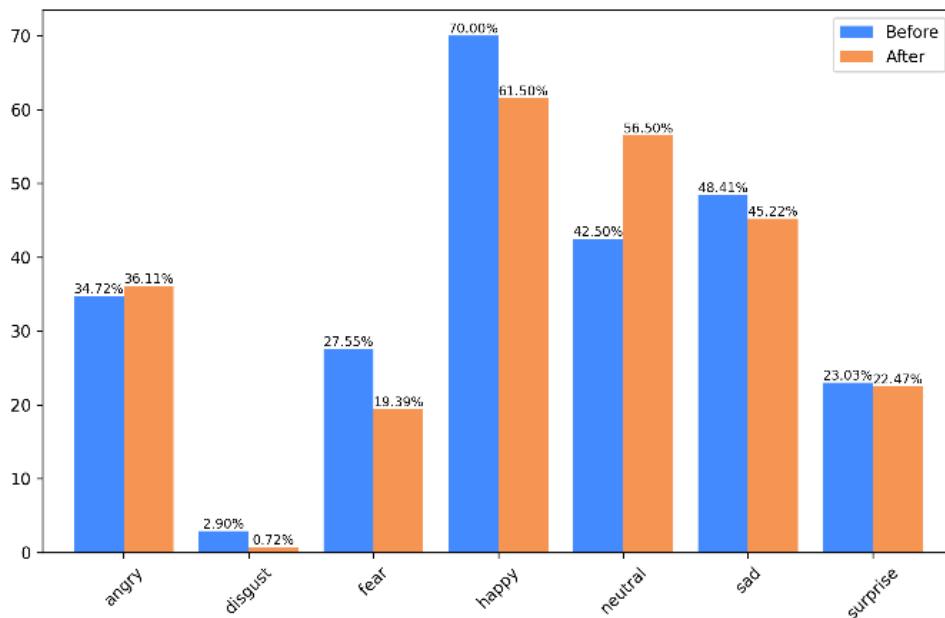
Tabel V.6 Hasil akurasi data wajah miring (*non-frontal*) dengan DeepFace

Emosi	Jumlah Wajah terdeteksi (%)	Sebelum ^a (%)	Setelah ^b (%)	Selisih (b - a)	Kesimpulan
Angry	100	34.72	36.11	1.39	Naik
Disgust	100	2.90	0.72	-2.18	Turun
Fear	100	27.55	19.39	-8.16	Turun
Happy	100	70.00	61.5	-8.50	Turun
Neutral	100	42.50	56.50	14.00	Naik
Sad	100	29.50	45.22	-3.19	Turun
Surprise	100	23.03	22.47	-0.56	Turun
Rata-rata				-1.03	Turun

a. Akurasi sebelum penerapan metode usulan

b. Akurasi setelah penerapan metode usulan

Adapun visualisasi hasil pengujian data sebelum dan sesudah penerapan metode usulan dengan model DeepFace dapat dilihat pada gambar V.8.



Gambar V.8 Hasil akurasi data wajah *non-frontal* dengan DeepFace

3. Hasil Akurasi pada Model Facelib

Hasil pengujian pada wajah miring dengan model FaceLib pada tabel V.7 menunjukkan bahwa metode usulan hanya berhasil menaikkan akurasi pada emosi *disgust* dan *sad* saja masing-masing sebesar 26.53% dan 12.44%, dengan sisanya mengalami penurunan. Jumlah rata-rata penurunan setelah penerapan metode usulan menggunakan FaceLib sebesar -5.12%.

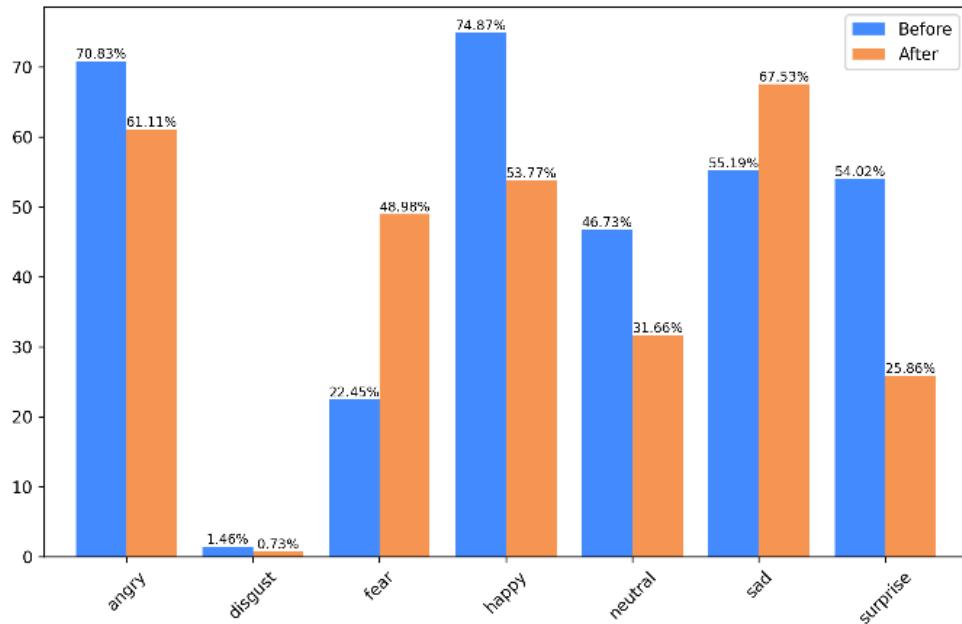
Tabel V.7 Hasil akurasi data wajah miring (*non-frontal*) dengan FaceLib

Emosi	Jumlah Wajah terdeteksi (%)	Sebelum ^a (%)	Setelah ^b (%)	Selisih (b - a)	Kesimpulan
Angry	100	70.83	51.11	-9.72	Turun
Disgust	99.27	1.46	0.73	-0.73	Turun
Fear	100	22.45	48.98	26.53	Naik
Happy	99.50	74.87	53.77	-21.10	Turun
Neutral	99.50	46.73	31.66	-15.10	Turun
Sad	98.08	55.19	67.63	12.44	Naik
Surprise	97.75	54.02	25.86	-28.20	Turun
Rata-rata				-5.12	Turun

a. Akurasi sebelum penerapan metode usulan

b. Akurasi setelah penerapan metode usulan

Adapun visualisasi hasil pengujian data sebelum dan sesudah penerapan metode usulan dengan model FaceLib dapat dilihat pada gambar V.9.



Gambar V.9 Hasil akurasi data wajah *non-frontal* dengan FaceLib

4. Hasil Akurasi pada Model SVM

Hasil pengujian pada wajah miring dengan model SVM pada tabel V.8 menunjukkan bahwa metode usulan berhasil menaikkan akurasi emosi pada emosi *happy* dan *sad* saja masing-masing sebesar 11.50% dan 23.57%, dengan sisanya mengalami penurunan. Walaupun demikian, secara keseluruhan hasilnya mengalami kenaikan rata-rata sebesar 2.21%.

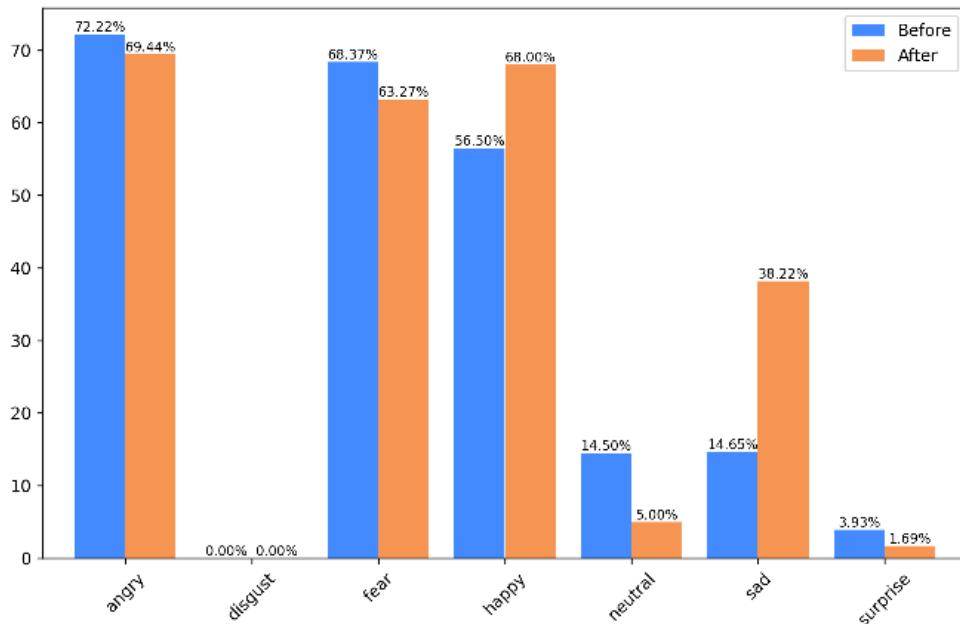
Tabel V.8 Hasil akurasi data wajah miring (*non-frontal*) dengan SVM

Emosi	Jumlah Wajah terdeteksi (%)	Sebelum ^a (%)	Setelah ^b (%)	Selisih (b - a)	Kesimpulan
Angry	100	72.22	69.44	-2.78	Turun
Disgust	100	0.00	0.00	0.00	Tetap
Fear	100	68.37	63.27	-5.10	Turun
Happy	100	56.50	68.00	11.50	Naik
Neutral	100	14.50	5.00	-9.50	Turun
Sad	100	14.65	38.22	23.57	Naik
Surprise	100	3.93	1.69	-2.24	Turun
Rata-rata				2.21	Naik

a. Akurasi sebelum penerapan metode usulan

b. Akurasi setelah penerapan metode usulan

Adapun visualisasi hasil pengujian data sebelum dan sesudah penerapan metode usulan dengan model SVM dapat dilihat pada gambar V.10.



Gambar V.10 Hasil akurasi data wajah *non-frontal* dengan SVM

V.1.3 Hasil pada Wajah yang Tertutup Oklusi Tangan (Horizontal)

Walaupun metode usulan tidak dirancang untuk kasus seperti pada gambar V.11, namun metode usulan ini juga tetap diujikan terhadap data wajah dengan oklusi tangan horizontal, karena kondisi tersebut juga kerap terjadi terutama dalam dunia pendidikan daring.



Gambar V.11 Kondisi wajah tertutup tangan pada bagian bawah (mulut)

Pengujian dilakukan hanya dengan *dataset* RafDB yang telah diseleksi dengan jumlah sesuai pada tabel V.9, dimana *dataset* tersebut tidak ikut dilatihkan pada *training* model CNN yang hanya menggunakan data *frontal face* saja.

Tabel V.9 Jumlah data uji wajah dengan oklusi tangan (horizontal)

Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise
106	62	73	200	200	200	200

Hasil penerapan metode usulan terhadap wajah dengan oklusi tangan di bagian mulut (gambar V.11) tervisualisasi pada gambar V.12, dimana hasil tangan yang menutupi bagian mulut dirubah menjadi sedikit *blur* dan seakan warnanya menyatu dengan *pixel* disekitarnya.



Gambar V.12 Hasil penerapan metode usulan terhadap wajah dengan oklusi tangan pada bagian bawah wajah

Hasil pada tabel V.10 menunjukkan performa keberhasilan model untuk mendeteksi objek wajah. Terlihat dari 4 model yang diujikan, hanya FaceLib yang tidak berhasil secara 100% untuk mendeteksi wajah dengan oklusi tangan pada bagian mulut.

Tabel V.10 Performa model dalam mendeteksi wajah pada wajah yang teroklusi tangan pada bagian mulut

Emosi	CNN (%)	DeepFace (%)	FaceLib (%)	SVM (%)
Angry	100	100	96.22	100
Disgust	100	100	100	100
Fear	100	100	93.15	100
Happy	100	100	99.50	100
Neutral	100	100	99.50	100
Sad	100	100	99.50	100
Surprise	100	100	99.50	100

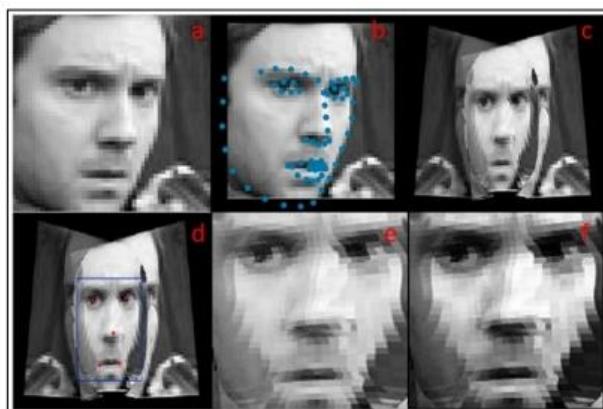
Pengujian dilakukan dengan menghitung selisih antara akurasi sebelum dan sesudah penerapan metode usulan. Hasilnya dapat dilihat pada tabel V.11, angka positif menunjukkan bahwa metode usulan berhasil menaikkan akurasi, sedangkan angka negatif berarti metode usulan membuat akurasinya menjadi memburuk.

Tabel V.11 Hasil pengujian metode usulan pada wajah yang teroklusi tangan pada bagian mulut

Emosi	CNN (%)	DeepFace (%)	FaceLib (%)	SVM (%)
Angry	-2.83	4.72	-0.98	-9.44
Disgust	-1.61	-3.23	0.00	0.00
Fear	-12.33	-5.48	14.71	0.00
Happy	-4.00	-6.50	4.02	-2.00
Neutral	32.00	-9.50	-41.70	-10.5
Sad	-21.5	-11.00	12.57	16.5
Surprise	12.00	10.5	-9.04	-0.5
Rata-rata	0.25	- 2.93	- 2.92	- 0.85

V.1.4 Hasil Komparasi terhadap Metode Frontalisasi Lain

Dalam rangka memperkuat validasi metode usulan, eksperimen juga dilakukan dengan dengan membandingkan metode usulan dengan metode frontalisasi yang dilakukan oleh Tsai dkk., (2021), dimana proses frontalisasinya juga dilakukan dengan hanya menggunakan pemrosesan gambar 2D (gambar V.13).



Gambar V.13 Metode frontalisasi wajah 2D oleh Tsai dkk., (2021)



Gambar V.14 Hasil komparasi, (A) gambar *input*, (B) metode usulan, (C) metode oleh Tsai dkk., (2021)

Tabel V.12 Hasil komparasi metode frontalisisasi

Emosi	CNN (%)		DeepFace (%)		FaceLib (%)		SVM (%)	
	a	b	a	b	a	b	a	b
Angry	-34.62	-26.39	-7.70	1.39	-14.29	-9.72	3.85	-2.78
Disgust	12.93	11.59	-3.40	-2.18	-2.04	-0.73	-0.68	0.00
Fear	17.82	28.57	-10.89	-8.16	16.83	26.53	-30.7	-5.10
Happy	29.00	14.50	-6.00	-8.50	4.53	-21.10	22.00	11.5
Neutral	28.00	28.50	17.50	14.00	-5.02	-15.07	1.00	-9.50
Sad	27.10	52.23	-15.49	-3.19	6.50	12.44	20.64	23.57
Surprise	25.34	6.74	-6.00	-0.56	-20.41	-28.16	-2.00	-2.24
Rata-rata	15.08	16.53	-4.57	-1.03	-1.99	-5.12	2.02	2.21

- a. Metode frontalisisasi oleh Tsai dkk., (2021)
- b. Metode usulan pada penelitian ini

Hasil pada tabel V.12 menunjukkan bahwa rata-rata dari perbandingan kedua metode tersebut tidak memiliki perbedaan yang terlalu jauh. Hasil komparasi ini terbatas hanya pada gambar wajah *non-frontal*, karena metode frontalisisasi yang dilakukan oleh Tsai dkk., (2021) tidak memiliki fitur untuk menghapus objek oklusi tangan.

V.2 Pembahasan

Bagian ini akan menjelaskan analisis mendalam serta beberapa temuan yang terjadi berdasarkan hasil eksperimen yang telah dijelaskan pada sub bab V.1.

V.2.1 Performa *Face Detector* pada Model

Tahap pertama dalam suatu sistem *Facial Expression Recognition* (FER) adalah proses deteksi wajah. Jika tahap ini gagal, seluruh *pipeline* selanjutnya akan berhenti karena sistem tidak dapat mengekstraksi fitur ekspresi dari citra yang tidak teridentifikasi sebagai wajah. Berdasarkan hasil penelitian yang dilakukan menggunakan empat metode *state-of-the-art* yaitu CNN, DeepFace, Facelib, dan SVM, hanya metode Facelib yang mengalami kegagalan dalam deteksi wajah, sementara tiga metode lainnya berhasil 100% mengenali wajah. Facelib menggunakan algoritma *Single Shot Multibox Detector* (SSD) pada tahap deteksi wajah.

SSD merupakan arsitektur *one-stage detector* yang mengombinasikan ekstraksi fitur berbasis CNN dengan mekanisme *multi-scale feature map* untuk mendeteksi

objek dalam berbagai ukuran secara simultan. Keunggulan utamanya terletak pada kecepatan inferensi yang tinggi karena proses klasifikasi dan *regresi bounding box* dilakukan dalam satu langkah. Namun, sejumlah penelitian menunjukkan bahwa efisiensi ini didapat dengan mengorbankan nilai kepresisan, khususnya dalam kondisi *occlusion* atau *pose variation*.

SSD bekerja optimal pada citra wajah *frontal* dan tanpa gangguan latar, tetapi performanya menurun drastis ketika sebagian wajah tertutup (misalnya oleh tangan atau masker) atau ketika wajah berada dalam posisi *non-frontal* (Qu dkk., 2022). Hasil pada table V.3, dan tabel V.7 membuktikan teori tersebut, dimana performa Facelib dengan detektor SSD pada wajah yang mengalami oklusi oleh tangan hanya mencapai 76% dan 88% pada kelas emosi *angry* dan *disgust*, sementara kelas lainnya (*fear*, *happy*, *neutral*, *sad*, dan *surprise*) menunjukkan hasil lebih tinggi, yaitu sekitar 99.5%. Untuk citra wajah non-frontal, performa deteksi masing-masing kelas adalah 100%, 99.27%, 100%, 99.50%, 99.50%, 98.08%, dan 97.75%. Penurunan performa deteksi pada wajah teroklusi cukup selaras dengan penelitian Jang dkk. (2019), yang menunjukkan bahwa varian algoritma Face-SSD hanya mampu mempertahankan akurasi di kisaran 80 - 95% tergantung derajat oklusi (Jang dkk., 2019).

Dengan demikian, meskipun pada penelitian lain disebutkan bahwa SSD menawarkan performa waktu nyata (*real-time performance*) yang unggul (Tariyal dkk., 2024), hasil pada penelitian ini menunjukkan bahwa tanpa peningkatan arsitektur dan tahap *pre-alignment*, performa SSD masih kurang memadai untuk aplikasi FER di lingkungan dunia nyata yang sarat dengan oklusi dan variasi *pose*.

V.2.2 Wajah yang Teroklusi oleh Tangan Sebagian

Adanya tangan yang menutupi sebagian wajah merupakan salah satu bentuk oklusi parsial yang paling umum terjadi dalam kondisi nyata, seperti ketika seseorang bereaksi secara spontan dalam emosi tertentu ataupun dalam keadaan bosan ketika mengikuti pelajaran daring, dimana tangan biasanya digunakan untuk menopang kepala atau wajah. Oklusi jenis ini secara signifikan menurunkan performa sistem FER karena menghalangi area wajah yang berperan penting dalam penentuan ekspresi, seperti mata, hidung, dan terutama mulut. Ketika area tersebut hilang dari

citra gambar wajah yang diujikan, representasi fitur yang dihasilkan dari proses ekstraksi menjadi tidak utuh, sehingga model kesulitan memetakan pola ke emosi yang benar.

Hasil eksperimental memperlihatkan efek tersebut dengan jelas. Pada data uji sebelum dilakukan pre-processing dengan metode usulan (data “*before*”), rata-rata akurasi FER hanya sebesar 24.71%, 23.57%, 38.12%, dan 30.08% masing-masing untuk model CNN, DeepFace, Facelib, dan SVM. Angka tersebut menunjukkan bahwa hampir dua pertiga klasifikasi gagal mengenali emosi dengan benar ketika wajah tertutup tangan. Terutama pada model berbasis deep learning seperti CNN dan DeepFace, performa yang rendah ini menunjukkan bahwa meskipun model mampu belajar fitur kompleks, model tersebut tetap sangat bergantung pada keutuhan informasi spasial wajah.

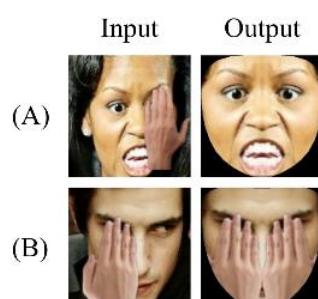
Setelah dilakukan *pre-processing* menggunakan metode usulan yang berfungsi untuk menghilangkan objek tangan pada gambar wajah, terjadi peningkatan akurasi yang signifikan. Nilai akurasi naik menjadi 58.5%, 37.9%, 48.8%, dan 36.6% untuk keempat model yang sama. Kenaikan rata-rata di atas 20% ini memperlihatkan efektivitas tahapan *pre-processing* dalam memulihkan informasi penting yang sebelumnya hilang. Proses ini membantu model menghasilkan representasi fitur yang lebih stabil sekalipun sebagian wajah tetap tidak terlihat karena tertutup oleh tangan. Emosi *neutral* dan *happy* adalah yang paling terdampak secara positif oleh metode yang diusulkan dalam kondisi wajah sebagian tertutup tangan secara vertikal. Kenaikan akurasi drastis pada emosi *neutral* (hingga +59.5% dengan CNN dan +45% dengan DeepFace) mengindikasikan bahwa representasi ekspresi *neutral* sangat terbantu dengan proses *pre-processing* menggunakan metode usulan pada area yang terhalang tangan, kemungkinan karena karakteristik ekspresi *neutral* yang kurang menonjol pada fitur-fitur dinamis seperti mulut, sehingga model lebih mudah mengenali struktur utama wajah meski sebagian tertutup.

Peningkatan tajam juga terjadi pada emosi *happy*, terlihat pada DeepFace (+54%), FaceLib (+32.16%), dan SVM (+45.5%). Hal ini mengindikasikan bahwa indikator visual *happy*, seperti pergerakan pipi dan kontraksi otot *orbicularis oculi* di sekitar mata, tetap dapat ditangkap model walaupun beberapa area wajah tertutup tangan.

Sebaliknya, emosi *surprise* dan *fear* merupakan yang paling terdampak secara negatif dengan penurunan akurasi cukup tajam, terutama pada model DeepFace dan FaceLib. Penurunan ini menandakan bahwa fitur ekspresi pada kedua emosi tersebut, seperti pelebaran mata atau perubahan tiba-tiba di area atas wajah, sangat rentan terhadap gangguan akibat oklusi tangan pada setengah bagian wajah. Dengan adanya tangan yang menutupi sebagian area mata atau dahi, pola ciri khas dari ekspresi *surprise* dan *fear* menjadi sulit dikenali oleh model. Temuan ini menunjukkan bahwa pentingnya area atas wajah dalam mendekripsi emosi yang sangat dinamis dan menekankan perlunya pengembangan algoritma yang lebih sensitif terhadap keberadaan oklusi di area penting ekspresi, seperti mata dan alis.

Dari sisi interpretasi sistem, peningkatan ini juga mengindikasikan bahwa pendekatan FER yang tangguh terhadap oklusi memerlukan penanganan secara keseluruhan pada setiap tahap deteksi atau normalisasi wajah. Dengan kata lain, bukan hanya arsitektur model yang menentukan performa akhir, tetapi juga kualitas dan keutuhan *input* yang diterima. Dalam konteks aplikasi dunia nyata penertapan FER pada *smart education* atau *smart campuss* ini, masalah tangan yang menutupi wajah harus dipertimbangkan sejak tahap desain *pipeline* agar sistem dapat tetap handal meskipun menghadapi objek wajah yang tidak sepenuhnya terbuka.

Ada kelemahan dalam metode yang diusulkan ini, dimana terdapat kondisi model *hand-detector* dari pustaka *MediaPipe* gagal dalam mendekripsi keberadaan tangan pada wajah sehingga mengagalkan proses setelahnya yang menjadikan *output* otomatis menjadi tidak seperti yang diharapkan (gambar V.15 (B)). Adapun jumlah kegagalan deteksi tangan adalah masing-masing sebanyak : *angry* 9, *disgust* 6, *fear* 13, *happy* 17, *neutral* 11, dan *surprise* 17 dari total 200 gambar per kelas emosi, jadi bisa disimpulkan tingkat kegagalan *hand-detection* di bawah 0.1%.

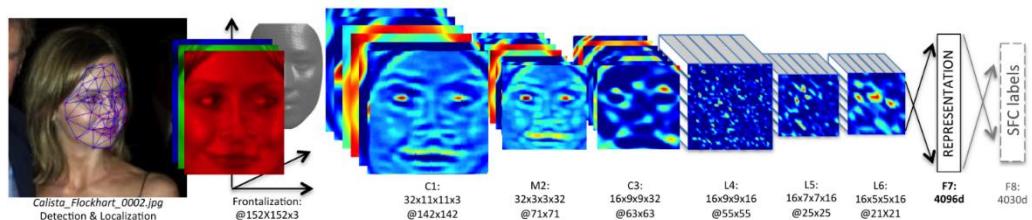


Gambar V.15 Hasil metode usulan, (A) kondisi berhasil, (B) kondisi gagal

Kegagalan model *hand-detector* dari *MediaPipe* dalam mendeteksi keberadaan tangan pada wajah umumnya disebabkan oleh kombinasi faktor teknis dan kondisi visual yang kompleks. Salah satu penyebab utama adalah *overlap* antara tangan dan wajah, dimana warna kulit yang serupa serta jarak tangan yang terlalu dekat membuat sistem kesulitan membedakan area tangan dari kontur wajah (Pu dkk., 2023). Kualitas citra dan pencahayaan juga berperan penting seperti pada kasus pencahayaan yang tidak merata, bayangan di area wajah, atau tingkat refleksi cahaya yang tinggi dapat menurunkan nilai *confidence score* di bawah ambang batas 0.7 yang telah ditetapkan oleh model *MediaPipe*. Kegagalan juga kerap terjadi ketika ukuran tangan dalam citra terlalu besar atau terlalu kecil akibat jarak kamera yang tidak sesuai dengan rasio deteksi ideal.

V.2.3 Wajah dengan Kondisi Miring (*Non-Frontal*)

Bentuk oklusi lainnya yang diuji pada penelitian ini adalah keadaan wajah *non-frontal* ataupun *pose-invariant*. Walaupun hasil *pre-processing* dengan metode usulan ini mampu menaikkan akurasi dari model FER pada data wajah yang tertutup tangan sebagian (vertikal) oleh tangan, namun metode usulan ini juga ternyata dapat menurunkan akurasi dari sebuah model FER pada kondisi wajah *non-frontal*. Hasil experimental (sub bab V.1.2) menunjukkan keadaan data “*before*” akurasi dari model CNN, DeepFace, FaceLib, dan SVM adalah masing-masing sebesar 24.05%, 35.59%, 46.51%, dan 32.88%. Setelah diterapkan *pre-processing* dengan metode usulan akurasinya menjadi 40.6%, 34.6%, 41.4%, dan 35.1%. Pengaruh positif hanya terjadi pada model CNN dan SVM, sedangkan pada model DeepFace dan FaceLib terjadi penurunan akurasi rata-rata. Kenaikan di atas 15% terjadi pada model CNN terjadi karena model dilatih dengan hanya data wajah *frontal* saja, sehingga data uji hasil *pre-processing* yang berupa wajah *frontal* dengan mudah dikenali oleh model.



Gambar V.16 Arsitektur DeepFace (Taigman dkk., 2014)

Adapun untuk DeepFace, terjadi penurunan akurasi sebesar -1.03%, karena pada DeepFace (Gambar V.16) sudah menerapkan *pre-processing* berupa proses frontalisasi pada arsitekturnya. Sehingga ketika diujikan dengan metode usulan terjadi dua kali frontalisasi yaitu pertama dengan metode usulan, lalu dilakukan frontalisasi kembali oleh Deepface. Proses ganda ini menyebabkan redundansi transformasi spasial, dimana wajah yang sebelumnya telah difrontalisasi kembali diproses pada tahap internal DeepFace, mengakibatkan perubahan kecil pada struktur wajah yang justru menurunkan ketepatan klasifikasi. Namun demikian, penurunan akurasi ini masih tergolong kecil dan tidak berdampak signifikan terhadap performa keseluruhan model. Hal yang sama terjadi pada FaceLib, dimana FaceLib sudah mempunyai *face align* pada arsitekturnya, hal tersebut tampak pada dokumentasi penggunaan kode FaceLib pada *source* resminya (Ayoubi, 2020).

Proses frontalisasi yang diterapkan dalam penelitian ini bertujuan untuk menormalkan orientasi wajah agar sejajar terhadap bidang pandang kamera. Namun demikian, proses tersebut berpotensi menimbulkan efek samping berupa penyamarataan tekstur pada area wajah tertentu (Hassner dkk., 2015). Dalam kondisi dimana salah satu sisi wajah tertutup tangan atau mengalami perbedaan distribusi pencahayaan, maka proses replikasi ini dapat menghasilkan tekstur buatan (*artificial texture*) pada area tersebut, dimana tekstur buatan ini dapat mengubah keaslian pola fitur lokal di sekitar area yang direkonstruksi, sehingga memengaruhi representasi citra yang digunakan oleh sistem pengenalan emosi. Dalam konteks ini, model FaceLib yang bergantung pada informasi spasial dan berdasarkan tekstur untuk mengekstraksi fitur ekspresi dapat terjadi salah penafsiran pola hasil replikasi tersebut sebagai karakteristik ekspresi yang berbeda dari kondisi sebenarnya. Akibatnya, proses klasifikasi emosi berpotensi mengalami penurunan akurasi akibat ketidaksesuaian struktur tekstur hasil frontalisasi.

Sejalan dengan hasil perbandingan metode usulan dengan metode frontalisasi lain yang dilakukan oleh K.Y. Tsai, et. al., (2021) (Tsai dkk., 2021) pada tabel V.12. Hasilnya menunjukkan bahwa baik metode usulan pada penelitian ini maupun metode pembanding tersebut memiliki kesimpulan yang sama yaitu adanya kenaikan akurasi pada CNN dan SVM, lalu terdapat penurunan akurasi pada DeepFace dan FaceLib, dimana seperti yang sudah dijelaskan bahwa DeepFace dan

FaceLib memiliki paket *pipeline* yang sudah memiliki proses penyesuaian (*align*) wajah di dalamnya, sehingga melakukan *pre-processing* sebelum menggunakan DeepFace dan FaceLib berpotensi untuk menurunkan akurasinya.

Secara umum, hasil ini memperlihatkan bahwa efek *pre-processing* terhadap kinerja model sangat bergantung pada desain dan karakteristik arsitektur yang digunakan. Jika tahap koreksi citra selaras dengan strategi ekstraksi fitur di dalam model, maka peningkatan akurasi dapat dicapai secara konsisten. Namun, jika transformasi awal justru berpotensi menggandakan fungsi yang telah ada pada *pipeline*, maka dampaknya bisa sebaliknya. Temuan ini menegaskan pentingnya keselarasan antara strategi *pre-processing* dan rancangan internal model agar hasil pengenalan emosi tetap optimal dalam berbagai kondisi wajah terutama pada wajah *non-frontal*.

V.2.4 Oklusi Parsial pada Area Mulut

Berdasarkan hasil perubahan rata-rata akurasi sebelum dan sesudah penerapan metode pada wajah dengan oklusi tangan di area mulut, terlihat bahwa penurunan yang terjadi relatif tidak signifikan pada sebagian besar model. Misalnya, CNN hanya mengalami peningkatan sangat kecil yaitu +0.25%, sementara DeepFace dan FaceLib justru menunjukkan penurunan rata-rata -2.93% dan -2.92%. SVM juga hanya turun sedikit sebesar -0.85%. Dapat disimpulkan bahwa metode yang diusulkan tidak memberikan dampak berarti dalam meningkatkan performa pengenalan emosi pada kasus oklusi tangan di area mulut, serta tidak memunculkan perbedaan besar dibandingkan kondisi awal. Hal ini menandakan bahwa karakteristik fitur ekspresi di area mulut yang tertutup tangan sulit dinormalkan melalui pendekatan yang digunakan, sehingga model tidak memperoleh informasi baru yang substansial untuk proses klasifikasi.

Hasil ini sejalan dengan fenomena yang telah ditemui pada penelitian pengenalan ekspresi wajah ketika area mulut tertutup, seperti pada kasus penggunaan masker medis. Berbagai studi telah membuktikan bahwa ekspresi seperti bahagia, sedih, dan marah lebih sulit dikenali jika bagian mulut dan hidung tertutup masker (Grahlow dkk., 2022)(Ruan dkk., 2022b). Penurunan akurasi juga dapat disebabkan oleh hilangnya beberapa karakteristik ekspresi penting di area tersebut seperti

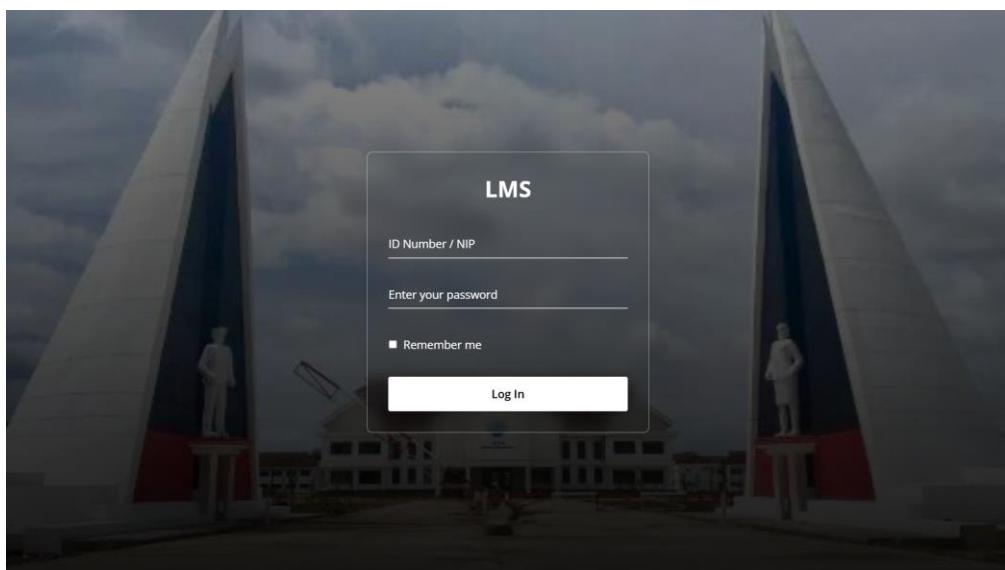
senyuman, pergerakan bibir, atau kerutan hidung, yang memiliki kontribusi besar pada deteksi emosi tertentu. Pada kondisi ini, sistem cenderung lebih bergantung pada fitur bagian atas wajah seperti mata dan alis, yang sebenarnya kurang informatif untuk sejumlah emosi.

V.3 Tampilan Antarmuka

Pada bagian ini akan dijelaskan hasil dari rancangan halaman antarmuka aplikasi LMS yang digunakan sebagai media integrasi sistem FER untuk mendapatkan data emosi siswa dalam pembelajaran *asynchronous*. Adapun bagian-bagian tampilan antarmukanya antara lain : halaman *login*, halaman *dashboard*, halaman akses materi pembelajaran *asynchronous*, dan halaman *engagement* siswa.

1. Halaman Login

Halaman *login* pada gambar V.17 merupakan tampilan pertama kali ketika pengguna membuka aplikasi dalam keadaan belum melakukan proses *login*. Adapun komponen yang terdapat pada halaman *login* antara lain, *form input username*, *password*, dan sebuah tombol *login*.

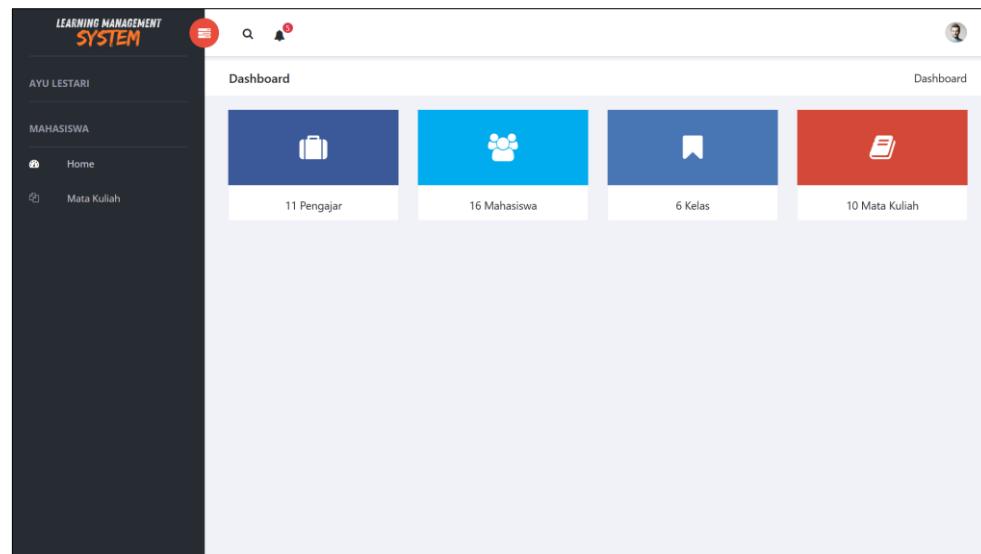


Gambar V.17 Halaman *login*

2. Halaman Dashboard

Halaman *dashboard* pada gambar V.18 adalah halaman setelah pengguna melakukan login sesuai hak aksesnya masing-masing. Halaman ini berisi

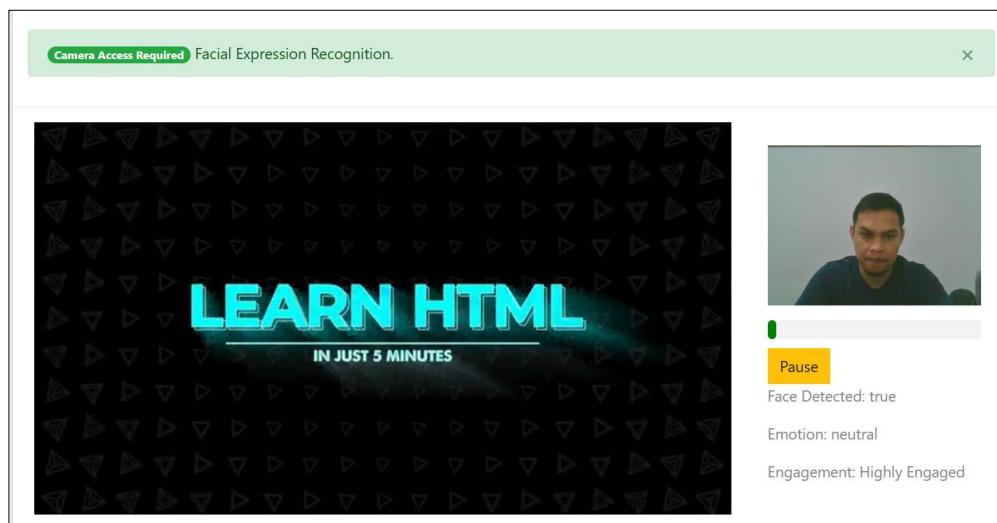
beberapa informasi seperti jumlah pengajar, jumlah mahasiswa, jumlah kelas, dan jumlah matakuliah.



Gambar V.18 Halaman *dashboard*

3. Halaman Akses Materi Pembelajaran *Asynchronous*

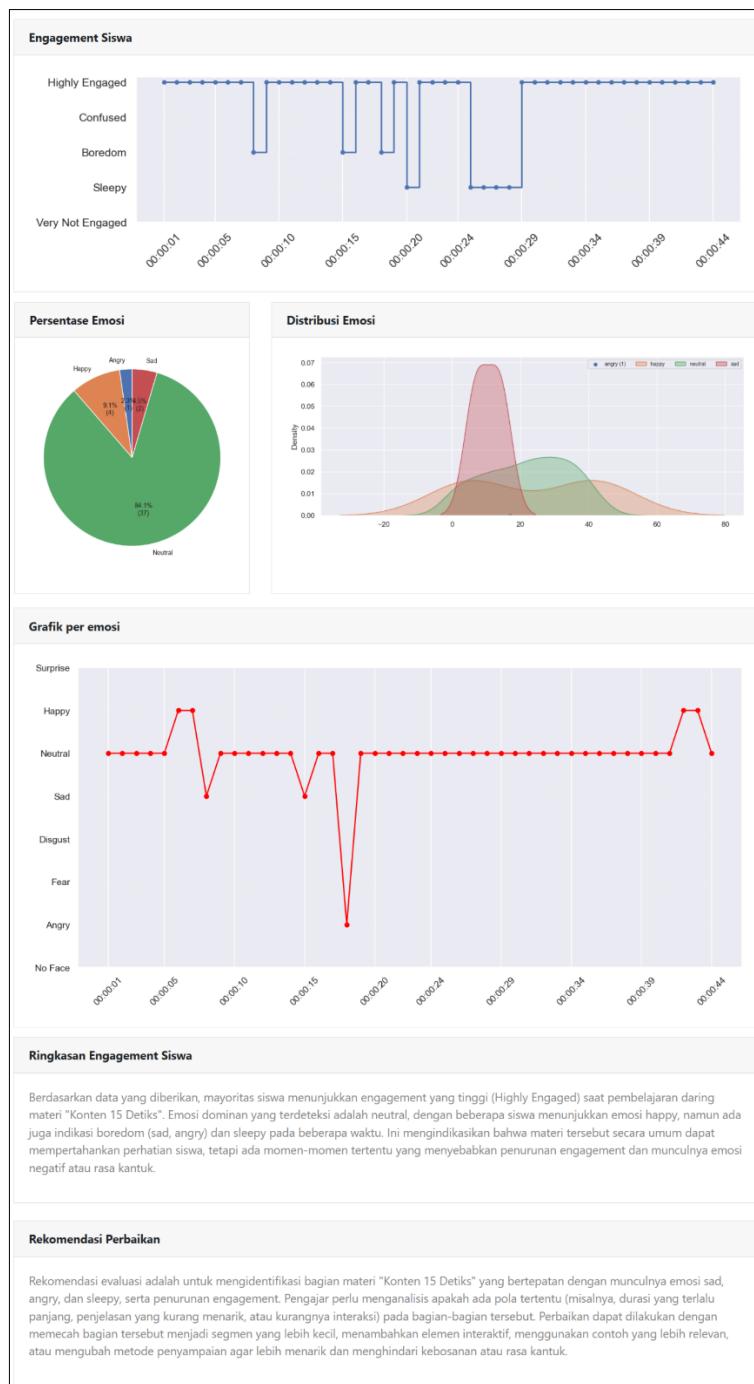
Halaman akses materi pembelajaran *asynchronous* pada gambar V.19 berisi beberapa komponen seperti *frame* video materi, *frame* tangkapan *webcam*, tombol *media-player*, dan beberapa keterangan analisis wajah yang ditangkap oleh *webcam*.



Gambar V.19 Halaman akses materi pembelajaran *asynchronous*

4. Halaman Ringkasan Engagement Siswa

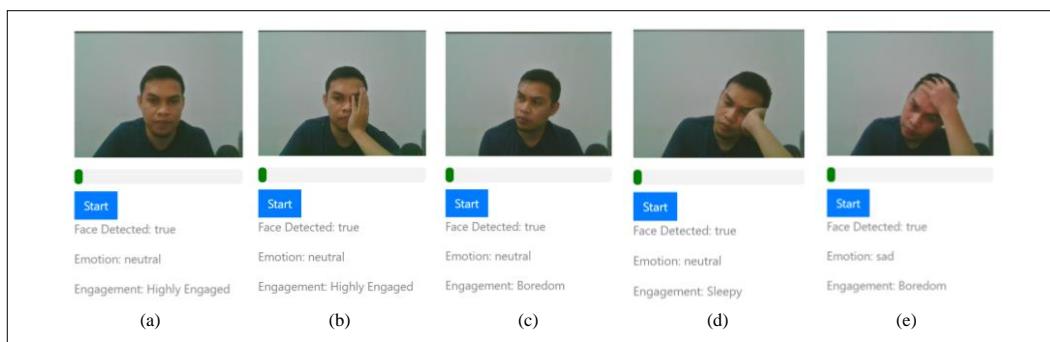
Halaman hasil ringkasan *engagement* siswa pada gambar V.20 berisi beberapa analisis dan statistik *engagement* dari siswa saat melakukan pembelajaran *asynchronous* antara lain, data fokus siswa, persentase dan distribusi emosi, grafik perubahan emosi siswa, dan sedikit interpretasi hasil menggunakan model LLM Gemini.



Gambar V.20 Halaman hasil ringkasan *engagement* siswa

V.4 Implementasi FER pada LMS

Tahapan pertama dari implementasi sistem FER pada LMS adalah menangkap wajah siswa dari sisi *front-end* yaitu menggunakan *javascript*. *Frame* wajah akan dikirimkan ke *back-end flask* untuk dilakukan *pre-processing* dengan metode usulan dan diklasifikasikan emosinya oleh model. DeepFace dipilih sebagai model deteksi emosi dalam aplikasi LMS. Beberapa pertimbangannya adalah kemampuan DeepFace yang bisa dijalankan hanya dengan menggunakan CPU (Taigman dkk., 2014) dan berbasis REST API, sehingga dapat digunakan pada perangkat yang tidak memiliki spesifikasi yang tinggi dan juga praktis dalam implementasinya. Hasil analisis wajah yang berupa data emosi, arah mata, dan kepala dikirimkan kembali oleh *back-end* ke *front-end* untuk ditampilkan dalam halaman antarmuka, dimana data tersebutlah yang akan disimpan dalam bentuk JSON dan akan menjadi bahan analisis oleh sistem dimana hasilnya akan diproses menggunakan model LLM Gemini untuk mendapat *insight* dari hasil pembelajaran siswa.



Gambar V.21 Keterangan hasil analisis wajah siswa per *frame*, (a) kondisi *highly engaged*, wajah lurus ke layar tanpa halangan, (b) kondisi *highly engaged*, wajah lurus ke layar dengan halangan tangan, (c) kondisi *boredom*, wajah tidak menghadap ke layar, (d) kondisi *sleepy*, mata tertutup, (e) kondisi *boredom* kepala ditopang tangan

Gambar IV.21 menunjukkan contoh beberapa kondisi wajah siswa saat pembelajaran *asynchronous*, yaitu kondisi ketika siswa menonton materi yang diberikan pengajar dalam bentuk video. Aplikasi hanya diberi fitur tombol *play*, *pause*, dan *replay* sehingga siswa tidak bisa mempercepat video pembelajaran. Tampak pada gambar IV.21 dimana ada beberapa skenario yang diujikan, mulai dengan *engagement highly engaged*, hingga pengujian dalam kondisi *boredom* atau bosan dan *sleepy*. Pengujian dilakukan dengan mengkondisikan beberapa gestur tangan sesuai dengan skenario yang dibuat.

Gambar V.21 (b) tampak sistem masih bisa melakukan klasifikasi emosi dengan tepat walaupun terdapat tangan yang menutupi sebagian wajah. Pada halaman antarmuka aplikasi juga tidak ditampilkan proses-proses *back-end*. Aplikasi hanya akan menampilkan kondisi wajah siswa saja dengan hasil yang sudah diklasifikasi oleh model FER. Hal tersebut atas pertimbangan bahwa ketika proses *back-end* ditampilkan, maka akan memperberat kinerja di bagian *front-end* sehingga bisa mengganggu kenyamanan pengguna.

BAB VI Kesimpulan

Deteksi tingkat *engagement* siswa pada pembelajaran daring berdasarkan emosi wajah dan sikap siswa yang berupa data arah kepala dan keadaan mata dapat memberikan gambaran keterlibatan peserta didik secara lebih objektif. Melalui kombinasi analisis ekspresi wajah, arah pandang, dan status mata, sistem dapat mengidentifikasi dinamika perhatian serta tingkat *engagement* siswa selama proses pembelajaran, khususnya pada pembelajaran berbasis video *asynchronous* yang memiliki kelemahan terhadap pengawasan siswa dalam proses belajarnya. Hasil pemantauan tersebut menunjukkan bahwa perilaku visual seperti tatapan yang mengarah ke layar dan mata terbuka disertai emosi positif seperti *neutral* dan *happy*, cenderung diasosiasikan dengan *engagement* yang tinggi. Sebaliknya, arah kepala yang menjauh atau tidak melihat layar, mata terpejam, atau emosi negatif, mengindikasikan *engagement* yang menurun. Dengan demikian, pengukuran multi-modal berbasis visual ini dapat digunakan sebagai indikator penting untuk mengevaluasi dan meningkatkan efektivitas interaksi siswa dalam pembelajaran daring, yang pada akhirnya turut berpengaruh terhadap capaian akademik mereka.

Adapun fokus penelitian ini menunjukkan bahwa metode yang diusulkan memberikan hasil yang paling kuat pada kasus wajah yang tertutup sebagian secara vertikal oleh tangan, dimana sebagian area wajah yang tertutup adalah pada posisi mata, pipi, alis dan sebagian mulut. Pada kondisi ini, beberapa model seperti CNN dan DeepFace menunjukkan peningkatan performa yang jelas, dengan rata-rata peningkatan akurasi mencapai 33.79% pada CNN dan 14.29% pada DeepFace. Sementara itu, untuk wajah dengan kemiringan *pose (non-frontal)*, metode ini juga memberikan kenaikan akurasi, meskipun bersifat tidak signifikan, menunjukkan potensi model untuk beradaptasi pada variasi orientasi wajah. Sebaliknya, pada wajah dengan oklusi di area mulut, metode ini tidak memberikan pengaruh yang berarti terhadap akurasi deteksi emosi, dengan beberapa model justru mengalami penurunan performa, seperti DeepFace (-2.93%) dan FaceLib (-2.92%). Hal ini menandakan bahwa kehilangan informasi pada area mulut sangat berpengaruh terhadap kemampuan sistem dalam menafsirkan ekspresi emosional, terutama

ekspresi seperti *happy*, *sad*, dan *fear* yang umumnya sangat bergantung pada gerakan bibir.

Meskipun demikian, hasil penelitian ini tetap memperlihatkan bahwa pendekatan yang diusulkan berpotensi meningkatkan *robustness* sistem pengenalan ekspresi terhadap kondisi oklusi parsial tertentu. Faktor lain seperti kegagalan deteksi tangan oleh *hand-detector MediaPipe* juga menjadi aspek yang memengaruhi kinerja keseluruhan sistem, khususnya pada kondisi pencahayaan rendah atau pose tangan ekstrem. Keterbatasan ini membuka ruang bagi pengembangan model oklusi-adaptif yang lebih sensitif terhadap variasi posisi tangan di masa mendatang.

Perlu dicatat bahwa hasil penelitian ini terbatas hanya pada *dataset* yang digunakan dalam eksperimen ini, yang memiliki karakteristik dan distribusi citra tertentu. Oleh karena itu, performa metode mungkin akan berbeda apabila diuji pada *dataset* dengan resolusi lebih beragam, tingkat oklusi lebih kompleks, atau latar visual yang berbeda. Penelitian lanjutan dengan menggunakan *dataset* skenario yang beragam dan pendekatan *multi-modality* disarankan untuk mengonfirmasi efektivitas metode secara lebih general serta memperluas penerapannya pada skenario dunia nyata.

DAFTAR PUSTAKA

- Afzal, S., Khan, H. A., Piran, M. J., & Lee, J. W. (2024). A Comprehensive Survey on Affective Computing: Challenges, Trends, Applications, and Future Directions. *IEEE Access*, 12, 96150–96168. <https://doi.org/10.1109/ACCESS.2024.3422480>
- Alotaibi, N. S. (2024). The Impact of AI and LMS Integration on the Future of Higher Education: Opportunities, Challenges, and Strategies for Transformation. *Sustainability (Switzerland)*, 16(23). <https://doi.org/10.3390/su162310357>
- Aly, M., Ghallab, A., & Fathi, I. S. (2023). Enhancing Facial Expression Recognition System in Online Learning Context Using Efficient Deep Learning Model. *IEEE Access*, 11(November), 121419–121433. <https://doi.org/10.1109/ACCESS.2023.3325407>
- Awais, M., Raza, M., Singh, N., Bashir, K., Manzoor, U., Islam, S. U., & Rodrigues, J. J. P. C. (2021). LSTM-Based Emotion Detection Using Physiological Signals: IoT Framework for Healthcare and Distance Learning in COVID-19. *IEEE Internet of Things Journal*, 8(23), 16863–16871. <https://doi.org/10.1109/JIOT.2020.3044031>
- Ayoubi, S. (2020). *FaceLib: Face Analysis*. <https://github.com/sajjadayobi/FaceLib>
- Bao, J., Tao, X., & Zhou, Y. (2024a). An Emotion Recognition Method Based on Eye Movement and Audiovisual Features in MOOC Learning Environment. *IEEE Transactions on Computational Social Systems*, 11(1), 171–183. <https://doi.org/10.1109/TCSS.2022.3221128>
- Bao, J., Tao, X., & Zhou, Y. (2024b). An Emotion Recognition Method Based on Eye Movement and Audiovisual Features in MOOC Learning Environment. *IEEE Transactions on Computational Social Systems*, 11(1), 171–183. <https://doi.org/10.1109/TCSS.2022.3221128>
- Beekman, R., Tijssen, C. C., Visser, L. H., & Schellens, R. L. L. A. (2002). Dropped head as the presenting symptom of primary hyperparathyroidism [3]. *Journal of Neurology*, 249(12), 1738–1739. <https://doi.org/10.1007/s00415-002-0898-7>
- Bekmanova, G., Yergesh, B., Sharipbay, A., & Mukanova, A. (2022). Emotional Speech Recognition Method Based on Word Transcription. *Sensors*, 22(5). <https://doi.org/10.3390/s22051937>
- Benabbes, K., Housni, K., Hmedna, B., Zellou, A., & Mezouary, A. El. (2023). A New Hybrid Approach to Detect and Track Learner's Engagement in e-Learning. *IEEE Access*, 11(June), 70912–70929. <https://doi.org/10.1109/ACCESS.2023.3293827>
- Bhardwaj, P., Gupta, P. K., Panwar, H., Siddiqui, M. K., Morales-Menendez, R., & Bhaik, A. (2021). Application of Deep Learning on Student Engagement in e-learning environments. *Computers and Electrical Engineering*, 93(August 2020), 107277. <https://doi.org/10.1016/j.compeleceng.2021.107277>
- Cao, T., Liu, C., & Chen, J. (2021). Nonfrontal Expression Recognition in the Wild Based on PRNet Frontalization and Muscle Feature Strengthening.

- Mathematical Problems in Engineering*, 2021, 1–21.
<https://doi.org/10.1155/2021/6620752>
- Cheong, J. H., Jolly, E., Xie, T., Byrne, S., Kenney, M., & Chang, L. J. (2023). Py-Feat: Python Facial Expression Analysis Toolbox. *Affective Science*, 4(4), 781–796. <https://doi.org/10.1007/s42761-023-00191-4>
- Costley, J., Southam, A., Bailey, D., & Haji, S. A. (2022). How use of learning management system mediates the relationships between learner interactions and learner outcomes. *Interactive Technology and Smart Education*, 19(2), 184–201. <https://doi.org/10.1108/ITSE-12-2020-0236>
- Cui, Y., Wang, S., & Zhao, R. (2021). Machine Learning-Based Student Emotion Recognition for Business English Class. *International Journal of Emerging Technologies in Learning*, 16(12), 94–107. <https://doi.org/10.3991/ijet.v16i12.23313>
- Das, A., Sarma, M. Sen, Hoque, M. M., Siddique, N., & Dewan, M. A. A. (2024). AVaTER: Fusing Audio, Visual, and Textual Modalities Using Cross-Modal Attention for Emotion Recognition. *Sensors*, 24(18), 5862. <https://doi.org/10.3390/s24185862>
- Engin, D., Ecabert, C., Ekenel, H. K., & Thiran, J. P. (2018). Face frontalization for cross-pose facial expression recognition. *European Signal Processing Conference, 2018-September*, 1795–1799. <https://doi.org/10.23919/EUSIPCO.2018.8553087>
- Fernandez, J., Martínez, R., Innocenti, B., & López, B. (2024). Contribution of EEG Signals for Students' Stress Detection. *IEEE Transactions on Affective Computing*, 1–12. <https://doi.org/10.1109/TAFFC.2024.3503995>
- Freire-obregón, J. S. J. L. D., & Castrillón-santana, M. (2024). Multimodal emotion recognition based on a fusion of audiovisual information with temporal dynamics. *Multimedia Tools and Applications*. <https://doi.org/10.1007/s11042-024-20227-6>
- Gamage, G., De Silva, D., Mills, N., Alahakoon, D., & Manic, M. (2024). Emotion AWARE: an artificial intelligence framework for adaptable, robust, explainable, and multi-granular emotion analysis. *Journal of Big Data*, 11(1). <https://doi.org/10.1186/s40537-024-00953-2>
- Grahlow, M., Rupp, C. I., & Derntl, B. (2022). The impact of face masks on emotion recognition performance and perception of threat. *PLOS ONE*, 17(2), e0262840. <https://doi.org/10.1371/journal.pone.0262840>
- Guan, Y. (2024). Characteristics of teaching English and American literature based on multimodality under language acquisition theory. *Applied Mathematics and Nonlinear Sciences*, 9(1), 1–16. <https://doi.org/10.2478/amns.2023.2.00978>
- Gupta, S., Kumar, P., & Tekchandani, R. (2023a). A multimodal facial cues based engagement detection system in e-learning context using deep learning approach. *Multimedia Tools and Applications*, 82(18), 28589–28615. <https://doi.org/10.1007/s11042-023-14392-3>
- Gupta, S., Kumar, P., & Tekchandani, R. (2023b). An optimized deep convolutional neural network for adaptive learning using feature fusion in multimodal data. *Decision Analytics Journal*, 8(May), 100277. <https://doi.org/10.1016/j.dajour.2023.100277>

- Gupta, S., Kumar, P., & Tekchandani, R. K. (2023c). Facial emotion recognition based real-time learner engagement detection system in online learning context using deep learning models. *Multimedia Tools and Applications*, 82(8), 11365–11394. <https://doi.org/10.1007/s11042-022-13558-9>
- Harb, A., Gad, A., Yaghi, M., Alhalabi, M., Zia, H., Yousaf, J., Khelifi, A., Ghoudi, K., & Ghazal, M. (2023). Diverse distant-students deep emotion recognition and visualization. *Computers and Electrical Engineering*, 111(PB), 108963. <https://doi.org/10.1016/j.compeleceng.2023.108963>
- Hassner, T., Harel, S., Paz, E., & Enbar, R. (2015). Effective face frontalization in unconstrained images. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 07-12-June*, 4295–4304. <https://doi.org/10.1109/CVPR.2015.7299058>
- Herrero-Alvarez, R., Callejas-Castro, E., Miranda, G., & Leon, C. (2024). Analysis of Sentiment Toward Computer Science in Pre-University Education. *IEEE Access*, 12(April), 71205–71218. <https://doi.org/10.1109/ACCESS.2024.3402991>
- Hunukumbure, A. D., Horner, P. J., Fox, J., & Thakerar, V. (2021). An online discussion between students and teachers: a way forward for meaningful teacher feedback? *BMC Medical Education*, 21(1), 1–9. <https://doi.org/10.1186/s12909-021-02730-8>
- Jain, V., & Learned-Miller, E. (2010). Fddb: A benchmark for face detection in unconstrained settings. *UMass Amherst Technical Report, January*. http://works.bepress.com/erik_learned_miller/55/
- Jang, Y., Gunes, H., & Patras, I. (2019). Registration-free Face-SSD: Single shot analysis of smiles, facial attributes, and affect in the wild. *Computer Vision and Image Understanding*, 182(July 2018), 17–29. <https://doi.org/10.1016/j.cviu.2019.01.006>
- Kanwal, S., & Asghar, S. (2021). Speech Emotion Recognition Using Clustering Based GA-Optimized Feature Set. *IEEE Access*, 9, 125830–125842. <https://doi.org/10.1109/ACCESS.2021.3111659>
- Khan, U. A., Xu, Q., Liu, Y., Lagstedt, A., Alamäki, A., & Kauttonen, J. (2024). Exploring contactless techniques in multimodal emotion recognition: insights into diverse applications, challenges, solutions, and prospects. In *Multimedia Systems* (Vol. 30, Nomor 3). Springer Berlin Heidelberg. <https://doi.org/10.1007/s00530-024-01302-2>
- Khoshnam, F., & Baraani-Dastjerdi, A. (2022). A dual framework for implicit and explicit emotion recognition: An ensemble of language models and computational linguistics. *Expert Systems with Applications*, 198(October 2021), 116686. <https://doi.org/10.1016/j.eswa.2022.116686>
- Kim, J., & Lee, D. (2023). Facial Expression Recognition Robust to Occlusion and to Intra-Similarity Problem Using Relevant Subsampling. *Sensors*, 23(5). <https://doi.org/10.3390/s23052619>
- Le, H. D., Lee, G. S., Kim, S. H., Kim, S., & Yang, H. J. (2023). Multi-Label Multimodal Emotion Recognition With Transformer-Based Fusion and Emotion-Level Representation Learning. *IEEE Access*, 11(February), 14742–14751. <https://doi.org/10.1109/ACCESS.2023.3244390>
- Li, S. (2025). Application of entertainment e-learning mode based on genetic algorithm and facial emotion recognition in environmental art and design

- courses. *Entertainment Computing*, 52(May 2024), 100798. <https://doi.org/10.1016/j.entcom.2024.100798>
- Li, S., & Deng, W. (2022a). A Deeper Look at Facial Expression Dataset Bias. *IEEE Transactions on Affective Computing*, 13(2), 881–893. <https://doi.org/10.1109/TAFFC.2020.2973158>
- Li, S., & Deng, W. (2022b). Deep Facial Expression Recognition: A Survey. *IEEE Transactions on Affective Computing*, 13(3), 1195–1215. <https://doi.org/10.1109/TAFFC.2020.2981446>
- Lin, H. C. K., Liao, Y. C., & Wang, H. T. (2022). Eye Movement Analysis and Usability Assessment on Affective Computing Combined with Intelligent Tutoring System. *Sustainability (Switzerland)*, 14(24). <https://doi.org/10.3390/su142416680>
- Liu, T., Wang, M., Yang, B., Liu, H., & Yi, S. (2025). ESERNet: Learning spectrogram structure relationship for effective speech emotion recognition with swin transformer in classroom discourse analysis. *Neurocomputing*, 612(July 2024), 128711. <https://doi.org/10.1016/j.neucom.2024.128711>
- Liu, Z. T., Rehman, A., Wu, M., Cao, W. H., & Hao, M. (2021). Speech emotion recognition based on formant characteristics feature extraction and phoneme type convergence. *Information Sciences*, 563, 309–325. <https://doi.org/10.1016/j.ins.2021.02.016>
- Llurba, C., Fretes, G., & Palau, R. (2024). Classroom Emotion Monitoring Based on Image Processing. *Sustainability (Switzerland)*, 16(2). <https://doi.org/10.3390/su16020916>
- Malakar, S., Chiracharit, W., & Chamnongthai, K. (2024). Masked Face Recognition with Generated Occluded Part using Image Augmentation and CNN Maintaining Face Identity. *IEEE Access*, 12(August), 126356–126375. <https://doi.org/10.1109/ACCESS.2024.3446652>
- Mamieva, D., Abdusalomov, A. B., Kutlimuratov, A., Muminov, B., & Whangbo, T. K. (2023). Multimodal Emotion Detection via Attention-Based Fusion of Extracted Facial and Speech Features. *Sensors*, 23(12). <https://doi.org/10.3390/s23125475>
- Mao, J., Qian, Z., & Lucas, T. (2023). Sentiment Analysis of Animated Online Education Texts Using Long Short-Term Memory Networks in the Context of the Internet of Things. *IEEE Access*, 11(August), 109121–109130. <https://doi.org/10.1109/ACCESS.2023.3321303>
- Mattioli, M., & Cabitza, F. (2024). Not in My Face: Challenges and Ethical Considerations in Automatic Face Emotion Recognition Technology. *Machine Learning and Knowledge Extraction*, 6(4), 2201–2231. <https://doi.org/10.3390/make6040109>
- Mehta, N. K., Prasad, S. S., Saurav, S., Saini, R., & Singh, S. (2022). Three-dimensional DenseNet self-attention neural network for automatic detection of student's engagement. *Applied Intelligence*, 52(12), 13803–13823. <https://doi.org/10.1007/s10489-022-03200-4>
- Mensah, J. A., Nortey, E. N. N., Ocran, E., Iddi, S., & Asiedu, L. (2024). De-occlusion and recognition of frontal face images: a comparative study of multiple imputation methods. *Journal of Big Data*, 11(1). <https://doi.org/10.1186/s40537-024-00925-6>

- Meriem, B., Benlahmar, H., Naji, M. A., Sanaa, E., & Wijdane, K. (2022). Determine the Level of Concentration of Students in Real Time from their Facial Expressions. *International Journal of Advanced Computer Science and Applications*, 13(1), 159–166. <https://doi.org/10.14569/IJACSA.2022.0130119>
- Moise, G., Dragomir, E. G., Şchiopu, D., & Iancu, L. A. (2024). Towards Integrating Automatic Emotion Recognition in Education: A Deep Learning Model Based on 5 EEG Channels. *International Journal of Computational Intelligence Systems*, 17(1). <https://doi.org/10.1007/s44196-024-00638-x>
- Morar, A., Moldoveanu, F., Moldoveanu, A., & Oana Balan, V. A. (2017). GPU accelerated 2D and 3D image processing. *Proceedings of the 2017 Federated Conference on Computer Science and Information Systems, FedCSIS 2017*, 11, 653–656. <https://doi.org/10.15439/2017F265>
- Ngo, D., Nguyen, A., Dang, B., & Ngo, H. (2024). Facial Expression Recognition for Examining Emotional Regulation in Synchronous Online Collaborative Learning. *International Journal of Artificial Intelligence in Education*, 34(3), 650–669. <https://doi.org/10.1007/s40593-023-00378-7>
- Orosoo, M., Rajkumari, Y., Ramesh, K., Fatma, G., Nagabhaskar, M., Gopi, A., & Rengarajan, M. (2024). Enhancing English Learning Environments Through Real-Time Emotion Detection and Sentiment Analysis. *International Journal of Advanced Computer Science and Applications*, 15(7), 875–889. <https://doi.org/10.14569/IJACSA.2024.0150787>
- Pan, T., Ye, Y., Zhang, Y., Xiao, K., & Cai, H. (2024). Online multi-hypergraph fusion learning for cross-subject emotion recognition. *Information Fusion*, 108(February), 102338. <https://doi.org/10.1016/j.inffus.2024.102338>
- Pascual, A. M., Valverde, E. C., Kim, J. I., Jeong, J. W., Jung, Y., Kim, S. H., & Lim, W. (2022). Light-FER: A Lightweight Facial Emotion Recognition System on Edge Devices. *Sensors*, 22(23), 1–9. <https://doi.org/10.3390/s22239524>
- Peterson, A. T. (2023). Asynchrony and promotive interaction in online cooperative learning. *International Journal of Educational Research Open*, 5(September), 100300. <https://doi.org/10.1016/j.ijedro.2023.100300>
- Pordoy, J., Farman, H., Dicheva, N., Anwar, A., Nasralla, M. M., Khilji, N., & Rehman, I. U. (2024). Multi-Frame Transfer Learning Framework for Facial Emotion Recognition in e-Learning Contexts. *IEEE Access*, 12(October), 151360–151381. <https://doi.org/10.1109/ACCESS.2024.3478072>
- Poux, D., Allaert, B., Ihaddadene, N., Bilasco, I. M., Djeraba, C., & Bennamoun, M. (2022). Dynamic Facial Expression Recognition under Partial Occlusion with Optical Flow Reconstruction. *IEEE Transactions on Image Processing*, 31, 446–457. <https://doi.org/10.1109/TIP.2021.3129120>
- Pu, M., Chong, C. Y., & Lim, M. K. (2023). Robustness Evaluation in Hand Pose Estimation Models using Metamorphic Testing. *2023 IEEE/ACM 8th International Workshop on Metamorphic Testing (MET)*, 31–38. <https://doi.org/10.1109/MET59151.2023.00012>
- Qu, Z., Shang, X., Xia, S. F., Yi, T. M., & Zhou, D. Y. (2022). A method of single-shot target detection with multi-scale feature fusion and feature

- enhancement. *IET Image Processing*, 16(6), 1752–1763. <https://doi.org/10.1049/ipr2.12445>
- Rathod, M., Dalvi, C., Kaur, K., Patil, S., Gite, S., Kamat, P., Kotecha, K., Abraham, A., & Gabralla, L. A. (2022). Kids' Emotion Recognition Using Various Deep-Learning Models with Explainable AI. *Sensors*, 22(20). <https://doi.org/10.3390/s22208066>
- Roy, S., Gaur, V., Raza, H., & Jameel, S. (2023). CLEFT: Contextualised Unified Learning of User Engagement in Video Lectures With Feedback. *IEEE Access*, 11(January), 17707–17720. <https://doi.org/10.1109/ACCESS.2023.3245982>
- Ruan, L., Han, Y., Sun, J., Chen, Q., & Li, J. (2022a). Facial expression recognition in facial occlusion scenarios: A path selection multi-network. *Displays*, 74(June), 102245. <https://doi.org/10.1016/j.displa.2022.102245>
- Ruan, L., Han, Y., Sun, J., Chen, Q., & Li, J. (2022b). Facial expression recognition in facial occlusion scenarios: A path selection multi-network. *Displays*, 74, 102245. <https://doi.org/10.1016/j.displa.2022.102245>
- Sassi, A., Jaafar, W., Cherif, S., Abderrazak, J. Ben, & Yanikomeroglu, H. (2023). Video Traffic Analysis for Real-Time Emotion Recognition and Visualization in Online Learning. *IEEE Access*, 11(July), 99376–99386. <https://doi.org/10.1109/ACCESS.2023.3313973>
- Shi, L. (2024). The Integration of Advanced AI-Enabled Emotion Detection and Adaptive Learning Systems for Improved Emotional Regulation. *Journal of Educational Computing Research*, 63(1), 173–201. <https://doi.org/10.1177/07356331241296890>
- Shobana, B. T., & Kumar, G. A. S. (2021). I-Quiz: An Intelligent Assessment Tool for Non-Verbal Behaviour Detection. *Computer Systems Science and Engineering*, 40(3), 1007–1021. <https://doi.org/10.32604/CSSE.2022.019523>
- Sowjanya, U. L., & Krithiga, R. (2024). Decoding Student Emotions: An Advanced CNN Approach for Behavior Analysis Application Using Uniform Local Binary Pattern. *IEEE Access*, 12(August), 106273–106284. <https://doi.org/10.1109/ACCESS.2024.3436531>
- Sukumaran, A., & Manoharan, A. (2024). Multimodal Engagement Recognition From Image Traits Using Deep Learning Techniques. *IEEE Access*, 12(December 2023), 25228–25244. <https://doi.org/10.1109/ACCESS.2024.3353053>
- Taigman, Y., Yang, M., Ranzato, M., & Wolf, L. (2014). DeepFace: Closing the gap to human-level performance in face verification. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1701–1708. <https://doi.org/10.1109/CVPR.2014.220>
- Tanko, D., Dogan, S., Burak Demir, F., Baygin, M., Engin Sahin, S., & Tuncer, T. (2022). Shoelace pattern-based speech emotion recognition of the lecturers in distance education: ShoePat23. *Applied Acoustics*, 190, 108637. <https://doi.org/10.1016/j.apacoust.2022.108637>
- Tao, X., Shannon-Honson, A., Delaney, P., Li, L., Dann, C., Li, Y., & Xie, H. (2022). Data Analytics on Online Student Engagement Data for Academic Performance Modeling. *IEEE Access*, 10(October), 103176–103186. <https://doi.org/10.1109/ACCESS.2022.3208953>

- Tariyal, S., Chauhan, R., Bijalwan, Y., Rawat, R., & Gupta, R. (2024). A comparative study of MTCNN, Viola-Jones, SSD and YOLO face detection algorithms. *2024 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE)*, 1–7. <https://doi.org/10.1109/IITCEE59897.2024.10467445>
- Tian, X., Nunes, B. P., Liu, Y., & Manrique, R. (2024). Predicting Student Engagement Using Sequential Ensemble Model. *IEEE Transactions on Learning Technologies*, 17, 939–950. <https://doi.org/10.1109/TLT.2023.3342860>
- Trabelsi, Z., Alnajjar, F., Parambil, M. M. A., Gochoo, M., & Ali, L. (2023). Real-Time Attention Monitoring System for Classroom: A Deep Learning Approach for Student's Behavior Recognition. *Big Data and Cognitive Computing*, 7(1), 1–17. <https://doi.org/10.3390/bdcc7010048>
- Tran, T. P., & Meacheam, D. (2020). Enhancing Learners' Experience through Extending Learning Systems. *IEEE Transactions on Learning Technologies*, 13(3), 540–551. <https://doi.org/10.1109/TLT.2020.2989333>
- Trigka, M., & Dritsas, E. (2025). A Comprehensive Survey of Deep Learning Approaches in Image Processing. *Sensors*, 25(2). <https://doi.org/10.3390/s25020531>
- Tsai, K. Y., Tsai, Y. W., Lee, Y. C., Ding, J. J., & Chang, R. Y. (2021). Frontalization and adaptive exponential ensemble rule for deep-learning-based facial expression recognition system. *Signal Processing: Image Communication*, 96(April), 116321. <https://doi.org/10.1016/j.image.2021.116321>
- Ülgen Sönmez, Yeşim Varol, A. (2024). In-depth investigation of speech emotion recognition studies from past to present –The importance of emotion recognition from speech signal for AI-. *Intelligent Systems with Applications*, 22(March). <https://doi.org/10.1016/j.iswa.2024.200351>
- Villegas-Ch, W., Garcia-Ortiz, J., & Sanchez-Viteri, S. (2024). Personalization of Learning: Machine Learning Models for Adapting Educational Content to Individual Learning Styles. *IEEE Access*, 12(July), 121114–121130. <https://doi.org/10.1109/ACCESS.2024.3452592>
- Wang, S. (2021). Online learning behavior analysis based on image emotion recognition. *Traitement du Signal*, 38(3), 865–873. <https://doi.org/10.18280/ts.380333>
- Wickline, V. B., Hall, A. S., Lavrisa, R., McCook, K., Woodcock, M., Bani, M., Strepparava, M. G., Russo, S., & Nowicki, S. (2025). Facial occlusion with medical masks: Impacts on emotion recognition rates for emotion types and intensities. *Quarterly Journal of Experimental Psychology*. <https://doi.org/10.1177/17470218241308569>
- Xu, Z. (2023). Application of Convolution Neural Network Algorithm in Online Education Emotion Recognition. *International Journal of Web-Based Learning and Teaching Technologies*, 18(2), 1–13. <https://doi.org/10.4018/IJWLTT.331077>
- Yu, S., Androsov, A., Yan, H., & Chen, Y. (2024). Bridging computer and education sciences: A systematic review of automated emotion recognition in online learning environments. *Computers and Education*, 220(3663), 105111. <https://doi.org/10.1016/j.compedu.2024.105111>

- Zaareer, Z., Alhatemi, S. A. A., Alotaibi, A. A., Soliman, A. F., Abdulghani, S. R., Hassan, S., Abouelnaga, H. M., Azaizeh, A., & Jaradat, A. (2024). Investigating the Impact of LMS Quality , Technical Support and Perceived Usefulness on Student Satisfaction in Saudi Universities. *Evolutionary Studies in Imaginative Culture*, 8.
- Zeng, H., Shu, X., Wang, Y., Wang, Y., Zhang, L., Pong, T. C., & Qu, H. (2021). EmotionCues: Emotion-Oriented Visual Summarization of Classroom Videos. *IEEE Transactions on Visualization and Computer Graphics*, 27(7), 3168–3181. <https://doi.org/10.1109/TVCG.2019.2963659>
- Zhang, N., Liu, N., Han, J., Wan, K., & Shao, L. (2023). Face De-Occlusion With Deep Cascade Guidance Learning. *IEEE Transactions on Multimedia*, 25, 3217–3229. <https://doi.org/10.1109/TMM.2022.3157036>
- Zhang, S., Yang, Y., Chen, C., Zhang, X., Leng, Q., & Zhao, X. (2024). Deep learning-based multimodal emotion recognition from audio, visual, and text modalities: A systematic review of recent advancements and future prospects. *Expert Systems with Applications*, 237(PC), 121692. <https://doi.org/10.1016/j.eswa.2023.121692>
- Zhang, X., & Qin, X. (2022). Research on Sentiment Analysis Algorithm for Comments on Online Ideological and Political Courses. *International Journal of Advanced Computer Science and Applications*, 13(11), 174–179. <https://doi.org/10.14569/IJACSA.2022.0131119>
- Zhang, X., Zhang, F., & Xu, C. (2022). Joint Expression Synthesis and Representation Learning for Facial Expression Recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(3), 1681–1695. <https://doi.org/10.1109/TCSVT.2021.3056098>
- Zhu-Zhou, F., Gil-Pita, R., García-Gómez, J., & Rosa-Zurera, M. (2022). Robust Multi-Scenario Speech-Based Emotion Recognition System. *Sensors*, 22(6). <https://doi.org/10.3390/s22062343>
- Zhu, C., Wan, X., Xie, S., Li, X., & Gu, Y. (2022). Occlusion-robust Face Alignment using A Viewpoint-invariant Hierarchical Network Architecture. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2022-June, 11102–11111. <https://doi.org/10.1109/CVPR52688.2022.01083>
- Zhu, H., Xu, H., Ma, X., & Bian, M. (2022). Facial Expression Recognition Using Dual Path Feature Fusion and Stacked Attention. *Future Internet*, 14(9). <https://doi.org/10.3390/fi14090258>
- Zhu, K., He, X., Lv, Z., Zhang, X., Hao, R., He, X., Wang, J., He, J., Zhang, L., & Mu, Z. (2023). A 3D Occlusion Facial Recognition Network Based on a Multi-Feature Combination Threshold. *Applied Sciences (Switzerland)*, 13(10). <https://doi.org/10.3390/app13105950>

LAMPIRAN

Lampiran A *Source code* fungsi utama metode usulan (*Half-Flip 2D*)

```
def half_flip(img):
    try:
        images = img
        img_h, img_w, _ = images.shape
        if img_h < 500:
            new_height = 500
            (h, w) = images.shape[:2]
            aspect_ratio = w / h
            new_width = int(new_height * aspect_ratio)
            images = cv2.resize(images, (new_width, new_height))

        output_list = []
        landm = images.copy()
        land, _ = landmark_wajah(landm)
        output_list.append((land, cv2.COLOR_RGB2BGR))

        img_ori = images.copy()
        points_, _ = get_face_mesh_3d(img_ori)
        pitch, roll, yaw = compute_head_angle(points_)

        img_roll, face_detected = correct_roll(img_ori, roll, yaw)
        img_r = img_roll.copy()
        img_rr = img_roll.copy()
        img_rr = blurring_hand(img_rr)

        if face_detected == True:
            points_3d, _ = get_face_mesh_3d(img_r)
            _, _, yaw_angle = compute_head_angle(points_3d)
            img_height, img_width, _ = img_roll.shape
            _, list_hand = landmark_tangan(img_roll)
            sum_of_hand = len(list_hand)
            if sum_of_hand > 0:
                hand_masked = masking_tangan_canvas_hitam(img_roll)
                left, right, occlusion = split_with_padding(hand_masked)
                combine_face_hand = 'True'
            else:
                occlusion = None
                combine_face_hand = False

            _, landmark_face = landmark_wajah(img_roll)
            dua_D = []
            for i in range(len(landmark_face)):
                ix = int(landmark_face[i].x * img_width)
                iy = int(landmark_face[i].y * img_height)
                dua_D.append([ix, iy])

            poin_wajah_full = list_poin_wajah(dua_D)
            wajah_full_masked = masking_img(img_roll, poin_wajah_full, 'putih')
            if combine_face_hand == False and yaw_angle < 2:
                frontal_miring = main_front(img_r)
                frontal_miring = cv2.cvtColor(frontal_miring, cv2.COLOR_BGR2RGB)
                return frontal_miring

            poin_wajah_kiri = wajah_kiri(dua_D)
            face_kiri_masked = masking_img(img_roll, poin_wajah_kiri, 'putih')
            landmark_warp_kiri = warp_poin(poin_wajah_kiri, yaw_angle)

            poin_wajah_kanan = wajah_kanan(dua_D)
            face_kanan_masked = masking_img(img_roll, poin_wajah_kanan, 'putih')
            landmark_warp_kanan = warp_poin(poin_wajah_kanan, yaw_angle)

            if combine_face_hand == 'True':
                mask_hand_full = np.any(hand_masked != [0, 0, 0], axis=-1)
                hasil_full = wajah_full_masked
                hasil_full[mask_hand_full] = hand_masked[mask_hand_full]

                mask_hand_kanan = np.any(right != [0, 0, 0], axis=-1)
```

```

hasil_kanan = face_kanan_masked.copy()
hasil_kanan[mask_hand_kanan] = right[mask_hand_kanan]

mask_hand_kiri = np.any(left != [0, 0, 0], axis=-1)
hasil_kiri = face_kiri_masked.copy()
hasil_kiri[mask_hand_kiri] = left[mask_hand_kiri]
else:
    hasil_kanan = face_kanan_masked.copy()
    hasil_kiri = face_kiri_masked.copy()

poin_wajah_kiri = wajah_kiri(dua_D)
landmark_warp_kiri = warp_poin(poin_wajah_kiri, yaw_angle)
warpp_kiri = images_warping(img_rr, poin_wajah_kiri,
landmark_warp_kiri)

poin_wajah_kanan = wajah_kanan(dua_D)
landmark_warp_kanan = warp_poin(poin_wajah_kanan, yaw_angle)
warpp_kanan = images_warping(img_rr, poin_wajah_kanan,
landmark_warp_kanan)

luas_kiri = luas_wajah(hasil_kiri)
luas_kanan = luas_wajah(hasil_kanan)
if occlusion == "right":
    occ = 2
    out_ = combine_wajah(warpp_kanan, landmark_warp_kanan, occ)
elif occlusion == "left":
    occ = 1
    out_ = combine_wajah(warpp_kiri, landmark_warp_kiri, occ)
else:
    if luas_kiri > luas_kanan:
        occ = 1
        out_ = combine_wajah(warpp_kiri, landmark_warp_kiri, occ)
    else:
        occ = 2
        out_ = combine_wajah(warpp_kanan, landmark_warp_kanan, occ)

flip_output = out_
flip_output = cv2.resize(flip_output, (224, 224))
flip_output = cv2.cvtColor(flip_output, cv2.COLOR_BGR2RGB)
return flip_output
else:
    return images
except Exception as e:
    return images

```