

BAB I

PENDAHULUAN

1.1 LATAR BELAKANG

Perkembangan pembelajaran digital menuntut sistem yang dapat membantu pendidik memahami keterlibatan peserta didik secara lebih objektif. *Student engagement* dalam *smart education* perlu dipahami melalui definisi, dataset, dan metode estimasi yang beragam [1]. *Engagement* juga mencakup dimensi *behavioral*, *cognitive*, *emotional*, dan *social* yang saling berkaitan dalam proses pembelajaran [2]. Pada lingkungan daring, deteksi keterlibatan peserta didik mulai dikembangkan melalui *computer vision* dan penggabungan fitur multidimensi agar analisis tidak hanya bergantung pada pengamatan langsung [3]. *Behavioral engagement* juga memerlukan indikator yang dapat diamati secara sistematis, sedangkan dimensi *emotional*, *behavioral*, dan *cognitive* perlu dipahami sebagai bagian dari aktivitas belajar yang saling berhubungan [4], [5].

Ekspresi wajah dapat digunakan sebagai salah satu sinyal visual untuk memahami kondisi emosional peserta didik dalam pembelajaran daring. Survei *deep facial expression recognition* menunjukkan bahwa ekspresi wajah banyak dimanfaatkan dalam *human-computer interaction* dan *affective computing* [6]. Riset *Facial Expression Recognition* (FER) berbasis *deep learning* menjelaskan bahwa performa pengenalan emosi dipengaruhi oleh karakteristik dataset, arsitektur model, dan tahapan *preprocessing* [7], [8]. *Deep neural network* mampu mempelajari representasi fitur wajah secara otomatis, sedangkan *Convolutional Neural Network* (CNN) dapat berfokus pada area penting ekspresi seperti mata, mulut, dan alis [9], [10].

Aplikasi deteksi *engagement* berbasis ekspresi wajah membutuhkan model yang ringan agar dapat berjalan secara *realtime* pada perangkat dengan sumberdaya terbatas. Arsitektur CNN yang lebih dalam seperti ResNet dan model hasil *scaling* seperti EfficientNet memang mampu memberikan akurasi yang tinggi, namun keduanya memiliki kompleksitas komputasi yang besar dengan jumlah parameter

mencapai jutaan (misalnya, ResNet-50 memiliki sekitar 25 juta parameter) [11], [12]. Mini-Xception diperkenalkan sebagai *realtime* CNN untuk klasifikasi emosi dan *gender* dengan desain yang efisien, sedangkan *lightweight* CNN juga terbukti relevan untuk deteksi ekspresi wajah *realtime* [13], [14]. Mini-Xception menawarkan efisiensi dengan mereduksi jumlah parameter hingga 80 kali lipat lebih sedikit dibandingkan arsitektur CNN tradisional [13]. Meskipun ResNet unggul dalam ekstraksi fitur yang sangat mendalam, Mini-Xception hanya memerlukan sekitar 60.000 parameter, sehingga jauh lebih cepat untuk dijalankan pada perangkat *consumer-grade* tanpa mengorbankan akurasi yang signifikan. Selain itu, MediaPipe dan BlazeFace dapat digunakan untuk mendukung deteksi wajah berlatensi rendah sehingga proses *cropping* dan klasifikasi ekspresi dapat dilakukan pada area wajah yang tepat [15], [16].

Pelatihan model ekspresi wajah memerlukan dataset *benchmark* yang dapat merepresentasikan variasi emosi secara terukur. FER-2013 merupakan dataset ekspresi wajah *grayscale* berukuran 48×48 piksel yang digunakan dalam kompetisi *Challenges in Representation Learning* dan banyak dijadikan acuan penelitian FER [17]. Namun, dataset ekspresi wajah di lingkungan nyata masih menghadapi variasi pencahayaan, pose, dan ambiguitas ekspresi yang dapat memengaruhi performa model [18]. Ketidakseimbangan kelas juga dapat membuat model lebih sulit mempelajari kelas emosi dengan jumlah sampel yang lebih sedikit [19].

Agar keluaran sistem sesuai dengan kebutuhan pembelajaran, hasil klasifikasi emosi perlu dipetakan menjadi kategori *engagement*. Kajian *automatic engagement estimation* menjelaskan bahwa level *engagement* dapat dikelompokkan berdasarkan *basic emotions* atau model Russell, serta dapat disederhanakan menjadi dua kelompok, yaitu *engaged* dan *disengaged* [1]. Model afektif Altuwairqi *et al.* menghubungkan ekspresi, emosi, dan level *engagement* untuk mendeteksi keterlibatan peserta didik pada *e-learning Massive Open Online Course* (MOOC) [20]. Penelitian Gupta *et al.* menggunakan informasi FER untuk menghitung *engagement index* dan memprediksi keadaan *Engaged* dan *Disengaged*, serta

menunjukkan bahwa *happy*, *neutral*, dan *surprise* cenderung berkaitan dengan *engaged*, sedangkan *sad*, *angry*, dan *fear* berkaitan dengan kondisi kurang terlibat [21], [22].

Berdasarkan uraian tersebut, penelitian ini mengusulkan aplikasi deteksi *student engagement realtime* berbasis ekspresi wajah dengan keluaran *Engaged* dan *Disengaged*. Model Mini-Xception digunakan untuk mengenali tujuh ekspresi pada FER-2013, yaitu *angry*, *disgust*, *fear*, *happy*, *neutral*, *sad*, dan *surprise*, kemudian hasil klasifikasi emosi dipetakan menjadi kategori *engagement* [13], [17]. Pada penelitian ini, *happy*, *surprise*, dan *neutral* dipetakan sebagai *Engaged*, sedangkan *angry*, *disgust*, *fear*, dan *sad* dipetakan sebagai *Disengage* berdasarkan literatur *engagement* berbasis emosi dan aturan operasional penelitian [1], [20], [22]. Sistem ini tetap diposisikan sebagai alat bantu atau sinyal pendukung, bukan keputusan tunggal, karena *engagement* peserta didik juga dipengaruhi oleh konteks pembelajaran, perilaku, kognisi, dan interaksi selama pembelajaran daring atau *hybrid* [1], [2].

1.2 PERUMUSAN MASALAH

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana merancang dan membangun aplikasi deteksi *student engagement realtime* berbasis ekspresi wajah menggunakan *input webcam*?
2. Bagaimana menerapkan aturan pemetaan ekspresi wajah menjadi dua kategori *engagement*?
3. Bagaimana mengevaluasi performa sistem berdasarkan metrik klasifikasi, hasil pemetaan *engagement*, stabilitas inferensi *realtime*, serta penggunaan sumberdaya komputasi?

1.3 TUJUAN

Dengan rumusan masalah yang disampaikan, penelitian ini bertujuan:

1. Merancang dan membangun aplikasi deteksi *student engagement realtime* berbasis ekspresi wajah menggunakan *webcam*
2. Menerapkan aturan pemetaan ekspresi wajah menjadi dua kategori *engagement*, yaitu *engaged* dan *disengaged*
3. Mengevaluasi performa sistem berdasarkan metrik klasifikasi, hasil pemetaan *engagement*, stabilitas inferensi *realtime*, serta penggunaan sumberdaya komputasi.

1.4 BATASAN MASALAH

Batasan-batasan perlu ditetapkan agar penelitian dapat lebih terarah. Batasan masalah yang ditetapkan pada penelitian ini antara lain:

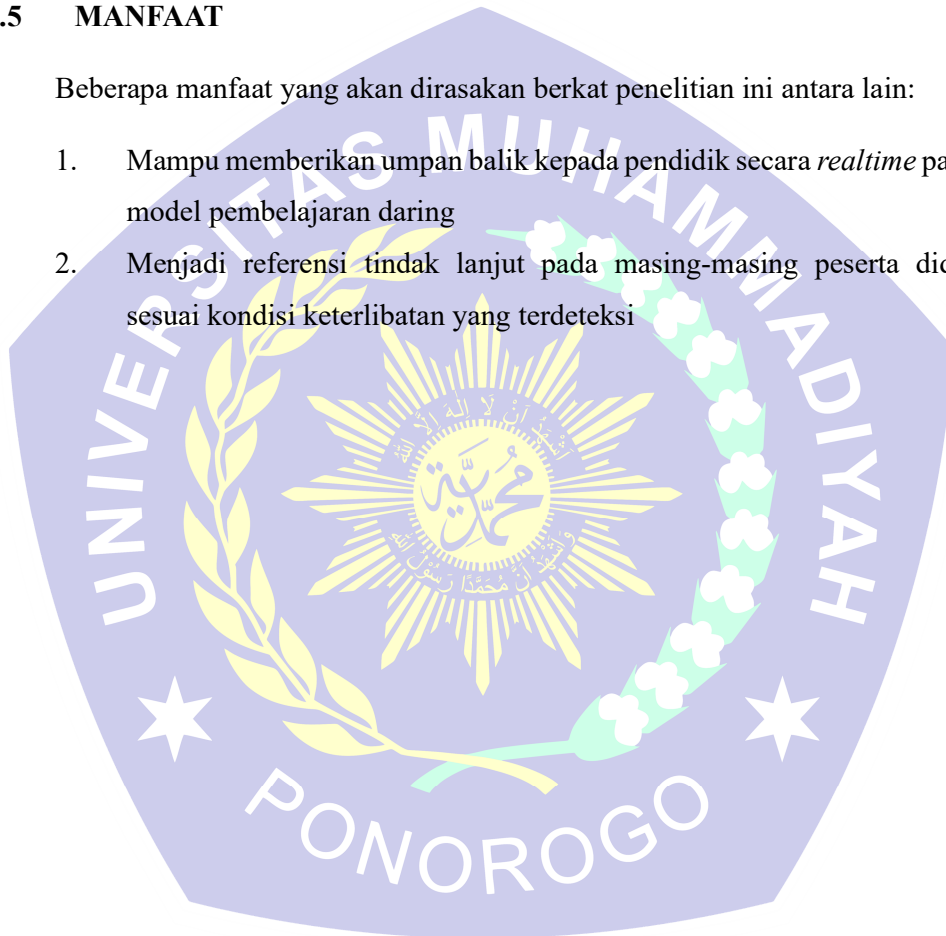
1. Penelitian berfokus pada deteksi *student engagement* berbasis ekspresi wajah atau emosi, bukan berdasarkan suara, *gaze direction*, interaksi LMS, gestur tubuh, atau aktivitas digital lainnya yang mendeteksi atensi kognitif
2. Dataset yang digunakan untuk pelatihan model adalah FER-2013 dengan citra wajah *grayscale* berukuran 48×48 piksel
3. Model klasifikasi ekspresi wajah yang digunakan adalah CNN dengan arsitektur Mini-Xception
4. Deteksi wajah pada aplikasi *realtime* dilakukan menggunakan MediaPipe *Face Detection*
5. Input *realtime* aplikasi berasal dari *webcam* dengan berfokus pada *feed* langsung (*oncam*) bukan foto yang dipasang di depan kamera
6. Kelas ekspresi wajah yang digunakan mengikuti tujuh label FER-2013, yaitu *angry*, *disgust*, *fear*, *happy*, *neutral*, *sad*, dan *surprise*
7. Pemetaan *engagement* pada penelitian ini dibatasi menjadi dua kategori, yaitu a) *engaged*: *happy*, *surprise*, dan *neutral*; dan b) *disengaged*: *angry*, *disgust*, *fear*, dan *sad*
8. Kategori *neutral* diposisikan sebagai bagian dari *engaged* berdasarkan aturan operasional penelitian, tetapi tetap dipahami sebagai kondisi *engagement* minimal atau normal, bukan bukti keterlibatan yang kuat

9. Hasil deteksi *engagement* diposisikan sebagai sinyal pendukung bagi pendidik, bukan sebagai keputusan mutlak terhadap tingkat keterlibatan peserta didik
10. Evaluasi sistem dibatasi pada performa klasifikasi, hasil pemetaan *engagement*, stabilitas deteksi wajah *realtime*, serta penggunaan sumber daya komputasi

1.5 MANFAAT

Beberapa manfaat yang akan dirasakan berkat penelitian ini antara lain:

1. Mampu memberikan umpan balik kepada pendidik secara *realtime* pada model pembelajaran daring
2. Menjadi referensi tindak lanjut pada masing-masing peserta didik sesuai kondisi keterlibatan yang terdeteksi

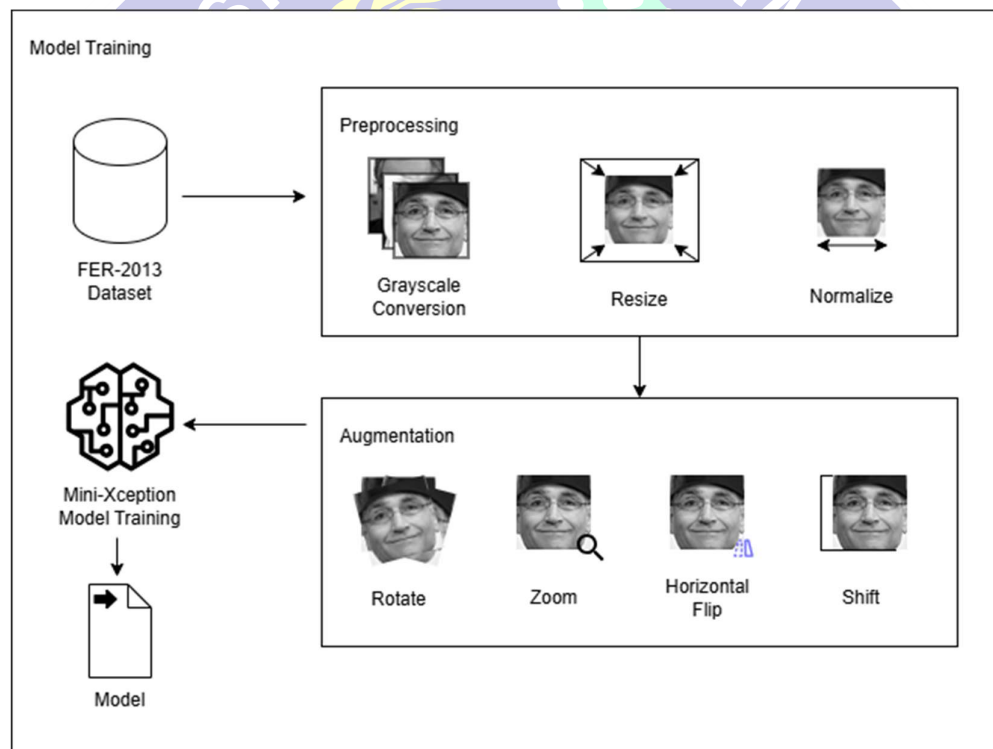


BAB II

PERANCANGAN SISTEM & ARSITEKTUR REKAYASA

2.1 DESAIN SISTEM

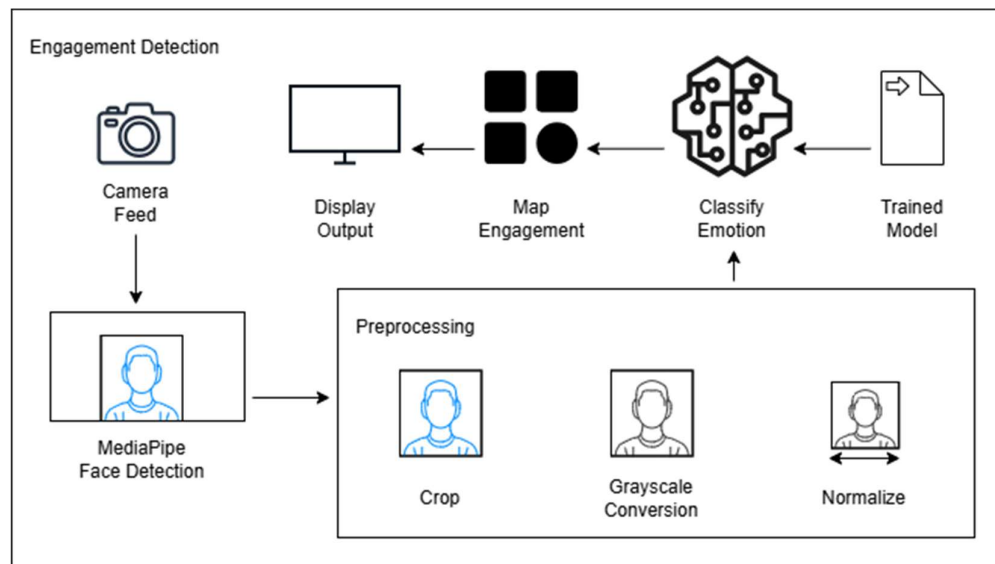
Sistem yang diajukan memiliki dua tahapan yaitu tahap pelatihan model dan tahap deteksi *engagement*. Tahap pelatihan model bertujuan membangun bobot model secara mandiri sebelum diintegrasikan ke aplikasi. Proses dimulai dari akuisisi dataset FER-2013, diikuti dengan augmentasi data (*rotate, zoom, shift, horizontal flip*) untuk mengatasi masalah *data imbalance*. Citra kemudian melalui proses *preprocessing* (*grayscale, resize, normalize*) sebelum melatih arsitektur Mini-Xception.



Gambar 2.1 Tahap Pelatihan Model

Pada tahap deteksi *engagement*, aplikasi memproses *input video* dari *webcam* secara *realtime*. Deteksi wajah dilakukan oleh MediaPipe dengan algoritma BlazeFace untuk efisiensi sumberdaya perangkat *consumer-grade*. Hasil deteksi

dipotong (*crop*), diubah ke *grayscale*, dan dinormalisasi agar identic dengan format pelatihan. Model kemudian melakukan klasifikasi emosi yang hasilnya dipetakan menjadi kategori *Engaged* (*happy, surprise, neutral*) atau *Disengaged* (*angry, disgust, fear, sad*).



Gambar 2.2 Tahap Deteksi *Engagement*

2.2 FER-2013

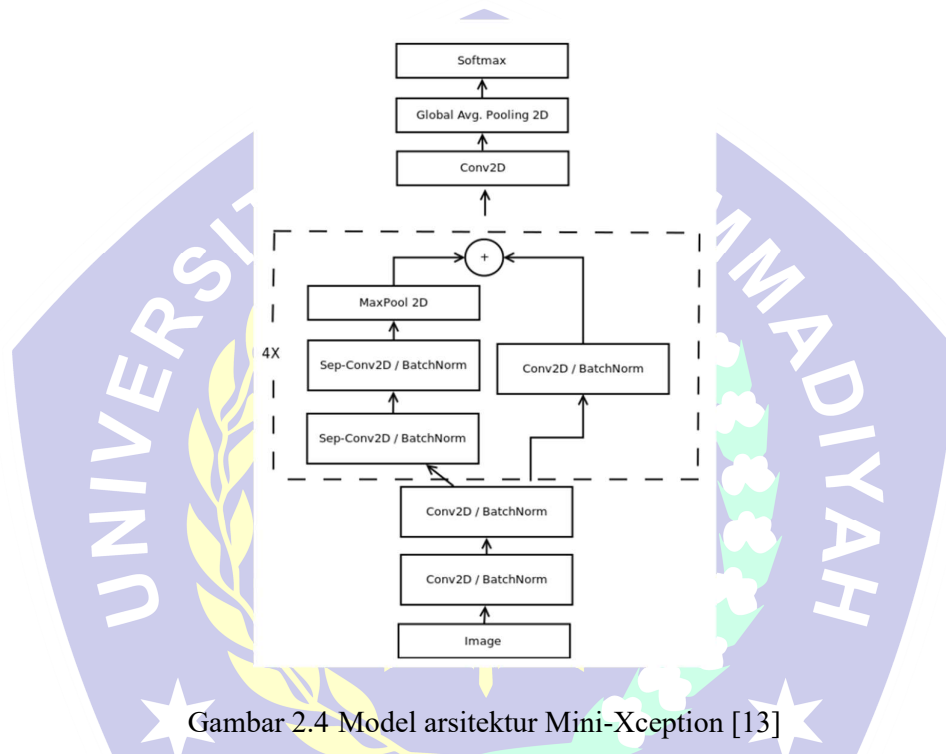


Gambar 2.3 Sampel dataset FER-2013

Dataset yang digunakan dalam penelitian ini adalah FER-2013, yaitu dataset ekspresi wajah *grayscale* berukuran 48×48 piksel yang digunakan dalam kompetisi *Challenges in Representation Learning* [17]. Dataset ini terdiri atas tujuh kelas emosi, yaitu *angry, disgust, fear, happy, neutral, sad, dan surprise*. FER-2013 dipilih karena banyak digunakan sebagai *benchmark* dalam penelitian pengenalan

ekspresi wajah, sehingga hasil pengujian model dapat dibandingkan dengan penelitian sebelumnya. Meskipun demikian, dataset ini memiliki tantangan seperti resolusi citra yang rendah dan ketidakseimbangan jumlah sampel pada beberapa kelas emosi, yang dapat memengaruhi kemampuan model dalam mengenali kelas tertentu [18], [19].

2.3 MINI-XCEPTION



Gambar 2.4 Model arsitektur Mini-Xception [13]

Mini-Xception merupakan arsitektur CNN ringan yang dirancang untuk klasifikasi emosi wajah secara *realtime*. Arsitektur ini diperkenalkan oleh Arriaga *et al.* untuk tugas klasifikasi emosi dan *gender* dengan tujuan menghasilkan model yang efisien tetapi tetap mampu mengenali pola ekspresi wajah secara baik [13]. Arsitektur Mini-Xception secara inovatif menghapus penggunaan *fully connected layers*, dan menggantinya dengan *Global Average Pooling*. Teknik ini secara signifikan mengurangi parameter model hingga 80 kali lipat dibandingkan CNN tradisional, namun tetap mempertahankan performa klasifikasi yang setara dengan kemampuan manusia dalam mengenali emosi. Mini-Xception relevan digunakan

pada system berbasis *webcam* karena aplikasi *realtime* membutuhkan model dengan waktu inferensi cepat dan kebutuhan komputasi yang tidak terlalu besar.

2.4 MEDIAPIPE

MediaPipe adalah kerangka kerja untuk membangun solusi pembelajaran mesin *cross-platform* [23]. Aplikasi ini mengintegrasikan *framework* MediaPipe yang di dalamnya menggunakan algoritma BlazeFace untuk deteksi wajah. BlazeFace adalah kerangka kerja deteksi wajah yang dioptimalkan untuk penggunaan sumberdaya perangkat *mobile* dan mengadaptasi kerangka kerja dari *the Single Shot Multibox Detector* (SSD) [16]. BlazeFace dipilih karena kemampuannya dalam mendeteksi area wajah dengan kecepatan sangat tinggi (submilidetik), yang memungkinkan proses *cropping* wajah dilakukan secara instan sebelum masuk ke tahap klasifikasi emosi.

2.5 PEMETAAN ENGAGEMENT

Pada penelitian ini, tujuh ekspresi dari FER-2013 dipetakan menjadi dua kategori *engagement*. Ekspresi *happy*, *surprise*, dan *neutral* dikelompokkan sebagai *Engaged*, sedangkan *angry*, *disgust*, *fear*, dan *sad* dikelompokkan sebagai *Disengaged*. Pemetaan emosi *neutral* ke dalam kategori *Engaged* didasarkan pada asumsi operasional bahwa siswa berada dalam kondisi focus normal saat menyimak materi. Hal ini sejalan dengan penelitian Gupta *et al.* yang mengelompokkan emosi netral sebagai indikator keterlibatan, meskipun secara teoritis tetap dipahami sebagai tingkat keterlibatan yang paling dasar. Emosi yang dipetakan pada Tabel 2.1 digunakan sebagai aturan operasional penelitian, sehingga hasil deteksi tetap dipahami sebagai sinyal pendukung, bukan Keputusan mutlak terhadap keterlibatan peserta didik.

Tabel 2.1 Pemetaan ekspresi wajah ke kategori *engagement*

Ekspresi	Kategori	Keterangan
FER-2013	<i>Engagement</i>	
<i>Happy</i>	<i>Engaged</i>	Menunjukkan respons positif atau ketertarikan
<i>Surprise</i>	<i>Engaged</i>	Dapat menunjukkan perhatian atau respons terhadap materi
<i>Neutral</i>	<i>Engaged</i>	Diposisikan sebagai kondisi normal atau minimal <i>engagement</i>
<i>Angry</i>	<i>Disengaged</i>	Dapat menunjukkan ketidaknyamanan atau hambatan belajar
<i>Disgust</i>	<i>Disengaged</i>	Emosi negatif: digunakan sebagai aturan operasional
<i>Fear</i>	<i>Disengaged</i>	Dapat menunjukkan kebingungan, tekanan, atau hambatan belajar
<i>Sad</i>	<i>Disengaged</i>	Dapat menunjukkan rendahnya keterlibatan emosional

Sumber tabel: diadaptasi dari Karimah dan Hasegawa, Altuwairqi *et al.*, serta Gupta *et al*

BAB III

IMPLEMENTASI, PENGUJIAN, DAN VALIDASI DATA

3.1 IMPLEMENTASI

Pada tahap implementasi, sistem deteksi *student engagement realtime* dibangun menggunakan kombinasi MediaPipe Face Detection dan model Mini-Xception untuk melakukan klasifikasi ekspresi wajah. Model dikembangkan menggunakan bahasa pemrograman Python dengan dukungan beberapa pustaka seperti TensorFlow, Numpy, Pandas, dan Scikit-learn. Implementasi sistem dikembangkan menggunakan bahasa pemrograman JavaScript dengan pustaka TensorFlow dan MediaPipe.

Pelatihan model dan pengujian model deteksi *student engagement* dilakukan menggunakan perangkat keras dan perangkat lunak dengan spesifikasi berikut:

Tabel 3.1 Spesifikasi Komputer Pelatihan

Komponen	Spesifikasi
Prosesor	Intel Core i7-8700T 6Cores @2.40GHz
RAM	16GB
Sistem Operasi	Windows 11
Bahasa Pemrograman	Python
Framework Deep Learning	TensorFlow 2./ Keras 2

Sedangkan implementasi deteksi *student engagement* secara realtime menggunakan *webcam* menggunakan perangkat keras dan perangkat lunak dengan spesifikasi berikut:

Tabel 3.2 Spesifikasi Komputer Implementasi

Komponen	Spesifikasi
Prosesor	Intel Core Ultra 7 155H 16Cores
RAM	16GB
Sistem Operasi	Windows 11
Bahasa Pemrograman	JavaScript
Framework Deep Learning	TensorFlow
Library Face Detection	MediaPipe

3.1.1 Implementasi Model Mini-Xception

Model klasifikasi emosi yang digunakan pada penelitian ini adalah Mini-Xception. Model dilatih menggunakan dataset FER-2013 yang terdiri dari tujuh kategori emosi, yaitu Angry, Disgust, Fear, Happy, Sad, Surprise, dan Neutral.

Sebelum proses pelatihan dilakukan, citra wajah terlebih dahulu melalui tahap preprocessing yang meliputi konversi ke grayscale, normalisasi piksel, dan perubahan ukuran citra menjadi 48x48 piksel sesuai kebutuhan input model.

Parameter pelatihan yang digunakan meliputi:

Tabel 3.3 Parameter Pelatihan

Komponen	Spesifikasi
Ukuran Input	48x48x1
Jumlah Kelas	7
Optimizer	Adam
Loss Function	Categorical Crossentropy
Batch Size	32
Epoch	52

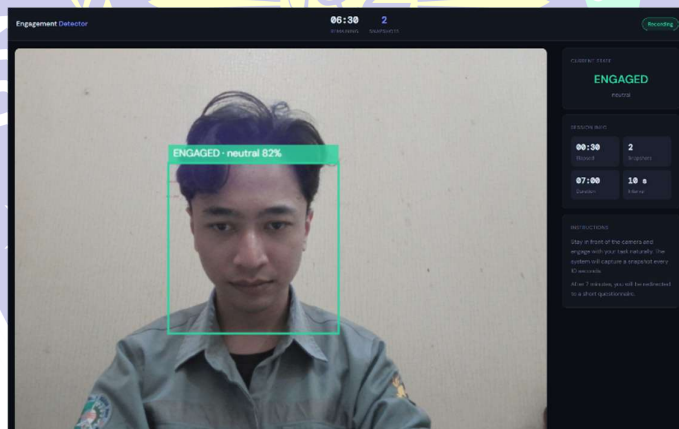
Pengaturan augmentasi data yang digunakan antara lain:

Tabel 3.4 Pengaturan Augmentasi Data

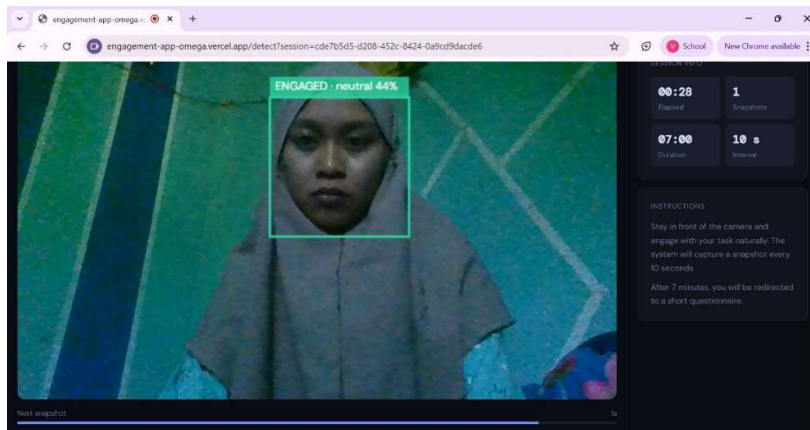
Komponen	Spesifikasi
Rescale	1./255
Rotation Range	15 degrees
Width Shift Range	0.12
Height Shift Range	0.12
Zoom Range	0.15
Shear Range	0.05
Brightness Range	[0.85, 1.15]
Horizontal Flip	True

3.1.2 Implementasi pada *Platform Website*

Berikut merupakan implementasi menggunakan *platform website*:

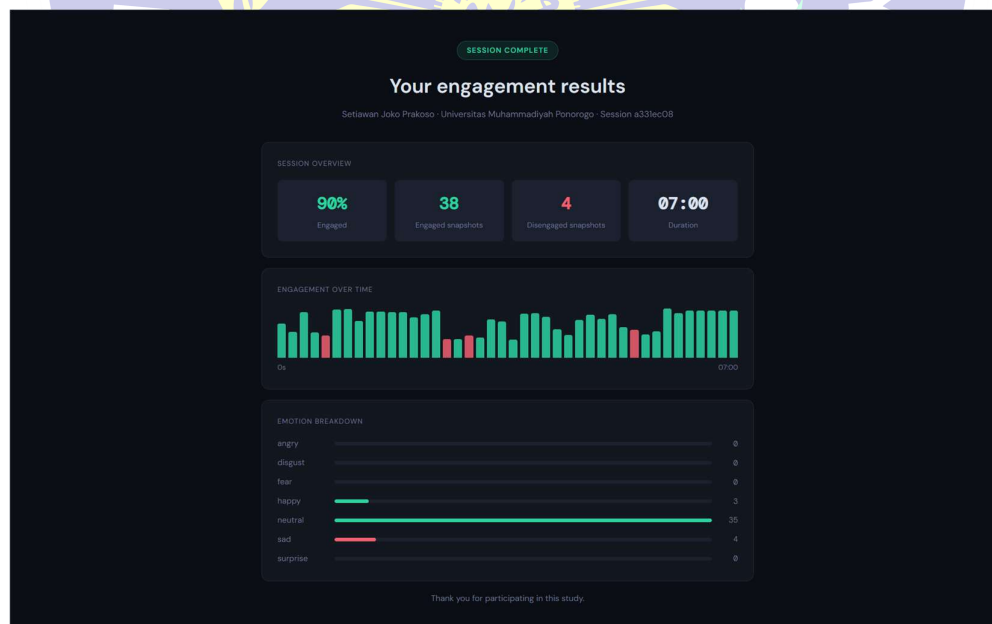


Gambar 3.1 Implementasi Deteksi *Engagement* (1)



Gambar 3.2 Implementasi Deteksi *Engagement* (2)

Implementasi deteksi *engagement* menggunakan *platform website* didesain untuk melakukan tangkapan data setiap 10 detik. Hal ini dapat membantu pendidik dalam mengawasi peserta didik dalam jangka waktu tertentu. Penyalahgunaan *feed video* dengan menggunakan foto statis di depan kamera juga dapat diamati melalui analisis yang dilakukan oleh pendidik.



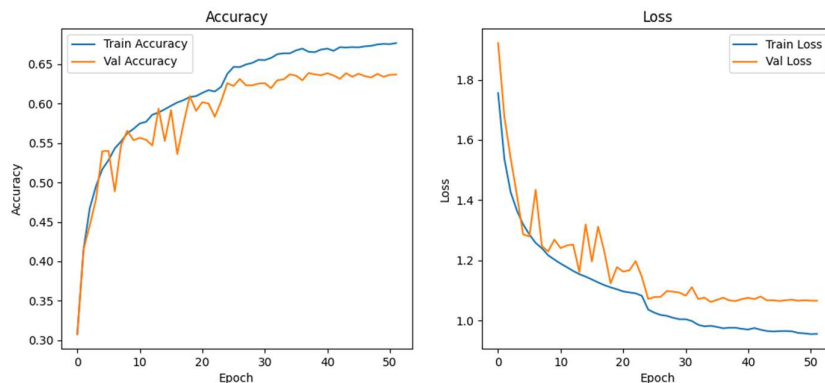
Gambar 3.3 Halaman Analisis Emosi dan *Engagement*

3.2 PENGUJIAN

3.2.1 Grafik Accuracy dan Loss

Berdasarkan perbandingan kurva *training* dan *validation*, tidak ditemukan indikasi *overfitting* yang serius. Hal ini dapat dilihat dari beberapa aspek berikut:

1. *Training accuracy* dan *validation accuracy* sama-sama mengalami peningkatan hingga akhir pelatihan;
2. *Validation loss* tidak menunjukkan kenaikan yang signifikan ketika training loss terus menurun;
3. Selisih antara *training accuracy* dan *validation accuracy* relatif kecil yaitu sekitar 3-4%; dan
4. *Validation accuracy* tetap stabil pada kisaran 63-64% setelah model mencapai konvergensi.

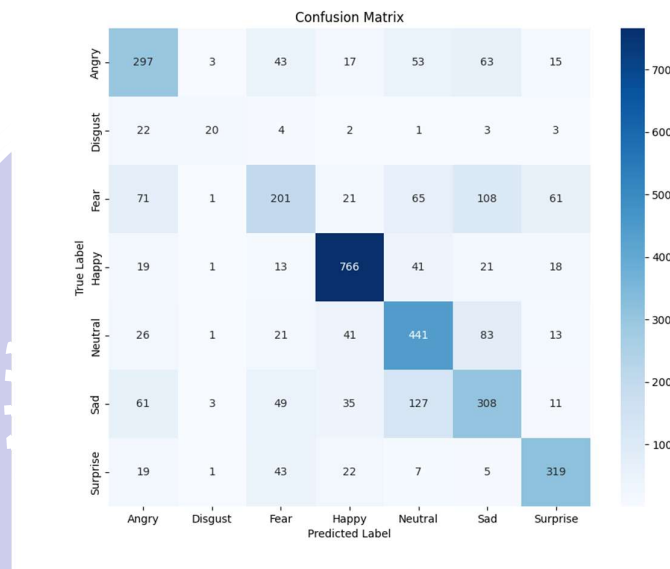


Gambar 3.4 Grafik Accuracy dan Loss

Meskipun demikian, terdapat indikasi *overfitting* ringan pada *epoch-epoch* akhir karena *training accuracy* masih meningkat hingga 67,69%, sedangkan *validation accuracy* cenderung stagnan di sekitar 63-64%. Namun, tingkat *overfitting* tersebut masih dapat diterima dan tidak menyebabkan penurunan performa validasi secara signifikan.

3.2.2 Confusion Matrix dan Classification Report

Berdasarkan *Confusion Matrix*, kesalahan klasifikasi paling banyak terjadi pada pasangan kelas *Fear-Sad*, *Sad-Neutral*, dan *Disgust-Angry*. Kondisi ini menunjukkan bahwa model masih mengalami kesulitan dalam membedakan ekspresi yang memiliki kemiripan pola wajah. Faktor lain yang turut memengaruhi adalah ketidakseimbangan jumlah data pada dataset FER-2013, terutama pada kelas *Disgust* yang memiliki jumlah sampel lebih sedikit dibandingkan kelas lainnya.



Gambar 3.5 Confusion Matrix

Hasil *Classification Report* menunjukkan bahwa model memperoleh *accuracy* sebesar 66% dengan *weighted F1-score* sebesar 0,65, yang mengindikasikan kemampuan klasifikasi yang cukup baik pada data pengujian. Nilai *precision*, *recall*, dan *F1-score* bervariasi pada setiap kelas, menunjukkan bahwa tingkat kesulitan klasifikasi tidak sama untuk seluruh emosi. Secara umum, model memberikan performa yang lebih baik pada kelas yang memiliki karakteristik visual yang kuat dan jumlah data yang relatif banyak.

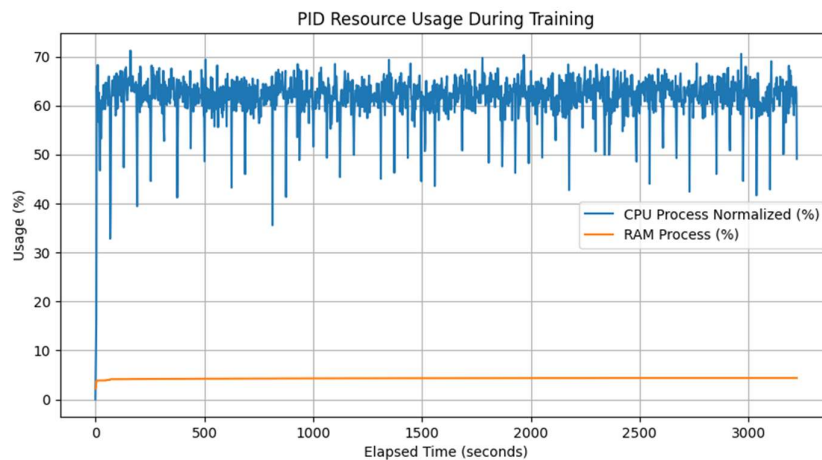
Tabel 3.5 Classification Report

	<i>precision</i>	<i>recall</i>	<i>f1-score</i>	<i>support</i>
Angry	0.58	0.60	0.59	491
Disgust	0.67	0.36	0.47	55
Fear	0.54	0.38	0.45	528
Happy	0.85	0.87	0.86	879
Neutral	0.60	0.70	0.65	626
Sad	0.52	0.52	0.52	594
Surpsise	0.72	0.77	0.75	416
accuracy			0.66	3589
macro avg	0.64	0.60	0.61	3589
weighted avg	0.65	0.66	0.65	3589

Berdasarkan Tabel 3.5, kelas *Happy* memiliki performa terbaik dengan *F1-score* sebesar 0,86, diikuti oleh kelas *Surprise* dengan *F1-score* sebesar 0,75. Sebaliknya, kelas *Fear* dan *Disgust* memperoleh *F1-score* terendah, masing-masing sebesar 0,45 dan 0,47, yang menunjukkan bahwa kedua kelas tersebut masih sulit dikenali secara konsisten oleh model. Perbedaan performa antar kelas ini sejalan dengan hasil yang terlihat pada *Confusion Matrix*, khususnya pada kelas-kelas yang sering mengalami kesalahan klasifikasi.

3.2.3 Resource Usage

Hasil pemantauan *resource usage* menunjukkan bahwa proses pelatihan model berjalan dengan menggunakan sumberdaya yang relatif stabil sepanjang proses *training*. Penggunaan CPU mendominasi proses komputasi dengan nilai yang sebagian besar berada pada kisaran 60-65%, sedangkan penggunaan RAM cenderung konstan pada kisaran 4-5%. Kondisi ini menunjukkan bahwa pelatihan dapat dijalankan tanpa mengalami lonjakan penggunaan sumberdaya yang signifikan.



Gambar 3.6 *Resource Usage* selama Pelatihan

Berdasarkan Gambar 3.6, penggunaan CPU meningkat secara cepat pada awal proses *training* dan kemudian mempertahankan pola yang stabil hingga akhir pelatihan. Meskipun terdapat beberapa penurunan sesaat (*spike downward*), CPU kembali ke tingkat penggunaan normal dalam waktu singkat sehingga tidak mengganggu proses pelatihan secara keseluruhan. Sementara itu, penggunaan RAM menunjukkan pola yang sangat stabil dengan variasi yang relatif kecil, yang mengindikasikan bahwa kebutuhan memori selama *training* dapat terkelola dengan baik dan tidak mengalami indikasi *memory leak* maupun konsumsi memori berlebihan.

3.3 VALIDASI

Pada pelatihan yang dilakukan dengan parameter sesuai pada Tabel 3.3 dan Tabel 3.4, diperoleh *epoch* terbaik dengan hasil *test accuracy* 0.6553, *test loss* 1.0212, dan *validation accuracy* 0.6389 pada *epoch* ke 39.

Untuk memahami perilaku model secara lebih mendalam, dilakukan analisis terhadap beberapa sampel citra yang berasal dari kelas *Angry*. Gambar 3.7 menampilkan label sebenarnya (*True*) dan hasil prediksi model (*Pred*). Dari sembilan sampel yang diamati, enam sampel berhasil diklasifikasikan dengan benar

sebagai *Angry*, sedangkan tiga sampel mengalami kesalahan klasifikasi. Dengan demikian, tingkat keberhasilan sampel yang ditampilkan mencapai sekitar 66,7%.



BAB IV

KESIMPULAN DAN SARAN

4.1 KESIMPULAN

Penelitian ini telah berhasil merancang dan membangun sebuah aplikasi deteksi keterlibatan siswa (*student engagement*) secara *realtime* dengan mengintegrasikan arsitektur model Mini-Xception dan MediaPipe *face detection*. Sistem ini menggunakan dataset FER-2013 untuk melatih kemampuan klasifikasi tujuh ekspresi wajah dasar yang kemudian hasilnya dipetakan ke dalam kategori biner, yaitu terlibat (*engaged*) dan tidak terlibat (*disengaged*). Pemilihan Mini-Xception terbukti efektif karena mereduksi jumlah parameter hingga 80 kali lipat dibandingkan arsitektur CNN konvensional melalui penerapan teknik *Global Average Pooling* dan *depthwise separable convolutions*, sehingga memungkinkan aplikasi berjalan lancar pada perangkat *consumer-grade*. Dalam aturan operasionalnya, ekspresi *Happy*, *Surprise* dan *Neutral* dikelompokkan ke dalam kategori terlibat, di mana emosi netral dipahami sebagai kondisi fokus minimal yang lazim ditemukan pada pembelajaran materi yang serius. Sementara itu, emosi seperti *Angry*, *Disgust*, *Fear*, dan *Sad* diposisikan sebagai indikator awal ketidakterlibatan atau adanya hambatan belajar kognitif yang memerlukan intervensi pendidik.

Model klasifikasi emosi ini mencapai tingkat akurasi yang kompetitif pada kisaran 62% hingga 67%, yang merupakan hasil optimal untuk kategori sistem ringan berbasis webcam. Meskipun sistem saat ini masih terbatas pada pemrosesan citra wajah yang bersifat statis per bingkai (*frame-by-frame*) dan belum menangkap dinamika temporal secara utuh, penelitian ini memberikan kontribusi signifikan dalam menyediakan alat bantu objektif bagi pendidik untuk memantau respons emosional siswa di lingkungan daring maupun hibrida. Melalui teknologi visi komputer ini, diharapkan efektivitas pembelajaran digital dapat meningkat dengan adanya sinyal visual pendukung yang membantu pendidik membaca kondisi internal siswa secara lebih tepat dan terukur.

4.2 SARAN

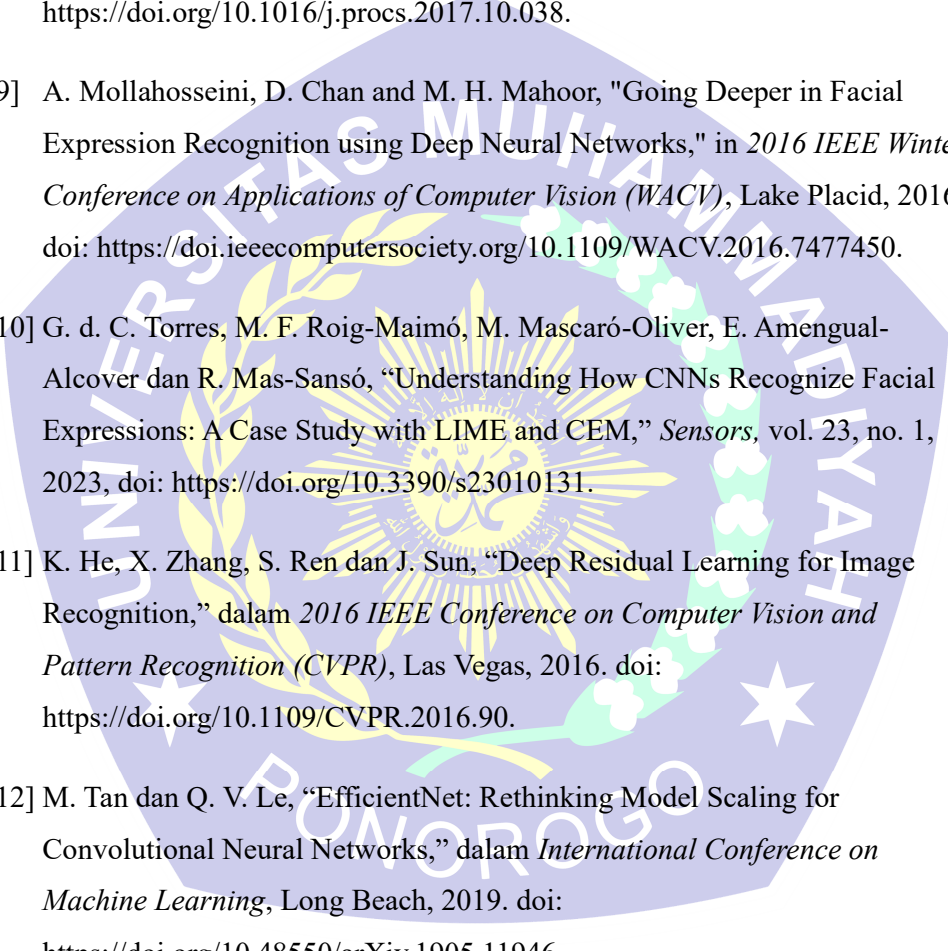
Penelitian ini memberikan gambaran bagaimana pengukuran keterlibatan peserta didik dilakukan pada perangkat *consumer-grade*. Akurasi deteksi ekspresi dapat ditingkatkan dengan menambah dataset pelatihan yang serupa. FER-2013 juga dikenal memiliki ketidakseimbangan data yang cukup signifikan, penambahan dataset akan mampu membantu menyelesaikan permasalahan tersebut. Peneliti selanjutnya disarankan untuk mengintegrasikan dataset *benchmark* lain yang lebih besar dan bervariasi, seperti AffectNet yang menyediakan lebih dari 1 juta citra atau RAF-DB, untuk meningkatkan sensitivitas model terhadap emosi minoritas dan mengurangi bias terhadap kelas mayoritas.

Selain itu, karena penelitian ini membatasi input analisis pada ekspresi emosi wajah saja, pengembangan ke arah sistem multimodal sangat direkomendasikan untuk mendapatkan penilaian keterlibatan yang lebih akurat. Melalui *video stream* yang sama, peneliti dapat mengekstraksi modalitas lain seperti arah pandangan (*gaze direction*), frekuensi kedipan mata, pergerakan alis, hingga posisi kepala (*head pose*). Penggunaan kerangka kerja seperti MediaPipe *Face Mesh* dapat dioptimalkan untuk mengekstraksi hingga 470 titik koordinat wajah guna memantau perilaku spesifik pada area mata dan mulut yang mencerminkan atensi kognitif peserta didik.

Integrasi dengan sumber data digital eksternal juga menjadi peluang riset yang signifikan. Penggabungan fitur visual wajah dengan rekaman aktivitas pada sistem MOOC atau *Learning Management System* (LMS) seperti log interaksi, partisipasi forum, dan durasi pengerjaan tugas akan memberikan pemahaman keterlibatan yang lebih holistik dari dimensi perilaku dan kolaborasi. Terakhir, sesuai dengan keterbatasan sistem saat ini yang memproses citra secara statis, disarankan bagi peneliti mendatang untuk menerapkan model temporal (video kontinu) menggunakan arsitektur RNN atau LSTM. Pendekatan ini memungkinkan sistem untuk menangkap dinamika perubahan emosi secara sekuensial dari waktu ke waktu, memberikan gambaran *engagement* yang lebih valid dibandingkan analisis per bingkai (*frame-by-frame*).

DAFTAR PUSTAKA

- [1] S. N. Karimah dan S. Hasegawa, "Automatic engagement estimation in smart education/learning settings: a systematic review of engagement definitions, datasets, and methods," *Smart Learning Environments*, vol. 9, no. 31, 2022. <https://doi.org/10.1186/s40561-022-00212-y>.
- [2] D. R. Joshi, K. P. Adhikari, B. Khanal, J. Khaddka and S. Belbase, "Behavioral, cognitive, emotional and social engagement in mathematics learning during COVID-19 pandemic," *PLoS One*, vol. 17, no. 11, Nov 2022. doi: <https://doi.org/10.1371/journal.pone.0278052>.
- [3] N. Xie, Z. Liu, Z. Li, W. Pang and B. Lu, "Student engagement detection in online environment using computer vision and multi-dimensional feature fusion," *Multimedia Systems*, vol. 29, pp. 3559-3577, 2023. doi: <https://doi.org/10.1007/s00530-023-01153-3>.
- [4] K. King, "Interventions to Enhance Behavioral Engagement," *Student Engagement*, pp. 133-156, Mar 2020. doi: https://doi.org/10.1007/978-3-030-37285-9_8.
- [5] M. R. Jimenez-Liso, A. Bellocchi, M. Martinez-Chico and R. Lopez-Gay, "A Model-Based Inquiry Sequence as a Heuristic to Evaluate Students' Emotional, Behavioural, and Cognitive Engagement," *Research in Science Education*, vol. 52, pp. 1313-1334, 2022. doi: <https://doi.org/10.1007/s11165-021-10010-0>.
- [6] S. Li dan W. Deng, "Deep Facial Expression Recognition: A Survey," *IEEE Transactions on Affective Computing*, vol. 13, no. 3, pp. 1195-1215, 2022, doi: <https://doi.org/10.1109/TAFFC.2020.2981446>.

- 
- [7] W. Mellouk dan W. Handouzi, "Facial emotion recognition using deep learning: review and insights," *Procedia Computer Science*, vol. 175, pp. 689-694, 2020, doi: <https://doi.org/10.1016/j.procs.2020.07.101>.
- [8] D. A. Pitaloka, A. Wulandari, T. Basarudin dan D. Y. Liliana, "Enhancing CNN with Preprocessing Stage in Automatic Emotion Recognition," *Procedia Computer Science*, vol. 116, pp. 523-529, 2017, doi: <https://doi.org/10.1016/j.procs.2017.10.038>.
- [9] A. Mollahosseini, D. Chan and M. H. Mahoor, "Going Deeper in Facial Expression Recognition using Deep Neural Networks," in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, Lake Placid, 2016, doi: <https://doi.ieeecomputersociety.org/10.1109/WACV.2016.7477450>.
- [10] G. d. C. Torres, M. F. Roig-Maimó, M. Mascaró-Oliver, E. Amengual-Alcover dan R. Mas-Sansó, "Understanding How CNNs Recognize Facial Expressions: A Case Study with LIME and CEM," *Sensors*, vol. 23, no. 1, 2023, doi: <https://doi.org/10.3390/s23010131>.
- [11] K. He, X. Zhang, S. Ren dan J. Sun, "Deep Residual Learning for Image Recognition," dalam *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, 2016. doi: <https://doi.org/10.1109/CVPR.2016.90>.
- [12] M. Tan dan Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," dalam *International Conference on Machine Learning*, Long Beach, 2019. doi: <https://doi.org/10.48550/arXiv.1905.11946>.
- [13] O. Arriaga, M. Valdenegro-Toro dan P. Plöger, "Real-time Convolutional Neural Network for emotion and gender classification," dalam *2019 European Symposium on Artificial Neural Networks*, Bruges, 2019. doi: <https://doi.org/10.48550/arXiv.1710.07557>.

- [14] N. Zhou, R. Liang dan W. Shi, "A Lightweight Convolutional Neural Network," *IEEE Access*, vol. 9, pp. 5573-5584, 2020, doi: 10.1109/ACCESS.2020.3046715.
- [15] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C.-L. Chang, M. G. Yong, J. Lee, W.-T. Chang, W. Hua, M. Georg dan M. Grundmann, "MediaPipe: A Framework for Building Perception Pipelines," *arXiv*, vol. abs/1906.08172, 2019. doi: <https://doi.org/10.48550/arXiv.1906.08172>.
- [16] V. Bazarevsky, Y. Kartynnik, A. Yakunov, K. Raveendran dan M. Grundmann, "BlazeFace: Sub-millisecond Neural Face Detection on Mobile GPUs," *ArXiv*, 2019. doi: <https://doi.org/10.48550/arXiv.1907.05047>.
- [17] G. Ian J., et. al., "Challenges in Representation Learning: A report on three machine learning contests," dalam *International Conference on Machine Learning Processing*, Daegu, 2013. doi: https://doi.org/10.1007/978-3-642-42051-1_16.
- [18] A. Mollahosseini, B. Hasani dan M. H. Mahoor, "AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild," *IEEE Transaction on Affective Computing*, vol. 10, no. 1, pp. 18-31, 2019, doi: <https://doi.org/10.1109/TAFFC.2017.2740923>.
- [19] Q. T. Ngo dan S. Yoon, "Facial Expression Recognition Based on Weighted-Cluster Loss and Deep Transfer Learning Using a Highly Imbalanced Dataset," *Sensors*, vol. 20, no. 9, 2020. doi: <https://doi.org/10.3390/s20092639>.
- [20] K. Altuwairqi, S. K. Jarraya, A. Allinjawy dan M. Hammami, "A new emotion-based affective model to detect student's engagement," *Journal of King Saud University - Computer and Information Sciences*, vol. 33, no. 1, pp. 99-109, 2021. doi: <https://doi.org/10.1016/j.jksuci.2018.12.008>.

- [21] S. Gupta, P. Kumar dan R. K. Tekchandani, "Facial emotion recognition based real-time learner engagement detection system in online learning context using deep learning models," *Multimedia Tools and Applications*, vol. 82, pp. 11365-11394, 2023. doi: <https://doi.org/10.1007/s11042-022-13558-9>.
- [22] S. Gupta, P. Kumar dan R. Tekchandani, "A multimodal facial cues based engagement detection system in e-learning context using deep learning approach," *Multimedia Tools and Applications*, vol. 82, pp. 28589-28615, 2023. doi: <https://doi.org/10.1007/s11042-023-14392-3>.
- [23] F. Zhang, V. Bazarevsky, A. Vakunov, A. Tkachenka, G. Sung, C.-L. Chang dan M. Grundmann, "MediaPipe Hands: On-device Real-time Hand Tracking," dalam *2020 CVPR Workshop on Computer Vision for Augmented and Virtual Reality*, Seattle, 2020. doi: <https://doi.org/10.48550/arXiv.2006.10214>.
- [24] A. Wakjira dan S. Bhattacharya, "Predicting Student Engagement in the Online Learning Environment," *International Journal of Web-Based Learning and Teaching Technologies*, vol. 16, no. 6, pp. 1-21, 2021. doi: <https://doi.org/10.4018/IJWLTT.287095>.
- [25] I. Goodfellow, Y. Bengio dan A. Courville, *Deep Learning*, MIT Press, 2016.
- [26] Purwono, A. Ma'arif, W. Rahmانيar, H. I. K. Fathurrahman, A. Z. K. Frisky dan Q. M. u. Haq, "Understanding of Convolutional Neural Network (CNN): A Review," *International Journal of Robotics and Control Systems*, vol. 2, no. 4, pp. 739-748, 2022. doi: <http://dx.doi.org/10.31763/ijrcs.v2i4.888>.