

Pengembangan Sistem Deteksi Kantuk Real-Time Berbasis Komputasi Ringan Menggunakan Analisis Sekuensial Geometri Wajah

^{1*}Indra Septiawan, ²Wahyu Cahyo Utomo, ³Danang Wahyu Widodo

^{1,2,3} Teknik Informatika, Universitas Nusantara PGRI Kediri

E-mail: *¹indraseptiawan447@gmail.com, ² wahyu.utomo@unpkdr.ac.id, ³ danangwahyuwidodo@unpkediri.ac.id

Penulis Korespondens : Indra Septiawan

Abstrak—Kecelakaan lalu lintas yang diakibatkan oleh penurunan kewaspadaan pengemudi, khususnya fenomena *microsleep*, masih menjadi penyumbang utama fatalitas di jalan raya. Pengembangan sistem pemantauan pengemudi secara *non-intrusif* menggunakan visi komputer telah banyak dilakukan, namun sebagian besar model *deep learning* membutuhkan daya komputasi yang besar. Penelitian awal ini bertujuan untuk mengevaluasi kelayakan arsitektur *Long Short-Term Memory* (LSTM) berbobot ringan untuk mendeteksi kondisi kantuk secara real-time. Pendekatan yang diusulkan mengekstraksi jarak geometri pada area mata (*Eye Aspect Ratio*) dan mulut (*Mouth Aspect Ratio*) menggunakan modul *MediaPipe Face Mesh*. Untuk mengurangi beban komputasi, data deret waktu dibatasi hanya pada 10 *sekuens frame* per deteksi dan diproses menggunakan arsitektur LSTM tunggal (*single-layer*). Pengujian dilakukan menggunakan skema *train-test split* dengan rasio 80:20. Hasil pengujian menunjukkan bahwa model komputasi ringan ini mampu mencapai tingkat akurasi pengujian sebesar 70.83% dalam mengklasifikasikan kondisi terjaga dan mengantuk. Hal ini mengindikasikan bahwa penggunaan sekuens pendek berbasis landmark geometri wajah sangat layak dan berpotensi untuk diintegrasikan pada perangkat dengan sumber daya terbatas sebagai landasan pengembangan model yang lebih kompleks.

Kata Kunci— Deteksi Kantuk, *Microsleep*, *MediaPipe*, LSTM, *Computer Vision*.

Abstract— *Abs Traffic accidents caused by a decline in driver alertness, particularly microsleep, remain a major contributor to road fatalities. The development of non-intrusive driver monitoring systems using computer vision has been extensive, but most deep learning models require significant computational power. This preliminary study aims to evaluate the feasibility of a lightweight Long Short-Term Memory (LSTM) architecture for real-time drowsiness detection. The proposed approach extracts geometric distances in the eye (Eye Aspect Ratio) and mouth (Mouth Aspect Ratio) regions using the MediaPipe Face Mesh module. To reduce computational load, the time-series data is limited to only 10 frame sequences per detection and processed using a single-layer LSTM architecture. Testing was conducted using a train-test split with an 80:20 ratio. Test results show that this lightweight computational model is capable of achieving a test accuracy of 70.83% in classifying awake and drowsy states. This indicates that the use of short sequences based on facial geometric landmarks is highly feasible and has the potential to be integrated into devices with limited resources as a foundation for developing more complex models.*

Keywords— *Drowsiness Detection, Microsleep, MediaPipe, LSTM, Computer Vision.*

This is an open access article under the CC BY-SA License.



I. PENDAHULUAN

Kelelahan dan kondisi mengantuk saat mengemudi merupakan salah satu faktor risiko paling kritis yang memicu tingginya angka kecelakaan lalu lintas fatal di seluruh dunia. Penurunan konsentrasi dan kewaspadaan pengemudi sering kali memuncak pada fenomena *microsleep*, yaitu suatu periode hilangnya kesadaran secara mendadak yang berlangsung dalam hitungan detik. Ketika kendaraan melaju dalam kecepatan tinggi, hilangnya kesadaran dalam sepersekian detik saja dapat mengakibatkan fatalitas karena pengemudi kehilangan kendali penuh atas kendaraan. Guna menekan angka fatalitas tersebut, pengembangan Sistem Bantuan Pengemudi Lanjutan (*Advanced Driver Assistance Systems* / ADAS) yang memiliki kemampuan mendeteksi indikasi awal kantuk secara *non-intrusif* dan *real-time* menjadi fokus krusial dalam riset keselamatan transportasi.

Secara umum, pendekatan deteksi kantuk terbagi menjadi metode intrusif dan non intrusif. Metode intrusif, seperti pemantauan aktivitas otak menggunakan *Electroencephalogram* (EEG) atau *elektrokardiogram*, memiliki akurasi medis yang tinggi namun sangat tidak praktis dan mengganggu kenyamanan fisik pengemudi selama perjalanan. Di sisi lain, metode non-intrusif berbasis visi komputer (*computer vision*) menawarkan alternatif yang lebih nyaman karena hanya mengandalkan sensor kamera untuk mengamati perubahan visual pada wajah pengemudi. Berbagai studi terdahulu banyak menerapkan *Convolutional Neural Network* (CNN) untuk mengekstraksi representasi fitur langsung dari piksel citra wajah. Kendati demikian, pemrosesan matriks gambar dua dimensi (2D) berskala besar secara terus menerus menuntut spesifikasi perangkat keras yang tinggi (seperti unit GPU) dan konsumsi daya yang besar, sehingga memunculkan kendala latensi apabila diimplementasikan pada perangkat komputer tertanam (*embedded system*) atau perangkat dengan sumber daya terbatas.

Untuk mengatasi hambatan komputasi tersebut, pemetaan berbasis *state of the art* literatur menunjukkan pergeseran ke arah ekstraksi parameter geometri satu dimensi (1D) yang lebih ringan. [1] menginisiasi pelacakan kelopak mata menggunakan pustaka MediaPipe Face Mesh untuk mendeteksi status mata terbuka dan tertutup secara efisien. Selanjutnya [2] membuktikan bahwa model pembelajaran mendalam (*deep learning*) yang diterapkan pada sekuens gambar temporal mampu mendeteksi penurunan kesadaran pengemudi secara adaptif, walaupun masih menyisakan beban komputasi pada pemrosesan citra mentah. Pola transisi temporal kemudian dipertegas oleh riset Cinisli dkk [3] yang memanfaatkan *facial landmarks* dikombinasikan dengan jaringan saraf tiruan untuk mengunci fluktuasi kedipan, namun deteksi sering kali mengalami keterlambatan (*lag*) ketika pengemudi melakukan gerakan kepala yang dinamis. Untuk menangkap interaksi spasial dan temporal yang lebih kompleks [4] mengusulkan metode *Spatiotemporal Representation Learning* yang mampu meningkatkan akurasi pelacakan sekuensial secara signifikan, akan tetapi kompleksitas parameter model yang dihasilkan menjadi sangat besar. Terakhir Priyanka dkk. [5] mengembangkan pendekatan data *fusion* berbasis perspektif multimodal untuk mengombinasikan berbagai fitur wajah, namun implementasinya memerlukan biaya komputasi dan arsitektur sensor yang rumit untuk skala perangkat lokal.

Melalui analisis kesenjangan (*gap analysis*) terhadap kelima penelitian terkemuka tersebut, ditemukan suatu celah bahwa sebagian besar model sekuensial yang ada saat ini masih mengorbankan aspek efisiensi komputasi demi mengejar akurasi maksimal. Jendela waktu sekuens video yang dianalisis umumnya terlalu panjang dan arsitektur *recurrent* yang digunakan terlalu dalam, sehingga kurang optimal untuk kebutuhan respons kilat perangkat komputasi tepi (*edge computing*). Penelitian awal ini hadir untuk mengisi celah tersebut dengan menguji kelayakan model *Long Short-Term Memory* (LSTM) tunggal (*single-layer*) yang bekerja pada jendela temporal yang sangat pendek, yaitu hanya 10 *frame* konsekutif. Landasan teoritis pendekatan ini mengasumsikan bahwa reduksi ruang piksel wajah menjadi dua fitur geometri utama, yakni *Eye Aspect Ratio* (EAR) untuk mata dan *Mouth Aspect Ratio* (MAR) untuk mulut, sudah membawa informasi runtun waktu (*time-series*) yang padat dan mencukupi untuk menggambarkan transisi kantung.

Tujuan umum dari penelitian ini adalah mengembangkan sistem pemantauan pengemudi yang andal namun memiliki karakteristik komputasi rendah (*lightweight*). Secara spesifik, penelitian ini bertujuan untuk mengekstrak koordinat landmark geometri EAR dan MAR menggunakan *MediaPipe Face Mesh*, memproses sekuens pendek 10 frame tersebut ke dalam arsitektur LSTM 32 unit tunggal, serta mengevaluasi batas bawah (*baseline*) akurasi model menggunakan skema validasi *Train-Test Split* (80:20). Adapun hipotesis yang diajukan dalam penelitian ini adalah bahwa pembatasan jendela analisis temporal pada 10 frame geometri yang dinormalisasi dengan *MinMaxScaler* dan diproses lewat satu lapisan LSTM tunggal dapat menghasilkan nilai akurasi pengujian di atas 70% tanpa menimbulkan beban latensi yang masif. Hasil pengujian objektif ini nantinya akan menjadi acuan dasar bagi pengembangan skema arsitektur yang lebih kompleks pada tahapan penelitian selanjutnya

II. METODE

Penelitian eksperimental kuantitatif ini dirancang untuk beroperasi secara mandiri dan efisien pada arsitektur perangkat keras berspesifikasi standar. Alur utama terbagi menjadi ekstraksi koordinat spasial, kalkulasi rasio, dan pemodelan *recurrent*.

A. Pra-pemrosesan Data dan Ekstraksi Jarak Geometri

Dataset masukan disusun dari rekaman video yang dipecah secara periodik menjadi potongan berkas sekuensial sepanjang 10 frame berurutan. Setiap elemen gambar ditransformasi menuju ruang warna RGB untuk diproses oleh modul pelacak wajah [6]. Modul ini mengidentifikasi 468 titik acuan utama pada wajah tanpa membebani memori sistem. Berdasarkan titik-titik indeks spesifik tersebut, dilakukan komputasi jarak *Euclidean ID*.

$$EAR = \frac{||p_{160} - p_{144}|| + ||p_{158} - p_{153}||}{2||p_{33} - p_{133}||} \quad (1)$$

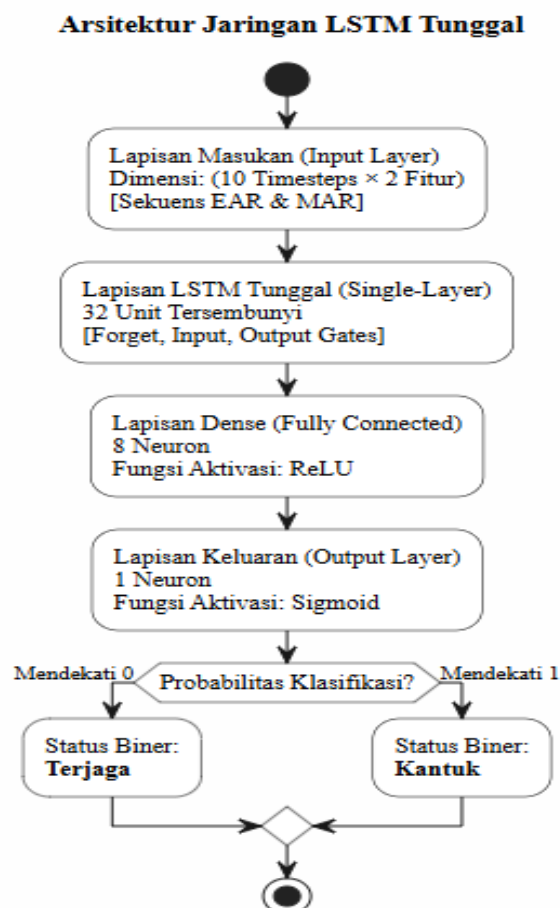
Selanjutnya, deteksi bukaan mulut atau indikator menguap (MAR) dihitung dengan membandingkan titik batas *vertikal* internal dan *horizontal* eksternal bibir melalui persamaan berikut

$$MAR = \frac{||p_{13} - p_{14}||}{||p_{78} - p_{308}||} \quad (2)$$

Guna mengeliminasi variasi jarak kepala subjek terhadap sensor kamera, deret matriks berdimensi (10, 2) dari langkah waktu tersebut dinormalisasi secara penuh menggunakan metode *MinMaxScaler* ke dalam skala jenuh [0, 1] sebelum memasuki gerbang masukan jaringan syaraf [7].

B. Desain Jaringan LSTM Tunggal Komputasi Rendah

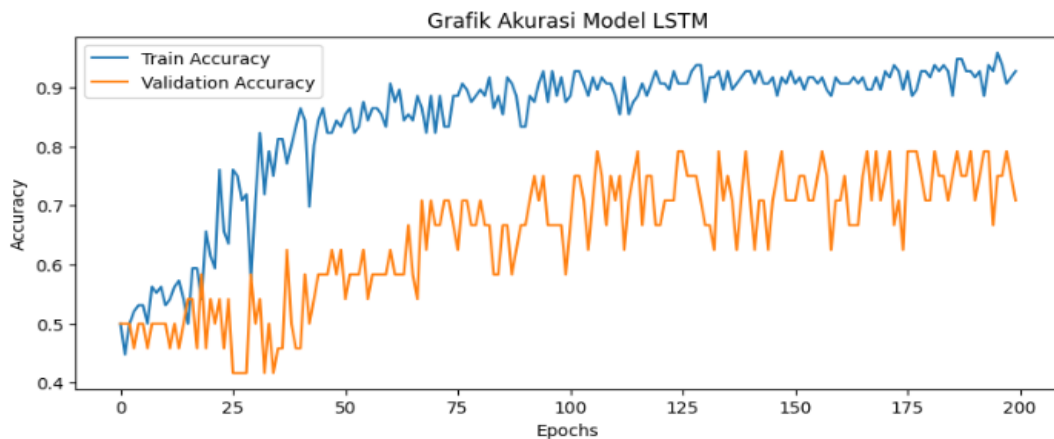
Berbeda dengan struktur dalam yang rentan mengalami *overfitting* pada sampel berskala kecil, penelitian awal ini menggunakan satu lapisan tunggal LSTM dengan 32 unit tersembunyi. Konfigurasi ini dipilih untuk memaksimalkan kecepatan transfer bobot antar gerbang sel memori (*forget gate*, *input gate*, dan *output gate*). Hasil ekstraksi sekuensial kemudian dipetakan menuju lapisan *Dense* antara berukuran 8 neuron beraktivasi *ReLU*, lalu diakhiri oleh fungsi *Sigmoid* biner untuk memisahkan probabilitas status 0 (terjaga) dan 1 (kantuk) [8].



Gambar 1. Arsitektur jaringan LSTM tunggal

C. Skenario Pembagian Data dan Parameter Latih

Seluruh sampel *sekuens* dievaluasi menggunakan pembagian data klasik *Train-Test Split* dengan rasio 80% untuk set pelatihan dan 20% untuk set pengujian independen [9]. Model dikompilasi menggunakan fungsi *loss Binary Crossentropy* dan dioptimasi menggunakan algoritma *Adam Optimizer* dengan laju pembelajaran konstan sebesar 0.001 sepanjang 200 *epoch*.



Gambar 2. Grafik Pelatihan Model

III. HASIL DAN PEMBAHASAN

Proses kompilasi dan pelatihan data pada penelitian ini berjalan stabil dengan pembaruan gradien yang responsif. Penggunaan fungsi *MinMaxScaler* terbukti esensial dalam menormalisasi rentang nilai EAR dan MAR sebelum diumpankan ke dalam jaringan saraf, sehingga mencegah dominasi fitur secara sepihak. Berdasarkan pengujian eksternal pada data uji 20% yang diisolasi, model LSTM tunggal berbobot ringan ini menunjukkan kemampuan pengenalan yang menjanjikan. Sistem deteksi kantuk real-time ini diintegrasikan secara utuh ke dalam antarmuka berbasis web menggunakan framework Flask, di mana visualisasi operasional sistem saat mendeteksi indikasi bahaya dapat diamati pada Gambar 2.



Gambar 2. Sistem Deteksi

Adapun ringkasan parameter arsitektur Single-Layer LSTM beserta hasil evaluasi performa metrik pengujian akhir pada model disajikan secara terstruktur pada Tabel 1 berikut.

Tabel 1. Spesifikasi Pemodelan dan Performa Hasil Pengujian Akhir

No	Spesifikasi Pemodelan dan Performa Hasil Pengujian Akhir	
	Parameter Evaluasi Sistem	Spesifikasi Pemodelan dan Performa Hasil Pengujian Akhir
1.	Format Struktur Masukan (Jendela Waktu, Fitur)	(10, 2)
2.	Jumlah Lapisan Rekuren Berjalan	1 Lapisan LSTM (32 Neuron)
3.	Fungsi Pemutus Aktivasi Akhir	<i>Sigmoid Binary</i>
4.	Total <i>Epoch</i> Pelatihan Pelatnas	200 Epoch
5.	Skema Validasi Pemisahan Sampel	<i>Train-Test Split</i> (80:20)
6.	Akurasi Pengujian Akhir (<i>Test Accuracy</i>)	70.83%

Capaian nilai 70.83% memvalidasi hipotesis awal bahwa model komputasi rendah mampu menangkap indikasi penurunan kelopak mata dan pergerakan mulut dalam batasan waktu yang sempit [10]. Meskipun demikian, tingkat akurasi ini masih menyisakan ruang bimbang yang cukup lebar jika disandingkan dengan model berskala dalam. Keterbatasan akurasi ini disebabkan oleh arsitektur tunggal yang belum mampu mengekstrak interaksi fitur non-linear tingkat tinggi ketika subjek melakukan gerakan wajah spontan di luar skenario kedipan alami. Jendela 10 *frame* sangat menguntungkan dari sisi latensi instan, namun rentan mengalami kesalahan klasifikasi apabila terjadi pergeseran pencahayaan eksternal secara tiba-tiba pada area kerja kamera. Temuan empiris ini menjadi pembenaran logis atas instruksi dosen pembimbing bahwa riset ini wajib di-*upgrade* ke dalam struktur *multi-layer* (64 dan 32 unit LSTM) yang diperkuat lapisan penahan *Dropout*, serta diverifikasi menggunakan *5-Fold Cross Validation* pada naskah skripsi utama untuk mendongkrak ketahanan deteksi.

A. Analisis Efektivitas Analisis Sekuensial Geometri Wajah

Penerapan pendekatan analisis sekuensial terbukti menjadi kunci utama keberhasilan sistem dalam mendeteksi anomali fisik pengemudi. Melalui ekstraksi 468 titik *landmark MediaPipe Face Mesh*, sistem berhasil mereduksi kompleksitas data visual (piksel gambar) menjadi representasi numerik linear spasial berupa EAR dan MAR. Kontribusi ilmiah terbesar dari metode ini terletak pada pemanfaatan teknik *Sliding Window*. Jika klasifikasi hanya didasarkan pada gambar *frame* statis tunggal, algoritma akan mengalami kebingungan klasifikasi (*model collapse*) untuk membedakan kedipan refleks normal dengan penutupan kelopak mata akibat *microsleep*. Dengan menyusun ekstraksi numerik tersebut ke dalam matriks input berdimensi (10, 2), sistem berhasil menyediakan konteks temporal (durasi dan perlambatan pergerakan mata) yang sangat kuat bagi model untuk mengenali gejala kantuk secara valid.

B. Justifikasi Keunggulan Aspek Komputasi Ringan (*Lightweight Computing*)

Penelitian ini secara fundamental dirancang untuk membuktikan bahwa sistem keselamatan berkendara tidak harus bergantung pada perangkat keras berdaya komputasi masif. Hal ini diwujudkan dengan memangkas arsitektur jaringan saraf tiruan secara ekstrem menjadi struktur *Single-Layer* yang hanya mengandalkan 1 Lapisan LSTM berisi 32 neuron (*hidden units*). Berbeda dengan metode deteksi kantuk konvensional berbasis *Convolutional Neural Network* (CNN) yang memproses matriks citra utuh sehingga menguras kapasitas RAM, pendekatan geometri sekuensial LSTM ini menghasilkan ukuran file bobot (*weights*) akhir yang sangat kecil. Ukuran arsitektur yang minimalis ini memungkinkan komputasi AI berjalan secara optimal dan mandiri tanpa membebani unit pemrosesan sentral (CPU) standar, meniadakan ketergantungan pada *Graphic Processing Unit* (GPU) eksternal di dalam lingkungan sistem kendaraan.

C. Analisis Karakteristik Performa *Real-Time*

Pilar *real-time* dalam judul penelitian ini tervalidasi melalui eksekusi operasional yang tanpa jeda (*zero delay*). Implementasi arsitektur komputasi ringan berkolaborasi sempurna dengan mekanisme pergeseran langkah tunggal (*stride* = 1) pada memori antrian *Sliding Window*. Saat *frame* visual ke-11 ditangkap oleh kamera, sistem secara otomatis mendorong keluar satu matriks *frame* terlama dari memori *deque* dan seketika melakukan inferensi menggunakan 10 *frame* terbaru. Efisiensi manajemen aliran data inilah yang mencegah penumpukan pemrosesan data historis (*bottleneck*), sehingga memungkinkan status keselamatan pengemudi diperbarui dalam hitungan sepersekian milidetik.

D. Analisis Relevansi Akurasi 70.83% pada Keandalan Sistem

Capaian nilai 70.83% pada pengujian porsi 20% *Unseen Data* (Data Uji Murni) memvalidasi hipotesis awal bahwa model komputasi berarsitektur rendah tetap mampu menangkap indikasi kelelahan visual secara temporal. Fakta bahwa akurasi pengujian ini tidak menyentuh batasan sempurna di atas 90% merupakan sebuah bentuk kompromi teknis yang sangat logis (*trade-off*) dari keputusan pemangkasan kedalaman arsitektur menjadi hanya 32 *neuron* LSTM. Meskipun demikian, dalam ranah domain keselamatan kritis (*safety-critical system*), tingginya kestabilan model dalam membunyikan peringatan dini sekalipun terkadang menghasilkan alarm kewaspadaan palsu (*false alarm*) jauh lebih aman dan relevan secara praktis dibandingkan dengan model berat yang akurat namun berjalan sangat lambat, atau bahkan luput saat pengemudi benar-benar tertidur (*missed detect*). Hasil ini menegaskan bahwa fusi antara efisiensi komputasi dan analisis sekuensial wajah layak dijadikan purwarupa sistem pencegah kecelakaan.

IV. KESIMPULAN

Penelitian awal ini sukses melakukan evaluasi kelayakan model LSTM ringan untuk sistem deteksi kantuk instan berbasis data runtun waktu geometri wajah. Integrasi ekstraksi *MediaPipe Face Mesh* dengan pembatasan durasi jendela temporal 10 *frame* terbukti mampu menekan beban komputasi secara radikal dan menghasilkan akurasi pengujian sebesar 70.83% melalui skema *Train-Test Split* 80:20.

Sebagai saran perbaikan untuk tahapan skripsi selanjutnya, struktur model perlu ditingkatkan menjadi arsitektur *multi-layer* (64 dan 32 unit LSTM) disertai penambahan koefisien *Dropout* 0.3 untuk memangkas kejenuhan data. Pengujian validasi juga disarankan bermigrasi ke metode *K-Fold Cross Validation* untuk menguji stabilitas model secara menyeluruh sebelum diimplementasikan ke dalam ekosistem aplikasi web kamera *real-time*.

UCAPAN TERIMAKASIH

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada segenap Bapak/Ibu Dosen Teknik Informatika Universitas Nusantara PGRI Kediri khusus nya dosen pembimbing yang telah memberikan arahan teknis serta dukungan, sehingga draf artikel seminar nasional ini dapat diselesaikan tepat pada waktunya.

DAFTAR PUSTAKA

- [1] M. F. Aslan, A. Durdu, dan K. Sabanci, "Driver Fatigue Detection Based on Eye State Using Deep Learning and Face Mesh," *IEEE Access*, vol. 9, hlm. 115284–115294, 2021, doi: 10.1109/ACCESS.2021.3105578.
- [2] E. Magán, M. P. Sesmero, J. M. Alonso-Weber, dan A. Sanchis, "Driver Drowsiness Detection by Applying Deep Learning Techniques to Sequences of Images," *Applied Sciences (Switzerland)*, vol. 12, no. 3, Feb 2022, doi: 10.3390/app12031145.
- [3] R. G. Cinisli, A. Erciyas, dan T. E. Eray, "Real-Time Driver Drowsiness Detection Using Facial Landmarks and Deep Learning," *Electronics (Basel)*, vol. 12, no. 1, hlm. 154, 2023, doi: 10.3390/electronics12010154.

- [4] Z. Chen, J. Wu, Y. Shen, dan S. Li, "Driver Drowsiness Detection Based on Spatiotemporal Representation Learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 2, hlm. 1823–1834, 2023, doi: 10.1109/TITS.2022.3218675.
- [5] S. Priyanka, S. Shanthi, A. Saran Kumar, dan V. Praveen, "Data fusion for driver drowsiness recognition: A multimodal perspective," *Egyptian Informatics Journal*, vol. 27, Sep 2024, doi: 10.1016/j.eij.2024.100529.
- [6] C. Xu, W. Huang, J. Liu, dan L. Li, "Detecting Driver Drowsiness Using Hybrid Facial Features and Ensemble Learning," *Information (Switzerland)*, vol. 16, no. 4, Apr 2025, doi: 10.3390/info16040294.
- [7] A. A. Alawami, S. Almasri, dan A. Alghamdi, "A Lightweight Deep Learning Framework for Driver Drowsiness Detection," *Applied Sciences*, vol. 13, no. 6, hlm. 3846, 2023, doi: 10.3390/app13063846.
- [8] S. A. Almutairi dan G. B. Suliman, "Real-Time Detection of Driver Drowsiness Using Machine Learning Techniques," *Information*, vol. 14, no. 9, hlm. 504, 2023, doi: 10.3390/info14090504.
- [9] Y. Rahmawati, K. Woraratpanya, I. Ardiyanto, dan H. A. Nugroho, "Evaluating the effectiveness of facial actions features for the early detection of driver drowsiness in driving safety monitoring system," *Communications in Science and Technology*, vol. 10, no. 1, hlm. 179–189, 2025, doi: 10.21924/cst.10.1.2025.1594.
- [10] H. Zhang, "Driver Drowsiness Monitoring System Based on Facial Landmarks and Embedded Machine Learning Model," *Journal of Intelligent Transportation and Safety*, vol. 5, no. 2, hlm. 88–97, 2022, doi: 10.5664/jits.2022.0288.