



JOURNAL CERITA:

**Creative Education of Research in Information
Technology and Artificial Informatics**

Vol. 11 No. 2 (2025) 255-263

e-ISSN: 2655 - 2574

Peningkatan Akurasi Pada Slowfast Network Menggunakan Multi-Head Self Attention Layer

Muhammad Reza¹, Arief Setyanto^{*2}

¹Program Pascasarjana, Universitas Amikom Yogyakarta, Yogyakarta, Indonesia

Email: muhammadreza348@students.amikom.ac.id¹; arief_s@amikom.ac.id^{*2}

Reza, M., & Setyanto, A. (2025). Peningkatan Akurasi Pada Slowfast Network Menggunakan Multi-Head Self Attention Layer. *Journal Cerita: Creative Education of Research in Information Technology and Artificial Informatics*, 11(2), 255-263

DOI: <https://doi.org/10.33050/cerita.v11i2.3870>

ABSTRAK

Keselamatan berkendara merupakan aspek krusial dalam sistem transportasi modern, di mana perilaku pengemudi seperti kantuk dan kehilangan fokus menjadi faktor utama penyebab kecelakaan. Penelitian ini mengusulkan sistem deteksi perilaku pengemudi berbasis video dengan menggunakan arsitektur SlowFast Network yang dikombinasikan dengan mekanisme Self-Attention. SlowFast Network memungkinkan pemrosesan dua jalur informasi temporal secara paralel, yaitu gerakan cepat dan lambat, untuk menangkap pola ekspresi dan dinamika visual pengemudi. Mekanisme Self-Attention ditambahkan untuk memperkuat kemampuan model dalam mengenali fitur penting secara kontekstual. Dataset yang digunakan adalah SUST Driver Drowsiness Dataset, yang telah melalui proses segmentasi dan normalisasi. Model dilatih menggunakan pembagian data pelatihan, validasi, dan pengujian, serta dievaluasi dengan metrik akurasi, presisi, recall, dan F1-score. Hasil pengujian menunjukkan bahwa penambahan Self-Attention meningkatkan performa model secara signifikan, dengan akurasi mencapai 96% pada data uji seimbang. Visualisasi attention map dan filter Conv3D mendukung interpretasi bahwa model mampu menangkap pola perilaku tanpa memerlukan deteksi eksplisit bagian wajah. Sistem ini menunjukkan potensi untuk diterapkan dalam sistem peringatan dini berbasis video guna meningkatkan keselamatan berkendara secara real-time.

Kata kunci: Deteksi Kantuk Pengemudi, Jaringan SlowFast, Self-Attention, Computer Vision.

ABSTRACT

Driver behavior is a critical factor in road safety, with conditions such as drowsiness and loss of focus being major contributors to traffic accidents. This study proposes a video-based driver behavior detection system using the SlowFast Network architecture combined with a Self-Attention mechanism. The SlowFast Network enables parallel processing of fast and slow temporal features to capture expression dynamics and behavioral patterns, while Self-Attention enhances the model's ability to contextually prioritize important visual features. The system was trained using the SUST Driver Drowsiness Dataset, which was preprocessed through video segmentation and pixel normalization. The model was evaluated on separate training, validation, and testing data splits using metrics such as accuracy, precision, recall, and F1-score. Experimental results show that the integration of Self-Attention significantly improves model performance, achieving up to 96% accuracy on balanced test data. Visualization of attention maps and Conv3D filters confirms the model's ability to extract relevant spatial-temporal patterns without relying on explicit facial landmark detection. This system demonstrates strong potential for real-time implementation in intelligent driver monitoring and early warning systems to enhance road safety.

Keywords: Driver Drowsiness Detection, SlowFast Network, Self-Attention, Computer Vision.

I. PENDAHULUAN

Keselamatan berkendara merupakan isu global yang terus menjadi perhatian, terutama dengan meningkatnya angka kecelakaan lalu lintas yang disebabkan oleh perilaku pengemudi yang berisiko, seperti mengantuk, kurang fokus, dan penggunaan ponsel saat berkendara. Menurut Organisasi Kesehatan Dunia (WHO), kecelakaan lalu lintas adalah salah satu penyebab utama kematian di dunia, di mana kelalaian pengemudi berkontribusi secara signifikan terhadap insiden tersebut (Ajay Kumar, 2023). Menurut WHO, kecelakaan lalu lintas mengakibatkan 1,19 juta kematian pada tahun 2021, dengan mayoritas korban berada di usia produktif (18–59 tahun). Hal ini memicu dampak sosial dan ekonomi yang signifikan, termasuk hilangnya produktivitas tenaga kerja dan meningkatnya beban finansial pada sistem kesehatan serta keluarga korban (WHO, Global Status report on road safety. 2023). Kecelakaan yang disebabkan oleh perilaku pengemudi tidak hanya berdampak pada korban jiwa, tetapi juga menyebabkan kerugian material yang besar. Oleh karena itu, diperlukan sistem deteksi perilaku pengemudi yang mampu mengidentifikasi dan memberikan peringatan dini terhadap perilaku berisiko sebelum kecelakaan terjadi (Fangming Qu, 2024).

Dalam beberapa tahun terakhir, teknologi computer vision telah banyak digunakan dalam deteksi perilaku pengemudi, khususnya dalam mengenali ekspresi wajah yang mengindikasikan kantuk atau kurangnya perhatian. Penelitian

sebelumnya telah mengimplementasikan algoritma berbasis Convolutional Neural Networks (CNN) untuk menganalisis ekspresi wajah dan gerakan kepala pengemudi. Namun, pendekatan ini memiliki beberapa keterbatasan, seperti kurangnya kemampuan dalam menangkap dinamika temporal perilaku pengemudi dan sensitivitas terhadap perubahan kondisi pencahayaan atau sudut kamera (J. Robert Theivadas, 2024). Beberapa metode lain, seperti Haar Cascade dan pendekatan berbasis fitur, misalnya Eye Aspect Ratio (EAR) dan Mouth Aspect Ratio (MAR), juga telah diterapkan, tetapi masih memiliki akurasi yang terbatas dalam kondisi pencahayaan rendah atau sudut pandang kamera yang tidak ideal (Jing Liu, 2021).

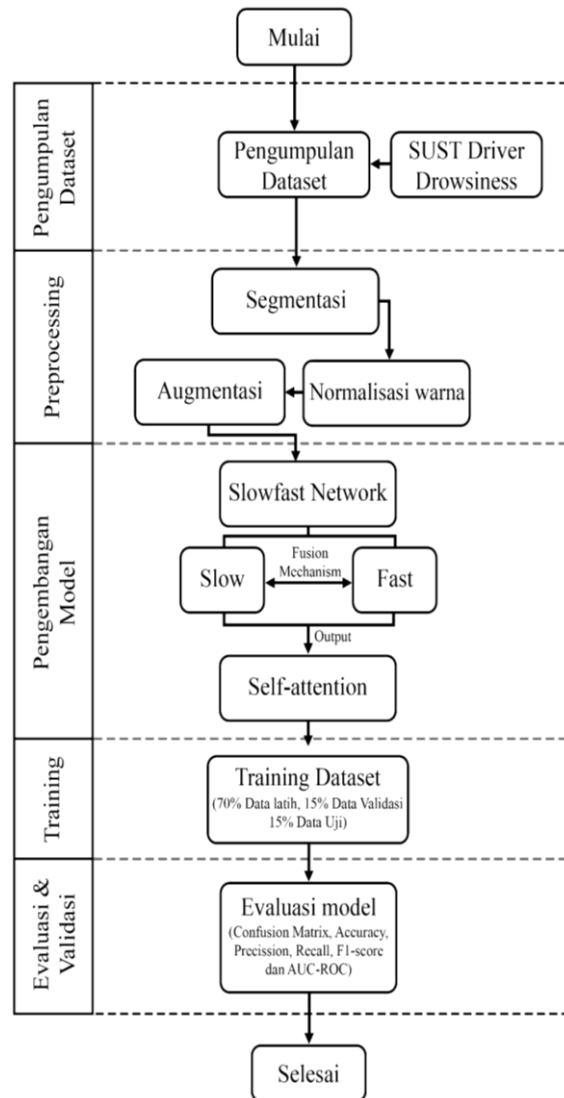
Untuk mengatasi keterbatasan tersebut, penelitian ini mengusulkan metode deteksi perilaku pengemudi berbasis SlowFast Network dan Self-Attention. SlowFast Network merupakan arsitektur jaringan saraf konvolusional yang memproses video input melalui dua jalur pemrosesan berbeda: jalur cepat (Fast) dan jalur lambat (Slow). Jalur cepat mengambil informasi frame yang banyak untuk menangkap dinamika temporal, sedangkan jalur lambat mengambil frame lebih sedikit sebagai konteks global. Mekanisme ini memungkinkan model untuk menganalisis perubahan pola visual secara menyeluruh dalam durasi pendek. Sementara itu, Self-Attention digunakan untuk menekankan fitur penting dari gabungan dua jalur tersebut, sehingga meningkatkan sensitivitas model terhadap indikasi kantuk secara umum tanpa bergantung pada segmentasi wajah secara

eksplisit (Myeongjun Kim, 2020). Penelitian ini mengintegrasikan Self-Attention ke dalam arsitektur SlowFast untuk mengevaluasi pengaruhnya terhadap peningkatan kemampuan model dalam mendeteksi perilaku kantuk. Penelitian ini meneliti seberapa besar kontribusi Self-Attention dalam meningkatkan performa dibandingkan model SlowFast standar. Penelitian ini juga akan berfokus pada deteksi perilaku pengemudi menggunakan video yang direkam di dalam kabin kendaraan dengan metode SlowFast Network dan Self-Attention. Dataset yang digunakan dalam penelitian ini adalah SUST Driver Drowsiness Dataset, yang mencakup berbagai kondisi pencahayaan dan sudut kamera untuk menguji keandalan sistem. Pengujian dilakukan dalam skenario simulasi dan rekaman video tanpa implementasi langsung pada kendaraan yang bergerak di jalan raya. Dengan pengembangan sistem ini, diharapkan deteksi perilaku pengemudi dapat dilakukan secara real-time dengan tingkat akurasi yang lebih tinggi.

II. METODE PENELITIAN

Penelitian ini disusun secara sistematis melalui serangkaian tahapan yang saling terkait untuk mencapai tujuan deteksi perilaku kantuk pengemudi berbasis video. Secara garis besar, tahapan penelitian meliputi pengumpulan dataset, preprocessing data, pengembangan model, serta pelatihan dan evaluasi model. Pada tahap pengumpulan dataset, data yang digunakan berasal dari SUST Driver Drowsiness Dataset yang dipilih karena mencakup variasi kondisi pencahayaan, ekspresi wajah, dan sudut pandang kamera. Pada tahap preprocessing, dilakukan segmentasi video menjadi klip pendek berdurasi 2 detik menggunakan metode sliding window, augmentasi data untuk memperbanyak variasi sampel, serta normalisasi warna dengan mengubah rentang piksel dari 0–255 menjadi 0–1 untuk mempercepat konvergensi pelatihan. Tahap pengembangan model dilakukan dengan membangun arsitektur SlowFast Network yang terdiri dari jalur cepat dan jalur lambat, diikuti dengan mekanisme Self-Attention untuk meningkatkan relevansi fitur sebelum klasifikasi. Terakhir, tahap pelatihan dan evaluasi melibatkan pembagian data menjadi 70% data latih, 15% data validasi, dan 15% data uji, serta evaluasi performa model menggunakan

Confusion Matrix, Akurasi, Precision, Recall, F1-Score, dan AUC-ROC..



Gambar 1. Tahapan Penelitian

A. Metode Penelitian

Penelitian ini merupakan penelitian eksperimental dengan pendekatan kuantitatif yang bertujuan untuk mendeteksi perilaku pengemudi menggunakan kombinasi metode SlowFast Network dan Self-Attention. Pendekatan ini dipilih karena mampu menganalisis video melalui dua jalur temporal yang berbeda yaitu jalur cepat dan jalur lambat untuk menangkap dinamika perilaku secara menyeluruh. Selain itu, mekanisme Self-Attention digunakan untuk menekankan fitur penting dari gabungan dua jalur tersebut dalam proses klasifikasi, tanpa ketergantungan pada deteksi bagian wajah secara eksplisit. Penelitian ini dilakukan secara rasional, empiris, dan sistematis. Pendekatan rasional digunakan dalam

pengembangan arsitektur model berdasarkan teori computer vision dan kecerdasan buatan. Pendekatan empiris diterapkan melalui pengujian model menggunakan dataset video pengemudi yang telah dikurasi. Sementara itu, pendekatan sistematis dilakukan melalui serangkaian tahapan, mulai dari pengumpulan dataset, preprocessing data, pengembangan model, pelatihan dan evaluasi, hingga pengujian sistem menggunakan video uji dari dataset yang telah dipisahkan.

B. Pengumpulan Dataset

Dataset yang digunakan dalam penelitian ini adalah SUST Driver Drowsiness Dataset, yang terdiri dari rekaman video ekspresi wajah pengemudi dalam berbagai kondisi pencahayaan dan sudut kamera (Esra Kavalci Yilmaz, 2022). Dataset ini dipilih karena telah digunakan dalam berbagai penelitian sebelumnya yang berfokus pada deteksi kantuk dan gangguan perhatian menggunakan teknik deep learning. Dataset ini terdiri dari dua kategori perilaku pengemudi, yaitu kondisi normal sebanyak 1.099 video, dan kondisi kantuk atau kehilangan fokus sebanyak 975 video. Setiap video berdurasi 10 detik dengan format .mp4 dan memiliki resolusi asli 720×1280 piksel. Dataset ini telah dibagi menjadi tiga subset, yaitu training set, validation set, dan testing set. Seluruh video diolah lebih lanjut melalui tahap preprocessing untuk menyiapkan input model berupa sequence frame sesuai dengan kebutuhan arsitektur SlowFast Network.

C. Preprocessing Data

Untuk meningkatkan kualitas data sebelum pelatihan model, dilakukan beberapa tahap preprocessing, yaitu:

1. Segmentasi Temporal

Setiap video berdurasi 10 detik dipotong menjadi klip pendek berdurasi sekitar 2 detik menggunakan metode sliding window dengan panjang 60 frame dan overlap 50%. Segmentasi ini bertujuan untuk mempertahankan kontinuitas pola temporal dan meningkatkan jumlah sampel pelatihan.

2. Normalisasi Citra

Setiap frame yang diambil dari video terlebih dahulu dikonversi dari format warna BGR (standar default pada OpenCV) menjadi format RGB yang umum digunakan dalam

pemrosesan citra. Kemudian, ukuran gambar diubah menjadi resolusi tetap 112×112 piksel agar sesuai dengan kebutuhan input model. Perubahan ukuran ini dilakukan melalui proses resampling piksel, sehingga seluruh gambar tetap dipertahankan namun dengan kemungkinan perubahan rasio aspek (distorsi kecil) apabila rasio panjang dan lebar aslinya tidak 1:1. Setelah itu, nilai warna pada setiap piksel, yang semula berada dalam rentang 0 hingga 255, disesuaikan dengan cara dibagi 255. Hasilnya adalah nilai yang berada pada skala antara 0 dan 1, yang dikenal sebagai normalisasi. Nilai yang telah dinormalisasi ini kemudian dikonversi ke tipe data float (desimal) agar dapat diproses lebih efisien oleh model deep learning. Proses ini bertujuan untuk menyamakan skala input, mengurangi variabilitas data, dan mempercepat proses pelatihan model.

D. Pengembangan Model

Model yang dikembangkan dalam penelitian ini menggunakan arsitektur SlowFast Network yang dimodifikasi dan dikombinasikan dengan mekanisme Self-Attention. Arsitektur ini dirancang untuk menangkap informasi temporal dan spasial dari video secara lebih efektif dalam konteks deteksi perilaku kantuk pengemudi.

1. SlowFast Network

Arsitektur SlowFast Network terdiri dari dua jalur pemrosesan paralel, yaitu Slow Pathway dan Fast Pathway, yang masing-masing dirancang untuk menangkap informasi visual dari video dengan kecepatan temporal yang berbeda.

- a) Fast Pathway menerima 30 frame berurutan dari video berdurasi pendek. Jalur ini dirancang untuk mendeteksi perubahan cepat seperti kedipan mata, gerakan kepala mendadak, atau aktivitas mikro lainnya dalam waktu singkat.
- b) Slow Pathway menerima 8 frame yang dipilih secara merata dari 30 frame tersebut. Jalur ini fokus pada perubahan lambat dan konteks global, seperti arah pandangan, penurunan aktivitas, atau ekspresi umum wajah.

Setiap jalur kemudian memproses inputnya menggunakan layer Conv3D, yang bertugas mengekstraksi pola spasial (bentuk) dan temporal (gerakan) secara bersamaan. Tensor

input pada setiap jalur direpresentasikan sebagai:

$$X \in R^{TxHxWxC}$$

Dengan:

T: jumlah frame,

H × W: resolusi frame,

C: jumlah channel warna (RGB = 3).

R: himpunan bilangan real, menunjukkan bahwa nilai piksel pada tensor X berada dalam domain bilangan riil (floating point), bukan bilangan bulat.

Operasi Conv3D dilakukan menggunakan kernel ukuran 3×3×3, yang dijalankan untuk seluruh channel:

$$Output(t,h,w,c)=\sum_{i,j,k} \sum_c X(t+i,h+j,w+k,c') \cdot W(i,j,k,c',c)$$

t: indeks waktu (frame ke-t dalam klip video).

h: indeks tinggi (baris pixel pada frame).

w: indeks lebar (kolom pixel pada frame).

c: channel warna input (misalnya, R, G, atau B).

i, j, k: offset/indeks lokal dari kernel 3D pada dimensi waktu (i), tinggi (j), dan lebar (k).

c': channel warna dari input X saat dilakukan konvolusi untuk menghasilkan channel output c.

Hasil dari Conv3D kemudian diringkas menggunakan Global Average Pooling (GAP) 3D untuk menghasilkan representasi vektor:

$$f_c = \frac{1}{T \cdot H \cdot W} \sum_{t=1}^T \sum_{h=1}^H \sum_{w=1}^W Output(t, h, w, c)$$

Masing-masing jalur akan menghasilkan vektor berdimensi tetap, misalnya 1×128. Kedua vektor dari Slow dan Fast Pathway kemudian digabungkan (concatenated) menjadi:

$$FusedFeature = [Ffast \parallel Fslow] \in R^{1 \times 256}$$

Vektor ini menjadi representasi akhir dari video, yang memuat informasi komprehensif dari kedua perspektif: cepat dan lambat. Representasi ini diteruskan ke tahap selanjutnya, yaitu Self-Attention dan klasifikasi.

2. Self-Attention

Setelah dua jalur fitur digabungkan, model menggunakan mekanisme Self-Attention untuk meningkatkan relevansi fitur sebelum klasifikasi akhir. Penelitian ini menggunakan

layer Multi-Head Self-Attention (MHSA), yang memungkinkan model menangkap hubungan antar fitur dari berbagai perspektif secara paralel. Pertama, vektor fitur gabungan $X \in R^{1 \times 256}$ diproyeksikan menjadi Query (Q), Key (K), dan Value (V) menggunakan bobot transformasi:

$$Q = XW^Q, K = XW^K, V = XW^V$$

Keterangan:

X: Vektor input (hasil penggabungan fitur Slow dan Fast) berukuran 1×256

W^Q, W^K, W^V : Matriks bobot transformasi (trainable parameters) untuk membentuk Query, Key, dan Value

Q, K, V: Vektor hasil proyeksi input ke ruang Query, Key, dan Value

Kemudian, perhatian dihitung menggunakan fungsi Softmax dari hasil dot-product antara Q dan K, dibagi dengan akar dari dimensi key (dk):

$$Attention(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{dk}}\right)V$$

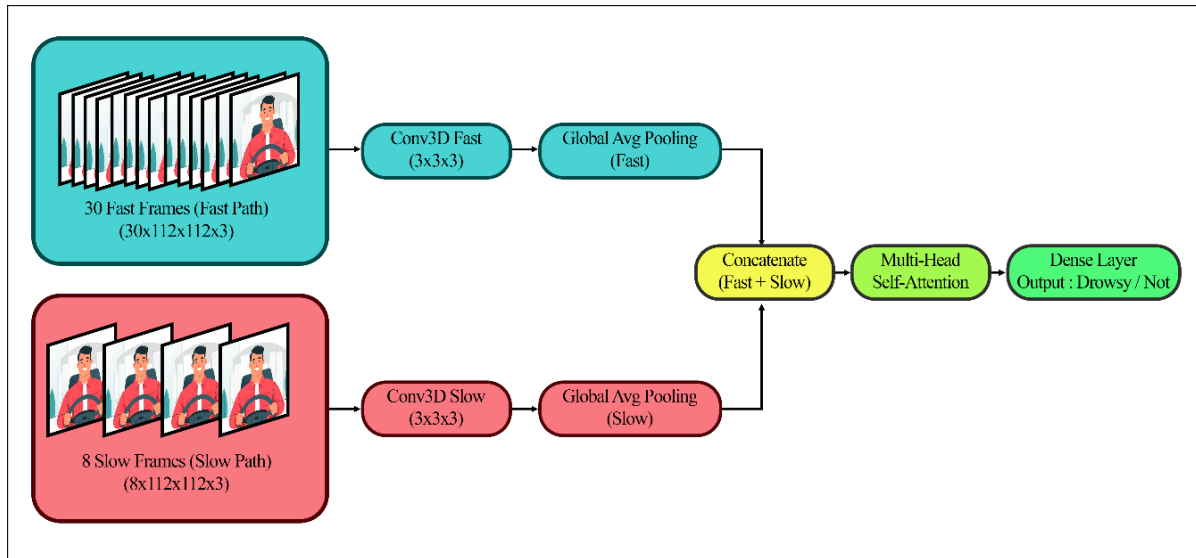
Mekanisme ini menghasilkan vektor fitur baru yang memprioritaskan bagian-bagian penting dari input gabungan, seperti ekspresi wajah atau gerakan kepala yang relevan terhadap deteksi kantuk. Setelah perhatian dihitung, vektor hasil Self-Attention diteruskan ke dense layer atau layer klasifikasi. Fungsi aktivasi sigmoid digunakan untuk menghasilkan nilai prediksi biner:

$$\hat{y} = \sigma(W_o \cdot AttentionOutput + b)$$

Jika $\hat{y} > 0.5$, maka model memprediksi drowsy; jika tidak, maka not drowsy. Gambar 2 menyajikan arsitektur model yang diusulkan dalam penelitian ini. Struktur dua jalur pemrosesan, yaitu Slow Pathway dan Fast Pathway, diadaptasi dari penelitian yang dilakukan oleh Feichtenhofer et al. (2019) terkait pengenalan video menggunakan SlowFast Networks. Penelitian ini melakukan modifikasi dengan mengintegrasikan mekanisme Self-Attention ringan setelah proses penggabungan fitur dari kedua jalur, dengan tujuan untuk

meningkatkan relevansi representasi fitur sebelum memasuki tahap klasifikasi akhir.

menggeneralisasi prediksi dengan baik terhadap data baru



Gambar 2. Arsitektur SlowFast dan Self-Attention

E. Pelatihan Model

Model SlowFast yang telah dikembangkan dilatih menggunakan dataset video yang telah melalui proses segmentasi dan normalisasi. Dataset dibagi menjadi tiga bagian utama, yaitu 70% untuk pelatihan (training), 15% untuk validasi (validation), dan 15% untuk pengujian (testing). Setiap video klip diolah menjadi dua jalur input, yaitu Fast Pathway yang terdiri dari 30 frame berurutan, dan Slow Pathway yang terdiri dari 8 frame yang diambil secara merata dari klip yang sama. Kedua jalur ini diolah secara paralel dalam arsitektur model.

Selama pelatihan, digunakan pendekatan efisien berbasis custom data generator (`tf.keras.utils.Sequence`) untuk memuat data secara bertahap dalam bentuk batch, guna menghindari beban memori yang berlebihan. Proses pelatihan model dilakukan menggunakan algoritma optimasi Adam dengan fungsi loss binary cross-entropy, yang sesuai untuk klasifikasi biner antara kondisi normal dan kondisi kantuk pada pengemudi. Setiap batch berisi 4 klip video, dengan ukuran frame yang disesuaikan menjadi 112×112 piksel. Model dilatih selama 10 epoch untuk memastikan konvergensi yang stabil dan menghindari overfitting. Tujuan utama dari proses pelatihan ini adalah agar model dapat belajar mengenali pola perilaku pengemudi berdasarkan informasi temporal dan spasial dari video, serta mampu

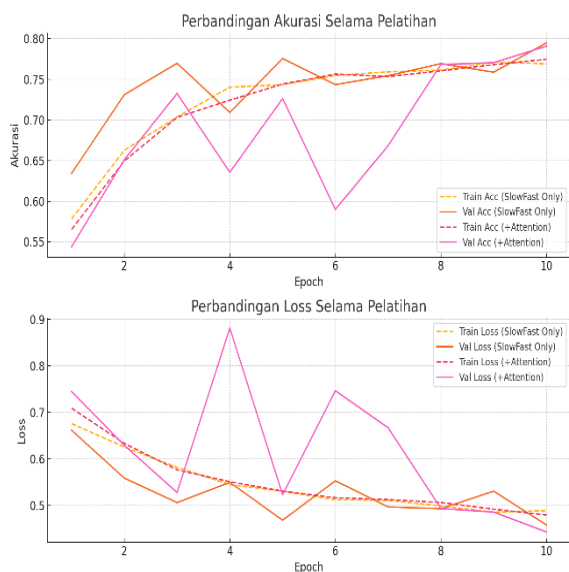
F. Evaluasi Model

Evaluasi model menggunakan data uji (test set) yang sebelumnya dipisahkan dan tidak digunakan selama pelatihan. Evaluasi dilakukan dengan menghitung berbagai metrik performa klasifikasi biner, yaitu akurasi, presisi, recall, dan F1-score. Metrik ini dihitung berdasarkan hasil prediksi model yang kemudian dibandingkan dengan label sebenarnya. Selain itu, hasil prediksi divisualisasikan dalam bentuk confusion matrix untuk melihat secara langsung jumlah kesalahan dan keberhasilan klasifikasi pada masing-masing kelas. Dengan menggunakan confusion matrix, dapat diamati sejauh mana model cenderung salah klasifikasi pada kelas tertentu, terutama pada kelas minoritas seperti kondisi kantuk. Evaluasi dilakukan terhadap dua model, yaitu model SlowFast tanpa Self-Attention, dan model SlowFast dengan Self-Attention.

III. HASIL DAN PEMBAHASAN

A. Hasil Penelitian Model

Proses pelatihan dilakukan selama 10 epoch pada dua varian model, yaitu SlowFast Network tanpa Self-Attention dan SlowFast Network dengan tambahan Self-Attention. Untuk setiap epoch, nilai akurasi dan loss dihitung pada data pelatihan dan validasi. Gambar 3 menunjukkan perbandingan nilai akurasi dan loss yang diperoleh kedua model selama proses pelatihan berlangsung.



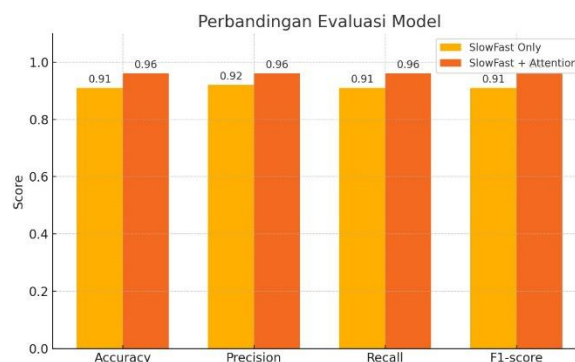
Gambar 3. Grafik Pelatihan Model
Sumber: diolah dari data primer

Dari grafik akurasi dapat diamati bahwa kedua model menunjukkan tren peningkatan akurasi yang stabil dari epoch pertama hingga terakhir. Model dengan Self-Attention secara umum memiliki nilai akurasi validasi yang lebih konsisten dan cenderung lebih tinggi pada akhir pelatihan.

Sementara itu, grafik loss menunjukkan bahwa model SlowFast + Attention mengalami penurunan loss yang lebih tajam dibandingkan model tanpa attention. Meskipun terdapat fluktuasi ringan pada beberapa epoch awal, secara keseluruhan nilai loss terus menurun, terutama pada data validas

B. Perbandingan Performa Dua Arsitekur

Nilai dari efektivitas arsitektur SlowFast dan pengaruh penambahan Self-Attention, dilakukan perbandingan performa kedua model berdasarkan beberapa metrik evaluasi: akurasi, presisi, recall, dan F1-score. Evaluasi dilakukan menggunakan subset data uji yang telah diseimbangkan secara proporsi (balanced sample), masing-masing terdiri dari 50 video drowsy dan 50 video not drowsy. Gambar 4 berikut memperlihatkan hasil evaluasi akhir dalam bentuk grafik batang



Gambar 4. Perbandingan Evaluasi Model Slowfast dan SlowFast + Attention

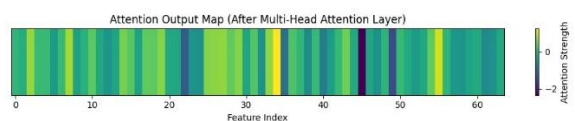
Sumber : diolah data dari data primer

Berdasarkan grafik tersebut, terlihat bahwa model SlowFast + Attention unggul pada semua metrik evaluasi. Model ini mencapai nilai akurasi sebesar 0.96 (96%), presisi 0.96 (96%), recall 0.96 (96%), dan F1-score 0.96 (96%), dibandingkan dengan model SlowFast saja yang memiliki skor rata-rata sekitar 0.91 (91%)–0.92(92%).

Peningkatan yang signifikan pada nilai recall dan F1-score menunjukkan bahwa Self-Attention membantu model untuk lebih sensitif dalam mendeteksi kondisi kantuk, tanpa mengorbankan presisi. Hal ini penting dalam konteks keselamatan berkendara, di mana deteksi yang cepat dan akurat terhadap perilaku berisiko sangat krusial. Dengan demikian, penambahan Self-Attention terbukti memberikan dampak positif terhadap performa model dalam klasifikasi perilaku pengemudi berdasarkan video input.

C. Visualisasi Attention Map

Visualisasi Attention Map berguna Untuk memahami bagaimana Self-Attention bekerja dalam model yang dikembangkan, dilakukan visualisasi attention map dari layer Multi-Head Self-Attention setelah proses pelatihan selesai. Visualisasi ini bertujuan untuk mengidentifikasi bagian fitur mana yang mendapatkan bobot perhatian lebih tinggi oleh model selama proses klasifikasi berlangsung. Gambar 5 menunjukkan salah satu hasil visualisasi attention map dari sampel video yang diuji menggunakan model SlowFast + Self-Attention:



Gambar 5. Visualisasi Attention Map pada Fitur Hasil Ekstraksi Model

Sumber: diolah dari data primer

Dari visualisasi tersebut terlihat bahwa bobot perhatian tersebar secara selektif pada wilayah tertentu dalam representasi fitur yang diproses. Area dengan intensitas warna yang lebih terang menunjukkan bagian fitur yang dianggap penting oleh model dalam pengambilan keputusan. Ini menunjukkan bahwa mekanisme self-attention bekerja dengan cara memfilter informasi visual dan memberikan prioritas pada pola temporal atau spasial yang relevan, seperti perubahan ekspresi wajah atau pergerakan kepala. Visualisasi ini mendukung klaim bahwa penambahan Self-Attention pada arsitektur SlowFast meningkatkan kemampuan model untuk mengidentifikasi pola perilaku kantuk secara lebih fokus dan efisien, tanpa tergantung pada fitur tunggal seperti mata atau mulut secara eksplisit.

IV. KESIMPULAN

Penelitian ini telah berhasil mengembangkan sistem deteksi perilaku pengemudi berbasis video dengan menggunakan arsitektur SlowFast Network yang dikombinasikan dengan mekanisme Self-Attention. Sistem ini dirancang untuk membedakan kondisi pengemudi antara kantuk dan tidak kantuk melalui analisis spasial dan temporal pada video berdurasi pendek. Dataset yang digunakan adalah SUST Driver Drowsiness, yang telah diproses melalui segmentasi dan normalisasi, serta dibagi menjadi data pelatihan, validasi, dan pengujian. Hasil pengujian menunjukkan bahwa model dengan Self-Attention mampu memberikan performa lebih baik dibandingkan model tanpa mekanisme tersebut, dengan nilai akurasi, recall, dan F1-score mencapai 96% pada pengujian data seimbang. Visualisasi dari attention map dan filter Conv3D mengindikasikan bahwa model mampu mengenali pola-pola visual penting terkait tanda-tanda kantuk, tanpa perlu deteksi eksplisit terhadap bagian wajah tertentu. Model ini masih memiliki keterbatasan, terutama pada ruang lingkup klasifikasi yang hanya mencakup dua kategori perilaku. Selain itu, sistem belum diuji dalam kondisi nyata di kendaraan. Penelitian selanjutnya dapat mempertimbangkan perluasan kategori perilaku dan pengujian langsung dalam lingkungan berkendara. Secara keseluruhan, model yang dikembangkan telah

menunjukkan kinerja yang baik dalam mendeteksi perilaku mengantuk berbasis video.

DAFTAR PUSTAKA

- [1]. Abdulrahman Abououf, Ibrahim Sobh, Mohammad Nasser, Omar Alsaqa, Omar Elezaby, & John F. W. Zaki. (2022). Multimodel System for Driver Distraction Detection and Elimination. (IEEE Access). <https://doi.org/10.1109/ACCESS.2022.3188715>
- [2]. Andi Asvin Mahersatillah Suradi, Samsu Alam, Mushaf, Muhammad Furqan Rasyid, Imran Djafar. (2023). Sistem Deteksi Kantuk Pengemudi Mobil Berdasarkan Analisis Rasio Mata Menggunakan Computer Vision. JUKI : Jurnal Komputer dan Informatika. <https://ioinformatic.org/index.php/JUKI/article/view/269>
- [3]. Balasubramani S, John Aravindhar D, P.N. Renjith, & K. Ramesh. (2024). DDSS: Driver decision support system based on the driver behaviour prediction to avoid accidents in intelligent transport system. International Journal of Cognitive Computing in Engineering. <https://doi.org/10.1016/j.ijcce.2023.12.001>
- [4]. Dawei Yang, Yan Wang, Ran Wei, Jiapeng Guan, Xiaohua Huang, Wei Cai, & Zhe Jiang. (2024). An efficient multi-task learning CNN for driver attention monitoring. Journal of Systems Architecture. ELSEVIER. <https://doi.org/10.1016/j.sysarc.2024.103085>
- [5]. Fangming Qu, Nolan Dang, Borko Furht & Mehrdad Nojournian. (2024). Comprehensive study of driver behavior monitoring systems using computer vision and machine learning techniques. Journal of Big Data. <https://doi.org/10.1186/s40537-024-00890-0>
- [6]. J Robert Theivadas & Suresh Ponnann. (2024). VigilEye: Machine learning-powered driver fatigue recognition for safer

- roads. Science direct, Measurement: Sensors.
<https://doi.org/10.1016/j.measen.2024.101186>
- [7]. Maykol Santosa, Paulo Jorge Coelho, Ivan Miguel Pires, Pedro Goncalves, & Goncalo Paiva Dias. (2024). An Overview of Machine Learning Algorithms to Reduce Driver Fatigue and Distraction-Related Traffic Accidents. Science direct, Procedia computer science.
<https://doi.org/10.1016/j.procs.2024.06.003>
- [8]. Md. Uzzol Hossain, Md. Aatur Rahman, Md. Manowarul Islam, Arnisha Akhter, Md. Ashraf Uddin, Bikash Kumar Paul. (2022). Automatic driver distraction detection using deep convolutional neural networks. Science direct, Intelligent Systems with Applications.
<https://doi.org/10.1016/j.iswa.2022.200075>
- [9]. Shahzeb Ansari, Fazel Naghdy, Haiping Du, & Yasmeen Naz Pahnwar. (2021). Driver Mental Fatigue Detection Based on Head Posture Using New Modified reLU-BiLSTM. IEEE Transaction on Intelligent Transportation System.
<https://doi.org/10.1109/TITS.2021.3098309>
- [10]. V. Uma Maheswari, Rajanikanth Aluvalu, Mvv Prasad Kantipudi, Krishna Keerthi Chennam, Ketan Kotecha, & Jatinderkumar R. Saini. (2022). Driver Drowsiness Prediction Based on Multiple Aspects Using Image Processing Techniques. IEEE Access.
<https://doi.org/10.1109/ACCESS.2022.3176451>
- [11]. Ward Ahmed Al-Hussein, Lip Yee Por, Miss Laiha Mat Kiah, & Bilal Bahaa Zaidan (2022). Driver Behavior Profiling and Recognition Using Deep-Learning Methods: In Accordance with Traffic Regulations and Experts Guidelines. International Journal of Environmental Research and Public Health. (MDPI).
<https://www.mdpi.com/1660-4601/19/3/1470>
- [12]. Zulqarnain H. Khattak, Wan Li, Thomas Karnowski, & Asad J. Khattak. (2024). The role of driver head pose dynamics and instantaneous driving in safety critical events: Application of computer vision in naturalistic driving. Accident Analysis and Prevention. ELSEVIER.
<https://doi.org/10.1016/j.aap.2024.107545>
- [13]. World Health Organization (WHO). (2023). Global Status Report on Road Safety
- [14]. Zulqarnain H. Khattak, Wan Li, Thomas Karnowski, & Asad J. Khattak. (2024). The role of driver head dynamics and instantaneous driving in safety critical events: Application of computer vision in naturalistic driving. Accident Analysis and Prevention. ELSEVIER