

DETEKSI KANTUK MENGGUNAKAN EAR DAN MAR BERBASIS DEEP LEARNING PADA CITRA SEKUENSIAL

1st Whisnumurty Galih Ananta
Informatika
Universitas Telkom
Surabaya, Indonesia
[whisnumurtyga@student.telkomunivers
ity.ac.id](mailto:whisnumurtyga@student.telkomunivers
ity.ac.id)

2nd Ahmad Wali Satria Bahari Johan
Informatika
Universitas Telkom
Surabaya, Indonesia
ahmadsatria@telkomuniversity.ac.id

3rd Pima Hani Safitri
Informatika
Universitas Telkom
Surabaya, Indonesia
phanisafitri@telkomuniversity.ac.id

Abstrak — Pada tahun 2023, kecelakaan lalu lintas mencapai angka tertinggi dalam lima tahun terakhir, dengan total 148.575 kasus, dimana 20% di antaranya disebabkan oleh kantuk, yang meningkatkan risiko kecelakaan hingga tiga kali lipat akibat penurunan kewaspadaan pengemudi. Penelitian ini mengusulkan metode ekstraksi fitur geometris *Eye Aspect Ratio* (EAR) dan *Mouth Aspect Ratio* (MAR) dari citra wajah sekuensial secara *real-time* menggunakan MediaPipe. EAR dan MAR dihitung berdasarkan koordinat *landmark* mata dan mulut, kemudian disusun dalam urutan temporal untuk menggambarkan perubahan kondisi subjek seiring waktu. Representasi ini efektif dalam menggambarkan transisi kantuk, yang dapat digunakan sebagai input dalam model deteksi berbasis *deep learning*. Penelitian ini melibatkan lima komponen utama: pengujian metode klasifikasi, pengolahan input data, perbaikan citra, augmentasi data, dan pipeline model. Data dari National Tsing Hua University Drowsiness Dataset (NTHU-DDD) dikelompokkan dalam window 60 *frame*, dengan fitur EAR dan MAR diekstraksi menggunakan MediaPipe. Hasil penelitian menunjukkan bahwa model CNN-LSTM efektif dalam memproses fitur EAR dan MAR secara sekuensial. Representasi penuh dengan input implisit (120, 1) memberikan performa terbaik, sementara teknik augmentasi SMOTE meningkatkan performa dengan menyeimbangkan distribusi kelas. Model CNN-LSTM-120FT tanpa perbaikan citra atau augmentasi menunjukkan performa paling stabil, dengan *accuracy* 85,59% dan *precision* 92,31%.

Kata kunci— kecelakaan lalu lintas, kantuk, EAR, MAR, *deep learning*, citra sekuensial

I. PENDAHULUAN

Pada tahun 2023, Indonesia mencatatkan jumlah kecelakaan lalu lintas tertinggi dalam lima tahun terakhir, dengan 148.575 kasus, lebih dari 94% di antaranya disebabkan oleh faktor manusia [1]. Kantuk merupakan salah satu penyebab utama, berkontribusi sebesar 20% terhadap kecelakaan tersebut dan meningkatkan risiko hingga tiga kali lipat [2], [3]. Kantuk, yang mengurangi kewaspadaan dan respons pengemudi akibat kelelahan atau durasi mengemudi yang panjang, menjadi faktor kritis yang meningkatkan kemungkinan kecelakaan, terutama di jalan tol [4].

Sebagai respons terhadap permasalahan tersebut, deteksi kantuk melalui analisis citra wajah pengemudi menggunakan fitur digital telah menjadi fokus utama dalam penelitian informatika. Dua parameter utama yang digunakan untuk mendeteksi kantuk adalah *Eye Aspect Ratio* (EAR) dan *Mouth Aspect Ratio* (MAR), yang masing-masing mengukur keterbukaan mata dan aktivitas menguap [5]. Meskipun kedua parameter ini terbukti efektif, pendekatan yang ada masih bergantung pada nilai ambang batas tetap yang kurang adaptif terhadap variasi kondisi individu dan lingkungan [6]. Untuk mengatasi keterbatasan ini, teknologi *deep learning*, khususnya *Recurrent Neural Network* (RNN), dapat digunakan untuk mengolah data sekuensial EAR dan MAR, yang memungkinkan deteksi kantuk lebih akurat dan adaptif [7].

Penelitian ini bertujuan mengembangkan metode deteksi kantuk yang lebih akurat dan adaptif dengan menggunakan fitur EAR dan MAR. Fitur ini akan diproses menggunakan *deep learning*, khususnya variasi RNN, untuk menanggulangi variasi kondisi lingkungan dan karakteristik individu. Citra sekuensial digunakan untuk menangkap perubahan aktivitas kantuk secara temporal. Dengan pendekatan ini, diharapkan dapat meningkatkan keandalan deteksi kantuk dan memberikan kontribusi pada pengembangan metode serupa.

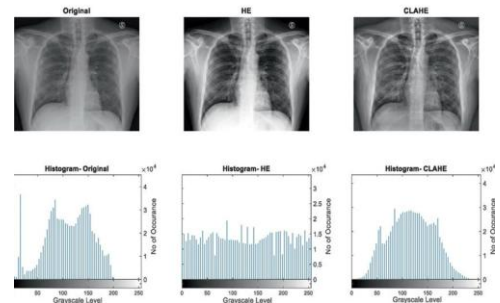
II. KAJIAN TEORI

A. Penelitian terdahulu

TABEL 1
(PENELITIAN TERDAHULU)

No	Judul	Perbedaan
1	Automatic Driver Drowsiness Detection Using Artificial Neural Network Based on Visual Facial Descriptors: Pilot Study [8]	Menggunakan perangkat invasif EEG berbasis <i>respiration</i> .

2	Non-Invasive Driver Drowsiness Detection System [3]	Tidak melibatkan fitur mata atau mulut. EAR dan MAR.
3	Driver's real-time Drowsiness Detection using Adaptable Eye Aspect Ratio and Smart Alarm System [9]	Hanya menggunakan pendekatan EAR serta MAR dengan nilai ambang batas statis.
2	A CNN-Based Approach for Driver Drowsiness Detection by Real-Time Eye State Identification [10]	Hanya menggunakan fitur mata, dan klasifikasi dilakukan per <i>frame</i> tunggal,
3	DrowsyNet: Multivariate Time Series Classification for Embedded Driver Drowsiness Detection [11]	Penggunaan konsep kantuk sebagai aktivitas sekuensial tanpa memanfaatkan fitur EAR dan MAR.



GAMBAR 2
(CARA KERJA HE DAN CLAHE)

HE mendistribusikan intensitas piksel secara merata pada histogram citra, menghasilkan peningkatan kontras global. Sementara itu, CLAHE, yang merupakan pengembangan dari HE, meningkatkan kontras secara lokal dengan membatasi amplifikasi intensitas, sehingga mengurangi *noise* berlebih. Teknik ini lebih efektif pada citra dengan detail halus dan kontras rendah [13].

B. Kantuk

Kantuk merupakan kondisi fisiologis yang ditandai dengan penurunan kewaspadaan serta kemampuan untuk mempertahankan perhatian, yang umumnya disebabkan oleh kurang tidur atau kelelahan. Kondisi ini tidak hanya memengaruhi kewaspadaan, tetapi juga berimplikasi pada penurunan fungsi kognitif, seperti proses pengambilan keputusan, memori, dan konsentrasi, yang pada gilirannya dapat menurunkan kinerja mental dan fisik individu [2], [4].

C. Citra Sekuensial

Citra sekuensial merujuk pada serangkaian gambar yang diambil dalam urutan waktu untuk merekam perubahan atau peristiwa secara bertahap. Karakteristik utamanya adalah kemampuan untuk merekam dinamika waktu, seperti gerakan atau pola aktivitas yang umumnya diperoleh dari *frame-by-frame* dari video. Integrasi pola spasial dan temporal dapat menghasilkan model deteksi kantuk dengan performa lebih baik [12].



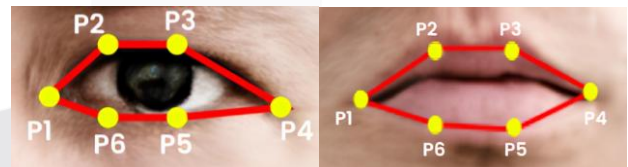
GAMBAR 1
(GAMBAR CITRA SEKUENSIAL)

D. Contrast Enhancement

Contrast enhancement adalah teknik pemrosesan citra yang bertujuan meningkatkan kualitas visual dengan memperbesar perbedaan intensitas antara area terang dan gelap pada citra. Dua metode yang sering digunakan dalam *contrast enhancement* adalah *Histogram Equalization* (HE) dan *Contrast Limited Adaptive Histogram Equalization* (CLAHE).

E. EAR dan MAR

Deteksi kantuk dapat dilakukan dengan menggunakan teknologi deteksi *landmark* wajah untuk memantau pergerakan mata dan mulut. *Eye Aspect Ratio* (EAR) mengukur sejauh mana mata tertutup, yang menunjukkan tingkat kantuk atau kelelahan, dihitung berdasarkan posisi *landmark* di sekitar mata [14]. Penurunan nilai EAR mengindikasikan mata yang menutup, gejala kantuk yang ditunjukkan dengan berkedip atau menutup mata berulang kali. *Mouth Aspect Ratio* (MAR) mengukur sejauh mana mulut terbuka, yang juga menunjukkan tingkat kantuk atau kelelahan, dihitung berdasarkan posisi *landmark* di sekitar mulut. Peningkatan nilai MAR mengindikasikan mulut yang lebih lebar terbuka, yang menunjukkan kantuk, ditandai dengan menguap atau membuka mulut sebagai respons terhadap kelelahan [15], [9].



GAMBAR 3
(VISUALISASI EAR DAN MAR)

$$EAR = MAR = \frac{\|P2 - P6\| + \|P3 - P5\|}{2\|P1 - P4\|} \quad (1)$$

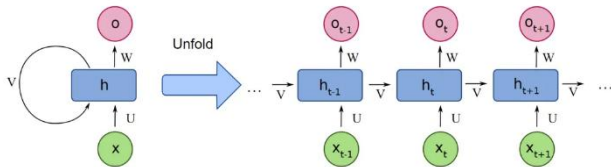
dimana:

- P_1 = Titik pada sudut mata/ mulut bagian dalam.
- P_2 = Titik pada ujung atas, di dekat sudut bagian dalam.
- P_3 = Titik pada ujung atas, di dekat sudut bagian luar.
- P_4 = Titik pada sudut bagian luar.
- P_5 = Titik pada ujung bawah, di dekat sudut bagian luar.
- P_6 = Titik pada ujung bawah, di dekat sudut bagian dalam.

F. Recurrent Network

Penelitian ini mengoptimalkan deteksi kantuk menggunakan *deep learning* dengan jaringan saraf berlapis *Deep Neural Networks* (DNN), yang meniru cara otak

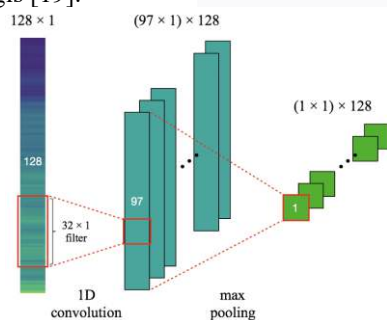
manusia dalam membuat keputusan kompleks [16]. Model yang digunakan berbasis *Recurrent Neural Network* (RNN), yang efektif dalam menangkap pola data sekuensial dengan mekanisme umpan balik untuk mempertahankan informasi dari langkah sebelumnya [17]. Namun, RNN kesulitan menangani ketergantungan jangka panjang akibat masalah *vanishing gradient*. Untuk mengatasi hal ini, *Long Short-Term Memory* (LSTM), varian RNN dengan mekanisme gerbang, memungkinkan pemrosesan informasi lebih lama, sementara *Bidirectional LSTM* (Bi-LSTM) memproses data sekuensial dari dua arah [18].



GAMBAR 4
(CARA KERJA RNN)

G. Convolutional Neural Network (CNN)

Dalam sistem deteksi kantuk yang menggunakan fitur fisiologis seperti EAR dan MAR, CNN 1D dapat digunakan untuk mengenali perubahan nilai fitur seiring waktu dengan menganalisis pola lokal pada data sekuensial. Dengan operasi konvolusi, CNN mendeteksi pola perubahan berulang, kestabilan, atau perubahan mendadak pada data, serta mengenali tanda-tanda awal kantuk seperti perubahan gerakan mata atau mulut, menjadikannya komponen penting dalam klasifikasi tingkat kantuk berdasarkan analisis urutan data fisiologis [19].



GAMBAR 5
(CARA KERJA CNN)

H. Attention Mechanism

Attention mechanism digunakan dalam pengolahan data sekuensial untuk memberikan bobot perhatian yang berbeda pada setiap langkah waktu berdasarkan relevansinya, memungkinkan model fokus pada segmen data penting, seperti penutupan mata atau peningkatan menguap yang mengindikasikan kantuk, sebagai alternatif untuk CNN dalam mendeteksi pola yang relevan [20].

I. Hybrid Model

Penggunaan satu jenis model dalam sistem klasifikasi sering kali tidak cukup untuk menangkap kompleksitas pola data, karena setiap algoritma memiliki keunggulan dan keterbatasan tertentu. Oleh karena itu, pendekatan *hybrid model* digunakan untuk menggabungkan beberapa model, memanfaatkan keunggulan masing-masing, dan meningkatkan representasi fitur, akurasi klasifikasi, serta kemampuan generalisasi model terhadap variasi data [21].

J. Augmentasi Data

Dalam *deep learning*, ketersediaan data yang memadai sangat penting untuk membangun model klasifikasi yang efektif dan dapat menggeneralisasi data baru. Untuk mengatasi keterbatasan jumlah dan keragaman data, augmentasi data digunakan untuk memperluas jumlah dan diversitas data pelatihan [22]. SMOTE, mengatasi ketidakseimbangan kelas melalui interpolasi sampel minoritas, namun memiliki keterbatasan pada data sekuensial karena mengabaikan ketergantungan temporal, sementara *Gaussian Feature Perturbation* (GFP) menambah keragaman data dengan menambahkan *noise Gaussian* pada nilai fitur setiap *timestep* tanpa mengganggu kontinuitas temporal. Sementara itu, *subsequence augmentation*, memanipulasi urutan waktu data sekuensial dengan mengambil segmen tertentu untuk memperbesar sampel pelatihan.

K. Evaluasi Performa Model

Evaluasi sistem deteksi kantuk dilakukan secara sistematis dengan menggunakan *confusion matrix*, yang menggambarkan performa model melalui distribusi prediksi yang mencakup *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN) (Lovell et al., 2022). Penggunaan *confusion matrix* memungkinkan analisis performa model melalui metrik evaluasi seperti *accuracy*, *precision*, *recall*, dan *F1-score* [23].

L. Cross-Validation dan Hold-Out Validation

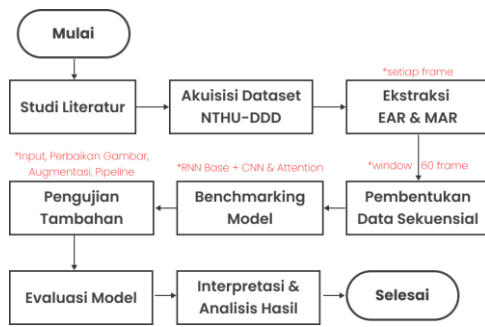
Setelah menentukan metrik evaluasi, langkah selanjutnya adalah memastikan evaluasi dilakukan secara adil dan representatif melalui strategi validasi data yang sesuai. *Cross-validation*, khususnya *k-fold cross-validation*, adalah teknik umum dalam *deep learning* untuk mengukur kemampuan generalisasi model dengan membagi data latih menjadi beberapa *fold* dan menguji model pada setiap *fold* secara bergiliran [24]. Sementara itu, *hold-out validation* membagi dataset menjadi dua subset, dengan proporsi 90% untuk pelatihan dan 10% untuk pengujian, namun metode ini lebih cocok untuk dataset besar dan representatif karena hasil evaluasi sangat bergantung pada cara pembagian data, yang dapat menyebabkan variabilitas tinggi dalam estimasi performa model [25].

III. METODE

A. Rancangan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan desain eksperimen untuk mengembangkan dan membandingkan berbagai arsitektur model berbasis RNN dalam mendeteksi kantuk. Variasi *hybrid model* juga digunakan, seperti kombinasi penggunaan CNN dan *Attention*, guna meningkatkan kemampuan model dalam menangkap pola lokal maupun temporal.

B. Prosedur Penelitian



GAMBAR 6
(PROSEDUR PENELITIAN)

Penelitian dimulai dari studi literatur, dilanjutkan akuisisi dataset NTHU-DDD dan ekstraksi fitur EAR & MAR dari setiap *frame*. Data disusun dalam *window* sekuensial dan digunakan untuk *benchmarking* 7 model berbasis RNN serta kombinasi CNN dan *attention*. Pengujian tambahan dilakukan pada variasi input, augmentasi, dan perbaikan gambar. Evaluasi dilakukan dengan *hold-out* dan *cross-validation* untuk memperoleh model paling andal.

C. Sumber Data

Sumber data yang digunakan adalah National Tsing Hua University Driver Drowsiness Dataset (NTHU-DDD) yang terpublikasi pada platform Kaggle. Dataset ini terdiri atas 66.520 *frame* yang telah dilabeli menjadi dua kelas: *drowsy* (36.030 *frame*) dan *not drowsy* (30.490 *frame*).

D. Metode Pengumpulan Data

Setiap *frame* diproses menggunakan MediaPipe untuk mendeteksi *landmark* wajah, kemudian dihitung nilai EAR dan MAR dari titik-titik pada mata dan mulut. Hasilnya disusun dalam bentuk urutan waktu sekuensial sepanjang 60 *frame*.

E. Metode Analisis Data

Analisis data dilakukan secara eksperimental terhadap hasil pengujian. Terdapat 5 tahapan pengujian, meliputi Model Klasifikasi, Input Data, Perbaikan Gambar, Augmentasi, *Pipeline* Model. Analisis data dilakukan dengan mengevaluasi performa model melalui metrik evaluasi yang dihasilkan dari *confusion matrix* hasil prediksi dengan mempertimbangkan *accuracy*, *precision*, *recall*, dan *f1-score*.

F. Pengujian Metode Klasifikasi

Model RNN digunakan sebagai *baseline* model, namun pendekatan *hybrid model* juga diadopsi dengan menambahkan CNN atau Attention untuk menyeimbangkan deteksi pola sekuensial dan lokal. CNN-LSTM mengalirkan output CNN ke LSTM [26], sedangkan LSTM-Attention memproses data dengan LSTM sebelum diberi bobot atensi [27].

TABEL 2
(SKEMA PENGUJIAN METODE KLASIFIKASI)

No	Arsitektur Model
1	RNN
2	LSTM
3	BiLSTM
4	CNN-LSTM

5	LSTM-Attention
6	CNN-BiLSTM
7	BiLSTM-Attention

G. Pengujian Metode Input Data

Variasi representasi input dapat memengaruhi kemampuan model dalam mengenali pola sekuensial. Oleh karena itu, dilakukan pengujian terhadap berbagai skema penyusunan fitur EAR dan MAR, baik terpisah, gabungan, maupun urutan panjang dengan atau tanpa penanda eksplisit menggunakan arsitektur model terbaik dari tahap sebelumnya.

TABEL 3
(SKEMA PENGUJIAN METODE INPUT DATA)

No	Metode Input	Input
1	Pasangan EAR & MAR	(60, 2)
2	Hanya EAR	(60, 1)
3	Hanya MAR	(60, 1)
4	Urutan EAR-MAR Implisit (120 FT)	(120, 1)
5	Urutan EAR-MAR Eksplisit (Concat)	(120, 1)

H. Pengujian Metode Perbaikan Gambar

Kualitas visual citra wajah memengaruhi *accuracy* deteksi kantuk, terutama dalam pencahayaan rendah. Untuk itu, diuji dua metode *contrast enhancement*, *Histogram Equalization* (HE) dan *Contrast Limited Adaptive Histogram Equalization* (CLAHE) untuk menilai dampaknya terhadap performa klasifikasi. Sebagai pembandingan, pengujian juga dilakukan pada citra asli tanpa perbaikan.

TABEL 4
(SKEMA PENGUJIAN METODE PERBAIKAN GAMBAR)

No	Metode
1	Tanpa Perbaikan Gambar
2	HE
3	CLAHE

I. Pengujian Metode Augmentasi

Pengujian augmentasi dilakukan untuk menilai dampak peningkatan variasi data sekuensial terhadap performa deteksi kantuk. Augmentasi diterapkan pada fitur EAR dan MAR tanpa mengubah struktur semantik. Dua jenis augmentasi diuji: *temporal* (*Subsequence Augmentation*) dan *feature space*, yaitu *Synthetic Minority Over-sampling Technique* (SMOTE) dan *Gaussian Feature Perturbation* (GFP). SMOTE menghasilkan sampel sintetik, sedangkan GFP menambahkan *noise* untuk meningkatkan keragaman lokal.

TABEL 5
(SKEMA PENGUJIAN METODE AUGMENTASI)

No	Metode	Jumlah Data	<i>Drowsy</i>	<i>Not Drowsy</i>
1	Tanpa Augmentasi	1.108	601	507
2	<i>Subsequence Augmentation</i>	1.500	750	750
3	SMOTE	1.500	750	750
4	GFP	1.500	750	750

J. Pengujian Pipeline Model

Pengujian *pipeline* model dilakukan untuk mengevaluasi kinerja sistem secara menyeluruh dari 51 kombinasi

representasi input, perbaikan citra, dan augmentasi, guna mengidentifikasi konfigurasi optimal dari interaksi antar komponen yang telah diuji secara terpisah.

TABEL 6
(SKEMA PENGUJIAN PIPELINE MODEL)

No	Perbaikan Gambar	Input	Augmentasi
1	-Tanpa	(60, 2)	Tanpa Augmentasi
2	Perbaikan - HE - CLAHE	120FT	Tanpa Augmentasi
		CONCAT	Subsequence
		EAR	GFP
		MAR	SMOTE

K. Metode Pelatihan dan Evaluasi

Proses pelatihan dan pengujian model dilakukan dengan membagi dataset menjadi 90% data latih dan 10% data uji. Evaluasi performa dilakukan melalui kombinasi *hold-out validation* untuk mengukur generalisasi model secara umum, serta *cross-validation* guna mengidentifikasi model yang paling andal secara konsisten.

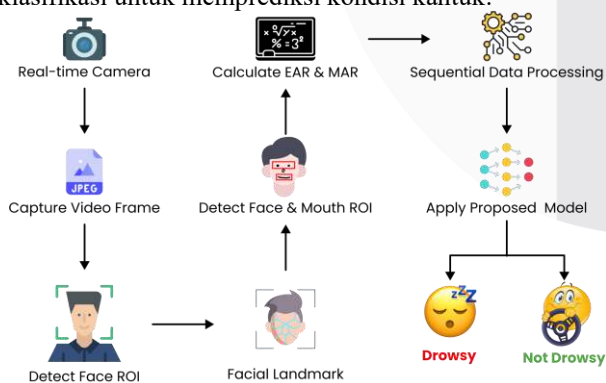
TABEL 7
(RINGKASAN KESELURUHAN PENGUJIAN DAN EVALUASI)

No	Metode Pengujian	Jenis Evaluasi
1	Klasifikasi	<i>Cross-Validation</i>
2	Input Data	<i>Hold-Out Validation</i>
3	Perbaikan Gambar	<i>Hold-Out Validation</i>
4	Augmentasi	<i>Hold-Out Validation</i>
5	Pipeline Model	<i>Hold-Out Validation, Cross-Validation</i>

IV. HASIL DAN PEMBAHASAN

A. Implementasi Arsitektur Sistem

Sistem deteksi kantuk dalam penelitian ini terdiri atas tahapan akuisisi citra wajah secara *real-time* menggunakan kamera, dilanjutkan dengan deteksi area wajah dan identifikasi *landmark* mata dan mulut. Berdasarkan koordinat *landmark* tersebut, dihitung nilai EAR dan MAR yang kemudian disusun dalam bentuk sekuensial untuk merepresentasikan perubahan kondisi subjek secara temporal. Data sekuensial ini digunakan sebagai input model klasifikasi untuk memprediksi kondisi kantuk.

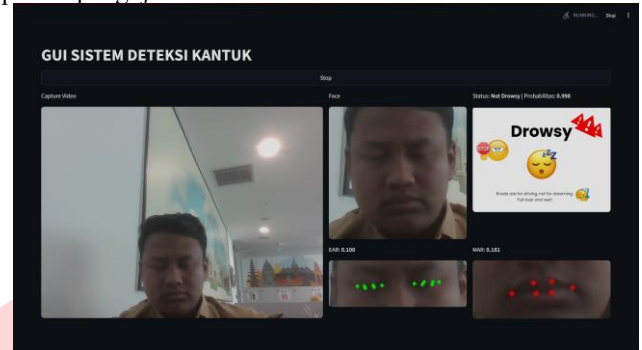


GAMBAR 7
(ARSITEKTUR SISTEM DETEKSI KANTUK)

B. Implementasi Graphical User Interface (GUI)

GUI pada sistem deteksi kantuk dirancang secara minimalis namun informatif untuk mendukung kemudahan interaksi. Pengguna cukup menekan tombol *Start/Run Detection* untuk memulai, dan sistem akan secara otomatis melakukan akuisisi *frame* video serta menjalankan proses

deteksi kantuk secara *real-time* di latar belakang. Tampilan utama menampilkan citra wajah subjek di sisi kiri, area deteksi wajah beserta ROI mata dan mulut di sisi kanan, serta hasil klasifikasi (*Drowsy* atau *Not Drowsy*) dengan probabilitas prediksi. Seluruh elemen disusun untuk memberikan umpan balik visual yang komprehensif selama proses pengujian.



GAMBAR 8
(GUI SISTEM DETEKSI KANTUK)

C. Hasil Pengujian Metode Klasifikasi

Pengujian metode klasifikasi dilakukan menggunakan skema *k-fold cross-validation* sebanyak 5 *fold* untuk memperoleh estimasi performa model yang lebih stabil serta meminimalkan risiko *overfitting*.

TABEL 8
(HASIL PENGUJIAN METODE KLASIFIKASI)

Arsitektur	Acc	Prec	Rec	F1
CNN-LSTM	80,44%	81,00%	85,40%	82,36%
CNN-BiLSTM	79,10%	84,49%	77,01%	79,91%
LSTM-Attention	72,02%	78,04%	71,41%	73,56%
LSTM	72,35%	79,88%	69,00%	73,02%
BiLSTM	72,95%	82,11%	65,32%	72,33%
BiLSTM-Attention	70,88%	78,71%	65,98%	70,78%
RNN	66,57%	74,31%	66,46%	67,21%

Model CNN-LSTM menunjukkan performa terbaik dengan *accuracy* 80,44% dan *F1-score* 82,36%, didukung oleh *precision* 81,00% dan *recall* 85,40%. Hal ini menunjukkan efektivitas CNN dalam mengekstraksi pola lokal dari urutan EAR dan MAR, yang kemudian diproses temporal oleh LSTM untuk menangkap ketergantungan jangka panjang pada transisi kondisi kantuk. Sebaliknya, model lain seperti CNN-BiLSTM (*accuracy* 79,10%, *F1-score* 79,91%) menunjukkan *trade-off* antara *precision* yang lebih tinggi (84,49%) dan *recall* yang lebih rendah (77,01%). Model LSTM-Attention, LSTM, BiLSTM, dan BiLSTM-Attention memiliki *accuracy* 70-72% dan penurunan *F1-score*, sementara model RNN (*accuracy* 66,57%) kesulitan dalam mempelajari dependensi jangka panjang.

D. Hasil Pengujian Metode Input Data

TABEL 9
(HASIL PENGUJIAN METODE INPUT DATA)

Model	Acc	Prec	Rec	F1
CNN-LSTM	84,68%	97,00%	73,33%	83,81%
CNN-LSTM-120FT	90,99%	90,32%	93,33%	91,80%
CNN-LSTM-CONCAT	81,98%	77,03%	95,00%	85,07%
CNN-LSTM-EAR	81,98%	75,00%	100,00%	85,71%
CNN-LSTM-MAR	87,39%	85,94%	91,67%	88,71%

Pengujian variasi representasi input menunjukkan bahwa skema penggabungan sekuensial implisit (CNN-LSTM-120FT) menghasilkan performa terbaik dengan *accuracy* 90,99% dan *F1-score* 91,80%. Kombinasi *precision* 90,32% dan *recall* 93,33% mengindikasikan bahwa representasi input sekuensial panjang memungkinkan model mengoptimalkan hubungan temporal antar fitur EAR dan MAR. Sebaliknya, skema penggabungan paralel (CNN-LSTM) mencatatkan *accuracy* 84,68% dengan *precision* 97,00%, namun *recall* rendah 73,33%, sementara skema dengan penyandian eksplisit (CNN-LSTM-CONCAT) menghasilkan akurasi 81,98% dan *F1-score* 85,07%, dengan *recall* tinggi (95,00%) tetapi *precision* turun (77,03%).

E. Hasil Pengujian Metode Perbaikan Gambar

TABEL 10
(HASIL PENGUJIAN METODE PERBAIKAN GAMBAR)

Metode	Acc	Prec	Recall	F1
Tanpa Perbaikan	90,99%	90,32%	93,33%	91,80%
HE	89,19%	90,00%	90,00%	90,00%
CLAHE	85,59%	82,35%	93,33%	87,50%

Pengujian menunjukkan bahwa model baseline tanpa perbaikan citra menghasilkan performa klasifikasi tertinggi dengan *accuracy* 90,99% dan *F1-score* 91,80%. Keseimbangan *precision* 90,32% dan *recall* 93,33% menunjukkan bahwa kualitas citra asli cukup stabil dalam menjaga konsistensi deteksi landmark EAR dan MAR, yang memungkinkan pembentukan representasi temporal yang optimal dalam pembelajaran model.

Sebaliknya, penerapan HE menurunkan performa dengan *accuracy* 89,19% dan *F1-score* 90,00%. Meskipun *precision* dan *recall* tetap seimbang pada angka 90,00%, redistribusi kontras secara global dapat menyebabkan keseragaman intensitas piksel, mengurangi ketepatan representasi fitur EAR dan MAR. Penurunan kinerja lebih signifikan pada CLAHE, dengan *accuracy* turun menjadi 85,59% dan *F1-score* 87,50%. Meskipun *recall* tetap tinggi (93,33%), penurunan *precision* (82,35%) mengindikasikan peningkatan kesalahan *false positive* akibat distorsi lokal dari CLAHE.

F. Hasil Pengujian Metode Augmentasi

TABEL 11
(HASIL PENGUJIAN METODE AUGMENTASI)

Metode	Acc	Prec	Rec	F1
Tanpa Augmentasi	90,99%	90,32%	93,33%	91,80%
Subsequence Augmentation	84,67%	77,66%	97,33%	86,39%
SMOTE	93,33%	92,21%	94,67%	93,42%
GFP	83,33%	91,67%	73,33%	81,48%

SMOTE menunjukkan performa terbaik di antara seluruh strategi augmentasi yang diuji, dengan *accuracy* 93,33% dan *F1-score* 93,42%. Peningkatan ini mencerminkan keberhasilan sintesis data minoritas secara numerik yang mengurangi bias terhadap kelas mayoritas dan mempertahankan struktur temporal sekuens fitur. *Precision* 92,21% dan *recall* 94,67% menunjukkan bahwa model mampu mendeteksi kondisi kantung secara andal tanpa menghasilkan *false detection* yang berlebihan. Keberhasilan SMOTE menegaskan bahwa augmentasi berbasis rekonstruksi distribusi antar kelas lebih efektif dibandingkan pendekatan berbasis gangguan fitur lokal atau segmentasi temporal.

Sebaliknya, metode GFP dan *Subsequence Augmentation* menunjukkan penurunan performa. GFP menghasilkan *precision* tinggi (91,67%) namun *recall* rendah (73,33%), menunjukkan *under-detection*, sementara *Subsequence Augmentation* memiliki *recall* tertinggi (97,33%) namun *precision* rendah (77,66%), mencerminkan *over-detection*. Ketidakseimbangan antara *precision* dan *recall* ini menunjukkan bahwa augmentasi tidak selalu meningkatkan performa klasifikasi.

G. Hasil Pengujian Pipeline Model

TABEL 12
(HASIL PENGUJIAN PIPELINE MODEL HOLD-OUT)

Nama Model	Acc	Prec	Rec	F1
CNN-LSTM	84,7%	97,0%	73,3%	83,8%
CNN-LSTM-120FT	88,1%	88,0%	89,7%	88,3%
CNN-LSTM-CONCAT	84,0%	84,3%	85,1%	84,3%
CNN-LSTM-EAR	81,8%	79,7%	87,0%	82,8%
CNN-LSTM-MAR	80,2%	74,0%	95,9%	83,2%
CNN-LSTM-SUBSEQ	79,3%	76,8%	86,0%	80,4%
CNN-LSTM-GFP	81,8%	82,5%	83,7%	82,2%
CNN-LSTM-SMOTE	87,3%	84,6%	93,0%	88,2%
HE-CNN-LSTM	87,4%	83,8%	95,0%	89,1%
HE-CNN-LSTM-120FT	88,3%	87,1%	92,2%	89,1%
HE-CNN-LSTM-CONCAT	87,2%	90,2%	92,2%	86,8%
HE-CNN-LSTM-EAR	82,3%	83,8%	82,9%	82,7%
HE-CNN-LSTM-MAR	74,1%	69,7%	87,4%	77,3%
HE-CNN-LSTM-SUBSEQ	77,8%	76,1%	83,7%	78,9%

HE-CNN-LSTM-GFP	82,7%	82,1%	86,3%	83,5%
HE-CNN-LSTM-SMOTE	87,0%	84,8%	91,7%	87,8%
CLAHE-CNN-LSTM	80,2%	73,2%	100,0%	84,5%
CLAHE-CNN-LSTM-120FT	88,9%	89,0%	89,7%	89,2%
CLAHE-CNN-LSTM-CONCAT	88,7%	88,6%	90,1%	89,0%
CLAHE-CNN-LSTM-EAR	82,8%	81,2%	87,7%	83,6%
CLAHE-CNN-LSTM-MAR	79,0%	74,4%	92,4%	79,4%
CLAHE-CNN-LSTM-SUBSEQ	79,5%	81,6%	79,0%	79,4%
CLAHE-CNN-LSTM-GFP	87,0%	84,4%	92,3%	87,9%
CLAHE-CNN-LSTM-SMOTE	91,5%	90,0%	94,3%	91,9%

Konfigurasi CLAHE-CNN-LSTM-SMOTE memberikan hasil terbaik di seluruh metrik, dengan *accuracy* 91,50% dan *F1-score* 91,92%, serta *precision* 90,01% dan *recall* 94,33% yang seimbang. Penggunaan peningkatan kontras lokal (CLAHE) membantu mempertajam fitur visual pada citra wajah, sementara representasi input berbasis sekuens dan augmentasi SMOTE memperkaya distribusi data tanpa mengganggu keseimbangan kelas. Konsistensi metrik evaluasi menunjukkan bahwa konfigurasi ini efektif mendeteksi kantuk dengan tingkat kesalahan minimum. Beberapa konfigurasi lain, seperti CLAHE-CNN-LSTM-120FT (*F1-score* 89,24%) dan HE-CNN-LSTM-120FT (89,13%), juga menunjukkan performa tinggi, meskipun sedikit lebih rendah dari konfigurasi dengan SMOTE.

Namun, beberapa konfigurasi dengan performa rendah, seperti CLAHE-CNN-LSTM-SUBSEQ (*F1-score* 79,44%), menunjukkan bahwa augmentasi berbasis subsekuens berpotensi mengganggu pola temporal jika tidak diimbangi struktur data yang stabil. Model dengan input tunggal seperti HE-CNN-LSTM-MAR (77,29%) dan CLAHE-CNN-LSTM-MAR (82,09%) juga menunjukkan penurunan performa, yang dapat dikaitkan dengan keterbatasan informasi dari fitur MAR.

TABEL 13
(HASIL PENGUJIAN PIPELINE MODEL CROSS-VALIDATION)

Model	Acc	Prec	Rec	F1
Hold-Out Validation				
CLAHE-CNN-LSTM-120FT-SMOTE	96,00%	97,26%	94,67%	95,95%
CLAHE-CNN-LSTM-120FT-GFP	92,00%	93,15%	90,67%	91,89%
CNN-LSTM-120FT	90,99%	90,32%	93,33%	91,80%
CNN-LSTM-120FT-SUBSEQ	84,67%	77,66%	97,33%	86,39%
Cross-Validation				

CLAHE-CNN-LSTM-120FT-SMOTE	82,00%	77,91%	89,33%	83,23%
CLAHE-CNN-LSTM-120FT-GFP	80,67%	85,94%	73,33%	79,14%
CNN-LSTM-120FT	85,59%	92,31%	80,00%	85,71%
CNN-LSTM-120FT-SUBSEQ	81,33%	88,52%	72,00%	79,41%

Model CNN-LSTM menunjukkan performa stabil pada skema *k-fold cross-validation*, dengan *accuracy* 85,59% dan *F1-Score* 85,71%, tertinggi dibandingkan model-model lain yang diuji. Meskipun tidak menggunakan teknik augmentasi atau praproses tambahan seperti CLAHE atau SMOTE, hasil ini menunjukkan bahwa arsitektur CNN-LSTM efektif memanfaatkan informasi sekuensial melalui input implisit (120,1). Nilai *precision* yang tinggi (92,31%) mengindikasikan bahwa model ini memiliki tingkat kesalahan prediksi positif yang rendah, sehingga cocok digunakan pada sistem deteksi kantuk yang memerlukan ketelitian tinggi.

Sebagian besar model yang menggunakan metode augmentasi mengalami penurunan performa saat diuji menggunakan *cross-validation*. Model seperti CLAHE-CNN-LSTM-120FT-SMOTE dan CLAHE-CNN-LSTM-120FT-GFP mengalami penurunan *F1-Score* lebih dari 12 poin dibandingkan hasil pada skema hold-out, menunjukkan bahwa beberapa metode augmentasi menyebabkan *overfitting*. Namun, model CNN-LSTM-120FT-SUBSEQ menunjukkan penurunan performa terkecil, sekitar 6,98 poin, menandakan bahwa augmentasi berbasis subsekuens lebih stabil.

H. Implikasi dan Pengembangan

Penelitian ini memiliki implikasi penting dalam pengembangan sistem deteksi kantuk berbasis pengolahan citra, yang dapat diterapkan pada sistem pemantauan pengemudi untuk meningkatkan keselamatan berkendara. Model CNN-LSTM yang menggabungkan fitur EAR dan MAR terbukti efektif dalam mendeteksi kantuk secara akurat dengan mempertimbangkan pola sekuensial dan temporal. Pengembangan lebih lanjut dapat difokuskan pada peningkatan generalisasi model melalui eksplorasi arsitektur yang lebih kompleks, teknik augmentasi data yang lebih canggih, dan pengujian pada dataset yang lebih beragam, termasuk variasi etnis dan kondisi pencahayaan, untuk meningkatkan ketahanan sistem dalam kondisi nyata.

V. KESIMPULAN

Kesimpulan dari penelitian ini menunjukkan bahwa model CNN-LSTM dengan representasi input fitur EAR dan MAR dalam bentuk sekuensial (CNN-LSTM-120FT) tanpa penggunaan perbaikan citra atau teknik augmentasi memberikan performa yang paling stabil dan andal. Meskipun teknik augmentasi seperti SMOTE dan perbaikan citra dengan CLAHE dapat memberikan peningkatan performa dalam beberapa skenario, penggunaan input data dalam bentuk sekuensial implisit terbukti menghasilkan hasil yang lebih optimal, dengan penurunan performa yang

minimal (5,4%) pada skema *cross-validation* dan akurasi 85,59%. Dengan demikian, pendekatan ini dapat dianggap sebagai konfigurasi terbaik untuk deteksi kantuk secara otomatis pada pengemudi.

REFERENSI

- [1] A. D. Ayuningtyas, "2023 Capai Angka Tertinggi Kecelakaan Lalu Lintas 5 Tahun Terakhir," [goodstats.id](https://goodstats.id/article/2023-capai-angka-tertinggi-kecelakaan-lalu-lintas-5-tahun-terakhir-z67bk). [Online]. Available: <https://goodstats.id/article/2023-capai-angka-tertinggi-kecelakaan-lalu-lintas-5-tahun-terakhir-z67bk>
- [2] A. Moradi, S. S. H. Nazari, and K. Rahmani, "Sleepiness and the risk of road traffic accidents: A systematic review and meta-analysis of previous studies," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 65. pp. 620–629, 2019. doi: 10.1016/j.trf.2018.09.013.
- [3] H. U. R. Siddiqui *et al.*, "Non-Invasive Driver Drowsiness Detection System." mdpi.com, 2021.
- [4] D. Lestari, "KNKT: 80 Persen Kecelakaan di Tol Akibat Mengantuk dan Letih." [Online]. Available: Siddiqui, Hafeez Ur Rehman; Saleem, Adil Ali; Bbowrn, Robert; Bademci, Bahattin; Lee, Enesto; Rustam, Furqan; Dudley, Sandra
- [5] D. C. S. Ram, D. J. V. . Koushik, and H. S. Pavan, "Drowsiness Detection using EAR (Eye Aspect Ratio) by Machine Learning," *Interantional Journal of Scientific Research in Engineering and Management*, vol. 08, no. 01. pp. 1–13, 2024. doi: 10.55041/ijrsrem19675.
- [6] L.Thulasimani, Poojeevan P, Prithashasni S P, "Real Time Driver Drowsiness Detection using OpenCV and Facial Landmarks," *Journal of Contemporary Issues in Business and Government*, vol. 27, no. 6. 2021. doi: 10.47750/cibg.2021.27.06.054.
- [7] S. Plubin, W. Bunyatisai, M. Khamkong, T. Mouktonglang, and B. Plubin, "Multilabel-Thai Text Classification with Transformer-Rnn in Thai Banking Classification," *Pakistan Journal of Life and Social Sciences*, vol. 23, no. 1. pp. 1–10, 2024. doi: 10.57239/PJLSS-2025-23.1.001.
- [8] P. Inkeaw *et al.*, "Automatic Driver Drowsiness Detection Using Artificial Neural Network Based on Visual Facial Descriptors: Pilot Study," *Nature and Science of Sleep*, vol. 14. pp. 1641–1649, 2022. doi: 10.2147/NSS.S376755.
- [9] J. Chandiwala and S. Agarwal, "Driver's real-time Drowsiness Detection using Adaptable Eye Aspect Ratio and Smart Alarm System," *2021 7th International Conference on Advanced Computing and Communication Systems, ICACCS 2021*. pp. 1350–1355, 2021. doi: 10.1109/ICACCS51430.2021.9441756.
- [10] R. Florez, F. Palomino-Quispe, R. J. Coaquira-Castillo, J. C. Herrera-Levano, T. Paixão, and A. B. Alvarez, "A CNN-Based Approach for Driver Drowsiness Detection by Real-Time Eye State Identification," *Applied Sciences (Switzerland)*, vol. 13, no. 13. 2023. doi: 10.3390/app13137849.
- [11] S. Zu, Y. Jin, D. Yang, and H. Xu, "DrowsyNet: Multivariate Time Series Classification for Embedded Driver Drowsiness Detection," *2022 8th International Conference on Control, Automation and Robotics, ICCAR 2022*. pp. 442–451, 2022. doi: 10.1109/ICCAR55106.2022.9782626.
- [12] F. Faraji, F. Lotfi, J. Khorramdel, A. Najafi, and A. Ghaffari, "Drowsiness Detection Based On Driver Temporal Behavior Using a New Developed Dataset." 2021. [Online]. Available: <http://arxiv.org/abs/2104.00125>
- [13] Saputra Desy, "Ratusan hektar tambak udang jebol akibat rob," [antaranews.com](https://www.antaranews.com/berita/289303/ratusan-hektar-tambak-udang-jebol-akibat-rob). [Online]. Available: <https://www.antaranews.com/berita/289303/ratusan-hektar-tambak-udang-jebol-akibat-rob>
- [14] J. Wan, H. Liu, Y. Wu, Z. Lai, W. Min, and J. Liu, "Precise facial landmark detection by Dynamic Semantic Aggregation Transformer," *Pattern Recognition*, vol. 156. 2024. doi: 10.1016/j.patcog.2024.110827.
- [15] V. S. K. Rupani, V. V. S. Thushar, and K. Tejith, "Real-Time Drowsiness Detection Using Eye Aspect Ratio and Facial Landmark Detection," *arXiv preprint arXiv:2408.05836*. 2024. [Online]. Available: <https://arxiv.org/abs/2408.05836> <https://arxiv.org/pdf/2408.05836>
- [16] J. Holdsworth and M. Scapicchio, "What is deep learning?" [Online]. Available: <https://www.ibm.com/topics/deep-learning>
- [17] M. Waqas and U. W. Humphries, "A Critical Review of RNN and LSTM Variants in Hydrological Time Series Predictions," *MethodsX*. p. 102946, 2024.
- [18] A. T. Mohan and D. V. Gaitonde, "A Deep Learning based Approach to Reduced Order Modeling for Turbulent Flow Control using LSTM Neural Networks," no. April, 2018, doi: 10.48550/arXiv.1804.09269.
- [19] A. M. Tahir *et al.*, "Deep Learning for Reliable Classification of COVID-19, MERS, and SARS from Chest X-ray Images," *Cognitive Computation*. 2022. doi: 10.1007/s12559-021-09955-1.
- [20] S. Hao, D. H. Lee, and D. Zhao, "Sequence to sequence learning with attention mechanism for short-term passenger flow prediction in large-scale metro system," *Transp. Res. Part C Emerg. Technol.*, vol. 107, no. August, pp. 287–300, 2019, doi: 10.1016/j.trc.2019.08.005.
- [21] A. Puente-Castro, E. Fernandez-Blanco, D. Rivero, and A. Molares-Ulloa, "Enhancing Orthopox Image Classification Using Hybrid Machine Learning and Deep Learning Models," 2025, [Online]. Available: <http://arxiv.org/abs/2506.06007>
- [22] Y. Dang *et al.*, "Data Augmentation for Sequential Recommendation: A Survey," 2024, [Online]. Available: <http://arxiv.org/abs/2409.13545>
- [23] D. Lovell, D. Miller, J. Capra, and A. Bradley, "Never mind the metrics -- what about the

- uncertainty? Visualising confusion matrix metric distributions.” 2022. [Online]. Available: <http://arxiv.org/abs/2206.02157>
- [24] J. Zou, C. Yuan, X. Zhang, G. Zou, and A. T. K. Wan, “Model Averaging for Support Vector Machine by Cross-Validation,” vol. 1, pp. 1–33, 2021, [Online]. Available: <http://arxiv.org/abs/2112.14578>
- [25] F. Aldi, I. Nozomi, and S. Soeheri, “Comparison of Drug Type Classification Performance Using KNN Algorithm,” *SinkrOn*, vol. 7, no. 3, pp. 1028–1034, 2022, doi: 10.33395/sinkron.v7i3.11487.
- [26] Muthukumar G. and Jyosna Philip, “CNN-LSTM Hybrid Deep Learning Model for Remaining Useful Life Estimation,” *Int. J. Innov. Res. Multidiscip. F.*, vol. 10, no. Special Issue-54, pp. 38–50, 2024, [Online]. Available: <https://www.ijirmf.com>
- [27] M. Cao, R. Yao, J. Xia, K. Jia, and H. Wang, “LSTM Attention Neural-Network-Based Signal Detection for Hybrid Modulated Faster-Than-Nyquist Optical Wireless Communications †,” *Sensors*, vol. 22, no. 22, 2022, doi: 10.3390/s22228992.

