

Tinjauan Sistematis terhadap Arsitektur Convolutional Neural Network untuk Deteksi Kantuk pada Pengemudi Berbasis Visual

Michael Sidharta Dharma ^{(1)*}, Aditiya Hermawan ⁽²⁾

Departemen Teknik Informatika, Universitas Buddhi Dharma, Tangerang, Indonesia
e-mail : {michael.sidharta.dharma,aditiya.hermawan}@ubd.ac.id.

* Penulis korespondensi.

Artikel ini diajukan 22 Desember 2025, direvisi 26 Juni 2026, diterima 1 Juli 2026, dan dipublikasikan 25 September 2026.

Abstract

This study presents a systematic literature review (SLR) on the application of Convolutional Neural Networks (CNNs) for vision-based driver drowsiness detection. A total of 23 peer-reviewed studies published between 2021 and 2025 were analyzed using the PICOC framework and PRISMA 2020 guidelines. The review aims to identify trends in model architectures, datasets, preprocessing strategies, and performance outcomes. The findings show that transfer learning models, such as MobileNetV2, ResNet, and InceptionV3, along with hybrid architectures like CNN-LSTM, dominate recent research, achieving accuracies predominantly above 90%. However, performance varies significantly depending on dataset characteristics and experimental conditions. Key challenges include sensitivity to low-light environments, occlusion (e.g., glasses), and head pose variations. In addition, most studies rely on controlled datasets, limiting real-world applicability and generalization. Emerging trends emphasize lightweight architectures, temporal modeling, and efficiency-oriented optimization for real-time deployment. This review highlights the need for cross-dataset evaluation, more diverse and representative data, and robust model design to ensure reliable performance in real-world driving scenarios.

Keywords: *Convolutional Neural Network, Driver Drowsiness Detection, Hybrid Architecture, Real-Time Detection, Systematic Review*

Abstrak

Penelitian ini menyajikan *systematic literature review* (SLR) terhadap penerapan Convolutional Neural Network (CNN) untuk deteksi kantuk pengemudi berbasis visual. Sebanyak 23 studi *peer-reviewed* yang dipublikasikan pada periode 2021–2025 dianalisis menggunakan kerangka PICOC dan pedoman PRISMA 2020. Kajian ini bertujuan mengidentifikasi tren arsitektur model, *dataset*, strategi *preprocessing*, serta performa yang dihasilkan. Hasil menunjukkan bahwa pendekatan *transfer learning* seperti MobileNetV2, ResNet, dan InceptionV3, serta arsitektur hibrida seperti CNN-LSTM, mendominasi penelitian dengan akurasi umumnya di atas 90%. Namun, performa sangat dipengaruhi oleh karakteristik *dataset* dan kondisi eksperimen. Tantangan utama meliputi sensitivitas terhadap pencahayaan rendah, *occlusion* (misalnya kacamata), serta variasi pose kepala. Selain itu, sebagian besar studi masih menggunakan *dataset* terkontrol sehingga membatasi kemampuan generalisasi pada kondisi dunia nyata. Tren penelitian terkini mengarah pada pengembangan arsitektur ringan, pemodelan temporal, dan optimasi efisiensi untuk mendukung implementasi *real-time*. Tinjauan ini menekankan pentingnya evaluasi lintas *dataset*, penggunaan data yang lebih representatif, serta desain model yang lebih *robust* untuk meningkatkan keandalan sistem dalam skenario berkendara nyata.

Kata Kunci: *Convolutional Neural Network, Deteksi Kantuk Pengemudi, Hybrid Architecture, Real-Time Detection, Systematic Review*

1. PENDAHULUAN

Mengemudi dalam keadaan mengantuk terbukti menurunkan konsentrasi dan memperlambat waktu reaksi pengemudi (Thoriq et al., 2024). Secara global, kondisi ini diperkirakan menjadi penyebab sekitar 20% kecelakaan lalu lintas (Salem & Waleed, 2024). Di Indonesia, 79% pengemudi dilaporkan pernah berkendara dalam keadaan mengantuk, dan 32% di antaranya hampir mengalami kecelakaan fatal akibat kondisi tersebut (Rahmadiyani & Widyanti, 2023).



Artikel ini didistribusikan mengikuti lisensi Atribusi-NonKomersial CC BY-NC sebagaimana tercantum pada <https://creativecommons.org/licenses/by-nc/4.0/>.

Fakta ini menegaskan pentingnya sistem yang mampu mendeteksi serta memberikan peringatan dini terhadap tanda-tanda kantuk secara otomatis.

Metode deteksi tradisional seperti pemantauan pola kemudi, penyimpangan jalur, atau sensor kendaraan masih terbatas pada penerapan dan tidak secara langsung mengidentifikasi tanda fisiologis maupun perilaku pengemudi (Namburi et al., 2024). Oleh karena itu, riset terkini beralih pada pendekatan *computer vision* dan *deep learning* yang menawarkan solusi adaptif, *non-intrusive*, dan dapat beroperasi secara *real-time*. Dalam konteks deteksi kantuk pengemudi, Convolutional Neural Network (CNN) masih menjadi fondasi utama karena kemampuannya dalam mengekstraksi fitur spasial dari citra wajah, yang selanjutnya dapat digunakan sebagai masukan bagi arsitektur Recurrent Neural Network (RNN) untuk memodelkan dinamika temporal (Alameen & Alhothali, 2023). Berbagai arsitektur *deep learning* seperti RNN juga telah diintegrasikan dengan CNN untuk menambah fitur analisis pola temporal pada data video atau sekuens wajah (Namburi et al., 2024). Hal ini menunjukkan bahwa kombinasi CNN dengan arsitektur *deep learning* lain mampu menghasilkan sistem deteksi kantuk pengemudi yang lebih andal, adaptif, dan efektif dalam mengenali tanda-tanda kantuk secara *real-time*.

Meskipun demikian, arsitektur hibrida CNN dan variannya masih menghadapi berbagai keterbatasan, seperti sensitivitas terhadap pencahayaan rendah, penggunaan kacamata, dan variasi ekspresi wajah pengemudi (Al-Gburi et al., 2025; Huang, 2025; Magán et al., 2022; T.-C. Phan et al., 2025). Hal tersebut mendorong munculnya beragam penelitian baru yang mencoba mengatasi keterbatasan tersebut. Oleh karena itu, diperlukan tinjauan sistematis untuk memetakan tren penelitian, mengevaluasi performa pendekatan yang ada, serta mengidentifikasi celah penelitian dalam penerapan CNN untuk deteksi kantuk pengemudi. Beberapa tinjauan literatur sistematis sebelumnya telah membahas deteksi kantuk pengemudi dari berbagai perspektif. Salah satunya telah berfokus pada pendekatan berbasis sinyal fisiologis seperti EEG dan ECG, sehingga tidak mencakup pendekatan *vision-based* berbasis CNN (Saleem et al., 2023). Sementara itu, tinjauan lainnya mengkaji berbagai pendekatan *deep learning* secara luas, termasuk CNN, RNN, model hibrida, dan *transformer* dalam skenario *multimodal* (Fonseca & Ferreira, 2025). Namun, karena cakupannya yang luas, analisis yang dilakukan belum secara mendalam mengeksplorasi arsitektur CNN secara spesifik, khususnya dalam kaitannya dengan desain model, karakteristik *dataset*, dan performa yang dihasilkan.

Tabel 1 Perbandingan Tinjauan Literatur Terdahulu dan Penelitian Ini

Aspek	Saleem et al. (2023)	Fonseca & Ferreira (2025)	Penelitian Ini
Fokus	Physiological signals (EEG, ECG, EOG, respiration, HR, dll)	<i>Deep learning</i> untuk drowsiness detection (<i>multimodal</i>)	Arsitektur CNN & kombinasinya
Rentang Tahun	2011–2021 (berdasarkan kriteria seleksi studi)	2015–2025	2021–2025
Analisis Model	Tidak berfokus pada arsitektur; mencakup berbagai ML/DL secara umum	CNN, RNN, hybrid, <i>transformer</i>	Fokus mendalam CNN & hybrid
<i>Dataset</i>	Beragam (berdasarkan studi yang direview; tidak menggunakan <i>dataset</i> tunggal)	Variatif (behavioral, physiological, <i>multimodal</i>)	Analisis kuantitatif + tren
Evaluasi Kinerja	Ringkasan hasil studi (tanpa evaluasi eksperimental langsung)	Komprehensif (accuracy, F1, AUC)	Komparatif antar studi
Gap	Tidak membahas pendekatan <i>vision-based</i> /CNN; fokus pada keterbatasan sensor fisiologis	Tidak fokus pada CNN	Fokus spesifik CNN

Berdasarkan limitasi tersebut, masih terdapat kebutuhan akan tinjauan sistematis yang secara khusus memfokuskan pada arsitektur CNN dalam deteksi kantuk pengemudi. Penelitian ini bertujuan untuk mengisi celah tersebut dengan melakukan analisis mendalam terhadap



perkembangan arsitektur CNN dan variannya pada periode 2021–2025, termasuk hubungan antara desain model, *dataset* yang digunakan, serta performa yang dicapai. Perbandingan antara tinjauan literatur sistematis (Systematic Literature Review/SLR) terdahulu dan posisi penelitian ini disajikan pada Tabel 1 untuk menegaskan kontribusi dan kebaruan studi yang diusulkan.

Penelitian ini berkontribusi secara teoretis dalam memperkaya pemetaan pengetahuan mengenai evolusi riset deteksi kantuk pengemudi berbasis CNN, serta secara praktis dalam memberikan panduan bagi pengembangan sistem deteksi kantuk yang lebih adaptif dan andal di lingkungan nyata. Struktur artikel ini disusun sebagai berikut: Bagian 2 menjelaskan metodologi SLR, termasuk strategi pencarian dan kriteria seleksi. Bagian 3 menyajikan hasil dan analisis tematik serta membahas tren dan tantangan penelitian, sedangkan Bagian 4 menyimpulkan temuan utama serta arah penelitian masa depan.

2. METODE PENELITIAN

Penelitian ini menggabungkan dua kerangka kerja yang saling melengkapi, yaitu PICOC (*Population, Intervention, Comparison, Outcome, Context*) dan PRISMA 2020, untuk memastikan proses tinjauan literatur dilakukan secara sistematis, terstruktur, dan transparan. Kerangka PICOC digunakan sebagai dasar konseptual dalam merumuskan pertanyaan penelitian (*research questions*) serta menentukan ruang lingkup kajian (Carrera-Rivera et al., 2022). Tabel 2 menunjukkan komponen PICOC yang digunakan dalam penelitian ini sebagai dasar untuk merumuskan fokus kajian serta membatasi ruang lingkup SLR.

Di sisi lain, pedoman PRISMA 2020 (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*) digunakan untuk mengatur alur seleksi studi secara transparan dan dapat direplikasi (Fonseca & Ferreira, 2025). PRISMA berperan dalam mendokumentasikan setiap tahapan proses tinjauan, mulai dari identifikasi literatur, penyaringan berdasarkan kriteria inklusi dan eksklusi, hingga penentuan studi akhir yang dianalisis.

Tabel 2 Rancangan Kerangka PICOC

Komponen PICOC	Deskripsi dalam Penelitian Ini
<i>Population</i> (P)	Studi yang meneliti topik deteksi kantuk pada pengemudi berbasis visual menggunakan arsitektur Convolutional Neural Network (CNN)
<i>Intervention</i> (I)	Pendekatan atau arsitektur CNN deteksi yang digunakan seperti model hibrida atau non-hibrida yang digunakan untuk mendeteksi kantuk.
<i>Comparison</i> (C)	Perbandingan antar pendekatan berdasarkan performa akurasi, kompleksitas model, serta kemampuan generalisasi lintas <i>dataset</i>
<i>Outcomes</i> (O)	Hasil utama berupa akurasi model pada penelitian yang menggunakan arsitektur CNN
<i>Context</i> (C)	<i>Driver drowsiness detection, deep learning, CNN, vision</i>

2.1 Pertanyaan Penelitian

Berdasarkan kesenjangan yang telah diidentifikasi pada bagian pendahuluan, penelitian deteksi kantuk pengemudi masih menghadapi beberapa keterbatasan. Studi tinjauan sebelumnya oleh Saleem et al. (2023) berfokus pada pendekatan berbasis sinyal fisiologis, sehingga tidak mencakup pendekatan *vision-based* berbasis Convolutional Neural Network (CNN). Sementara itu, tinjauan oleh Fonseca & Ferreira (2025) membahas berbagai pendekatan *deep learning* secara luas, tetapi belum secara spesifik mendalami arsitektur CNN, khususnya dalam kaitannya dengan desain model, karakteristik *dataset*, serta performa yang dihasilkan. Selain itu, variasi pendekatan dan skenario pengujian antarstudi, termasuk perbedaan arsitektur, *dataset*, dan teknik *preprocessing*, menyebabkan sulitnya memperoleh pemahaman yang konsisten mengenai efektivitas model CNN dalam berbagai kondisi implementasi. Sebagian besar penelitian juga melaporkan performa dalam bentuk akurasi tanpa adanya standar evaluasi yang seragam, sehingga perbandingan antarstudi menjadi terbatas.



Kesenjangan tersebut menjadi dasar dalam perumusan empat pertanyaan penelitian (*Research Questions/RQ*) pada studi ini. Pertanyaan penelitian difokuskan pada empat aspek utama, yaitu perkembangan dan variasi arsitektur CNN, performa dan reliabilitas model pada berbagai kondisi eksperimental, karakteristik *dataset* dan teknik pra-pemrosesan serta pengaruhnya terhadap performa model, dan tantangan implementasi serta arah pengembangan metode CNN di masa depan. Keempat pertanyaan penelitian tersebut disajikan pada Tabel 3.

Tabel 3 Pertanyaan Penelitian

Pertanyaan Penelitian (RQ)	Motivasi Utama
RQ1. Model arsitektur CNN apa yang digunakan dalam penelitian pada deteksi kantuk periode 2021 – 2025?	Untuk memetakan tren dan variasi arsitektur CNN terkini, termasuk model <i>hybrid</i> dan <i>lightweight</i> , dalam sistem deteksi kantuk pengemudi.
RQ2. Bagaimana tingkat akurasi dan reliabilitas model CNN dalam mendeteksi kantuk pada berbagai kondisi eksperimental?	Untuk mengevaluasi performa model CNN dan melihat sejauh mana model tetap reliabel di kondisi nyata.
RQ3. Jenis <i>dataset</i> dan teknik <i>preprocessing</i> apa yang umum digunakan dalam deteksi kantuk berbasis arsitektur CNN?	Untuk menyintesis informasi terkait dataset dan teknik pra-pemrosesan yang paling sering digunakan serta pengaruhnya terhadap kinerja model.
RQ4. Apa tantangan utama dan kesenjangan penelitian yang masih terdapat dalam studi deteksi kantuk, serta arah penelitian apa yang disarankan untuk masa depan?	Untuk mengidentifikasi keterbatasan metodologis yang berulang, mengungkap area yang belum tergarap, dan memberikan arah penelitian lanjutan guna meningkatkan efektivitas deteksi.

2.2 Kriteria Inklusi dan Eksklusi

Kriteria inklusi dan eksklusi perlu ditentukan sebelum penelitian dimulai untuk mencegah bias pada hasil (Carrera-Rivera et al., 2022). Tinjauan ini difokuskan pada penelitian yang menelaah sistem deteksi kantuk pengemudi berbasis CNN. Kriteria inklusi dan eksklusi yang digunakan untuk memastikan bahwa studi yang dianalisis relevan, sesuai dengan ruang lingkup, serta memenuhi standar kualitas penelitian disajikan pada Tabel 4.

Tabel 4 Kriteria Inklusi dan Eksklusi Pemilihan Studi

Inklusi	Eksklusi
Studi yang berfokus pada partisipan pengemudi (komersial, pribadi, atau profesional).	Partisipan non-pengemudi seperti mahasiswa, tenaga medis, atau pilot.
Menggunakan arsitektur CNN (<i>hybrid</i> atau <i>non-hybrid</i>) untuk mendeteksi kantuk.	Tidak melibatkan CNN atau hanya membahas metode fisiologis/perangkat keras tanpa evaluasi model.
Melaporkan setidaknya satu metrik kinerja kuantitatif (misalnya akurasi, presisi, <i>recall</i> , atau F1-score).	Tidak mencantumkan metrik kinerja kuantitatif.
Artikel jurnal ilmiah berbahasa Inggris yang dipublikasikan pada jurnal bereputasi Q1–Q4 berdasarkan SCImago Journal Rank serta tersedia dalam akses terbuka (open access) untuk memungkinkan proses ekstraksi dan verifikasi data secara menyeluruh.	Artikel non-Inggris, abstrak konferensi, opini, atau editorial, dan tidak <i>open access</i> .
Deteksi kantuk berbasis citra visual dalam konteks mengemudi.	Tidak relevan dengan konteks mengemudi atau hanya berbasis sinyal fisiologis.



2.3 Strategi Pencarian

Pencarian literatur dalam penelitian ini dilakukan secara sistematis pada bulan September 2025 dengan cakupan publikasi dari Januari 2021 hingga September 2025. Proses identifikasi studi difokuskan pada dua basis data utama, yaitu Scopus dan IEEE Xplore. Pemilihan kedua basis data ini didasarkan pada kualitas indeksasi serta relevansinya terhadap bidang penelitian. Scopus dipilih karena memiliki cakupan indeksasi yang luas terhadap jurnal bereputasi lintas penerbit, termasuk Elsevier, Springer, dan Wiley, sehingga secara tidak langsung telah mencakup sebagian besar literatur yang juga tersedia pada platform seperti ScienceDirect dan SpringerLink. Di sisi lain, IEEE Xplore secara khusus dipilih untuk memastikan keterwakilan literatur teknis di bidang teknik elektro, komputer, dan kecerdasan buatan yang mungkin tidak seluruhnya terindeks secara optimal di Scopus. Dengan kombinasi tersebut, penelitian ini berupaya menyeimbangkan antara cakupan literatur yang luas dan relevansi domain spesifik. Selain itu, penggunaan strategi pencarian yang komprehensif serta kriteria inklusi yang ketat diharapkan mampu meminimalkan potensi bias seleksi meskipun jumlah basis data yang digunakan terbatas.

Untuk meminimalkan potensi bias seleksi, strategi pencarian dirancang menggunakan kombinasi kata kunci yang komprehensif namun tetap terarah. Kata kunci mencakup tiga aspek utama, yaitu kondisi yang diteliti (*drowsiness/fatigue*), metode yang digunakan (Convolutional Neural Network), serta modalitas data (*vision-based*). Selain itu, kriteria inklusi dan eksklusi diterapkan secara ketat guna memastikan bahwa hanya studi yang relevan dan berkualitas yang disertakan dalam analisis. Pendekatan ini sejalan dengan praktik dalam berbagai studi *systematic literature review* (SLR) sebelumnya yang mengombinasikan beberapa basis data untuk meningkatkan cakupan literatur sekaligus menjaga konsistensi metodologi (Fonseca & Ferreira, 2025; Saleem et al., 2023).

Query pencarian disusun menggunakan operator Boolean untuk mengidentifikasi artikel yang relevan secara semantik pada kedua basis data. Query yang digunakan pada masing-masing basis data adalah sebagai berikut:

Scopus: *TITLE-ABS-KEY(("drowsiness detection" OR "driver drowsiness" OR "driver sleepiness") AND ("Convolutional Neural Network" OR "cnn") AND ("driver" OR "driving") AND ("vision-based" OR "image" OR "video" OR "facial" OR "eye"))*

IEEE Xplore: *("All Metadata": "drowsiness detection" OR "driver drowsiness") AND ("All Metadata": "Convolutional Neural Network" OR "cnn") AND ("All Metadata": "vision" OR "facial" OR "eye")*

Variasi istilah tersebut digunakan untuk meningkatkan sensitivitas pencarian tanpa mengorbankan spesifisitas terhadap fokus penelitian, yaitu deteksi kantuk pengemudi berbasis Convolutional Neural Network (CNN). Dalam penelitian ini, teknik pencarian tambahan seperti *citation tracking* dan *snowballing* tidak digunakan, dengan tujuan menjaga konsistensi prosedur serta memastikan bahwa seluruh tahapan tetap terstruktur, transparan, dan dapat direplikasi.

2.4 Pengumpulan dan Ekstraksi Data

Proses pengumpulan data dan ekstraksi data penting untuk mencari informasi yang diperlukan dalam langkah menjawab pertanyaan penelitian (Carrera-Rivera et al., 2022). Proses pengumpulan dan ekstraksi data dilakukan dan diimpor ke dalam platform *Mendeley*, yang digunakan untuk menghapus duplikasi secara otomatis. Penyaringan judul dan abstrak dilakukan berdasarkan kriteria inklusi dan eksklusi yang telah ditentukan. Studi yang memenuhi kriteria awal selanjutnya ditelaah secara penuh untuk menentukan inklusi akhir. Untuk ekstraksi data, digunakan lembar kerja terstruktur dalam Microsoft Excel guna mengumpulkan informasi utama dari setiap studi. Hal ini meliputi *Problem Statement*, *Finding & Novelty*, *Methodology*, *Result*, *Conclusion*, *Limitation*, dan *Dataset*. Proses ekstraksi didukung oleh *ChatGPT* (OpenAI, San Francisco, CA, USA, GPT-5, 2025) untuk membantu identifikasi dan organisasi data. Semua data



yang diekstraksi diverifikasi secara manual melalui pembacaan penuh terhadap setiap artikel sebelum disertakan dalam analisis akhir, guna menjaga akurasi dan integritas data.

2.5 Quality Assessment

Quality assessment perlu dilakukan untuk mengevaluasi kualitas metodologis setiap penelitian yang memenuhi kriteria inklusi setelah proses seleksi studi primer selesai. Tahap ini bertujuan untuk memastikan bahwa studi yang disintesis memiliki informasi yang memadai mengenai pengembangan model, proses eksperimen, dan pelaporan hasil, sehingga dapat mendukung sintesis yang lebih reliabel. Proses penilaian dilakukan menggunakan kriteria yang diadaptasi dari pedoman *systematic literature review* untuk penelitian ilmu komputer (Carrera-Rivera et al., 2022). Kriteria *quality assessment* yang digunakan dalam penelitian ini disusun untuk mengevaluasi kelengkapan pelaporan metodologi pada setiap studi primer. Penilaian difokuskan pada aspek-aspek yang dianggap penting dalam penelitian deteksi kantuk pengemudi berbasis *deep learning*, meliputi pelaporan *dataset*, arsitektur CNN, tahapan *preprocessing*, metode evaluasi, hasil eksperimen, serta pembahasan keterbatasan penelitian. Kriteria yang digunakan dalam proses penilaian disajikan pada Tabel 5.

Tabel 5 Kriteria Quality Assessment

Kode	Kriteria Penilaian
QA1	Karakteristik <i>dataset</i> dijelaskan secara memadai, meliputi sumber data serta informasi yang relevan untuk mendukung proses evaluasi model.
QA2	Arsitektur CNN yang digunakan dijelaskan secara memadai, termasuk model utama atau modifikasi arsitektur yang diterapkan.
QA3	Tahapan <i>preprocessing</i> dijelaskan secara memadai, termasuk proses persiapan data sebelum pelatihan model
QA4	Prosedur evaluasi dan metrik performa dijelaskan secara memadai, termasuk metode pengujian atau metrik yang digunakan untuk mengevaluasi model.
QA5	Hasil evaluasi dilaporkan menggunakan metrik yang memadai untuk mendukung interpretasi performa model.
QA6	Keterbatasan penelitian dibahas secara eksplisit sehingga ruang lingkup dan potensi kelemahan metodologi penelitian dapat dipahami

Setiap kriteria pada *quality assessment* diberikan skor berdasarkan tingkat kelengkapan informasi yang dilaporkan oleh masing-masing studi. Skor diberikan secara bertingkat sehingga dapat menggambarkan tingkat kualitas pelaporan metodologi dari setiap penelitian. Skema penilaian yang digunakan dalam penelitian ini disajikan pada Tabel 6.

Tabel 6 Quality Score Penelitian

Skor	Kriteria Penilaian
1	Kriteria dijelaskan secara lengkap.
0,5	Kriteria dijelaskan sebagian, tetapi masih terdapat informasi yang belum lengkap.
0	Kriteria tidak dijelaskan.

Total skor diperoleh dari penjumlahan seluruh kriteria *quality assessment*. Berdasarkan total skor tersebut, kualitas studi diklasifikasikan ke dalam tiga kategori, yaitu kualitas tinggi (5,0–6,0), kualitas sedang (3,0–4,5), dan kualitas rendah (<3,0). Selain melakukan *quality assessment*, penelitian ini juga mengidentifikasi potensi *risk of bias* secara naratif pada setiap studi primer. Identifikasi dilakukan berdasarkan karakteristik metodologi yang dilaporkan dalam masing-masing penelitian, dengan mempertimbangkan aspek seperti karakteristik *dataset*, rancangan eksperimen, prosedur evaluasi, serta keterbatasan yang diungkapkan oleh penulis. Hasil identifikasi tersebut digunakan sebagai salah satu pertimbangan dalam proses sintesis dan interpretasi temuan penelitian.



2.6 Sintesis Data

Sintesis data merupakan proses menyatukan hasil berbagai studi untuk menghasilkan kesimpulan atau pemahaman baru yang terstruktur dan bermakna (Carrera-Rivera et al., 2022). Pada tahap ini dilakukan pendekatan naratif terstruktur yang didasarkan pada empat pertanyaan penelitian utama. Mengingat heterogenitas tinggi antar studi khususnya dalam tujuan pemodelan, jenis input, lingkungan evaluasi, dan metrik keluaran analisis meta-kuantitatif tidak dimungkinkan. Oleh karena itu, hasil disajikan secara tematik untuk menangkap pola berulang, tren metodologis, serta variasi kinerja model arsitektur CNN dalam deteksi kantuk.

Hasil sintesis disajikan dalam bentuk tabel yang dibagi ke dalam beberapa kolom, meliputi karakteristik studi, arsitektur model, metrik performa yang dilaporkan, sifat *dataset*, dan tantangan implementasi. Struktur ini memungkinkan keterkaitan langsung antara temuan dan pertanyaan riset, sekaligus mendukung interpretasi komparatif antar studi. Pendekatan sintesis ini memberikan gambaran komprehensif mengenai kondisi terkini literatur mengenai deteksi kantuk pengemudi (Driver Drowsiness Detection/DDD) berbasis CNN, termasuk keberagaman metodologi, hambatan implementasi, serta kesenjangan penelitian yang masih perlu dijembatani. Dengan demikian, hasil tinjauan ini memberikan dasar berbasis bukti untuk menilai reliabilitas, skalabilitas, dan potensi penerapan model CNN dalam konteks deteksi kantuk pengemudi.

3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil sintesis dari 23 penelitian yang memenuhi kriteria inklusi dan eksklusi serta kerangka PICOC yang telah dirancang sebelumnya. Analisis mencakup karakteristik umum penelitian, distribusi arsitektur CNN yang digunakan, serta reliabilitas dan akurasi model. Selain itu, analisis mencakup persebaran *dataset* serta tantangan dan keterbatasan penerapan sistem deteksi kantuk berbasis CNN.

3.1 Pemilihan Studi

Proses pemilihan penelitian dalam tinjauan sistematis ini mengikuti pedoman PRISMA 2020 yang mencakup empat tahap utama, yaitu identifikasi, penyaringan, penilaian kelayakan, dan inklusi. Seluruh proses seleksi didasarkan pada kriteria inklusi dan eksklusi yang telah dirumuskan sebelumnya pada Tabel 4. Kriteria tersebut mengacu pada kerangka PICOC untuk memastikan konsistensi dan relevansi penelitian yang dianalisis.

Pada tahap identifikasi, sebanyak 1259 studi diperoleh dari basis data yang digunakan. Selanjutnya, dilakukan proses penyaringan awal dengan mengeliminasi 1068 studi berdasarkan kriteria eksklusi, yang meliputi studi di luar rentang tahun publikasi yang ditetapkan sebanyak 221, artikel non-Inggris 3, serta tipe dokumen yang tidak relevan berjumlah 838, meliputi *conference proceedings*, *review articles*, editorial, dan dokumen non-penelitian lainnya. Selain itu, sebanyak 6 studi duplikat yang teridentifikasi antar basis data juga dihapus. Setelah tahap ini, tersisa 191 studi untuk proses penyaringan lebih lanjut.

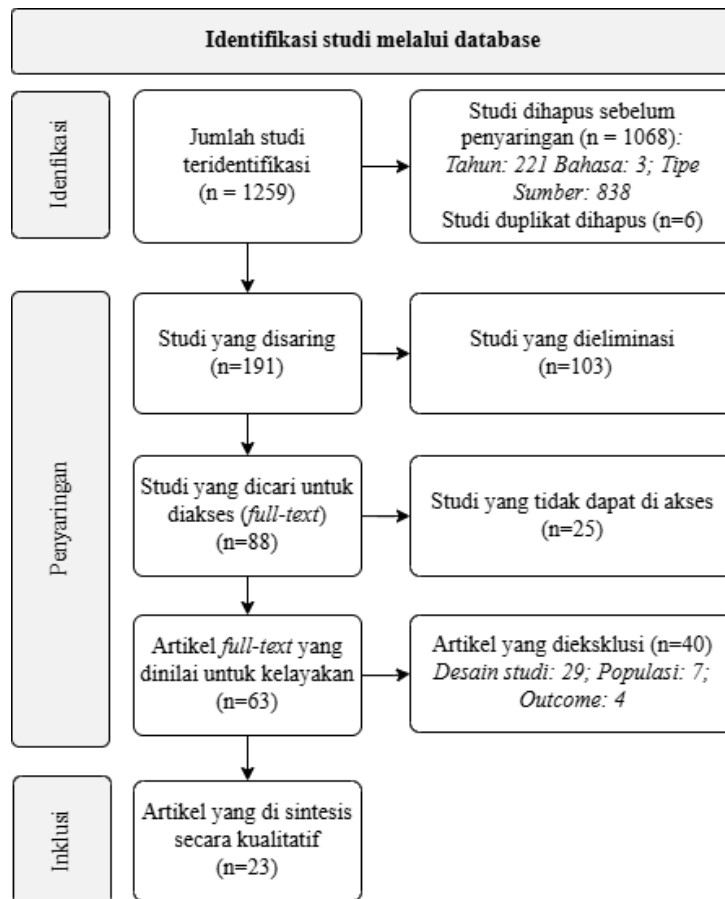
Pada tahap penyaringan, dilakukan evaluasi terhadap judul dan abstrak dari 191 studi untuk menilai kesesuaiannya dengan kriteria inklusi. Sebanyak 103 studi dieliminasi karena tidak memenuhi kriteria, terutama tidak berfokus pada konteks pengemudi, tidak menggunakan pendekatan berbasis CNN, atau tidak berbasis data visual. Beberapa studi juga dieliminasi karena hanya menggunakan data fisiologis tanpa melibatkan citra visual. Tahap ini menghasilkan 88 studi yang dinilai relevan untuk dianalisis lebih lanjut. Pada tahap pencarian teks lengkap (*full-text retrieval*), dilakukan upaya untuk mengakses keseluruhan artikel dari 88 studi tersebut. Sebanyak 25 artikel tidak dapat diakses, sehingga dikeluarkan dari proses seleksi. Dengan demikian, sebanyak 63 artikel dilanjutkan ke tahap penilaian kelayakan.

Pada tahap penilaian kelayakan, setiap artikel dianalisis secara mendalam berdasarkan kriteria inklusi dan eksklusi yang dirancang menggunakan kerangka PICOC. Dari total 63 artikel, sebanyak 42 studi dieliminasi karena tidak memenuhi kriteria yang ditetapkan. Secara rinci, 29



studi dieliminasi pada aspek *study design* karena tidak menggunakan arsitektur CNN atau tidak menyajikan evaluasi model secara eksperimental. Sebanyak 7 studi dieliminasi pada aspek *population* karena tidak berfokus pada pengemudi, sementara 4 studi dieliminasi pada aspek *outcome* karena tidak melaporkan metrik kinerja kuantitatif yang memadai.

Pada akhirnya, sebanyak 23 studi dinyatakan memenuhi seluruh kriteria inklusi dan diikutsertakan dalam sintesis kualitatif. Studi-studi tersebut merepresentasikan berbagai pendekatan berbasis CNN dalam deteksi kantuk pengemudi, mencakup variasi arsitektur, *dataset*, serta strategi evaluasi yang digunakan. Alur lengkap proses seleksi studi ini disajikan pada Gambar 1.



Gambar 1 Proses Identifikasi, Penyaringan, dan Inklusi Studi dengan PRISMA 2020

3.2 Gambaran Umum Penelitian

Bagian ini memperkenalkan karakteristik utama dari 23 studi yang dianalisis dalam tinjauan ini. Studi-studi tersebut menunjukkan variasi yang cukup besar dalam tahun publikasi, *venue* penerbitan, arsitektur model, serta jenis *dataset* yang digunakan. Namun terdapat kecenderungan metodologis yang jelas: dominasi teknik berbasis CNN dan fokus pada evaluasi sesuai metrik yang telah ditentukan. Tabel 11 pada Lampiran A menyajikan 23 studi yang merepresentasikan berbagai pendekatan dan arsitektur CNN terkini dalam mendeteksi kantuk pada pengemudi.

3.2.1 Quality Assessment Studi

Seluruh studi yang memenuhi kriteria inklusi dievaluasi menggunakan instrumen *quality assessment*. Penilaian ini bertujuan untuk mengevaluasi kualitas pelaporan metodologi setiap



penelitian sehingga hasil sintesis yang diperoleh dapat diinterpretasikan secara lebih objektif. Hasil penilaian untuk masing-masing studi disajikan pada Tabel 7.

Tabel 7 Hasil Penilaian Quality Assessment Seluruh Studi

Author	QA1	QA2	QA3	QA4	QA5	QA6	Total	Kategori
Majeed et al. (2023)	1	1	1	1	1	1	6.0	Tinggi
Alameen & Alhothali (2023)	1	1	1	1	1	1	6.0	Tinggi
Delwar et al. (2025)	1	0.5	1	1	1	1	5.5	Tinggi
Florez et al. (2023)	0.5	1	1	1	1	0.5	5.0	Tinggi
Sohail et al. (2024)	0.5	1	1	0.5	0.5	0.5	4.0	Sedang
Magán et al. (2022)	1	1	1	1	0.5	1	5.5	Tinggi
T.-C. Phan et al. (2025)	1	1	1	1	1	1	6.0	Tinggi
Al-Gburi et al. (2025)	1	1	1	1	1	0.5	5.5	Tinggi
Yang & Yi (2024)	0.5	1	0.5	0.5	1	0.5	4.0	Sedang
Nandyal & Sharanabasappa (2024)	1	0.5	1	0.5	0.5	0.5	4.0	Sedang
Salem & Waleed (2024)	1	1	1	1	1	1	6.0	Tinggi
Namburi et al. (2024)	0.5	1	1	1	1	1	5.5	Tinggi
Huang (2025)	0.5	1	1	1	1	0.5	5.0	Tinggi
Rajput et al. (2025)	1	1	1	0.5	0.5	1	5.0	Tinggi
Vijaypriya & Uma (2023)	1	1	1	1	1	0.5	5.5	Tinggi
Adhithyaa et al. (2023)	1	1	1	1	1	1	6.0	Tinggi
Jebraeily et al. (2024)	0.5	1	0.5	1	1	0.5	4.5	Sedang
Kumral & Küçükmanisa (2022)	0.5	1	1	0.5	0.5	0.5	4.0	Sedang
A.-C. Phan et al. (2021)	0.5	1	1	1	1	1	5.5	Tinggi
Minhas et al. (2022)	0.5	0.5	0.5	1	1	0.5	4.0	Sedang
Winarno et al. (2025)	1	1	1	1	1	1	6.0	Tinggi
Ahmed et al. (2023)	0.5	1	1	0.5	1	1	5.0	Tinggi
Venkateswarlu & Chirra (2024)	1	1	1	1	1	1	6.0	Tinggi

Hasil *quality assessment* pada metodologis penelitian yang ditinjau tergolong baik, dengan 17 dari 23 studi (73,9%) memperoleh kategori tinggi (skor $\geq 5,0$), sedangkan 6 studi (26,1%) berada pada kategori sedang (skor 4,0–4,5). Tidak terdapat studi yang memperoleh kategori rendah, sehingga seluruh artikel dinilai memiliki kualitas yang memadai untuk dipertahankan dalam proses sintesis. Sebanyak tujuh studi memperoleh skor maksimum 6,0, yaitu Majeed et al. (2023), Alameen & Alhothali (2023), T.-C. Phan et al. (2025), Salem & Waleed (2024), Adhithyaa et al. (2023), Winarno et al. (2025), serta Venkateswarlu & Chirra (2024). Studi-studi tersebut umumnya memberikan pelaporan yang lengkap mengenai karakteristik *dataset*, arsitektur CNN, tahapan *preprocessing*, metode evaluasi, metrik performa, serta keterbatasan penelitian. Sebaliknya, studi dengan skor lebih rendah umumnya kehilangan poin pada aspek pelaporan *dataset*, kelengkapan metode evaluasi, atau penyampaian keterbatasan penelitian secara eksplisit.

Identifikasi terhadap potensi *risk of bias* juga dilakukan untuk membantu menginterpretasikan hasil sintesis secara lebih kritis. Penilaian ini tidak dimaksudkan untuk mengecualikan studi, melainkan untuk mengidentifikasi karakteristik metodologis yang berpotensi memengaruhi validitas hasil yang dilaporkan. Berdasarkan hasil observasi, sebagian besar studi telah menyajikan pelaporan metodologi yang memadai, sehingga risiko *reporting bias* relatif rendah. Namun, beberapa potensi bias masih ditemukan pada aspek desain penelitian dan strategi evaluasi model.

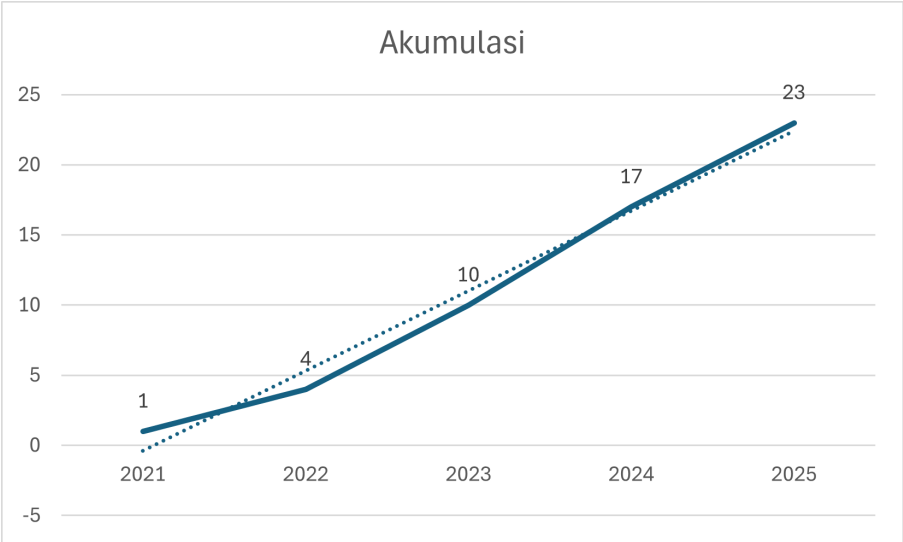
Potensi bias yang paling sering dijumpai adalah *dataset bias* dan *evaluation bias*. Sebanyak sembilan dari 23 studi menggunakan *dataset* YawDD sebagai sumber data utama, sedangkan hanya enam studi yang mengevaluasi model pada lebih dari satu *dataset*. Kondisi ini menunjukkan bahwa sebagian besar model masih divalidasi pada lingkungan data yang terbatas sehingga kemampuan generalisasinya terhadap *dataset* lain belum banyak dibuktikan. Selain itu,



mayoritas studi masih menggunakan satu kali pembagian data latih dan data uji (*single train-test split*), sementara penerapan *k-fold cross-validation* maupun validasi lintas *dataset* masih relatif terbatas. Beberapa studi juga menggunakan *dataset* privat atau *dataset* dengan jumlah sampel yang terbatas, yang berpotensi meningkatkan risiko *sampling bias*. Oleh karena itu, akurasi yang tinggi pada sebagian besar penelitian perlu diinterpretasikan dengan mempertimbangkan karakteristik *dataset* dan strategi evaluasi yang digunakan, karena kedua aspek tersebut dapat memengaruhi kemampuan model ketika diterapkan pada kondisi yang berbeda dari lingkungan pengujiannya.

3.2.2 Karakteristik Studi

Analisis distribusi temporal mengungkapkan peningkatan jumlah publikasi yang konsisten selama beberapa tahun terakhir: dari 1 artikel pada 2021, meningkat menjadi 3 pada 2022, 6 pada 2023, 7 pada 2024, dan 6 artikel pada 2025, sehingga total mencapai 23 studi. Pola kenaikan ini mencerminkan perhatian riset yang bertambah pada pengembangan sistem deteksi kantuk berbasis pembelajaran mendalam, kemungkinan seiring dengan kemajuan arsitektur CNN dan kemudahan akses ke *dataset* citra wajah/ekspresi. Gambar 2 memvisualisasikan akumulasi jumlah penelitian deteksi kantuk pengemudi berbasis CNN.



Gambar 2 Grafik Akumulasi Penelitian Pada Model CNN Deteksi Kantuk Pengemudi

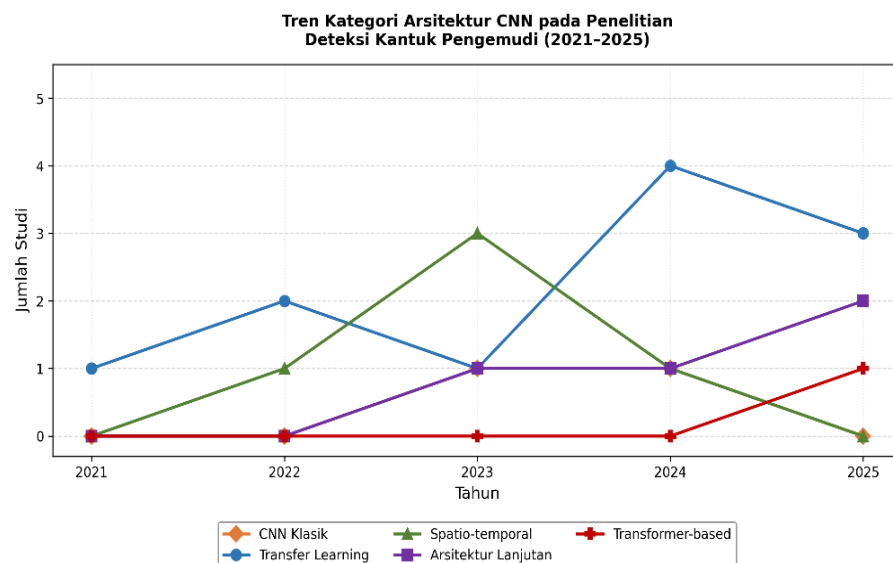
3.3 Arsitektur CNN Dalam Deteksi Kantuk (RQ1)

Berdasarkan hasil analisis terhadap 23 studi yang ditinjau, terlihat bahwa arsitektur berbasis Convolutional Neural Network (CNN) masih menjadi pendekatan utama dalam pengembangan sistem deteksi kantuk pengemudi. Namun, implementasi CNN dalam penelitian-penelitian tersebut menunjukkan variasi yang cukup beragam, baik dari sisi kompleksitas model maupun integrasi dengan metode lain. Tabel 8 menyajikan distribusi arsitektur yang telah digunakan pada studi sebelumnya dari tahun 2021-2025.

Tabel 8 Distribusi Arsitektur CNN pada Penelitian Deteksi Kantuk Pengemudi

Kategori Arsitektur	Kriteria	Jumlah
CNN Klasik	CNN tanpa pretrained	3
Transfer learning CNN	MobileNet, ResNet, Inception, EfficientNet	10
Spatio-temporal CNN	CNN + GRU atau CNN + LSTM	5
Arsitektur Lanjutan	Hybrid kompleks (CNN+KNN, MCNN-FSA, dll)	4
Transformer-based	Vision Transformer (ViT)	1
Total		23

Gambar 3 menggambarkan tren masing-masing kategori arsitektur dari tahun ke tahun. Pendekatan *transfer learning* muncul sebagai kategori dominan, digunakan dalam 10 studi. Model seperti MobileNetV2, ResNet, InceptionV3, dan EfficientNet banyak dipilih karena kemampuannya dalam memanfaatkan pengetahuan dari *dataset* besar, sehingga mampu meningkatkan performa model meskipun dilatih pada *dataset* yang relatif terbatas (Rajput et al., 2025; Salem & Waleed, 2024). Selain itu, pendekatan ini juga lebih efisien dari sisi waktu pelatihan dan cenderung lebih adaptif terhadap variasi kondisi lingkungan. Pendekatan dengan jumlah terbanyak kedua adalah spatio-temporal yang digunakan dalam 5 studi. Pendekatan ini mencakup model CNN–LSTM, CNN–GRU, serta 3D-CNN yang dirancang untuk menangkap dinamika temporal dari perilaku pengemudi, seperti pola berkedip, durasi mata tertutup, dan gerakan kepala (Alameen & Alhothali, 2023; Namburi et al., 2024). Meskipun menggunakan mekanisme yang berbeda, pendekatan-pendekatan tersebut secara umum bertujuan untuk mengintegrasikan informasi spasial dan temporal guna meningkatkan akurasi deteksi kantuk.



Gambar 3 Tren Arsitektur CNN Untuk Deteksi Kantuk Pengemudi

Pendekatan CNN klasik masih digunakan dalam 3 studi, umumnya dengan arsitektur yang lebih sederhana tanpa memanfaatkan model pralatih. Meskipun lebih ringan secara komputasi, pendekatan ini cenderung memiliki keterbatasan dalam menangani variasi kondisi dunia nyata, terutama pada aspek pencahayaan dan pose wajah (Ahmed et al., 2023; Sohail et al., 2024). Selain itu, beberapa penelitian mengembangkan arsitektur lanjutan dengan mengintegrasikan mekanisme tambahan di luar konvolusi standar, seperti *attention mechanism*, metode klasifikasi *non-deep learning*, serta teknik optimasi berbasis fitur. Pendekatan ini muncul dalam 4 studi dan menunjukkan upaya untuk meningkatkan performa model melalui eksplorasi *pipeline* yang lebih kompleks (Huang, 2025; Winarno et al., 2025).

Terakhir, terdapat pendekatan berbasis *transformer* dalam tinjauan ini yang terintegrasi dalam skema hibrida bersama CNN. Studi tersebut menggabungkan Vision Transformer dengan ekstraksi fitur berbasis CNN untuk meningkatkan representasi spasial dan global secara simultan (T.-C. Phan et al., 2025). Oleh karena itu, kategori ini lebih tepat diklasifikasikan sebagai *CNN–Transformer hybrid* daripada sebagai arsitektur *transformer* murni. Penyesuaian ini memberikan representasi yang lebih akurat terhadap arah perkembangan metodologi dalam deteksi kantuk pengemudi berbasis *deep learning*. Hasil ini menunjukkan bahwa penelitian deteksi kantuk berbasis CNN tidak hanya berfokus pada peningkatan akurasi melalui arsitektur yang lebih dalam, tetapi juga mengarah pada integrasi informasi temporal serta eksplorasi pendekatan baru untuk meningkatkan *kerobustan* model. Hal ini menandakan bahwa perkembangan arsitektur



tidak lagi bersifat linear, melainkan menunjukkan keberagaman dan kemampuan adaptasi terhadap kebutuhan penerapan di dunia nyata.

3.4 Reliabilitas dan Akurasi Model (RQ2)

Dalam tinjauan ini, akurasi digunakan sebagai metrik utama karena dilaporkan secara konsisten oleh seluruh studi, sehingga memungkinkan perbandingan yang lebih sistematis antar studi. Model yang ditinjau dalam berbagai studi menunjukkan variasi performa yang dipengaruhi oleh jenis arsitektur, strategi *preprocessing*, serta karakteristik *dataset* yang digunakan. Rentang akurasi model yang dianalisis dalam tinjauan ini menunjukkan variasi yang cukup signifikan, yaitu berkisar antara sekitar 63% hingga 99,71%, tergantung pada kompleksitas arsitektur, kualitas *dataset*, serta kondisi pengujian yang digunakan. Nilai yang lebih rendah umumnya ditemukan pada studi dengan keterbatasan data atau skenario evaluasi yang lebih kompleks. Tabel 11 pada Lampiran A telah menunjukkan akurasi dari masing-masing arsitektur CNN beserta *dataset* yang digunakan.

3.4.1 Benchmark Akurasi Arsitektur CNN

Sebagian besar penelitian menunjukkan bahwa model berbasis CNN mampu mencapai performa yang tinggi dalam deteksi kantuk pengemudi. Arsitektur CNN sederhana, seperti kombinasi CNN dengan metode deteksi wajah seperti Viola–Jones menghasilkan akurasi 96,69% (Sohail et al., 2024). Hal ini menunjukkan bahwa bahkan arsitektur dasar tetap kompetitif ketika didukung oleh *preprocessing* yang tepat.

Pendekatan berbasis *transfer learning* menjadi salah satu strategi yang paling dominan. Model seperti MobileNetV2, InceptionV3, dan ResNet50V2 secara konsisten menunjukkan akurasi di atas 96%, bahkan mencapai hingga 99% pada *dataset* dengan skala besar dan variasi kondisi yang beragam (Florez et al., 2023; Rajput et al., 2025; Salem & Waleed, 2024; Venkateswarlu & Chirra, 2024; Yang & Yi, 2024). ResNet-50 dan variannya, termasuk yang dilengkapi mekanisme *attention*, juga menunjukkan performa stabil pada rentang 93–97%, terutama dalam menghadapi variasi pencahayaan dan pose wajah (Huang, 2025; Minhas et al., 2022; A.-C. Phan et al., 2021).

Dalam konteks pemodelan temporal, pendekatan berbasis CNN dan Long Short-Term Memory (LSTM) yang menggabungkan informasi spasial dan temporal menunjukkan kinerja yang konsisten dengan akurasi pada kisaran 93–96% (Alameen & Alhothali, 2023; Namburi et al., 2024). Integrasi informasi temporal dari data video membantu meningkatkan stabilitas prediksi, khususnya dalam mendeteksi perubahan kondisi kantuk secara bertahap. Model berbasis 3D-CNN menunjukkan performa yang bervariasi, terutama ketika digunakan sebagai arsitektur tunggal. Pada beberapa studi, pendekatan ini menghasilkan akurasi yang relatif lebih rendah, yaitu sekitar 75–77%, yang umumnya disebabkan oleh tingginya kompleksitas model serta keterbatasan *dataset* (Adhithyaa et al., 2023).

Perkembangan arsitektur yang lebih modern juga menunjukkan peningkatan performa yang signifikan. Model seperti Vis-DenNet dan pendekatan berbasis transformer mencapai akurasi hingga 98,94%, sementara kombinasi EfficientNet dengan metode klasifikasi K-Nearest Neighbors (KNN) bahkan mencapai akurasi 99,52% pada himpunan data tertentu (T.-C. Phan et al., 2025; Winarno et al., 2025). Di sisi lain, EfficientNetB0 yang digabungkan dengan Gated Recurrent Unit (GRU) menunjukkan kinerja yang jauh lebih rendah, yaitu sekitar 63%, terutama pada kondisi simulasi (Magán et al., 2022). Secara umum, banyak studi melaporkan akurasi di atas 90%, dengan beberapa model mencapai lebih dari 95% terutama pada pendekatan berbasis *transfer learning* dan arsitektur hibrida. Bahkan, pendekatan berbasis optimasi seperti kombinasi Genetic Algorithm dan *transfer learning* mampu mencapai akurasi hingga 99,79% pada *dataset* tertentu, meskipun dengan konsekuensi peningkatan kompleksitas komputasi dan waktu pelatihan (Jebraeily et al., 2024). Tren penelitian juga mengarah pada integrasi mekanisme *attention*, pemodelan temporal, serta optimasi fitur untuk meningkatkan performa model. Untuk menyederhanakan interpretasi terhadap variasi performa yang telah dibahas, ringkasan distribusi akurasi berdasarkan kategori arsitektur disajikan pada Tabel 9.



Tabel 9 Ringkasan *Benchmark* Akurasi Berdasarkan Kategori Arsitektur

Kategori Arsitektur	Rentang Akurasi	Rata-rata	Kondisi Dominan
CNN Klasik	96.69–99,23%	~97,7%	Terkontrol
Transfer Learning	85–99,79%	~95,6%	Terkontrol / Semi-real
Spatio-temporal	63–96,69%	~87%	Bervariasi
Arsitektur Lanjutan	93,5–99,52%	~97,1%	Terkontrol
Transformer-based	98,94%	98,94%	Multi-dataset

Arsitektur CNN klasik dan model berbasis transfer learning menunjukkan rata-rata akurasi yang tinggi, masing-masing sekitar 97,7% dan 95,6%, terutama pada kondisi pengujian yang terkontrol. Hal ini mengindikasikan bahwa pendekatan berbasis fitur spasial masih sangat efektif ketika variabilitas data relatif rendah. Sebaliknya, pendekatan spatio-temporal menunjukkan rata-rata akurasi yang lebih rendah (~87%) dengan variasi performa yang lebih besar. Temuan ini mengindikasikan bahwa meskipun integrasi informasi temporal secara teoretis meningkatkan kemampuan model dalam menangkap dinamika kantuk, kinerjanya sangat bergantung pada kualitas dan ukuran himpunan data berbasis urutan yang digunakan. Menariknya, arsitektur lanjutan dan *transformer-based* menunjukkan performa yang tinggi (>97%), namun hasil ini sebagian besar diperoleh dari skenario pengujian yang terbatas atau *dataset* tertentu. Hal ini menunjukkan bahwa performa tinggi pada kategori ini belum tentu mencerminkan generalisasi yang kuat di lingkungan dunia nyata.

3.4.2 Validitas Perbandingan Performa Antar Studi

Hasil *benchmark* pada Tabel 9 memberikan gambaran mengenai kecenderungan performa berbagai kategori arsitektur CNN dalam deteksi kantuk pengemudi. Namun, interpretasi terhadap nilai akurasi perlu dilakukan secara hati-hati karena setiap penelitian menggunakan rancangan eksperimen yang berbeda. Perbedaan tersebut mencakup penggunaan *dataset*, jumlah sampel, teknik *preprocessing*, *region of interest* (ROI), serta strategi pembagian data pelatihan dan pengujian. Selain itu, karakteristik *dataset* yang digunakan juga bervariasi, baik dari sisi kondisi pencahayaan, sudut pengambilan gambar, variasi subjek, maupun skenario pengambilan data. Oleh karena itu, hasil *benchmark* pada penelitian ini lebih tepat digunakan untuk mengidentifikasi kecenderungan performa berbagai kategori arsitektur CNN daripada sebagai dasar untuk menetapkan satu arsitektur yang secara universal memiliki performa terbaik.

Pengaruh kondisi pengujian terhadap performa model terlihat pada beberapa studi yang ditinjau. Sebagai contoh, model 3D-CNN–LSTM yang mencapai akurasi 96% pada *dataset* YawDD mengalami penurunan hingga sekitar 75% ketika diuji pada *dataset* 3MDAD yang memiliki karakteristik pengambilan data berbeda (Alameen & Alhothali, 2023). Sebaliknya, pendekatan *Multiscale* CNN menunjukkan performa yang relatif konsisten pada dua *dataset*, yaitu YawDD dan NTHU-DDD, dengan akurasi di atas 98% (Vijaypriya & Uma, 2023). Meskipun beberapa penelitian telah melakukan evaluasi menggunakan lebih dari satu *dataset*, sebagian besar pengujian masih dilakukan pada *dataset* yang terbatas atau memiliki karakteristik tertentu, sehingga bukti mengenai kemampuan generalisasi model pada kondisi yang lebih beragam masih memerlukan pengujian lebih lanjut. Interpretasi performa juga perlu mempertimbangkan karakteristik *dataset* dan keterbatasan yang dilaporkan pada masing-masing penelitian. Beberapa studi yang melaporkan akurasi di atas 99% juga membahas beberapa keterbatasan, seperti ukuran *dataset* yang relatif kecil, penggunaan *dataset* yang dikembangkan sendiri, maupun pengujian pada kondisi laboratorium yang terkontrol (Florez et al., 2023; Jebraeily et al., 2024; Venkateswarlu & Chirra, 2024; Winarno et al., 2025). Temuan tersebut menunjukkan bahwa nilai akurasi yang tinggi tetap perlu diinterpretasikan sesuai dengan konteks eksperimen dan karakteristik data yang digunakan. Dengan demikian, hasil *benchmark* dalam penelitian ini memberikan gambaran mengenai perkembangan performa berbagai arsitektur CNN berdasarkan kondisi pengujian yang dilaporkan dalam literatur, bukan sebagai perbandingan yang absolut untuk menentukan arsitektur terbaik.



3.5 Dataset dan Metode *Preprocessing* (RQ3)

3.5.1 Dataset

Bagian ini memberikan tinjauan terstruktur terhadap *dataset* yang digunakan dalam berbagai studi terkait deteksi kantuk. Aksesibilitas *dataset* memiliki pengaruh besar terhadap reproduktibilitas, komparabilitas, serta skalabilitas model *deep learning*. Berdasarkan berbagai penelitian yang dianalisis, dua kategori utama *dataset* dapat diidentifikasi: *dataset* publik (*open-source*) dan *dataset* khusus (*proprietary* atau *private*) yang dikembangkan secara mandiri untuk kebutuhan eksperimental tertentu. Tabel 10 menyajikan distribusi *dataset* pada studi deteksi kantuk menggunakan arsitektur CNN.

Tabel 10 Distribusi Dataset Penelitian Deteksi Kantuk

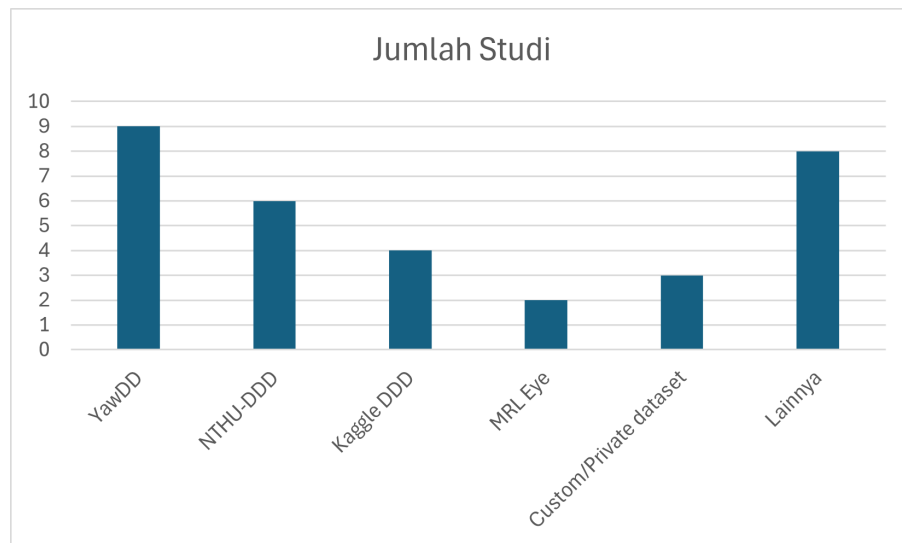
<i>Dataset</i>	Jumlah Studi (n)	Jenis
YawDD	9	Publik
NTHU-DDD	6	Publik
Kaggle DDD	4	Publik
MRL Eye	2	Publik
DROZY	1	Publik
UTA-RLDD	1	Publik
NITYMED	1	Publik
Driver Drowsiness Dataset	1	Publik
Health Informatics	1	Publik
SUST-DDD	1	Publik
KEC-DDD	1	Privat
CJI/RMFD	1	Privat
CEW/Glasses/OC	1	Privat
Custom/Private dataset	3	Privat

Gambar 4 menunjukkan distribusi *dataset* yang digunakan, di mana penelitian dalam bidang ini masih didominasi oleh penggunaan *dataset* publik. YawDD dan NTHU-DDD muncul sebagai *dataset* yang paling dominan, diikuti oleh Kaggle Drowsiness Dataset dan MRL Eye Dataset. Dominasi ini bukan tanpa alasan. *Dataset* publik umumnya telah melalui proses kurasi dan anotasi yang relatif baik, serta menyediakan skenario yang relevan dengan kondisi kantuk, seperti mata tertutup, menguap, dan variasi perilaku pengemudi. Selain itu, beberapa *dataset* seperti NTHU-DDD juga menawarkan variasi kondisi pencahayaan dan lingkungan yang lebih beragam, sehingga sering dijadikan acuan dalam evaluasi model. Di sisi lain, penggunaan *dataset* privat masih terbatas dan biasanya dikembangkan untuk kebutuhan yang lebih spesifik, misalnya untuk menyesuaikan dengan kondisi lingkungan tertentu atau karakteristik pengguna lokal. *Dataset* jenis ini memiliki keunggulan dalam merepresentasikan kondisi dunia nyata secara lebih kontekstual. Namun, keterbatasan dalam jumlah data, variasi kondisi, serta standar anotasi yang belum tentu konsisten sering kali menjadi kendala, terutama ketika hasil penelitian ingin dibandingkan dengan studi lain.

Untuk memberikan gambaran yang lebih konkret, dilakukan analisis kuantitatif sederhana dengan membandingkan performa model berdasarkan jenis *dataset* yang digunakan. Berdasarkan data pada Tabel 11 pada Lampiran A, model yang dievaluasi menggunakan *dataset* publik umumnya melaporkan akurasi tinggi, banyak di antaranya berada di atas 90% dan bahkan mendekati 99% pada beberapa studi. Sementara itu, model yang menggunakan *dataset* privat menunjukkan variasi performa yang lebih besar, dengan akurasi berkisar dari sekitar 75% hingga di atas 90%. Variasi ini menunjukkan bahwa perbedaan performa tidak hanya dipengaruhi oleh jenis *dataset*, tetapi juga oleh karakteristik data, ukuran sampel, serta desain eksperimen yang digunakan. *Dataset* publik biasanya memiliki jumlah sampel yang lebih besar, anotasi yang lebih konsisten, serta telah digunakan secara luas dalam berbagai penelitian, sehingga model cenderung terlatih dengan baik terhadap pola yang ada. Sebaliknya, *dataset* privat sering kali lebih terbatas dan



belum tentu mencakup variasi kondisi yang cukup luas, sehingga performa model menjadi lebih fluktuatif.



Gambar 4 Distribusi Dataset Pada Penelitian Untuk Deteksi Kantuk Pengemudi

Meskipun demikian, akurasi yang tinggi pada *dataset* publik belum tentu mencerminkan performa model ketika diterapkan di dunia nyata. Beberapa studi menunjukkan bahwa model yang bekerja sangat baik pada satu *dataset* dapat mengalami penurunan performa ketika diuji pada *dataset* lain atau pada kondisi lingkungan yang berbeda. Dalam konteks ini, *dataset* privat justru memiliki nilai penting karena mampu merepresentasikan skenario penggunaan yang lebih realistis, meskipun dengan segala keterbatasannya. Perlu juga dicatat bahwa perbandingan ini bersifat indikatif, mengingat adanya variasi dalam desain eksperimen, teknik *preprocessing*, serta protokol evaluasi antar studi. Faktor-faktor tersebut dapat memengaruhi nilai akurasi yang dilaporkan. Oleh karena itu, ke depannya diperlukan pendekatan evaluasi yang lebih komprehensif, seperti pengujian lintas *dataset* (*cross-dataset evaluation*), agar model yang dikembangkan tidak hanya unggul pada satu *dataset* tertentu, tetapi juga memiliki kemampuan adaptasi yang baik terhadap berbagai kondisi dunia nyata.

3.5.2 Metode Preprocessing

Tahapan *preprocessing* juga memiliki peran yang sangat penting dalam menentukan performa model deteksi kantuk. Dalam beberapa studi, perbedaan performa antar studi tidak hanya dipengaruhi oleh arsitektur model, tetapi juga oleh strategi *preprocessing* yang digunakan. Berdasarkan Tabel 11 pada Lampiran A, sebagian besar penelitian memulai *pipeline* dengan deteksi wajah dan ekstraksi *region of interest* (ROI), khususnya pada area mata dan mulut yang berkaitan langsung dengan indikator utama kantuk, seperti durasi mata tertutup dan aktivitas menguap.

Metode yang paling umum digunakan pada studi sebelumnya adalah DLib 68-*point landmark* dan Haar cascade (Majeed et al., 2023; Salem & Waleed, 2024; Sohail et al., 2024; Yang & Yi, 2024). Pendekatan ini relatif ringan dan cukup efektif untuk mengisolasi area penting wajah, sehingga sering dipilih untuk sistem yang membutuhkan efisiensi, termasuk implementasi *real-time*. Namun, metode ini masih rentan terhadap variasi kondisi, seperti perubahan pencahayaan, sudut wajah, dan penggunaan kacamata. Hal ini sejalan dengan temuan pada beberapa studi yang melaporkan penurunan akurasi ketika model diuji pada kondisi yang lebih bervariasi.

Seiring perkembangan teknologi, beberapa penelitian mulai beralih ke metode yang dapat meningkatkan kemampuan agar lebih akurat. Salah satunya adalah MediaPipe Face Mesh yang



mampu mendeteksi *landmark* wajah secara lebih detail, sehingga membantu menghasilkan fitur yang lebih stabil (Florez et al., 2023). Di sisi lain, pendekatan berbasis deteksi objek seperti YOLOv8 memberikan fleksibilitas dalam mendeteksi wajah pada berbagai posisi dan skala, serta lebih adaptif untuk skenario *real-time* yang dinamis (T.-C. Phan et al., 2025). Meskipun demikian, peningkatan presisi ini biasanya diikuti oleh kebutuhan komputasi yang lebih besar, sehingga memunculkan *trade-off* yang perlu diperhitungkan antara akurasi dan efisiensi model.

Pada tahapan selanjutnya, metode seperti normalisasi, *resizing*, dan augmentasi data juga hampir selalu diterapkan dalam berbagai studi. *Resizing* dilakukan untuk menyesuaikan ukuran input dengan kebutuhan model, sementara normalisasi membantu menjaga kestabilan distribusi data. Augmentasi data seperti rotasi, *flipping*, dan variasi pencahayaan, terbukti mampu meningkatkan keragaman data dan membantu model menjadi lebih *robust*. Beberapa penelitian juga mencoba pendekatan yang lebih spesifik dalam meningkatkan kualitas citra, seperti penggunaan *bilateral filter* dan *Dual-Tree Complex Wavelet Transform* (DTCWT) (Vijaypriya & Uma, 2023), serta *histogram equalization* pada (Magán et al., 2022). Pendekatan ini bertujuan untuk memperbaiki kontras dan mengurangi *noise* pada citra. Namun, hasil yang diperoleh menunjukkan bahwa peningkatan kompleksitas pada tahap *preprocessing* tidak selalu menghasilkan peningkatan akurasi yang konsisten.

Secara umum, *pipeline preprocessing* yang terlalu sederhana cenderung menghasilkan model yang sensitif terhadap gangguan eksternal, sementara *pipeline* yang terlalu kompleks tidak selalu memberikan peningkatan performa yang sebanding. Oleh karena itu, keseimbangan antara efisiensi dan kualitas ekstraksi fitur menjadi aspek yang sangat penting dalam perancangan sistem deteksi kantuk. Selain itu, variasi dalam strategi *preprocessing* juga menjadi salah satu faktor utama yang menjelaskan perbedaan performa antar studi, terutama ketika model dihadapkan pada kondisi dunia nyata yang lebih beragam dan tidak terkontrol.

3.6 Tantangan dan Limitasi (RQ4)

Berdasarkan analisis terhadap 23 studi yang dirangkum pada Tabel 11 pada Lampiran A, berbagai tantangan dan keterbatasan masih ditemukan dalam pengembangan sistem deteksi kantuk berbasis CNN. Keterbatasan tersebut tidak hanya berkaitan dengan aspek teknis model. Keterbatasan tersebut juga mencakup hambatan implementasi di dunia nyata serta isu terkait *robustness* dan efisiensi sistem.

3.6.1 Tantangan Teknis

Sebagian besar keterbatasan teknis yang teridentifikasi dalam studi berkaitan dengan sensitivitas model terhadap kondisi lingkungan, kualitas *dataset*, serta kemampuan generalisasi. Berdasarkan Tabel 11 pada Lampiran A, banyak penelitian secara eksplisit melaporkan bahwa performa model menurun pada kondisi pencahayaan rendah, posisi kamera, serta penggunaan aksesoris seperti kacamata (Alameen & Alhothali, 2023; Kumral & Küçükmanisa, 2022; Rajput et al., 2025; Sohail et al., 2024). Selain itu, beberapa studi melaporkan keterbatasan pada *dataset* yang digunakan, dari segi jumlah sampel, distribusi kelas yang tidak seimbang antara kondisi mengantuk dan tidak mengantuk, serta kurangnya variasi demografis (Adhithyaa et al., 2023; Namburi et al., 2024; A.-C. Phan et al., 2021). Kondisi-kondisi ini mengindikasikan kemungkinan terjadinya *overfitting* atau keterbatasan generalisasi, yang disebabkan oleh data pelatihan yang belum sepenuhnya merepresentasikan variasi dunia nyata. Di samping itu, tidak semua penelitian menerapkan teknik validasi yang memadai, seperti *k-fold cross-validation*, sehingga reliabilitas performa model menjadi sulit untuk diverifikasi secara menyeluruh.

Keterbatasan juga muncul pada tahap pra-pemrosesan dan ekstraksi fitur wajah. Sistem berbasis CNN sangat bergantung pada keberhasilan deteksi wajah dan *landmark*. Ketika terjadi *occlusion* akibat kacamata, tangan, atau bayangan, performa sistem dapat menurun secara signifikan (Al-Gburi et al., 2025; Delwar et al., 2025; Majeed et al., 2023). Selain itu, model yang hanya memanfaatkan citra statis cenderung kehilangan informasi temporal yang penting dalam mendeteksi pola kantuk, seperti durasi mata tertutup atau frekuensi kedipan. Hal ini mendorong



penggunaan arsitektur hibrida seperti CNN–LSTM atau 3D-CNN untuk menangkap dinamika temporal (Alameen & Alhothali, 2023; Kumral & Küçükmanisa, 2022). Dari sisi komputasi, kompleksitas model juga menjadi tantangan tersendiri. Beberapa arsitektur memerlukan sumber daya komputasi yang tinggi, termasuk penggunaan GPU untuk proses pelatihan (Jebraeily et al., 2024; Nandyal & Sharanabasappa, 2024; T.-C. Phan et al., 2025; Winarno et al., 2025). Pada perangkat dengan keterbatasan sumber daya, seperti CPU atau sistem tertanam, waktu pelatihan dan inferensi menjadi lebih lambat, sehingga membatasi potensi implementasi secara *real-time*.

3.6.2 Hambatan Implementasi di Dunia Nyata

Meskipun banyak model menunjukkan performa yang baik dalam lingkungan terkontrol, implementasi di dunia nyata masih menghadapi berbagai tantangan. Salah satu kendala utama yang sering dilaporkan adalah sensitivitas model terhadap kondisi lingkungan, seperti pencahayaan ekstrem, posisi kamera, serta variasi orientasi wajah pengemudi. Pada kondisi malam hari atau pencahayaan rendah, fitur penting seperti area mata menjadi sulit dideteksi secara akurat, sehingga meningkatkan kemungkinan kesalahan klasifikasi. Selain itu, penggunaan aksesoris seperti kacamata atau masker, serta perubahan posisi kepala, dapat mengganggu proses ekstraksi fitur wajah. Faktor-faktor ini menyebabkan penurunan performa model ketika diterapkan di lingkungan yang lebih dinamis dibandingkan kondisi laboratorium.

Keterbatasan perangkat keras juga menjadi hambatan dalam implementasi sistem secara *real-time*. Banyak model masih memerlukan komputasi yang relatif tinggi, sehingga sulit diintegrasikan ke dalam perangkat berdaya rendah atau sistem kendaraan secara langsung. Selain itu, sebagian besar studi masih dilakukan menggunakan data video statis atau lingkungan simulatif, sehingga belum sepenuhnya merepresentasikan kondisi berkendara yang sesungguhnya, seperti getaran kendaraan, perubahan pencahayaan yang cepat, serta interaksi kompleks antara pengemudi dan lingkungan. Ketergantungan pada kamera RGB standar tanpa dukungan sensor tambahan, seperti *infrared*, juga menjadi keterbatasan, terutama dalam kondisi pencahayaan yang tidak stabil. Di samping itu, aspek optimasi waktu inferensi dan desain sistem yang modular masih belum banyak dibahas, padahal hal ini penting untuk memastikan sistem dapat berjalan secara efisien dalam implementasi nyata.

3.6.3 Strategi untuk Meningkatkan *Robustness* dan Aplikabilitas Model

Berbagai strategi telah diusulkan dalam literatur untuk meningkatkan performa dan kemampuan generalisasi model CNN. Salah satu pendekatan yang paling umum adalah augmentasi data, yang bertujuan untuk meningkatkan keragaman data melalui transformasi seperti rotasi, variasi pencahayaan, *flipping*, serta penambahan *noise*. Pendekatan ini membantu model mengenali pola wajah dalam berbagai kondisi yang lebih beragam. Selain itu, penggunaan *transfer learning* dari model pralatih seperti MobileNetV2, VGG16, dan ResNet50 telah terbukti efektif dalam meningkatkan performa, khususnya pada *dataset* dengan jumlah sampel terbatas. Arsitektur hibrida, seperti CNN–LSTM atau CNN–Transformer, juga semakin banyak digunakan untuk menggabungkan informasi spasial dan temporal secara simultan. Dari sisi efisiensi, pengembangan arsitektur ringan (*lightweight models*) serta teknik kompresi model seperti *pruning*, *quantization*, dan *knowledge distillation* menjadi solusi potensial untuk memungkinkan implementasi pada perangkat dengan sumber daya terbatas. Pendekatan ini bertujuan untuk mengurangi kompleksitas model tanpa mengorbankan performa secara signifikan. Evaluasi lintas *dataset* (*cross-dataset evaluation*) serta pengujian dalam kondisi dunia nyata untuk memastikan keandalan model juga sangat penting. Strategi ini menunjukkan arah pengembangan menuju sistem deteksi kantuk yang tidak hanya akurat, tetapi juga adaptif, efisien, dan siap diterapkan dalam sistem kendaraan cerdas.

3.7 Keterbatasan Penelitian

Penelitian ini memiliki beberapa keterbatasan yang perlu dipertimbangkan dalam menginterpretasikan hasil sintesis. Pertama, proses penelusuran literatur hanya dilakukan pada dua basis data, yaitu Scopus dan IEEE Xplore. Meskipun keduanya merupakan sumber publikasi



bereputasi yang banyak digunakan dalam bidang ilmu komputer dan kecerdasan buatan, penggunaan dua basis data tersebut berpotensi menyebabkan beberapa studi relevan yang terindeks pada basis data lain tidak tercakup dalam tinjauan ini.

Kedua, proses identifikasi artikel tidak menggunakan teknik *backward* maupun *forward snowballing*. Seluruh artikel diperoleh melalui strategi pencarian berbasis kata kunci dan proses penyaringan berdasarkan kriteria inklusi serta eksklusi. Akibatnya, terdapat kemungkinan bahwa beberapa penelitian yang relevan tetapi tidak teridentifikasi melalui kata kunci pencarian tidak termasuk dalam proses sintesis.

Ketiga, penelitian ini hanya mempertimbangkan artikel *full-text* yang dapat diakses secara terbuka (*open access*). Pendekatan ini dipilih untuk menjamin transparansi proses ekstraksi data dan memungkinkan seluruh informasi metodologis dapat dianalisis secara konsisten. Namun demikian, keputusan tersebut juga berpotensi mengecualikan beberapa penelitian berkualitas yang tersedia melalui akses berbayar.

Terakhir, sintesis pada penelitian ini bersifat naratif karena tingginya heterogenitas antar studi, baik dari sisi arsitektur CNN, karakteristik dataset, strategi preprocessing, maupun protokol evaluasi yang digunakan. Kondisi tersebut menyebabkan hasil penelitian tidak memungkinkan untuk digabungkan melalui pendekatan *meta-analysis*. Dengan demikian, temuan yang dihasilkan lebih menekankan pada identifikasi pola, tren, serta kesenjangan penelitian dibandingkan estimasi efek kuantitatif secara statistik.

3.8 Arah Penelitian Selanjutnya

Temuan dalam tinjauan ini menunjukkan bahwa pengembangan penelitian selanjutnya perlu diarahkan pada peningkatan kemampuan generalisasi model melalui strategi evaluasi yang lebih representatif. Dari 23 studi yang dianalisis, sebagian besar masih menggunakan satu *dataset* sebagai dasar pengujian, sedangkan hanya enam studi yang melakukan evaluasi pada lebih dari satu *dataset*. Dominasi penggunaan *dataset* seperti YawDD juga menunjukkan bahwa validasi model masih banyak dilakukan pada lingkungan yang relatif terkontrol. Oleh karena itu, penelitian mendatang perlu memperluas penggunaan *dataset* dengan karakteristik yang lebih beragam atau menerapkan evaluasi lintas *dataset* (*cross-dataset evaluation*) untuk memperoleh gambaran yang lebih realistis mengenai ketahanan model terhadap perubahan kondisi operasional.

Standar evaluasi model juga perlu diperkuat dikarenakan sebagian besar penelitian masih menggunakan pembagian data pelatihan dan pengujian secara *single split*. Sementara itu, penggunaan teknik validasi yang lebih *robust*, seperti *k-fold cross-validation*, masih relatif terbatas. Penerapan protokol evaluasi yang lebih konsisten akan meningkatkan reliabilitas hasil serta memudahkan perbandingan performa antarpelatihan.

Di sisi lain, tren penggunaan arsitektur hibrida yang menggabungkan CNN dengan mekanisme spasio-temporal, seperti LSTM, GRU, maupun Transformer, menunjukkan potensi dalam memanfaatkan informasi temporal dari video. Namun, hasil sintesis memperlihatkan bahwa peningkatan kompleksitas arsitektur tidak selalu menghasilkan performa yang lebih baik apabila kualitas dan keragaman data belum memadai. Oleh karena itu, pengembangan model sebaiknya diimbangi dengan penyediaan *dataset* video yang lebih representatif terhadap berbagai kondisi pencahayaan, sudut pandang kamera, penggunaan aksesoris wajah, serta karakteristik pengemudi yang beragam.

Efisiensi komputasi juga tetap menjadi aspek penting untuk mendukung implementasi pada perangkat nyata. Beberapa penelitian mulai mengeksplorasi arsitektur ringan seperti MobileNet dan EfficientNet, namun masih diperlukan upaya lebih lanjut untuk memperoleh keseimbangan antara akurasi, kecepatan inferensi, dan kebutuhan sumber daya komputasi. Dengan demikian, penelitian selanjutnya diharapkan tidak hanya berorientasi pada peningkatan akurasi, tetapi juga pada pengembangan sistem yang lebih robust, efisien, dan tervalidasi pada kondisi berkendara yang mendekati situasi dunia nyata.



4. KESIMPULAN

Convolutional Neural Network (CNN) merupakan fondasi utama dalam pengembangan model deteksi kantuk pengemudi berbasis *deep learning*. Hal ini disebabkan oleh kemampuannya dalam mengekstraksi fitur spasial secara efektif serta fleksibilitasnya untuk diintegrasikan dengan berbagai arsitektur lain. Berdasarkan analisis terhadap 23 studi yang dipublikasikan pada periode 2021–2025, pendekatan berbasis *transfer learning* dan arsitektur hibrida, seperti CNN–LSTM, merupakan pendekatan yang paling banyak digunakan dan umumnya mampu mencapai akurasi di atas 90% pada berbagai *dataset*. Temuan ini menunjukkan bahwa integrasi CNN dengan mekanisme ekstraksi fitur yang lebih adaptif maupun pemodelan informasi temporal berpotensi meningkatkan performa sistem deteksi kantuk.

Meskipun demikian, hasil sintesis juga menunjukkan bahwa tingginya akurasi yang dilaporkan belum sepenuhnya mencerminkan keandalan model pada kondisi operasional yang sesungguhnya. Performa model masih sangat dipengaruhi oleh karakteristik *dataset*, strategi *preprocessing* data, serta protokol evaluasi yang digunakan. Berbagai penelitian masih melaporkan keterbatasan berupa sensitivitas terhadap pencahayaan rendah, *occlusions*, variasi pose wajah, serta penurunan performa ketika model diuji pada kondisi yang berbeda dari data pelatihannya. Selain itu, peningkatan kompleksitas arsitektur sering kali diikuti oleh kebutuhan komputasi yang lebih tinggi, sehingga menimbulkan pertukaran antara akurasi dan efisiensi implementasi pada sistem *real-time*.

Berdasarkan hasil tinjauan ini, masih terdapat beberapa aspek yang perlu difokuskan pada penelitian berikutnya. Sebanyak 9 dari 23 studi menggunakan *dataset* YawDD sebagai data utama, sedangkan hanya 6 studi yang mengevaluasi model pada lebih dari satu *dataset*. Kondisi tersebut menunjukkan bahwa pengujian model masih didominasi oleh lingkungan data yang terbatas sehingga kemampuan model untuk bekerja secara konsisten pada data yang berbeda belum banyak dibuktikan. Oleh karena itu, penelitian selanjutnya perlu memanfaatkan *dataset* yang lebih beragam dan menerapkan *cross-dataset evaluation* agar kemampuan generalisasi model dapat dievaluasi secara lebih menyeluruh. Selain itu, penggunaan k-fold cross-validation juga perlu dipertimbangkan karena sebagian besar studi masih mengandalkan satu kali pembagian data latih dan data uji. Di sisi lain, pengembangan model tidak hanya perlu diarahkan pada peningkatan akurasi, tetapi juga pada efisiensi komputasi serta ketahanan terhadap variasi pencahayaan, pose wajah, dan *occlusions*, sehingga sistem yang dihasilkan lebih siap diterapkan pada skenario berkendara yang nyata.

DAFTAR PUSTAKA

- Adhithyaa, N., Tamilarasi, A., Sivabalaselvamani, D., & Rahunathan, L. (2023). Face Positioned Driver Drowsiness Detection Using Multistage Adaptive 3D Convolutional Neural Network. *Information Technology and Control*, 52(3), 713–730. <https://doi.org/10.5755/j01.itc.52.3.33719>
- Ahmed, M. I. B., Alabdulkarem, H., Alomair, F., Aldossary, D., Alahmari, M., Alhumaidan, M., Alrassan, S., Rahman, A., Youldash, M., & Zaman, G. (2023). A Deep-Learning Approach to Driver Drowsiness Detection. *Safety*, 9(3), Article ID: 65. <https://doi.org/10.3390/safety9030065>
- Alameen, S. A., & Alhothali, A. M. (2023). A Lightweight Driver Drowsiness Detection System Using 3DCNN with LSTM. *Computer Systems Science and Engineering*, 44(1), 895–912. <https://doi.org/10.32604/csse.2023.024643>
- Al-Gburi, S. H., Al-Sammak, K. A., Marghescu, I., Oprea, C. C., Drăgulescu, A.-M. C., Alduais, N. A. M., Alheeti, K. M. A., & Al-Sammak, N. A. H. (2025). EffRes-DrowsyNet: A Novel Hybrid Deep Learning Model Combining EfficientNetB0 and ResNet50 for Driver Drowsiness Detection. *Sensors*, 25(12), Article ID: 3711. <https://doi.org/10.3390/s25123711>
- Carrera-Rivera, A., Larrinaga, F., & Lasa, G. (2022). Context-Awareness for the Design of Smart-Product Service Systems: Literature Review. *Computers in Industry*, 142, Article ID: 103730. <https://doi.org/10.1016/j.compind.2022.103730>



- Delwar, T. S., Singh, M., Mukhopadhyay, S., Kumar, A., Parashar, D., Lee, Y., Rahman, M. H., Sejan, M. A. S., & Ryu, J. Y. (2025). AI- and Deep Learning-Powered Driver Drowsiness Detection Method Using Facial Analysis. *Applied Sciences*, 15(3), Article ID: 1102. <https://doi.org/10.3390/app15031102>
- Florez, R., Palomino-Quispe, F., Coaquira-Castillo, R. J., Herrera-Levano, J. C., Paixão, T., & Alvarez, A. B. (2023). A CNN-Based Approach for Driver Drowsiness Detection by Real-Time Eye State Identification. *Applied Sciences*, 13(13), Article ID: 7849. <https://doi.org/10.3390/app13137849>
- Fonseca, T., & Ferreira, S. (2025). Drowsiness Detection in Drivers: A Systematic Review of Deep Learning-Based Models. *Applied Sciences*, 15(16), Article ID: 9018. <https://doi.org/10.3390/app15169018>
- Huang, Z. (2025). Integrating Attention Mechanisms and ResNet-50 For Enhanced Driver Sleepiness Detection. *Informatica*, 49(15), 135–144. <https://doi.org/10.31449/inf.v49i15.7977>
- Jebraeily, Y., Sharafi, Y., & Teshnehlab, M. (2024). Driver Drowsiness Detection Based on Convolutional Neural Network Architecture Optimization Using Genetic Algorithm. *IEEE Access*, 12, 45709–45726. <https://doi.org/10.1109/ACCESS.2024.3381999>
- Kumral, F., & Küçükmanisa, A. (2022). Temporal Analysis Based Driver Drowsiness Detection System Using Deep Learning Approaches. *Sakarya University Journal of Science*, 26(4), 710–719. <https://doi.org/10.16984/saufenbilder.1071863>
- Magán, E., Sesmero, M. P., Alonso-Weber, J. M., & Sanchis, A. (2022). Driver Drowsiness Detection by Applying Deep Learning Techniques to Sequences of Images. *Applied Sciences*, 12(3), Article ID: 1145. <https://doi.org/10.3390/app12031145>
- Majeed, F., Shafique, U., Safran, M., Alfarhood, S., & Ashraf, I. (2023). Detection of Drowsiness among Drivers Using Novel Deep Convolutional Neural Network Model. *Sensors*, 23(21), Article ID: 8741. <https://doi.org/10.3390/s23218741>
- Minhas, A. A., Jabbar, S., Farhan, M., & Najam ul Islam, M. (2022). A Smart Analysis of Driver Fatigue and Drowsiness Detection Using Convolutional Neural Networks. *Multimedia Tools and Applications*, 81(19), 26969–26986. <https://doi.org/10.1007/s11042-022-13193-4>
- Namburi, A., Sitpasert, P., & Duang-Onnam, W. (2024). A CNN-LSTM Approach for Accurate Drowsiness and Distraction Detection in Drivers. *ICIC Express Letters*, 18(9), 907–917. <https://doi.org/10.24507/icicel.18.09.907>
- Nandyal, S., & Sharanabasappa, S. (2024). Deep ResNet 18 and Enhanced Firefly Optimization Algorithm for On-Road Vehicle Driver Drowsiness Detection. *Journal of Autonomous Intelligence*, 7(3), Article ID: 975. <https://doi.org/10.32629/jai.v7i3.975>
- Phan, A.-C., Nguyen, N.-H.-Q., Trieu, T.-N., & Phan, T.-C. (2021). An Efficient Approach for Detecting Driver Drowsiness Based on Deep Learning. *Applied Sciences*, 11(18), Article ID: 8441. <https://doi.org/10.3390/app11188441>
- Phan, T.-C., Phan, A.-C., & Nguyen, N.-H.-Q. (2025). A Novel Approach of Drowsiness Levels Detection Using Vis-Net Combined with Facial Emotion. *Systems and Soft Computing*, 7, Article ID: 200288. <https://doi.org/10.1016/j.sasc.2025.200288>
- Rahmadiyah, R., & Widyanti, A. (2023). Prevalence of Drowsy Driving and Modeling Its Intention: An Indonesian Case Study. *Transportation Research Interdisciplinary Perspectives*, 19, Article ID: 100824. <https://doi.org/10.1016/j.trip.2023.100824>
- Rajput, I. S., Tyagi, S., Pandey, V., Upreti, A., & Kukreja, J. K. (2025). Enhancing Driver Safety with MobileNetV2-Based Transfer Learning for Drowsiness Detection. *Procedia Computer Science*, 259, 1239–1248. <https://doi.org/10.1016/j.procs.2025.04.079>
- Saleem, A. A., Siddiqui, H. U. R., Raza, M. A., Rustam, F., Dudley, S., & Ashraf, I. (2023). A Systematic Review of Physiological Signals Based Driver Drowsiness Detection Systems. *Cognitive Neurodynamics*, 17(5), 1229–1259. <https://doi.org/10.1007/s11571-022-09898-9>
- Salem, D., & Waleed, M. (2024). Drowsiness Detection in Real-Time via Convolutional Neural Networks and Transfer Learning. *Journal of Engineering and Applied Science*, 71(1), Article ID: 122. <https://doi.org/10.1186/s44147-024-00457-z>
- Sohail, A., Shah, A. A., Ilyas, S., & Alshammry, N. (2024). A CNN-based Deep Learning Framework for Driver's Drowsiness Detection. *International Journal of Advanced Computer Science and Applications*, 15(3), 169–178. <https://doi.org/10.14569/IJACSA.2024.0150317>



- Thoriq, M. F., Ramdhana, M. F., Rahma, D. N., Putri, N. A., Zahir, R. H., Mindara, G. P., & Giri, E. P. (2024). Implementasi Sistem Deteksi Kantuk Secara Real-Time Bagi Pengemudi Menggunakan Metode Eye Aspect Ratio. *Jurnal Sistem Informasi dan Ilmu Komputer*, 2(4), 70–85. <https://doi.org/10.59581/jusiik-widyakarya.v2i4.4226>
- Venkateswarlu, M., & Chirra, V. R. R. (2024). DrowsyDetectNet: Driver Drowsiness Detection Using Lightweight CNN with Limited Training Data. *IEEE Access*, 12(1), 110476–110491. <https://doi.org/10.1109/ACCESS.2024.3440585>
- Vijaypriya, V., & Uma, M. (2023). Facial Feature-Based Drowsiness Detection with Multi-Scale Convolutional Neural Network. *IEEE Access*, 11, 63417–63429. <https://doi.org/10.1109/ACCESS.2023.3288008>
- Winarno, S., Alzami, F., Naufal, M., Al Azies, H., Soeleman, M. A., & Malim, N. H. A. H. (2025). EfficientNet-KNN for Real-Time Driver Drowsiness Detection via Sequential Image Processing. *Ingénierie des Systèmes d'Information*, 30(7), 1703–1713. <https://doi.org/10.18280/isi.300703>
- Yang, E., & Yi, O. (2024). Enhancing Road Safety: Deep Learning-Based Intelligent Driver Drowsiness Detection for Advanced Driver-Assistance Systems. *Electronics*, 13(4), Article ID: 708. <https://doi.org/10.3390/electronics13040708>



LAMPIRAN A

Tabel 11 Analisis Studi Deteksi Kantuk Pengemudi yang Menggunakan CNN

Author	Model	Preprocessing	Akurasi	Dataset	Limitasi
Majeed et al. (2023)	Deep CNN vs CNN + LSTM	DLib 68-point landmark, ROI mulut (MAR), resize 64×64, normalisasi, augmentasi (rotasi, flipping, zoom, shear), split 70:15:15	96,69%	YawDD	Sensitif cahaya, pose, dan penggunaan kacamata
Alameen & Alhothali (2023)	3D-CNN LSTM	Resize 60×60×3, sequence 50 frame, augmentasi (28 transformasi: rotasi, flipping, brightness), tanpa ROI spesifik	96%	YawDD	Tanpa deteksi wajah otomatis, sensitif sudut kamera
Delwar et al. (2025)	MobileNet, CNN, VGG16	Haar cascade + landmark, ROI mata–mulut, resize 145×145, grayscale conversion, resizing, labeling 4 kelas	92–93%	Kaggle, Health Informatics	Sensitif terhadap cahaya rendah dan gagal deteksi kacamata hitam
Florez et al. (2023)	CNN (InceptionV3, VGG16, ResNet50V2 – TL)	MediaPipe Face Mesh (468 landmark), ROI mata + koreksi ROI, resize 112×112, normalisasi, augmentasi	99,31–99,71%	NITYMED	Dataset kecil, potensi overfitting, hanya fokus ROI mata, belum robust terhadap occlusion/kacamata
Sohail et al. (2024)	CNN + Viola-Jones	Viola–Jones face detection, SMOTE (class balancing), ekstraksi fitur statistik (Hahn moment, central moment)	96,69%	YawDD	Cahaya rendah, pose kepala
Magán et al. (2022)	EfficientNetB0 + GRU	DLib (HOG + SVM) face detection, Gaussian blur, histogram equalization, ImageNet mean subtraction,	63%	UTA-RLDD, YawDD	Akurasi rendah, tidak stabil, dan bergantung pada dataset simulasi



Author	Model	Preprocessing	Akurasi	Dataset	Limitasi
T.-C. Phan et al. (2025)	Vision Transformer + Variasi CNN	sequence ≈60 detik, fitur temporal YOLOv8 face detection, resize 224×224, augmentasi (rotasi, zoom), filtering frame berbasis emosi	98,94%	Kaggle, CUI, RMFD	Belum sepenuhnya real-time dan komputasi tinggi untuk implementasi
Al-Gburi et al. (2025)	EfficientNetB0 + ResNet50	Dlib 68 landmark, EAR & MAR, implicit normalization (ImageNet)	92–97%	SUST-DDD, NTHU-DDD, YawDD	Bergantung pada akurasi landmark wajah yang rentan pada variasi pose ekstrem, pencahayaan, dan occlusion Terbatas pada klasifikasi biner
Yang & Yi (2024)	ShuffleNet + ELM	Median filtering, tanpa ROI, resize tidak disebutkan secara eksplisit Min–max normalization, fitur EFF	97,05%	YawDD	
Nandyal & Sharanabasappa (2024)	CNN–EFF–ResNet18	Haar Cascade (face & eye detection), augmentasi (rotasi, zoom, flip), resize tidak disebutkan secara eksplisit	91,3%	NTHU-DDD	Komputasi tinggi serta bersiko overfitting akibat optimasi fitur yang intensif
Salem & Waleed (2024)	CNN, InceptionV3, MobileNetV2	92–99,9%	YAWDD, CEW, Glasses, MRL Eye, NTHU-DDD, OC-Dataset	Sensitif pencahayaan dan variasi kondisi wajah	
Namburi et al. (2024)	CNN–LSTM	MediaPipe FaceMesh (468 landmark) + Hand (21 titik), EAR, MAR, IOU, HeadPose, window 16 frame	93,60%	Private Dataset	Generalisasi terbatas karena menggunakan dataset privat
Huang (2025)	ResNet-50 + Attention	Resize 224×224, normalisasi, augmentasi (rotasi ±15°, flipping, contrast adjustment), feature scaling	93,5%	Driver Drowsiness Dataset	Perlu sensor tambahan
Rajput et al. (2025)	MobileNetV2 (TL)	Haar Cascade (ROI mata), resize 128×128, grayscale,	98,5%	MRL Eye	Sensitif pencahayaan dan variasi kondisi wajah



Author	Model	Preprocessing	Akurasi	Dataset	Limitasi
Vijaypriya & Uma (2023)	Multiscale CNN + FSA	augmentasi (rotasi, flipping, shifting) DLlib landmark detection, Cross-Guided Bilateral Filtering (denoising)	98,3%	YAWDD, NTHU-DDD	Pipeline preprocessing kompleks membatasi implementasi real-time
Adhithyaa et al. (2023)	Adaptive 3D-CNN	CNN-based face positioning (I-Net, P-Net, O-Net), NIOUS (overlap suppression), spatio-temporal 3D input (video sequence)	75-77%	KEC-DDD, NTHU-DDD	Performa relatif rendah akibat keterbatasan dataset serta kompleksitas model 3D
Jebraeily et al. (2024)	CNN hasil optimasi Genetic Algorithm + Transfer Learning	Ekstraksi video frame, deteksi wajah, konversi grayscale, resize 48×48, labeling (normal, fatigue, drowsy)	99,79%	Dataset buatan (7342 citra) + FER-2013 (35,887 citra)	Training lama, butuh komputasi tinggi, bergantung kualitas dataset dan desain arsitektur CNN
Kumral & Kükükmanisa (2022)	SSD MobileNet + CNN	SSD MobileNetV2 FPNLite (640×640) object detection, deteksi 7 kelas wajah, sequence 50 frame (≈2 detik), decision matrix 50×50	85-92%	YawDD, NTHU-DDD	Performa menurun pada kondisi pencahayaan rendah dan penggunaan kacamata
A.-C. Phan et al. (2021)	MobileNetV2/ResNet50V2	SSD ResNet-10 face detection, resize 224×224, normalisasi, facial landmark (EAR + lip distance, adaptive threshold)	96-97%	Custom Dataset	Sensitif terhadap variasi sudut wajah, pencahayaan, dan perubahan ekspresi
Minhas et al. (2022)	ResNet-50, InceptionV3	Resize tidak disebut eksplisit, augmentasi (rotasi, blur, noise, flipping), tanpa ROI spesifik	93,69%	Dataset privat (1000 citra)	Generalisasi terbatas akibat ukuran dataset kecil dan kurang variasi kondisi
Winarno et al. (2025)	EfficientNet + KNN	Haar Cascade face detection, resize 128×128, normalisasi, sequence segmentation (5	99,52%	DROZY	Bias dataset dan keterbatasan pengujian pada kondisi laboratorium



Author	Model	Preprocessing	Akurasi	Dataset	Limitasi
Ahmed et al. (2023)	CNN vs VGG16	detik = 25 frame), time-distributed input Face cropping, resize 145×145 (CNN) dan 64×64 (VGG16), one-hot encoding, augmentasi, regularisasi (L2)	97% (CNN), 74% (VGG16)	Kaggle Drowsiness Dataset	Tidak mencakup variasi pose ekstrem dan occlusion
Venkateswarlu & Chirra (2024)	Shallow CNN vs VGG19, ResNet50, MobileNetV2, InceptionV3	Dlib 68-point landmark, deteksi wajah dan ekstraksi ROI mata, cropping area mata, resize 128×128	99,23% (D1), 99,14% (D2)	Dataset-1 (324 citra), Kaggle Yawn Eye Dataset (1452 citra)	Dataset kecil, kurang variasi kondisi nyata (cahaya, pose, occlusion)

