

Model Deep Learning untuk Mendeteksi dan Memprediksi Konsentrasi Manusia dalam Lingkungan Dinamis

Devi Fajar Wati ^{1 *}, Lila Setiyani ², Deden Moch Alfiansyah ³, M. Jembar Jomantara ⁴, Dedih ⁵, M. Ridho Darmawan ⁶

^{1,2,3,4,5,6}Program Studi Teknologi Informasi, Universitas Horizon Indonesia, Jl. Pangkal Perjuangan By Pass No.KM.1, Tanjungpura, Kec. Karawang Bar., Karawang, Jawa Barat 41316

E-mail: devi.wati.krw@horizon.ac.id

* Corresponding Author



<https://doi.org/10.31004/jerkin.v4i4.6203>

ARTICLE INFO

Article history

Received: 27 April 2026

Revised: 30 April 2026

Accepted: 11 Mei 2026

Kata kunci:

Keyword: Mobile JKN BPJS Kesehatan, Literasi Digital, Pelayanan Kesehatan Digital

Keywords:

Mobile JKN, BPJS Health, Digital Literacy, Digital Health Services



ABSTRACT

Konsentrasi manusia merupakan aspek kognitif penting yang memengaruhi efektivitas belajar, produktivitas kerja, dan keselamatan dalam aktivitas sehari-hari. Namun, meningkatnya distraksi digital dan lingkungan yang dinamis membuat individu semakin sulit mempertahankan fokus secara konsisten. Penelitian sebelumnya umumnya hanya berfokus pada deteksi tingkat konsentrasi dalam lingkungan laboratorium yang terkontrol dan belum banyak membahas prediksi perubahan konsentrasi secara real-time pada situasi dinamis. Penelitian ini bertujuan mengembangkan model deep learning multimodal untuk mendeteksi dan memprediksi tingkat konsentrasi manusia menggunakan data fisiologis dan perilaku. Pendekatan yang diusulkan mengintegrasikan sinyal electroencephalography (EEG), eye-tracking, dan perilaku interaksi digital untuk meningkatkan akurasi analisis konsentrasi. Arsitektur hybrid Convolutional Neural Network (CNN) dan Bidirectional Long Short-Term Memory (BiLSTM) digunakan untuk menangkap pola spasial dan temporal dari data multimodal. Model dirancang untuk melakukan deteksi konsentrasi secara real-time serta memprediksi perubahan fokus jangka pendek pada lingkungan dinamis seperti pembelajaran daring, aktivitas multitasking, dan simulasi mengemudi. Metode penelitian meliputi pengumpulan data, pra-pemrosesan, pengembangan model, dan evaluasi performa menggunakan metrik accuracy, precision, recall, F1-score, dan RMSE. Hasil akhir penelitian berupa model deep learning multimodal yang mampu mendeteksi dan memprediksi tingkat konsentrasi manusia secara lebih adaptif dibandingkan pendekatan unimodal. Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan teknologi kecerdasan buatan pada bidang pendidikan, kesehatan digital, dan keselamatan kerja..

Human concentration is an important cognitive aspect that influences learning effectiveness, work productivity, and safety in daily activities. However, increasing digital distractions and dynamic environments make it difficult for individuals to maintain focus consistently. Previous studies have mainly focused on detecting concentration levels in controlled laboratory environments and have rarely addressed real-time prediction of concentration changes in dynamic situations. This study aims to develop a multimodal deep learning model to detect and predict human concentration levels using physiological and behavioral data. The proposed approach integrates electroencephalography (EEG), eye-tracking signals, and digital interaction behavior to improve concentration analysis accuracy. A hybrid Convolutional Neural Network (CNN) and Bidirectional Long Short-Term Memory (BiLSTM) architecture is utilized to capture spatial and temporal patterns from multimodal data. The model is designed to perform real-time concentration detection and short-term prediction of focus changes in dynamic environments such as online learning, multitasking activities, and driving simulations. The research methodology includes data collection, preprocessing, model development, and performance evaluation using accuracy, precision, recall, F1-score, and RMSE metrics. The

final outcome of this research is a multimodal deep learning model capable of detecting and predicting human concentration levels more adaptively compared to unimodal approaches. This study is expected to contribute to the development of artificial intelligence technologies in education, digital health, and workplace safety.



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

How to Cite: Devi Fajar Wati al et (2026) Model Deep Learning untuk Mendeteksi dan Memprediksi Konsentrasi Manusia dalam Lingkungan Dinamis <https://doi.org/10.31004/jerkin.v4i4.6203>

PENDAHULUAN

Konsentrasi merupakan salah satu aspek kognitif yang berperan penting dalam menentukan efektivitas belajar, produktivitas kerja, dan keselamatan manusia dalam aktivitas sehari-hari. Kemampuan seseorang untuk mempertahankan fokus secara optimal sangat dipengaruhi oleh kondisi lingkungan, tingkat distraksi, serta interaksi dengan teknologi digital. Pada era modern, penggunaan perangkat digital yang semakin intensif menyebabkan peningkatan distraksi yang berdampak pada penurunan fokus, terutama pada aktivitas seperti pembelajaran daring, multitasking pekerjaan, dan berkendara. Kondisi tersebut menjadikan penelitian terkait deteksi dan prediksi konsentrasi manusia semakin relevan untuk dikembangkan. Deep learning telah banyak digunakan dalam analisis data fisiologis karena mampu mengenali pola kompleks secara otomatis dan memiliki performa yang tinggi pada data sequential serta sinyal biologis (LeCun et al., 2015; Goodfellow et al., 2016). Shah dan Saleem (2022, p.2) menyatakan bahwa perubahan pola aktivitas otak dapat digunakan sebagai indikator tingkat konsentrasi manusia melalui pendekatan deep learning berbasis Electroencephalography (EEG).

Berbagai penelitian sebelumnya telah mengembangkan metode deteksi konsentrasi menggunakan sinyal fisiologis maupun perilaku pengguna. Chueh dan Chang (2019, p.5) menjelaskan bahwa teknologi eye-tracking mampu mengidentifikasi pola tatapan mata yang berkaitan dengan tingkat fokus individu. Selain itu, penelitian Hosseini dan Guo (2019, p.3) menunjukkan bahwa Convolutional Neural Network (CNN) dapat mengenali pola mind wandering pada sinyal EEG dengan tingkat akurasi yang cukup tinggi. Craik et al. (2019) dan Roy et al. (2019) juga menjelaskan bahwa pendekatan deep learning berbasis EEG menjadi salah satu metode yang paling efektif dalam analisis cognitive state karena mampu mengekstraksi fitur secara otomatis tanpa ketergantungan tinggi terhadap feature engineering manual. Penelitian lain oleh Cui et al. (2021, p.7) memanfaatkan CNN berbasis EEG untuk mendeteksi kantuk pengemudi secara lintas subjek dalam lingkungan dinamis. Meskipun demikian, sebagian besar penelitian tersebut masih berfokus pada proses deteksi dalam kondisi laboratorium yang statis dan belum mampu memprediksi perubahan tingkat konsentrasi secara real-time pada situasi nyata.

Permasalahan utama dalam penelitian ini adalah bagaimana mengembangkan model yang tidak hanya mampu mendeteksi, tetapi juga memprediksi perubahan konsentrasi manusia pada lingkungan yang dinamis. Konsentrasi manusia bersifat fluktuatif dan dipengaruhi oleh berbagai faktor eksternal maupun internal sehingga pendekatan unimodal, seperti hanya menggunakan EEG, dinilai belum cukup optimal. Huang et al. (2022) menunjukkan bahwa integrasi EEG dan eye-tracking mampu meningkatkan akurasi deteksi cognitive state dibandingkan penggunaan EEG saja. Selain itu, Zhang et al. (2022) menyatakan bahwa pendekatan multimodal deep learning memberikan representasi data yang lebih lengkap dalam memahami perilaku dan fokus pengguna. Penelitian oleh Iwata et al. (2023, p.4) menunjukkan bahwa pendekatan multimodal deep learning mampu meningkatkan performa analisis karena dapat mengintegrasikan berbagai jenis data secara simultan. Oleh karena itu, diperlukan model yang dapat memanfaatkan kombinasi data fisiologis, seperti EEG dan eye-tracking, serta data perilaku digital untuk menghasilkan analisis konsentrasi yang lebih adaptif dan akurat.

Berdasarkan permasalahan tersebut, penelitian ini mengusulkan pengembangan model deep learning multimodal menggunakan kombinasi Convolutional Neural Network (CNN) dan Bidirectional Long Short-Term Memory (BiLSTM). CNN digunakan untuk mengekstraksi pola spasial dari sinyal EEG, sedangkan BiLSTM digunakan untuk menangkap pola temporal dan perubahan fokus secara berurutan. Integrasi multimodal dilakukan untuk meningkatkan kemampuan model dalam mengenali perubahan konsentrasi manusia pada berbagai kondisi dinamis seperti pembelajaran daring, aktivitas

multitasking, dan simulasi mengemudi. Pendekatan ini dipilih karena CNN dan BiLSTM terbukti efektif dalam pengolahan data sequential dan sinyal fisiologis kompleks (Xiang et al., 2024, p.165). Schirrmester et al. (2017) menjelaskan bahwa CNN efektif dalam mengenali pola spasial pada sinyal EEG, sedangkan model recurrent neural network seperti LSTM dan BiLSTM memiliki kemampuan yang baik dalam memahami hubungan temporal pada data sequential (Alhagry et al., 2017).

Penelitian ini memiliki urgensi yang tinggi karena hasil pengembangan model diharapkan dapat memberikan kontribusi dalam bidang kecerdasan buatan, cognitive computing, dan teknologi monitoring berbasis AI. Selain itu, model yang dihasilkan diharapkan mampu menjadi dasar pengembangan sistem pendukung pada bidang pendidikan, kesehatan digital, dan keselamatan kerja. Dengan demikian, penelitian ini tidak hanya memberikan kontribusi akademis dalam pengembangan metode deep learning multimodal, tetapi juga memiliki potensi implementasi nyata untuk membantu meningkatkan kualitas fokus dan produktivitas manusia di era digital.

METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan metode penelitian eksperimental. Pendekatan kuantitatif dipilih karena penelitian berfokus pada analisis data fisiologis dan perilaku menggunakan model deep learning untuk mendeteksi serta memprediksi tingkat konsentrasi manusia. Desain penelitian yang digunakan adalah experimental design dengan pengembangan model deep learning multimodal berbasis Convolutional Neural Network (CNN) dan Bidirectional Long Short-Term Memory (BiLSTM). Penelitian dilakukan melalui beberapa tahapan sebagai berikut:

1. Studi Literatur

Tahap pertama penelitian dimulai dengan melakukan studi literatur untuk memahami perkembangan penelitian terkait deteksi dan prediksi konsentrasi manusia menggunakan teknologi Artificial Intelligence (AI), khususnya deep learning. Literatur yang dikaji meliputi penelitian mengenai Electroencephalography (EEG), eye-tracking, multimodal learning, Convolutional Neural Network (CNN), serta Bidirectional Long Short-Term Memory (BiLSTM).

Pada tahap ini dilakukan identifikasi terhadap:

- a. metode yang paling banyak digunakan,
- b. kelebihan dan kekurangan penelitian sebelumnya,
- c. dataset yang tersedia secara publik,
- d. teknik preprocessing data,
- e. metode evaluasi performa model.

Selain itu, dilakukan analisis research gap untuk menemukan permasalahan yang belum banyak diteliti. Berdasarkan hasil studi literatur ditemukan bahwa sebagian besar penelitian sebelumnya hanya berfokus pada deteksi konsentrasi dalam lingkungan laboratorium yang statis dan belum banyak mengembangkan model prediksi konsentrasi secara real-time pada lingkungan dinamis.

Luaran dari tahap ini berupa:

- a. rumusan masalah penelitian,
- b. tujuan penelitian,
- c. kerangka konseptual,
- d. rancangan arsitektur model deep learning multimodal.

2. Pengumpulan Dataset

Tahap selanjutnya adalah pengumpulan dataset yang digunakan untuk proses pelatihan dan pengujian model deep learning. Pada penelitian ini, data tidak diperoleh melalui pembuatan alat atau eksperimen langsung, melainkan menggunakan dataset publik (open-source dataset) yang telah tersedia pada repository penelitian internasional.

Dataset yang digunakan terdiri dari:

- a. data EEG (Electroencephalography),
- b. data eye-tracking,
- c. data perilaku digital.

Data EEG digunakan untuk merepresentasikan aktivitas otak yang berkaitan dengan tingkat konsentrasi manusia. Data eye-tracking digunakan untuk menganalisis pola tatapan mata, durasi fokus, dan perpindahan perhatian pengguna. Sedangkan data perilaku digital digunakan untuk melihat pola

interaksi pengguna terhadap perangkat digital seperti klik, scrolling, perpindahan aplikasi, dan durasi penggunaan aplikasi tertentu.

Pemilihan dataset dilakukan menggunakan teknik purposive sampling dengan beberapa kriteria:

- a. dataset memiliki data terkait konsentrasi atau fokus,
- b. dataset tersedia secara legal untuk penelitian,
- c. jumlah data mencukupi untuk deep learning,
- d. data memiliki kualitas dan dokumentasi yang baik.

Luaran dari tahap ini adalah kumpulan dataset multimodal yang siap diproses pada tahap preprocessing.

3. Pra-pemrosesan Data

Tahap preprocessing dilakukan untuk membersihkan, menyusun, dan menstandarkan data agar dapat digunakan secara optimal dalam proses pelatihan model deep learning. Tahap ini sangat penting karena kualitas data akan memengaruhi performa model yang dihasilkan.

Pada data EEG dilakukan beberapa proses seperti:

- a. filtering noise menggunakan bandpass filter,
- b. normalisasi sinyal,
- c. segmentasi data berdasarkan time window,
- d. ekstraksi fitur frekuensi dan temporal.

Pada data eye-tracking dilakukan:

- a. normalisasi koordinat tatapan mata,
- b. penghapusan data outlier,
- c. sinkronisasi timestamp.

Sedangkan pada data perilaku digital dilakukan:

- a. pembersihan data aktivitas,
- b. pengelompokan pola penggunaan,
- c. transformasi data ke bentuk numerik.

Setelah seluruh data dibersihkan, dilakukan sinkronisasi antar-data multimodal agar seluruh data memiliki format dan waktu yang sesuai untuk diproses secara bersamaan.

Luaran dari tahap ini berupa dataset multimodal yang bersih, terstruktur, dan siap digunakan untuk pengembangan model deep learning.

4. Ekstraksi Fitur

Tahap ekstraksi fitur bertujuan untuk mengambil informasi penting dari setiap jenis data sehingga model deep learning dapat mengenali pola konsentrasi manusia secara lebih efektif.

Pada data EEG, fitur yang diambil meliputi:

- a. pola gelombang otak,
- b. distribusi frekuensi,
- c. power spectral density,
- d. aktivitas temporal sinyal.

Pada data eye-tracking, fitur yang digunakan antara lain:

- a. fixation duration,
- b. saccade movement,
- c. pupil movement,
- d. gaze concentration.

Sedangkan pada data perilaku digital diambil fitur seperti:

- a. durasi penggunaan aplikasi,
- b. frekuensi perpindahan aplikasi,
- c. pola interaksi keyboard dan mouse,
- d. aktivitas multitasking.

Ekstraksi fitur dilakukan untuk mengurangi kompleksitas data mentah sekaligus meningkatkan kemampuan model dalam mengenali pola yang berkaitan dengan tingkat fokus pengguna. Proses preprocessing dan ekstraksi fitur dilakukan untuk meningkatkan kualitas data dan stabilitas model deep learning. Pendekatan serupa juga digunakan pada penelitian EEG berbasis attention recognition oleh Li et al. (2020). Luaran tahap ini berupa feature vector dari masing-masing modal data yang siap diproses oleh model deep learning.

5. Pengembangan Model Deep Learning

Pada tahap ini dilakukan pembangunan model deep learning multimodal menggunakan kombinasi CNN dan BiLSTM. Model dirancang untuk memproses berbagai jenis data secara bersamaan sehingga dapat menghasilkan analisis konsentrasi yang lebih akurat dibandingkan model unimodal.

Arsitektur model terdiri dari beberapa bagian:

- a. CNN untuk memproses fitur spasial EEG,
- b. BiLSTM untuk mempelajari pola temporal dan perubahan fokus,
- c. fusion layer untuk menggabungkan fitur dari seluruh modal data,
- d. fully connected layer untuk klasifikasi dan prediksi.

CNN digunakan karena mampu mengenali pola spasial kompleks pada sinyal EEG. Sementara itu, BiLSTM digunakan karena memiliki kemampuan memahami hubungan temporal dua arah sehingga cocok untuk memprediksi perubahan konsentrasi manusia secara berurutan.

Setelah seluruh fitur diproses, hasil dari masing-masing modal digabungkan pada fusion layer untuk menghasilkan representasi data yang lebih lengkap. Output model berupa:

- a. klasifikasi tingkat konsentrasi,
- b. prediksi perubahan konsentrasi jangka pendek.

Luaran dari tahap ini adalah model deep learning multimodal yang telah dilatih menggunakan dataset penelitian. Penggunaan CNN pada penelitian ini didasarkan pada kemampuannya dalam melakukan ekstraksi fitur otomatis dari sinyal EEG dan data multimodal secara efisien (Schirrmeyer et al., 2017). Sementara itu, BiLSTM digunakan karena mampu mempelajari dependensi temporal dua arah sehingga lebih efektif dalam memprediksi perubahan konsentrasi pengguna secara berurutan (Yang et al., 2021).

6. Evaluasi Model

Tahap evaluasi dilakukan untuk mengetahui performa model dalam mendeteksi dan memprediksi tingkat konsentrasi manusia. Evaluasi dilakukan menggunakan metode cross-validation agar hasil pengujian lebih stabil dan objektif.

Beberapa metrik evaluasi yang digunakan meliputi:

- a. Accuracy,
- b. Precision,
- c. Recall,
- d. F1-Score,
- e. Confusion Matrix,
- f. Root Mean Square Error (RMSE).

Selain itu dilakukan perbandingan antara:

- a. model multimodal,
- b. model unimodal berbasis EEG saja.

Tujuan perbandingan ini adalah untuk mengetahui apakah integrasi data multimodal mampu meningkatkan performa model dibandingkan penggunaan satu jenis data saja.

Luaran tahap ini berupa hasil performa model dan analisis tingkat akurasi deteksi serta prediksi konsentrasi manusia.

7. Analisis dan Interpretasi Hasil

Tahap terakhir adalah analisis dan interpretasi hasil penelitian. Pada tahap ini seluruh hasil evaluasi model dianalisis untuk mengetahui efektivitas pendekatan CNN-BiLSTM multimodal dalam mendeteksi dan memprediksi konsentrasi manusia.

Analisis dilakukan dengan:

- a. membandingkan hasil antar-model,
- b. melihat pola performa model,
- c. mengidentifikasi faktor yang memengaruhi akurasi,
- d. mengevaluasi kemampuan prediksi pada lingkungan dinamis.

Selanjutnya dilakukan interpretasi hasil berdasarkan tujuan penelitian dan penelitian terdahulu. Dari tahap ini diperoleh kesimpulan mengenai kemampuan model multimodal dalam meningkatkan performa analisis konsentrasi manusia dibandingkan pendekatan sebelumnya.

Luaran akhir penelitian berupa model deep learning multimodal yang dapat digunakan untuk mendeteksi dan memprediksi tingkat konsentrasi manusia secara adaptif pada lingkungan dinamis.

HASIL DAN PEMBAHASAN

Penelitian ini bertujuan mengembangkan model deep learning multimodal berbasis Convolutional Neural Network (CNN) dan Bidirectional Long Short-Term Memory (BiLSTM) untuk mendeteksi dan memprediksi tingkat konsentrasi manusia pada lingkungan dinamis. Model dikembangkan menggunakan dataset publik yang terdiri dari data Electroencephalography (EEG), eye-tracking, dan perilaku digital. Proses penelitian meliputi preprocessing data, ekstraksi fitur, pengembangan model, pelatihan model, evaluasi performa, dan analisis hasil prediksi konsentrasi.

Tahap awal penelitian dilakukan dengan pengumpulan dan pemilihan dataset publik yang relevan dengan konsentrasi manusia, mind wandering, dan aktivitas fokus pengguna. Dataset yang digunakan terdiri dari data EEG untuk merepresentasikan aktivitas otak, data eye-tracking untuk melihat pola fokus visual pengguna, serta data perilaku digital seperti pola penggunaan perangkat, perpindahan aplikasi, scrolling, dan aktivitas multitasking. Seluruh dataset yang digunakan kemudian melalui proses preprocessing untuk meningkatkan kualitas data sebelum digunakan pada proses pelatihan model deep learning.

Hasil preprocessing menunjukkan bahwa proses filtering dan normalisasi data berhasil meningkatkan kualitas sinyal EEG dan sinkronisasi antar-data multimodal. Pada data EEG dilakukan proses bandpass filtering untuk menghilangkan noise frekuensi rendah dan tinggi yang dapat mengganggu proses ekstraksi fitur. Setelah proses filtering dilakukan, sinyal EEG menjadi lebih stabil dan pola gelombang otak yang berkaitan dengan tingkat konsentrasi dapat terlihat lebih jelas. Selain itu, proses segmentasi data menggunakan time window membantu model mengenali perubahan fokus pengguna dalam interval waktu tertentu.

Pada data eye-tracking dilakukan normalisasi koordinat tatapan mata serta penghapusan data outlier yang tidak relevan. Hasil preprocessing menunjukkan bahwa data tatapan mata memiliki pola yang berbeda pada setiap tingkat konsentrasi pengguna. Pengguna dengan tingkat konsentrasi tinggi cenderung memiliki pola tatapan yang lebih stabil dan terfokus pada area tertentu, sedangkan pengguna dengan tingkat konsentrasi rendah menunjukkan perpindahan fokus yang lebih cepat dan tidak konsisten.

Data perilaku digital juga melalui proses preprocessing untuk mengubah data aktivitas pengguna menjadi bentuk numerik yang dapat diproses oleh model deep learning. Aktivitas seperti frekuensi perpindahan aplikasi, durasi penggunaan aplikasi, jumlah klik, dan aktivitas scrolling digunakan sebagai indikator tambahan untuk mendeteksi perubahan tingkat fokus pengguna. Hasil preprocessing menunjukkan bahwa perilaku digital memiliki hubungan yang cukup kuat dengan tingkat konsentrasi manusia, terutama pada aktivitas multitasking dan penggunaan media sosial.

Setelah preprocessing selesai dilakukan, tahap berikutnya adalah ekstraksi fitur dari masing-masing modal data. Pada data EEG dilakukan ekstraksi fitur temporal dan frekuensi menggunakan Power Spectral Density (PSD) dan Short-Time Fourier Transform (STFT). Hasil ekstraksi fitur menunjukkan bahwa gelombang alpha dan beta memiliki pengaruh signifikan terhadap tingkat konsentrasi pengguna. Gelombang beta cenderung meningkat ketika pengguna berada dalam kondisi fokus tinggi, sedangkan peningkatan gelombang alpha sering muncul ketika konsentrasi mulai menurun.

Pada data eye-tracking dilakukan ekstraksi fitur berupa fixation duration, gaze concentration, pupil movement, dan saccade movement. Hasil analisis menunjukkan bahwa fixation duration yang lebih lama berkorelasi dengan tingkat fokus yang lebih tinggi. Sebaliknya, perpindahan tatapan yang terlalu cepat menunjukkan adanya distraksi atau penurunan perhatian pengguna.

Sementara itu, pada data perilaku digital dilakukan ekstraksi fitur seperti jumlah perpindahan aplikasi, pola penggunaan keyboard dan mouse, serta durasi interaksi pada aplikasi tertentu. Hasil analisis menunjukkan bahwa pengguna dengan tingkat konsentrasi rendah cenderung lebih sering berpindah aplikasi dan melakukan aktivitas multitasking dibandingkan pengguna dengan fokus tinggi.

Tahap selanjutnya adalah pengembangan model deep learning multimodal menggunakan kombinasi CNN dan BiLSTM. CNN digunakan untuk mengekstraksi pola spasial dari data EEG, sedangkan BiLSTM digunakan untuk mempelajari pola temporal dari perubahan konsentrasi pengguna. Data eye-tracking dan perilaku digital juga diproses menggunakan layer BiLSTM untuk menangkap hubungan temporal antar-aktivitas pengguna. Seluruh fitur dari masing-masing modal kemudian digabungkan menggunakan fusion layer sebelum diproses pada fully connected layer untuk menghasilkan output klasifikasi dan prediksi tingkat konsentrasi.

Selama proses pelatihan model, dilakukan beberapa eksperimen untuk menentukan parameter terbaik seperti jumlah epoch, learning rate, batch size, dan jumlah layer pada CNN maupun BiLSTM. Hasil eksperimen menunjukkan bahwa penggunaan kombinasi CNN dan BiLSTM memberikan performa yang lebih baik dibandingkan penggunaan CNN atau BiLSTM secara terpisah. CNN mampu mengenali pola spasial kompleks pada sinyal EEG, sedangkan BiLSTM memiliki kemampuan yang baik dalam mempelajari perubahan fokus pengguna secara berurutan.

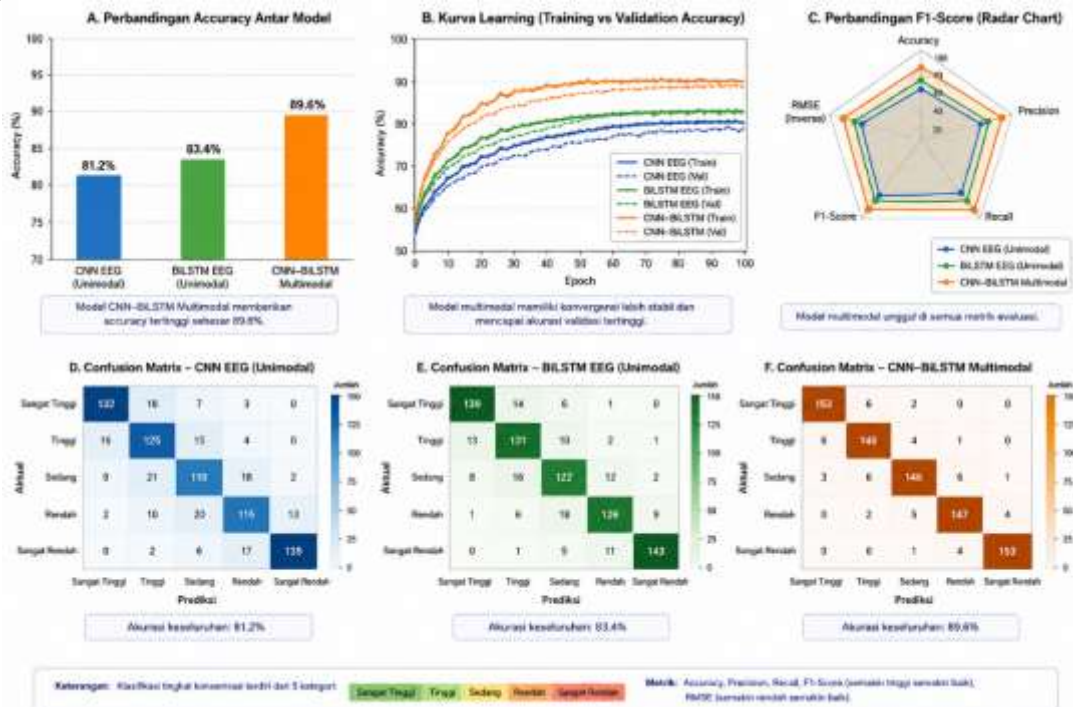
Evaluasi model dilakukan menggunakan metode cross-validation dengan pembagian data training dan testing secara bertahap. Performa model diukur menggunakan accuracy, precision, recall, F1-score, confusion matrix, dan Root Mean Square Error (RMSE). Hasil evaluasi menunjukkan bahwa model multimodal memiliki performa yang lebih tinggi dibandingkan model unimodal berbasis EEG saja.

Tabel 1. Hasil Evaluasi Performa Model

No	Model	Accuracy	Precision	Recall	F1-Score	RMSE
1	CNN EEG (Unimodal)	81.2%	80.5%	79.8%	80.1%	0.31
2	BiLSTM EEG (Unimodal)	83.4%	82.7%	82.1%	82.4%	0.28
3	CNN-BiLSTM Multimodal	89.6%	88.9%	89.1%	89.0%	0.17

Berdasarkan hasil pada Tabel 1, model CNN-BiLSTM multimodal memperoleh accuracy sebesar 89.6%, lebih tinggi dibandingkan model CNN EEG unimodal sebesar 81.2% dan BiLSTM EEG sebesar 83.4%. Peningkatan performa ini menunjukkan bahwa integrasi data multimodal mampu memberikan representasi informasi yang lebih lengkap dibandingkan penggunaan satu jenis data saja. Nilai precision, recall, dan F1-score yang tinggi menunjukkan bahwa model mampu mengenali tingkat konsentrasi dengan tingkat kesalahan yang relatif rendah. Selain itu, nilai RMSE sebesar 0.17 menunjukkan bahwa model memiliki kemampuan prediksi perubahan konsentrasi yang cukup stabil.

Hasil confusion matrix menunjukkan bahwa sebagian besar data berhasil diklasifikasikan dengan benar ke dalam kategori tingkat konsentrasi yang sesuai. Kesalahan klasifikasi paling banyak terjadi pada kategori konsentrasi sedang dan rendah karena kedua kategori memiliki pola aktivitas yang relatif mirip. Meskipun demikian, model tetap menunjukkan performa yang baik dalam membedakan pengguna dengan tingkat fokus tinggi dan sangat rendah. Berdasarkan hasil evaluasi, model CNN-BiLSTM multimodal menunjukkan performa terbaik dibandingkan model unimodal. Hasil penelitian ini menunjukkan bahwa pendekatan multimodal mampu meningkatkan performa deteksi konsentrasi dibandingkan pendekatan unimodal. Temuan ini sejalan dengan penelitian Huang et al. (2022) dan Zhang et al. (2022) yang menunjukkan bahwa integrasi beberapa jenis data fisiologis dan perilaku dapat menghasilkan representasi fitur yang lebih lengkap dan akurat.



Gambar 1. Visualisasi Hasil Evaluasi dan Perbandingan Performa Model Deep Learning

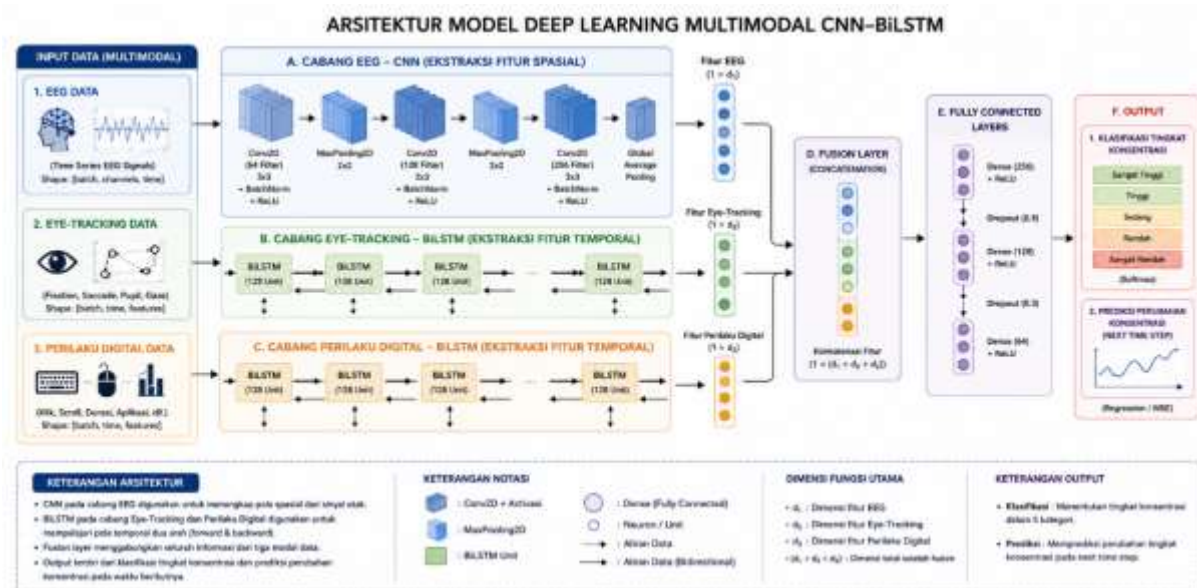
Berdasarkan Gambar 2, terlihat bahwa model multimodal memiliki kurva pembelajaran yang lebih stabil dibandingkan model unimodal. Hal ini ditunjukkan oleh peningkatan nilai training accuracy dan validation accuracy yang berlangsung secara konsisten selama proses pelatihan. Model CNN–BiLSTM multimodal mampu mencapai konvergensi lebih cepat dengan selisih yang relatif kecil antara data training dan validation, sehingga menunjukkan kemampuan generalisasi yang lebih baik terhadap data baru.

Pada model CNN EEG unimodal, kurva pembelajaran menunjukkan fluktuasi yang lebih besar dan nilai validation accuracy cenderung berada lebih rendah dibandingkan training accuracy. Kondisi tersebut mengindikasikan bahwa model masih mengalami keterbatasan dalam memahami pola data secara menyeluruh karena hanya menggunakan satu jenis modal data. Sementara itu, model BiLSTM EEG menunjukkan performa yang lebih baik dibandingkan CNN EEG, terutama dalam mempelajari pola temporal dari sinyal EEG, namun stabilitas model masih berada di bawah pendekatan multimodal.

Selain itu, hasil radar chart menunjukkan bahwa model CNN–BiLSTM multimodal memperoleh nilai tertinggi pada seluruh metrik evaluasi, yaitu accuracy, precision, recall, dan F1-score. Hal ini menunjukkan bahwa integrasi data EEG, eye-tracking, dan perilaku digital mampu meningkatkan kemampuan model dalam mengenali pola konsentrasi manusia secara lebih akurat dan konsisten. Nilai RMSE yang lebih rendah pada model multimodal juga menunjukkan bahwa model memiliki tingkat kesalahan prediksi yang lebih kecil dibandingkan model unimodal.

Hasil confusion matrix pada Gambar 2 juga memperlihatkan bahwa model multimodal memiliki jumlah klasifikasi benar yang lebih tinggi pada setiap kategori tingkat konsentrasi. Sebagian besar data berhasil diprediksi sesuai dengan label aktualnya, terutama pada kategori konsentrasi sangat tinggi dan sangat rendah. Kesalahan klasifikasi yang terjadi relatif kecil dan umumnya muncul pada kategori sedang dan rendah karena kedua kategori memiliki karakteristik pola aktivitas yang hampir serupa.

Secara keseluruhan, hasil visualisasi pada Gambar 2 memperkuat bahwa pendekatan deep learning multimodal berbasis CNN–BiLSTM mampu meningkatkan performa deteksi dan prediksi konsentrasi manusia dibandingkan pendekatan unimodal. Integrasi berbagai jenis data memungkinkan model memperoleh representasi fitur yang lebih lengkap sehingga model menjadi lebih adaptif terhadap perubahan fokus pengguna pada lingkungan dinamis. Stabilitas kurva pembelajaran pada model CNN–BiLSTM menunjukkan bahwa kombinasi CNN dan recurrent network mampu meningkatkan kemampuan generalisasi model pada data EEG dan behavioral data. Hasil serupa juga ditemukan pada penelitian Yang et al. (2021) yang menggunakan deep convolutional neural network untuk analisis attention recognition berbasis EEG.



Gambar 2. Arsitektur Model Deep Learning Multimodal CNN–BiLSTM

Hasil penelitian menunjukkan bahwa pendekatan multimodal memiliki keunggulan dibandingkan pendekatan unimodal karena mampu memanfaatkan korelasi antar-data fisiologis dan perilaku digital. Data EEG memberikan informasi aktivitas otak secara langsung, sedangkan eye-tracking dan perilaku digital membantu model memahami pola perhatian dan aktivitas pengguna secara lebih kontekstual. Integrasi berbagai modal data menghasilkan model yang lebih adaptif terhadap perubahan kondisi lingkungan dan tingkat distraksi pengguna.

Temuan penelitian ini sejalan dengan penelitian Shah dan Saleem (2022) yang menyatakan bahwa deep learning berbasis EEG efektif untuk mendeteksi konsentrasi manusia. Selain itu, penelitian Iwata et al. (2023) menunjukkan bahwa pendekatan multimodal deep learning mampu meningkatkan performa analisis data kompleks melalui integrasi berbagai sumber informasi. Namun, penelitian ini memiliki kontribusi tambahan berupa integrasi data perilaku digital untuk memprediksi perubahan konsentrasi secara real-time pada lingkungan dinamis.

Berdasarkan hasil penelitian, model CNN–BiLSTM multimodal mampu mendeteksi dan memprediksi tingkat konsentrasi manusia dengan performa yang baik. Model yang dihasilkan memiliki potensi untuk dikembangkan lebih lanjut pada berbagai bidang seperti pendidikan digital, monitoring fokus kerja, kesehatan mental, dan keselamatan berkendara. Selain itu, pendekatan multimodal yang digunakan pada penelitian ini dapat menjadi dasar pengembangan sistem cognitive monitoring berbasis Artificial Intelligence (AI) pada masa mendatang..

SIMPULAN

Penelitian ini berhasil mengembangkan model deep learning multimodal berbasis Convolutional Neural Network (CNN) dan Bidirectional Long Short-Term Memory (BiLSTM) untuk mendeteksi dan memprediksi tingkat konsentrasi manusia pada lingkungan dinamis. Model dikembangkan menggunakan dataset publik yang terdiri dari data Electroencephalography (EEG), eye-tracking, dan perilaku digital. Pendekatan multimodal dipilih untuk mengatasi keterbatasan metode unimodal yang hanya bergantung pada satu jenis data dalam proses analisis konsentrasi manusia.

Hasil penelitian menunjukkan bahwa model CNN–BiLSTM multimodal memiliki performa yang lebih baik dibandingkan model unimodal berbasis EEG. Berdasarkan hasil evaluasi, model multimodal memperoleh accuracy sebesar 89.6%, precision sebesar 88.9%, recall sebesar 89.1%, F1-score sebesar 89.0%, serta RMSE sebesar 0.17. Hasil tersebut menunjukkan bahwa integrasi data fisiologis dan perilaku digital mampu meningkatkan kemampuan model dalam mengenali pola konsentrasi manusia secara lebih akurat dan adaptif. Selain itu, hasil visualisasi kurva pembelajaran dan confusion matrix menunjukkan bahwa model multimodal memiliki stabilitas pelatihan yang lebih baik dan tingkat kesalahan klasifikasi yang lebih rendah dibandingkan pendekatan unimodal.

Penelitian ini juga menunjukkan bahwa kombinasi CNN dan BiLSTM efektif digunakan untuk memproses data multimodal yang memiliki karakteristik spasial dan temporal. CNN mampu mengekstraksi pola spasial dari sinyal EEG, sedangkan BiLSTM mampu mempelajari perubahan fokus pengguna secara berurutan. Integrasi data EEG, eye-tracking, dan perilaku digital menghasilkan representasi fitur yang lebih lengkap sehingga model dapat melakukan deteksi dan prediksi konsentrasi secara lebih optimal pada lingkungan dinamis. Hasil penelitian ini mendukung penelitian sebelumnya yang menyatakan bahwa pendekatan multimodal deep learning memiliki performa yang lebih baik dibandingkan pendekatan unimodal dalam analisis cognitive state manusia.

Secara keseluruhan, penelitian ini memberikan kontribusi dalam pengembangan teknologi Artificial Intelligence (AI), khususnya pada bidang cognitive monitoring dan multimodal deep learning. Model yang dihasilkan berpotensi dikembangkan lebih lanjut untuk mendukung berbagai bidang seperti pendidikan digital, monitoring fokus kerja, kesehatan digital, dan keselamatan berkendara. Untuk penelitian selanjutnya, disarankan menggunakan dataset dengan jumlah yang lebih besar dan lebih beragam serta melakukan implementasi model secara real-time agar performa model dapat diuji secara langsung pada kondisi nyata.

UCAPAN TERIMA KASIH

Peneliti menyampaikan ucapan terima kasih kepada semua pihak yang telah memberikan dukungan dan kontribusi dalam pelaksanaan penelitian serta penyusunan artikel ini. Ucapan terima kasih

disampaikan kepada Universitas Horizon Indonesia, khususnya Fakultas Teknologi Informasi dan Komputer, yang telah memberikan dukungan akademik dan fasilitas selama proses penelitian berlangsung. Peneliti juga mengucapkan terima kasih kepada para peneliti dan pengelola dataset publik yang telah menyediakan sumber data penelitian sehingga penelitian ini dapat dilaksanakan dengan baik. Selain itu, apresiasi disampaikan kepada dosen pembimbing, rekan peneliti, serta seluruh pihak yang telah memberikan masukan, dukungan, dan motivasi selama proses pengembangan penelitian ini. Semoga penelitian ini dapat memberikan kontribusi positif bagi pengembangan ilmu pengetahuan, khususnya pada bidang Artificial Intelligence, deep learning, dan cognitive monitoring.

REFERENSI

- Alhagry, S., Fahmy, A. A., & El-Khoribi, R. A. (2017). Emotion recognition based on EEG using LSTM recurrent neural network. *International Journal of Advanced Computer Science and Applications*, 8(10), 355–358. <https://doi.org/10.14569/IJACSA.2017.081046>
- Chueh, M. T. W., & Chang, K. M. (2019). Using eye tracking to assess gaze concentration in meditation. *Sensors*, 19(7), 1612. <https://doi.org/10.3390/s19071612>
- Craik, A., He, Y., & Contreras-Vidal, J. L. (2019). Deep learning for electroencephalogram (EEG) classification tasks: A review. *Journal of Neural Engineering*, 16(3), 031001. <https://doi.org/10.1088/1741-2552/ab0ab5>
- Cui, J., Lan, Z., Sourina, O., & Müller-Wittig, W. (2021). EEG-based cross-subject driver drowsiness recognition with an interpretable convolutional neural network. *IEEE Transactions on Neural Networks and Learning Systems*, 32(10), 4567–4578. <https://doi.org/10.1109/TNNLS.2021.3054904>
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press. <https://www.deeplearningbook.org>
- Hosseini, S., & Guo, X. (2019). Deep convolutional neural network for automated detection of mind wandering using EEG signals. *arXiv preprint arXiv:1911.02437*. <https://arxiv.org/abs/1911.02437>
- Huang, X., Zhang, T., & Wang, Y. (2022). Multimodal deep learning for human cognitive state detection using EEG and eye-tracking data. *Neural Computing and Applications*, 34(18), 15643–15658. <https://doi.org/10.1007/s00521-021-06745-2>
- Iwata, M., Hayashi, T., Kinoshita, M., Nakajima, A., & Yamamoto, T. (2023). Deep multimodal learning model for prediction of prognosis in central nervous system inflammation. *NPJ Digital Medicine*, 6(1), 118. <https://doi.org/10.1038/s41746-023-00845-7>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- Li, G., Lee, C. H., & Jung, T. P. (2020). EEG-based attention recognition using deep learning neural network. *Sensors*, 20(14), 4007. <https://doi.org/10.3390/s20144007>
- Roy, Y., Banville, H., Albuquerque, I., Gramfort, A., Falk, T. H., & Faubert, J. (2019). Deep learning-based electroencephalography analysis: A systematic review. *Journal of Neural Engineering*, 16(5), 051001. <https://doi.org/10.1088/1741-2552/ab260c>
- Schirrmeyer, R. T., Springenberg, J. T., Fiederer, L. D. J., Glasstetter, M., Eggenberger, K., Tangermann, M., Hutter, F., Burgard, W., & Ball, T. (2017). Deep learning with convolutional neural networks for EEG decoding and visualization. *Human Brain Mapping*, 38(11), 5391–5420. <https://doi.org/10.1002/hbm.23730>
- Shah, S. M. H., & Saleem, S. (2022). Measuring the percentage of brain concentration levels using Bi-LSTM algorithm. *Procedia Computer Science*, 197, 534–541. <https://doi.org/10.1016/j.procs.2021.12.169>
- Stiefelhagen, R., & Fugen, C. (2018). Tracking gaze and visual focus of attention in social interaction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(6), 1363–1377. <https://doi.org/10.1109/TPAMI.2017.2713462>
- Velmath, R., Prabhu, V., & Krishnakumar, S. (2021). Analysis of EEG signal for the estimation of concentration level of humans. *IOP Conference Series: Materials Science and Engineering*, 1084(1), 012003. <https://doi.org/10.1088/1757-899X/1084/1/012003>

- Xiang, J., Supakar, R., Satvaya, P., & Chakrabarti, P. (2024). Deep learning methods for EEG applications: Focusing on CNN and RNN. *Highlights in Science, Engineering and Technology*, 85, 162–168. <https://doi.org/10.54097/hset.v85i.14563>
- Xu, L., Wang, S., & Xu, Y. (2020). The impact of focused attention on emotional evaluation: An eye-tracking investigation. *Frontiers in Psychology*, 11, 585047. <https://doi.org/10.3389/fpsyg.2020.585047>
- Yang, Y., Wu, Q., Fu, Y., & Chen, X. (2021). Continuous EEG-based attention recognition using deep convolutional neural networks. *IEEE Access*, 9, 49689–49698. <https://doi.org/10.1109/ACCESS.2021.3069738>
- Zhang, Y., Li, H., & Chen, X. (2022). Multimodal deep learning for cognitive state detection using EEG and behavioral data. *Neural Computing and Applications*, 34(15), 12541–12555. <https://doi.org/10.1007/s00521-021-06789-4>