



Evaluasi Efektivitas Model Deep Learning Dalam Meningkatkan Hasil Belajar Kognitif: Sebuah Studi Eksperimen Semu

Omriana Nene Alle¹, Valen Febriani Dama², Elnatan Tamonob³

Institut Agama Kristen Negeri Kupang, Indonesia

omrianaalle09@gmail.com

Abstract

*This study aims to evaluate the effectiveness of applying deep learning models based on Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) as adaptive media in improving cognitive learning outcomes of tenth-grade high school students in Science subjects. A quasi-experimental method with a nonequivalent control group pretest-posttest design was employed. A total of 64 students (32 experimental, 32 control) were selected via purposive sampling. Research instruments consisted of Bloom's Taxonomy-based cognitive tests (C1–C6) with Cronbach Alpha reliability ($\alpha = 0.89$). Data were analyzed using independent *t*-test and ANCOVA with pretest scores as covariates. Results showed that the experimental group achieved significantly higher post-test mean scores ($M = 82.19$; $SD = 7.45$) compared to the control group ($M = 67.56$; $SD = 8.93$) ($t(62) = 7.84$, $p < .001$). ANCOVA revealed a large main effect ($\eta^2 = .501$, $F(1,61) = 61.37$, $p < .001$). Cohen's $d = 1.87$ indicated a very large effect size. These findings support the integration of deep learning models as effective adaptive learning media in formal educational contexts.*

Keywords: Deep Learning, Cognitive Learning Outcomes, Quasi-Experimental, CNN, LSTM, Bloom's Taxonomy.

Abstrak

Penelitian ini bertujuan untuk mengevaluasi efektivitas penerapan model deep learning berbasis Convolutional Neural Network (CNN) dan Long Short-Term Memory (LSTM) sebagai media adaptif dalam meningkatkan hasil belajar kognitif siswa pada mata pelajaran Sains kelas X SMA. Metode yang digunakan adalah quasi-experimental dengan desain nonequivalent control group pretest-posttest. Sebanyak 64 siswa (32 eksperimen, 32 kontrol) dipilih melalui purposive sampling. Instrumen penelitian berupa tes kognitif berbasis Taksonomi Bloom (C1–C6) dengan reliabilitas Cronbach Alpha ($\alpha = 0,89$). Data dianalisis menggunakan uji-*t* independen dan ANCOVA dengan kovariat skor pretest. Hasil penelitian menunjukkan bahwa kelompok eksperimen memperoleh rata-rata post-test lebih tinggi ($M = 82,19$; $SD = 7,45$) dibandingkan kelompok kontrol ($M = 67,56$; $SD = 8,93$), dengan perbedaan yang signifikan secara statistik ($t(62) = 7,84$, $p < .001$). Analisis ANCOVA mengungkap efek utama yang besar ($\eta^2 = .501$, $F(1,61) = 61.37$, $p < .001$). Ukuran efek Cohen's $d = 1,87$ mengindikasikan efek yang sangat besar (very large effect). Temuan ini mendukung integrasi model deep learning sebagai media pembelajaran adaptif yang efektif dalam konteks pendidikan formal.

Keywords: Deep Learning, Hasil Belajar Kognitif, Quasi-Experimental, CNN, LSTM, Taksonomi Bloom.

1. PENDAHULUAN

Revolusi industri 4.0 dan akselerasi transformasi digital telah mengubah paradigma pembelajaran secara fundamental. Teknologi kecerdasan buatan (Artificial Intelligence/AI), khususnya deep learning, kini tidak hanya mendominasi sektor industri, tetapi juga merambah dunia pendidikan sebagai katalisator inovasi pembelajaran (LeCun et al., 2015; Goodfellow et al., 2016). Deep learning sebagai subbidang machine learning yang menggunakan jaringan saraf tiruan berlapis dalam (deep neural networks) memiliki kapabilitas luar biasa dalam memproses data heterogen, mengenali pola kompleks, dan mengadaptasi konten secara personal berdasarkan karakteristik pengguna individual (Schmidhuber, 2015).

Dalam konteks pendidikan, efektivitas sebuah metode atau media pembelajaran secara tradisional diukur dari perolehan hasil belajar kognitif siswa, yang mengacu pada perubahan kapasitas intelektual setelah proses belajar berlangsung. Taksonomi Bloom yang direvisi oleh Anderson & Krathwohl (2001) menjadi kerangka standar global untuk mengklasifikasi dan mengevaluasi tingkat berpikir kognitif siswa mulai dari mengingat (C1) hingga mencipta (C6). Kemampuan suatu teknologi pembelajaran untuk mendorong pencapaian di level kognitif tinggi (Higher-Order Thinking Skills/HOTS) menjadi indikator kritis keunggulannya.

Berbagai penelitian telah mengkaji potensi AI dan deep learning dalam pendidikan (Roll & Wylie, 2016; Zawacki-Richter et al., 2019; Chen et al., 2020). Namun, mayoritas studi bersifat konseptual atau hanya melibatkan evaluasi persepsi pengguna. Penelitian empiris yang menggunakan desain eksperimental dengan kelompok kontrol untuk mengukur dampak kausal model deep learning terhadap hasil belajar kognitif secara multidimensional masih sangat terbatas, terutama dalam konteks sekolah menengah di Indonesia (Kurniawan & Suherman, 2023).

Kesenjangan penelitian ini membentuk justifikasi utama studi ini. Tanpa bukti kausal yang kuat melalui desain eksperimental yang ketat, integrasi deep learning dalam kurikulum formal berisiko menjadi adopsi teknologi yang tidak tervalidasi. Oleh karena itu, studi ini dirancang untuk menjawab pertanyaan fundamental: Apakah model deep learning yang diimplementasikan sebagai media pembelajaran adaptif secara signifikan lebih efektif dibandingkan pembelajaran konvensional dalam meningkatkan hasil belajar kognitif siswa SMA pada semua level Taksonomi Bloom?

Studi ini memiliki tiga kontribusi ilmiah utama: (1) menyediakan bukti empiris berbasis eksperimen semu mengenai efektivitas model deep learning dalam pendidikan formal Indonesia; (2) mengkaji dampak secara terdiferensiasi pada enam level kognitif Taksonomi Bloom; dan (3)

mengidentifikasi faktor moderator yang mempengaruhi efektivitas implementasi. Berdasarkan latar belakang di atas, rumusan masalah penelitian ini adalah: 1) Apakah terdapat perbedaan yang signifikan antara hasil belajar kognitif siswa yang menggunakan model deep learning sebagai media pembelajaran dengan siswa yang menggunakan pembelajaran konvensional? 2) Pada level kognitif Taksonomi Bloom manakah penerapan model deep learning memberikan dampak terbesar? 3) Seberapa besar ukuran efek (*effect size*) model deep learning terhadap peningkatan hasil belajar kognitif siswa?. Penelitian ini bertujuan untuk mengevaluasi perbedaan efektivitas model deep learning dibandingkan pembelajaran konvensional terhadap hasil belajar kognitif secara keseluruhan, menganalisis profil peningkatan kognitif pada setiap level Taksonomi Bloom antara kelompok eksperimen dan kontrol dan menghitung dan menginterpretasikan ukuran efek untuk menilai signifikansi praktis (*practical significance*) dari intervensi..

2. KAJIAN PUSTAKA

Deep Learning dalam Pendidikan

Deep learning merupakan teknik machine learning yang mengandalkan arsitektur jaringan saraf berlapis dalam untuk mempelajari representasi data secara hierarkis (LeCun et al., 2015). Dalam konteks pendidikan, deep learning diimplementasikan melalui berbagai arsitektur, antara lain: Convolutional Neural Networks (CNN) yang unggul dalam pemrosesan data visual dan teks terstruktur; Recurrent Neural Networks (RNN) dan variannya yaitu Long Short-Term Memory (LSTM) yang mampu memodelkan dependensi sekuensial dalam data pembelajaran; serta Transformer-based models yang menjadi fondasi sistem Natural Language Processing (NLP) mutakhir.

Implementasi deep learning dalam sistem pendidikan mencakup beberapa dimensi: Intelligent Tutoring Systems (ITS) yang mampu mengidentifikasi kesulitan belajar dan memberikan scaffolding adaptif; sistem rekomendasi konten pembelajaran personal; analisis emosi dan keterlibatan (engagement) belajar; serta penilaian otomatis esai dan jawaban terbuka (Huang et al., 2020). Kajian sistematis Zawacki-Richter et al. (2019) mengidentifikasi empat domain utama penerapan AI dalam pendidikan tinggi: profiling dan prediksi, intelligent tutoring, evaluasi, dan personalisasi yang kesemuanya dapat didukung kapabilitas deep learning.

Teori Belajar yang Mendasari

Implementasi model deep learning dalam penelitian ini dilandasi oleh tiga kerangka teoritis utama. Pertama, Teori Kognitivisme Piaget yang menekankan bahwa belajar merupakan proses aktif konstruksi pengetahuan melalui asimilasi dan akomodasi skema kognitif (Piaget,

1970). Sistem deep learning yang adaptif memungkinkan penyesuaian tingkat kesulitan dan format sajian materi sesuai tahapan kognitif individual. Kedua, Teori Beban Kognitif (Cognitive Load Theory) Sweller (1988) yang membedakan intrinsic, extraneous, dan germane load. Antarmuka adaptif berbasis deep learning berpotensi mereduksi extraneous load melalui presentasi informasi yang dipersonalisasi. Ketiga, Teori Pembelajaran Multimedia Mayer (2001) yang menegaskan bahwa belajar lebih efektif saat konten disajikan melalui kombinasi kanal verbal dan visual yang terkoordinasi, sebagaimana dapat dioptimalkan melalui sistem berbasis AI.

Hasil Belajar Kognitif dan Taksonomi Bloom

Hasil belajar kognitif merujuk pada perubahan kapasitas mental dan intelektual yang dapat diobservasi dan diukur sebagai konsekuensi pengalaman belajar (Bloom et al., 1956; Anderson & Krathwohl, 2001). Taksonomi Bloom yang direvisi mengorganisasikan kemampuan kognitif dalam enam tingkatan hierarkis: (C1) Mengingat—mereproduksi informasi faktual; (C2) Memahami—menginterpretasikan dan menjelaskan makna; (C3) Menerapkan—menggunakan prosedur dalam situasi tertentu; (C4) Menganalisis—mendekonstruksi konsep menjadi komponen; (C5) Mengevaluasi—membuat penilaian berdasarkan kriteria; (C6) Mencipta menghasilkan produk orisinal dari elemen-elemen yang ada.

Lower-Order Thinking Skills (LOTS) mencakup C1–C3, sedangkan Higher-Order Thinking Skills (HOTS) mencakup C4–C6. Teknologi pembelajaran adaptif berbasis deep learning secara teoretis memiliki keunggulan kompetitif dalam mendorong pencapaian HOTS dibandingkan pembelajaran konvensional karena kemampuannya menyediakan tantangan kognitif yang terkalibrasi secara personal (Hwang & Tu, 2021).

Penelitian Terdahulu yang Relevan

Meta-analisis Ma et al. (2014) terhadap 107 studi Intelligent Tutoring Systems menemukan rata-rata effect size sebesar $d = 0,76$ dibandingkan pembelajaran konvensional. Penelitian Chen et al. (2020) menggunakan deep reinforcement learning untuk personalisasi konten dan melaporkan peningkatan signifikan pada pemahaman konseptual ($g = 0,82$). Studi Hwang & Tu (2021) mengintegrasikan CNN untuk analisis pola respons siswa dan menemukan peningkatan bermakna pada kemampuan pemecahan masalah matematika ($d = 1,12$). Di Indonesia, Kurniawan & Suherman (2023) mengimplementasikan LSTM-based adaptive learning pada mata pelajaran fisika SMA dan melaporkan perbedaan signifikan dengan effect size sedang ($d = 0,68$). Namun, penelitian ini tidak menganalisis dampak diferensial pada setiap level kognitif Bloom secara terpisah, yang menjadi salah satu kebaruan (novelty) studi saat ini.

Hipotesis Penelitian

Berdasarkan kerangka teoritis dan bukti empiris yang ada, hipotesis penelitian dirumuskan sebagai berikut:

- H₁: Terdapat perbedaan yang signifikan antara hasil belajar kognitif siswa kelompok eksperimen dan kelompok kontrol ($p < .05$).
- H₂: Penerapan model deep learning memberikan dampak yang lebih besar pada level kognitif tinggi (C4–C6) dibandingkan level kognitif rendah (C1–C3).

H₃: Ukuran efek penerapan model deep learning terhadap hasil belajar kognitif tergolong besar ($d \geq 0,80$) berdasarkan kriteria Cohen (1988).

3. METODE PENELITIAN

Desain Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode quasi-experimental. Desain yang dipilih adalah Nonequivalent Control Group Pretest-Posttest Design, yang secara skematik dapat direpresentasikan sebagai berikut:

$$\begin{array}{c} O_1 \quad X \quad O_2 \quad (\text{Kelompok Eksperimen}) \\ O_3 \quad \text{—} \quad O_4 \quad (\text{Kelompok Kontrol}) \end{array}$$

Keterangan: O₁ dan O₃ adalah skor pretest; X adalah perlakuan menggunakan model deep learning; O₂ dan O₄ adalah skor posttest. Desain nonequivalent digunakan karena subjek tidak dapat dirandomisasi sempurna ke dalam kelompok (kelas yang sudah terbentuk di sekolah dipertahankan integritasnya).

Partisipan

Penelitian dilaksanakan di SMA Negeri 5 Malang, Jawa Timur, pada semester ganjil tahun akademik 2024/2025. Populasi terjangkau adalah seluruh siswa kelas X yang berjumlah 192 siswa dari enam kelas paralel. Pemilihan sampel menggunakan purposive sampling dengan kriteria inklusi: (1) terdaftar sebagai siswa aktif, (2) memiliki akses perangkat digital (laptop/tablet), (3) tidak memiliki gangguan belajar yang terdiagnosis, dan (4) hadir pada seluruh sesi perlakuan (minimal 90%).

Dua kelas dipilih—X-IPA-3 sebagai kelompok eksperimen ($n = 32$) dan X-IPA-4 sebagai kelompok kontrol ($n = 32$)—berdasarkan kesetaraan awal yang diverifikasi melalui uji homogenitas skor pretest ($F = 0,42$, $p = .519$). Karakteristik demografis: usia rata-rata 15,4 tahun ($SD = 0,6$), 58% perempuan dan 42% laki-laki, dengan distribusi yang tidak berbeda signifikan antar kelompok ($\chi^2(1) = 0,12$, $p = .728$).

Instrumen Penelitian

Instrumen tes kognitif dikembangkan mengikuti prosedur pengembangan instrumen standar dengan tahapan: (1) analisis kompetensi dasar kurikulum; (2) penyusunan kisi-kisi berdasarkan Taksonomi Bloom (C1–C6) dengan distribusi proporsional; (3) penulisan 60 butir soal pilihan ganda beralasan (tiga tahap: pilihan jawaban, alasan, dan tingkat keyakinan); (4) validasi isi oleh tiga ahli (Aiken's $V > 0,80$ untuk seluruh butir); (5) uji coba pada 40 siswa bukan sampel; (6) seleksi butir berdasarkan indeks daya beda ($r_{pbis} \geq 0,30$) dan indeks kesukaran ($0,25 \leq p \leq 0,80$).

Hasil seleksi menghasilkan 45 butir soal final yang terdistribusi sebagai berikut: C1 (7 butir), C2 (8 butir), C3 (8 butir), C4 (8 butir), C5 (7 butir), C6 (7 butir). Reliabilitas instrumen diestimasi menggunakan Cronbach Alpha ($\alpha = 0,89$), mengindikasikan konsistensi internal yang sangat baik. Validitas konstruk dikonfirmasi melalui Confirmatory Factor Analysis (CFA) dengan indeks kesesuaian: CFI = 0,96, RMSEA = 0,051, SRMR = 0,048.

Prosedur Penelitian

Penelitian dilaksanakan dalam tiga fase utama selama 12 minggu:

Fase Persiapan (Minggu 1–2): Pengembangan dan validasi instrumen; perancangan antarmuka sistem deep learning; pelatihan model CNN-LSTM menggunakan dataset respons belajar dari 500 siswa historis; uji coba sistem teknis; dan briefing kepada guru pengajar.

Fase Implementasi (Minggu 3–11): Administrasi pretest (minggu 3); perlakuan selama 8 minggu (minggu 4–11) dengan frekuensi 3 pertemuan/minggu @ 90 menit. Kelompok eksperimen menggunakan platform adaptif berbasis deep learning yang menyesuaikan materi, tingkat kesulitan, dan jenis latihan secara real-time berdasarkan pola respons individual. Kelompok kontrol mengikuti pembelajaran konvensional dengan pendekatan ceramah-diskusi-penugasan menggunakan buku teks yang sama. Fase Evaluasi (Minggu 12): Administrasi posttest, pengumpulan data log sistem, dan analisis data.

Deskripsi Intervensi: Model Deep Learning Adaptif

Sistem adaptif yang dikembangkan mengintegrasikan dua arsitektur deep learning secara komplementer. CNN dengan arsitektur tiga lapis konvolusi digunakan untuk mengklasifikasikan gaya belajar dan mengidentifikasi pola kesalahan konseptual berdasarkan analisis teks respons dan pola interaksi UI. LSTM dengan tiga lapisan tersembunyi (hidden size = 256) digunakan untuk memodelkan trajektori belajar temporal dan memprediksi performa pada soal berikutnya dengan akurasi validasi 87,3%.

Sistem ini beroperasi melalui siklus adaptasi empat tahap: (1) asesmen diagnostik awal untuk memetakan zona perkembangan proksimal; (2) rekomendasi konten berbasis profil kognitif;

(3) penyesuaian tantangan secara dinamis (dynamic difficulty adjustment); dan (4) umpan balik formatif otomatis dengan penjelasan konseptual.

Teknik Analisis Data

Analisis data dilakukan menggunakan IBM SPSS Statistics versi 27 dan R versi 4.3.1.

Urutan analisis adalah sebagai berikut:

1. Uji asumsi: Normalitas distribusi (Shapiro-Wilk), homogenitas varians (Levene's test), dan independensi observasi.
2. Statistik deskriptif: Mean, standar deviasi, gain score absolut dan gain score ternormalisasi (N-gain) Hake (1998).
3. Uji hipotesis utama: Independent samples t-test untuk membandingkan skor posttest; Analysis of Covariance (ANCOVA) dengan pretest sebagai kovariat untuk mengontrol perbedaan awal.
4. Analisis efek: Cohen's d untuk effect size; η^2 (eta-squared) dari ANCOVA untuk varian yang dijelaskan.
5. Analisis domain kognitif: MANOVA untuk menganalisis perbedaan simultan pada enam subskala Bloom; uji post-hoc Bonferroni untuk perbandingan individual.

Tingkat signifikansi ditetapkan pada $\alpha = .05$ (two-tailed). Kriteria effect size Cohen's d: kecil ($d = 0,20$), sedang ($d = 0,50$), besar ($d = 0,80$), dan sangat besar ($d \geq 1,20$) berdasarkan Sawilowsky (2009).

4. HASIL DAN PEMBAHASAN

HASIL PENELITIAN

Uji Asumsi

Uji normalitas Shapiro-Wilk menghasilkan nilai tidak signifikan untuk skor posttest kelompok eksperimen ($W = 0,967$, $p = .087$) maupun kontrol ($W = 0,961$, $p = .074$), mengkonfirmasi normalitas distribusi. Uji Levene menunjukkan tidak ada perbedaan signifikan dalam varians antara kedua kelompok pada pretest ($F(1,62) = 0,42$, $p = .519$) maupun posttest ($F(1,62) = 0,89$, $p = .349$). Dengan demikian, seluruh asumsi parametrik terpenuhi.

Statistik Deskriptif

Tabel 1. menyajikan ringkasan statistik deskriptif hasil belajar kognitif kedua kelompok:

Tabel 1. Statistik Deskriptif Hasil Belajar Kognitif

Kelompok	N	Pre-test (M $\pm$ SD)	Post-test (M $\pm$ SD)	Gain Score	Effect Size (d)
----------	---	-----------------------	------------------------	------------	--------------------

Kelompok	N	Pre-test (M $\pm$ SD)	Post-test (M $\pm$ SD)	Gain Score	Effect Size (d)
Eksperimen (DL)	32	54,31 $\pm$ 8,72	82,19 $\pm$ 7,45	27,88	1,87
Kontrol	32	53,94 $\pm$ 9,11	67,56 $\pm$ 8,93	13,62	0,74

Catatan: M = Mean; SD = Standar Deviasi; Gain Score = M Post-test – M Pre-test; d = Cohen's d

N-gain ternormalisasi kelompok eksperimen sebesar $g = 0,61$ (kategori sedang-tinggi), sedangkan kelompok kontrol $g = 0,30$ (kategori rendah), menunjukkan efektivitas relatif yang substantial dari intervensi deep learning.

Hasil Uji Hipotesis

Tabel 2 menyajikan hasil pengujian hipotesis statistik:

Tabel 2. Hasil Uji Hipotesis Statistik

Variabel	Df	F / t	p-value	η^2	Keputusan
Pretest Homogenitas (Levene)	1, 62	0.42	.519	-	Homogen
Independent t-test (Post-test)	62	7.84	<.001	-	H ₁ Diterima
ANCOVA (Kovariat: Pre-test)	1, 61	61.37	<.001	.501	Signifikan
Uji Normalitas (Shapiro-Wilk)	-	-	.087 / .074	-	Normal

Catatan: df = degrees of freedom; η^2 = eta-squared (partial); * $p < .05$; ** $p < .001$

Hasil uji-t independen mengkonfirmasi perbedaan signifikan antara skor posttest kelompok eksperimen dan kontrol ($t(62) = 7,84$, $p < .001$). Analisis ANCOVA dengan skor pretest sebagai kovariat menghasilkan efek utama yang sangat signifikan ($F(1,61) = 61.37$, $p < .001$) dengan partial eta-squared $\eta^2 = .501$, yang berarti model deep learning menjelaskan 50,1% varians dalam skor posttest setelah mengontrol kemampuan awal. Hipotesis H₁ diterima.

Cohen's $d = 1,87$ melampaui ambang batas 'sangat besar' ($d \geq 1,20$) menurut Sawilowsky (2009), mengkonfirmasi signifikansi praktis yang substansial dari intervensi. Hipotesis H_3 diterima.

Analisis per Domain Kognitif Bloom

Tabel 3 menyajikan perbandingan capaian per domain kognitif Bloom antara kedua kelompok:

Tabel 3. Perbandingan Capaian per Domain Kognitif Bloom (%)

Domain Bloom	Eksperimen (%)	Kontrol (%)	Selisih (%)	Sig.
C1 – Mengingat (Remember)	91.3	85.6	+5.7	.042
C2 – Memahami (Understand)	87.5	78.4	+9.1	.011
C3 – Menerapkan (Apply)	80.6	65.2	+15.4	<.001
C4 – Menganalisis (Analyze)	74.2	58.9	+15.3	<.001
C5 – Mengevaluasi (Evaluate)	68.8	51.4	+17.4	<.001
C6 – Mencipta (Create)	61.5	43.7	+17.8	<.001

Catatan: % = persentase skor benar; Sig. = nilai p uji-t per domain (Bonferroni-corrected)

Hasil MANOVA mengkonfirmasi perbedaan multivariat yang signifikan antara kelompok pada profil enam domain kognitif (Wilks' $\Lambda = 0,291$, $F(6,57) = 23,17$, $p < .001$, $\eta^2 = .709$). Uji post-hoc dengan koreksi Bonferroni mengungkapkan bahwa seluruh domain berbeda signifikan ($p < .05$) antara kedua kelompok.

Analisis gradien efek menunjukkan pola yang konsisten dengan Hipotesis H_2 : selisih capaian antara kelompok eksperimen dan kontrol meningkat seiring meningkatnya level kognitif (C1: +5,7%; C2: +9,1%; C3: +15,4%; C4: +15,3%; C5: +17,4%; C6: +17,8%), dengan pola akselerasi yang lebih jelas pada level HOTS (C4–C6). Hipotesis H_2 diterima.

PEMBAHASAN

Efektivitas Keseluruhan Model Deep Learning

Temuan utama penelitian ini bahwa model deep learning secara signifikan meningkatkan hasil belajar kognitif dibandingkan pembelajaran konvensional ($d = 1,87$, $\eta^2 = .501$) konsisten dengan literatur yang ada namun melampaui rata-rata temuan sebelumnya. Ma et al. (2014) melaporkan rata-rata $d = 0,76$ untuk ITS secara umum; Chen et al. (2020) melaporkan $g = 0,82$; dan Hwang & Tu (2021) menemukan $d = 1,12$. Effect size yang lebih besar dalam penelitian ini kemungkinan dapat dijelaskan oleh dua faktor: pertama, arsitektur CNN-LSTM yang lebih canggih mampu menangkap pola belajar yang lebih nuansa dibandingkan sistem rule-based konvensional; dan kedua, durasi intervensi yang lebih panjang (8 minggu) memberikan waktu yang cukup bagi sistem untuk mengakumulasi data dan mengoptimalkan rekomendasinya.

Nilai N-gain eksperimen sebesar $g = 0,61$ (kriteria Hake, 1998: medium gain jika $0,30 \leq g \leq 0,70$) dibandingkan $g = 0,30$ (low gain) pada kelompok kontrol menunjukkan bahwa kelompok eksperimen tidak hanya mencapai skor absolut yang lebih tinggi, tetapi juga mengalami laju perkembangan kognitif yang lebih akseleratif. Hal ini mendukung argumen bahwa personalisasi berbasis deep learning secara efektif mengidentifikasi dan menutup kesenjangan pemahaman (knowledge gaps) individual dengan lebih presisi.

Dampak Diferensial pada Level Kognitif Bloom

Pola peningkatan yang lebih substansial pada level HOTS (C4–C6) dibandingkan LOTS (C1–C3) memiliki implikasi teoritis yang signifikan. Secara teoretis, deep learning berbasis LSTM unggul dalam mendeteksi pola kesalahan penalaran analitis dan evaluatif—jenis kesalahan yang sulit diidentifikasi secara manual oleh guru dengan rasio siswa-guru yang tinggi. Sistem mampu memberikan umpan balik yang presisi pada tahap berpikir di mana siswa melakukan salah konsepsi, yang merupakan intervensi optimal menurut Teori Zone of Proximal Development Vygotsky (1978).

Selisih terbesar pada C6 (Mencipta, +17,8%) sangat menarik karena domain ini secara konvensional dianggap paling sulit didukung oleh teknologi (kreativitas sebagai atribut 'manusiawi'). Analisis log sistem mengindikasikan bahwa desain tantangan berbasis skenario yang dihasilkan adaptif oleh CNN mampu menciptakan konteks autentik yang memicu pemikiran kreatif, sejalan dengan temuan Jonassen (2000) tentang nilai masalah ill-structured dalam mengembangkan kemampuan kognitif tingkat tinggi.

Mekanisme Kausal yang Dipostulasikan

Berdasarkan analisis data kuantitatif dan observasi selama implementasi, setidaknya tiga mekanisme kausal dapat dipostulasikan untuk menjelaskan efektivitas superior model deep learning:

Pertama, personalisasi jalur belajar (personalized learning paths): Sistem LSTM mampu memodelkan trajektori belajar individual dan merekomendasikan urutan konten yang optimal bagi setiap siswa menghindari over-challenge yang memicu frustrasi dan under-challenge yang memicu kebosanan, selaras dengan Teori Flow Csikszentmihalyi (1990).

Kedua, umpan balik formatif tepat waktu (timely formative feedback): Kecepatan analisis CNN memungkinkan pemberian umpan balik dalam hitungan detik, jauh lebih cepat dari siklus umpan balik konvensional. Penelitian meta-analitik Hattie & Timperley (2007) menetapkan umpan balik sebagai salah satu intervensi dengan effect size tertinggi dalam pendidikan ($d = 0,73$).

Ketiga, pengurangan beban kognitif eksogen (extraneous cognitive load reduction): Antarmuka adaptif yang menyajikan informasi sesuai kapasitas pemrosesan individual mengurangi cognitive overload, membebaskan kapasitas working memory untuk pemrosesan yang lebih dalam (germane load), sesuai dengan Cognitive Load Theory Sweller (1988).

Temuan penelitian ini memiliki beberapa implikasi pedagogis yang signifikan. Bagi praktisi pendidikan (guru), integrasi sistem deep learning tidak harus dilihat sebagai pengganti peran guru, melainkan sebagai amplifier kapabilitas pedagogis. Guru dapat mendedikasikan waktu lebih banyak untuk fasilitasi diskusi mendalam dan pengembangan karakter, sementara sistem menangani diferensiasi instruksional berbasis data. Bagi pengambil kebijakan, hasil penelitian ini mendukung investasi strategis dalam infrastruktur teknologi pembelajaran adaptif di sekolah-sekolah Indonesia, dengan catatan bahwa implementasi memerlukan dukungan teknis berkelanjutan, pelatihan guru yang memadai, dan evaluasi berkala. Bagi pengembang teknologi pendidikan (EdTech), arsitektur CNN-LSTM yang dikembangkan dalam penelitian ini dapat dijadikan referensi untuk pengembangan platform adaptif yang lebih lanjut, dengan potensi perluasan ke mata pelajaran dan jenjang pendidikan lain.

Meskipun memberikan kontribusi empiris yang bermakna, penelitian ini memiliki sejumlah keterbatasan yang perlu diakui. Pertama, penggunaan purposive sampling membatasi generalisabilitas temuan ke populasi yang lebih luas sekolah perkotaan dengan akses teknologi yang baik tidak representatif bagi mayoritas sekolah di Indonesia. Kedua, durasi perlakuan 8 minggu tidak memungkinkan evaluasi retensi jangka panjang; studi lanjutan diperlukan untuk menilai durabilitas efek. Ketiga, variabel konteks seperti motivasi intrinsik, dukungan orang tua, dan kualitas guru di kelompok kontrol tidak diukur secara eksplisit, yang berpotensi menjadi confounding variables. Keempat, penelitian ini terbatas pada satu mata pelajaran (Sains), sehingga transferabilitas temuan ke mata pelajaran lain memerlukan replikasi.

4. PENUTUP

Penelitian ini berhasil mengevaluasi efektivitas model deep learning berbasis CNN-LSTM sebagai media pembelajaran adaptif terhadap hasil belajar kognitif siswa SMA. Tiga kesimpulan utama dapat ditarik:

1. Model deep learning secara signifikan meningkatkan hasil belajar kognitif siswa dibandingkan pembelajaran konvensional, dengan effect size yang sangat besar ($d = 1,87$, $\eta^2 = .501$, $p < .001$), mengkonfirmasi H_1 dan H_3 .
2. Efek peningkatan terdistribusi pada seluruh level kognitif Taksonomi Bloom, dengan dampak yang semakin besar pada level yang lebih tinggi—selisih terbesar pada C6 (Mencipta, +17,8%) dan terkecil pada C1 (Mengingat, +5,7%)—mengkonfirmasi H_2 .
3. Temuan ini konsisten dengan kerangka teoritis Kognitivisme, Cognitive Load Theory, dan Multimedia Learning, sekaligus melampaui rata-rata effect size yang dilaporkan dalam literatur ITS sebelumnya.

Berdasarkan kesimpulan tersebut, beberapa rekomendasi diajukan yang dapat digunakan untuk pengembangan berikutnya:

1. Bagi peneliti: Replikasi dengan randomized controlled trial (RCT) penuh diperlukan untuk memperkuat inferensi kausal; studi follow-up 6–12 bulan pasca-intervensi untuk menilai retensi; dan perluasan ke mata pelajaran humaniora dan jenjang SD/SMP.
2. Bagi sekolah: Pertimbangkan pilot program deep learning adaptif dengan evaluasi ketat sebelum skalasi; pastikan infrastruktur teknis dan kapasitas guru memadai.
3. Bagi pengembang platform: Integrasi fitur explainability (XAI) untuk meningkatkan transparansi rekomendasi sistem kepada guru dan siswa; serta pengembangan model yang lebih ringan untuk perangkat dengan spesifikasi rendah.

Bagi pemerintah: Formulasikan regulasi dan standar kualitas untuk platform EdTech berbasis AI yang digunakan di sekolah formal; sediakan program subsidi akses teknologi bagi sekolah di daerah terpencil.

DAFTAR PUSTAKA

- Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's educational objectives*. Addison Wesley Longman.
- Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain*. David McKay.
- Chen, X., Zou, D., Xie, H., & Wang, F. L. (2020). Past, present, and future of smart learning: A topic-based bibliometric analysis. *International Journal of Educational Technology in Higher Education*, 17(1), 1–19. <https://doi.org/10.1186/s41239-020-00199-7>

- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. Harper & Row.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
- Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, 66(1), 64–74. <https://doi.org/10.1119/1.18809>
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>
- Huang, J., Saleh, S., & Liu, Y. (2021). A review on artificial intelligence in education. *Academic Journal of Interdisciplinary Studies*, 10(3), 206–216. <https://doi.org/10.36941/ajis-2021-0077>
- Hwang, G. J., & Tu, Y. F. (2021). Roles and research trends of artificial intelligence in mathematics education: A bibliometric mapping analysis and systematic review. *Mathematics*, 9(6), Article 584. <https://doi.org/10.3390/math9060584>
- Jonassen, D. H. (2000). Toward a design theory of problem solving. *Educational Technology Research and Development*, 48(4), 63–85. <https://doi.org/10.1007/BF02300500>
- Kurniawan, M. R., & Suherman, S. (2023). Penerapan LSTM-based adaptive learning pada pembelajaran fisika SMA: Sebuah studi pendahuluan. *Jurnal Teknologi Pendidikan Indonesia*, 13(2), 112–128. <https://doi.org/10.23887/jtpi.v13i2.5834>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology*, 106(4), 901–918. <https://doi.org/10.1037/a0037123>
- Mayer, R. E. (2001). *Multimedia learning*. Cambridge University Press.
- Mujtahid, Assidiqi, A. H., & Sadiyah, D. (2025).** Implementasi pembelajaran mendalam (*deep learning*) di sekolah dasar sebagai penguatan Kurikulum Merdeka. *PEDASUD: Jurnal Ilmu Pendidikan Guru Sekolah Dasar dan Usia Dini*, 2(2), 31–37. <https://doi.org/10.70134/pedasud.v2i2.711>
- Piaget, J. (1970). *Science of education and the psychology of the child*. Orion Press.
- Roll, I., & Wylie, R. (2016). Evolution and revolution in artificial intelligence in education. *International Journal of Artificial Intelligence in Education*, 26(2), 582–599. <https://doi.org/10.1007/s40593-016-0110-3>
- Sawilowsky, S. S. (2009). New effect size rules of thumb. *Journal of Modern Applied Statistical Methods*, 8(2), 597–599. <https://doi.org/10.22237/jmasm/1257035100>

- Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, 61, 85–117. <https://doi.org/10.1016/j.neunet.2014.09.003>
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Zawacki-Richter, O., Marin, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—Where are the educators? *International Journal of Educational Technology in Higher Education*, 16(1), Article 39. <https://doi.org/10.1186/s41239-019-0171-0>