

Deep Learning Applications in Primary Education: A Systematic Literature Review of Emerging Trends, Challenges, and Opportunities

Ricki Fahma Adi Saputra^{1,*}, Muhammad Ridha¹, & Jamiu Temitope Sulaimon²

¹Elementary School Teacher Education Study Program, Universitas Ahmad Dahlan, Indonesia

²Department of Adult and Primary Education, University of Ilorin, Nigeria

*Corresponding email: 2200005209@webmail.uad.ac.id

Received: 10 August 2025

Accepted: 31 August 2025

Published: 13 September 2025

Abstract: Deep Learning Applications in Primary Education: A Systematic Literature Review of Emerging Trends, Challenges, and Opportunities. Objectives: Despite significant advances in healthcare and finance, Deep Learning (DL) remains underutilized in primary education, with only about 12% of AI-related studies focusing on K–6. Given the formative role of early schooling in cognitive and social development, this systematic review analyzed recent empirical studies to identify key trends, challenges, and opportunities in applying DL to primary education. **Methods:** Following the PRISMA 2020 guidelines, 21 studies published between 2021 and 2025 across 11 countries were analyzed using a structured coding sheet and further examined with bibliometric mapping, *Microsoft Excel*, and *Python*-based visualization tools. The review included diverse sources focusing on DL use cases across global primary education systems. **Findings:** The findings reveal a global increase in DL adoption, led by China and South Korea, with growing contributions from Indonesia, India, and the Philippines. Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Generative Adversarial Networks (GANs), and transformer-based models are among the most commonly applied architectures to tasks such as personalized learning, early detection of learning difficulties, assessment automation, and curriculum development. However, there are still problems in the context of ethical concerns, including data privacy, algorithmic bias, and equity of access. Technical barriers involve dataset complexity, model generalization, and resource limitations, while pedagogical issues center on aligning DL applications with developmental needs and classroom realities. Despite these obstacles, DL demonstrates significant potential to enhance personalization, foster engagement, and support holistic educational outcomes. **Conclusion:** This review contributes a strategic roadmap for integrating DL in primary education by balancing innovation with pedagogy and ethics. Future research should prioritize cross-disciplinary collaboration, greater geographic diversity, and improvements in scalability and interpretability to ensure DL supports equitable, future-ready learning.

Keywords: deep learning, primary education, personalized learning, systematic literature review, educational technology.

To cite this article:

Saputra, R. F. A., Ridha, M., & Sulaimon, J. T. (2025). Deep Learning Applications in Primary Education: A Systematic Literature Review of Emerging Trends, Challenges, and Opportunities. *Jurnal Pendidikan Progresif*, 15(3), 1785-1810. doi: 10.23960/jpp.v15i3.pp1785-1810.

■ INTRODUCTION

Deep Learning (DL) has transformed critical domains such as healthcare, finance, and transportation through predictive analytics, real-

time decision-making, and automation (Razzaq & Shah, 2025; Sarker, 2021; Taye, 2023). Its capacity to process massive amounts of unstructured data and generate high-accuracy

predictions has revolutionized medical diagnostics, credit risk assessment, and autonomous systems. These successes illustrate DL's transformative potential across knowledge domains. However, paradoxically, its integration into education, particularly primary education (K–6), remains fragmented and underdeveloped.

A fundamental barrier is the persistent conceptual confusion between DL as a technology defined through neural network architectures such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) and deep learning as pedagogy, which emphasizes critical thinking, collaboration, creativity, and deep engagement with content (Kovač, Nome, Jensen, & Skreland, 2025; Pan et al., 2023; Yue, Jong, & Dai, 2022). This overlap has limited the coherence of program design. Some initiatives prioritize technological sophistication without considering the pedagogical dimension, while others emphasize pedagogy but lack technological feasibility (Rui, Nasri, & Mahmud, 2024). Our systematic review seeks to reconcile these two interpretations by clarifying how DL technologies can be aligned with pedagogical objectives in the primary school context.

The urgency of addressing this gap is underscored by the formative importance of early education. Primary schooling lays the foundation for lifelong learning, shaping children's cognitive schemas, socio-emotional development, problem-solving skills, and intrinsic motivation to learn (Bernacki, Greene, & Lobczowski, 2021; Van Schoors, Elen, Raes, & Depaepe, 2021). Interventions that fail to align technological tools with children's developmental needs risk cognitive overload and superficial engagement. Alarming, recent evidence shows that only about 12% of AI-based educational studies focus on K–6 learners, while almost 70% are devoted to higher education (Bhutoria, 2022; Naseer, Khan, Tahir, Addas, & Aejaz, 2024). This underrepresentation is concerning, given the role of

primary education in cognitive and socio-emotional development.

Despite its promise, most research positions DL narrowly within educational data mining and algorithmic design, overlooking the psychological, ethical, and pedagogical dimensions that are crucial in primary classrooms (Lu, Griffin, Sadler, Laffey, & Goggins, 2025; Ozyurt, Ozyurt, & Mishra, 2023). For example, studies on personalized learning often treat students as data points rather than as developing individuals with diverse cognitive and socio-emotional needs. However, international agendas such as UNESCO's Education 2030 Framework and the United Nations' Sustainable Development Goal 4 (SDG 4) explicitly call for inclusive, equitable, and quality education. Achieving these goals requires DL applications that are developmentally appropriate, teacher-friendly, and pedagogically meaningful.

In parallel, global education policies are increasingly calling for personalized, inclusive, and equitable learning models that adapt to the diverse needs of learners (Aamer, Ba-Alawi, Kang, Lee, & Jo, 2025; Almuhan, 2025; Imran, Almusharraf, Ahmed, & Mansoor, 2024). DL must be carefully adapted to children's developmental stages and teacher readiness to realize this vision. Our review, therefore, seeks to provide a roadmap for reconciling technological and pedagogical interpretations of DL, ensuring that innovation strengthens rather than disrupts the mission of early education.

Existing literature broadly discusses generic DL applications in education, but often neglects the unique pedagogical, ethical, and practical considerations of primary classrooms (Almalawi, Soh, Li, & Samra, 2024; López-Meneses, López-Catalán, Pelicano-Piris, & Mellado-Moreno, 2025; López-Meneses, Mellado-Moreno, Gallardo Herrerías, & Pelicano-Piris, 2025; Tzimas & Demetriadis, 2021). Several systematic reviews have examined AI or DL in

education, yet most of them focus on higher education or secondary schools, with limited attention to early schooling (Aravantos et al., 2024; S. J. Lee & Kwon, 2024; Martínez-Comesaña et al., 2023; Ouyang, Zheng, & Jiao, 2022). For instance, reviews of AI in personalized learning primarily emphasize university-level adoption, overlooking the distinct developmental, pedagogical, and ethical considerations required in K–6 settings. This underrepresentation highlights an urgent need for scholarship tailored to primary education.

Addressing this knowledge void is critical for at least three reasons. First, from a theoretical perspective, existing studies rarely articulate frameworks that integrate DL's technological and pedagogical interpretations coherently. This review seeks to bridge that divide by explicitly examining how both perspectives can be aligned to support primary education. Second, the policy relevance of this work cannot be understated. Educational policymakers worldwide are under increasing pressure to implement digital transformation strategies in line with Sustainable Development Goal 4 (SDG 4). However, there is still a lack of empirical evidence and practical guidance for applying DL in primary schools. Third, the practical implications are equally pressing, as teachers and schools are increasingly exposed to AI-based tools but remain uncertain about their developmental appropriateness, integration into classroom management, and training requirements necessary for successful adoption.

Building on these rationales, the present review introduces three key novelties absent from prior studies. First, it offers a specific and exclusive focus on DL applications in primary education (K–6), a domain that has been systematically overlooked in existing reviews. Second, it provides a time-sensitive coverage of recent literature published between 2021 and 2025, thereby capturing the latest innovations and

reflecting the acceleration of educational technology adoption in the post-pandemic era. Third, it adopts a rigorous synthesis approach by combining PRISMA 2020 guidelines with bibliometric mapping, enabling the identification of thematic clusters, empirical trends, and research gaps that cannot be captured through narrative reviews alone. Accordingly, this study is guided by three research questions:

- RQ1. What are the current trends in DL applications for primary education?
- RQ2. What ethical, technical, and pedagogical challenges arise when implementing DL at this level?
- RQ3. What opportunities exist to leverage DL for personalized learning, student engagement, and holistic outcomes in primary schools?

This study makes three significant contributions to the field. First, at the theoretical level, it advances the discourse by integrating technological and pedagogical perspectives of DL into a coherent framework for primary education, addressing a conceptual gap that has limited effective program design. Second, at the methodological level, it demonstrates the combined use of PRISMA 2020 protocols and bibliometric mapping, providing a transparent and replicable synthesis model that enhances the reliability of the findings compared to prior narrative reviews. Third, at the practical and policy level, it generates actionable insights for teachers, school leaders, and policymakers, particularly in guiding digital transformation strategies aligned with SDG 4. By highlighting issues of developmental appropriateness, ethical considerations, and teacher readiness, this review provides a roadmap for implementing DL responsibly in K–6 contexts.

By addressing these questions and contributions, the present systematic review not only consolidates current evidence but also

advances theoretical, policy, and practical discourse. It offers educators, researchers, and policymakers a roadmap of what has been achieved, what remains limited, and where future research and interventions should be directed. Ultimately, by conceptualizing DL both as a technology and a pedagogy, this review argues for responsible, developmentally sensitive, and equity-oriented integration of DL in primary classrooms.

■ METHOD

Research Design

This study employed a Systematic Literature Review (SLR) methodology to explore the application of Deep Learning (DL) approaches in primary school education, with a particular focus on their potential for personalized learning, student engagement, and improved educational outcomes. The SLR method was selected because it allows for a comprehensive, transparent, and reproducible synthesis of empirical findings across diverse studies (Krüger, Lausberger, von Nostitz-Wallwitz, Saake, & Leich, 2020; Paul, Lim, O’Cass, Hao, & Brescian i, 2021). The review strictly utilized the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) protocol (Page et al., 2021), ensuring methodological rigor and adherence to international standards for conducting systematic reviews.

Search Strategy

The primary database selected for this review was Scopus, chosen for its extensive coverage of education technology, computer science, and pedagogy-related publications. A comprehensive search was conducted on August 1, 2025, covering the period from January 2021

to August 2025 to capture the most recent and relevant developments. The search used Boolean operators across the title, abstract, and keyword fields (TITLE-ABS-KEY). The search terms combined DL-related concepts with primary education terms to ensure both precision and breadth of coverage. Specifically, the Boolean search string applied in Scopus was as follows. (TITLE-ABS-KEY (“deep learning” OR “machine learning” OR “artificial intelligence”)) AND TITLE-ABS-KEY (“primary education” OR “elementary school” OR “K-6” OR “early childhood education”)) AND TITLE-ABS-KEY (“personalized learning” OR “educational technology”)) AND PUBYEAR > 2021 AND PUBYEAR < 2025 AND (LIMIT-TO (DOCT YPE, “ar”)) AND (LIMIT-TO (LANGUAGE, “English”)) AND (LIMIT-TO (SRCTYPE, “j”)) AND (LIMIT-TO (PUBSTAGE, “final”)).

This query ensured that only peer-reviewed journal articles, published in English between 2021 and 2025 and finalized for publication, were included in the dataset.

Inclusion and Exclusion Criteria

A set of explicit inclusion and exclusion criteria was carefully developed to ensure that only the most relevant and methodologically sound studies were incorporated into this review. These criteria served as a guiding framework for screening and evaluating all retrieved articles, allowing the review to maintain a clear focus on Deep Learning (DL) applications within primary education while filtering out irrelevant or low-quality publications. By applying these standards consistently, the study aimed to enhance the credibility, transparency, and replicability of the review process. The criteria are summarized in Table 1.

Table 1. Inclusion and exclusion criteria

Criteria Type	Inclusion Criteria	Exclusion Criteria
Document Type	Only peer-reviewed journal articles and reputable conference proceedings	Book chapters, theses/dissertations, editorials, white papers, and any non-

	are retrieved from Scopus, ensuring academic rigor and credibility.	peer-reviewed or grey literature.
Publication Timeframe	Publications from January 2021 to August 2025 will reflect the most current developments and trends in DL and primary education.	Studies published before 2021 or after August 2025 are excluded from the review window.
Language	Publications written in English should be clear and consistent in interpretation.	Non-English publications should avoid translation bias and ensure accuracy.
Research Methodology	Empirical, experimental, or conceptual studies with a clearly defined DL application and transparent methodology relevant to education.	Purely technical studies focusing on algorithm development or model optimization without an educational context.
Educational Stage	Research explicitly addresses primary or elementary education (K–6), including personalized learning, student engagement, or assessment using DL.	Studies focusing exclusively on secondary schools, higher education, vocational training, or non-educational sectors.
Quality and Rigor	Studies demonstrating clear research objectives, robust study design, and well-documented findings.	Studies with vague aims, unclear methodology, or insufficient reporting on outcomes.

Quality Assurance Process

A multi-stage quality assurance strategy was applied throughout this SLR to ensure methodological rigor and maintain the integrity of the review process (Mwogosi & Mambile, 2025; Naghib, Gharehchopogh, & Zamanifar, 2025). First, two independent reviewers screened all retrieved titles and abstracts, consistently applying the inclusion and exclusion criteria. Any discrepancies in selection were resolved through consensus meetings, with a third reviewer adjudicating unresolved cases to minimize bias.

At the full-text review stage, each study was examined for methodological soundness, relevance to primary education, and clear articulation of Deep Learning (DL) applications. To strengthen the assessment, we applied the Critical Appraisal Skills Programme (CASP) checklist to evaluate the quality of all 21 included studies. The CASP framework enabled us to systematically assess research clarity, methodological appropriateness, data collection

rigor, ethical considerations, and the validity of reported findings (Long, French, & Brooks, 2020). Two independent reviewers conducted the CASP appraisal, and all disagreements were discussed until consensus was reached.

Furthermore, all data extraction sheets, including coding of DL models, educational purposes, datasets, and reported challenges, were cross-validated between reviewers to ensure consistency and accuracy. The results of the CASP appraisal were summarized in the Supplementary Material, providing a transparent overview of the methodological quality of the included studies. Finally, the synthesis phase was conducted collaboratively to guarantee that thematic interpretations and trend analyses were grounded in high-quality evidence and a shared understanding of the reviewed literature.

PRISMA Flow and Study Selection

This review adhered strictly to the PRISMA 2020 protocol (Page et al., 2021) to ensure a

transparent, reproducible, and rigorous process of identifying and selecting relevant studies. The initial database search on Scopus yielded 682 records, as shown in Figure 1. As part of the screening process, 190 duplicate entries were

removed, leaving 492 unique records for further examination. From these, 273 records were automatically excluded by Scopus filters for irrelevance or incomplete metadata, resulting in 219 full-text reports retrieved for assessment.

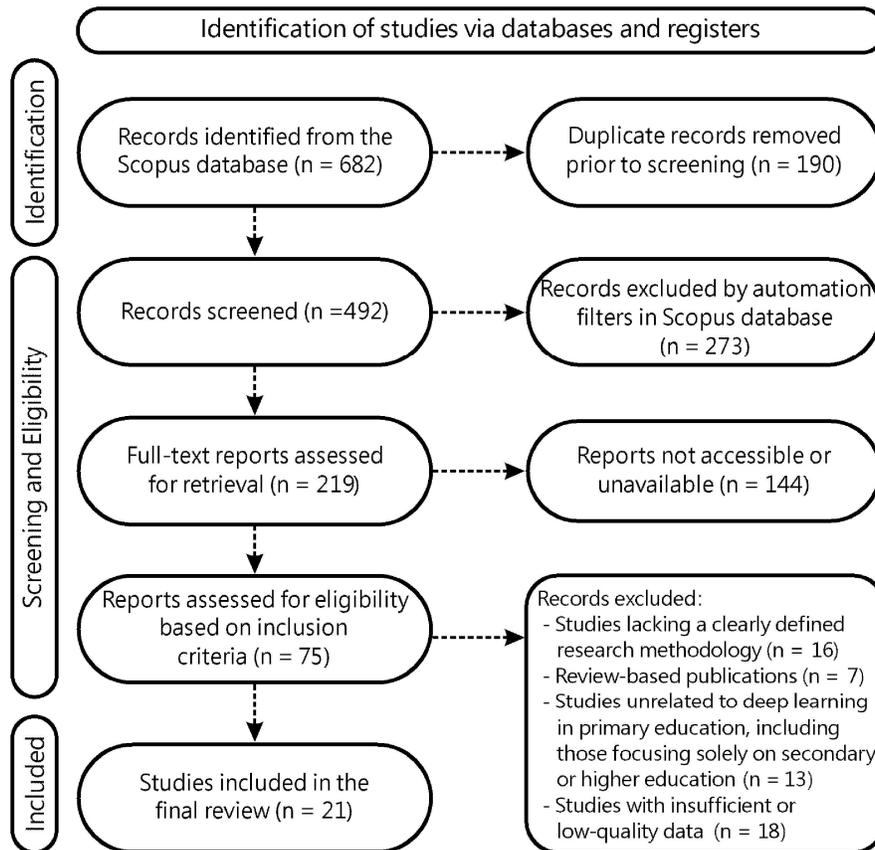


Figure 1. PRISMA flow diagram of study identification and selection workflow

However, due to technical and institutional limitations, 144 full-text articles could not be accessed. Most of these inaccessible articles (approximately 71%) were behind paywalls, 23% were unavailable due to limited institutional access, and 6% were not retrievable due to broken links or outdated repositories. As a result, 75 full-text studies were eligible for comprehensive screening based on predefined inclusion and exclusion criteria.

During the eligibility review, several studies were excluded for specific reasons: 16 lacked a clearly defined research methodology, 13 focused

on educational levels beyond primary education, 18 presented insufficient or poor-quality data, and 7 were review papers rather than original empirical or conceptual work. Additionally, 54 articles were excluded because they mentioned deep learning superficially without providing methodological applications relevant to primary education. After this systematic screening process, 21 studies met all inclusion criteria and were included in the final synthesis.

This refined selection process underscores the methodological rigor of the review, ensuring that only studies with firm empirical grounding,

contextual relevance to primary education (K–6), and clear, deep learning implementation were analyzed in depth.

Data Extraction and Synthesis Approach

For the 21 studies included in this review, information was systematically collected using a structured coding sheet to ensure organization and consistency. The coding process began with a deductive approach, guided by the three research questions (RQ1–RQ3), but allowed for inductive refinement when new patterns emerged. Each study was coded for attributes such as author, year, country, type of Deep Learning (DL) model (CNN, RNN, GAN, transformer, or hybrid), educational purpose (e.g., personalized learning, engagement, assessment), dataset type, preparation methods, and reported findings and challenges.

Two independent reviewers conducted the coding to strengthen reliability. Inter-rater reliability was calculated using Cohen's Kappa ($\hat{\kappa} = 0.82$), indicating strong agreement. When discrepancies arose, such as whether a study should be classified under "technical challenge" or "pedagogical limitation," these were resolved through discussion. A third reviewer was consulted in two cases to reach a final consensus.

The thematic analysis followed the six systematic phases of Braun & Clarke (2006): (1) familiarization with the data, (2) generating initial codes, (3) searching for themes, (4) reviewing themes, (5) defining and naming themes, and (6) producing the report (Byrne, 2022). This ensured transparency and methodological rigor in building themes directly aligned with the three guiding research questions.

The synthesis combined qualitative thematic analysis with quantitative mapping. Excel was used to tabulate the frequencies of DL models (Miñan, Moreno, & Fernández, 2023), dataset sources, and domains of application (RQ1). Barriers (ethical, technical, and pedagogical) were categorized (RQ2), and innovative opportunities

were clustered (RQ3). Additionally, Python, combined with NLTK (Natural Language Toolkit), was used to conduct text mining on abstracts and keywords. Specifically, the FreqDist function was used to calculate and analyze keyword frequency distributions. At the same time, matplotlib was employed to generate the visualization, where the yearly publication counts were plotted as line charts to illustrate the distribution and trend of keyword-related publications over time (Khandare, Agarwal, Bodhankar, Kulkarni, & Mane, 2023; Lavanya et al., 2023).

These visualizations, presented in the results and supplementary materials, highlight dominant research patterns and provide more transparency to the synthesis. By integrating systematic coding, inter-rater reliability testing, resolution of discrepancies, and advanced visualization, this approach ensures the synthesis is methodologically rigorous and interpretively robust, offering a reliable foundation for drawing insights on DL in primary education.

■ RESULT AND DISCUSSION

RQ1. What are the current trends in DL applications for primary education?

Deep Learning (DL) is a subset of machine learning that utilizes multi-layered artificial neural networks to automatically extract and learn high-level representations from raw data, enabling advanced pattern recognition, prediction, and decision-making. In educational contexts, DL can process complex inputs such as images, text, audio, and sensor data to support adaptive learning systems, automate assessments, and provide personalized feedback. Its ability to handle large-scale and unstructured data makes DL highly relevant for addressing diverse learning needs in primary education.

There has been a notable surge in scholarly interest in applying DL to primary education in recent years. This growing body of research reflects a global shift in educational priorities,

where data-driven methodologies and intelligent systems are increasingly viewed as essential tools for personalizing learning, automating assessment, and supporting early interventions. As traditional teaching methods adapt to 21st-century demands, DL offers promising solutions for addressing complex educational challenges, ranging from

cognitive development to emotional monitoring and environmental responsiveness. To better understand these emerging directions, Figure 2 presents the global distribution of DL-related studies from 2021 to 2025, providing a temporal lens into the evolution of this research landscape.

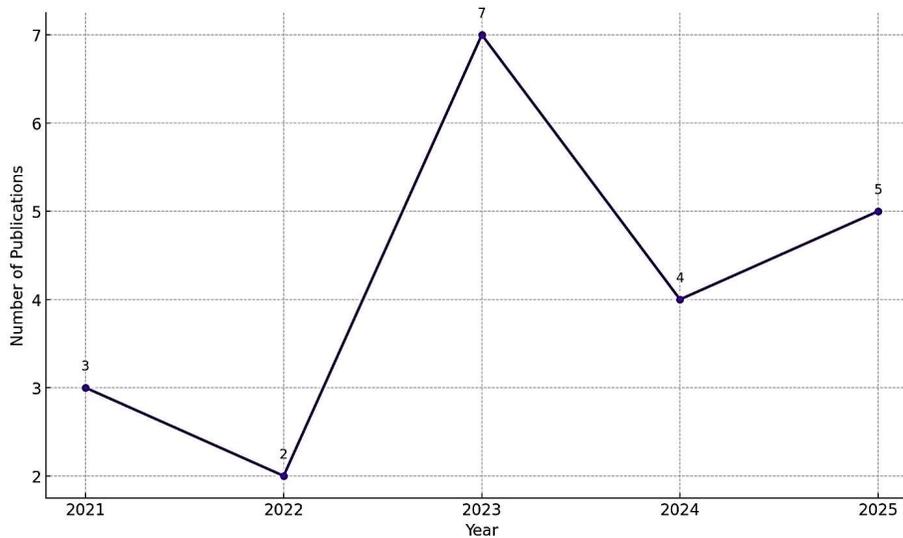


Figure 2. Global distribution of DL research

The trend of deep learning (DL) publications in primary education from 2021 to 2025, as shown in Figure 2, demonstrates a steady increase in research interest over time. Beginning with three publications in 2021, the number rose slightly to 2 in 2022, followed by a significant peak in 2023 with seven studies. Although there was a modest decline to 4 publications in 2024, the number rebounded to 5 in 2025, indicating sustained academic engagement. This pattern suggests a growing recognition of DL's potential in enhancing learning environments, particularly in personalized education, automated assessment, and the early detection of learning needs. The peak in 2023 may also reflect momentum gained from post-pandemic educational reforms and the acceleration of AI-driven innovations in schooling systems.

While the temporal analysis presented in Figure 3 reveals a consistent increase in research

outputs over time, a more in-depth examination of thematic and methodological patterns offers profound insights into how deep learning (DL) is operationalized in primary education, aligning with Research Question 1 (RQ1). Table 2 consolidates findings from diverse studies conducted across 11 countries, illustrating the broad scope of DL applications, ranging from environmental monitoring and curriculum enhancement to emotional diagnostics and personalized learning feedback. These studies draw upon various datasets and implement a broad spectrum of DL architectures, including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Generative Adversarial Networks (GANs), and transformer-based models. Each model is contextually adapted to address specific educational challenges and learning objectives, reflecting the growing sophistication and relevance of DL in primary education.

Table 2. Overview of deep learning trends in primary education

No	Author and Country	Trend Category	Dataset Used	Deep Learning Model
1	(Aamer et al., 2025)/ Korea	DL for Environmental Health & Safety in Schools	Real-world PM2.5 measurements from 3 schools (ES, MS, HS) & 9 nearby air quality stations (576+ data points)	ADCAE (Attention Deep Convolutional Autoencoder)
2	(Boob & Radke, 2025)/ India	Curriculum-Focused DL in Geometry	ElementaryCQT (380,000 images, 76 classes of 2D shapes)	Sequential CNN, VGG16, ResNet50, InceptionV3, Inception-ResNetV2
3	(Lin, Zhao, Wang, & Chen, 2025)/ China	AI Curriculum Development with DL Concepts	Hands-on activities with 5th-grade students at Huangshi Chinese-English School; CNN image recognition tasks	Convolutional Neural Network (CNN)
4	(L. Tian, Ding, Tian, Chen, & Wang, 2025)/ China	DL for Intelligent Educational Assessment	4,481 valid responses from Chinese 6th-graders answering scientific argumentation tests	TextCNN & Bi-LSTM+Attention
5	(R. Zhang et al., 2024)/ China	Urban Environmental Color Analytics for Child Psychology	Baidu Street View Images (497,088 images), OSM Road Network, POI Data, Statistical Yearbook, VR Questionnaire Data	DeepLab V3+ (Semantic Segmentation) + Lasso Regression
6	(Moon, Kim, Kim, & Kim, 2024)/ Korea	AI Literacy-Oriented DL Education	Pre/post-test AI literacy assessments (20 elementary gifted students)	Deep Learning (via Machine Learning for Kids)
7	(Deng, Huang, & Ren, 2024)/ China	Psychological Variable-Driven DL Models	147,210 elementary students (LegiLexi exam 2016–2021)	OSO-DDNN (Owl Search Optimized Dynamic Deep Neural Network)
8	(Muhathir, Maqfirah, El Akmal, Ula, & Sahputra, 2024)/ Indonesia	Facial Recognition for Early Screening	200 facial images of autistic and non-autistic students (70:30 & 80:20 split)	SVM with HOG & SURF feature extraction
9	(X. Zhang, Wang, & Chen, 2025)/ China	Cognitive Analytics for Art-based Learning	1968 coded peer-assessment comments (49,869 words) from 40 first graders	<i>Epistemic Network Analysis</i> (ENA – treated analogously to DL-based cognitive mapping)
10	(Zirak, Saeedian, Zomorodian, & Tahsildoost, 2023)/ Iran	Predictive Modeling for Educational Space Design	4×864 + 4×858 simulated layouts of Tehran primary schools (12 configurations)	GAN (pix2pix) for image-to-image translation; CNN (VGG16, VGG19) for feature extraction & regression
11	(Baharuddin & Naufal, 2023)/	Automation of Exam Question	449 Indonesian elementary school multiple-choice	Fine-tuned IndoBERT

	Indonesia	Classification	questions, labeled by Bloom's Taxonomy (C1–C4)	
12	(Ong et al., 2023)/ Philippines	DLNN in Online Learning Evaluation	385 Filipino primary school students (ages 7–13), purposive sample	Deep Learning Neural Network (DLNN)
13	(Lomurno et al., 2023)/ Italy	Early Dysgraphia Risk Detection	Longitudinal handwriting-related signals from 247 children (kindergarten → 2nd grade) using the <i>Play-Draw-Write</i> iPad app	Ensemble DL model + Quasi-SVM meta-model
14	(Neha & Kumar, 2023)/ India	Graduate Performance Prediction with Correlated Features	Dataset of 2,497 students (SSC, intermediate, undergraduate semesters) collected via interviews and academic records.	GIPA-PA-ECFSM (Graduate Interlinked Precedent Academic-based Performance Analysis using Enhanced Correlated Feature Set Model)
15	(Winje & Løndal, 2023)/ Norway	Pragmatist-Based Deep Learning in Outdoor Education (<i>Uteskole</i>)	3 months of fieldwork: participatory observations & qualitative interviews of pupils (grades 2, 5–7) in two Norwegian primary schools	Pragmatist theoretical framework (Dewey's Transaction & Continuity) operationalised via Nicol's Multimodal "Model of Knowing"
16	(Y. Lee, 2023)/ Korea	AI-based Predictive Learning Models for Math Achievement	301 first-grade students' math characteristics & achievement data (number size/order/counting, computational fluency, cognitive processing)	RNN with Seq2Seq architecture
17	(Kim et al., 2022)/ Korea	Household-Level Predictive Modeling using Smart Data	Smart water meter readings (2017–2019) + weather data + day-type data (weekend/weekday) for four household types	LSTM (Long Short-Term Memory) (compared to ARIMA)
18	(X. Tian, Zhao, & Nguyen, 2022)/ China	Integration of Deep Learning Concepts into Math Pedagogy	Classroom-based observations & theoretical literature on deep learning in mathematics teaching	Deep Neural Network (conceptual analogy, not computational)
19	(Hava, 2021)/ Turkey	Flipped Classroom as a Vehicle for Deep Learning	97 first-year undergraduates in Turkey (7-week quasi-experimental study)	<i>Not computational DL, but pedagogical deep learning strategies</i>
20	(Yang & Hong, 2021)/ Korea	Neuroimaging + DL fusion for cognitive detection	Resting-state fNIRS signals from 24 participants (15 MCI, nine healthy controls)	Pre-trained CNNs (VGG16, VGG19, ResNet, AlexNet) with Transfer Learning (FRTL & CTL)
21	(Piol, Lacatan, & Pulumbarit, 2021)/ Philippines	Predictive Analytics for School Enrolment	DepEd Batangas Division enrollment data (K–6) for SY 2016–2019	Deep Learning, Decision Tree, Random Forest, Gradient Boosted Tree, SVM, and Linear Regression (compared)

The synthesis of 21 studies in Table 3 highlights a multidimensional evolution in the application of deep learning (DL) in primary education. DL has been harnessed across diverse geographical contexts, spanning Asia, Europe, and the Middle East, for a broad spectrum of educational objectives, illustrating its versatility and increasing relevance in addressing pedagogical and systemic challenges.

A prominent cluster of studies focuses on curriculum-specific innovations, such as the use of CNN, RNN, and VGG architectures to enhance geometry instruction, AI literacy, and concept-based mathematics learning (Boob & Radke, 2025; Moon et al., 2024; X. Tian et al., 2022). Another cluster emphasizes predictive modeling and early detection, with applications in academic performance prediction, dysgraphia risk screening, autism identification, and broader cognitive outcomes (Deng et al., 2024; Y. Lee, 2023; Lomurno et al., 2023; Muhathir et al., 2024; Neha & Kumar, 2023; Yang & Hong, 2021). These strands highlight how DL supports both targeted interventions and scalable assessment frameworks.

In parallel, several studies demonstrate DL's role in automating educational processes, such as test classification and evaluation of scientific argumentation using IndoBERT, TextCNN, and Bi-LSTM with attention mechanisms (Baharuddin & Naufal, 2023; L. Tian et al., 2025). Complementing this, research has extended DL applications into non-traditional learning environments, including outdoor learning, environmental design of school facilities, and artistic or creative development, thereby underscoring DL's holistic potential in primary education (Winje & Løndal, 2023; X. Zhang et al., 2025; Zirak et al., 2023).

Finally, a smaller yet critical group of studies explores novel and hybrid DL architectures, such as attention-based autoencoders for air quality monitoring, GANs for architectural optimization,

and Quasi-SVM ensembles for longitudinal writing analysis (Aamer et al., 2025; Lomurno et al., 2023; Zirak et al., 2023). These are complemented by research integrating socio-economic and environmental data for student modeling (Kim et al., 2022), pointing toward a shift to interdisciplinary and context-aware DL systems that balance technical robustness with the nuanced needs of children and policy frameworks (Aamer et al., 2025; Lin et al., 2025; Ong et al., 2023; Piol et al., 2021; R. Zhang et al., 2024).

This clustering not only synthesizes individual findings but also builds broader arguments that DL research in primary education is evolving along three major directions: (1) curriculum innovation, (2) predictive and diagnostic modeling, and (3) expansion into holistic and interdisciplinary educational domains.

These empirical directions are geographically distributed in a way that reinforces the global nature of DL adoption in primary education. As shown in Figure 3, China (7 studies) and South Korea (5 studies) dominate the research landscape, underlining their strategic investment in AI-integrated education (Irwanto, 2025). The dominance of China and South Korea in the selected studies reflects these countries' national AI education strategies, such as South Korea's "AI for All" curriculum and China's large-scale investments in AI policy for public education. These initiatives create fertile ground for DL experimentation and adoption. However, this also means that the findings may not be directly applicable to contexts like Indonesia, where internet access, infrastructure, and teacher readiness vary significantly. Emerging contributions from India, Indonesia, and the Philippines reveal a rising interest in harnessing DL for socially diverse and context-sensitive learning environments (Baharuddin & Naufal, 2023; Boob & Radke, 2025; Muhathir et al., 2024; Neha & Kumar, 2023; Piol et al., 2021).

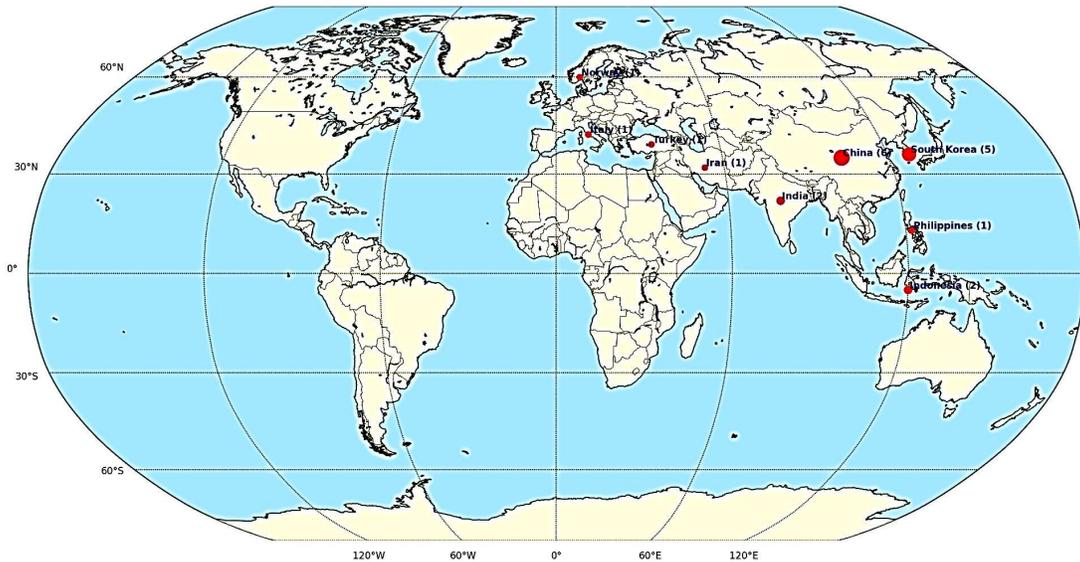


Figure 3. Global distribution of DL research in primary education

Meanwhile, European and Middle Eastern countries, including Norway, Italy, Iran, and Turkey, contribute more localized innovations, such as outdoor learning and spatial analytics.

These regional patterns illustrate that DL's role in primary education is not confined to advanced economies, but is increasingly seen as a transformative tool across educational systems worldwide

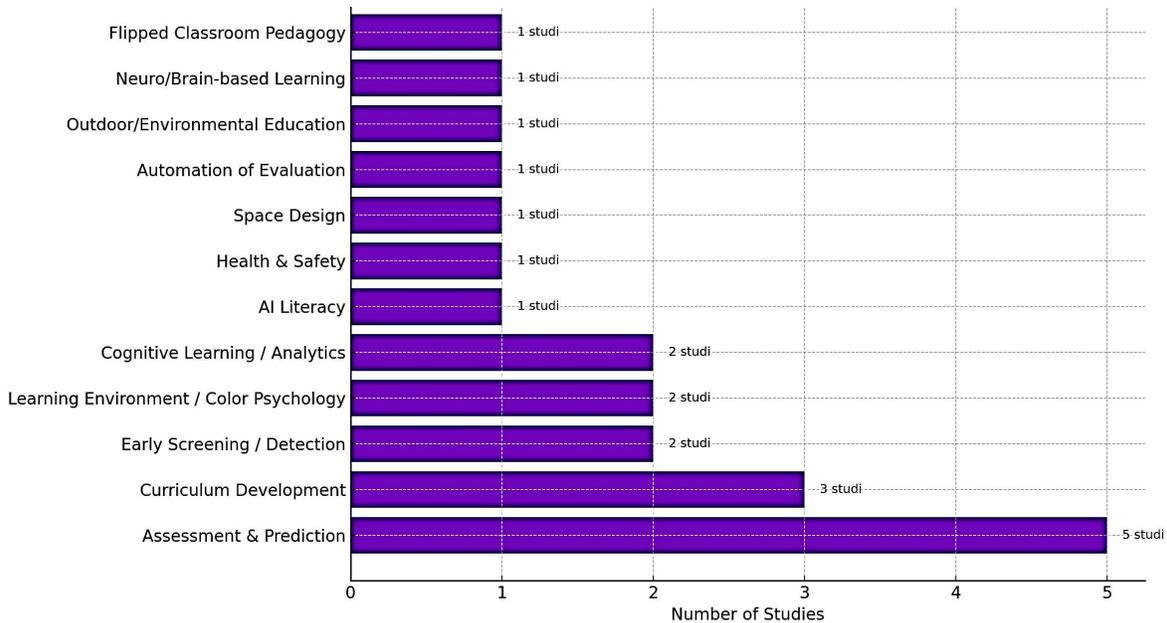


Figure 4. Dominant trends in DL applications for primary education

Figure 4 complements this perspective by illustrating the thematic distribution of DL applications. Assessment & prediction (5 studies)

is the most dominant category, which aligns with global priorities on measurable learning outcomes and early academic interventions. *Curriculum*

Development (3 studies) highlights DL’s contribution to instructional design and AI-infused pedagogy. Mid-tier trends, such as Early Screening/Detection, Learning Environment/Color Psychology, and Cognitive Learning/Analytics (with two studies), indicate a growing interest in personalized, emotionally attuned education. Less frequent yet innovative applications such as *AI Literacy*, *Health & Safety*,

Space Design, *Automation of Evaluation*, and *Outdoor/Environmental Education* demonstrate exploratory efforts in niche domains. These themes suggest that DL is being leveraged to enhance academic instruction and enrich the overall learning experience.

To further understand how DL models are mapped to educational objectives, Figure 5 presents a cross-comparison of model usage by

Trend	Assessment & Prediction	1	1	0	1	0	1	1	1	0	1
	Curriculum Development	1	0	0	0	0	0	0	0	0	1
	Early Screening / Detection	0	0	0	1	0	0	0	0	1	0
	Learning Environment / Color Psychology	1	0	0	0	0	0	0	0	0	1
	AI Literacy	0	0	0	0	0	0	0	0	0	1
	Health & Safety	0	0	0	0	0	0	0	0	0	1
	Cognitive Learning / Analytics	0	0	0	0	0	0	0	1	0	1
	Space Design	1	0	1	0	0	0	0	0	0	1
	Automation of Evaluation	0	0	0	0	1	0	0	0	0	0
	Outdoor/Environmental Education	0	0	0	0	0	0	0	0	0	1
		CNN	RNN	GAN	SVM	BERT	AE	LSTM	DNN	-semble	Other

Figure 5. Cross-Mapping DL models with educational trends

trend. CNNs emerge as the most adaptable, applied across domains such as *Assessment and Prediction*, *Curriculum Development*, and *Space Design*, reflecting their strength in handling visual and spatial data. RNN and LSTM models were notably used for sequential learning tasks tied to performance prediction, while SVMs and GANs were applied in more specialized contexts such as screening and generative design. IndoBERT appeared uniquely in *Automation of Evaluation*, showcasing the rise of transformer-based NLP models tailored to language-specific educational tasks. Interestingly, Cognitive Learning/Analytics and Health & Safety categories often rely on hybrid or custom architectures labeled “Other,” emphasizing the contextual flexibility required for nuanced educational goals. This diversity in model selection

underscores that DL’s impact depends on computational power and how well the architecture aligns with pedagogical needs and data types.

The “Other” category encompasses hybrid and custom architectures, such as transformer-based NLP models (e.g., IndoBERT) used for automated evaluation tasks, as well as combined CNN–RNN frameworks applied to multimodal learning contexts. This highlights the contextual flexibility necessary to address nuanced educational goals that a single architecture cannot accommodate. Moreover, the limited mapping observed, for example, RNN/LSTM primarily linked to prediction tasks and GANs to space/environmental design, reflects both the inherent strengths of these models and the current narrow scope of empirical applications. The abundance

of empty cells in the cross-mapping table suggests underexplored opportunities for future research, such as extending generative models to assessment domains or applying sequential models to adaptive learning environments.

RQ1. What ethical, technical, and pedagogical challenges arise when implementing DL at this level?

RQ2. What opportunities exist to leverage DL for personalized learning, student engagement, and holistic outcomes in primary schools?

Building upon the thematic and geographical mapping of DL research in primary education, this section addresses two pivotal aspects: the challenges (RQ2) and opportunities (RQ3) that

arise when implementing deep learning (DL) at the primary school level. As the use of DL becomes increasingly prominent in early education, scholars have begun to critically examine its ethical implications, technical barriers, and pedagogical complexities, alongside its transformative potential to personalize learning, increase student engagement, and promote holistic development.

Table 3 presents a synthesized overview of 21 selected studies, categorizing each by its identified challenge, corresponding opportunity, and the main takeaway or contribution. This synthesis offers a nuanced understanding of how DL is both reshaping and being shaped by the practical realities of primary education, and it serves as the empirical foundation for the following discussion.

Table 3. Summary of challenges and opportunities in applying deep learning (DL) for Primary education

No	Author	Challenge	Opportunity	Main Synthesis
1	(Aamer et al., 2025)	Pedagogical & Ethical Concerns	Smart School Monitoring & Well-being	DL connects smart infrastructure and primary education, supporting healthier, data-informed learning environments while saving costs and promoting equity.
2	(Boob & Radke, 2025)	Technical – Dataset Complexity	Personalized & Interactive Learning	DL can transform early geometry lessons into adaptive, visually engaging experiences, enabling personalized feedback, higher engagement, and potentially holistic outcomes for young learners.
3	(Lin, Zhao, Wang, & Chen, 2025)	Ethical & Equity Concerns	Foundational AI Literacy & Engagement	DL concepts (CNN, digitization, algorithmization) can serve as a gateway to early AI literacy, sparking curiosity and paving the way for deeper computational and critical thinking in the primary years.
4	(L. Tian, Ding, Tian, Chen, & Wang, 2025)	Pedagogical Challenges	Automated Assessment & Teacher Support	DL models can transform assessment workflows, moving teachers from manual scoring to higher-value pedagogical roles, while providing students with quicker feedback for better learning.
5	(R. Zhang et al., 2024)	Technical	Personalized Emotional Mapping	Enables child-specific design of school neighborhoods, reducing stress & improving emotional well-being.
6	(Moon, Kim, Kim, & Kim, 2024)	Pedagogical Challenge	Engagement through Block-Based DL	DL models can be simplified into playful learning environments, making abstract concepts accessible.

7	(Deng, Huang, & Ren, 2024)	Technical Challenge	Psychological Integration for Engagement	DL can provide academic and socio-emotional support.
8	(Muhathir, Maqfirah, El Akmal, Ula, & Sahputra, 2024)	Ethical Challenges	Integration with Future DL Models	Opens path for next-gen DL to enhance sensitivity & reduce manual tuning.
9	(X. Zhang, Wang, & Chen, 2025)	Ethical	Personalized Learning Pathways	DL models could create individualized learning journeys, adjusting complexity based on cognitive stage.
10	(Zirak, Saeedian, Zomorodian, & Tahsildoost, 2023)	Pedagogical	Personalized Space Design	DL can personalize the physical environment, improving students' focus, mood, and health.
11	(Baharuddin & Naufal, 2023)	Ethical	Teacher Workflow Automation	DL can act as a "cognitive assistant" to reduce admin load and empower teachers to focus on higher-value teaching tasks.
12	(Ong et al., 2023)	Technical	Personalization	DL enables adaptive learning paths aligned to cognitive & emotional readiness.
13	(Lomurno et al., 2023)	Technical	Personalized Early Intervention	Deep learning can shift interventions years earlier, personalizing support and preventing academic/emotional struggles.
14	(Neha & Kumar, 2023)	Technical	Early Performance Prediction	DL enables predictive support systems, allowing schools to intervene early for students at risk.
15	(Winje & Løndal, 2023)	Pedagogical Scope	Holistic Deep Learning Framework	Outdoor education (<i>uteskole</i>) can turn abstract concepts into lived experiences, embedding DL beyond cognition.
16	(Y. Lee, 2023)	Pedagogical Alignment	Customized Learning Pathways	DL enables tailored tutoring, adjusting pace & content for each child.
17	(Kim et al., 2022)	Technical – Overfitting & Model Generalization	Predictive Behavioral Modeling	DL can create early-warning systems for at-risk learners by spotting deviations in learning "consumption" patterns.
18	(X. Tian, Zhao, & Nguyen, 2022)	Ethical Dimension – Equity of Access	Cross-Disciplinary Integration	Positions DL as a bridge for holistic education, combining academic, cognitive, and socio-emotional growth.
19	(Hava, 2021)	Technical Access & Equity	Enhanced Cognitive & Emotional Engagement	DL-inspired flipped models can stimulate intrinsic motivation and critical thinking when thoughtfully scaffolded.
20	(Yang & Hong, 2021)	Technical Barriers &	Transfer Learning for	This approach could help schools or researchers utilize DL without relying

		Imbalance	Small Datasets	on massive labeled datasets, thereby democratizing advanced AI tools.
21	(Piol, Lacatan, & Pulumbarit, 2021)	Algorithm Comparability & Interpretability	Forecasting Resource Needs	It shows that DL/ML can optimize educational planning at a system level.

Table 3 presents a synthesized overview of 21 selected studies, categorizing each by its identified challenge, corresponding opportunity, and the main takeaway or contribution. This synthesis offers a nuanced understanding of how DL is both reshaping and being shaped by the practical realities of primary education, serving as the empirical foundation for the following discussion.

In synthesizing the findings, a clear relationship emerges between the challenges (RQ2) and opportunities (RQ3) of integrating Deep Learning (DL) in primary education. Technical barriers such as dataset complexity, overfitting, and infrastructure limitations highlight the difficulty of implementing DL at scale. However, these same obstacles have spurred opportunities for innovation, such as using transfer learning, resource-efficient algorithms, and

predictive assessment tools that enable early intervention.

Similarly, pedagogical challenges, ranging from teacher readiness to curriculum alignment, have created opportunities for more personalized pathways, adaptive feedback systems, and block-based engagement strategies that strengthen teacher support rather than replace it. Ethical concerns, including equity, privacy, and transparency, remain significant; however, they also drive the development of inclusive design principles and cross-disciplinary integration, ensuring that DL applications remain sensitive to children’s developmental needs. The interplay of RQ2 and RQ3 suggests that the barriers identified by researchers often serve as catalysts for innovation, turning constraints into opportunities that foster more equitable, engaging, and future-ready learning environments.

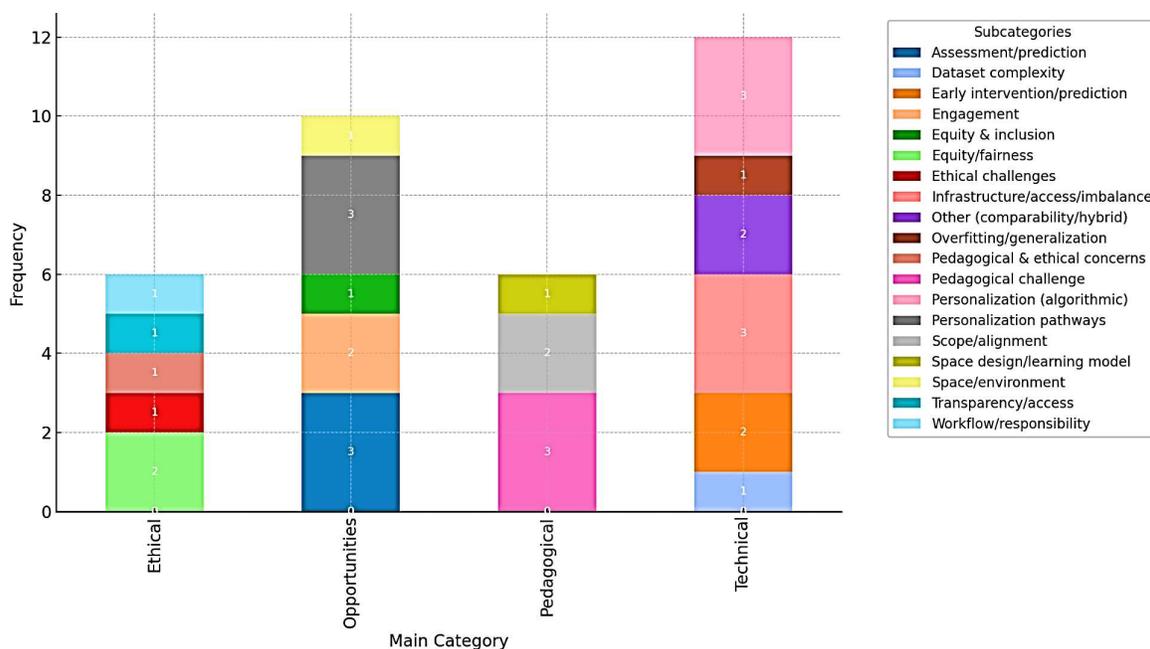


Figure 6. Frequency of challenges and opportunities in deep learning applications for primary education

The reviewed studies highlight various ethical, technical, and pedagogical challenges associated with the integration of DL in primary education, as outlined in Table 3 and Figure 6. Ethical and pedagogical concerns emerge prominently in several studies (Aamer et al., 2025; Lin et al., 2025; Muhathir et al., 2024), revealing dilemmas related to data privacy, algorithmic bias, and equitable access. These concerns are particularly heightened in applications that involve monitoring, automation, or personalization, where children's data and developmental sensitivity demand thoughtful governance. However, these same studies suggest that DL, when ethically managed, can foster smarter, more responsive educational environments that support well-being, equity, and early-stage AI literacy, marking a significant step toward inclusive digital transformation in schools.

From a technical perspective, barriers such as dataset complexity, overfitting, and limited generalizability continue to challenge researchers and practitioners. Studies by Boob & Radke (2025), Kim et al. (2022), and Yang & Hong (2021) underline how the requirement for large, labeled datasets and robust models can restrict the implementation of DL in resource-limited educational settings. Studies have started leveraging transfer learning to mitigate overfitting and limited dataset availability (Yang & Hong, 2021), using pre-trained models adapted for smaller educational datasets. Additionally, model simplification through ensemble techniques and hybrid systems (Lomurno et al., 2023) offers a practical path forward in resource-constrained environments. Nonetheless, innovative responses to these technical limitations are also evident. For example, transfer learning, model simplification, and interpretability frameworks are helping to democratize DL, allowing even smaller schools to benefit from adaptive and scalable AI tools (Piol et al., 2021; Yang & Hong, 2021).

Pedagogical challenges center around the alignment of DL technologies with young learners'

developmental needs and cognitive styles. Studies by Y. Lee (2023) and Moon et al. (2024) emphasize the difficulty of integrating DL into early-stage curricula without oversimplifying or overwhelming students. However, several studies demonstrate that DL can enhance engagement, foster curiosity, and support differentiated instruction when properly adapted, such as through block-based learning environments or customized tutoring systems. DL's capacity to automate assessments (L. Tian et al., 2025) and personalize content delivery (Ong et al., 2023) empowers educators to shift from administrative burdens to more meaningful pedagogical roles.

The opportunities DL presents in personalizing learning are possibly the most compelling finding across these studies. Several researchers (Lomurno et al., 2023; Neha & Kumar, 2023; X. Zhang et al., 2025) illustrate how DL enables early detection of learning difficulties and tailoring educational pathways to individual cognitive and emotional profiles. This personalization extends beyond content, affecting spatial and affective dimensions of learning environments. Works like those of R. Zhang et al. (2024) and Zirak et al. (2023) highlight how DL can support emotional mapping and personalized classroom design, contributing to improved mental health and focus among students.

Finally, a growing vision of DL as a holistic framework transcends traditional disciplinary boundaries. Studies by X. Tian et al. (2022) and Winje & Løndal (2023) position DL as a tool that supports academic performance and fosters socio-emotional development, environmental awareness, and real-world readiness. The convergence of AI with outdoor education, flipped classrooms (Hava, 2021), and system-level forecasting (Piol et al., 2021) reveals a future where DL is not just an instructional supplement, but a core infrastructure for cultivating resilient, adaptive, and emotionally intelligent learners in the primary years. Together, these insights

underscore the dual imperative of addressing foundational challenges while embracing the vast potential DL holds for shaping the future of education.

Technology–Pedagogy Tension in Deep Learning

The synthesis of 21 studies reveals a persistent tension in how Deep Learning (DL) is conceptualized and implemented in primary education: as a *technological construct* versus a *pedagogical principle*. On the technological side, DL architectures such as CNNs, RNNs, and GANs have demonstrated strong capabilities in pattern recognition, predictive modeling, and assessment automation. Studies such as Boob & Radke (2025), Lin et al. (2025), and Moon et al. (2024) highlight the power of DL to personalize curricula and deliver adaptive pathways. However, these efforts often privilege algorithmic efficiency over child-centered pedagogical values. On the pedagogical side, studies such as Hava (2021) and Winje & Løndal (2023) foreground experiential learning and holistic engagement but lack the technological robustness needed for scalable application. This divergence underscores a fundamental challenge: innovation oscillates between technocentric solutions and pedagogically rich but technically underpowered approaches.

This tension is especially critical in primary education, where developmental appropriateness must remain a central focus. A technocentric approach risks exposing young learners to data overload or inappropriate personalization models that ignore socio-emotional needs. Conversely, a purely pedagogical perspective without technological feasibility may fail to meet the demands of 21st-century learning systems (Asad et al., 2021; Moninoor & Haider, 2024). The studies reviewed demonstrate that the most promising outcomes occur when these two orientations are integrated rather than polarized. For example, automated assessment systems not only improve grading efficiency but also provide

teachers with timely insights for pedagogical decision-making, while block-based DL applications translate complex algorithmic concepts into playful, developmentally appropriate tools for student engagement (Baharuddin & Naufal, 2023; Moon et al., 2024).

The evidence suggests that DL should be reframed as an *enabling infrastructure* that empowers teachers, rather than as an automated replacement. The role of DL in primary education should be to amplify teacher agency, create adaptive yet pedagogically grounded pathways, and ensure that personalization aligns with children's cognitive and emotional growth. Such a reframing transforms the technology–pedagogy tension into a productive synergy, where algorithms support pedagogical goals instead of being substitutes for them.

Controversies and Unresolved Issues

Beyond the broad promise of DL integration, the reviewed studies also expose unresolved controversies and dilemmas that demand closer scrutiny. Ethical concerns are among the most pressing. Research employing biometric or behavioral data (e.g., facial recognition for autism detection in Muhathir et al. (2024); school air-quality monitoring in Aamer et al. (2025) raises serious questions about privacy, consent, and algorithmic fairness. Critics argue that these systems could reproduce or even exacerbate educational inequalities without strict governance. The data that enable personalization may simultaneously generate risks of surveillance and discrimination.

Pedagogical debates are equally pronounced. Some scholars emphasize that DL can reduce teachers' administrative burden and free them to focus on higher-order facilitation (Assaly & Jabarin, 2024; Culver, Kezar, & Koren, 2023). Others, however, warn that over-reliance on automation could diminish the human dimensions of teaching, such as empathy, creativity, and socio-emotional guidance that no

algorithm can replace (Henriksen, Creely, Gruber, & Leahy, 2025). These divergent perspectives illustrate the contested role of teachers in DL-driven classrooms: are they redefined as facilitators empowered by intelligent systems, or marginalized by technologies that assume their evaluative functions.

Technical controversies also persist. Most DL applications require large, well-labeled datasets, creating barriers for under-resourced educational systems. Studies by Kim et al. (2022), Lomurno et al. (2023), and Yang & Hong (2021) revealed that overfitting and limited generalizability remain significant challenges, especially in small-sample or heterogeneous school contexts. Solutions such as transfer learning and hybrid ensemble models offer potential pathways, but adoption is still in its infancy. The paradox is apparent: those educational contexts that most need adaptive, scalable innovations are often the least equipped to implement them. These controversies highlight that DL's transformative potential is inseparable from its risks, requiring a careful balance between innovation and accountability.

Implications for Developing Countries: The Case of Indonesia

The reviewed studies also illuminate how contextual differences shape the feasibility of DL adoption. While China and South Korea dominate the landscape with large-scale AI curricula and state-backed policy infrastructures, replicating these strategies directly in developing contexts like Indonesia is neither practical nor desirable. Significant disparities in internet connectivity, teacher readiness, and school-level infrastructure create barriers to the adoption of resource-intensive models. For example, GANs for architectural optimization or multimodal analytics for urban learning environments may be aspirational but remain unrealistic in the short term (R. Zhang et al., 2024; Zirak et al., 2023).

Nevertheless, several DL applications identified in this review are well-suited for

Indonesia's current conditions. Automated assessment systems such as IndoBERT-based test classification (Baharuddin & Naufal, 2023) or TextCNN for evaluating scientific argumentation (L. Tian et al., 2025) align directly with the country's large-scale examination needs and competency-based curriculum reforms. Predictive modeling tools for early detection of learning difficulties (Y. Lee, 2023; Lomurno et al., 2023) also hold strong potential for integration into existing school information systems, enabling earlier interventions without requiring advanced infrastructure. Furthermore, playful and low-resource engagement models such as block-based DL (Moon et al., 2024) could be adapted to support the *Merdeka* Curriculum's emphasis on differentiated and student-centered learning.

These findings suggest a staged adoption strategy: Indonesia should prioritize *low-resource, high-impact* applications, automated assessments, predictive analytics, and engagement tools that deliver immediate benefits while laying the foundation for more complex DL systems in the future. At the policy level, this implies the need for a national roadmap that emphasizes capacity-building, teacher training, and robust ethical regulation. In this way, DL adoption in Indonesia can remain contextually relevant, pedagogically grounded, and technologically feasible, ensuring that innovations strengthen rather than disrupt the foundational goals of primary education.

LIMITATIONS OF THE REVIEW

Despite rigorous adherence to PRISMA guidelines and the application of a systematic literature review (SLR) methodology, this study has limitations. Firstly, the review was confined to publications indexed in Scopus between January 2021 and August 2025. While Scopus provides extensive coverage across education, computer science, and interdisciplinary fields, it inevitably excludes studies indexed in other reputable databases such as Web of Science, ERIC, or IEEE Xplore. This may have omitted relevant contributions, particularly highly technical

works in engineering and computer science or pedagogical studies from specialized education databases. The exclusive focus on English-language peer-reviewed journal articles and conference proceedings further introduces potential language and publication biases, limiting representation from non-English-speaking contexts and excluding valuable insights from grey literature such as reports, dissertations, or policy briefs.

Secondly, accessibility barriers presented a significant limitation. Of the 219 full-text studies retrieved for eligibility screening, 144 articles (66%) could not be accessed due to paywalls, restricted institutional access, or outdated repositories. This substantially reduced the final sample size, raising the possibility that relevant findings were overlooked. To mitigate this, a metadata analysis of inaccessible articles (titles, abstracts, authors, and affiliations) was conducted, which indicated that most originated from East Asia and North America regions already strongly represented in the 21 included studies. Their abstracts suggested a broad focus on AI in education, rather than primary (K–6) contexts, implying that their exclusion may not have drastically altered the thematic scope of this review. Nonetheless, this introduces the risk of geographic and contextual bias, underscoring that the findings must be interpreted cautiously regarding global generalizability.

Thirdly, the geographical concentration of studies poses an additional limitation. Most of the included works were in China and South Korea, countries with advanced national AI policies and robust digital infrastructures. While this dominance reflects their leadership in AI-driven education, it limits the transferability of findings to developing contexts such as Indonesia, where infrastructure, teacher readiness, and resource availability vary widely. Although contributions from India, Indonesia, the Philippines, and selected European and Middle Eastern countries provide diversity, the overall synthesis may still underrepresent challenges and opportunities

unique to low-resource or culturally diverse settings.

Finally, the scope of the reviewed studies was inherently constrained by the methodological focus of the included articles. Many studies framed DL primarily within algorithmic development or educational data mining, with less emphasis on psychological, ethical, or socio-emotional dimensions. This imbalance may skew interpretations toward technological feasibility while underrepresenting child-centered pedagogy and equity concerns. Moreover, the systematic review design remains dependent on the quality and reporting rigor of the original studies, which varied in methodological robustness, sample size, and contextual detail.

These limitations highlight the need for caution in generalizing the results globally. They also point to future research directions, including cross-database reviews, multilingual inclusion strategies, open-access collaborations, and region-specific studies that better capture the diversity of global primary education systems.

■ IMPLICATION AND FUTURE INSIGHT

The results of this systematic review offer significant implications for researchers, educators, policymakers, and technology developers invested in the future of primary education. The demonstrated versatility and transformative potential of DL underscore the urgency of integrating ethically sound, pedagogically informed, and technically robust frameworks within educational practices. Policymakers and education leaders should prioritize creating guidelines to mitigate ethical risks, such as data privacy and algorithmic fairness, ensuring equitable access, and fostering stakeholder trust.

From a pedagogical perspective, it is critical to encourage greater teacher involvement in the design and implementation stages of DL applications, fostering professional development to enhance digital literacy and readiness. Educators can thus be empowered as active

contributors rather than passive recipients of technology. Similarly, DL applications must be contextualized to align closely with students' developmental stages, cognitive abilities, and socio-emotional needs, thus ensuring genuinely personalized and holistic educational outcomes.

For future research, there is a compelling need to expand geographical and methodological diversity. More empirical studies from underrepresented regions, including Europe, Africa, and the Americas, would significantly enrich the global understanding of DL's potential and limitations in primary education. Additionally, longitudinal and action research methodologies could offer deeper insights into DL interventions' long-term effectiveness and adaptability.

Technologically, future innovations should focus on enhancing the scalability, interpretability, and generalizability of DL models. This includes leveraging techniques such as transfer learning, hybrid modeling, and automated interpretability frameworks to overcome limitations related to dataset availability, computational resource constraints, and model transparency. Moreover, cross-disciplinary collaborations integrating insights from psychology, educational theory, and data science are essential to building more holistic DL-driven educational ecosystems.

Future development should also prioritize teacher training through stackable micro-credentials that equip educators with practical competencies in using AI tools effectively in the classroom. Additionally, establishing ethical frameworks specifically designed for DL in K–6 settings, drawing inspiration from child-centered standards such as the General Data Protection Regulation for Kids (GDPR-K), is crucial to ensure children's data is handled responsibly and equitably. Equally important is the co-design of hybrid learning ecosystems by educators and developers, to ensure that DL technologies are pedagogically relevant, culturally contextualized, and feasible to implement across diverse educational environments, including resource-constrained settings.

Ultimately, the strategic integration of DL within primary education has the potential to redefine early learning environments. By addressing existing limitations and capitalizing on identified opportunities, stakeholders can collaboratively advance toward educational systems characterized by personalized instruction, empowered teachers, inclusive accessibility, and enhanced learner engagement, ensuring sustainable educational development aligned with global 21st-century competencies.

■ CONCLUSION

This systematic literature review addresses the critical research gap identified in the introduction: the conceptual fragmentation between Deep Learning (DL) as a technological architecture and a pedagogical philosophy. By synthesizing 21 recent empirical studies (2021–2025), this review demonstrates how DL applications in primary education have begun to reconcile this divide, aligning algorithmic capabilities (e.g., CNNs, RNNs, GANs, transformers) with pedagogical priorities such as personalization, engagement, and holistic learning. In doing so, this SLR maps current global practices and fills the void left by earlier reviews that largely overlooked K–6 education.

Findings reveal three principal directions in DL adoption: (1) curriculum innovation through AI-infused learning and assessment, (2) predictive and diagnostic modeling for early interventions, and (3) expansion into holistic domains such as emotional well-being, spatial-environmental design, and outdoor learning. Yet, persistent challenges, ethical concerns (data privacy, equity of access), technical limitations (dataset complexity, overfitting, infrastructural disparities), and pedagogical misalignment highlight that DL integration remains uneven and context-sensitive. Importantly, this review shows that these challenges often catalyze innovation, spurring solutions such as transfer learning, low-resource engagement models, and teacher-empowering assessment tools.

One critical recommendation emerges: future adoption of DL in primary education must prioritize teacher capacity-building as the central lever for sustainable and ethical integration. Without digitally literate, pedagogically informed teachers, even the most advanced DL systems risk misalignment with developmental needs. Therefore, stakeholders, including policymakers, curriculum designers, and AI developers, should prioritize national frameworks for teacher training, supported by context-sensitive, low-resource DL applications that are both scalable and inclusive. By positioning educators as empowered agents rather than passive recipients, DL can evolve from a fragmented innovation into a coherent, equitable, and future-ready infrastructure for primary education worldwide.

■ ACKNOWLEDGMENT

The authors thank the Study Program of Elementary School Teacher Education at Universitas Ahmad Dahlan for academic support, as well as colleagues, lecturers, and reviewers for their valuable feedback. Grammarly AI was used solely for proofreading to improve clarity and readability; the content itself was entirely written and finalized by the authors, who take full responsibility for its accuracy and integrity.

■ REFERENCES

- Aamer, H., Ba-Alawi, A. H., Kang, S., Lee, T., & Jo, Y. M. (2025). Prediction of school PM2.5 by an attention-based deep learning approach informed with data from nearby air quality monitoring stations. *Chemosphere*, 375, 144241. <https://doi.org/10.1016/j.chemosphere.2025.144241>
- Almalawi, A., Soh, B., Li, A., & Samra, H. (2024). Predictive models for educational purposes: a systematic review. *Big Data and Cognitive Computing*, 8(12), 1–42. <https://doi.org/10.3390/bdcc8120187>
- Almuhanna, M. A. (2025). Teachers' perspectives of integrating AI-powered technologies in K-12 education for creating customized learning materials and resources. *Education and Information Technologies*, 30(8), 10343–10371. <https://doi.org/10.1007/s10639-024-13257-y>
- Aravantinos, S., Lavidas, K., Voulgari, I., Papadakis, S., Karalis, T., & Komis, V. (2024). Educational approaches with aé in primary school settings: a systematic review of the literature available in scopus. *Education Sciences*, 14(7), 1–25. <https://doi.org/10.3390/educsci14070744>
- Asad, M. M., Aftab, K., Sherwani, F., Churi, P., Moreno-Guerrero, A. J., & Pourshahian, B. (2021). Techno-Pedagogical skills for 21st century digital classrooms: an extensive literature review. *Education Research International*, 1–12. <https://doi.org/10.1155/2021/8160084>
- Assaly, I., & Jabarin, A. (2024). Arab Israeli EFL teachers' perceptions and practices vis-à-vis teaching higher-order thinking skills: A complicated relationship. *Language Teaching Research*, 28(4), 1635–1655. <https://doi.org/10.1177/13621688211032426>
- Baharuddin, F., & Naufal, M. F. (2023). Fine-Tuning IndoBERT for indonesian exam question classification based on bloom's taxonomy. *Journal of Information Systems Engineering and Business Intelligence*, 9(2), 253–263. <https://doi.org/10.20473/jisebi.9.2.253-263>
- Bernacki, M. L., Greene, M. J., & Lobczowski, N. G. (2021). A systematic review of research on personalized learning: personalized by whom, to what, how, and for what purpose(s)? In *Educational Psychology Review* (Vol. 33). Educational Psychology Review. <https://doi.org/10.1007/s10648-021-09615-8>
- Bhutoria, A. (2022). Personalized education and Artificial Intelligence in the United States, China, and India: A systematic review using

- a Human-In-The-Loop model. *Computers and Education: Artificial Intelligence*, 3(April), 100068. <https://doi.org/10.1016/j.caeai.2022.100068>
- Boob, A., & Radke, M. (2025). ElementaryCQT: A New dataset and its deep learning analysis for 2d geometric shape recognition. *SN Computer Science*, 6(1), 1–14. <https://doi.org/10.1007/s42979-024-03521-w>
- Braun, V., & Clarke, V. (2006). Qualitative research in psychology using thematic analysis in psychology using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. Retrieved from <http://www.tandfonline.com/action/journalInformation?journalCode=uqrp20>
- Byrne, D. (2022). A worked example of Braun and Clarke's approach to reflexive thematic analysis. *Quality and Quantity*, 56(3), 1391–1412. <https://doi.org/10.1007/s11135-021-01182-y>
- Culver, K., Kezar, A., & Koren, E. R. (2023). Improving access and inclusion for VITAL Faculty in the scholarship of teaching and learning through sustained professional development programs. *Innovative Higher Education*, 48(6), 1071–1094. <https://doi.org/10.1007/s10755-023-09672-7>
- Deng, J., Huang, X., & Ren, X. (2024). A multidimensional analysis of self-esteem and individualism: A deep learning-based model for predicting elementary school students' academic performance. *Measurement: Sensors*, 33(April), 101147. <https://doi.org/10.1016/j.measen.2024.101147>
- Hava, K. (2021). The effects of the flipped classroom on deep learning strategies and engagement at the undergraduate level. *Participatory Educational Research*, 8(1), 379–394. <https://doi.org/10.17275/per.21.22.8.1>
- Henriksen, D., Creely, E., Gruber, N., & Leahy, S. (2025). Social-Emotional learning and generative ai: a critical literature review and framework for teacher education. *Journal of Teacher Education*, 76(3), 312–328. <https://doi.org/10.1177/00224871251325058>
- Imran, M., Almusharraf, N., Ahmed, S., & Mansoor, M. I. (2024). Personalization of E-Learning: Future trends, opportunities, and challenges. *International Journal of Interactive Mobile Technologies*, 18(10), 4–18. <https://doi.org/10.3991/ijim.v18i10.47053>
- Irwanto, I. (2025). Research trends on artificial intelligence in K-12 education in Asia: a bibliometric analysis using the Scopus database (1996–2025). *Discover Artificial Intelligence*, 5(1), 1–42. <https://doi.org/10.1007/s44163-025-00389-4>
- Khandare, A., Agarwal, N., Bodhankar, A., Kulkarni, A., & Mane, I. (2023). Study of Python libraries for NLP. *International Journal of Data Analysis Techniques and Strategies*, 15(1–2), 116–128. <https://doi.org/10.1504/IJDATS.2023.132564>
- Kim, J., Lee, H., Lee, M., Han, H., Kim, D., & Kim, H. S. (2022). Development of a deep learning-based prediction model for water consumption at the household level. *Water (Switzerland)*, 14(9), 1–17. <https://doi.org/10.3390/w14091512>
- Kovač, V. B., Nome, D., Jensen, A. R., & Skreland, L. L. (2025). The why, what and how of deep learning: critical analysis and additional concerns. *Education Inquiry*, 16(2), 237–253. <https://doi.org/10.1080/20004508.2023.2194502>
- Krüger, J., Lausberger, C., von Nostitz-Wallwitz, I., Saake, G., & Leich, T. (2020). Search. review. repeat? an empirical study of threats to replicating SLR searches. *Empirical Software Engineering*, 25(1), 627–677. <https://doi.org/10.1007/s10664-019-09763-0>

- Lavanya, A., Gaurav, L., Sindhuja, S., Seam, H., Joydeep, M., Uppalapati, V., ... S.D, V. (2023). Assessing the performance of python data visualization libraries: a review. *International Journal of Computer Engineering in Research Trends*, 10(1), 28–39. <https://doi.org/10.22362/ijcert/2023/v10/i01/v10i0104>
- Lee, S. J., & Kwon, K. (2024). A systematic review of AI education in K-12 classrooms from 2018 to 2023: Topics, strategies, and learning outcomes. *Computers and Education: Artificial Intelligence*, 6, 100211. <https://doi.org/10.1016/j.caeai.2024.100211>
- Lee, Y. (2023). Development of AI predictive model for mathematics learning achievement using deep learning. *Journal of Theoretical and Applied Information Technology*, 101(23), 7760–7768.
- Lin, P., Zhao, F., Wang, X., & Chen, Y. (2025). Initiating a novel elementary school artificial intelligence-related image recognition curricula. *Multimedia Tools and Applications*, 84(20), 22425–22439. <https://doi.org/10.1007/s11042-024-19982-3>
- Lomurno, E., Dui, L. G., Gatto, M., Bollettino, M., Matteucci, M., & Ferrante, S. (2023). Deep learning and procrustes analysis for early dysgraphia risk detection with a tablet application. *Life*, 13(3), 1–20. <https://doi.org/10.3390/life13030598>
- Long, H. A., French, D. P., & Brooks, J. M. (2020). Optimising the value of the critical appraisal skills programme (CASP) tool for quality appraisal in qualitative evidence synthesis. *Research Methods in Medicine & Health Sciences*, 1(1), 31–42. <https://doi.org/10.1177/2632084320947559>
- López-Meneses, E., López-Catalán, L., Pelicano-Piris, N., & Mellado-Moreno, P. C. (2025). Artificial intelligence in educational data mining and human-in-the-loop machine learning and machine teaching: analysis of scientific knowledge. *Applied Sciences (Switzerland)*, 15(2), 1–21. <https://doi.org/10.3390/app15020772>
- López-Meneses, E., Mellado-Moreno, P. C., Gallardo Herrerías, C., & Pelicano-Piris, N. (2025). Educational data mining and predictive modeling in the age of artificial intelligence: an in-depth analysis of research dynamics. *Computers*, 14(2), 1–26. <https://doi.org/10.3390/computers14020068>
- Lu, W., Griffin, J., Sadler, T. D., Laffey, J., & Goggins, S. P. (2025). Game-Based learning prediction model construction: toward validated stealth assessment implementation. *Journal of Learning Analytics*, 12(1), 293–321. <https://doi.org/10.18608/jla.2025.8105>
- Martínez-Comesaña, M., Rigueira-Díaz, X., Larrañaga-Janeiro, A., Martínez-Torres, J., Ocarranza-Prado, I., & Kreibel, D. (2023). Impact of artificial intelligence on assessment methods in primary and secondary education: Systematic literature review. *Revista de Psicodidáctica (English Ed.)*, 28(2), 93–103. <https://doi.org/10.1016/j.psicoe.2023.06.002>
- Miñan, G. S., Moreno, J. A., & Fernández, X. D. (2023). LIA method for the application of microsoft excel in data tabulation in systematic reviews. *CEUR Workshop Proceedings*, 3691, 44–55.
- Moninoor, M., & Haider, Z. (2024). Teaching 21st-century skills in rural secondary schools: From theory to practice. *Heliyon*, 10(9), e30769. <https://doi.org/10.1016/j.heliyon.2024.e30769>
- Moon, W., Kim, B., Kim, B., & Kim, J. (2024). Development of artificial intelligence education programs centered on deep learning principles. *Nanotechnology Perceptions*, 20(S2), 62–79. <https://doi.org/10.62441/nano-ntp.v20iS2.6>
- Muhathir, M., Maqfirah, D. R., El Akmal, M.,

- Ula, M., & Sahputra, I. (2024). Facial-Based autism classification using support vector machine method. *International Journal of Computing and Digital Systems, 16*(1), 875–886. <https://doi.org/10.12785/ijcds/160163>
- Mwogosi, A., & Mambile, C. (2025). AI integration in EHR systems in developing countries: a systematic literature review using the TCCM framework. *Information Discovery and Delivery, (November 2024)*. <https://doi.org/10.1108/IDD-07-2024-0097>
- Naghieb, A., Gharehchopogh, F. S., & Zamanifar, A. (2025). A comprehensive and systematic literature review on intrusion detection systems in the internet of medical things: current status, challenges, and opportunities. *Artificial Intelligence Review, 58*(4). <https://doi.org/10.1007/s10462-024-11101-w>
- Naseer, F., Khan, M. N., Tahir, M., Addas, A., & Aejaaz, S. M. H. (2024). Integrating deep learning techniques for personalized learning pathways in higher education. *Heliyon, 10*(11), e32628. <https://doi.org/10.1016/j.heliyon.2024.e32628>
- Neha, K., & Kumar, R. (2023). Enhancing graduate academic performance prediction and classification: an analysis using the enhanced correlated feature set model. *Ingénierie des Systèmes d'Information, 28*(6), 1605–1612. <https://doi.org/10.18280/isi.280617>
- Ong, A. K. S., Cuares, J. C., Custodio, J. P. F., Gumasing, E. Y. J., Pascual, P. N. A., & Gumasing, M. J. J. (2023). Investigating preceding determinants affecting primary school students' online learning experience utilizing deep learning neural network. *Sustainability (Switzerland), 15*(4), 1–24. <https://doi.org/10.3390/su15043517>
- Ouyang, F., Zheng, L., & Jiao, P. (2022). Artificial intelligence in online higher education: A systematic review of empirical research from 2011 to 2020. *Education and Information Technologies, 27*(6), 7893–7925. <https://doi.org/10.1007/s10639-022-10925-9>
- Ozyurt, O., Ozyurt, H., & Mishra, D. (2023). Uncovering the educational data mining landscape and future perspective: a comprehensive analysis. *IEEE Access, 11*(November), 120192–120208. <https://doi.org/10.1109/ACCESS.2023.3327624>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *The BMJ, 372*, 1–9. <https://doi.org/10.1136/bmj.n71>
- Pan, Q., Zhou, J., Yang, D., Shi, D., Wang, D., Chen, X., & Liu, J. (2023). Mapping knowledge domain analysis in deep learning research of global education. *Sustainability (Switzerland), 15*(4), 1–22. <https://doi.org/10.3390/su15043097>
- Paul, J., Lim, W. M., O'Cass, A., Hao, A. W., & Bresciani, S. (2021). Scientific procedures and rationales for systematic literature reviews (SPAR-4-SLR). *International Journal of Consumer Studies, (45)*, 1–16. <https://doi.org/10.1111/ijcs.12695>
- Pirol, E. L., Lacatan, L. L., & Pulumbarit, J. P. (2021). Predictive analysis of the enrolment of elementary schools using regression algorithms. *International Journal of Emerging Technology and Advanced Engineering, 11*(11), 184–188. https://doi.org/10.46338/IJETAE1121_21
- Razzaq, K., & Shah, M. (2025). Machine learning and deep learning paradigms: from techniques to practical applications and research frontiers. *Computers, 14*(3), 1–27. <https://doi.org/10.3390/computers14030093>
- Rui, L., Nasri, N. B. M., & Mahmud, N. D. B.

- (2024). The role of self-directed learning in promoting deep learning processes: a systematic literature review. *International Journal of Academic Research in Progressive Education and Development*, 13(4), 1–24. <https://doi.org/10.6007/ijarped/v13-i4/24374>
- Sarker, I. H. (2021). Deep Learning: A comprehensive overview on techniques, taxonomy, applications, and research directions. *SN Computer Science*, 2(6), 1–20. <https://doi.org/10.1007/s42979-021-00815-1>
- Taye, M. M. (2023). Understanding of machine learning with deep learning: architectures, workflow, applications, and future directions. *Computers*, 12(91), 1–26. <https://doi.org/10.3390/computers12050091>
- Tian, L., Ding, Y., Tian, X., Chen, Y., & Wang, J. (2025). Design and implementation of an intelligent assessment technology for elementary school students' scientific argumentation ability. *Assessment in Education: Principles, Policy and Practice*, 32(2), 231–251. <https://doi.org/10.1080/0969594X.2025.2467688>
- Tian, X., Zhao, J., & Nguyen, K. T. (2022). Practical research on primary mathematics teaching based on deep learning. *Scientific Programming*, 1–7. <https://doi.org/10.1155/2022/7899180>
- Tzimas, D., & Demetriadis, S. (2021). Ethical issues in learning analytics: a review of the field. *Educational Technology Research and Development*, 69(2), 1101–1133. <https://doi.org/10.1007/s11423-021-09977-4>
- Van Schoors, R., Elen, J., Raes, A., & Depaepe, F. (2021). An overview of 25 years of research on digital personalised learning in primary and secondary education: A systematic review of conceptual and methodological trends. *British Journal of Educational Technology*, 52(5), 1798–1822. <https://doi.org/10.1111/bjet.13148>
- Winje, Ø., & Løndal, K. (2023). ‘Wow! is that a birch leaf? In the picture, it looked totally different: ‘a pragmatist perspective on deep learning in Norwegian uteskole’. *Education 3-13*, 51(1), 142–155. <https://doi.org/10.1080/03004279.2021.19559>
- Yang, D., & Hong, K. S. (2021). Quantitative assessment of resting-state for mild cognitive impairment detection: a functional near-infrared spectroscopy and deep learning approach. *Journal of Alzheimer's Disease*, 80(2), 647–663. <https://doi.org/10.3233/JAD-201163>
- Yue, M., Jong, M. S. Y., & Dai, Y. (2022). Pedagogical design of k-12 artificial intelligence education: a systematic review. *Sustainability (Switzerland)*, 14(23), 1–29. <https://doi.org/10.3390/su142315620>
- Zhang, R., Huang, Q., Peng, Z., Zhang, X., Shang, L., & Yang, C. (2024). Evaluating the impact of elementary school urban neighborhood color on children's mentalization of emotions through multi-source data. *Buildings*, 14(10), 1–23. <https://doi.org/10.3390/buildings14103128>
- Zhang, X., Wang, Y., & Chen, H. (2025). A study on the effect of peer assessment on children's knowledge construction processes based on Epistemic Network Analysis. *Studies in Educational Evaluation*, 84, 101441. <https://doi.org/10.1016/j.stueduc.2024.101441>
- Zirak, A. G., Saeedian, A., Zomorodian, Z. S., & Tahsildoost, M. (2023). Enhancing visual comfort prediction in space layouts using deep learning models. *Intelligent Buildings International*, 15(6), 266–284. <https://doi.org/10.1080/17508975.2024.2351800>