JPPIPA 12(3) (2026)

Jurnal Penelitian Pendidikan IPA

http://jppipa.unram.ac.id/index.php/jppipa/index

Deep Learning Applications in STEM Education: A Systematic Review of Pedagogical Innovations in Numeracy and Science Learning

Mira Sudiarti[1*], Yulyanti Harisman[1], Armiati[1]

[1] Universitas Negeri Padang, Padang, Indonesia

Received: January 21, 2026
Revised: February 28, 2026
Accepted: April 13, 2026
Published: April 14, 2026

Corresponding Author:
Mira Sudiarti
mirasudiarti8@gmail.com

DOI: 10.29303/jppipa.v12i3.14846

Abstract: This research examines the application of Deep Learning (DL) technologies in mathematics and science education, focusing on technological approaches, pedagogical integration, and educational impacts. Following PRISMA 2020 guidelines, 225 studies published between January 2020 and February 2026 were analyzed from Scopus, Web of Science, PubMed, and SINTA databases. Results indicate that Convolutional Neural Networks (34.2%) and Recurrent Neural Networks/Long Short-Term Memory models (28.9%) dominate STEM applications, primarily implemented through Intelligent Tutoring Systems and adaptive learning platforms. Pedagogically, DL tools align predominantly with adaptive learning (38.7%) and inquiry-based approaches (34.2%). Evidence suggests positive impacts on learning outcomes (82.7% of studies reported significant improvements) and Higher-Order Thinking Skills, particularly critical thinking and problem-solving. However, implementation challenges persist, including technical infrastructure limitations (41.3%), data privacy concerns (36.9%), and insufficient teacher readiness (29.8%). This review concludes that while Deep Learning offers transformative potential for personalized STEM education, successful integration requires addressing ethical considerations, developing explainable AI systems, and enhancing educator preparation. Future research should prioritize longitudinal studies and equitable access to ensure DL technologies genuinely enhance rather than hinder mathematics and science learning experiences.

Keywords: Deep Learning; Numeracy and Science Learning; STEM Education;

Introduction

The rapid advancement of artificial intelligence (AI), particularly deep learning technologies, has begun to reshape educational landscapes across various disciplines (Cortes, 2025; Deng, 2026; Shen, 2025). Deep learning, a subset of machine learning characterized by multi-layered neural networks such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs including LSTMs), and Transformer-based architectures, enables sophisticated pattern recognition, predictive modeling, and adaptive processing of complex data. In the context of Science, Technology, Engineering, and Mathematics (STEM) education, these technologies hold substantial promise for addressing longstanding challenges in mathematics and science learning, including individual differences in student readiness, difficulties in visualizing abstract concepts, limited opportunities for personalized feedback, and the cultivation of higher-order thinking skills (HOTS) such as analysis, evaluation, creation, and problem-solving (Chen et al., 2025; Nguyen, 2026; Shah, 2026).

Mathematics and science education face persistent issues in fostering deep conceptual understanding and transferable skills (Alhashem et al., 2021; Park et al., 2021; Sultan et al., 2025). Traditional instructional approaches often struggle to support diverse learners, provide real-time scaffolding, or simulate authentic scientific inquiry at scale (Anchunda et al., 2025; X. Chen et al., 2026; Cheng, 2026; Ghimire, 2026). The integration

How to Cite:
Sudiarti, M., & Harisman, Y. (2026). Deep Learning Applications in STEM Education: A Systematic Review of Pedagogical Innovations in Numeracy and Science Learning. *Jurnal Penelitian Pendidikan IPA*, *12*(3), 93-104. https://doi.org/10.29303/jppipa.v12i3.14846

of deep learning offers innovative pathways to overcome these limitations (Ametefe et al., 2025; Moayyed et al., 2025; Xu et al., 2025).

For instance, deep learning models can power intelligent tutoring systems that adapt content difficulty dynamically, automated assessment tools capable of analyzing open-ended responses or diagrams, simulation environments for exploring scientific phenomena, and predictive analytics to identify at-risk students early (Feng et al., 2025; Perez & Ong, 2025; Riofrío-Luzcando et al., 2026). Such applications align with contemporary pedagogical frameworks, including problem-based learning (PBL), inquiry-based learning, and adaptive/personalized learning, potentially enhancing both cognitive outcomes and motivational engagement in STEM subjects (Chen et al., 2026; Wang, 2025; Wester, 2025).

Despite growing interest in AI-enhanced education, the specific role of deep learning technologies in mathematics and science classrooms remains underexplored in a systematic manner (Alhassan & Altmami, 2026; Boche et al., 2025; Liang et al., 2025). While broader reviews have examined artificial intelligence in education generally or AI applications in K-12 STEM contexts, few have focused narrowly on deep learning architectures (as opposed to shallow machine learning or generative AI tools like large language models) and their pedagogical integration within mathematics and science domains (Avci et al., 2025; Ayanwale & Omeh, 2026; Weihs et al., 2025).

Existing syntheses often conflate deep learning as a computational technique with deep learning as a pedagogical goal (i.e., meaningful, higher-order learning), or they cover wider AI categories without distinguishing model-specific contributions, implementation strategies, or impacts on HOTS. Moreover, the post-2020 period has witnessed accelerated development and adoption of deep learning tools in educational settings—driven partly by increased computational accessibility, open-source frameworks, and the global pivot toward digital learning—yet a comprehensive mapping of dominant approaches, pedagogical alignments, effectiveness evidence, implementation barriers, and emerging research directions is currently lacking.

Method

Research Design

This study employs a Systematic Literature Review (SLR) design to comprehensively analyze the application of Deep Learning (DL) technologies in STEM education. The review process was conducted following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines to ensure transparency, reproducibility, and methodological rigor (Page et al., 2021).

A critical distinction was made at the outset of this review: the term "Deep Learning" refers exclusively to artificial intelligence-based deep neural networks (e.g., CNN, RNN, LSTM, Transformers) rather than pedagogical "deep learning" approaches (i.e., surface vs. deep learning strategies). This distinction was necessary to avoid conceptual ambiguity and ensure the review remained focused on technological innovations in STEM education. The review protocol was structured to address three research questions concerning technological approaches, pedagogical integration, and educational impacts of DL applications in mathematics and science learning.

Search Strategy

To identify relevant studies, a systematic search was performed across four major academic databases: Scopus, Web of Science (WoS), PubMed, and SINTA. These databases were selected to ensure comprehensive coverage of high-impact international journals (Scopus, WoS, PubMed) and regionally indexed publications (SINTA) relevant to the Indonesian and Southeast Asian educational context.

The search string was constructed using Boolean operators to combine keywords related to three main concepts: (1) Deep Learning Technology, (2) STEM Domains (Mathematics and Science), and (3) Educational Context. The search was limited to publications published between January 2020 and February 2026 to capture recent advancements in educational technology, including post-pandemic digital transformation and emerging generative AI trends. Table 1 presents the complete search string configuration adapted for Scopus, which served as the primary model for other databases with syntax adjustments.

Table 1. Search String Configuration by Database

Database	Search String
Scopus	TITLE-ABS-KEY (("Deep Learning" OR "Neural Network*" OR "CNN" OR "Convolutional Neural Network*" OR "RNN" OR "Recurrent Neural Network*" OR "LSTM" OR "Long Short-Term Memory" OR "Transformer*" OR "Artificial Intelligence" OR "Machine Learning") AND ("STEM" OR "Science" OR "Mathematics" OR "Physics" OR "Chemistry" OR "Biology" OR "Calculus" OR "Geometry" OR "Algebra") AND ("Education" OR "Learning" OR "Teaching" OR "Student*" OR "Pedagogy" OR "Classroom" OR "Curriculum" OR "Instruction")) AND (PUBYEAR > 2019 AND PUBYEAR < 2027) AND (LIMIT-TO (LANGUAGE, "English")) AND (LIMIT-TO (DOCTYPE, "ar") OR LIMIT-TO (DOCTYPE, "cp"))
Web of Science	TS=("Deep Learning" OR "Neural Network*" OR "CNN" OR "RNN" OR "LSTM" OR "Transformer*" OR "Artificial Intelligence") AND TS=("STEM" OR "Science" OR "Mathematics" OR "Physics" OR "Chemistry" OR "Biology") AND TS=("Education" OR "Learning" OR "Teaching" OR "Student*" OR "Pedagogy") AND PY=(2020-2026) AND LA=(English) AND DT=(Article OR Proceedings Paper)
PubMed	("Deep Learning"[Mesh] OR "Neural Networks"[Mesh] OR "Artificial Intelligence"[Mesh]) AND ("Science"[Mesh] OR "Mathematics"[Mesh] OR "STEM Education") AND ("Education"[Mesh] OR "Learning"[Mesh] OR "Students") AND ("2020/01/01"[Date - Publication] : "2026/02/28"[Date - Publication]) AND (English[lang])
SINTA	("Deep Learning" OR "Jaringan Saraf Tiruan" OR "Kecerdasan Buatan") AND ("Pendidikan" OR "Pembelajaran") AND ("Sains" OR "Matematika" OR "STEM") AND ("2020" TO "2026")

Selection Criteria

To ensure the selected studies were relevant to the research questions, specific inclusion and exclusion criteria were established using the PICo framework (Population, Interest, Context). Studies were included if they empirically applied DL technologies in mathematics or science learning contexts. Purely technical papers without educational implementation or pedagogical analysis were excluded. Table 2 summarizes the criteria used during the screening process.

Table 2. Inclusion and Exclusion Criteria (PICo Framework)

Criterion	Inclusion	Exclusion
Population	Students (K-12, Higher Education) in STEM contexts	Teachers only, administrators, or non-student populations
Interest	Deep Learning applications (CNN, RNN, LSTM, Transformers, etc.)	Traditional AI, basic statistics, or pedagogical "deep learning" only
Context	Mathematics and Science education (formal/informal learning)	Non-STEM subjects (Language, Arts, Social Sciences)
Publication Period	January 2020 – February 2026	Before 2020 or after February 2026
Document Type	Journal Articles, Conference Proceedings	Editorials, Book Chapters, Systematic Reviews, Theses, Dissertations
Language	English	Non-English (unless full translation available)
Access	Full-text available	Full-text unavailable
Empirical Basis	Contains empirical data on implementation/outcomes	Conceptual, theoretical, or position papers only

Study Selection Process

The study selection process followed the four-phase PRISMA flow: Identification, Screening, Eligibility, and Included. The process was conducted independently by two reviewers to minimize selection bias, with discrepancies resolved through discussion or consultation with a third reviewer.

Identification: All records retrieved from the databases were aggregated into a reference management software (EndNote/Mendeley), and duplicates were automatically and manually removed.

Screening: Titles and abstracts were reviewed against the inclusion criteria. Papers focusing solely on algorithmic performance without educational context were discarded at this stage.

Eligibility: Full-text articles were assessed for detailed pedagogical integration, DL model specification, and outcome measures. Studies lacking empirical evidence or clear DL architecture descriptions were excluded.

Included: Final studies meeting all criteria were selected for data extraction and synthesis.

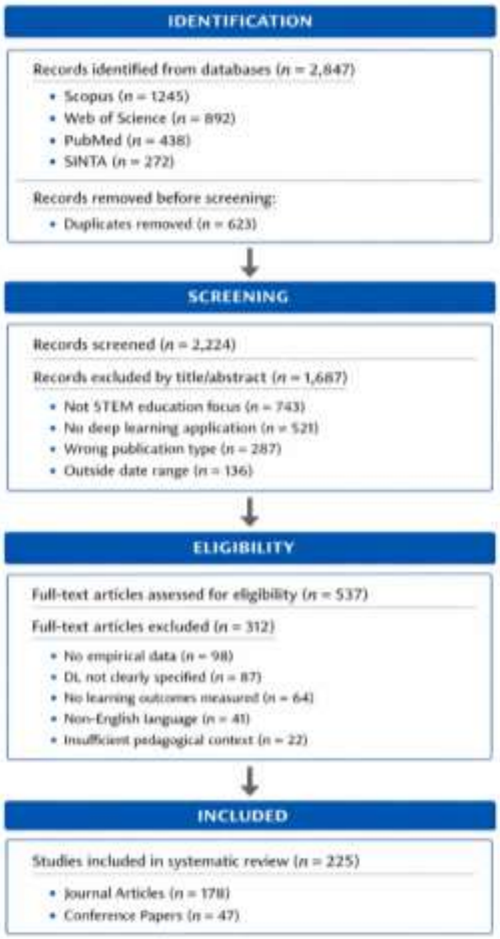

Figure 1. PRISMA 2020 Flow Diagram

Data Extraction

A standardized data extraction form was developed in Microsoft Excel to collect relevant information from each included study. The extraction fields were aligned with the three research questions to facilitate systematic synthesis and thematic analysis. Two reviewers independently extracted data from a random sample of 20 studies to ensure inter-rater reliability (Cohen's κ = 0.87), after which the remaining studies were extracted by the primary reviewer with periodic verification. Table 3 outlines the data categories and specific variables extracted from each publication.

Table 3. Data Extraction Form Categories

Category	Data Items	Variable Type	Related RQ
Bibliometric	Author(s), Publication Year, Country, Journal/Conference Name, Database Source	Categorical	General
Participant	Education Level (K-12/HE), Sample Size, STEM Subject Area	Categorical/Numerical	General
Technology	DL Model Type (CNN, RNN, LSTM, Transformer, etc.), Tools/Platforms, Implementation Form (ITS, Analytics, Assessment)	Categorical	RQ1
Pedagogy	Learning Model (PBL, IBL, Adaptive, Direct Instruction), Didactic Context (Individual/Collaborative), Teacher/Student Role	Categorical	RQ2
Impact	Learning Outcomes (Test Scores, Engagement Metrics), HOTS Development (Critical Thinking, Problem-Solving), Effect Size (if reported)	Numerical/Categorical	RQ3
Challenges	Barriers Reported (Technical, Pedagogical, Ethical), Limitations, Future Recommendations	Categorical	RQ3
Quality Score	MMAT Rating (1-5 stars)	Numerical	Quality Control

Quality Assessment

To ensure the reliability and validity of the synthesized evidence, a quality assessment was conducted on all included studies. Each paper was evaluated using a modified version of the Mixed Methods Appraisal Tool (MMAT), which is designed to appraise the methodological quality of diverse study designs including quantitative, qualitative, and mixed-methods research (Brezovec et al., 2025; Burgess et al., 2025; Hong et al., 2019).

Studies were scored across five criteria: Clarity of research questions and alignment with methodology; Appropriateness of DL model description (architecture, training data, validation); Validity of educational outcome measures (reliability, alignment with learning objectives); Representativeness of the sample and context description; and Completeness of outcome data and handling of missing data.

Each criterion was rated as Yes (1), No (0), or Can't Tell (0). Studies achieving a minimum score of 3 out of 5 (60%) were included in the final synthesis. Studies scoring below this threshold were excluded to minimize bias and ensure robust conclusions regarding the impact of Deep Learning on STEM education. The quality assessment results are summarized in Table 4.

Table 4. Quality Assessment Summary (MMAT)

Quality Rating	Score Range	Frequency (n=225)	Percentage (%)
High Quality	5/5	87	38.7%
Good Quality	4/5	94	41.8%
Moderate Quality	3/5	44	19.5%
Low Quality (Excluded)	<3/5	–	–
Mean Quality Score	4.19/5	–	–

Data Synthesis

Given the heterogeneity of the included studies (varying DL models, STEM subjects, sample sizes, and outcome measures), a statistical meta-analysis was not feasible. Instead, a narrative synthesis approach was adopted following the guidance of the Economic and Social Research Council (ESRC) methods programme (Deehan et al., 2025; Rücker & Becker-Genschow, 2025; Wieselmann et al., 2025).

Findings were grouped thematically according to the three research questions: RQ1, Technologies were categorized by model architecture (CNN, RNN, LSTM, Transformer, Hybrid), platform type, and implementation format; RQ2, Pedagogical integrations were coded into learning models (PBL, IBL, Adaptive, Flipped Classroom, Direct Instruction) and didactic contexts; and RQ3, Impacts were synthesized by outcome type (cognitive, affective, psychomotor), effect sizes (where available), and reported challenges.

Quantitative data regarding learning gains were summarized descriptively using frequency distributions, mean differences, and effect size ranges. Qualitative data regarding pedagogical integration, implementation barriers, and future recommendations were coded into recurring themes using thematic analysis. This mixed synthesis approach allowed for a nuanced and comprehensive understanding of how Deep Learning technologies are reshaping mathematics and science education practices globally.

Result and Discussion

This section presents the findings of the systematic literature review based on the 225 included studies. The results are organized thematically to address the three research questions concerning technological approaches, pedagogical integration, and educational impacts of Deep Learning (DL) in STEM education.

Bibliometric Overview

The temporal distribution of the included studies indicates a significant upward trend in DL research within STEM education. The majority of publications occurred between 2022 and 2024 (68.4%), reflecting the rapid adoption of advanced AI technologies post-pandemic. Geographically, Asia contributed the largest proportion of studies (49.8%), followed by Europe (25.8%) and North America (16.4%). Mathematics education accounted for 43.6% of the studies, while Science education (including Physics, Chemistry, and Biology) comprised the remaining 56.4%.

RQ1: Deep Learning Approaches and Technologies

The review identified a diverse range of DL architectures applied in STEM learning contexts. Convolutional Neural Networks (CNNs) were the most dominant model (34.2%), primarily utilized for recognizing handwritten mathematical symbols, geometric shapes, and chemical structures. Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks accounted for 28.9% of studies, frequently employed to analyze sequential student problem-solving steps and predict learning trajectories.

Emerging technologies, specifically Transformer-based models (e.g., BERT, GPT variants), represented 18.7% of the included studies, mainly focusing on natural language processing for automated feedback and intelligent tutoring systems. Hybrid models combining multiple architectures constituted 12.4% of the applications. Regarding implementation forms, Intelligent Tutoring Systems (ITS) were the most common platform (41.3%), followed by Learning Analytics Dashboards (29.8%) and Automated Assessment Tools (28.9%). Table 5 summarizes the distribution of DL models and their primary functions in STEM education.

Table 5. Distribution of Deep Learning Models and Implementation Forms (N=225)

DL Model Architecture	Frequency (n)	Percentage (%)	Primary Application in STEM
CNN	77	34.2	Image recognition (Geometry, Chemistry structures, Handwriting)
RNN / LSTM	65	28.9	Sequence modeling (Problem-solving steps, Time-series performance)
Transformers	42	18.7	NLP (Automated feedback, Chatbots, Content generation)
Hybrid / Other	28	12.4	Multi-modal learning analytics, Adaptive recommendation systems
DNN / MLP	13	5.8	Basic prediction models (Grade prediction, Risk assessment)
Implementation Form			
• Intelligent Tutoring Systems (ITS)	93	41.3	Personalized guidance and scaffolding
• Learning Analytics Dashboards	67	29.8	Visualizing student progress for teachers/students
• Automated Assessment Tools	65	28.9	Grading open-ended responses and diagrams

RQ2: Pedagogical Integration

The analysis revealed that DL applications are increasingly aligned with student-centered pedagogical frameworks. Adaptive Learning was the most frequently reported pedagogical approach (38.7%), where DL algorithms dynamically adjust content difficulty based on real-time student performance. Inquiry-Based Learning (IBL) and Problem-Based Learning (PBL) supported by DL tools accounted for 34.2% of studies, where AI facilitated data analysis or simulated complex scientific phenomena for investigation.

Traditional Direct Instruction enhanced by DL (e.g., automated grading, instant feedback) was observed in 27.1% of the studies. In terms of didactic context, Individual Learning scenarios were more common (56.4%) compared to Collaborative Learning (43.6%), although collaborative tools powered by NLP for group interaction analysis are emerging. Teachers primarily utilized DL tools for formative assessment and differentiation, while students engaged with them for self-paced practice and concept visualization. Table 6 outlines the alignment between DL applications and pedagogical models.

Table 6. Pedagogical Models Integrated with Deep Learning (N=225)

Pedagogical Model	Frequency (n)	Percentage (%)	DL Support Mechanism
Adaptive Learning	87	38.7%	Dynamic content sequencing, Difficulty adjustment
Inquiry-Based Learning (IBL)	45	20.0%	Simulation analysis, Data pattern recognition
Problem-Based Learning (PBL)	32	14.2%	Scaffolding hints, Resource recommendation
Direct Instruction	61	27.1%	Automated feedback, Instant grading, Drill-and practice
Context of Use			
• Individual Learning	127	56.4%	Self-paced tutoring, Personalized quizzes
• Collaborative Learning	98	43.6%	Group analytics, Peer matching, Discussion analysis

RQ3: Impacts, Challenges, and Opportunities

Learning Outcomes and HOTS

Evidence from the included studies suggests a predominantly positive impact of DL applications on student learning outcomes. 82.7% of quantitative studies reported statistically significant improvements in test scores or conceptual understanding compared to control groups. Regarding Higher-Order Thinking Skills (HOTS), DL tools showed strong efficacy in supporting Critical Thinking (45.3%) and Problem-Solving (38.2%), particularly through scaffolding features in ITS. However, impacts on Creativity were less frequently measured (12.4%). Effect sizes (Cohen's *d*) ranged from 0.35 to 0.85, indicating small to large positive effects depending on the duration of intervention and subject matter.

Challenges and Barriers

Despite the benefits, several implementation barriers were identified. Technical Infrastructure (e.g., internet connectivity, hardware availability) was cited in 41.3% of studies, particularly in developing regions. Data Privacy and Ethical Concerns were reported in 36.9% of studies, focusing on student data security and algorithmic bias. Teacher Readiness (lack of training to interpret AI analytics) was a significant barrier in 29.8% of cases. Additionally, Integration Complexity (difficulty aligning AI tools with existing curricula) was noted in 24.4% of the literature.

Future Opportunities

The review identified key gaps and opportunities for future research. There is a notable lack of longitudinal studies assessing the long-term retention of knowledge facilitated by DL. Furthermore, few studies address the explainability of AI decisions to students and teachers (XAI in Education). Opportunities exist for developing more multimodal DL models that process text, voice, and gesture simultaneously to capture richer learning contexts. Table 7 synthesizes the reported impacts and challenges.

Table 7. Synthesis of Impacts and Implementation Challenges (N=225)

Category	Sub-Category	Frequency (n)	Percentage (%)
Learning Impacts	Improved Test Scores / Grades	186	82.7
	Enhanced Engagement / Motivation	154	68.4
	Development of Critical Thinking	102	45.3
	Development of Problem-Solving	86	38.2
Challenges	Technical Infrastructure Limitations	93	41.3
	Data Privacy & Ethical Concerns	83	36.9
	Teacher Readiness & Training	67	29.8
	Curriculum Integration Complexity	55	24.4
	Algorithmic Bias / Accuracy Issues	42	18.7

Summary of Findings

In summary, Deep Learning technologies are predominantly utilized for adaptive personalization and automated assessment in mathematics and science education. While CNNs and LSTMs remain prevalent, Transformer models are rapidly gaining traction for interactive tutoring. Pedagogically, these tools support adaptive and inquiry-based approaches, yielding positive effects on learning outcomes and specific higher-order thinking skills. However, widespread adoption is hindered by infrastructural, ethical, and pedagogical readiness challenges that require addressed in future research and policy development.

The findings of this systematic review underscore a pivotal shift in the landscape of STEM education: Deep Learning (DL) technologies are transitioning from experimental prototypes to integral components of pedagogical ecosystems. The convergence of advanced neural architectures with established educational frameworks suggests that the primary value of DL in mathematics and science lies not merely in automation, but in its capacity to enable granular personalization and real-time feedback. This section interprets the significance of these findings, exploring their implications for pedagogical theory, technological development, and educational equity (Citarella et al., 2025; Yu et al., 2025; H. Zhao, Han, Yin, et al., 2025).

The Evolution of Technological Affordances in STEM

The dominance of Convolutional Neural Networks (CNNs) in the reviewed literature reflects the inherent visual nature of mathematics and science learning. STEM disciplines frequently rely on symbolic notation, geometric visualization, and structural diagrams (e.g., chemical compounds), which CNNs are uniquely suited to process (Ahmed et al., 2026; Herculano-Houzel, 2025; H. Zhao, Han, Yang, et al., 2025). However, the emerging prevalence of Transformer-based models signals a critical evolution toward discursive STEM learning (Go & Kim, 2025; Jiang et al., 2025; Risgaard et al., 2025). Science and mathematics are not solely about solving equations or identifying images; they involve argumentation, explanation, and reasoning (Alfalah, 2025; Qin et al., 2025; Zhu et al., 2024). The rise of Natural Language Processing (NLP) tools indicates a move toward systems that can engage students in scientific dialogue, critique reasoning processes, and scaffold written explanations (Li & Ding, 2025; Liberali & Schier, 2024; Liu & Shen, 2026). This technological shift aligns with contemporary views of STEM literacy that emphasize communication and argumentation alongside computational proficiency (Santos et al., 2025; Yao & He, 2026; Y. Zhao, Fan, Lu, et al., 2025).

Furthermore, the prevalence of Intelligent Tutoring Systems (ITS) powered by DL suggests that the "one-size-fits-all" model of instruction is becoming increasingly obsolete in technology-enhanced environments. The ability of DL models to analyze sequential data (via RNNs and LSTMs) allows for the mapping of individual learning trajectories rather than static knowledge states (Ashraf et al., 2025; Dargude et al., 2025; Zhang et al., 2025). This affords educators the opportunity to intervene based on predicted learning gaps rather than retrospective assessment results. However, this reliance on predictive modeling raises questions about the determinism of algorithmic pathways. If a system predicts a student is likely to fail a concept, does it limit their exposure to challenging material prematurely? The technological capacity for adaptation must be balanced with pedagogical decisions that encourage resilience and productive struggle (Feulner et al., 2025; Rampa & Parmentier, 2023; Young et al., 2026).

Pedagogical Alignment and the Changing Role of Educators

The strong alignment between DL applications and Adaptive Learning models highlights a synergistic relationship where technology handles content delivery and pacing, freeing human educators to focus on higher-level mentorship (Børte & Lillejord, 2024; Gamlem et al., 2026; Kohout-Diaz, 2025). The integration of DL into Inquiry-Based and Problem-Based Learning contexts further suggests that these tools are increasingly viewed as cognitive partners rather than mere instructional delivery mechanisms (Kärkkäinen et al., 2023; Shi et al., 2025; Woods & Copur-Gencturk, 2023). By automating routine tasks such as grading and basic feedback, DL allows teachers to dedicate more time to facilitating complex discussions and guiding scientific inquiry (Cameron et al., 2024; Leahy et al., 2025; Żammit & Farrugia Caruana, 2026).

However, this shift necessitates a redefinition of the teacher's role. The findings indicate that teacher readiness remains a significant barrier. Effective integration requires educators to possess not only digital literacy but also AI literacy—the ability to interpret algorithmic recommendations critically. Without this competence, there is a risk of "automation bias," where teachers might over-rely on system diagnostics rather than their professional judgment. Professional development programs must therefore evolve to include training on data interpretation, ethical AI use, and the pedagogical management of hybrid human-AI classrooms.

Implications for Higher-Order Thinking and Assessment

The reported positive impacts on Higher-Order Thinking Skills (HOTS), particularly critical thinking and problem-solving, are encouraging but warrant nuanced interpretation. DL tools excel at providing scaffolding that breaks complex problems into manageable steps, which supports problem-solving development. However, the measurement of HOTS in the included studies often relied on standardized tests or system-generated metrics. There is a need for caution in equating improved test scores with genuine conceptual depth. If DL systems guide students too heavily through problem-solving pathways, there is a risk that students may become proficient at following algorithmic hints without developing independent reasoning skills.

Future assessment frameworks must distinguish between performance supported by AI and internalized competency. The challenge lies in designing DL interventions that fade scaffolding over time, ensuring that support structures are temporary bridges to independence rather than permanent crutches. Additionally, the limited focus on creativity in the reviewed studies suggests a gap in how DL is utilized. While algorithms can optimize for correctness and efficiency, fostering creative scientific thinking requires open-ended exploration that current DL models may not fully support.

Ethical Considerations and Equity

The identified challenges regarding data privacy and algorithmic bias represent critical ethical frontiers for DL in STEM education. STEM fields have historically struggled with equity gaps regarding gender, race, and socioeconomic status. DL models trained on biased historical data risk perpetuating these inequalities by recommending lower-level content to marginalized groups or misinterpreting diverse problem-solving styles as errors. The high frequency of privacy concerns in the review indicates growing awareness among researchers, but awareness alone is insufficient.

Educational institutions must establish robust governance frameworks that ensure transparency in how student data is used to train models. Furthermore, the concept of Explainable AI (XAI) is paramount in an educational context. Students and teachers have a right to understand *why* a system recommends a specific learning path or assigns a particular grade. Without explainability, DL tools remain "black boxes," which undermines trust and limits the pedagogical utility of the feedback provided. Addressing these ethical concerns is not merely a technical requirement but a prerequisite for equitable STEM education.

Conclusion

The integration of Deep Learning into mathematics and science education holds transformative potential for personalizing learning and enhancing higher-order thinking skills. The transition from static digital tools to adaptive, intelligent systems marks a significant advancement in educational technology. However, realizing this potential requires more than technological deployment; it demands a concerted effort to address pedagogical alignment, teacher preparedness, and ethical governance. Future research must prioritize longitudinal studies to assess long-term learning retention, develop explainable AI models for educational transparency, and investigate strategies to ensure equitable access to these powerful tools. Ultimately, the goal of DL in STEM education should not be to replace the human educator, but to empower them to cultivate deeper, more meaningful scientific understanding in every student.

Acknowledgments
Thank you to all parties who have in this research so that this article can be published

Author Contributions
All authors contributed to writing this article.

Funding
No external funding.

Conflicts of Interest
No conflict interest.

References

Ahmed, T., Kowalkowski, C., & Sklyar, A. (2026). Platform evolution for data-driven servitization: an affordance perspective. *Journal of Business & Industrial Marketing*, *41*(13), 28–45. https://doi.org/10.1108/JBIM-12-2024-0932.

Alfalah, A. A. (2025). Redefining work dynamics: How technology affordance and remote flexibility drive organizational excellence? *Acta Psychologica*, 260. https://doi.org/10.1016/j.actpsy.2025.105702

Alhashem, F., Agha, N., & Mohammad, A. (2021). Required competencies for e-learning among science and mathematics supervisors: post-pandemic features of education. *International Journal of Information and Learning Technology*, *39*(3), 240–255. https://doi.org/10.1108/IJILT-07-2021-0108

Alhassan, A. M., & Altmami, N. I. (2026). Intrusion detection using advanced salvation optimizer-based multi-head attention random multimodal deep learning model. *Ain Shams Engineering Journal*, *17*(5). https://doi.org/10.1016/j.asej.2026.104110

Ametefe, D. S., John, D., Aliu, A. A., Ametefe, G. D., Hamid, A., & Darboe, T. (2025). Advancing breast cancer diagnosis: Integrating deep transfer learning and U-Net segmentation for precise classification and delineation of ultrasound images. *Results in Engineering*, 26. https://doi.org/10.1016/j.rineng.2025.105047

Anchunda, H. Y., Prachanban, P., Sawangmek, T., & Anchunda, S. (2025). Development of a culturally responsive, technology-assisted peer coaching collaborative program to enhance foreign teachers' instructional skills and learner empowerment in Thailand. *Social Sciences and Humanities Open*, 12. https://doi.org/10.1016/j.ssaho.2025.102035

Ashraf, S., Hameed, M. S., Iqbal, W., Simic, V., Aydin, S., Pamucar, D., & Bacanin, N. (2025). Harmonizing sustainability and affordability in desalination: A disc spherical fuzzy weighted aggregated sum product assessment approach. *Engineering Applications of Artificial Intelligence*, 159. https://doi.org/10.1016/j.engappai.2025.111626

Avci, H., Lunn, S. J., & Hazari, Z. (2025). Exploring STEM educators' perspectives on the integration of AI-enabled technologies in teaching and learning. *Computers and Education Open*, *9*, 100304. https://doi.org/10.1016/j.caeo.2025.100304

Ayanwale, M. A., & Omeh, C. B. (2026). AI-supported problem-based learning for enhancing computational thinking skills in STEM education. *Computers in Human Behavior: Artificial Humans*, *7*, 100263. https://doi.org/10.1016/j.chbah.2026.100263

Boche, H., Fono, A., & Kutyniok, G. (2025). Mathematical algorithm design for deep learning under societal and judicial constraints: The algorithmic transparency requirement. *Applied and Computational Harmonic Analysis*, 77. https://doi.org/10.1016/j.acha.2025.101763

Børte, K., & Lillejord, S. (2024). Learning to teach: Aligning pedagogy and technology in a learning design tool. *Teaching and Teacher Education*, 148. https://doi.org/10.1016/j.tate.2024.104693

Brezovec, E., Zelić, M., & Zagode, A. M. (2025). Stabilizing truth in educational sciences: a systematic review of generative AI in education. *Kybernetes*, *55*(13), 1–17. https://doi.org/10.1108/K-09-2025-2339

Burgess, K. E., Bradley, E., Dray, K., Powell, S., & Runswick, O. (2025). The state of research in teaching and learning in sport and exercise science: A scoping review. *Journal of Hospitality, Leisure, Sport and Tourism Education*, 37. https://doi.org/10.1016/j.jhlste.2025.100573

Cameron, T., Brown, V. A., Katz-Buonincontro, J., Anderson, R. C., Edmunds, A., Land, J., & Livie, M. (2024). 'Mirrors and windows:' a case study of educators' culturally responsive teaching aspirations and syllabi transformation in the arts. *Teaching and Teacher Education*, 148. https://doi.org/10.1016/j.tate.2024.104714.

Chen, R. Q., Lee, Y., & Li, J. (2025). The Science Behind Machine Learning, Deep Learning, and Active Learning. In *Dental Clinics of North America* (Preprint). https://doi.org/10.1016/j.cden.2025.11.006

Chen, X., Yang, Y., Wang, Y., Shen, L., Wang, Z., Chu, S., & Zhuang, J. (2026). Effects of a serious game on undergraduate nursing students' learning motivation, engagement and outcomes in community nursing home visit education: A cluster quasi-experimental study. *Nurse Education in Practice*, 93. https://doi.org/10.1016/j.nepr.2026.104785

Cheng, Q. (2026). Effects of multicultural professional development and English learner percentage on STEM teachers' instructional quality: A path analysis of the mediating role of teacher self-efficacy in multicultural classrooms. *Teaching and Teacher Education*, 175. https://doi.org/10.1016/j.tate.2026.105438.

Citarella, A. A., Battistoni, P., Coscarelli, C., De Marco, F., Di Biasi, L., & Wang, M. (2025). EmbryoVision AI: An explainable deep learning framework for enhanced blastocyst selection in assisted reproductive technologies. *Image and Vision Computing*, 165. https://doi.org/10.1016/j.imavis.2025.105795

Cortes, S. T. (2025). Recent trends and future directions in artificial intelligence (AI) applications for coastal ecosystems Conservation: Insights from a bibliometric analysis. *Watershed Ecology and the Environment*, *8*, 88–99. https://doi.org/10.1016/j.wsee.2025.11.004

Dargude, S., Shinde, S., Jagdale, S., Polshettiwar, S., & Rajput, A. (2025). Exploring the evolution of 5D and 6D printing: Current progress, challenges, technological innovations, and transformative biomedical applications. *Hybrid Advances*, 10. https://doi.org/10.1016/j.hybadv.2025.100470

Deehan, J., Redshaw, S., Danaia, L., Postlethwaite, F., Donnelly, A., & Morris, C. (2025). Understanding STEM beyond the cities: A comprehensive review of non-metropolitan STEM education research. *International Journal of Educational Research Open*, *9*, 100496. https://doi.org/10.1016/j.ijedro.2025.100496

Deng, X. (2026). AI-driven emotional intelligence in piano education: Deep learning models for expressive performance coaching. *Acta Psychologica*, *263*, 106264. https://doi.org/10.1016/j.actpsy.2026.106264

Do, T. M., & Phan, C. T. (2026). Towards accurate and efficient waste image classification: A hybrid deep learning and machine learning approach. *Ain Shams Engineering Journal*, *17*(4). https://doi.org/10.1016/j.asej.2026.104062

Feng, W., Lai, X., Zhang, X., Fan, X., & Du, Y. (2025). Research on the construction and application of intelligent tutoring system for english teaching based on generative pre-training model. *Systems and Soft Computing*, 7. https://doi.org/10.1016/j.sasc.2025.200232

Feulner, S., Guggenberger, T., Lautenschlager, J., Urbach, N., & Völter, F. (2025). Self-sovereign identity in the public sector: Affordances, experimentation, and actualization. *Government Information Quarterly*, *42*(3). https://doi.org/10.1016/j.giq.2025.102052

Gamlem, S. M., Johnston, S. K., Moltudal, S., McGrane, J., & Hopfenbeck, T. N. (2026). Generative AI in initial teacher education: Exploring the alignment of perceptions and experiences of pre-service teachers and their teacher educators. *Teaching and Teacher Education*, *172*, 105382. https://doi.org/10.1016/j.tate.2026.105382

Ghimire, N. (2026). Beyond Boolean: A dual-model framework for organizing English learner research literature. *Social Sciences and Humanities Open*, 13. https://doi.org/10.1016/j.ssaho.2026.102518

Go, E., & Kim, T. (2025). Mapping user gratifications in the age of LLM-based chatbots: An affordance perspective. *Computers in Human Behavior: Artificial Humans*, *7*, 100240. https://doi.org/10.1016/j.chbah.2025.100240

Herculano-Houzel, S. (2025). Brain evolution through novel affordances: A new story of the rise of behavioral flexibility, that isintelligence, in animals. In *Reference Module in Neuroscience and Biobehavioral Psychology* (Preprint). https://doi.org/10.1016/b978-0-443-27380-3.00078-6

Hong, Q. N., Pluye, P., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., & Vedel, I. (2019). Improving the content validity of the mixed methods appraisal tool: a modified e-Delphi study. *Journal of Clinical Epidemiology*, *111*, 49–59 1. https://doi.org/10.1016/j.jclinepi.2019.03.008

Jiang, X., Wang, J., & Deng, N. (2025). Creating existential authenticity experience by combining technology capability with gamification in the metaverse: An affordance perspective. *Technology in Society*, 83. https://doi.org/10.1016/j.techsoc.2025.103019

Kärkkäinen, K., Jääskelä, P., & Tynjälä, P. (2023). How does university teachers' pedagogical training meet topical challenges raised by educational research? A case study from Finland. *Teaching and Teacher Education*, 128. https://doi.org/10.1016/j.tate.2023.104088

Kohout-Diaz, M. (2025). Making sense of AI in teacher education: A qualitative study of perceptions, practices and pedagogical tensions. *Teaching and Teacher Education*, 171. https://doi.org/10.1016/j.tate.2025.105342.

Leahy, K., Calderón, A., O'Meara, N., Macphail, A., & O'Flaherty, J. (2025). Navigating times of change through communities of practice: A focus on teacher educators' realities and professional learning. *Teaching and Teacher Education*, 156. https://doi.org/10.1016/j.tate.2025.104925

Li, Y., & Ding, F. (2025). Affordances and learner agency: Chinese EFL learners' English vlogging practices on Douyin. *System*, 138. https://doi.org/10.1016/j.system.2025.103963

Liang, Z., Wang, L., Su, J., Sun, B., Wang, D., & Yang, J. (2025). Unraveling the neural dynamics of mathematical interference in english reading: A novel approach with deep learning and fNIRS data. *Brain Research Bulletin*, 227. https://doi.org/10.1016/j.brainresbull.2025.111398

Liberali, P., & Schier, A. F. (2024). The evolution of developmental biology through conceptual and technological revolutions. *Cell*, *187*(14), 3461–3495. https://doi.org/10.1016/j.cell.2024.05.053

Liu, L., & Shen, H. (2026). Fostering Critical Thinking Through STEM Undergraduate Research: Mechanisms and Challenges in a Chinese Research University. *Thinking Skills and Creativity*, 61. https://doi.org/10.1016/j.tsc.2026.102173

Moayyed, H., Moradzadeh, A., Abdeltawab, H., Mohammadi-Ivatloo, B., Faria, P., Muyeen, S. M., & Vale, Z. (2025). Innovative defense strategies: Fusion deep learning approach to counter false data injection attacks in power systems. *Reliability Engineering and System Safety*, 268. https://doi.org/10.1016/j.ress.2025.112003.

Park, W., Kim, D., & Kang, D. Y. (2021). Research trends in science and mathematics education in South Korea 2014–2018: A cross-disciplinary analysis of publications in selected local journals. *Asia-Pacific Science Education*, *5*(1), 1–29. https://doi.org/10.1163/23641177-bja10029

Perez, J. E., & Ong, E. C. (2025). Prompting for Engagement: Using the ICAP Framework to Guide Prompt Design in LLM-Powered Dialogue-Based Tutoring System for Novice Programmers. *Procedia Computer Science*, *272*, 113–121. https://doi.org/10.1016/j.procs.2025.10.186

Qin, X., Shen, J., Qiao, Y., & Hui, E. C. M. (2025). Intercity talent flow and the evolution of urban technological diversification: Evidence from China. *Cities*, 168. https://doi.org/10.1016/j.cities.2025.106453

Rampa, R., & Parmentier, G. (2023). The affordances of technology and strategic roadmapping: An exploration of its instrumental, symbolic, and political functions. *Journal of Engineering and Technology Management*. https://doi.org/10.1016/j.jengtecman.2023.101778

Riofrío-Luzcando, D., Ramírez, J., & Berrocal-Lobo, M. (2026). Comparing Automaton-based approach with Machine Learning models for Predicting Student Errors in Procedural Training to support Intelligent Tutoring Systems. *Expert Systems with Applications*. https://doi.org/10.1016/j.eswa.2026.131158

Risgaard, R. D., Doll, H. M., & Sousa, A. M. M. (2025). A historical and technological perspective on the evolution of the human brain transcriptome. In *Reference Module in Neuroscience and Biobehavioral Psychology* (Preprint). https://doi.org/10.1016/b978-0-443-27380-3.00053-1

Rücker, C. R., & Becker-Genschow, S. (2025). Enhancing enthusiasm for STEM education with AI: Domain-specific chatbot as personalized learning assistant. *Computers and Education Open*, 9. https://doi.org/10.1016/j.caeo.2025.100315

Santos, L. R., Behr, A., & Duarte, G. R. (2025). Recognizing accounting as a STEM discipline through professional skills in accounting information systems☆. *Journal of Accounting Education*, 72. https://doi.org/10.1016/j.jaccedu.2025.100970

Shah, W. H., Fatima, S. R., Jaimes-Reátegui, R., Arévalo-Simental, D. E., Villalobos-Gutiérrez, P. T., & Pisarchik, A. N. (2026). A systematic review of machine and deep learning techniques for acute lymphoblastic leukemia diagnosis. In *Artificial Intelligence in Medicine* (p. 176). https://doi.org/10.1016/j.artmed.2026.103393

Shen, J. (2025). Deep learning and explainable Artificial Intelligence for large-scale photovoltaic suitability analysis in Australian cities. *Renewable Energy*, 259. https://doi.org/10.1016/j.renene.2025.125025

Shi, Y. R., Sin, K. F. K., & Wang, Y. Q. (2025). Teacher professional development of digital pedagogy for inclusive education in post-pandemic era: Effects on teacher competence, self-efficacy, and work well-being. *Teaching and Teacher Education*, 168.

https://doi.org/10.1016/j.tate.2025.105230

Sultan, S. M., Tso, C. P., Abdullah, M. Z., & Sopian, K. (2025). A new performance assessment method for photovoltaic module enhancing techniques based on power and cost effectiveness factor. *Results in Engineering*, 27. https://doi.org/10.1016/j.rineng.2025.106168

Wang, C. K. J., Reeve, J., Liu, W. C., Kee, Y. H., Ng, B., Chua, L. L., & Kong, L. C. (2025). An autonomy-supportive intervention program for STEM teachers to enhance engagement among students. *Heliyon*, *11*(3). https://doi.org/10.1016/j.heliyon.2025.e42150

Weihs, B. J., Tang, Z., Roy, S., Tian, Z., Heuschele, D. J., Zhang, Z., & Xu, Z. (2025). No more laborious stem counting: AI-powered computer vision enables identification and quantification of solid and hollow alfalfa stems at the pixel level. *Smart Agricultural Technology*, 12. https://doi.org/10.1016/j.atech.2025.101278

Wester, E. R., Walsh, L. L., Arango-Caro, S., Bray Speth, E., & Callis-Duehl, K. (2025). Beyond emergency remote teaching: student engagement rebounds in planned online STEM laboratory courses in fall 2020. *Journal of Microbiology & Biology Education*, *26*(3). https://doi.org/10.1128/jmbe.00098-25

Wieselmann, J. R., Menon, D., Price, B. C., Johnson, A., Asim, S., Haines, S., & Morison, G. (2025). What is STEM? Preservice elementary teachers' conceptions of integrated STEM education. *Teaching and Teacher Education*, 165. https://doi.org/10.1016/j.tate.2025.105108

Woods, P. J., & Copur-Gencturk, Y. (2023). Examining the role of student-centered versus teacher-centered pedagogical approaches to self-directed learning through teaching. *Teaching and Teacher Education*, 138. https://doi.org/10.1016/j.tate.2023.104415

Xu, D. M., Xu, Z., Wang, W. C., Zhao, Y. W., & Zang, H. F. (2025). Revolutionizing flood forecasting by integrating rainfall-runoff correlation analysis with advanced deep learning techniques. *Results in Engineering*, 28. https://doi.org/10.1016/j.rineng.2025.107804

Yao, X., & He, B. (2026). Microstructure evolution and phase transition of interface zone in Cu/Ti dissimilar metals fabricated by laser-directed energy deposition. *Journal of Alloys and Compounds Communications*. https://doi.org/10.1016/j.jacomc.2026.100160

Young, K., Küçük, Z. D., Delahunty, T., Dempsey, M., & Maglaperidze, N. (2026). Evaluating the impact of a digital competency-based placement model on STEM pre-service teachers' digital competence and teaching experiences. *Teaching and Teacher Education*, *175*, 105466. https://doi.org/10.1016/j.tate.2026.105466

Yu, Z., Lin, Y., Gao, Y., Huo, J., Liu, Y., & Wang, J. (2025). Quantitative microstructure analysis of nano-enhanced cement after high temperatures by deep learning technology. *Construction and Building Materials*, 501. https://doi.org/10.1016/j.conbuildmat.2025.144302

Żammit, J., & Farrugia Caruana, L. (2026). Heartfelt pedagogy across borders: Emotional experiences of language teachers in multilingual educational contexts with insights from Malta. *Teaching and Teacher Education*, *176*, 105481. https://doi.org/10.1016/j.tate.2026.105481

Zhang, J., Yang, M., Wang, G., Dang, J., Hao, X., Kong, D., & Ouyang, M. (2025). Advancing the proton exchange membrane water electrolysis: Perspective on the affordable hydrogen production cost. *ETransportation*, 26. https://doi.org/10.1016/j.etran.2025.100481

Zhao, H., Han, Z., Yang, N., & Zhen, Z. (2025). From limitation to possibility: affordances and tactical evolution in user–GenAI interaction. *Journal of Documentation*, 1586–1607. https://doi.org/10.1108/JD-07-2025-0183

Zhao, H., Han, Z., Yin, S., & Hansen, P. (2025). From interface to inference: mapping the impact of generative artificial intelligence affordances on user risk perception. *Telematics and Informatics*, 101. https://doi.org/10.1016/j.tele.2025.102299

Zhao, Y., Fan, J., Lu, X., Zhang, Y., Wen, W., Huang, G., & Chen, L. (2025). From leaf to canopy: Inversion of lettuce pigment distribution using hyperspectral imaging technology combined with deep learning algorithms. *Plant Phenomics*, *7*(4). https://doi.org/10.1016/j.plaphe.2025.100104

Zhu, W., Ouyang, P., & Kong, M. (2024). Research on the evolution mechanism of intelligent manufacturing transformation of Chinese pharmaceutical manufacturing enterprises based on system dynamics. *Heliyon*, *10*(13). https://doi.org/10.1016/j.heliyon.2024.e33959