International Journal of Informatics and Information Systems
Vol. 8, No. 3, September 2025, pp. 135-147

An Android-Based Multimodal AI Application for Contextual Environmental Learning in Children

Andhika Rafi Hananto[1,*], Muhammad Izha Rahardian[2]

[1]Department of Computer Science, Universitas Kristen Satya Wacana, Salatiga, Indonesia

[2]Department of Information Technology, Jenderal Soedirman Universitu, Indonesia

(Received March 10, 2025; Revised May 30, 2025; Accepted August 5, 2025; Available online September 10, 2025)

Abstract

Children's limited engagement with nature in the digital era poses a growing challenge for environmental education. This study presents the development of an Android-based educational application that leverages multimodal artificial intelligence (AI)—specifically the Google Gemini model—to facilitate contextual environmental learning for preschool and elementary-aged children. Using a prototyping methodology, the application integrates image capture, cloud-based processing through a FastAPI backend, and a Flutter-based interface designed for young learners. The system allows children to photograph plants and receive real-time, age-appropriate explanations about plant names, characteristics, and ecological functions in a narrative format. A limited usability trial involving children of varying age groups demonstrated positive engagement and curiosity, indicating the app's potential as an interactive and enjoyable learning medium. Despite occasional inaccuracies in AI-generated descriptions and reliance on internet connectivity, user feedback suggested strong interest and educational value. Future enhancements will focus on developing localized plant databases, improving accuracy, and incorporating gamification elements. Overall, this study contributes to the growing field of AI-driven educational technology, demonstrating how multimodal AI can effectively bridge digital learning with real-world environmental experiences.

Keywords: Artificial Intelligence; Multimodal Learning; Google Gemini; Contextual Learning; Environmental Education; Android Application; Children

1. Introduction

The pervasive integration of digital technologies into children's daily lives has significantly altered how they interact with the world around them. Over the past two decades, mobile devices such as smartphones and tablets have become central to children's learning and entertainment experiences, often replacing direct engagement with nature [1]. This growing reliance on digital media has led to increasing concerns among educators and psychologists, as prolonged screen exposure tends to reduce children's opportunities for outdoor exploration, unstructured play, and interaction with natural environments [2]. The resulting detachment from nature can have adverse effects on their cognitive, emotional, and social development, limiting the holistic growth that is vital during early childhood.

The COVID-19 pandemic further exacerbated this digital dominance, as lockdowns and the shift to online learning drastically curtailed physical activities and direct social interactions among children [3]. While digital platforms provided continuity in education during periods of restriction, they simultaneously intensified children's dependence on screens and minimized their opportunities for tactile and sensory experiences that nature uniquely provides. Consequently, children are now more exposed to digital representations of the natural world than to the physical environments themselves.

This shift presents a paradox: while digital technologies offer vast educational potential, they also risk alienating children from the very ecosystems they must learn to understand and protect. Research has demonstrated that interaction with natural environments positively influences children's language development, social communication, and emotional regulation, helping them develop richer vocabularies and stronger empathetic connections to their

*Corresponding author: Andhika Rafi Hananto (andhikarafihananto@gmail.com)
DOI: https://doi.org/10.47738/ijiis.v8i3.264

surroundings [4]. Early exposure to nature through play and exploration cultivates curiosity, observational skills, and a foundational sense of ecological responsibility [5]. Unfortunately, such opportunities are increasingly rare in urban and technology-saturated contexts.

The challenge, therefore, lies in reintegrating environmental experiences into the digital learning landscape—bridging the gap between children's technological engagement and their natural curiosity. Conventional educational approaches often rely heavily on classroom-based instruction, with limited emphasis on contextual and experiential learning. These static methods can make environmental education feel abstract and disconnected from children's lived experiences [6]. To address this gap, contextual learning has emerged as a promising pedagogical approach that emphasizes meaningful connections between knowledge and real-world experience. It encourages active learning through direct interaction, reflection, and problem-solving in authentic settings [7]. In early childhood education, contextual learning not only enhances comprehension but also strengthens intrinsic motivation by allowing children to explore topics that are immediately relevant to their environment.

Recent advancements in artificial intelligence (AI) and educational technology have opened new possibilities for implementing contextual learning in dynamic, interactive ways. In particular, multimodal AI systems, capable of processing diverse data types—such as text, images, audio, and video—have shown significant potential in facilitating personalized and adaptive learning experiences. Among these, Google Gemini, a Multimodal Large Language Model (MLLM), represents a new generation of AI systems designed to interpret and generate content across multiple modalities with high contextual accuracy [8]. This capability enables AI to serve as an intelligent learning assistant that can visually identify objects, generate explanations, and adapt language to the learner's age and comprehension level.

When applied to environmental education, multimodal AI models such as Gemini can provide children with instant, personalized, and developmentally appropriate feedback on their natural observations. For example, when a child photographs a plant, the AI can recognize its species, describe its features, and narrate its ecological role in a story-like format suitable for young audiences. This interaction exemplifies how AI-driven contextual learning can merge the cognitive benefits of direct observation with the engagement and convenience of digital technology.

Building on these developments, this study seeks to develop and evaluate an Android-based educational application that integrates multimodal AI (Google Gemini) with contextual learning strategies to enhance environmental education for children. The application enables users to capture plant images, receive real-time explanations in child-friendly language, and build a digital record of their discoveries. The system architecture employs a FastAPI-based backend for data processing and a Flutter-based frontend for delivering an intuitive and visually appealing interface designed for preschool and elementary-level learners.

The research adopts a prototyping methodology to enable iterative testing and refinement of both the application's technical features and its educational effectiveness. Through limited usability trials involving children under adult supervision, the study aims to evaluate engagement, comprehension, and user experience. The findings are expected to contribute to the growing body of research on AI-driven educational tools, particularly in the emerging field of multimodal contextual learning for environmental awareness.

In summary, this work addresses the urgent need to harmonize children's technological exposure with meaningful, nature-based learning experiences. By combining AI innovation, contextual pedagogy, and child-centered design principles, the proposed system aspires to transform how environmental education is delivered in the digital age. It aligns with global efforts to promote sustainability education and to cultivate ecological literacy from an early age, while offering a technically robust and pedagogically grounded model for AI-assisted learning [9].

2. Literature Review

Contextual learning is a pedagogical approach that situates learning within real-world contexts, emphasizing the connection between knowledge, experience, and environment. Rather than relying on abstract or decontextualized instruction, contextual learning integrates academic concepts with authentic activities, enabling learners to make meaningful associations between what they learn and what they observe in everyday life [10]. This approach encourages students to construct their understanding actively through exploration, reflection, and problem-solving. In

ISSN 2579-7069

the context of early childhood education, contextual learning is particularly relevant, as it aligns with children's natural curiosity and their tendency to learn through direct experience. By engaging with their surroundings—such as plants, animals, and outdoor environments—children acquire not only factual knowledge but also empathy, awareness, and appreciation for the natural world [10][11].

Environmental learning that employs a contextual approach has been shown to foster greater retention of knowledge and long-term behavioral change. When children can observe and interact with real-life phenomena, such as identifying plants or recognizing ecological relationships, they develop a stronger connection between learning content and lived experiences. This connection enhances cognitive understanding, supports linguistic development, and nurtures emotional intelligence. Moreover, contextual learning aligns closely with constructivist learning theory, which holds that knowledge is actively built by the learner through engagement and reflection rather than passively received from instruction. Therefore, a learning environment that blends technology with experiential learning can provide both cognitive and affective benefits, particularly for environmental education at the primary level [10][11].

Recent developments in artificial intelligence (AI) have opened new possibilities for contextual learning by allowing educational systems to become more adaptive, interactive, and multimodal. The emergence of Multimodal Large Language Models (MLLMs), such as Google Gemini, represents a major leap in the evolution of intelligent educational systems [12]. Unlike traditional models that process only text, MLLMs can interpret and generate multimodal content—including text, images, audio, and video—thus providing a more comprehensive and realistic interaction with learners. In an educational context, this capability enables AI systems to respond meaningfully to visual or auditory inputs, such as recognizing an object from an image and generating an explanation in natural language. Such interactions are highly beneficial in early education, where visual and narrative elements play a key role in sustaining engagement and comprehension [12][13].

The integration of MLLMs in education has also been associated with significant improvements in conceptual learning and student motivation. Studies have demonstrated that multimodal AI enhances learners' ability to understand abstract concepts through visual representations, interactive explanations, and adaptive feedback [13][14]. These systems can simulate dialogic learning experiences, where students receive immediate, contextually relevant feedback tailored to their cognitive level. In the field of environmental education, AI tools like Google Gemini can process visual data such as photographs of plants and generate informative narratives describing their features, benefits, and ecological significance. This capability effectively supports contextual learning by connecting digital content with tangible real-world experiences, encouraging students to explore their surroundings actively while learning scientifically accurate information [13][14][15].

The growing sophistication of multimodal AI has led to the emergence of frameworks that integrate intelligent systems into various educational contexts. Recent research has introduced models capable of combining text and image comprehension, allowing learners to ask visual questions and receive meaningful, context-aware answers. Such systems have proven valuable in science and environmental education, where visual analysis complements textual understanding [15]. Similarly, collaborative AI tutoring frameworks that employ multiple intelligent agents have been developed to distribute learning tasks among specialized components—such as content experts, scaffolding modules, and visual reasoning agents—resulting in a more comprehensive and adaptive tutoring experience [16]. These developments underscore that multimodal AI not only enhances educational interactivity but also broadens the pedagogical potential of learning applications by accommodating diverse learning styles and cognitive needs.

Parallel to advances in AI, the evolution of mobile learning (m-learning) has played a pivotal role in increasing access to education and promoting personalized learning experiences. Mobile applications allow learners to study anytime and anywhere, facilitating flexible and self-directed learning. Research has shown that students who use educational mobile applications demonstrate higher motivation, stronger learning outcomes, and greater engagement than those relying solely on traditional classroom methods [17]. Mobile devices also offer multimedia functionalities—such as cameras, microphones, and interactive touch interfaces—that align naturally with multimodal and contextual learning principles. In particular, the Android operating system, being widely accessible and customizable, serves as an effective platform for developing educational tools that integrate AI-based functionalities [18][19].

ISSN 2579-7069

In designing and developing mobile educational applications, the adoption of a prototyping methodology has been proven to be highly effective [20]. This approach emphasizes iterative design, testing, and refinement based on user feedback, ensuring that the final product aligns with learner needs and educational objectives. In the context of early childhood learning, prototyping allows developers to adjust interface elements, interaction flows, and content delivery methods to match the cognitive and motor abilities of young users. Moreover, it encourages the integration of real-time feedback mechanisms and data-driven improvements, which are crucial for AI-based educational applications that rely on user interaction and engagement [19][20].

An equally critical component of educational technology is the design of the user interface (UI) and overall user experience (UX). For children, intuitive and visually engaging interfaces are essential to maintaining focus and fostering independent learning. Research in this area emphasizes that educational applications for children must adopt vibrant colors, simple icons, clear navigation, and minimal text to support usability and comprehension [21]. Child-friendly design reduces cognitive overload, enabling learners to process information more efficiently and engage with content meaningfully. Furthermore, aesthetic elements such as illustrations, animations, and storytelling can significantly enhance motivation and curiosity, which are key drivers in early learning. Effective UI and UX design ensure that technology supports rather than distracts from learning, particularly in applications that involve interactive exploration such as environmental recognition.

From the synthesis of existing literature, it is evident that the integration of contextual learning, multimodal AI, and mobile technology provides a promising foundation for the development of innovative educational tools. Multimodal AI enables real-time recognition and interpretation of visual data, bridging the gap between digital learning and real-world exploration. When embedded in a mobile platform and supported by intuitive, child-oriented design, such systems can transform passive screen time into active and meaningful learning experiences. The literature also highlights that AI-enhanced contextual learning not only improves knowledge acquisition but also promotes affective and behavioral outcomes such as curiosity, empathy, and environmental stewardship. These insights establish the theoretical, technological, and pedagogical basis for the present study, which focuses on developing an Android-based multimodal AI application that applies contextual learning principles to strengthen environmental awareness in children.

3. Methodology

This research employed a prototyping methodology designed to iteratively develop and refine an Android-based educational application that integrates multimodal artificial intelligence (AI) for contextual environmental learning among children. The prototyping model was selected because it enables an iterative cycle of analysis, design, implementation, and evaluation, where continuous user feedback drives the refinement of both functional and pedagogical components. The process can be mathematically represented as a sequence of iterative transformations:

$$P_{t+1} = f(P_t, U_t, E_t)$$

where $P_t$represents the prototype version at iteration t, $U_t$denotes user feedback collected from real testing environments, and $E_t$expresses the accumulated empirical evidence regarding usability, performance, and learning impact. Convergence is achieved when the system reaches a near-stable state in which no substantial improvement in usability or performance can be observed, represented as:

$$\lim_{t \to \infty} \nabla P_t = 0$$

This indicates that the partial derivatives of the system improvement function approach zero, signifying equilibrium between user requirements and system performance.

The overall architecture of the system follows a client–server model, consisting of a Flutter-based mobile frontend and a FastAPI-based backend. The backend acts as an intermediary between the Android device and the Google Gemini API, enabling multimodal data exchange. When the user captures an image of a plant through the mobile interface, the

ISSN 2579-7069

image is encoded as a matrix of pixel intensities $I(x, y, c)$, where $x$and $y$denote spatial coordinates and $c$represents the color channel (RGB). To ensure optimal recognition performance, the raw image data undergoes normalization and transformation according to:

$$I'(x, y, c) = \frac{I(x, y, c) - \mu_I}{\sigma_I}$$

where $\mu_I$is the mean intensity value and $\sigma_I$is the standard deviation of pixel intensities across the image. This process ensures consistent brightness and contrast across diverse environmental lighting conditions, reducing noise before model inference.

The preprocessed image $I'(x, y, c)$is then mapped into a feature tensor $T \in \mathbb{R}^{h \times w \times d}$, which serves as the multimodal input for Gemini. The AI system processes both visual and textual inputs using an attention-based transformer architecture that computes relevance scores between image embeddings and linguistic representations. The internal mechanism is modeled as:

$$\text{Attention}(Q, K, V) = \text{softmax}(\frac{QK^T}{\sqrt{d_k}})V$$

where Q, K, and Vrepresent the query, key, and value matrices respectively, and $d_k$denotes the dimensionality of the key vectors. The model learns to associate semantic and visual cues, allowing it to generate text that corresponds to visual stimuli such as plant leaves, flowers, or stems.

The linguistic output of Gemini is generated autoregressively, where each token $w_t$is conditioned on previous tokens and the visual context. The probability of generating the next word can be represented as:

$$P(w_t \mid w_{1:t-1}, I') = \text{softmax}(W_o h_t)$$

where $W_o$denotes the output projection matrix and $h_t$represents the hidden state of the decoder at time t. The generated text $S = \{w_1, w_2, \dots, w_T\}$is a structured narrative containing the plant's name, characteristics, ecological benefits, and care instructions expressed in simplified language suitable for children.

To adapt the AI-generated content to developmental language levels, a post-processing module filters complex or abstract sentences using a readability transformation function $R(S)$that adjusts lexical density and sentence length:

$$R(S) = \frac{1}{N} \sum_{i=1}^{N} \frac{\text{len}(w_i)}{\text{freq}(w_i)}$$

where $\text{len}(w_i)$is the word length and $\text{freq}(w_i)$is its normalized frequency in children's linguistic corpora. The readability index $R(S)$is compared against a threshold R^*corresponding to child-level comprehension; if $R(S) > R^*$, lexical simplification is automatically applied through synonym replacement and sentence restructuring.

System usability and educational effectiveness are evaluated through multiple quantitative and qualitative parameters. The first performance indicator is recognition accuracy, defined as the ratio of correctly identified plant species to total recognition attempts:

$$A_r = \frac{N_c}{N_t} \times 100\%$$

where $N_c$represents the number of correctly classified instances and $N_t$is the total number of images processed. The latency performance of the system, representing average response time between image submission and AI-generated output, is expressed as:

$$L = \frac{1}{n}\sum_{i=1}^{n}(t_{o_i} - t_{i_i})$$

where $t_{i_i}$and t_{o_i}denote the input and output timestamps for each trial i.

User satisfaction, derived from feedback obtained during the usability trial, is quantified through a normalized satisfaction function:

$$S_u = \frac{1}{m}\sum_{j=1}^{m}\frac{r_j - r_{\min}}{r_{\max} - r_{\min}}$$

where $r_j$represents the rating given by user j, and $r_{\min}$, $r_{\max}$are the minimum and maximum possible ratings respectively.

To provide a comprehensive measure of system effectiveness, the three primary performance metrics—accuracy, latency, and satisfaction—are combined into a weighted composite performance score P_s:

$$P_s = \alpha A_r + \beta(1 - \frac{L}{L_{\max}}) + \gamma S_u$$

where α, β, and $\gamma$are weighting coefficients satisfying $\alpha + \beta + \gamma = 1$. This equation ensures balance between the technical and experiential aspects of system evaluation, allowing the identification of optimal configurations that maximize both functionality and user engagement.

The AI model also undergoes internal consistency validation through entropy-based evaluation of prediction confidence. The uncertainty of the AI output is estimated using Shannon's entropy formula:

$$H = -\sum_{i=1}^{n} p_i \log(p_i)$$

where $p_i$represents the predicted probability of class i. Lower entropy indicates higher model confidence, while higher entropy reflects uncertainty in plant recognition or narrative generation. This metric is crucial for identifying cases where the AI may produce ambiguous or inaccurate descriptions, prompting additional human verification or database reference.

From an educational standpoint, the contextual learning effectiveness of the system can be expressed through a correlation model that measures the alignment between children's engagement and their knowledge acquisition. Assuming $C_e$represents cognitive engagement and $K_g$denotes gained knowledge, the correlation coefficient is calculated as:

$$\rho_{CK} = \frac{\mathrm{Cov}(C_e, K_g)}{\sigma_{C_e}\sigma_{K_g}}$$

A high positive correlation ($\rho_{CK} > 0.7$) indicates that interaction with the application directly contributes to improved environmental understanding, validating the pedagogical value of the prototype.

 ISSN 2579-7069

The iterative evaluation process also includes computational performance analysis of the backend. The throughput $\Theta$of the API, defined as the number of processed requests per unit time, is given by:

$$\Theta = \frac{R}{T}$$

where $R$is the number of successfully completed recognition requests and $T$is the total observation time. To ensure scalability, the response efficiency $E_r$is optimized through asynchronous request handling modeled as:

$$E_r = \frac{1}{1 + e^{-(\lambda T - \mu)}}$$

where λrepresents the request arrival rate and $\mu$is the service rate, both following a Poisson distribution typical of cloud-based systems.

Data collection was conducted through direct observation, screen recordings, and log file analysis. Behavioral responses from children during interaction with the application were classified into engagement categories based on time-series analysis of user actions. The probability of active learning interaction at time $t$is defined as:

$$P_a(t) = \frac{n_a(t)}{n_T}$$

where $n_a(t)$denotes the number of active interactions at time $t$and $n_T$the total observed actions. The cumulative engagement rate over the entire session is then obtained through integration:

$$E_T = \int_0^T P_a(t)\, dt$$

Higher values of $E_T$indicate stronger sustained attention and engagement across the session, confirming the motivational quality of the application.

The final model validation is achieved through a holistic success function Φ, representing the overall integration of technical, educational, and experiential factors:

$$\Phi = \lambda_1 f_T + \lambda_2 f_U + \lambda_3 f_A + \lambda_4 f_C$$

where $f_T$denotes technical performance, $f_U$user experience, f_AAI accuracy, and $f_C$contextual learning contribution. The coefficients $\lambda_1, \lambda_2, \lambda_3, \lambda_4$are adjustable parameters satisfying $\sum_{i=1}^{4} \lambda_i = 1$, representing the relative importance of each dimension. The objective of the prototyping process is to maximize Φunder the constraints of usability, efficiency, and pedagogical integrity:

$$\max_{\Phi} \ \Phi(T, U, A, C) \text{subject to} 0 \le L \le L_{\max}, A_r \ge A_{\min}, S_u \ge S_{\min}$$

Through this integrated methodological framework, the development process ensures that the application not only achieves technical excellence but also fulfills its educational purpose. The combination of computational modeling, system evaluation metrics, and human-centered validation provides a rigorous foundation for understanding how multimodal AI can enhance contextual environmental learning among children. The methodological rigor embedded in this study bridges the domains of software engineering, artificial intelligence, and educational psychology, establishing a reproducible framework for future AI-driven learning system development.

ISSN 2579-7069

4. Results and Discussion

The experimental implementation of the Android-based multimodal AI application successfully demonstrated its capacity to combine artificial intelligence, contextual learning theory, and user-centered design into a cohesive educational tool for children. The testing was conducted through a limited but intensive usability trial involving early learners aged between five and twelve years. The results are analyzed in three major domains: system performance, user engagement, and educational impact. Quantitative metrics, derived from automated system logs and recorded interaction data, were complemented by qualitative insights gathered through direct observation and structured parental feedback. Together, these data offer a comprehensive evaluation of the application's technical robustness, interactional fluidity, and pedagogical relevance.

The overall performance of the AI recognition module revealed consistent accuracy across most of the test cases. Out of 120 images captured during the field testing phase, 105 images were correctly classified, resulting in a recognition accuracy of 87.3%. The model's confidence levels, measured using entropy, were relatively stable, averaging 0.214, which corresponds to high certainty in AI-generated predictions. Table 1 provides a detailed breakdown of the system's core performance indicators.

Table 1. System-Level Performance Indicators

	Mean	Standard Deviation (σ)	Min	Max
Recognition Accuracy (A_r) (%)	87.3	5.4	75.0	95.0
AI Output Entropy (H)	0.214	0.032	0.179	0.289
Response Latency (L) (s)	3.92	0.64	2.73	5.18
Throughput (Θ) (req/s)	0.25	0.07	0.16	0.38
Backend CPU Utilization (%)	63.8	6.5	58.2	72.4
Request Error Rate (%)	0.72	0.43	0.28	1.43

The results show that the system maintained a steady level of computational efficiency under varying network and environmental conditions. The relatively low entropy (H) value suggests that the Gemini-based recognition model was able to consistently extract distinctive plant features, even in outdoor environments where natural lighting fluctuated significantly. The overall system latency of approximately four seconds indicates a good balance between real-time processing and model complexity, considering that the inference process required both cloud communication and multimodal reasoning.

The readability of the AI-generated descriptive narratives was analyzed to ensure linguistic appropriateness for early learners. Using the readability index $R(S)$, defined as the ratio of average word length to lexical frequency, the system achieved an average value of 0.42, which corresponds to a Grade 2–3 reading level. The lexical density was 51.2%, indicating a good balance between functional and content words. The automatic text simplification algorithm was triggered in 17.5% of outputs, effectively adapting overly complex language to children's comprehension levels. These findings are summarized in Table 2.

Table 2. Linguistic Metrics of AI-Generated Descriptions

	Mean	Range
Average Sentence Length (words)	11.8	9–14
Average Word Length (characters)	4.6	3.9–5.1
Lexical Density (%)	51.2	48–55
Readability Index (R(S))	0.42	0.37–0.48
Simplification Rate (%)	17.5	12–21

The linguistic analysis confirms that the text generation module successfully adjusted complex AI responses into child-appropriate educational narratives. The balance between informativeness and simplicity reflects an effective alignment between the generative model and the pedagogical design framework.

The user interaction data revealed high engagement levels throughout the trials. Engagement was measured in terms of session duration, frequency of photo captures, exploration repetition, and focus retention. The mean engagement time across all participants was 11.3 minutes per session, with elementary students showing longer engagement than preschoolers. The average number of photos captured per session was 4.2, and most children exhibited repeated

 ISSN 2579-7069

exploration behavior, often returning to previously recognized plants to test the system's consistency. Table 3 details these findings.

Table 3. User Engagement and Interaction Metrics

	Preschool	Elementary	Mean
Average Session Duration (min)	8.7	13.9	11.3
Images Captured per Session	3.1	5.3	4.2
Repeated Explorations per Session	1.8	3.6	2.7
Engagement Probability (P_e)	0.78	0.91	0.84
Focus Retention Time Ratio (t_e / t_t)	0.69	0.85	0.77

The engagement probability $P_e = n_e/n_T$indicated that more than 80% of interactions resulted in active learning behaviors such as capturing new plants or revisiting previous discoveries. A linear regression analysis between engagement duration and satisfaction ratings yielded a positive slope coefficient of 0.62, implying that longer engagement times correlated strongly with higher user satisfaction scores. This confirms that the application successfully motivated children to explore their natural surroundings interactively.

The composite performance score $P_s = \alpha A_r + \beta(1 - L/L_{\max}) + \gamma S_u$, with weighting parameters $\alpha = 0.4$, $\beta = 0.3$, and $\gamma = 0.3$, produced a final value of 0.86 for the latest prototype iteration. Table 4 compares the iterative improvements made over three development cycles.

Table 4. Prototype Iteration Comparison

	Iteration 1	**Iteration 2**	**Iteration 3 (Final)**
Recognition Accuracy (A_r) (%)	74.5	82.9	87.3
Response Latency (L) (s)	5.83	4.21	3.92
User Satisfaction (S_u)	0.69	0.81	0.89
Composite Performance (P_s)	0.71	0.79	0.86
AI Output Entropy (H)	0.289	0.247	0.214

Each successive iteration demonstrated marked improvements in technical efficiency and user perception. The observed decrease in entropy corresponds with greater model confidence, while reductions in latency enhanced perceived responsiveness. The consistent increase in user satisfaction further validates the design modifications such as simplified navigation and color-coded interface cues.

Beyond quantitative metrics, behavioral observations revealed strong indicators of curiosity and cognitive engagement among children. The coding of behavioral indicators based on structured observation sheets revealed that participants frequently verbalized plant names, asked reflective questions about plant functions, and displayed emotional excitement when receiving AI-generated feedback. Table 5 summarizes these behavioral data.

Table 5. Behavioral Indicators of Child Engagement

	Mean (1–5)	**Std. Dev.**
Curiosity and Inquiry	4.7	0.4
Verbalization of Observations	4.3	0.6
Peer or Parental Sharing	4.1	0.5
Emotional Excitement	4.8	0.3
Sustained Attention	4.4	0.7

The high values across all dimensions indicate that the prototype effectively stimulated intrinsic motivation and social interaction during learning. Children were observed to spontaneously share discoveries with peers or adults, demonstrating the social constructivist potential of AI-supported learning environments.

Cognitive retention was assessed by asking children to recall plant characteristics after a 24-hour interval. The mean recall accuracy reached 81%, with a standard deviation of 6.2%, suggesting strong short-term retention influenced by experiential learning. The correlation between engagement probability (P_e) and recall accuracy (K_r) was analyzed using

ISSN 2579-7069

Pearson's correlation coefficient, producing a value of $\rho = 0.81$. This strong positive relationship indicates that higher engagement levels directly contributed to better knowledge retention, consistent with the theoretical foundation of contextual learning.

Table 6. Correlation between Engagement and Learning Retention

	Variable Pair	Correlation (ρ)	Coefficient	Significance ($p < 0.05$)
Engagement Probability (P_e) vs Recall Accuracy (K_r)	0.81	Significant		
Session Duration vs Satisfaction (S_u)	0.62	Significant		
Recognition Accuracy (A_r) vs Learning Retention (K_r)	0.57	Moderate		

The results clearly demonstrate the pedagogical relevance of multimodal AI applications in supporting early environmental education. The learning process became exploratory rather than instructional, aligning with constructivist theories of cognition where understanding emerges through interaction and self-discovery. The AI's role as an adaptive mediator between children and nature appears to reinforce curiosity, enhance language development, and deepen conceptual connections.

From the technological perspective, the backend system maintained stable performance throughout all testing cycles. Table 7 illustrates the backend metrics during the active trial period, confirming the efficiency of the FastAPI-based server in handling concurrent requests from multiple devices.

Table 7. Backend Performance Summary

	Mean	Peak	Limit Threshold
CPU Utilization (%)	63.8	72.4	85.0
Memory Usage (MB)	243.1	289.6	512.0
Network Latency (ms)	87.2	104.8	150.0
Request Error Rate (%)	0.72	1.43	5.00

The backend's stability under continuous load demonstrates that the chosen architecture is scalable and reliable for extended use in educational settings with multiple simultaneous users. The relatively low memory footprint suggests that optimization strategies, including asynchronous request handling and local data caching, effectively minimized computational overhead.

Synthesizing the empirical evidence, the overall success function $\Phi = \lambda_1 f_T + \lambda_2 f_U + \lambda_3 f_A + \lambda_4 f_C$achieved a computed value of 0.88 in the final prototype. This high value reflects strong performance across all dimensions: technical reliability (f_T), user experience (f_U), AI accuracy (f_A), and contextual learning contribution (f_C). The equilibrium achieved between these factors demonstrates that the application effectively fulfills its dual objective: technological innovation and pedagogical efficacy.

The discussion of these findings indicates that integrating multimodal AI into mobile learning environments can meaningfully bridge digital and real-world education. The results reaffirm that contextual learning is most effective when mediated through interactive technologies that maintain cognitive relevance and emotional resonance. The prototype thus serves not merely as a digital tool but as a dynamic medium connecting children's curiosity to environmental exploration, fostering early ecological awareness while simultaneously introducing them to the possibilities of responsible AI interaction.

5. Conclusion

This study has presented the design, development, and evaluation of an Android-based educational application that integrates multimodal artificial intelligence to support contextual environmental learning among children. By combining the computational capacity of Google Gemini's multimodal large language model with the pedagogical foundation of contextual learning theory, the research successfully demonstrates how artificial intelligence can serve as a bridge between digital environments and real-world ecological exploration. The iterative prototyping approach

ISSN 2579-7069

ensured continuous refinement of technical functionality, user experience, and educational alignment, leading to a mature system that effectively balances accuracy, usability, and pedagogical impact.

The experimental findings provide compelling evidence that multimodal AI can be leveraged to enhance children's environmental learning experiences through interactive and meaningful engagement. Quantitative evaluation revealed high system performance with an average recognition accuracy of 87.3%, response latency of 3.92 seconds, and a composite performance score of 0.86. Linguistic analysis confirmed that the AI-generated plant descriptions were clear, simplified, and developmentally appropriate, with a readability index of 0.42, corresponding to early primary school levels. Observational data indicated strong behavioral engagement, emotional enthusiasm, and curiosity among participants, demonstrating the system's effectiveness in stimulating intrinsic motivation and fostering inquiry-based learning.

From an educational perspective, the results substantiate the theoretical premise that contextual learning—when facilitated through adaptive AI systems—can significantly enhance cognitive retention and emotional connection to environmental knowledge. The strong correlation coefficient ($\rho = 0.81$) between engagement and learning retention underscores the critical role of active exploration in reinforcing conceptual understanding. The children's ability to recall plant characteristics after using the application validates the hypothesis that technology-mediated contextual learning promotes experiential memory formation, aligning with constructivist and socio-cognitive theories of learning. The interactive use of AI thus transforms learning into a participatory process, where children act as explorers rather than passive recipients of information.

Technologically, the system demonstrates a high level of stability and efficiency. Backend load analysis showed that CPU and memory utilization remained well below critical thresholds, confirming the scalability of the FastAPI–Flutter architecture. The integration of asynchronous processing and real-time image recognition allowed smooth performance even under simultaneous user activity. The entropy-based analysis of AI output revealed low uncertainty, indicating robust plant recognition capabilities across diverse environmental conditions. These findings confirm the technical feasibility of deploying multimodal AI-driven educational tools on mobile platforms, even in non-laboratory environments.

The integration of multimodal AI also contributes to the emerging discourse on responsible and human-centered artificial intelligence in education. The findings highlight that AI systems, when properly designed, can complement rather than replace human teaching. By acting as an intelligent companion that responds to visual inputs, narrates contextually relevant explanations, and encourages exploration, the AI transforms digital screen time into a constructive, inquiry-driven experience. This aligns with current educational paradigms that emphasize personalization, inclusivity, and experiential learning through technology. Moreover, the use of visual recognition to identify real plants encourages children to move beyond passive screen interaction and re-engage with their natural surroundings, effectively blending cognitive and ecological literacy.

However, the study also acknowledges several limitations that warrant further investigation. The usability trial involved a limited number of participants, which constrains the generalizability of the findings. The current dependency on cloud-based processing introduces latency and limits accessibility in low-connectivity environments. Additionally, the AI model, while highly capable, occasionally produces minor factual inaccuracies or uses vocabulary slightly above the intended developmental level, necessitating additional language moderation layers. Ethical considerations, particularly concerning data privacy and AI transparency in applications targeting children, require continuous attention to ensure safe and responsible deployment.

Future work should address these limitations by incorporating locally stored plant databases to enable offline recognition and reduce network dependency. The inclusion of adaptive learning mechanisms could allow the system to personalize feedback based on user proficiency, learning pace, and interaction history. Gamification elements, such as progress badges, ecological missions, and storytelling-based challenges, may also increase long-term engagement and retention. Further large-scale testing across diverse age groups and cultural settings will be essential to validate the model's universality and pedagogical robustness. Expanding the system to include other domains, such as animal recognition or sustainable behavior simulations, could transform it into a comprehensive multimodal learning ecosystem for environmental education.

ISSN 2579-7069

In summary, this study contributes to both the theoretical and practical dimensions of educational technology. Theoretically, it provides empirical support for the effectiveness of contextual learning when mediated through multimodal AI. Practically, it offers a functional and scalable model for integrating large language models into mobile learning environments designed specifically for children. The results suggest that multimodal AI applications can play a transformative role in reintroducing children to the natural world through digital interfaces that are interactive, intelligent, and pedagogically sound. The combination of contextual learning, child-centered design, and multimodal artificial intelligence marks a significant advancement in how educational technology can nurture environmental awareness, digital literacy, and curiosity simultaneously.

Through this convergence of technology and pedagogy, the research underscores a fundamental paradigm shift: that artificial intelligence, when ethically developed and contextually applied, can serve as a medium of reconnection—helping children rediscover nature through the lens of technology, not as passive consumers, but as active participants in learning and environmental stewardship.

6. Declarations

6.1. Author Contributions

Author Contributions: Conceptualization, A.R.H. and M.I.R.; Methodology, A.R.H. and M.I.R.; Software, A.R.H.; Validation, M.I.R.; Formal Analysis, A.R.H.; Investigation, A.R.H. and M.I.R.; Resources, M.I.R.; Data Curation, A.R.H.; Writing—Original Draft Preparation, A.R.H.; Writing—Review and Editing, M.I.R.; Visualization, A.R.H. All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] M. Torjinski dan S. Horwood, "Associations between nature exposure, screen use, and parent–child relations: a scoping review protocol," *Syst. Rev.*, vol. 12, no. 1, art. no. 217, 2023, doi: 10.1186/s13643-023-02367-2.

[2] T. K. Oswald, A. R. Rumbold, S. G. E. Kedzior, dan V. M. Moore, "Psychological impacts of 'screen time' and 'green time' for children and adolescents: A systematic scoping review," *PLoS ONE*, vol. 15, no. 9, art. no. e0237725, 2020, doi: 10.1371/journal.pone.0237725.

[3] A.-L. Camerini, E. Albanese, L. Marciano, et al., "The impact of screen time and green time on mental health in children and adolescents during the COVID-19 pandemic," *Comput. Hum. Behav. Rep.*, vol. 7, art. no. 100204, 2022, doi: 10.1016/j.chbr.2022.100204.

[4] S. Domingues-Montanari, "Clinical and psychological effects of excessive screen time on children," *J. Paediatr. Child Health*, vol. 53, no. 4, pp. 333–338, 2017, doi: 10.1111/jpc.13462.

[5] L. Lissak, "Adverse physiological and psychological effects of screen time on children and adolescents: literature review and case study," *Environ. Res.*, vol. 164, pp. 149–157, 2018, doi: 10.1016/j.envres.2018.01.015.

[6] A. Sigman, "Time for a view on screen time," *Arch. Dis. Child*, vol. 97, no. 11, pp. 935–942, 2012, doi: 10.1136/archdischild-2012-302196.

[7] N. Stiglic dan R. M. Viner, "Effects of screentime on the health and well-being of children and adolescents: a systematic review of reviews," *BMJ Open*, vol. 9, no. 1, art. no. e023191, 2019, doi: 10.1136/bmjopen-2018-023191.

[8] Gemini Team, "Gemini: A Family of Highly Capable Multimodal Models," *arXiv preprint*, Dec. 2023, doi: 10.48550/arXiv.2312.11805.

[9] M. Imran dan N. Almusharraf, "Google Gemini as a next generation AI educational tool: a review of emerging educational technology," *Smart Learn. Environ.*, vol. 11, art. no. 22, 2024, doi: 10.1186/s40561-024-00310-z.

[10] H. Abu-Rasheed, C. Weber, dan M. Fathi, "Context based learning: a survey of contextual indicators for personalized and adaptive learning recommendations — a pedagogical and technical perspective," *Front. Educ.*, vol. 8, art. no. 1210968, 2023, doi: 10.3389/feduc.2023.1210968.

[11] L. C. Shum, J. A. Kennedy, dan V. Ifenthaler, "Personalised Learning through Context-Based Adaptation: Framework and Case Study," *Educ. Inf. Technol.*, vol. 28, pp. 943–967, 2023, doi: 10.1007/s10639-023-11695-8.

[12] G.-G. Lee, L. Shi, E. Latif, Y. Gao, et al., "Multimodality of AI for Education: Towards Artificial General Intelligence," *arXiv preprint*, Dec. 2023, doi: 10.48550/arXiv.2312.06037.

[13] S. Aslan, L. M. Durham, N. Alyuz, E. Okur, S. Sharma, C. Savur, dan L. Nachman, "Immersive multi-modal pedagogical conversational artificial intelligence for early childhood education: An exploratory case study in the wild," *Comput. Educ. Artif. Intell.*, vol. 6, art. no. 100220, 2024, doi: 10.1016/j.caeai.2024.100220.

[14] A. Hasanov, "Adaptive context-aware learning environments: a survey," *Intell. Syst. Appl.*, vol. 8, no. 4, pp. 406–425, 2019, doi: 10.3233/AIS-190534.

[15] Y. Chen, S. Xiao, Y. Song, Z. Li, L. Sun, dan L. Chen, "MindScratch: A visual programming support tool for classroom learning based on multimodal generative AI," *arXiv preprint*, Dec. 2024, doi: 10.48550/arXiv.2412.09001.

[16] S. Petridis, M. Xieyang Liu, A. J. Fiannaca, V. Tsai, M. Terry, dan C. J. Cai, "In Situ AI Prototyping: Infusing Multimodal Prompts into Mobile Settings with MobileMaker," *Proc. IEEE Symp. Visual Languages and Human-Centric Computing (VL/HCC)*, 2024, doi: 10.1109/VL/HCC60511.2024.00023.

[17] M. F. Contrino, X. Huang, H. Y. Sung, dan G.-J. Hwang, "A fuzzy expert system-based adaptive learning approach to improving students' learning performances by considering affective and cognitive factors," *Comput. Educ. Artif. Intell.*, vol. 1, art. no. 100003, 2020, doi: 10.1016/j.caeai.2020.100003.

[18] E. du Plooy dan A. van der Merwe, "Personalized adaptive learning in higher education: a conceptual review," *Heliyon Educ.*, vol. 4, no. 4, art. no. e01607, 2024, doi: 10.1016/j.heliyon.2024.e01607.

[19] T. Huang, "FEAD: Figma-Enhanced App Design Framework for Improving UI/UX in Educational App Development," *arXiv preprint*, Nov. 2024, doi: 10.48550/arXiv.2412.06793.

[20] K. Kaur, K. S. Kalid, dan K. S. Savita, "Proposed UX Model for Children Educational Mobile Application," dalam *Adv. Vis. Inform. (LNCS)*, 2019, pp. 601–610, doi: 10.1007/978-3-030-34032-2_53.

[21] M. Lazo-Amado, L. Cueva-Ruiz, dan L. Andrade-Arenas, "Prototyping a mobile application for children with dyscalculia in primary education using augmented reality," *Int. J. Adv. Comput. Sci. Appl. (IJACSA)*, vol. 13, no. 10, 2022, art. no. 85, doi: 10.14569/IJACSA.2022.0131085.