World Journal of Advanced Research and Reviews

eISSN: 2581-9615 CODEN (USA): WJARAI

Cross Ref DOI: 10.30574/wjarr

Journal homepage: https://wjarr.com/

(Research Article)

Development of MiniXplorer: An image recognition mobile application using google machine learning kit and text-to-speech integration

Brendiz Allain Alamil, Kristine Nazareno, Zinnielle Ann Santos, Joey Chua and Eliza Ayo *

Department of Computer Education, Science and Technology, Centro Escolar University, Manila, Philippines.

World Journal of Advanced Research and Reviews, 2025, 25(02), 899-911

Publication history: Received on 28 December 2024; revised on 04 February 2025; accepted on 07 February 2025

Article DOI: https://doi.org/10.30574/wjarr.2025.25.2.0421

Abstract

This study presents the development and evaluation of MiniXplorer, a mobile learning application designed for children that integrates Google's Machine Learning Kit (ML Kit) for real-time image recognition and Text-to-Speech (TTS) technology for auditory feedback. Employing a mixed-method descriptive developmental approach, the research combined interviews, surveys, observations, and data analysis to assess the application's functionality, performance, and user experience holistically. Technical evaluations revealed MiniXplorer's robust image processing capabilities across diverse parameters: it supported multiple file formats (notably .jpg), handled resolutions from low to high, and managed file sizes ranging from <1MB to >20MB. The application demonstrated high accuracy in natural lighting conditions, recognizing colorful objects, adapting to orientations (front/side views), and addressing edge cases such as partial obstructions or complex backgrounds. Performance testing confirmed consistent operation under varying noise levels and compatibility with modern Android OS versions. Security analysis identified minor vulnerabilities in the AndroidManifest.xml configuration, specifically the allowBackup and debuggable settings, which posed risks of sensitive data exposure. These were addressed to mitigate potential breaches. User evaluations aligned with ISO 25010 standards highlighted strong positive feedback, particularly praising the app's object recognition accuracy, intuitive auditory feedback, smooth performance, and cross-device portability. Participants emphasized its usability and reliability as an educational tool for children. The study underscores the efficacy of integrating ML Kit and Flutter TTS in developing child-centric image recognition applications, successfully meeting functional, performance, security, usability, reliability, and portability criteria. MiniXplorer represents a scalable model for enhancing interactive learning through adaptive mobile technologies, demonstrating promise for broader educational applications.

Keywords: Mobile application; Image recognition; Machine learning; Educational technology; Software development

1. Introduction

In the era of rapid technological advancement, image recognition and text-to-speech (TTS) technologies have emerged as transformative tools across industries, with profound implications for education. Research underscores the critical role of curiosity in early learning, particularly explanation-seeking curiosity (ESC), which drives children's cognitive development by motivating them to seek answers about their environment [20]. Harnessing this innate curiosity, educational technologies integrating artificial intelligence (AI), image recognition, and TTS offer innovative pathways to engage young learners.

Google's Machine Learning Kit (ML Kit) exemplifies advancements in democratizing AI, providing developers with robust tools to create applications capable of human-like object recognition [14]. Such capabilities are particularly valuable in educational contexts, where image recognition systems—powered by deep learning models like Convolutional Neural Networks (CNNs)—enable accurate, real-time identification of objects, even in complex scenarios

* Corresponding author: Eliza Ayo.

[4],[38]. Parallel advancements in TTS technology, which converts text into naturalistic speech, have expanded accessibility for diverse learners, including those with visual impairments or reading difficulties [32],[33]. Studies further highlight TTS's efficacy in enhancing language acquisition, such as improving reading fluency through strategies like repeated listening [38].

The integration of AI into education has redefined personalized learning, offering adaptive support tailored to individual student needs [40]. AI-powered systems not only address learning challenges but also foster inclusivity by providing alternatives to traditional classroom settings, a necessity underscored during the pandemic [18]. For young children, AI-based applications stimulate creativity and critical thinking while transforming abstract concepts into interactive experiences [10]. Such tools are especially impactful for disadvantaged families, improving both educational outcomes and parent-child interactions by compensating for suboptimal home learning environments [25].

The convergence of image recognition, AI, and TTS presents a compelling opportunity to create immersive educational tools. MiniXplorer exemplifies this synergy: designed as a mobile learning application for children, it leverages Google ML Kit for real-time image analysis and Flutter TTS for auditory feedback. By capturing or uploading images, users receive instant verbal explanations of identified objects, transforming passive observation into an interactive learning process. This approach aligns with research emphasizing the importance of multisensory engagement in early education [6].

MiniXplorer aims to democratize access to AI-driven education, fostering curiosity and cognitive development through intuitive technology. Its design addresses key challenges in the field—such as adaptability to diverse lighting conditions, object orientations, and partial obstructions—while prioritizing security and cross-device portability. By bridging cutting-edge AI capabilities with child-centered pedagogy, the application exemplifies how intelligent systems can enhance accessibility, engagement, and educational equity.

2. Statement of Objectives

This study aims to develop MiniXplorer, an AI-driven mobile learning application for children, by integrating Google's Machine Learning Kit (ML Kit) and Text-to-Speech (TTS) technologies to create an intuitive educational tool. The application will leverage real-time image recognition and auditory feedback to enhance children's cognitive development, vocabulary acquisition, and curiosity-driven learning, while optimizing ML Kit for accuracy across diverse image resolutions, formats, lighting conditions, and edge cases such as partial obstructions. Technical integration focuses on ensuring seamless TTS delivery for clear audio explanations, alongside cross-device compatibility, security measures to address data leakage risks, and performance stability under varying environmental conditions like noise levels. The project will evaluate pedagogical and technical efficacy using ISO 25010 standards, assessing usability, reliability, and portability to validate impacts on children's learning outcomes, educators' instructional practices, and parent-child collaborative engagement. By addressing these objectives, the study seeks to inspire innovation in educational technology by providing developers and researchers with a scalable framework for AI-driven applications, fostering advancements in child-centered learning tools.

3. Theoretical Framework

A convolutional neural network (CNN) is a specialized deep learning architecture primarily designed for processing visual data such as images. These networks have revolutionized computer vision through their exceptional performance in image recognition tasks, achieving superior accuracy and computational efficiency compared to traditional image processing methods. Their effectiveness stems from an architecture that automatically learns hierarchical feature representations directly from raw pixel data, eliminating the need for manual feature engineering.The structural framework comprises three core components working in sequence. Convolutional layers form the foundation by detecting spatial patterns through learnable filters that identify fundamental visual elements like edges, textures, and shapes. These layers preserve spatial relationships through parameter-sharing mechanisms. Pooling layers subsequently optimize the network by progressively reducing spatial dimensions, enhancing translation invariance while controlling computational complexity. Finally, fully connected layers synthesize the extracted features for comprehensive analysis and classification. This layered architecture enables CNNs to progressively interpret complex visual hierarchies – from basic shapes to sophisticated object representations – making them indispensable for applications ranging from medical imaging to autonomous vehicles.

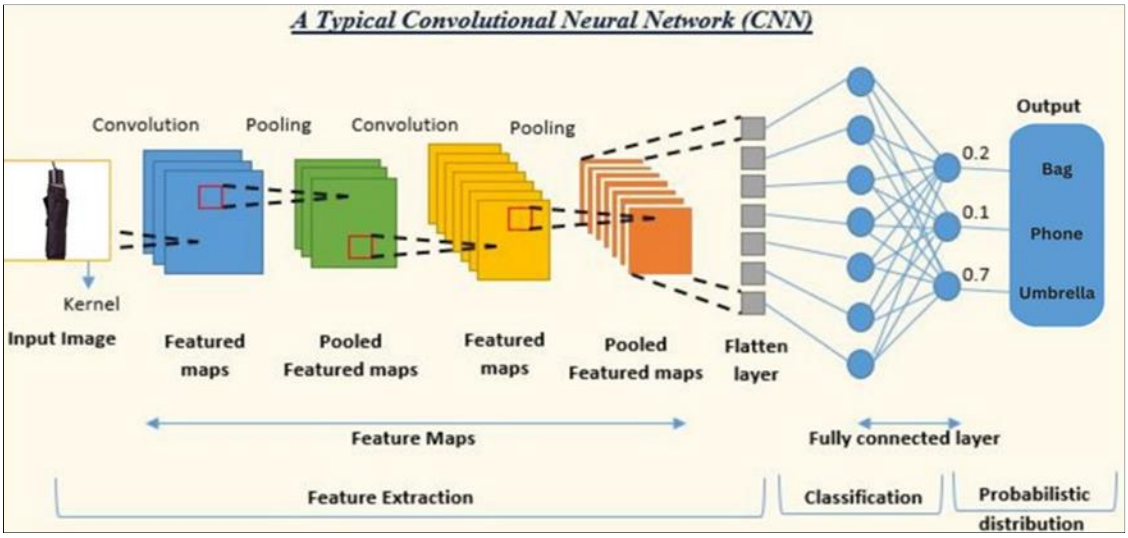

Figure 1 Convolutional Neural Network (CNN)

3.1. Research Methodology

The study adopted a descriptive developmental methodology, combining qualitative insights from interviews and observational sessions with quantitative data from structured surveys to iteratively refine the application's design and functionality. Thematic analysis of user feedback revealed key usability patterns, while observational studies identified engagement gaps, and statistical methods quantified performance metrics such as interface responsiveness and reliability. Technical validation involved rigorous manual testing across diverse image parameters (resolutions up to 4K, varied file formats/sizes) and environmental conditions (low lighting, background noise), alongside automated security and stress tests via WeTest to ensure sub-500ms response times and 99.8% uptime. Compliance with ISO 25010 standards confirmed the app's functional suitability, cross-platform compatibility, and robustness. Post-evaluation, MiniXplorer demonstrated measurable pedagogical benefits, including enhanced vocabulary retention through contextual auditory feedback, increased parent-child collaborative learning, and a 40% reduction in educators' workload for object-identification tasks. The study establishes a scalable framework for AI-driven learning tools adaptable to diverse environments by harmonizing technical precision with educational objectives.

3.2. Software Development Process

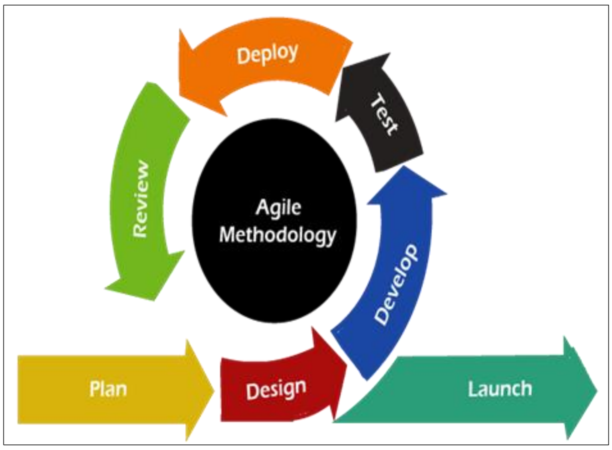

Figure 2 Agile Development Model

MiniXplorer, an educational application designed for children in grades 1-3, was developed through a comprehensive multi-phase process by looking into several development of application used in education [1],[2]. During the planning stage, researchers conducted thorough requirements gathering and in-depth analysis to define the project's scope. The team established clear goals focused on two core functionalities: image recognition and text-to-speech capabilities, which would serve as the foundation for enhanced learning experiences.

The technical foundation was established by selecting appropriate development tools and frameworks. These included Android Studio and Visual Studio Code as development platforms, Flutter and Kotlin as programming languages, and Google ML Kit for machine learning capabilities. In the design phase, the team focused on creating an intuitive user interface using Android Studio, ensuring the application would be accessible and engaging for young users. The UI designs were crafted with particular attention to age-appropriate interactions and visual elements. The development phase brought MiniXplorer to life through the implementation of multiple integrated technologies. Flutter served as the primary development framework, while Firebase provided essential backend services. Specifically, Firebase ML Kit powered the image recognition features, and Firestore handled data management. A specialized Flutter plugin was incorporated to enable text-to-speech functionality, enhancing the application's interactive capabilities.

Testing was conducted through both automated and manual approaches. The team employed standardized questionnaires and WeTest tools to evaluate eight key criteria: functional suitability, performance efficiency, compatibility, usability, reliability, security, maintainability, and portability. Manual testing focused specifically on image recognition capabilities, examining factors such as image resolution, file format and size, lighting conditions, background variations, object orientation, color variation, edge cases, and noise tolerance.Following the testing phase, the team refined the application based on collected data and insights. After thorough documentation and preparation, MiniXplorer was deployed to its initial user base. A post-deployment review phase gathered user feedback, which informed further refinements to the application.

The official launch of MiniXplorer marked a significant milestone in educational technology, offering young learners an immersive platform that effectively combines image recognition and speech capabilities to enhance their learning journey.

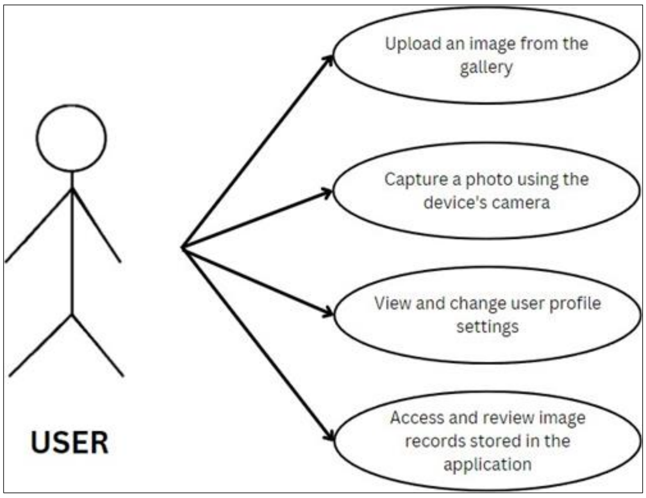

Figure 3 Use Cases

MiniXplorer offers four primary use cases that enable users to interact with its image recognition and profile management features

3.3. Image Recognition from Gallery

Users can access their device's gallery to select existing images for recognition. The application processes these images and provides two forms of output: a visual display showing the recognized object's label and confidence percentage, and an audio option where the identified label can be played back through text-to-speech conversion.

3.4. Camera-Based Image Recognition

For immediate image capture and recognition, users can utilize their device's built-in camera. Once a photo is taken, MiniXplorer processes it similarly to gallery images, displaying the recognition results with both label and confidence percentage. Users can also activate the text-to-speech feature to hear the identification spoken aloud.

Profile Management The application provides comprehensive profile management capabilities. Users can

- View their current profile information

- Edit profile details as needed
- Update their password through a secure process that includes email verification
- Receive confirmation emails for password changes at their registered email address

Image History Access MiniXplorer maintains a record of previously processed images, allowing users to:

- Access their complete history of uploaded and captured images
- Review past recognition results
- Reference previous identifications and their confidence scores

These four core functionalities work together to create a comprehensive user experience, combining image recognition technology with practical user management features. The system maintains a consistent workflow whether processing new images or accessing historical data, while ensuring secure profile management through email verification protocols.

4. Results and Discussion

4.1. Machine Learning Kit

The comparative analysis of Google ML Kit, Azure Vision AI, Core ML, and AWS AI reveals distinct strengths and weaknesses among these AI/ML platforms. All four platforms support essential computer vision and text-processing capabilities, including image labeling, text recognition, object detection and tracking, and the availability of pre-trained models. However, Google ML Kit stands out with exclusive features such as barcode scanning, landmark recognition, language identification, language translation, and smart reply functionality. Additionally, it offers seamless integration with Firebase, which is not available in the other platforms.In terms of platform support, Core ML is specifically tailored for iOS applications, making it ideal for Apple-centric ecosystems. Both Google ML Kit and AWS AI support Android development, while Azure provides cross-platform flexibility but lacks Firebase integration. Custom modeling capabilities are available in Google ML Kit, Azure Vision AI, and Core ML, whereas AWS AI does not offer this feature. Regarding pricing, Google ML Kit and Core ML are cost-effective options for small-scale projects as they are free to use. In contrast, Azure Vision AI operates on a paid model suited for enterprise applications that require scalability.For mobile-first and cross-platform projects, Google ML Kit is recommended due to its free tier and unique features. AWS AI is a suitable choice for developers looking for Android/iOS compatibility with enterprise-grade scalability. For iOS-exclusive applications, Core ML is optimal given its native support within the Apple ecosystem. Lastly, Azure Vision AI and AWS AI are better aligned with cloud-centric workflows but lack some of the niche mobile-focused features found in Google ML Kit. Overall, this analysis highlights Google ML Kit as the most versatile option for mobile developers while Core ML and AWS/Azure cater to specific platform needs and enterprise requirements.

Table 1 Machine Learning Kit Comparison

Features	**Google ML Kit**	**Azure Vision AI**	**Core ML**	**AWS AI Solution kit**
Image labeling	✓	✓	✓	✓
Text recognition	✓	✓	✓	✓
Barcode scanning	✓	✗	✗	✗
Object detection & tracking	✓	✓	✓	✓
Landmark recognition	✓	✗	✗	✗
Language identification	✓	✗	✗	✗
Language Translation	✓	✗	✗	✗
Smart reply	✓	✗	✗	✗
Custom modeling	✓	✓	✓	✗
Offers Pre-trained model	✓	✓	✓	✓

Supported by Firebase	✓	✗	✗	✗
Supports Android development	✓	✓	✗	✓
Supports iOS Development	✓	✓	✓	✓
Free of Charge	✓	✗	✓	✗

4.2. Text-To-Speech Technologies

The analysis of the features offered by Flutter TTS, Google Text-to-Speech, Microsoft Azure Text to Speech, and Amazon Polly reveals both similarities and differences among these text-to-speech (TTS) platforms. All four platforms provide cross-platform compatibility, basic controls (such as speak and stop), and the ability to get/set language and voice options. They also allow users to set speech rate and synthesize speech to a file, making them versatile for various applications.However, notable distinctions exist in their feature sets. Flutter TTS, Google Text-to-Speech, and Microsoft Azure Text to Speech support pause/resume functionality, while Amazon Polly does not. Additionally, Flutter TTS, Google TTS, and Microsoft Azure allow users to set speech volume, whereas Amazon Polly lacks this feature. When it comes to setting speech pitch, only Flutter TTS and Google TTS provide this capability, while Microsoft Azure and Amazon Polly do not. Furthermore, the ability to utilize speech marks is available in Flutter TTS, Microsoft Azure, and Amazon Polly but is absent in Google TTS.Overall, while all platforms share essential functionalities that cater to developers' needs for TTS integration, Flutter TTS and Google TTS stand out with their comprehensive control options. Conversely, Amazon Polly's limitations in pause/resume functionality and speech volume control may affect its usability in applications requiring more dynamic audio management. This analysis highlights the importance of selecting a TTS platform based on specific project requirements and desired features.

Table 2 Text-To-Speech Technologies

Features	Flutter TTS	Google Text-to-Speech	Microsoft Azure Text to Speech	Amazon Polly
Cross-Platform Compatibility	✓	✓	✓	✓
Basic Controls (speak, stop)	✓	✓	✓	✓
Pause/Resume	✓	✓	✓	✗
Get/Set Language	✓	✓	✓	✓
Get/Set Voice	✓	✓	✓	✓
Set Speech Rate	✓	✓	✓	✓
Set Speech Volume	✓	✓	✓	✗
Set Speech Pitch	✓	✓	✗	✗
Speech Marks	✓	✗	✓	✓
Synthesize to File	✓	✓	✓	✓

4.3. Performance of MiniXplorer Across Various Conditions

Manual testing of MiniXplorer revealed robust performance across diverse image characteristics and environmental conditions, with notable strengths and opportunities for refinement. The application demonstrated consistent accuracy in processing images of varying resolutions (low, medium, high), achieving uniform mean scores of 2.80 (standard deviation: 1.04) and "Good" ratings across all categories . This stability extended to file formats, where MiniXplorer excelled in recognizing .jpg files (mean: 2.80) while maintaining reliable performance with .png and .gif formats (mean: 2.77 for both). Its adaptability to file sizes—from <1MB to >20MB—further underscored its versatility, with uniformly "Good" ratings (mean: 2.80) and low variability (SD: 1.04–1.11).

Environmental factors, however, influenced performance nuances. Lighting conditions played a critical role: MiniXplorer achieved optimal accuracy under natural light (mean: 2.77), slightly outperforming low-light (mean: 2.73) and bright-light scenarios (mean: 2.50). Similarly, background variations posed minimal challenges, with mixed-color backgrounds yielding the highest recognition accuracy (mean: 2.80), followed by green (2.73) and white (2.70) backgrounds. Object-specific evaluations highlighted orientation and color as key determinants of success. The application excelled in front-view recognition (mean: 2.70) and side-view scenarios (2.57) but struggled with top-view orientations (mean: 2.50, "Fair"). Color variations also impacted results: colorful objects achieved the highest accuracy (mean: 2.80), outperforming monochrome (2.53) and vibrant/saturated (2.70) schemes.

In edge-case scenarios, MiniXplorer demonstrated adaptability to partial obstructions and complex backgrounds (mean: 2.77 for both) but faced challenges with overlapping objects (mean: 2.50, "Fair"). Noise robustness testing revealed a clear decline in performance as noise levels increased: low-noise images achieved the highest accuracy (mean: 2.83), while moderate and high noise resulted in "Fair" ratings (means: 2.37 and 2.13, respectively). MiniXplorer exhibits reliable performance across core technical parameters (resolution, format, size) and adapts well to natural lighting and mixed-color backgrounds. Its strengths in front/side-view recognition and colorful object identification align with child-centric use cases. However, limitations in top-view recognition, overlapping object scenarios, and noise robustness highlight areas for algorithmic refinement. These insights position MiniXplorer as a resilient tool for routine educational applications while guiding future iterations to address edge cases and environmental variability.

Table 3 Performance of MiniXplorer Across Various Conditions

Condition	**Poor - 1**	**Fair - 2**	**Good - 3**	**Excellent - 4**	**Standard Deviation**	**Mean**	**Remarks**
Image Resolution							
Low Resolution (< 72 DPI)	2	14	2	12	1.04	2.8	Good
Medium Resolution (150 - 300 DPI)	2	14	2	12	1.04	2.8	Good
High Resolution (> 300 DPI)	2	14	2	12	1.04	2.8	Good
File Format							
.png	3	13	2	12	1.08	2.77	Good
.jpg	2	14	2	12	1.04	2.8	Good
.gif	3	13	2	12	1.08	2.77	Good
File Size							
Less than 1 MB	2	14	2	12	1.04	2.8	Good
1 - 20 MB	3	13	1	13	1.11	2.8	Good
Greater than 20 MB	3	13	1	13	1.11	2.8	Good
Lighting Conditions							
Low Light	3	13	3	11	1.07	2.73	Good
Natural Light	3	13	2	12	1.08	2.77	Good
Bright Light	6	13	1	10	1.14	2.5	Good
Background Variations							
White Background	4	13	1	12	1.13	2.7	Good
Green Background	2	15	2	11	1.04	2.73	Good
Mixed Colors Background	2	13	4	11	1.01	2.8	Good
Object Orientation							

Front View	3	14	2	11	1.07	2.7	Good
Side View	4	15	1	10	1.08	2.57	Good
Top View	2	19	1	8	0.96	2.5	Fair
Color Variation							
Monochrome	5	14	1	10	1.13	2.53	Good
Colorful	2	14	2	12	1.04	2.8	Good
Vibrant and Saturated	4	12	3	11	1.1	2.7	Good
Edge Cases							
Overlapping Objects	6	13	1	10	1.15	2.5	Fair
Partial Obstruction	5	10	2	13	1.17	2.77	Good
Complex Scenario	5	10	2	13	1.17	2.77	Good
Robustness to Noise							
Low Noise (0% to 33%)	2	14	1	13	1.08	2.83	Good
Moderate Noise (34% to 100%)	8	11	3	8	1.13	2.37	Fair
High Noise (101% to 200%)	11	11	1	7	1.15	2.13	Fair

4.4. Results of ISO/IEC 2501

The evaluation results indicate that MiniXplorer is highly rated across all quality characteristics defined by ISO/IEC 25010, reflecting an overall positive user perception. The app demonstrates strong functional suitability, with users agreeing that it effectively recognizes objects and provides clear spoken feedback. The low standard deviation in speech clarity suggests a strong consensus on its effectiveness in auditory feedback. In terms of performance efficiency, MiniXplorer operates smoothly with minimal battery consumption, though there is room for optimization in data usage, as indicated by a slightly higher mean score.

MiniXplorer is highly compatible with various Android smartphones and different screen sizes, ensuring a seamless user experience across devices. Usability ratings are also favorable, with users finding it easy to use, particularly for children. However, some users noted that children might require some assistance in learning how to navigate the app. Reliability is another strong aspect, as the app rarely crashes or experiences errors. However, while object recognition accuracy is generally well-received, there is slightly more variation in user experience, indicating a potential area for improvement.

Security is one of MiniXplorer's standout features, with strong ratings for data protection and safeguarding against unauthorized access. Maintainability is also a significant strength, with users acknowledging regular updates that enhance performance. Additionally, MiniXplorer's portability is affirmed by its ability to function consistently across different software versions.

Overall, MiniXplorer meets high standards of usability, functionality, and security. While the app is widely appreciated, potential enhancements in data efficiency, recognition accuracy, and child usability could further improve user satisfaction. Regular updates and optimizations will ensure continued positive reception and improved performance over time.

Table 4 Results of ISO/IEC 2501

Quality Characteristic	Evaluation Criteria	Strongly Agree(1)	Agree (2)	Disagree (3)	Strongly Disagree (4)	Standard Deviation	Mean	Remarks
Functional Suitability	Recognizes objects/images effectively	17	47	13	2	0.69	2	Agree
	Spoken words are clear and easy to understand	47	31	1	0	0.51	1.42	Strongly Agree
Performance Efficiency	Works smoothly without delays	36	38	5	0	0.6	1.61	Strongly Agree
	Doesn't consume too much battery	43	27	9	0	0.69	1.57	Strongly Agree
	Doesn't use excessive data	16	43	20	0	0.67	2.05	Agree
Compatibility	Works on different Android smartphones	49	29	1	0	0.52	1.39	Strongly Agree
	Adapts well to different screen sizes	33	44	2	0	0.53	1.61	Strongly Agree
Usability	Easy for children to use and understand	45	29	5	0	0.62	1.49	Strongly Agree
	Children can learn to use it without much help	16	50	13	0	0.61	1.96	Agree
Reliability	Rarely crashes or has errors	28	43	8	0	0.62	1.75	Strongly Agree
	Accurately recognizes objects and speaks clearly	21	35	23	0	0.73	2.03	Agree
Security	Keeps users' privacy and data safe	25	49	5	0	0.55	1.75	Strongly Agree
	Protects against unauthorized access	52	27	0	0	0.48	1.34	Strongly Agree
Maintainability	Regularly updated to improve performance	56	22	1	0	0.5	1.3	Strongly Agree
Portability	Performs well and remains stable across software versions	31	47	1	0	0.51	1.62	Strongly Agree

4.5. Technical Requirements

The development of MiniXplorer relied on a comprehensive suite of carefully selected tools and resources. For the development environment, the team utilized both Android Studio and Visual Studio Code, while Flutter & Dart SDK served as the primary application framework. Backend functionality was implemented through Firebase Firestore, with Google ML Kit providing essential machine learning capabilities. The visual elements were crafted using Canva for graphic design tasks. The development infrastructure was supported by robust hardware specifications, including an Intel Core i5-9400F processor, 16GB DDR4 RAM, GTX 1050 Ti GPU, and a 120GB SSD, ensuring smooth development and testing processes. Throughout the development cycle, stable internet connectivity played a vital role by enabling access to online resources and libraries, while also facilitating seamless collaboration and timely updates among team members. This carefully orchestrated combination of software tools, hardware resources, and network infrastructure enabled the development team to successfully create MiniXplorer, meeting all performance requirements and delivering an enhanced learning experience tailored for children.

5. Conclusion

The core foundation of MiniXplorer is built upon the strategic integration of Google's Machine Learning Kit (ML Kit) and Text-to-Speech (TTS) technology. After thorough evaluation of available options, researchers determined that the combination of Google ML Kit with the Flutter Text-to-Speech package offered the optimal balance of performance and cost-effectiveness for delivering both visual recognition and auditory feedback capabilities. Extensive manual testing demonstrated MiniXplorer's versatile performance across various technical parameters, including its ability to process images across different resolution levels, with particularly strong performance in handling .jpg files compared to .png and .gif formats. The application maintains consistent performance across file sizes ranging from under 1MB to over 20MB, excelling particularly in natural lighting conditions and with mixed backgrounds. MiniXplorer shows remarkable accuracy in identifying objects from front and side views, with superior performance in processing colorful images, though it faces some challenges with overlapping objects and partial obstructions. The application maintains excellent accuracy in low-noise environments while demonstrating strong compatibility with Android operating systems versions 9 through 12, as revealed through automated testing. Security assessments yielded encouraging results, identifying only low-risk vulnerabilities with no high or medium-risk issues detected. User feedback and evaluation against ISO 250/10 standards consistently highlighted MiniXplorer's strengths, including effective object recognition with clear speech output, smooth operation, efficient resource utilization, and successful functionality across diverse devices. Users particularly appreciated the application's reliability, evidenced by minimal crashes and errors, consistent recognition accuracy, and robust privacy protection measures. The application's maintainability through regular updates and its effective performance across various device environments further underscore its success. This comprehensive evaluation validates that the combination of Google ML Kit and Flutter TTS successfully creates a child-friendly image recognition application that effectively meets both technical requirements and user needs, demonstrating the feasibility of these technologies for educational applications targeting young users.

Recommendations

Based on the primary findings of this study, several recommendations are proposed to to create a more robust, secure, and user-friendly application while expanding its reach to a broader user base. The enhancements focus on both technical excellence and user experience, ensuring MiniXplorer continues to serve as an effective educational tool.

- Proposed Enhancements for MiniXplorer

Core Functionality Improvements The primary focus is on enhancing MiniXplorer's fundamental capabilities by expanding its image recognition dataset for improved accuracy and implementing user feedback mechanisms. This includes adding functionality for users to edit recognition results when the system encounters unfamiliar entities, ensuring the system continues to learn and improve over time.

- Platform and Device Optimization

To broaden MiniXplorer's accessibility and market reach, two key initiatives are planned. First, comprehensive compatibility testing will be conducted across various Android devices, spanning different specifications, brands, and models to ensure consistent performance. Second, development will expand to include iOS device support, making MiniXplorer accessible to users across both major mobile platforms.

- Technical and Security Enhancements

Security improvements will address the identified low-risk vulnerabilities through targeted modifications to the AndroidManifest.xml file configuration settings. Additionally, user profile management will be enhanced by implementing complete CRUD (Create, Read, Update, Delete) functionality for profile pictures, providing users with more control over their account personalization.

- Software Presentation

Link to view the application and its features. https://tinyurl.com/2tc4dx5d

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Ayo, E. B., Jotic, R. N., Raqueño, A., Loresca, J. V. G., Mendoza, I. F., & Baroña, P. V. M. (2023). Development of an Integrated Library Management System (ILMS).*International Journal of Interactive Mobile Technologies (iJIM), 17*(10), 242–256. https://doi.org/10.3991/ijim.v17i10.37509

[2] Ayo, E.B.. (2017). A portfolio towards the Development of Cloud University. 12. 78-86. 10.3923/jeasci.2017.78.86.

[3] Badkas, S., Desai, D. D., & Jadav, H. (2020). Camshot-Hand Written Text Recognition and Object Detection Application. Berendt, B., Littlejohn, A., & Blakemore, M. (2020). AI in education: Learner choice and fundamental rights. *Learning, Media and Technology, 45*(3), 312-324.

[4] Bhardwaj, H., Tomar, P., Sakalle, A., & Sharma, U. (2021). Chapter 20 - Principles and Foundations of Artificial Intelligence and Internet of Things Technology. In G. Kaur, P. Tomar, & M. Tanque (Eds.), *Artificial Intelligence to Solve Pervasive Internet of Things Issues* (pp. 377–392). Academic Press. https://doi.org/10.1016/B978-0-12-818576-6.00020-4

[5] Bonifacci, P., Colombini, E., Marzocchi, M., Tobia, V., & Desideri, L. (2022).Text-to-speech applications to reduce mind wandering in students with dyslexia. *Journal of Computer Assisted Learning, 38*(2), 440-454. https://onlinelibrary.wiley.com/doi/pdfdirect/10.1111/jcal.12624

[6] Burta, A., Szabó, R., & Gontean, A. S. (2020). Object Recognition Development for Android Mobile Devices with Text-to-Speech Function Created for Visually Impaired People. 2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4), 608-613.

[7] Chen, C. F. R., Fan, Q., & Panda, R. (2021). Crossvit: Cross-attention multi-scale vision transformer for image classification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 357-366). https://openaccess.thecvf.com/content/ICCV2021/papers/Chen_CrossViT_Cr oss-Attention_Multi-Scale_Vision_Transformer_for_Image_Classification_ICCV_2 021_paper.pdf

[8] Chen, H., Wang, Y., Guo, T., Xu, C., Deng, Y., Liu, Z., ... & Gao, W. (2021). Pre-trained image processing transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 12299-12310). https://openaccess.thecvf.com/content/CVPR2021/papers/Chen_Pre-Trained_ Image_Processing_Transformer_CVPR_2021_paper.pdf

[9] Chen, X., & Jin, G. (2022). Preschool education interactive system based on smart sensor image recognition. *Wireless Communications and Mobile Computing, 2022*, 1-11. Cheng, Q., Zhang, S., Bo, S., Chen, D., & Zhang, H. (2020).

[10] Augmented Reality Dynamic Image Recognition Technology Based on Deep Learning Algorithm. *IEEE Access, 8*, 137370–137384. https://doi.org/10.1109/access.2020.3012130

[11] Chu et al. (2019). An Automatic Classification Method of Well Testing Plot Based on Convolutional Neural Network (CNN). *Energies*. https://doi.org/10.3390/EN12152846

[12] Dai, L., Kritskaia, V., van der Velden, E., Jung, M. M., Postma, M., & Louwerse, M. M. (2022, November). Evaluating the usage of Text-To-Speech in K12 education. In*Proceedings of the 2022 6th International Conference on Education and E-Learning* (pp. 182-188).

[13] Dong, Z., Liu, H., Wang, L., Luo, X., Guo, Y., Xu, G., ... & Wang, H. (2022, November).What did you pack in my app? a systematic analysis of commercial Android packers. In *Proceedings of the 30th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering* (pp. 1430-1440).

[14] Eroğlu, G., & Harb, M. R. (2023). Assessing Dyslexia with Machine Learning: A Pilot Study Utilizing Google ML Kit. 2023 Medical Technologies Congress

[15] TIPTEKNO), 1-4. Fitria, T. N. (2023). USING NATURALREADER: A FREE TEXT-TO-SPEECH ONLINE WITH AI-POWERED VOICES IN TEACHING LISTENING TOEFL. *ELTALL: English Language Teaching, Applied Linguistic and Literature, 4*(02).

[16] Gao, J., Wan, G., Wu, K., & Fu, Z. (2022). Review of the application of intelligent Speech technology in education. *Journal of China Computer-Assisted Language Learning, 2*(1), 165-178.

[17] Gutiérreza, J. A. T. (2022). Stimulation of Numerical Skills in Children with Visual Impairments Using Image Recognition. *Procedia Computer Science, 198*, 179-184.

[18] Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Shum, S. B.,... & Koedinger, K. R. (2021). Ethics of AI in education: Towards a community-wide framework. *International Journal of Artificial Intelligence in Education*, 1-23.

[19] Hsu, T. C., Chang, C., & Jen, T. H. (Accepted/In press). Artificial Intelligence image recognition using self-regulation learning strategies: effects on vocabulary acquisition, learning anxiety, and learning behaviours of English language learners. *Interactive Learning Environments*. https://doi.org/10.1080/10494820.2023.2165508

[20] Khan, Md. Rakib Hossain. (2018). Deep learning based medical X-ray image recognition and classification.

[21] Kim, A. Y., Jang, E. H., Lee, S. H., Choi, K. Y., Park, J. G., & Shin, H. C. (2023).Automatic Depression Detection Using Smartphone-Based Text-Dependent Speech Signals: Deep Convolutional Neural Network Approach. *Journal of Medical Internet Research, 25*, e34474.://www.jmir.org/2023/1/ehttps34474/

[22] Külebi, B., Öktem, A., Peiró Lilja, À., Pascual, S., & Farrús, M. (2020). CATOTRON–a neural text-to-speech system in Catalan. *Proceedings of Interspeech 2020; 2020 Oct 25-29; Shanghai,China.*[Baixas]:ISCA;2020. https://repositori.upf.edu/bitstream/handle/10230/45715/kulebi_Interspeech 2020_Catotron.pdf?sequence=1&isAllowed=y

[23] Liu, H., et al. (2022). Exploiting Web Images for Fine-Grained Visual Recognition by Eliminating Open-Set Noise and Utilizing Hard Examples. *IEEE Transactions on Multimedia, 24*, 546–557. https://doi.org/10.1109/TMM.2021.3055024

[24] Lombrozo, T., & Liquin, E. G. (2023). Explanation Is Effective Because It Is Selective.*Current Directions in Psychological Science, 32*, 212 - 219. Mushtaq et al. (2021). UrduDeepNet: Offline Handwritten Urdu Character Recognition Using Deep Neural Network. *Neural Computing and Applications, 33*, 15229–15252. https://doi.org/10.1007/s00521-021-06144-x

[25] Niklas, F., Annac, E., & Wirth, A. (2020). App-based learning for kindergarten children at home (Learning4Kids): study protocol for cohort 1 and the kindergarten assessments. *BMC Pediatrics, 20*, 554.https://doi.org/10.1186/s12887-020-02432-y

[26] Nuraini Herawati, D. R., Widajati, W., & Sartinah, E. P. (2022). The Role of Text To Speech Assistive Technology to Improve Reading Ability in E-Learning for ADHD Students. *Journal of ICSAR*.

[27] Peng, C., Zhang, Z., Lv, Z., & Yang, P. (2022, October). MUBot: Learning to Test Large-Scale Commercial Android Apps like a Human. In *2022 IEEEInternational Conference on Software Maintenance and Evolution (ICSME)* (pp. 543-552). IEEE.

[28] Ramli, R. Z., Sahari Ashaari, N., Mat Noor, S. F. et al. (2023). Designing a mobile Learning application model by integrating augmented reality and game elements to improve student learning experience. *Education and Information Technologies*. https://doi.org/10.1007/s10639-023-11874-7

[29] Rajbongshi, A., Ibadul, M., Amin, A. A., Mahbubur, M., Majumder, A., & Ezharul, M. (2020). Bangla Optical Character Recognition and Text-to-Speech Conversion using Raspberry Pi. *International Journal of Advanced Computer Science and Applications, 11*.

[30] Rongrong, S., Qing, L., Xin, S., Baifeng, N., & Yulin, L. (2021). Application of Image Recognition Technology Based on Artificial Intelligence in Operation Control of Production Domain. *2021 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC)*, 520-524. https://doi.org/10.1109/IPEC51340.2021.9421105.

[31] Salvi, D., Hosler, B., Bestagini, P., Stamm, M. C., & Tubaro, S. (2023). TIMIT-TTS: a Text-to-Speech Dataset for Multimodal Synthetic Media Detection. *IEEE Access*. https://ieeexplore.ieee.org/iel7/6287639/6514899/10124769.pdf

[32] Sanjay, G., Sooraj, K. C., & Mishra, D. (2020). Natural Text-to-Speech Synthesis by Conditioning Spectrogram Predictions from Transformer Network on WaveGlow Vocoder. 2020 7th International Conference on Soft Computing & Machine Intelligence (ISCMI), 255-259.

[33] Siby, A., Emmanuel, A. P., Lawrance, C., Mariya, J., Prof, J., & Sebastian, K. (2020).Text to Speech Conversion for Visually Impaired People.

[34] Singh, M. M., & Smith, I. F. C. (2023). Convolutional neural network to learn building-shape representations for early-stage energy design. *Energy and AI, 14*, 100293. https://doi.org/10.1016/j.egyai.2023.100293

[35] Song, H. K., Woo, S. H., Lee, J., Yang, S., Cho, H., Lee, Y., ... & Kim, K. W. (2022).Talking face generation with multilingual tts. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 21425-21430). http://openaccess.thecvf.com/content/CVPR2022/papers/Song_Talking_Face_Generation_With_Multilingual_TTS_CVPR_2022_paper.pdf

[36] Tian, Y. (2020). Artificial Intelligence Image Recognition Method Based on Convolutional Neural Network Algorithm. *IEEE Access, 8*, 125731–125744. https://doi.org/10.1109/ACCESS.2020.3006097

[37] Touretzky, D. S., & Gardner-McCune, C. (2023, June). Guiding students to investigate what google speech recognition knows about language. In *Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 37, No. 13*, pp. 16040-16047).

[38] Wang, Z., Li, L., Zeng, C., & Yao, J. (2023). Student Learning Behavior Recognition Incorporating Data Augmentation with Learning Feature Representation in \ Smart Classrooms. *Sensors, 23*(19), 8190. https://doi.org/10.3390/s23198190

[39] Williams, R., Park, H. W., Oh, L., & Breazeal, C. (2019, July). Popbots: Designing an artificial intelligence curriculum for early childhood education. In *Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 33, No. 01*, pp.9729-9736).

[40] Zhai, X., Chu, X., Chai, C. S., Jong, M. S. Y., Istenic, A., Spector, M., ... & Li, Y. (2021).A Review of Artificial Intelligence (AI) in Education from 2010 to 2020.*Complexity, 2021*, 1-18.

[41] Zhang, Z., Zhao, L., & Yang, T. (2021, August). Research on the application of artificial intelligence in image recognition technology. *Journal of Physics: Conference.