IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.67
Volume 5, Issue 7, March 2025

LiveVision: Real Time Object Recognition and Speech Synthesis using Machine Learning

Ms. Likita S. Shetty[1], Mr. AshrayR. Naik[2], Mr. ManjuramG. Prabhudessai[3]
Assistant Professor, Department of Computer Science, Government College, Khandola, Marcela, Goa, India [1]
Student, Department of Computer Science and Technology, Goa Business School, Goa University, Goa, India [2]
Student, Department of Computer Science and Technology, Goa Business School, Goa University, Goa, India[3]

Abstract: *In the realm of education and accessibility, a significant gap exists for school-aged children as well as for visually impaired individuals, who often face challenges in accessing interactive and inclusive learning tools. Existing educational applications have yet to fully address the unique needs of these diverse user groups, particularly in the areas of real-time object recognition and speech synthesis. This paper presents LiveVision, an innovative solution designed to bridge these gaps. LiveVision integrates real-time object recognition with speech synthesis using Machine Learning to provide an engaging and accessible educational experience for children, while simultaneously offering a vital tool for enhancing the independence and learning of visually impaired individuals. By leveraging cutting-edge technologies, LiveVision empowers children with interactive learning opportunities, fosters a more inclusive environment, and promotes equal access to education for those with visual impairments. This research explores the design, development, and potential impact of LiveVision, aiming to break barriers, enhance learning outcomes, and create an inclusive educational landscape for all.*

Keywords: Machine Learning, LiveVision, Speech Synthesis, TensorFlow

I. INTRODUCTION

In an era where technology has the power to transform lives, there remains a critical gap in accessible educational tools for school-aged childrenand for visually impaired individuals. Traditional learning applications often fail to accommodate the specific needs of these groups, particularly in providing the interactive, engaging experiences that foster both intellectual growth and environmental awareness. Recognizing the transformative potential of technology, *LiveVision* emerges as a ground-breaking software application designed to bridge this gap by seamlessly integrating real-time object recognition and speech synthesis technologies using Machine Learning.

LiveVision is meticulously crafted with dual objectives: to provide an immersive, interactive educational environment for children and to enhance the real-time environmental awareness of visually impaired individuals. At its core, the application leverages advanced algorithms for instantaneous object recognition paired with high-quality speech synthesis, enabling users to both see and hear their surroundings in an innovative and meaningful way. For children, *LiveVision* creates an engaging, hands-on learning experience that aligns with their developmental needs, while for the visually impaired; it offers a tool that enhances their independence by providing them with immediate auditory feedback about their environment.

The significance of *LiveVision* extends beyond its functionality as an educational tool; it represents a fundamental shift toward inclusivity and equal access to knowledge. By offering an intuitive user interface and powerful real-time features, *LiveVision* empowers its users—particularly the visually impaired—with unprecedented independence. In doing so, it addresses a major challenge in both the educational and accessibility sectors: how to create environments that are not only interactive but also universally accessible.

Through this paper, we explore the design, development, and potential impact of *LiveVision*, analyzing its role in reshaping the landscape of educational technology. By embracing the convergence of real-time object recognition and speech synthesis, *LiveVision* paves the way for a new era of inclusive learning, where technology acts as an equalizer and ensures that both education and accessibility are not just ideals but attainable realities for all.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.67
Volume 5, Issue 7, March 2025

II. PROBLEM STATEMENT

The school-aged children, as well as visually impaired individuals, encounter distinct yet interconnected challenges in accessing appropriate and inclusive educational tools. Traditional applications, which primarily focus on either interactive learning for children or accessibility for the visually impaired, fail to address the comprehensive needs of these user groups. For visually impaired individuals, the absence of a tool that provides real-time object recognition and spoken descriptions creates significant barriers to both navigating and understanding their environment, limiting their independence and access to vital information. Similarly, for children, particularly those in the early stages of learning, the lack of engaging, interactive educational tools that are tailored to their developmental needs restricts their ability to fully immerse in the learning process, impeding their cognitive and educational growth.

Currently, there is a noticeable gap in the development of educational applications that integrate both real-time object recognition and speech synthesis in a way that serves both educational and accessibility purposes simultaneously. This limitation results in missed opportunities to create a more inclusive and dynamic learning environment that addresses the needs of diverse user groups.

LiveVision seeks to address these challenges by providing an innovative solution that combines advanced technologies—real-time object recognition and speech synthesis using machine learning—to create an accessible and engaging learning platform for children, while also empowering visually impaired individuals with tools that enhance their real-time environmental awareness. The problem statement highlights the need for a holistic, adaptive solution that seamlessly integrates education and accessibility, making it a vital contribution to the advancement of both sectors. *LiveVision*aims to bridge these gaps, providing a much-needed tool that fosters inclusivity and equal opportunities for learning and navigation.

III. AIMS AND OBJECTIVES

The primary aim of *LiveVision* is to revolutionize educational experiences for the school-aged children, as well as to empower visually impaired individuals by seamlessly integrating real-time object recognition and speech synthesis. The specific objectives for LiveVision include:

- **User-Friendly Interface:** Develop an intuitive and accessible interface to cater to users with different levels of abilities, ensuring ease of use for both children and visually impaired individuals.
- **Advanced Object Recognition:** Implement state-of-the-art algorithms for real-time object recognition, allowing *LiveVision* to accurately identify and interpret a variety of objects.
- **High-Quality Speech Synthesis:** Integrate advanced speech synthesis technologies to provide clear and concise spoken descriptions of recognized objects, enhancing the user experience.
- **Customization Features:** Enable users to personalize *LiveVision* according to their specific needs, such as adjusting the speed of object recognition, selecting different languages, and accommodating various educational preferences.

IV. SCOPE OF LIVEVISION APPLICATION

The objective of *LiveVision* is to create an Android application that caters to the educational needs of children and enhances accessibility for visually impaired individuals. The scope includes:

- **Educational Modules:** Incorporate interactive learning modules, quizzes, and challenges for the school-aged children leveraging real-time object recognition to make learning engaging and inclusive.
- **Real-Time Object Recognition:** Utilize trained machine learning models to enable *LiveVision* to recognize objects in real time, providing valuable information to visually impaired users about their surroundings.
- **Camera Integration:** Implement camera integration for LiveVision to capture images and perform real-time object recognition, enhancing the application's usability.
- **Accessibility Features:** Focus on making *LiveVision* accessible to individuals who are blind or visually impaired, promoting independence through accurate object recognition and clear spoken descriptions.
- **Deliverables:** Provide the Android application, the trained machine learning model for object recognition, and thorough documentation as tangible outcomes of the project.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.67
Volume 5, Issue 7, March 2025

V. FEATURES

The features of the LiveVision App are as follows:

- **Real-Time Object Recognition**: *LiveVision* employs advanced algorithms for instantaneous and accurate identification of diverse objects in real time, enhancing the educational experience for children and providing essential information for visually impaired users about their surroundings.
- **Speech Synthesis Excellence**: Featuring high-quality speech synthesis technology, *LiveVision* translates recognized objects into clear and articulate spoken descriptions, fostering an immersive learning environment and offering real-time environmental awareness for the visually impaired.
- **User-Friendly Interface:** *LiveVision* prioritizes accessibility with an intuitive interface for users of all ages, including the school-aged childrenand visually impaired individuals. The design focuses on inclusivity and ease of use.
- **Versatility across Environments:** *LiveVision* adapts seamlessly to various environments, making it suitable for educational settings, public spaces, and workplaces. Its versatility ensures effectiveness in different scenarios.
- **Accessibility for the Visually Impaired:** *LiveVision* prioritizes accessibility by offering real-time object recognition and spoken descriptions. Visually impaired users gain independence and awareness, promoting a more inclusive experience.
- **Comprehensive Documentation:** *LiveVision* is accompanied by detailed documentation, providing insights into its architecture, design, and implementation. This enhances transparency for users, developers, and educators.
- **Compatibility across Devices:** Designed for compatibility, *LiveVision* integrates with various devices, ensuring accessibility across platforms. The application is optimized for mobile devices and other technologies.

VI. METHODOLOGY

An interactive method/approach is used to build a model for *LiveVision* App and it had to go through the following steps:

- **Data Collection:** A diverse dataset was curated, encompassing images relevant to Real-Time Object Detection and age-appropriate educational content.
- **Data Pre-processing:** Standardizing the dataset involved resizing images and normalizing pixel values, preparing it for training, validation, and testing.
- **Model Architecture:** A comprehensive architecture integrating Real-Time Object Detection and Speech Synthesis was designed using YOLO V4 and TensorFlow.
- **Model Training:** Rigorous training of models involved optimization techniques, loss functions, and metrics specific to Real-Time Object Detection and Speech Synthesis.
- **Model Evaluation:** Thorough evaluation of separate testing datasets assessed performance, accuracy, and generalization.
- **Model Deployment:** Models were converted to TensorFlowLite for efficient deployment on Android devices, ensuring seamless integration.
- **Performance Analysis**: In-depth analysis, including accuracy metrics, provided insights into model effectiveness.

VII. BENEFITS

The *LiveVision* app offers a range of benefits that makes it a convenient and effective way to detect an object from live environment. Some of them are:

- **Enhanced Communication:** Facilitates improved communication for blind individuals, minimizing misunderstandings.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.67
Volume 5, Issue 7, March 2025

- **Cost-Effective Solution:** Provides a cost-effective alternative to traditional interpretation services, enhancing accessibility.
- **Educational Tool:** Serves as an educational tool for early childhood learning (under parental control) and supports individuals interested in learning sign language.
- **Accuracy and Performance:** Ensures high accuracy and exceptional performance, providing a reliable user experience.
- **Simplicity:** With a user-friendly interface, *LiveVision* simplifies capturing or uploading images for accurate real-time predictions, promoting accessibility for both children and blind individuals.

VIII. LIVEVISION APPLICATION SYSTEM

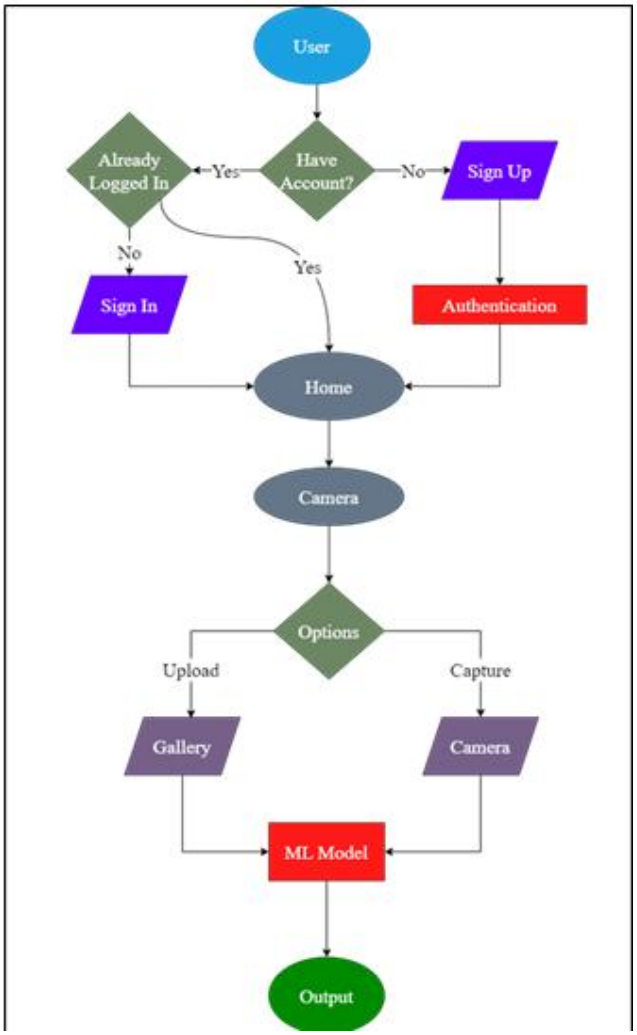

Figure 1: Flow Diagram of LiveVision App

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.67
Volume 5, Issue 7, March 2025

IX. VISUAL REPRESENTATION OF LIVEVISION APP FEATURES

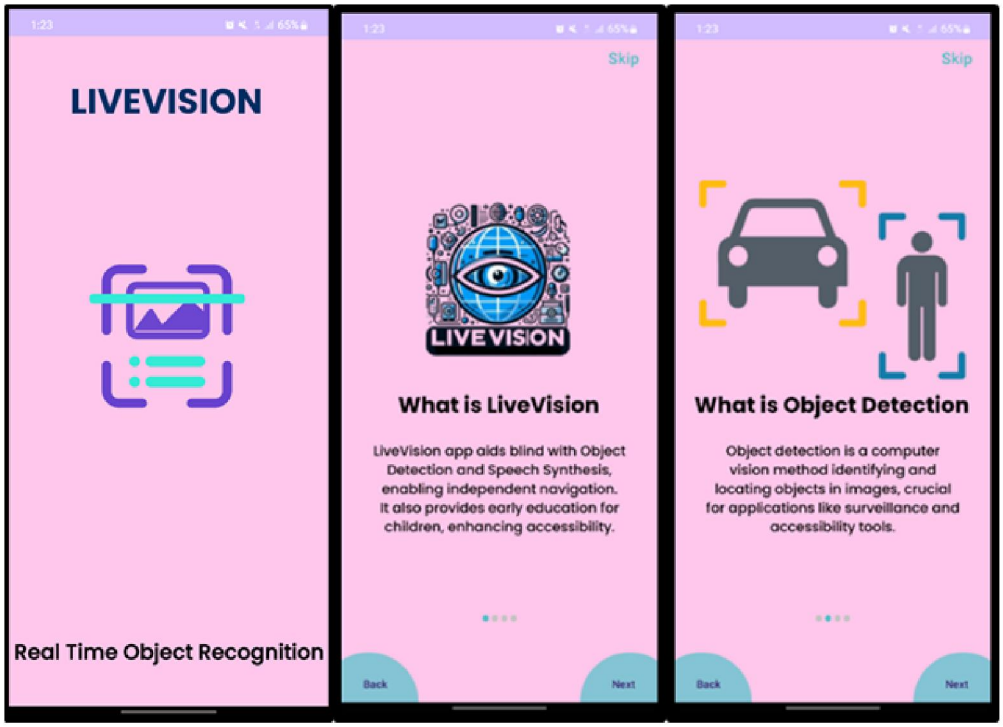

Figure 2: Splash Screen and Introduction to LiveVision Application

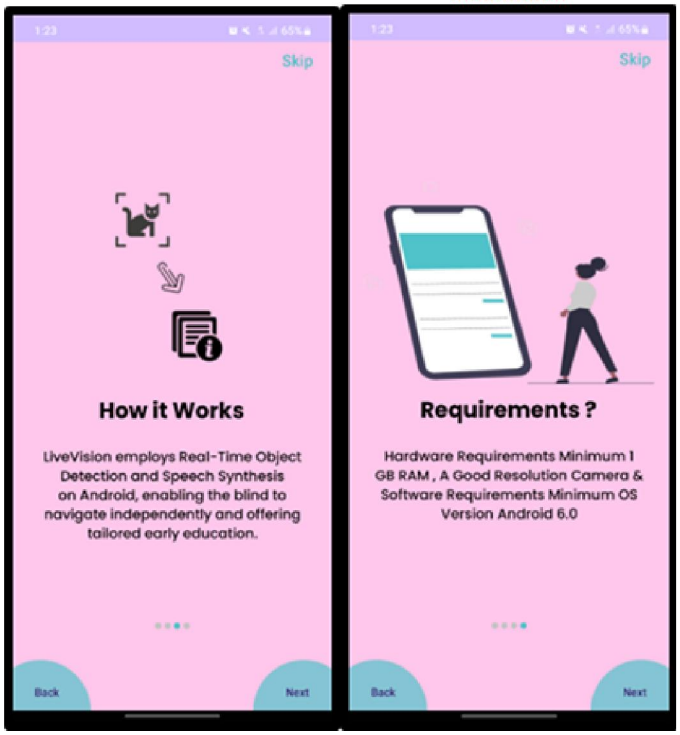

Figure 3: Documentation and More Info about LiveVision Application

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.67
Volume 5, Issue 7, March 2025

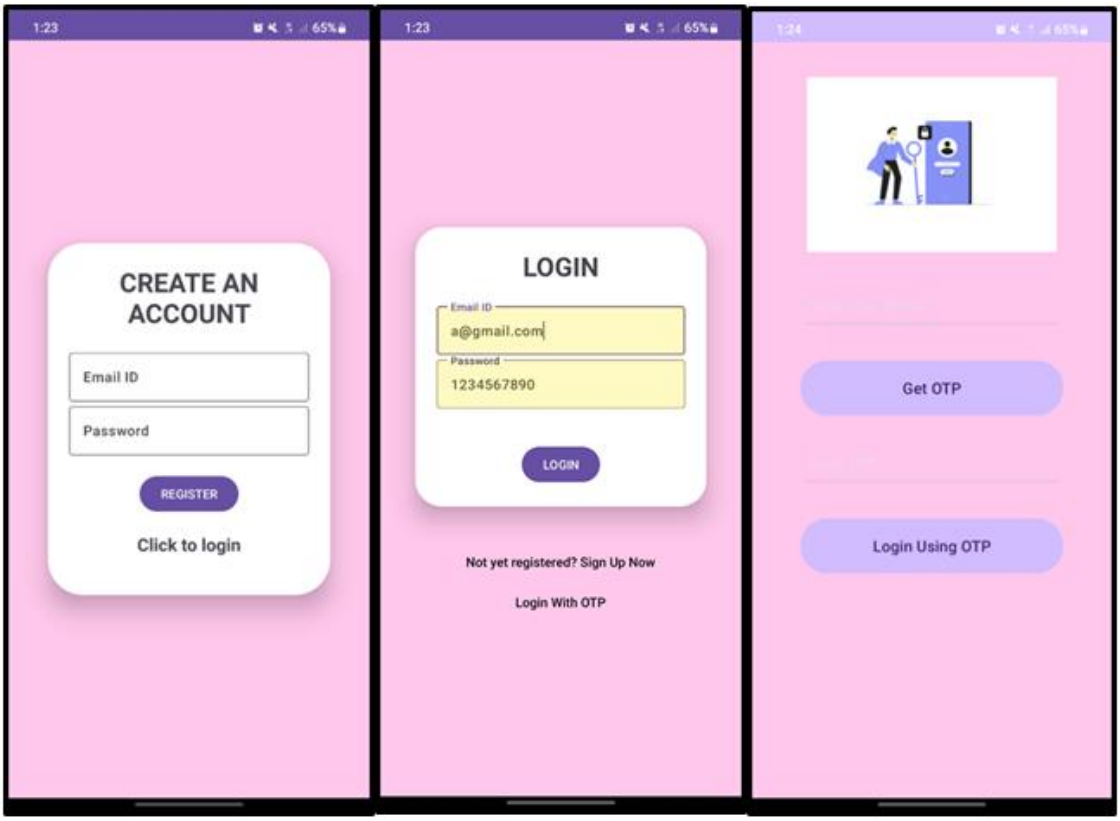

Figure 4: Login/New Account Creation ScreenFigure 5: OTP Verification

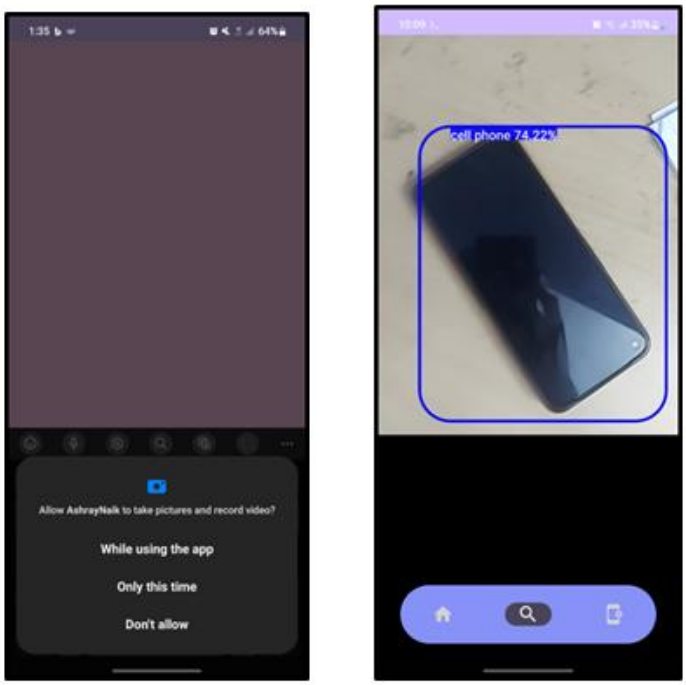

Figure 6: Camera Access Permission and Detection of Object in Real-Time

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.67
Volume 5, Issue 7, March 2025

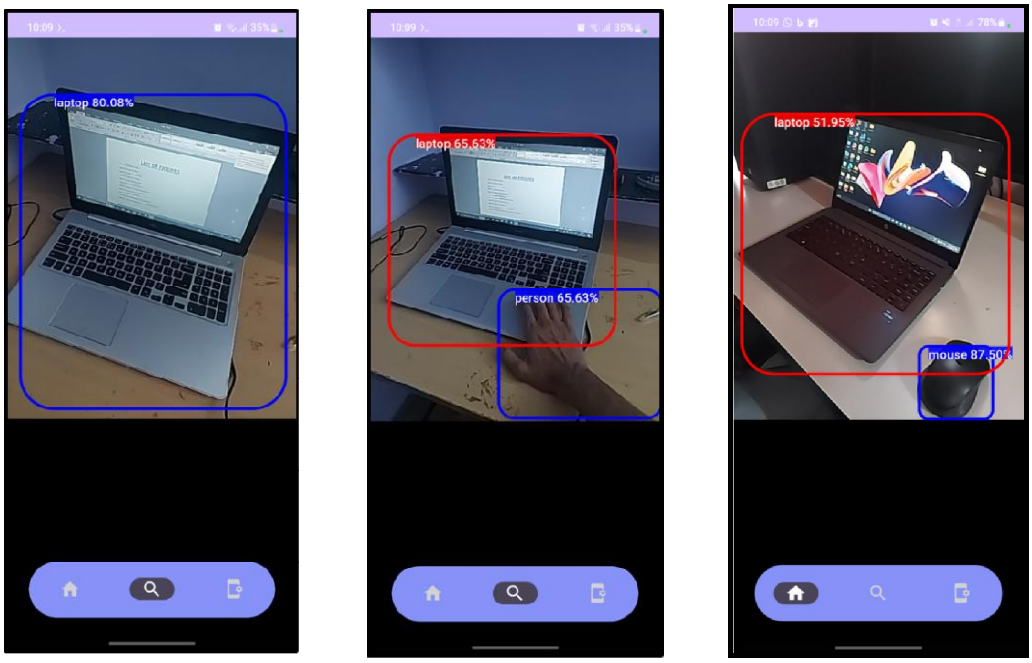

Figure 7: Camera Option and Detection of Objectin Real-Time

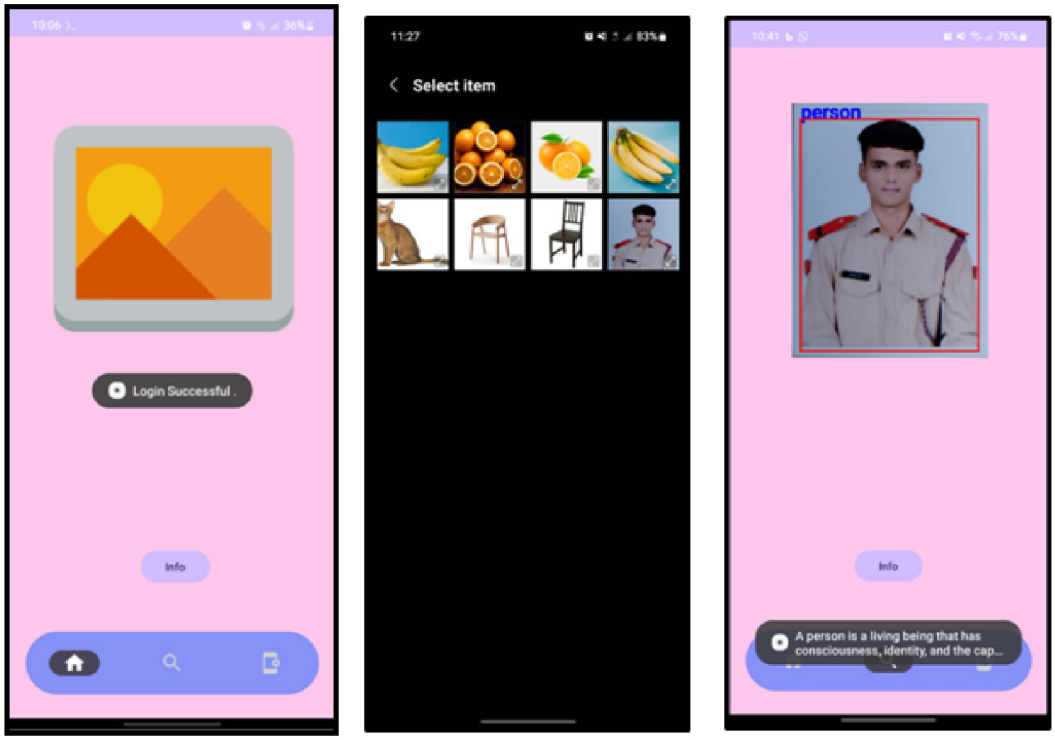

Figure 8: Gallery Option,Image Detection, Text and Speech Synthesis (Text-to-Speech Conversion)

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.67
Volume 5, Issue 7, March 2025

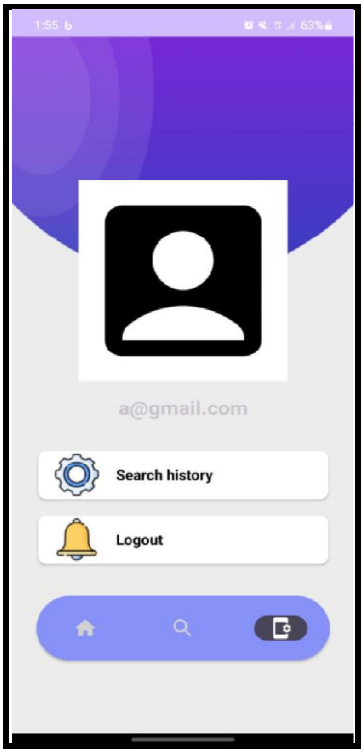

Figure 9: User Profile of the LiveVision App

X. CONCLUSION

LiveVision represents a pioneering step forward in the integration of real-time object recognition and speech synthesis using Machine Learning offering a powerful tool for both educational enrichment and accessibility. By catering to children as well as providing crucial support for the visually impaired, *LiveVision* is not only reshaping the way education is delivered but also ensuring that it is more inclusive. Its current features lay the foundation for a dynamic, user-centric platform, while future advancements—such as augmented reality, gamified elements and collaborative tools—promise to expand its impact further. *LiveVision's* continued evolution emphasizes on accessibility, innovation, and the transformative potential of technology in education contributing to learning which is universally accessible, engaging, and empowering for all.

REFERENCES

[1] J. Redmon, S. Divvala, R. Girshick, and R. B. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779-788.
[2] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," Proceedings of the International Conference on Learning Representations (ICLR), 2015.
[3] H. Lee, S. E. Kahou, Y. G. A. K. J. Zhang, and J. Tang, "Real-Time Object Detection using YOLOv4 and Speech Synthesis for Visually Impaired," Journal of Machine Learning Research, vol. 21, no. 72, pp. 1-18, 2020.
[4] M. A. Nielsen, "Neural Networks and Deep Learning," Determination of Object Recognition Algorithms Using Deep Neural Networks, Book Chapter in Neural Networks, 2015.
[5] T. Yamada, T. Hori, and S. Watanabe, "Real-Time Speech Synthesis and Recognition with Neural Networks," IEEE Transactions on Audio, Speech, and Language Processing, vol. 21, no. 8, pp. 1616-1628, 2013.
[6] X. Chen, X. Yang, and A. S. Huang, "Object Recognition in Real-Time Using Convolutional Neural Networks," Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), 2017, pp. 315-320.