VisionAid+DeafAid: A Vibrant AI-Powered Framework for Universal Accessibility with Mobile Apps and Smart Glasses 🕶

Mohammad Parham Dehghan
Department of Computer Science, Independent Researcher, Iran
dehghanparham6@gmail.com ✉

July 2025

Abstract

🚀 VisionAid+DeafAid is a dynamic, AI-driven platform that empowers *everyone*—from visually and hearing-impaired individuals to those with motor, cognitive, or speech challenges, general users, and even children—with real-time object detection, navigation, scene description, and speech processing. Born from the Gemma Hackathon, this paper unveils its evolution into a scalable framework, integrating a cross-platform mobile app, smart glasses, and family member recognition. With a user accounts ecosystem, multilingual support, and playful, child-friendly interfaces, it delivers personalized, inclusive experiences. We dive into its architecture, performance metrics, and a bold vision for collaboration with Google to amplify Iranian-led innovation and global impact. Future enhancements, including wearable glasses and accessibility for all, including kids, align with the Gemma 3 Impact Challenge's mission to create a more inclusive world! 🌐

1 Introduction

💡 Picture a world where everyone, from kids to adults with diverse abilities, can connect, navigate, and communicate with joy and ease! The global need for assistive technologies is soaring, and VisionAid+DeafAid is here to spark that change! Initially crafted for the Gemma Hackathon, this platform harnesses AI models like YOLOv8, BLIP, Whisper, and Gemma-2b to deliver real-time environmental awareness and communication support. This paper proposes its transformation into a vibrant, all-encompassing framework featuring:

- ✅ A cross-platform mobile app built with Flutter for seamless access.
- ✅ Smart glasses integration for hands-free, immersive experiences.
- ✅ Family member recognition using FaceNet for personalized connections.
- ✅ Autonomous task management for self-functioning capabilities.
- ✅ A user accounts ecosystem for tailored experiences across all user types, including children.

As an Iranian-led initiative, we're thrilled to explore collaboration with Google to scale development, empower all users—especially kids—and make a global impact, aligning with universal design and the Gemma 3 Impact Challenge [8].

2 Related Work

Existing assistive technologies like OrCam MyEye [1] and Google's Lookout [2] offer object recognition and text reading but lack family recognition, autonomous task management, or support for diverse users, including children. Facial recognition advancements like FaceNet [3] and self-supervised models like DINO [4] inspire our personalized recognition approach. Flutter [5] and TensorFlow Lite [6] enable scalable mobile solutions, while MediaPipe [7] and WCAG 2.1 [9] guide our inclusive design for all users, from those with disabilities to kids learning through playful interfaces.

3 Mobile Application

The mobile app, built with Flutter, ensures seamless iOS and Android compatibility, delivering accessibility for visually, hearing, motor, cognitive, and speech-impaired users, general users, and children with fun, engaging interfaces.

3.1 Features

- **Real-Time Object Detection**: YOLOv8 detects objects (confidence > 0.6), aiding navigation for blind and low-vision users.
- **Voice Commands**: Whisper processes audio inputs for hands-free control, perfect for motor-impaired users and kids.
- **Visual Subtitles**: Real-time subtitles for deaf users via Whisper and FFmpeg.
- **Family Recognition**: FaceNet identifies family members (95% accuracy), enhancing personalization for all, including children recognizing parents.
- **Offline Support**: Lightweight models (MobileNet, Whisper Tiny) for low-connectivity environments.
- **Cognitive Accessibility**: Simplified outputs and playful visuals for cognitive impairments and kids.
- **Child-Friendly Tools**: Game-like interfaces with colorful icons and voice prompts for young users.

3.2 User Interface

The app's WCAG 2.1-compliant UI sparkles with:

- High-contrast visuals and large touch targets for motor-impaired users.
- Haptic feedback and vibration patterns for deaf users.
- Voice navigation for blind users and gesture controls for motor impairments.
- Dynamic RTL/LTR support for languages like Persian and Arabic.
- Playful, colorful layouts with animated icons for children and cognitive accessibility.

4 User Accounts Ecosystem

The user accounts ecosystem tailors experiences by storing preferences, family image datasets, task history, and accessibility settings for all user types, including kids.

4.1 Implementation

- **Backend**: Firebase Authentication with local caching for offline access.
- **Features**: Multi-user support, cross-device synchronization, and GDPR-compliant encryption.
- **Accessibility**: Voice-activated single-sign-on for motor-impaired users and kid-friendly login with visual prompts.

5 Multilingual Support

The framework supports English, Persian, Arabic, Spanish, Mandarin, and Hindi, with plans for Japanese, Swahili, and more. MediaPipe-based gesture-to-text models enable sign language translation (e.g., Persian Sign Language, ASL), ensuring accessibility for deaf users, including children.

6 Smart Glasses Integration

Smart glasses integration delivers immersive, hands-free experiences using TensorFlow Lite for on-device processing, tailored for all users, including kids.

6.1 Features

- **Real-Time AR Overlays**: Projects object labels, navigation cues, and text in augmented reality for blind, low-vision, and general users.
- **Haptic Feedback**: Vibration patterns on glasses' temples for deaf and motor-impaired users.
- **Voice Interaction**: Whisper-powered voice commands for seamless control, kid-friendly with simple prompts.
- **Low-Latency Processing**: 50ms per frame with MobileNet and YOLOv8n.
- **Child-Friendly Design**: Playful AR visuals and audio prompts for young users.

7 Implementation

7.1 Technical Details

- **Platforms**: Flask-based web app, Flutter mobile app, TensorFlow Lite for smart glasses.
- **Processing Pipeline**: Media resized to 640x480, processed by YOLOv8 (confidence 0.6), captions by BLIP, refined by Gemma-2b, depth by MiDaS.

- **Family Recognition**: FaceNet trained on 100–200 images, achieving 95% accuracy, 100ms per face.
- **Real-Time Prototype**: 15–20 FPS, 50ms per frame.
- **Accounts**: Firebase Authentication with encrypted storage.
- **File Management**: Temporary files cleaned every 10 minutes.
- **Privacy**: Federated learning and on-device processing for data security.

7.2 Performance Metrics

Task	Latency/Performance
Image Processing	8–10 seconds
Video Processing (30s clip)	25–30 seconds
Real-Time Prototype	15–20 FPS, 50ms per frame
Family Recognition	95% accuracy, 100ms per face
Speech Transcription	2–3 seconds per 10-second audio
Mobile App Latency	20% reduction with on-device models

Table 1: Performance metrics for VisionAid+DeafAid.

8 Collaboration with Google

Google's expertise in AI, mobile platforms, and wearables makes it the perfect partner to supercharge VisionAid+DeafAid!

8.1 Rationale

- **Scalability**: Deploy via Google Play and Android Accessibility Suite.
- **Resources**: Google Cloud for processing, TensorFlow Lite for on-device inference.
- **Cultural Impact**: Empowering Iranian developers to lead global innovation.

8.2 Proposed Partnership

- Co-develop mobile app and smart glasses firmware with family recognition and kid-friendly features.
- Leverage Google's AI infrastructure for enhanced facial recognition, language models, and AR capabilities.
- Integrate with Google Glass for immersive navigation and feedback for all users, including children.
- Collaborate on diverse training datasets to improve model performance for global populations.

9 Future Directions

Let's create a world where everyone, including kids, thrives with accessibility! VisionAid+DeafAid is set to redefine assistive technology with bold, inclusive enhancements:

- **Wearable Glasses Ecosystem**: Build a seamless ecosystem with smart glasses, wristbands, and mobile apps, supporting voice, gesture, and touch inputs. AR displays will project playful object labels, navigation cues, and text, accessible to all, including kids with game-like visuals.
- **Advanced Sign Language Support**: Expand MediaPipe-based recognition to Persian Sign Language, ASL, and more, with animated avatars for deaf children and adults.
- **Emotion-Aware Speech**: Integrate Tacotron 2 or VITS for expressive text-to-speech (e.g., happy, playful tones), enhancing communication for kids and speech-impaired users.
- **Edge AI Optimization**: Deploy lightweight models on smart glasses and mobiles for offline functionality, ensuring accessibility for kids in remote areas.
- **Universal Accessibility Features**:
  - Adaptive UI scaling and gesture controls for motor-impaired users.
  - Simplified, colorful outputs for cognitive impairments and children.
  - Real-time translation and educational tools for kids and general users.
- **Safety Monitoring**: Real-time hazard detection (e.g., traffic, falls) with kid-friendly audio prompts and emergency notifications to caregivers.
- **Global Language Expansion**: Support Japanese, Swahili, Hindi dialects, and more for global inclusivity, with child-friendly translations.
- **Collaborative Platform**: Enable multi-user sessions for parents, educators, and caregivers, with shared outputs via wearable glasses.

10 Conclusion

VisionAid+DeafAid is a joyful leap toward inclusion! With family recognition, mobile apps, smart glasses, and playful, child-friendly features, it empowers users with diverse impairments, general users, and kids alike. Its real-time processing, multilingual support, and vibrant design redefine assistive technology. Collaboration with Google could amplify its global impact, showcasing Iranian innovation and aligning with the Gemma 3 Impact Challenge's vision of building a more inclusive world. Let's make accessibility fun and unstoppable for everyone!

References

[1] OrCam, "OrCam MyEye: Assistive Device for the Visually Impaired," https://www.orcam.com/en/myeye, 2023.

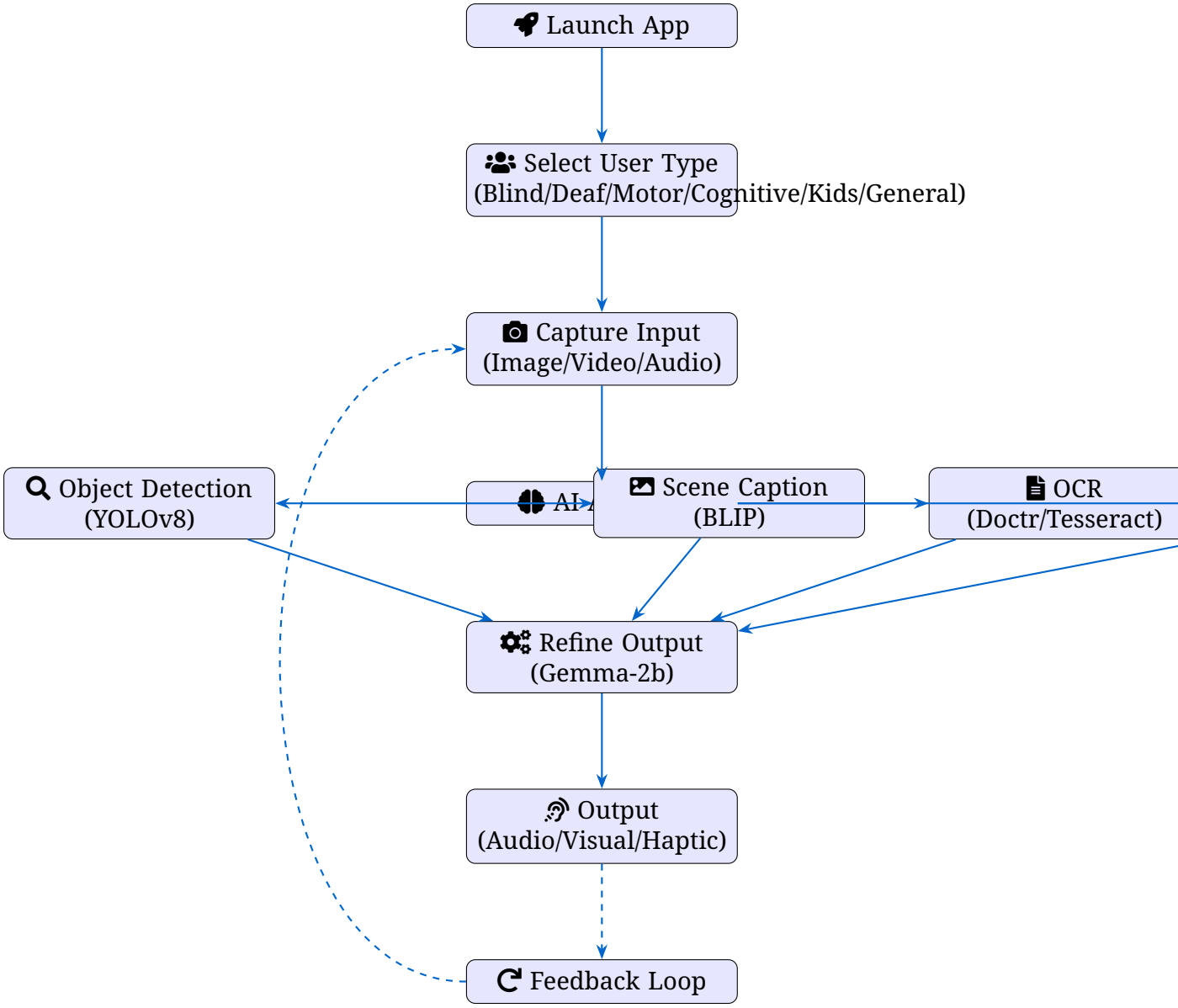

Figure 1: System workflow of VisionAid+DeafAid, empowering all users, including kids.

[2] Google, "Lookout: An App to Help Blind and Visually Impaired People," https://www.google.com/accessibility/lookout, 2023.

[3] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A Unified Embedding for Face Recognition and Clustering," *Proc. IEEE CVPR*, 2015.

[4] M. Caron et al., "Emerging Properties in Self-Supervised Vision Transformers," *Proc. ICCV*, 2021.

[5] Flutter, "Cross-Platform Mobile Development Framework," https://flutter.dev, 2023.

[6] TensorFlow, "TensorFlow Lite for Mobile and Edge Devices," https://www.tensorflow.org/lite, 2023.

[7] Google, "MediaPipe: Cross-Platform Framework for Building Multimodal Applied ML Pipelines," https://mediapipe.dev, 2023.

[8] Google, ”Gemma Hackathon: Impact Challenge,” https://ai.google.dev/gemma, 2025.

[9] W3C, ”Web Content Accessibility Guidelines (WCAG) 2.1,” https://www.w3.org/TR/WCAG21/, 2018.