AI-Assisted Object Detection for Individuals with Visual Impairments

Devyaan Bordloi (Avon High School, Avon CT, USA)

Abstract:

Visual impairment impacts over 250 million people worldwide, causing day-to-day difficulties for these individuals. While traditional tools like white canes and guide dogs are essential for mobility, they cannot verbally identify specific objects or obstacles in the user's environment. This project introduces a proof of concept for an AI-assisted object detection system designed to bridge that gap. We utilized Google's Teachable Machine and TensorFlow.js to create a browser-based model that identifies objects and announces them via a text-to-speech engine. We tested the system using various optical setups, including webcams and smartphones, and image data was processed locally on hardware ranging from high-performance laptops to standard Chromebooks. Our results showed that the model achieved extremely high accuracy for specific object classes while maintaining real-time performance on low-cost devices. This study demonstrates that accessible, browser-based computer vision technologies can effectively enhance safety and independence for the visually impaired without requiring expensive, specialized hardware.

Keywords:

Accessibility, Affordability, AI-Assistance, Computer Vision

Introduction:

Visual impairment has a far-reaching impact across the globe (~253 million affected individuals), causing a significant quality of life and economic impact for individuals and society at large (Ackland et al.; Marques et al.). Visual impairment can impact an individual at the health, social, and economic levels. Individuals with visual impairment are known to have difficulty performing daily tasks (West et al.) and require higher levels of dependency on informal and formal care (Hong et al.; West et al.). Accordingly, they also experience higher levels of mental health difficulties (Klauke et al.; Zheng et al.), lower educational achievement (Ma et al.), workplace productivity (Reddy et al.), and overall quality of life (Brown et al.). Collectively, these effects at the individual level amount to a global economic burden of $411 billion USD annually (Ackland et al.). This burden is partly driven by healthcare costs associated with visual impairment (Javitt et al.; Evans et al.; Wang et al.) along with loss of productivity for those afflicted (Reddy et al.).

Visual impairment is a broad category that covers many conditions ranging from short-sightedness (which can be remedied with corrective lenses) all the way to legal blindness. Numerous subclassifications are based on the severity of visual disability, cause, type of impairment, and range of vision. According to Cleveland Clinic (“Blindness”), blindness affects 43 million people (about 0.5% of the global population) and is driven by a combination of genetic and environmental factors. In comparison, mild visual impairments such as myopia, hyperopia, and astigmatism, are far more prevalent, with Cleveland Clinic (“Hyperopia”) estimating that

hyperopia alone affects 30% of adults globally, and are able to be corrected to an extent using eyeglasses, contacts, and surgeries.

A very common tool used to assist those with more severe visual impairments are white canes, which can be used by the user tapping the cane around their surroundings to identify if there are any obstacles or objects nearby. However, these canes do not have the ability to discern objects and provide this information to the user. Another common tool for those with visual impairments is the use of seeing-eye canines, which help their owners travel through complex environments. However, these dogs do have many limitations when it comes to identifying objects and communicating them to their user. While it is possible to train the dogs to recognize certain objects and alert their owner (typically by barking), there is an intrinsic language barrier that exists across species. This language barrier can be problematic, especially in situations when there are many unique objects nearby, and their proper identification is critical for the user.

Considering all the difficulties with existing visual impairment tools, we utilized Google's Teachable Machine to train a computer vision model capable of recognizing specific objects and audibly describe those objects to the user. We implemented this model using TensorFlow.js to ensure efficient, real-time performance within a web browser. The system paired this detection capability with a text-to-speech engine that audibly states the name of the recognized item. This integration is particularly valuable when developing assistive equipment for individuals with visual impairments. It allows users to identify objects with certainty rather than guessing based on tactile cues like shape or texture. The software effectively translated visual data into spoken audio without significant latency. Ultimately, AI-powered vision models have the potential to enhance user safety and foster more independent interaction with the daily environment.

Materials & Methods:

Our goal was to develop a real-life proof of concept for an AI-assisted device to help visually impaired individuals. As depicted in Figure 1, a hypothetical user with visual impairments would be able to simply look at an object, have it be identified by the goggles, and listen to the object's name.

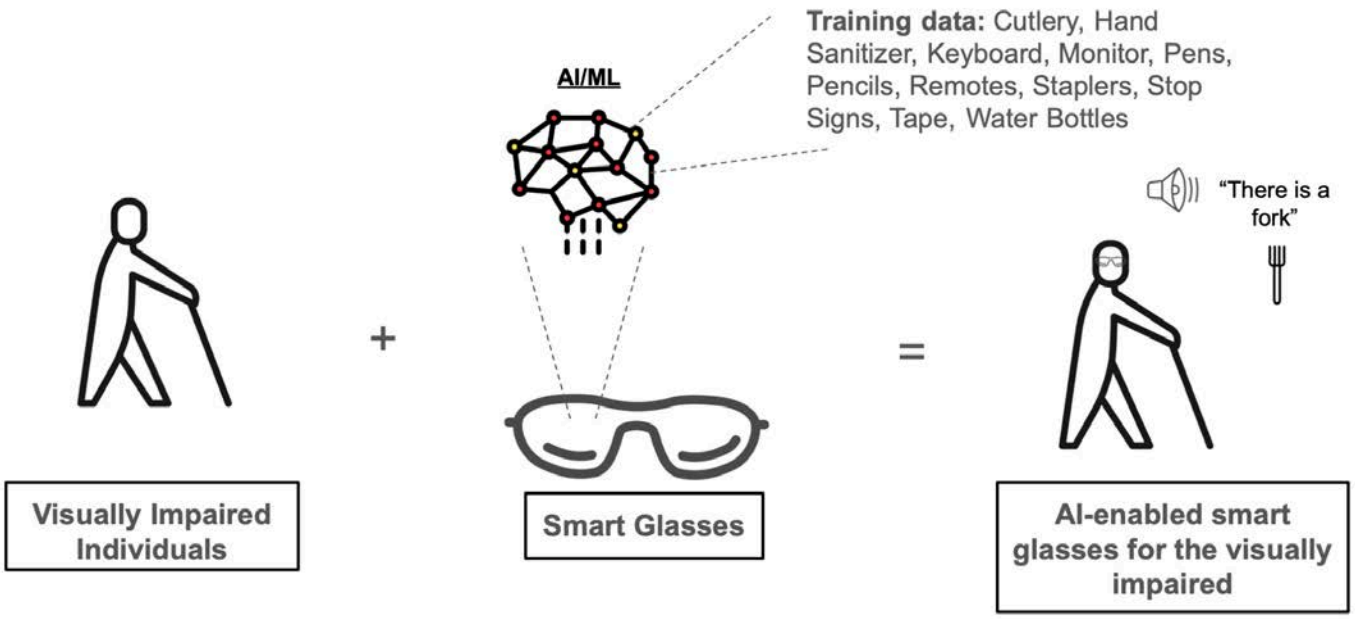

Figure 1. *A graphical representation of the overarching theme and application of the technology, enabling object recognition through smart glasses for those with visual impairments. A person with visual impairments would wear a pair of smart glasses connected to an artificial intelligence/machine learning system (AI/ML) system with an integrated camera on the glasses to provide a video feed to the AI/ML system. The AI/ML system would leverage previous training data for object detection of household items. Upon high fidelity detection, the system would utilize a text-to-speech, or TTS program, along with the integrated speaker in order to announce to the user what the object is.*

We utilized both Google Teachable Machine and TensorFlow.js to train and power the model. This specific configuration was chosen following an empirical evaluation of several different frameworks. Google Teachable Machine ultimately provided the most turnkey user experience. It also featured fast training speeds and was among the most accurate options for recognizing images, as Teachable Machine was able to correctly identify a trained object within one second 99% of the time while Microsoft's Lode platform was only able to identify objects at the same accuracy within 5 seconds. The Google platform works by training the model directly on the user's device. Additionally, it utilizes pre-trained models built on large datasets to make the training process much quicker. Once the model learns from the specialized training data, it is exported into a TensorFlow.js format. This efficient model type made it simple to implement the Text-to-Speech system within a lightweight HTML site.

For the training data, we compiled images via a combination of original captures and publicly sourced data available on the internet. The images were selected based on their relevance to the types of objects that an average user could encounter in normal environments, such as office supplies, kitchen utensils, obstacles, and signs that convey important information (Figure 2). Once it was trained, all thirteen classes measured at an accuracy of about .9999 out of 1 (Table 1, Figure 3, Figure 4).

Figure 2. *A collection of images that were either collected in person or from the internet and uploaded to the image recognition model as training data.*

Table 1. *Data on the accuracy of the machine learning model during testing.*

Table 1 (model performance)

Object:	Accuracy:
Hand Sanitizer	.9999
Keyboard	.9999
Monitor	.8999
Pen	.9999
Pencil	.9999
Remote	.9999
Stapler	.9999
Tape	.9999
Water	.9999
Fork	.9999
Spoon	.9999
Knife	.9999
Stop Sign	.9999

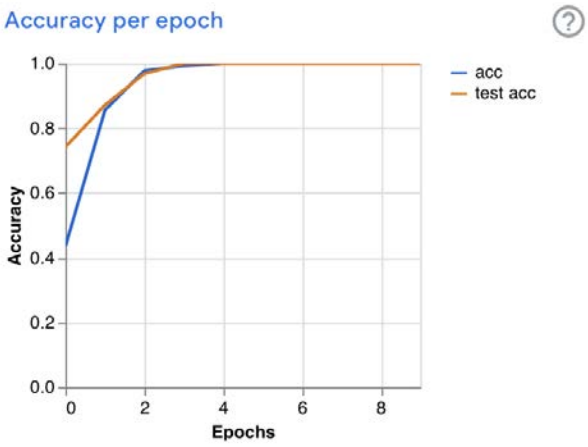

Figure 3. *A graph of the machine learning model's accuracy as a decimal for each epoch it was trained. In the first epoch, the accuracy was at approximately .85, or 85%, and this number continued to rise until the fourth epoch, after which it stabilized at about .99, meaning it had an accuracy of 99%.*

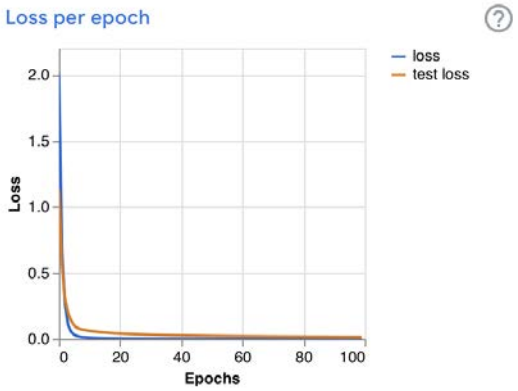

Figure 4. *A graph showing the convergence of training loss and test loss over 100 epochs, indicating successful model training without significant overfitting.*

We then began our testing process with a standard computer webcam to establish a baseline. From there, we explored both wired and wireless portable solutions to see how the model would adapt. First, we connected an Insta360 camera directly to the computer using a USB-C cable. We chose this wired setup to minimize latency while taking advantage of the camera's 48-megapixel main sensor, which captured video at a steady 30 frames per second.

Next, we wanted to test a completely wireless approach. We connected an iPhone 15 Pro over Wi-Fi. Although this wireless connection introduced a slight delay, it allowed us to leverage the phone's 48-megapixel sensor to push the frame rate much higher, reaching 60 frames per second.
To ensure our image recognition model had plenty of overhead and would not be bottlenecked during these initial trials, we processed all video input locally on a robust machine. We relied on a MacBook Pro equipped with an M4 Pro CPU and 48 gigabytes of RAM.

However, proving the model worked on a powerful system was only half the goal. We also needed to make sure it remained accessible for everyday users. To verify this, we conducted a final round of testing on a Chromebook featuring an Intel Celeron N4120 CPU and 4 gigabytes of RAM. This crucial step confirmed that the image recognition model can successfully run even on lower-end hardware.

We constructed the graphical user interface, or GUI, with very foundational HTML5 elements, CSS3, and basic JavaScript. This was a deliberate choice to avoid heavy and unnecessary frameworks, thus resulting in a minimalist design and a very lightweight codebase. The main benefit was improved performance, shown in its fast load times and low usage of system resources such as CPU cores and RAM. This basic design also allowed for maximum accessibility for the end-user, since it can be run on virtually any device with an internet connection and a modern web browser, and the lack of distracting elements make it so anyone can quickly learn how to use the interface.

The full code can be found at the following GitHub repository:
https://github.com/devyaanb/Image-Recognition-TTS

Results:

When testing the object detection system using the images in Figure 5, the model demonstrated performance that significantly varied based on environmental conditions. In controlled, indoor environments with ideal lighting, the model consistently and correctly recognized the trained objects. In these conditions, the model achieved an extremely high accuracy of around .9998. In outdoor environments with harsher lighting and more objects in the background, a model had a lower accuracy of around .9288. The accuracies were similar regardless of the type of camera being used.

Figure 5. *A collection of images that were either collected in person or from the internet and uploaded to the image recognition model as testing data. We chose these images to be the same types of objects as the training data while being distinct enough to show that the model can recognize the object between differing images.*

However, the system's accuracy noticeably decreased when tested outdoors or in environments with extreme lighting conditions. In scenarios with very high or low light, the model was less accurate and struggled to consistently identify the trained objects. This drop in performance indicates sensitivity to lighting variations found outside of controlled indoor settings. Additionally, testing setups with better cameras were able to more effectively recognize the objects in these conditions, indicating that higher-quality sensors may be necessary for outdoor usage.

Discussion:

Visual impairment impacts hundreds of millions of people globally and creates significant daily challenges, leading to a massive economic and social burden. The severity ranges from easily correctable issues to total blindness, and while traditional aids like white canes and guide dogs offer crucial mobility support, they cannot explicitly identify specific objects. To bridge this gap, we developed an AI-powered computer vision model using Google's Teachable Machine and TensorFlow.js, which pairs real-time object detection with a text-to-speech engine. By audibly announcing recognized items, this technology provides visually impaired users with precise information about their surroundings, ultimately fostering greater independence and safety in their daily lives.

This proof of concept demonstrates the ability of browser-based computer vision models to successfully identify daily objects with high accuracy using standard consumer hardware. An important finding was the model's ability to perform real-time image detection on low-power and low-cost devices such as Chromebooks, showing that expensive high-end processors are not needed for effective assistance. This technology goes beyond the limitations of conventional tools, such as white canes by providing users with specific and audible information about their surroundings instead of simply alerting them to nearby obstacles. As such, this approach has the potential to significantly enhance independence for individuals with visual impairments while reducing the economic burden associated with current high-cost assistive technologies. Despite these promising results, the system currently faces some limitations regarding false positives, portability, and speed. Additionally, the project is currently restricted to a software interface and lacks the custom wearable hardware integration necessary for a truly portable, hands-free user experience. While further optimization is required to handle more complex environments, this study confirms the fundamental feasibility of accessible and AI-driven visual aids.

With additional funding and time dedicated to development, the next phase of this project would be to expand the object detection model to include a much broader and more diverse selection of environments through a larger dataset and possibly another, more complex model framework. Funding would also allow for the creation of a fully wearable prototype equipped with integrated cameras and an onboard processor rather than having to rely on a separate device to load the model. Another possibility would be to utilize a much larger pre-trained model and leveraging quantization technology to effectively shrink the model down into something that could easily be run on consumer devices.

A significant financial investment is the essential next step for this project. It would provide the necessary capital to move beyond the planning stages and create the first physical prototypes of

the camera-equipped goggles, turning the core concept into a tangible product. Beyond prototyping, such an investment could also provide the foundation needed to launch a full-scale mass production phase. The transition to mass production is the most critical element for ensuring widespread accessibility. By manufacturing the goggles in large quantities, we could dramatically lower the production cost for each individual unit through efficiencies in the assembly process and the ability to purchase materials in bulk at a lower price. This cost-saving would be passed directly to the consumer, resulting in a final price that is affordable for the vast majority of people with visual impairments. This approach directly confronts the primary issue with the current market, where competing solutions have prices that are prohibitively high for the average person (Global Report on Assistive Technology). These existing devices are often so expensive that they are not a realistic option, especially for those who either don't have insurance or have plans which don't cover such items.

Acknowledgments:

I would like to thank my mentor, Rahul Patel, for his valuable guidance throughout the research and writing process.

Works Cited

Ackland, Peter, et al. “World Blindness and Visual Impairment: Despite Many Successes, the Problem Is Growing.” *Community Eye Health*, vol. 30, no. 100, 2017, pp. 71–73.

“Blindness (Vision Impairment): Types, Causes and Treatment.” *Cleveland Clinic*, https://my.clevelandclinic.org/health/diseases/24446-blindness. Accessed 21 July 2026.

Brown, Melissa M., et al. “Quality of Life Associated with Visual Loss: A Time Tradeoff Utility Analysis Comparison with Medical Health States.” *Ophthalmology*, vol. 110, no. 6, June 2003, pp. 1076–81. *PubMed*, https://doi.org/10.1016/S0161-6420(03)00254-9.

Evans, Jennifer R., et al. “Hospital Admissions in Older People with Visual Impairment in Britain.” *BMC Ophthalmology*, vol. 8, Sept. 2008, p. 16. *PubMed*, https://doi.org/10.1186/1471-2415-8-16.

“Farsightedness: Causes & Symptoms.” *Cleveland Clinic*, https://my.clevelandclinic.org/health/diseases/hyperopia-farsightedness. Accessed 21 July 2026.

Global Report on Assistive Technology. 1st ed, World Health Organization, 2022.

Hong, Thomas, et al. “Visual Impairment and Subsequent Use of Support Services among Older People: Longitudinal Findings from the Blue Mountains Eye Study.” *American Journal of Ophthalmology*, vol. 156, no. 2, Aug. 2013, pp. 393-399.e1. *PubMed*, https://doi.org/10.1016/j.ajo.2013.04.002.

Javitt, Jonathan C., et al. “Association between Vision Loss and Higher Medical Care Costs in Medicare Beneficiaries Costs Are Greater for Those with Progressive Vision Loss.” *Ophthalmology*, vol. 114, no. 2, Feb. 2007, pp. 238–45. *PubMed*, https://doi.org/10.1016/j.ophtha.2006.07.054.

Klauke, Susanne, et al. “The Impact of Low Vision on Social Function: The Potential Importance of Lost Visual Social Cues.” *Journal of Optometry*, vol. 16, no. 1, 2023, pp. 3–11. *PubMed*, https://doi.org/10.1016/j.optom.2022.03.003.

Ma, Xiaochen, et al. “Effect of Providing Free Glasses on Children’s Educational Outcomes in China: Cluster Randomized Controlled Trial.” *BMJ (Clinical Research Ed.)*, vol. 349, Sept. 2014, p. g5740. *PubMed*, https://doi.org/10.1136/bmj.g5740.

Marques, Ana Patricia, et al. “Estimating the Global Cost of Vision Impairment and Its Major Causes: Protocol for a Systematic Review.” *BMJ Open*, vol. 10, no. 9, Sept. 2020, p. e036689. *PubMed*, https://doi.org/10.1136/bmjopen-2019-036689.

Reddy, Priya Adhisesha, et al. “Effect of Providing near Glasses on Productivity among Rural Indian Tea Workers with Presbyopia (PROSPER): A Randomised Trial.” *The Lancet. Global Health*, vol. 6, no. 9, Sept. 2018, pp. e1019–27. *PubMed*, https://doi.org/10.1016/S2214-109X(18)30329-2.

“Effect of Providing near Glasses on Productivity among Rural Indian Tea Workers with Presbyopia (PROSPER): A Randomised Trial.” *The Lancet. Global Health*, vol. 6, no. 9, Sept. 2018, pp. e1019–27. *PubMed*, https://doi.org/10.1016/S2214-109X(18)30329-2.

Wang, Meng-Ting, et al. “Analysis of Excess Direct Medical Costs of Vision Impairment in Taiwan.” *Value in Health Regional Issues*, vol. 2, no. 1, May 2013, pp. 57–63. *PubMed*, https://doi.org/10.1016/j.vhri.2013.01.006.

West, Sheila K., et al. “How Does Visual Impairment Affect Performance on Tasks of Everyday Life? The SEE Project. Salisbury Eye Evaluation.” *Archives of Ophthalmology*

(Chicago, Ill.: 1960), vol. 120, no. 6, June 2002, pp. 774–80. *PubMed*, https://doi.org/10.1001/archopht.120.6.774.

"How Does Visual Impairment Affect Performance on Tasks of Everyday Life? The SEE Project. Salisbury Eye Evaluation." *Archives of Ophthalmology (Chicago, Ill.: 1960)*, vol. 120, no. 6, June 2002, pp. 774–80. *PubMed*, https://doi.org/10.1001/archopht.120.6.774.

Zheng, Yajing, et al. "The Prevalence of Depression and Depressive Symptoms among Eye Disease Patients: A Systematic Review and Meta-Analysis." *Scientific Reports*, vol. 7, Apr. 2017, p. 46453. *PubMed*, https://doi.org/10.1038/srep46453.