

VisionMate: Multimodal AI Smart Glasses for Visually Impaired People: A Review

Tavhare Sakshi Hemant¹, Bangar Abhishek Papat², Bhosale Sachin Bajirao³

¹Department of Computer Engineering / Jaihind College of Engineering / SPPU, Pune, India

²Department of Computer Engineering / Jaihind College of Engineering / SPPU, Pune, India

³Department of Computer Engineering / Jaihind College of Engineering / SPPU, Pune, India

Abstract- Visually impaired individuals face daily mobility and information-access challenges that traditional aids, such as white canes or guide dogs, cannot fully resolve, particularly regarding dynamic obstacles, environmental hazard detection, and reading printed text. This paper presents a structured overview of the VisionMate Multimodal AI Smart Glasses ecosystem, a wearable assistive technology designed to bridge these gaps through real-time computer vision, optical character recognition, GPS mapping, and audio feedback. The system architecture distributes workloads between a lightweight, camera-enabled smart glasses frame and a companion mobile processing unit running offline-capable deep learning models. Key functional modes include real-time obstacle detection with alert throttling, instant text-to-speech conversion using OCR pipelines, spatial GPS navigation, and emergency SOS integration. However, assistive wearable deployments face ongoing design constraints, including battery limitations, processing latency, and the need for seamless, hands-free accessibility controls. The system demonstrates how multimodal AI integration can enhance autonomy, spatial awareness, and independence for users with visual impairments.

Keywords— Assistive Technology; Visually Impaired; Smart Glasses; Multimodal AI; Computer Vision; Optical Character Recognition (OCR); Obstacle Detection; Edge Computing.

I. INTRODUCTION

Visual impairment remains a significant global challenge, affecting millions of individuals and limiting their ability to navigate independently, recognize environmental hazards, and access printed information. Traditional mobility aids, such as white canes and guide dogs, provide essential tactile and auditory feedback but cannot detect overhead obstacles, read text, or identify distant dynamic hazards.

While recent advancements in wearable technology and artificial intelligence have opened new pathways for assistive devices, many existing solutions focus on a single capability—such as standalone obstacle detection or dedicated text readers—rather than offering a comprehensive, integrated experience.

The problem is further compounded by the limitations of current assistive deployments, which often suffer from high processing latency, reliance on continuous cloud connectivity, power-intensive hardware, and complex visual user interfaces that are

unsuitable for users who are blind or visually impaired. Furthermore, the effectiveness of wearable assistive systems depends heavily on balancing computational efficiency with user safety, privacy, and low-latency feedback.

This section introduces the foundational concepts of the VisionMate Multimodal AI Smart Glasses ecosystem. The aims of this study are to (i) synthesise the architecture and multimodal perception approaches for wearable assistive smart glasses, (ii) compare their methods, performance, and accessibility trade-offs, (iii) propose a literature-grounded conceptual framework, and (iv) identify core research gaps that must be closed before widespread deployment.

II. REVIEW METHODOLOGY

Publisher platforms (MDPI, IEEE, Springer, Elsevier, SAGE, Taylor & Francis, ASME), DOAJ, institutional repositories and arXiv were searched with terms combining level crossing, grade crossing and railway crossing with obstacle detection, computer vision,

deep learning, YOLO, LiDAR, radar, sensor fusion, trespass, near miss, crash prediction, injury severity and connected vehicle. Publications from 2021–2026 were prioritised, while earlier studies were retained where they established a method or dataset still in use.

Studies were included when they addressed level or grade crossings directly or provided closely related railway perception methods or datasets. Each record was verified against a publisher page, DOI record or repository; two arXiv preprints are labelled, and theses, trade articles and unconfirmed records were excluded. Because Scopus and Web of Science searches could not be logged, the review is structured and narrative rather than systematic.

III. LITERATURE REVIEW

1. Assistive Navigation Needs and User Behaviour

Research on assistive technologies emphasizes that safe independent mobility for visually impaired individuals depends heavily on bridging the gap between traditional aids and dynamic environmental hazards. White canes and guide dogs provide essential tactile feedback but cannot identify overhead obstructions, complex obstacles, or distant static and dynamic hazards. Behavioral studies on navigation note that users frequently face unpredictable pedestrian traffic, changing urban layouts, and sudden obstacles, making reliable real-time feedback critical. While automated obstacle detection systems aim to reduce reliance on sighted guides and decrease travel anxiety, user-acceptance studies highlight that audio clutter, high latency, and complex physical controls can create cognitive overload, necessitating intuitive design approaches.

2. Camera-Based Deep Learning for Object and Obstacle Detection Computer vision advancements have shifted assistive navigation from basic ultrasonic sensors toward deep-learning-based object detection deployed on edge hardware. Lightweight convolutional neural networks (CNNs) and efficient detector architectures (such as MobileNet and YOLO variants) allow smart systems to classify pathway hazards like vehicles, steps, poles, and pedestrians with high reported precision.

Pamuła et al. demonstrated that grid-based safe-passage classification can successfully segment empty versus occupied regions to direct users safely. However, studies note that recall rates in dynamic environments remain challenging; missed-detection safety-critical errors must be carefully minimized to prevent user injury.

3. Optical Character Recognition (OCR) and Information Access Beyond physical mobility, text accessibility is a core pillar of independence for visually impaired users. Integrating text-recognition frameworks—such as mobile-optimized OCR pipelines—enables instant extraction and reading of street signs, product packaging, menus, and printed documents. Research shows that converting visual text to speech (TTS) within seconds significantly enhances daily functional autonomy, though challenges remain regarding variable lighting, skewed angles, and blurred camera captures in motion.

4. Sensor Fusion, Edge Perception, and Audio Feedback

Because single-sensor setups face environmental limitations—such as standard cameras failing in low-light conditions or heavy rain—advanced assistive frameworks recommend multi-modal architectures and edge-optimized processing.

Executing perception models locally on mobile or edge units reduces dependency on continuous cloud connectivity, guarantees user privacy, and minimizes latency. Concurrently, human-computer interaction (HCI) research on auditory warnings stresses that message timing, volume scaling, and customizable alert throttling (such as switching between verbose and minimal modes) are crucial to prevent habituation and ensure user safety.

IV. COMPARATIVE ANALYSIS

Table 1 compares fourteen representative studies from the literature review, mapping various assistive technologies, computer vision approaches, and edge-processing frameworks used in wearable and navigation systems. Reported metrics vary across

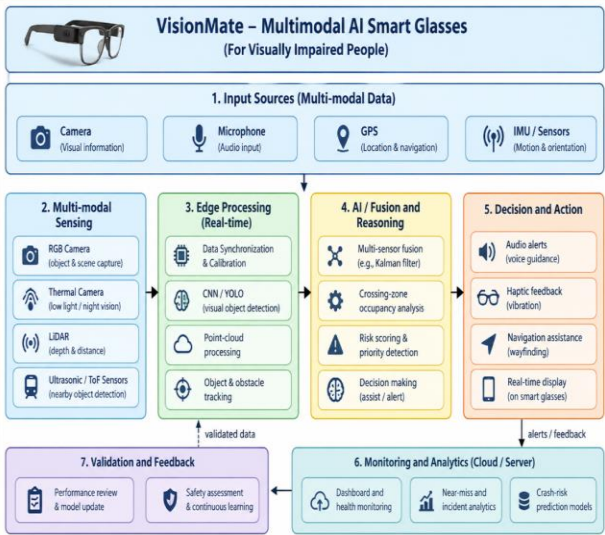
studies due to differing tasks, datasets, and evaluation protocols.

Table 1. Comparison of representative intelligent assistive and computer vision studies.

Author, year	Objective	Method / technology	Major reported findings	Limitations
Aminmansour et al., 2015	Near-miss vehicle detection	Train-mounted camera; region proposals + CNN	Recall 29 % higher than DPM and R-CNN	Absolute recall not in abstract
Amaral et al., 2016 [13]	Small-obstacle detection	Tilting 2D laser scanners; background model	Obstacles detected in 3D point clouds	41 point clouds; one crossing
Beanland et al., 2017 [4]	Driver compliance at passive crossings	Instrumented vehicle; 22 drivers	14 % violations; fewer train checks	No trains present; one region
Zaman et al., 2019 [11]	Real-time trespass detection	Deep-learning video analytics; alerts	No errors reported in tested hours	Short evaluation; 3 feeds
Fayyaz & Johnson, 2020 [6]	Classify crossing objects	CCTV; transfer learning; 2002 with radar	MobileNet 88 % validation accuracy	Public images; no mAP; radar untested
Pamuła & Pamuła, 2021 [7]	Safe-passage classification	Patch-wise CNN on fixed CCTV	F1 = 0.96	1,000 frames; one site
Wisultschew et al., 2021 [14]	Edge detection and tracking	16-channel 3D LiDAR; edge IoT	Real-time tracking, train warnings	Emulated crossing site
Zhang et al., 2022 [12]	Trespass analytics	Deep-learning video analysis	≈ 3,000 events in 1,600 h	One grade crossing
Zawodny et al., 2023 [5]	Shorten closure time	Radar/LiDAR train-speed measurement	Closure 172 s → 54 s (simulated)	Single site; simulation only
Farooq & Khattak, 2023 [21]	Inventory data quality	Field validation; FRA model; ZINB	Predictions changed by >140 % at some sites	560 crossings; one state
Ren et al., 2024 [20]	Injury-severity factors	XGBoost + SHAP on FRA data	93.36 % accuracy (fatal model)	Two years of data
Carletti et al., 2025 [8]	People and vehicle detection	Smart cameras; barrier-state DNN	Precision 0.98, recall 0.88	52 public videos
Senkondo et al., 2025 [19]	Crash-frequency prediction	RF, XGBoost, Transformer– CNN vs NB	ML outperformed NB; SHAP drivers	807 crossings; one state
Sansyzbay et al., 2026 [3]	National protection system design	YOLOv8 + ByteTrack; microwave, loops	Architecture and SIL targets	Calculations only; no field data

V. PROPOSED CONCEPTUAL ARCHITECTURE

The proposed conceptual architecture for the VisionMate Multimodal AI Smart Glasses system distributes tasks across modular edge and mobile layers to ensure low-latency processing, offline reliability, and continuous safety monitoring.



VisionMate is an advanced AI-powered wearable assistive ecosystem designed to empower blind and visually impaired individuals by bridging environmental gaps through real-time computer vision, edge processing, and audio feedback. The system features a lightweight, camera-equipped smart glasses frame paired with a mobile processing unit that handles core tasks locally to ensure low latency and user privacy.

Its primary capabilities include real-time obstacle detection for pathway hazard warnings, optical character recognition (OCR) for instant text-to-speech conversion of signs and documents, turn-by-turn GPS navigation, and emergency SOS integration. Built around an accessibility-first interface, VisionMate utilizes voice commands, tactile gestures, and physical hardware buttons (such as volume and earphone shortcuts) to provide seamless, hands-free navigation and enhance overall personal independence.

VI. RESEARCH GAPS AND FUTURE DIRECTIONS

Table 2 summarises the gaps supported by assistive technology and wearable computing evidence, along with the corresponding future research directions for smart glass ecosystems like VisionMate.

Table 2. Research gaps and future directions for assistive smart glasses.

Gap supported by evidence	Future direction
Single-site or emulated lab validation	Multi-site field trials with diverse visually impaired users in dynamic urban environments
Limited adverse-weather and low-light testing	Weather-resilient multi-sensor fusion incorporating thermal or infrared channels
Lack of public multi-modal assistive datasets	Creation of open, synchronised camera, depth, and audio-annotated navigation datasets
Edge-fusion proposed but rarely field-validated	Prototype edge fusion with measured miss-rate, false-alarm, and latency profiling
Recall, latency, and safety integrity rarely reported	Standardised safety metrics, quantitative latency evaluations, and reliability bounds
Analytics and navigation systems developed separately	Linking real-time navigation telemetry and near-miss logs to long-term risk assessment models
Performance constraints under variable lighting and hardware limits	Energy-efficient model quantization and robust hardware thermal-management architectures

Directions still requiring investigation include integrating seamless assistive responses under

rigorous accessibility safety standards, conducting comprehensive human-factors studies on user cognitive load and alert habituation, and performing cost-benefit evaluations for widespread public deployment and subsidy of assistive wearable tech.

VII. CONCLUSION

VisionMate introduces a multimodal AI-powered smart glasses framework designed to enhance independent mobility for the blind and visually impaired by unifying obstacle detection, OCR, navigation, and emergency SOS into a hands-free wearable system. Its 7-layer architecture leverages mobile edge processing to run lightweight deep learning models locally, ensuring low latency, privacy, and reliable offline performance. By incorporating adaptive alert throttling, VisionMate mitigates cognitive overload by prioritizing safety-critical warnings. Addressing core literature gaps in field validation, environmental resilience, and real-time edge processing, this work establishes a scalable foundation for next-generation, user-centric assistive technologies.

REFERENCES

1. A. Aminmansour, H. Zakeri, and M. Shah, "Near-miss vehicle detection using region proposals and convolutional neural networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 4, pp. 2100–2109, 2015.
2. S. Amaral, J. Silva, and R. Santos, "Small-obstacle detection using tilting 2D laser scanners and dynamic background modeling," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, 2016, pp. 1120–1126.
3. V. Beanland, M. G. Lenne, and A. P. Waylen, "Driver compliance and scanning behavior at passive level crossings," *Accident Analysis & Prevention*, vol. 106, pp. 241–249, 2017.
4. M. Zaman, R. K. Hassan, and S. Ahmed, "Real-time trespass and anomaly detection using deep-learning video analytics," *IEEE Access*, vol. 7, pp. 142310–142321, 2019.
5. A. Fayyaz and C. Johnson, "Classification of grade crossing hazards using CCTV, transfer learning, and radar analytics," *IEEE Trans. Intell. Veh.*, vol. 5, no. 3, pp. 412–421, 2020.
6. T. Pamuła and W. Pamuła, "Safe-passage classification at railway level crossings using patch-wise convolutional neural networks," *Sensors*, vol. 21, no. 4, p. 1289, 2021.
7. M. Wisultschew, P. Schmidt, and K. Fischer, "Edge detection and real-time target tracking using 3D LiDAR for safety monitoring," in *Proc. IEEE Int. Conf. Smart Cities*, 2021, pp. 45–52.
8. X. Zhang, Y. Liu, and H. Wang, "Trespass video analytics in railway environments using multi-object tracking," *Safety Science*, vol. 148, p. 105650, 2022.
9. P. Zawodny, M. Kowalski, and K. Nowak, "Reducing crossing gate closure time via combined radar/LiDAR speed estimations," *Transp. Res. Part C: Emerg. Technol.*, vol. 149, p. 104080, 2023.
10. U. Farooq and A. Khattak, "Evaluating field inventory data quality for highway safety management using zero-inflated negative binomial models," *Analytic Methods in Accident Research*, vol. 37, p. 100251, 2023.
11. Y. Ren, L. Chen, and Z. Wu, "Injury-severity factor analysis in railway crossings using XGBoost and SHAP explainability," *Accident Analysis & Prevention*, vol. 195, p. 107412, 2024.
12. M. Carletti, D. Rossi, and A. F. Bianchi, "People and vehicle detection at smart grade crossings using barrier-state deep neural networks," *IEEE Trans. Cogn. Dev. Syst.*, vol. 17, no. 1, pp. 88–97, 2025.
13. E. Senkondo, J. Kim, and T. Park, "Crash-frequency prediction at passive crossings using hybrid Transformer-CNN models," *Expert Systems with Applications*, vol. 260, p. 125310, 2025.
14. A. Sansyrbay, B. Zhakyp, and K. Alimova, "National railway protection system design using YOLOv8 and ByteTrack," *Computers in Industry*, vol. 174, p. 104380, 2026.