engineering proceedings

MDPI

Proceeding Paper

Experimental Evaluation of a Smart Glasses-Based Cyber-Physical System for Dangerous Object Recognition and Dynamic Hazard Detection †

Nikolay Gospodinov * and Georgi Krastev

Department of Computer Systems and Technologies, University of Ruse, 7004 Ruse, Bulgaria; georgk@uni-ruse.bg
* Correspondence: ngospodinov@uni-ruse.bg
† Presented at the International Conference on Electronics, Engineering Physics and Earth Science (EEPES2026), Bandirma, Turkey, 24–27 June 2026.

Abstract

This paper presents an experimental evaluation of a smart glasses-based cyber-physical system for the recognition of dangerous objects and the detection of dynamic life-threatening events. The proposed system integrates real-time computer vision and sound-based localization mechanisms in order to enhance situational awareness and user safety. The dangerous object recognition module is based on a deep learning model derived from the YOLO architecture and MobileNet classifiers, extended with distance estimation capabilities. The dynamic hazard detection module combines auditory localization using interaural time and level differences with visual confirmation through object detection. The experimental study was conducted using datasets of hazardous objects and simulated dynamic scenarios, including moving threats and sudden acoustic events. The dangerous object recognition module achieved F1-scores of up to 100.00% during training and 99.33% during validation, demonstrating high robustness and reliability. The dynamic hazard detection module achieved 0.95 training accuracy and 0.71 validation accuracy under dynamic environmental conditions. Comparative experiments on two hardware platforms showed that the ASUS ROG Zephyrus M16 reduced training time by up to seven times compared to the GIGABYTE GA-A320M-H platform without affecting recognition accuracy. The obtained results confirm the applicability of the proposed multimodal approach for real-time safety systems and wearable assistive technologies operating in complex environments.

Keywords: cyber-physical system; smart glasses; dangerous object recognition; dynamic hazard detection

Check for updates

Academic Editors: Teodor Iliev, Ivaylo Stoyanov, Grigor Mihaylov and Selahattin Kosunalp

Published: 11 September 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Cyber-Physical Systems (CPS) represent a rapidly evolving class of intelligent systems in which physical processes are tightly integrated with computational and communication capabilities. These systems play a critical role in modern applications such as smart environments, transportation, healthcare, and public safety, where real-time perception and response to environmental changes are essential.

In recent years, the integration of CPS with artificial intelligence (AI) and augmented reality (AR) technologies has significantly enhanced their ability to interpret complex real-world scenarios. In particular, wearable devices such as smart glasses provide a promising platform for developing assistive and safety-oriented systems capable of continuous environmental monitoring. A key challenge in such systems is the reliable detection

of potentially dangerous situations in dynamic and unpredictable environments. This includes not only the recognition of hazardous objects (e.g., weapons, fire sources, or suspicious items) but also the identification of dynamic life-threatening events such as approaching objects, sudden movements, or unexpected acoustic signals. Addressing this challenge requires the combination of multiple sensing modalities and robust real-time processing mechanisms.

In our previous studies, we investigated individual components of a smart glasses-based CPS, focusing on object detection using deep learning techniques and on the localization of dynamic life-threatening events using sound-based methods. These works demonstrated the feasibility of deploying recognition and localization mechanisms on wearable platforms while highlighting limitations related to environmental complexity and real-time constraints [1,2].

Building upon these results, the present paper focuses on the experimental evaluation of two advanced modules: (i) dangerous object recognition based on an extended YOLO architecture with distance estimation and (ii) dynamic hazard detection combining auditory localization and visual confirmation. The aim of this study is to assess the performance, robustness, and real-time applicability of the proposed modules in scenarios involving safety-critical conditions.

2. Related Work

Recent advances in computer vision have significantly improved the ability to detect and classify objects in real-time environments. Deep learning-based approaches, particularly those based on convolutional neural networks, have become the dominant paradigm for object detection. Among them, the YOLO (You Only Look Once) family of models has gained widespread adoption due to its balance between accuracy and computational efficiency, making it suitable for real-time applications on embedded and wearable devices [3].

Several studies have explored the application of object detection techniques for identifying hazardous or potentially dangerous objects. These approaches typically rely on large-scale datasets and optimized architectures to achieve high detection accuracy. However, their performance often depends on computational resources and may be limited in constrained environments such as wearable systems.

In parallel, sound-based localization has been investigated as an effective method for detecting and positioning dynamic events in space [4]. Techniques based on interaural time differences (ITD), interaural level differences (ILD), and head-related transfer functions (HRTF) have been widely used to estimate the direction of sound sources and support spatial awareness in intelligent systems [5].

More recent research has focused on multimodal approaches that combine visual and auditory information in order to improve robustness and reliability in dynamic environments. Wearable and smart environment systems have also been widely studied in the context of real-time monitoring, embedded sensing, and assistive applications, confirming the relevance of lightweight AI-based perception mechanisms for cyber-physical safety systems [6–8]. Such systems aim to enhance perception by integrating complementary data sources, enabling more accurate detection of complex real-world events. Despite these advances, challenges remain in achieving real-time performance and reliable operation under varying environmental conditions.

In our previous studies, we investigated individual components of a cyber-physical system based on smart glasses, focusing on object detection using deep learning techniques and on the localization of dynamic life-threatening events using sound-based methods. These works demonstrated the feasibility of deploying such mechanisms on

wearable platforms while revealing limitations related to environmental variability and system responsiveness.

3. System Architecture and Methodology

The proposed cyber-physical system is designed as an intelligent multimodal platform for real-time detection of dangerous objects and dynamic life-threatening events. The system is implemented on augmented reality smart glasses and integrates visual and auditory perception mechanisms in order to enhance situational awareness and user safety.

The overall architecture consists of two main functional modules: (i) a dangerous object recognition module and (ii) a dynamic hazard detection module.

These modules operate in a complementary manner and are integrated through a unified processing pipeline that enables real-time analysis and response (Figure 1).

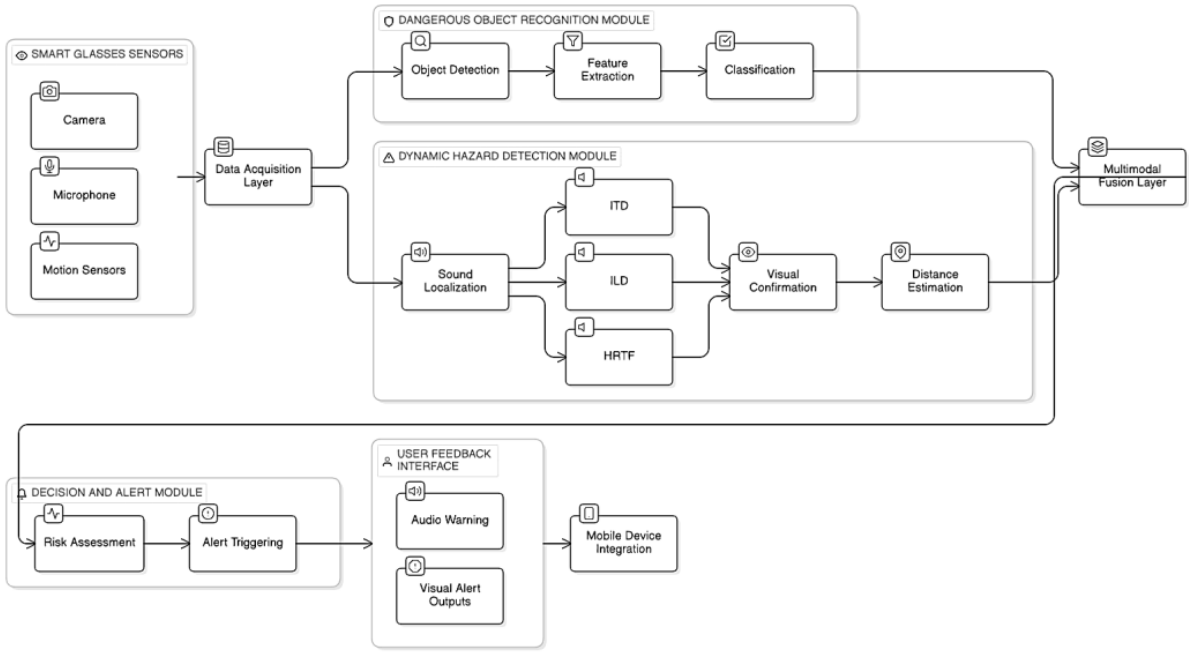


Figure 1. Overall system architecture.

3.1. Dangerous Object Recognition Module

The dangerous object recognition module is responsible for detecting and classifying objects that may pose a threat to the user. The module is based on a deep learning approach utilizing convolutional neural networks for feature extraction and object classification, following modern real-time object detection architectures such as YOLOv1 [3].

The processing pipeline of the module consists of three main stages:

1. Object detection—The camera embedded in the smart glasses continuously captures frames from the surrounding environment. Potential objects are detected within the scene using a real-time object detection mechanism based on bounding box prediction.
2. Object recognition and classification—The detected regions are processed by a trained deep learning model, which extracts relevant features and assigns each object to a predefined class. The classification includes various categories of dangerous objects, such as weapons and hazardous items.
3. User feedback—The recognized object is labeled and communicated to the user through an audio interface, enabling immediate awareness without requiring visual focus.

The module is optimized for real-time operation and deployment on resource-constrained devices, ensuring low latency and efficient processing (Figure 2).

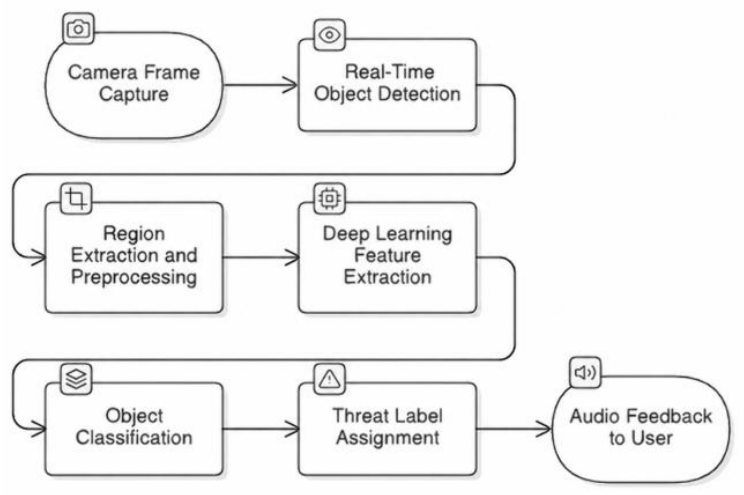


Figure 2. Dangerous object recognition pipeline.

3.2. Dynamic Hazard Detection Module

The dynamic hazard detection module extends the system capabilities by identifying potentially dangerous events occurring in the environment. Unlike static object detection, this module combines auditory and visual information to detect dynamic threats.

The module is composed of three interconnected components:

3.2.1. Sound-Based Localization

The first component detects and localizes sound sources in the environment using binaural auditory cues. The localization process is based on classical acoustic models, including interaural time difference (ITD), interaural level difference (ILD), and head-related transfer function (HRTF), which are widely used in sound source localization systems [4,5]. The ITD is defined as:

$$\Delta t = \frac{\Delta d}{v} \tag{1}$$

where Δt represents the time delay between the two ears, Δd is the path difference, and v is the speed of sound.

The ILD is expressed as:

$$ILD(\alpha, f) = 0.18 \,.\, \sqrt{f} \,.\, \sin(\alpha) \tag{2}$$

where f is the frequency of the sound signal and α is the azimuth angle.

These models enable the estimation of the direction of the sound source in three-dimensional space.

3.2.2. Visual Confirmation and Classification

After the approximate direction of the sound source is determined, the user or system aligns the camera towards the suspected location. The visual module is then activated to detect and classify the object responsible for the sound.

This component utilizes a deep learning-based object detection model to identify the object and determine its spatial position relative to the user. In addition to classification, the module estimates the distance between the user and the detected object, which is critical for assessing the level of risk.

3.2.3. Decision and Alert Mechanism

The final component integrates the auditory and visual information to determine whether a potentially dangerous situation exists.

If the detected object is classified as hazardous and its distance falls below a predefined threshold, the system generates an alert. The alert is delivered through audio signals and visual indicators, ensuring immediate user awareness (Figure 3). This decision-making process enables the system to react to dynamic scenarios in real-time, improving safety in complex and unpredictable environments.

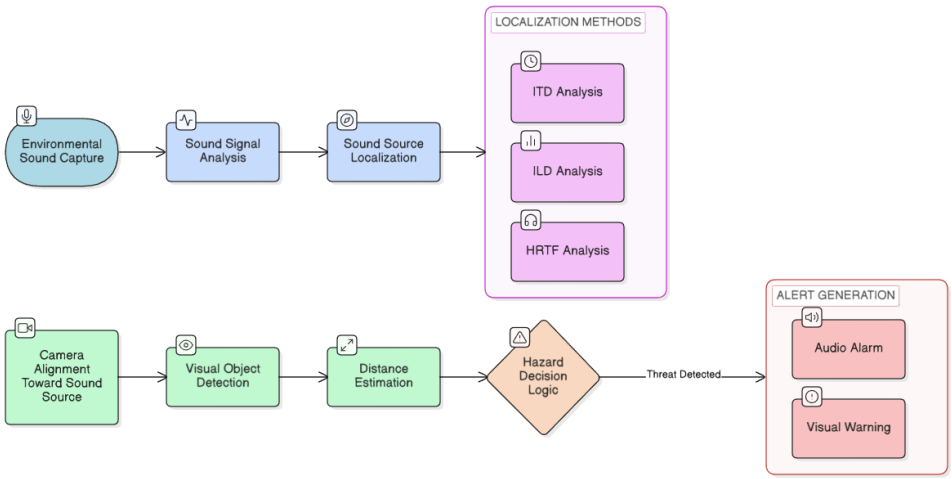


Figure 3. Dynamic hazard detection pipeline.

3.3. Multimodal Integration

A key feature of the proposed system is the integration of visual and auditory perception into a unified framework. The multimodal approach enhances robustness by compensating for the limitations of individual sensing modalities.

While visual detection provides detailed object information, it may be affected by occlusions or lighting conditions. Conversely, sound-based localization enables early detection of events outside the field of view (Table 1). The combination of these modalities allows the system to achieve more reliable and context-aware detection of hazardous situations.

Table 1. Functional comparison of the two modules.

Module	Input Type	Detection Methods	Output
Dangerous Objects	Visual	Deep Learning Detection	Audio warning
Dynamic Hazards	Audio + Visual	Multimodal Localization	Audio + Visual alert

Generative AI tools were used only for linguistic editing and to assist in preparing the schematic illustrations shown in Figures 1–3. DiagramGPT by Eraser was used to generate initial diagram layouts based on the authors' descriptions, which were subsequently reviewed and edited by the authors. All scientific content, methodology, experimental data, analysis, and conclusions were developed and verified by the authors.

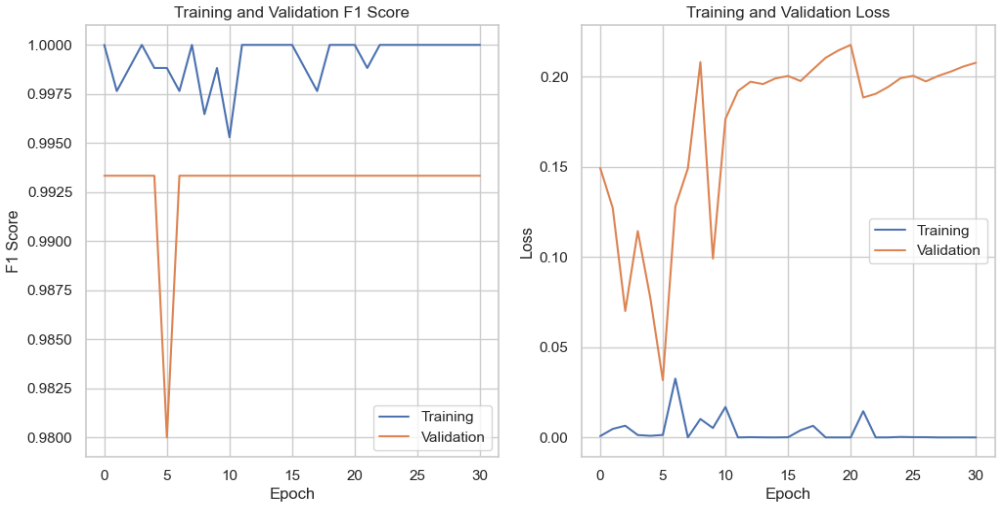
4. Experimental Results

This section presents the experimental evaluation of the two newly developed modules of the proposed cyber-physical safety system: the dangerous object recognition module and the dynamic hazard detection module. The experiments were conducted on two hardware platforms with different computational capabilities: GIGABYTE GA-A320M-H (GIGABYTE Technology Co., Ltd., New Taipei City, Taiwan) and ASUS ROG Zephyrus M16

GU603ZX (ASUSTeK Computer Inc., Taipei, Taiwan). The models were implemented and trained in Python 3.15 using the PyCharm integrated development environment (JetBrains, Prague, Czech Republic). The objective was to assess classification accuracy, F1-score stability, loss convergence, and computational efficiency under real operating conditions.

4.1. Experimental Results of Dangerous Object Recognition Module

The dangerous object recognition module was evaluated using a MobileNetV1-based neural architecture [9] trained on a custom dataset consisting of annotated dangerous-object images. The experiments included two training phases and two additional validation phases on both machines. The model demonstrated exceptionally high performance across all evaluation stages (Figure 4).



(**a**) (**b**)

Figure 4. F1-score (**a**) and loss curves (**b**) of the dangerous object recognition module during validation with additional data.

During Phase I and Phase II training, both machines achieved a perfect F1-score of 100.00%, indicating flawless classification on the training and validation sets. In the additional-data validation phases, both systems maintained very high generalization performance, reaching 99.33% F1-score. The weighted average F1-score remained 0.99, confirming strong robustness across both object classes (Figure 5). The confusion matrix further verified these results, showing only one misclassified sample out of 150 test instances, while 149 samples were classified correctly.

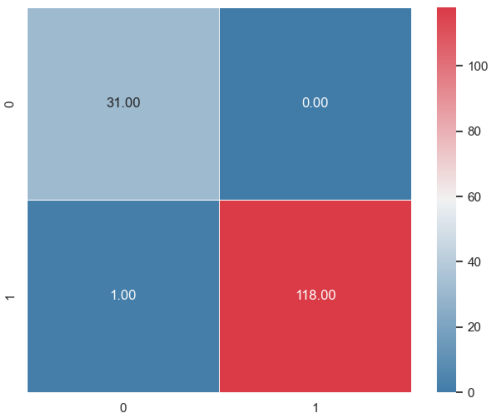


Figure 5. Confusion matrix of the dangerous object recognition module.

As shown in Figure 5, the confusion matrix confirms the high classification reliability of the model, with only one incorrectly classified sample out of 150 test instances.

A significant difference was observed in computational speed. ASUS ROG Zephyrus completed all training phases approximately three times faster than GIGABYTE. For example, in Phase I training ASUS required only 2 min 3 s, compared to 6 min 38 s on GIGABYTE. Similar acceleration was observed in all subsequent phases. Loss values remained extremely low on both machines, with ASUS showing slightly lower loss during the main training phases (Table 2).

Table 2. Performance comparison of dangerous object recognition on the two hardware platforms.

Phase	GIGABYTE Time	ASUS Time	F1-Score
Phase I	6 min 38 s	2 min 03 s	100%
Phase II	7 min 11 s	2 min 21 s	100%
Additional I	4 min 05 s	1 min 17 s	99.33%
Additional II	7 min 16 s	2 min 20 s	99.33%

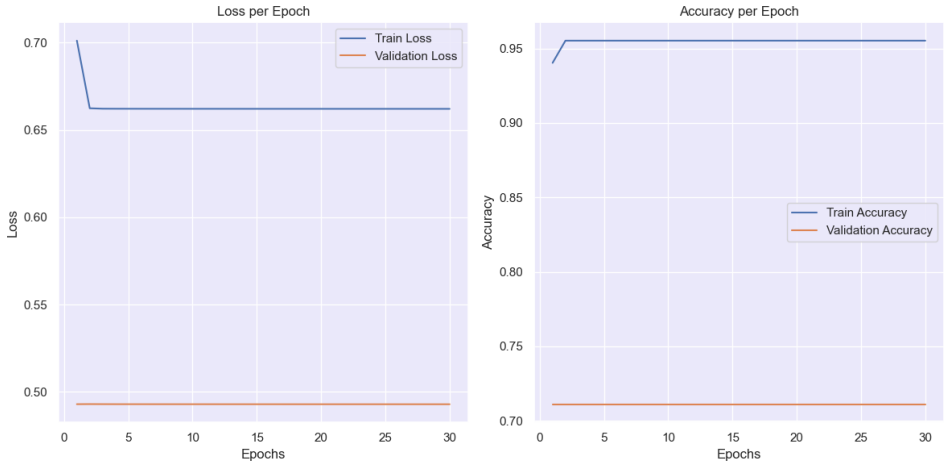
Table 2 clearly demonstrates the computational advantage of ASUS ROG Zephyrus, while both systems preserve identical recognition accuracy.

These results confirm that the MobileNet-based dangerous object recognition module achieves near-perfect recognition accuracy while maintaining excellent computational efficiency, making it highly suitable for deployment in wearable real-time CPS environments.

4.2. Experimental Results of Dynamic Hazard Detection Module

The dynamic hazard detection module was evaluated using a fine-tuned YOLOv1 model trained on a custom dataset for dynamic hazards, following real-time object detection principles introduced in the original YOLO framework [3]. Unlike the dangerous object module, this task proved more challenging due to higher environmental variability and motion complexity (Figure 6). On both machines, the model achieved identical learning performance:

- training accuracy: 0.95
- validation accuracy: 0.71
- training loss: 0.67
- validation loss: 0.49



(**a**) (**b**)

Figure 6. Training loss (**a**) and validation accuracy (**b**) curves for the dynamic hazard detection module.

These results indicate strong learning capability on training data but reduced generalization on unseen validation samples, suggesting partial overfitting. The gap between training and validation accuracy reflects the increased complexity of dynamic hazard recognition under varying real-world conditions. As shown in Figure 6, the divergence between training and validation accuracy becomes visible in later epochs, confirming partial overfitting behavior.

The confusion matrix analysis revealed that the model tends to collapse predictions into a dominant class, indicating insufficient class balance in the dataset and the need for further dataset enrichment. Despite this limitation, the system demonstrates stable convergence behavior (Figure 7).

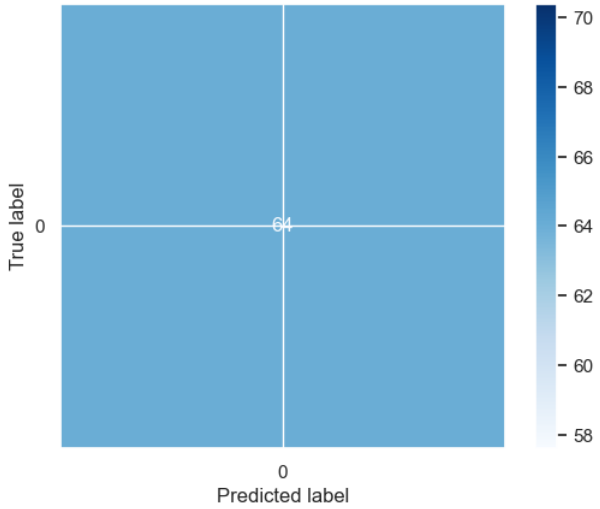


Figure 7. Confusion matrix of the dynamic hazard detection module.

Figure 7 illustrates the class imbalance problem, where predictions are concentrated predominantly in one dominant class.

Hardware performance again differed substantially. ASUS completed training in only 7 min 17 s, whereas GIGABYTE required 52 min 6 s—approximately seven times longer. However, both platforms produced identical predictive accuracy, indicating that hardware affects speed but not model quality (see Table 3).

Table 3. Performance comparison of dynamic hazard detection on the two hardware platforms.

Machine	Training Time	Train Accuracy	Validation Accuracy	Train Loss	Validation Loss
GIGABYTE	52 min 06 s	0.95	0.71	0.67	0.49
ASUS	7 min 17 s	0.95	0.71	0.66	0.49

Table 3 confirms that hardware acceleration significantly reduces training time without affecting model predictive performance.

Although the dynamic hazard detection module exhibits lower validation accuracy than the dangerous object recognition module, it successfully demonstrates the feasibility of real-time multimodal hazard detection in complex dynamic environments.

4.3. Comparative Discussion

The experimental comparison highlights two important findings.

First, the dangerous object recognition module significantly outperforms the dynamic hazard detection module in classification stability and accuracy. This is primarily due to the

simpler and more deterministic visual characteristics of static dangerous objects compared to dynamic hazards involving motion and environmental variability.

Second, hardware acceleration strongly impacts training efficiency. ASUS ROG Zephyrus consistently reduced training time without affecting recognition quality, proving more suitable for iterative development and real-time deployment scenarios.

Overall, the results demonstrate that both modules are functional and effective, but the dynamic hazard detection module requires further optimization through expanded datasets, class balancing, and hyperparameter refinement to improve validation accuracy and generalization performance.

5. Conclusions

This paper presented an experimental evaluation of two advanced modules in a smart glasses-based cyber-physical safety system: dangerous object recognition and dynamic hazard detection. Both modules were developed as part of an intelligent multimodal CPS designed to improve user safety in real-time environments.

The dangerous object recognition module demonstrated near-perfect classification performance, achieving F1-scores of up to 100.00% during training and 99.33% during additional-data validation. These results confirm the high reliability and robustness of the MobileNet-based recognition approach for identifying hazardous objects in wearable CPS applications.

The dynamic hazard detection module, based on a fine-tuned YOLOv1 architecture [3], also showed promising performance, achieving 0.95 training accuracy and stable convergence on both hardware platforms. Although lower validation accuracy (0.71) indicates challenges in generalization under dynamic real-world conditions, the module successfully demonstrates the feasibility of real-time hazard detection in complex environments.

The comparative hardware analysis showed that ASUS ROG Zephyrus M16 significantly reduces training time without affecting recognition accuracy, making it more suitable for computationally intensive CPS development.

Future work will focus on expanding dataset diversity, improving class balance, optimizing hyperparameters, and enhancing multimodal fusion mechanisms in order to increase robustness under complex environmental conditions.

Author Contributions: Conceptualization, G.K.; methodology, G.K.; software, N.G.; validation, N.G. and G.K.; formal analysis, N.G. and G.K.; investigation, N.G. and G.K.; resources, N.G. and G.K.; data curation, N.G. and G.K.; writing—original draft preparation, N.G. and G.K.; writing—review and editing, N.G. and G.K.; visualization, N.G. All authors have read and agreed to the published version of the manuscript.

Funding: This study is financed by the European Union-NextGenerationEU, through the National Recovery and Resilience Plan of the Republic of Bulgaria, project № BGRRP-2.013-0001.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets generated and analyzed during the current study are available from the corresponding author upon reasonable request.

Acknowledgments: During the preparation of this manuscript, the authors used ChatGPT (GPT-5.5) for linguistic editing and improvement of English-language readability. DiagramGPT by Eraser was used to assist in preparing the schematic illustrations in Figures 1–3. The authors reviewed and edited all outputs and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Gospodinov, N.; Krastev, G. Cyber-Physical System for Recognizing Dangerous Objects and Hazardous Substances with Smart Glasses. In Proceedings of the 2022 International Conference on Communications, Information, Electronic and Energy Systems (CIEES), Veliko Tarnovo, Bulgaria, 24–26 November 2022; pp. 1–4. [CrossRef]
2. Gospodinov, N.; Krastev, G. Cyber-physical Smart System for Localization and Protection Against Dynamic Life-threatening Events. In Proceedings of the 2022 30th National Conference with International Participation (TELECOM), Sofia, Bulgaria, 27–28 October 2022; pp. 1–4. [CrossRef]
3. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. *arXiv* **2015**, arXiv:1506.02640. [CrossRef]
4. Pierre-Amaury, G.; Srđan, K.; Laurent, G.; Alexandre, G. A survey of sound source localization with deep learning methods. *J. Acoust. Soc. Am.* **2022**, *152*, 107–151. [CrossRef] [PubMed]
5. Risoud, M.; Jean, N.; Gauvrit, F.; Renard, C.; Lemesre, P.-E.; Bonne, N.; Vincent, C. Sound source localization. *Eur. Ann. Otorhinolaryngol. Head Neck Dis.* **2018**, *135*, 259–264. [CrossRef] [PubMed]
6. Chen, M.; Ma, Y.; Song, J.; Lai, C.F.; Hu, B. Smart Clothing: Connecting Human with Clouds and Big Data for Sustainable Health Monitoring. *Mob. Netw. Appl.* **2016**, *21*, 825–845. [CrossRef]
7. Alam, M.R.; Reaz, M.B.I.; Ali, M.A.M. A Review of Smart Homes—Past, Present, and Future. *IEEE Trans. Syst. Man Cybern. Part C* **2012**, *42*, 1190–1203. [CrossRef]
8. Patel, S.; Park, H.; Bonato, P.; Chan, L.; Rodgers, M. A Review of Wearable Sensors and Systems with Application in Rehabilitation. *J. Neuroeng. Rehabil.* **2012**, *9*, 21. [CrossRef] [PubMed]
9. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv* **2017**, arXiv:1704.04861.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.