Assisting blind people with adaptive, explainable, and multimodal ML systems using smart glasses

Carlos Romano Gómez

University of Essex

Abstract

Machine learning (ML) algorithms have shown promise in the medical field, including assisting blind individuals. However, the lack of explicability in deep learning algorithms poses a challenge, hindering transparency and trust. This proposal proposes a system focusing on interpretability, multimodality, and dynamic adaptability, prioritizing an enhanced user experience. The system will employ a hybrid algorithm combining Coupled Matrix Tensor Factorization (CMTF) and Linear Dynamic Systems (LDS) for object detection, with Probabilistic Latent Component Analysis (PLCA) for decision-making. The objective is an efficient ML system on resource-constrained devices without compromising accuracy.

The algorithm will be integrated into smart glasses, assessing comfort and responsiveness. Evaluation metrics will include ML metrics (accuracy, recall, precision) and user experience metrics. Additionally, comparative analysis against deep learning algorithms will be conducted. While, user experience will be evaluated based on comfort, frequency of system failures in obstacle detection, and reaction speed in notifying users. This comprehensive analysis will enable understanding of system performance and effectiveness in real-world scenarios. The contributions of this research will include developing algorithms addressing deep learning limitations, offering interpretability and adaptability. Bridging the gap between ML and user experience advances the field, enabling more interpretable and user-friendly ML systems to assist blind individuals.

Key words: *Smart Glasses, ML, explicability, LDS, blind people, CMTF, multimodality.*

1 Introduction

In recent years, there has been a notable surge in technological advancements, coinciding with the emergence of innovative technological devices like smart glasses and significant progress in artificial intelligence. These technologies, built upon advance machine learning techniques, such as object and sound recognition, have the potential to make a revolutionary shift in the lives of individuals; particularly benefiting those with disabilities, such as partially or fully blind people.

As new technologies emerge in the field of assistive navigation, it becomes crucial to adapt to the evolving needs of visually impaired people, by providing them with adaptive systems with a great user experience. Some of the main challenges remain to be developing a system built upon dynamic, multimodal, explainable Machine Learning (ML) systems, that can also provide a great user experience. It also needs to consider the trade-off between computational power and accuracy, given that the system needs to operate on a resource constrained device.

To date, many studies failed to address these challenges. One of the reasons is the inherent nature of deep learning algorithms, considered as "Black box systems" due to their nested nonlinear structure and lack of transparency in the decision-making process [24]. Furthermore, these algorithms often struggle to adapt their parameters to new environments beyond the scope of their training data. While traditional ML algorithms are effective in certain applications, they may not possess the necessary ability and explanatory power for assistive device in various environments. Therefore, it becomes imperative exploring alternative ML approaches to overcome these challenges and create more robust and adaptable devices for blind individuals.

In this proposal, I propose system that addresses these challenges by employing a hybrid machine learning system that combines Linear Dynamic Systems (LDS) and Tensor Factorization techniques to ensure adaptability, explainability, and multimodality. The proposed system will be enhanced using Smart glasses to provide a user-friendly, comfortable experience that seamlessly integrates into daily life activities. Leveraging visual and binaural inputs, it will provide a complete perception of the surroundings to the users. In addition, an intelligent algorithm will be incorporated to prioritize and act upon the gathered information, helping users navigate with greater confidence and safety. My research study will aim to address the following questions:

1. Can a dynamic, explainable, multimodal system perform as well as deep learning methods?
2. Can such system efficiently work on a resource constrained device?
3. Can using smart glasses provide a better user experience for blind people?

With this proposal I would like to seek PhD opportunities under the supervision of Dr Sefki Kolozali in the Embedded and Intelligent Systems (EIS) research group. As the research presented in this proposal centers around the development of a smart glasses device that utilizes advanced ML algorithms to aid blind individuals in navigation and obstacle detection, the EIS lab's dedicated focus on ML aligns with the necessary development support and technology required for the implementation of the proposed device. Moreover, the lab's emphasis on advanced systems and architectures specifically tailored for real-time critical applications, including robotics and image processing, holds significance for the advancement of the smart glasses system. Furthermore, it is noteworthy to mention that Sefki Kolozali has previously collaborated with the author as a supervisor for a group project on a related topic and is currently supervising the author's dissertation project.

2 Related Work

2.1 ML challenges

Recent advancements in ML have witnessed the successful application of deep learning algorithms in various domains, including medical applications [11] and autonomous game playing [16]. However, a significant limitation of these algorithms is their lack of interpretability and inability to provide explanations for their decisions, which has raised concerns among researchers. Recognizing this limitation, some studies have proposed ideas to explain the process behind ML algorithms. For instance, [8] introduces the concept of falsification as a central methodology in ML, suggesting that ML models reject hypotheses based on disconfirming evidence and replace them with simpler and more appropriate ones. Consequently, researchers have focused on addressing this challenge by exploring alternative models that offer explanations for the decisions made by ML algorithms [25].

The development of new learning algorithms has been driven by studies like [18] and [6], emphasizing the need for users to have a deeper understanding of how raw data is transformed into outcomes that can affect them positively or negatively. By providing explainable ML that reveal the underlying decision-making process, users can access information about the parameters influencing the algorithm's decisions, thereby increasing trust and transparency. Various approaches have been proposed to enhance the explainability of ML algorithms. These include Feature Importance, which identifies the most influential features in the decision-making process, allowing users to comprehend the key factors driving the algorithm's predictions [4]. Counter-Factual Explanations provide alternative scenarios that demonstrate how changes in input variables affect the output, aiding in understanding the causal relationships within the model [20]. Adversarial training, on the other hand, involves training a model on both clean data and adversarial examples, which are modified versions of the original data designed to cause misclassifications or degrade performance [13]. Nevertheless, these approaches does not satisfy the need in its entirely, needing more research in the area.

Other of the many challenges of ML, is the curse of dimensionality when dealing with high-dimensional data. The curse of dimensionality refers to the phenomenon where the performance of certain ML algorithms deteriorates as the number of features or dimensions in the data increases. To address this limitation, [9] proposes the use of LDS, as they represent the evolution of a system over time through a set of state variables and linear equations, offering a solution to the curse of dimensionality in certain contexts. In [22], LDS is described as a mathematical model that can determine a system using initial variable values.

Furthermore, the utilisation of multiset learning techniques, such as, Coupled Matrix Tensor Factorization (CMTF) can help substantially analyse, matrix and tensor datasets. Since, data analysis from multiple sources requires handling of data sets of different orders, this approach makes a clustering for each different class, where every relation is represented as a matrix or tensor [2]. Due to this coupled nature and ability to fuse datasets, CMTF has been applied in multiple domains such as object detection [17], multiway clustering [5] and path prediction [10]. Although the combination of LDS and CMTF has not been specifically tested, it is expected that their integration could enhance the tracking and precision capabilities of the proposed system.

2.2 Assitive navigation challenges

Assistive navigation approaches often explore the use of integrated camera and computer vision systems, such as Convolutional Neural Networks (CNN), to recognize objects and provide alerts to users. Studies like [3], [14] and [23], propose systems that employ smart glasses equipped with stereo cameras, capable of detecting multiple objects encountered in daily life, with an average of 9 objects. While they offer a comfortable user experience, they rely on pretrained networks, limiting their applicability to specific use cases for blind individuals. Furthermore, the systems require considerable computational power, while achieving accuracies of only 69%, 76% and 72%, respectively, which falls short of the ideal performance required for assisting blind individuals. Therefore, it is essential for these studies to strike a balance between computational power and accuracy to enhance the device's performance.

The white cane is a commonly used device among blind individuals, which has been enhanced with innovative software to improve functionality and usability. Several studies [21], [26], [12], and [19], have explored the integration of ultrasonic sensors, Global Positioning Systems (GPS), water detection systems, and even mono cameras integrated into the white cane. These endeavours have resulted in the development of basic navigation aids that are reliable and affordable. However, these systems primarily rely on common stimuli, such as detecting water on the floor or obstacles at non-head levels, thereby failing to detect obstacles beyond these limited parameters. The restricted detection range of these systems presents a significant disadvantage, as it overlooks a crucial aspect of obstacle detection and avoidance.

To overcome these limitations, future research and development needs to prioritize revising assistive technologies for blind people by developing dynamic, multimodal, and explainable ML systems with superior user experience. By addressing these challenges, the proposed device will exhibit enhanced efficiency, adaptability to new environments, and the ability to meet the diverse needs of users in real time. The integration of multiple modalities, such as visual and auditory, will not only augment the user experience but also provide a comprehensive understanding of the environment. Furthermore, ensuring the explainability of the system will enable users and service providers to comprehend the rationale behind its actions, thereby fostering trust and usability.

3 Methodology

The primary objective of this project is to develop an advanced system that effectively assists blind people in navigating and avoiding obstacles. The system will be designed with four key focus areas: user experience, adaptability, explicability, and multimodality. To achieve the user experience, the system will use a pair of smart glasses capable of providing real-time feedback to the user about the direction and distance of objects in their surroundings. This approach ensures portability and comfort for the user. The adaptability aspect will be tackled through the implementation of a hybrid ML system.

3.1 System architecture

To realize the proposed, the device will have four main structures: Vision, binaural, attention and navigation systems. The system structure is presented in Figure 1. The device will hold the components for the system's functionality with a custom 3D printed pair of glasses frame. The main component will be a stereo depth camera for object detection and distance calculation. Strategically positioned microphones will enable binaural perception of object direction.

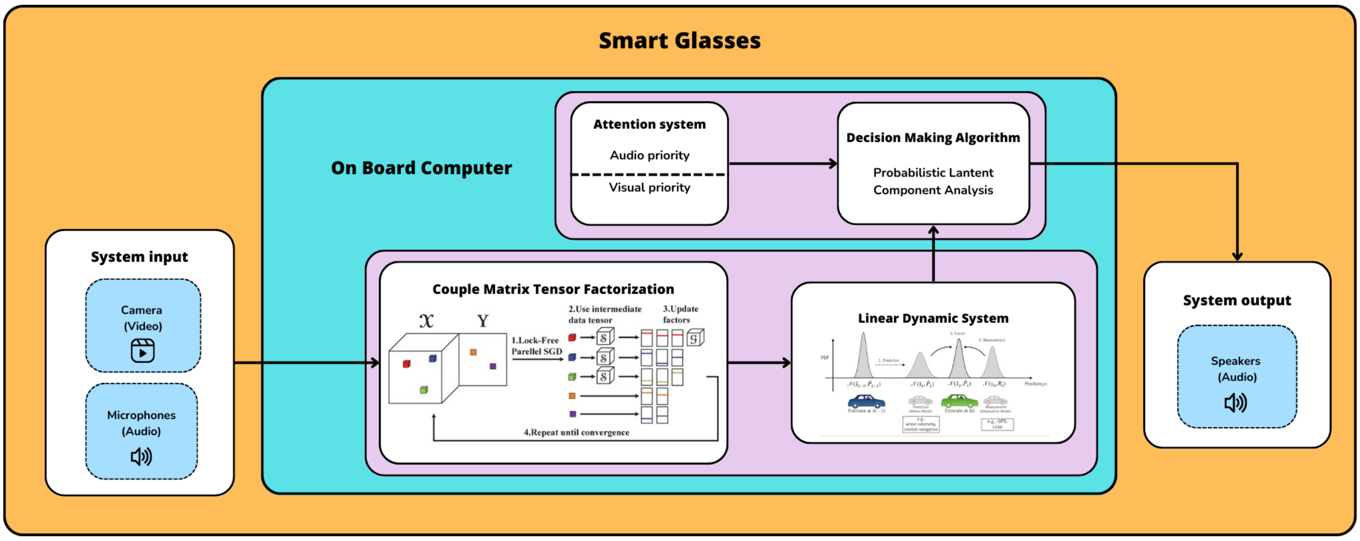

Figure 1: System architecture. It system is divided into 3 components: Input/Output, Decision making and Audiovisual cognition. Each is represented in a different box level.

The integration of the ZED Mini camera and microphones on each side of the 3D printed frame is achieved through precise Computer Aid Design (CAD) modelling to ensure a design that achieves an effective, comfortable, and user-friendly solution for blind individuals, aiding their navigation and obstacle avoidance.

In the other hand, to balance computational power and portability, the system incorporates the Jetson Orin Nano single board computer. This miniaturized computer offers significant computational capabilities while maintaining a compact form. The increased computational power provided by the Jetson Orin Nano enhances the overall efficiency of the system, enabling it to handle complex calculations and deliver real-time responses. Furthermore, it also ensures compatibility with the camera and support for ML algorithm development facilitate the system's implementation.

3.2 Audiovisual cognition component: CMTF & LDS

To accomplish a dynamic, explainable, and multimodal assistive system, the proposed system incorporates LDS in the form of a Kalman filter, which is a recursive filter used to predict future system states based on historical data. This approach is motivated by the findings of [15], which highlight the enhanced accuracy achieved by integrating the Kalman filter in distance prediction within the system. Additionally, the system will be integrated with a CMTF algorithm, which enables the analysis of audiovisual data. This approach facilitates the analysis of different temporal orders of evolving data, enhancing the system's capability to process and interpret the multimodal information [1]. Moreover, a comparison between the proposed system and deep learning algorithms, such as Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) like Long Short-Term Memory (LSTM), will be made to determine the superiority of the hybrid algorithm.

3.3 Decision making component: PLCA & Attention list

The system will use an advance algorithm, Probabilistic Latent Component Analysis (PLCA), to determine the appropriate action based on the detected objects. The attention system compiles object information into a priority list, while the navigation system's decision-making algorithm provides advice to the user. Furthermore, the PLCA algorithm, as a probabilistic model, it offers a cost-effective assessment framework [7]. By integrating these components, the proposed system aims to enhance dynamicity, enabling effective assistance for visually impaired users across diverse situations and environments.

4 Evaluation metrics

The evaluation of the proposed model and its integration will encompass two primary metrics: machine learning (ML) metrics and user experience metrics. ML metrics will focus on algorithm-related measurements, specifically accuracy, recall, and precision, which provide insights into the model's performance in terms of true and false positive values. This evaluation aims to determine whether the hybrid algorithm can achieve comparable performance to deep learning methods.

Meanwhile, the user experience evaluation, will consider several metrics. Firstly, the comfortability of the system during usage will be assessed by comparing its weight to that of other smart glasses. The weight of the device is crucial for prolonged use by blind individuals, as a lighter and more comfortable system will result in an improved user experience.

Additionally, the frequency of how many times the system fails to inform the user about obstacles will be measured. This metric aims to evaluate the system's reliability in detecting and notifying the user of obstacles in the surroundings. Furthermore, the reaction speed of the system in notifying the user about detected obstacles will be considered. A fast and efficient notification system can significantly enhance the user's sense of awareness and facilitate prompt action. These metrics will provide an understanding of the system's effectiveness and responsiveness in real-world scenarios, contributing to the overall user experience evaluation.

By employing these evaluation metrics, the model's performance in terms of accuracy, F1 score, user comfort, system reliability, and reaction speed will be comprehensively assessed, enabling a comprehensive analysis of its integration and effectiveness.

5 Conclusions

The application of ML algorithms in medical domains, including assisting blind individuals, has shown promising results. However, a significant challenge lies in the lack of explicability in current deep learning algorithms. Understanding the inner workings and reasons to decision-making of these algorithms is crucial for ensuring transparency and building trust with users. Therefore, there is a need for developing approaches that provide explicability in ML models, thereby facilitating their application in the medical field and improving outcomes for blind individuals.

In response to these challenges, this proposal proposes the development of a system that focuses on interpretability, multimodality, and dynamic adaptability, while also prioritizing an enhanced user experience. The proposed system will employ a hybrid algorithm that combines CMTF and LDS for object detection, along with a PLCA algorithm for decision-making. This combination of algorithms aims to address the limitations of current deep learning models by providing a more interpretable and adaptable system.

To evaluate the performance of the proposed algorithm, it will be integrated into smart glasses, which serve as the interface for the user. The evaluation will consider factors such as comfort and responsiveness to ensure a positive user experience. By integrating the algorithm into smart glasses, the system can be assessed in real-world scenarios, replicating the conditions that blind individuals encounter in their daily lives.

The goal of this research is to demonstrate that the hybrid system can achieve efficiency on a resource-constrained device without compromising accuracy. The contributions of this research lie in exploring new algorithms that improve upon or surpass existing approaches in terms of multimodality, adaptiveness of parameters and explicability. By addressing the limitations of current deep learning algorithms, this research seeks to advance the field and provide a foundation for the development of more interpretable and user-friendly ML systems for assisting blind individuals.

References

[1] E. Acar, D. M. Dunlavy, and T. G. Kolda. Link prediction on evolving data using matrix and tensor factorizations. In *2009 IEEE International Conference on Data Mining Workshops*, pages 262–269, 2009.

[2] E. Acar, T. G. Kolda, and D. M. Dunlavy. All-at-once optimization for coupled matrix and tensor factorizations, 2011.

[3] F. Ashiq, M. Asif, M. B. Ahmad, S. Zafar, K. Masood, T. Mahmood, M. T. Mahmood, and I. H. Lee. Cnn-based object recognition and tracking system to assist visually impaired people. *IEEE Access*, 10:14819–14834, 2022.

[4] D. Baehrens, T. Schroeter, S. Harmeling, M. Kawanabe, K. Hansen, and K.-R. Müller. How to explain individual classification decisions, 2010.

[5] A. Banerjee, S. Basu, and S. Merugu. Multi-way clustering on relation graphs. *Proceedings of the 2007 SIAM International Conference on Data Mining*, 2007.

[6] U. Bhatt, A. Xiang, S. Sharma, A. Weller, A. Taly, Y. Jia, J. Ghosh, R. Puri, J. M. Moura, and P. Eckersley. Explainable machine learning in deployment. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 2020.

[7] M. Brannstrom, F. Sandblom, and L. Hammarstrand. A probabilistic framework for decision-making in collision avoidance systems. *IEEE Transactions on Intelligent Transportation Systems*, 14(2):637–648, 2013.

[8] O. Buchholz and E. Raidl. A falsificationist account of artificial neural networks. *The British Journal for the Philosophy of Science*, 2022.

[9] W. E. A proposal on machine learning via dynamical systems. *Communications in Mathematics and Statistics*, 5(1):1–11, 2017.

[10] B. Ermiş, E. Acar, and A. T. Cemgil. Link prediction via generalized coupled tensor factorisation, 2012.

[11] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun. Dermatologist-level classification of skin cancer with deep neural networks, Jan 2017.

[12] D. E. Gbenga, A. I. Shani, and A. L. Adekunle. Smart walking stick for visually impaired people using ultrasonic sensors and arduino. *International Journal of Engineering and Technology*, 9(5):3435–3447, 2017.

[13] A. Ilyas, S. Santurkar, D. Tsipras, L. Engstrom, B. Tran, and A. Madry. Adversarial examples are not bugs, they are features, 2019.

[14] B. Jiang, J. Yang, Z. Lv, and H. Song. Wearable vision assistance system based on binocular sensors for visually impaired users. *IEEE Internet of Things Journal*, 6(2):1375–1383, 2019.

[15] D. Kalita and P. Lyakhov. Moving object detection based on a combination of kalman filter and median filtering. *Big Data and Cognitive Computing*, 6(4):142, 2022.

[16] J. H. Lee, J. Shin, and M. J. Realff. Machine learning: Overview of the recent progresses and implications for the process systems engineering field. *Computers & amp; Chemical Engineering*, 114:111–121, 2018.

[17] Y.-R. Lin, J. Sun, P. Castro, R. Konuru, H. Sundaram, and A. Kelliher. Metafac. *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2009.

[18] S. M. Mathews. Explainable artificial intelligence applications in nlp, biomedical, and malware classification: A literature review. *Advances in Intelligent Systems and Computing*, page 1269–1292, 2019.

[19] V. V. Meshram, K. Patil, V. A. Meshram, and F. C. Shu. An astute assistive device for mobility and object recognition for visually impaired people. *IEEE Transactions on Human-Machine Systems*, 49(5):449–460, 2019.

[20] P. Mishra. Counterfactual explanations for xai models. *Practical Explainable AI Using Python*, page 265–278, 2021.

[21] N. M. Nazri, S. S. Fauzi, R. A. Gining, T. R. Razak, and M. N. Jamaluddin. Smart cane for visually impaired with obstacle, water detection and gps. *International Journal of Computing and Digital Systems*, 10(1):1347–1354, 2021.

[22] J. Ouaknine. Specification and verification of linear dynamical systems: Advances and challenges. *Frontiers of Combining Systems*, page 197–197, 2013.

[23] M. Poggi and S. Mattoccia. A wearable mobility aid for the visually impaired based on embedded 3d vision and deep learning. *2016 IEEE Symposium on Computers and Communication (ISCC)*, 2016.

[24] W. Samek and K.-R. Müller. Towards explainable artificial intelligence. *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*, page 5–22, 2019.

[25] M. Turek. Explainable artificial intelligence (xai) (archived), Apr 2019.

[26] N. U. Design of smart cane with integrated camera module for visually impaired people. *International Journal for Research in Applied Science and Engineering Technology*, 10(6):3387–3391, 2022.