Explainable fatigue detection in assembly tasks through graph neural networks

Vishwesh Vishwesh[1], Maximillian Becker[2*], Pascal Birnstill[1], and Jürgen Beyerer[1,2]

[1] Fraunhofer IOSB, Karlsruhe, Germany
[2] Vision and Fusion Laboratory, Karlsruhe Institute of Technology, Karlsruhe, Germany
*maximilian.becker@kit.edu

Abstract Fatigue during assembly tasks can have a negative effect on subjective as well as objective quality of work. We recorded a novel dataset for the purpose of detecting fatigue in assembly scenarios. Participants were instructed to assemble and disassemble model cars with the help of a robot arm. The recordings consist of video, depth video, EEG and eye tracking data as well as questionnaires on the participants' fatigue. The dataset can be provided to researchers on demand. In addition to recording a dataset, we implemented a proof of concept system to detect fatigue solely on image data. In our approach the eye tracking data was used to label the participants' fatigue. Afterwards, a graph neural network was trained on poses extracted from the video data and the generated labels. The classifications of the model are made transparent through the use of explainable AI using saliency maps and GradCAM. This work can have a positive impact on human-machine interaction and assistance systems. Through explainability, we aim to increase the acceptance of such systems by workers and industries.

Keywords GNN, XAI, GradCAM, saliency maps, fatigue detection, dataset

1 Introduction

As the pace of industrial work intensifies, understanding and mitigating the effects of fatigue on human performance has emerged as a challenge. Fatigue can significantly impair cognitive and physical capabilities, leading to reduced productivity, increased error rates and potentially hazardous working conditions. Therefore, detecting fatigue is essential for enhancing productivity and improving the safety and comfort of workers. The goal of this work is to develop a system that can accurately detect fatigue levels in workers during assembly tasks and provide simple explanations for the decisions made by the system. By doing so, we aim to contribute to the development of more adaptive and worker-friendly industrial environments that are optimized for both efficiency and safety.

To explore this issue, we recorded a dataset where participants performed assembly tasks in a controlled environment. We used two modalities of the dataset: eye tracking and video data. The eye tracking data is employed to generate fatigue labels for each timestamp with pupil diameter variability (PDV) as the indicator, which has been empirically validated as a reliable marker of overall fatigue [1]. For the fatigue detection we only use video data. Video data does not disrupt the worker as opposed to wearable sensors and is often already available as it is needed for many assistance systems. On the video data, pose estimation is performed using Mediapipe [2], a tool that extracts human poses from video frames. The resulting pose data is then used to train a Graph Convolution Network (GCN), which is designed to predict fatigue levels based on body posture.

Incorporating transparency into the decision-making process of AI systems is critical, particularly in industrial contexts where the acceptance and trust in assistance systems are paramount. Furthermore, the European AI Act demands transparency if AI systems are used in ”work-related relationships [...] to allocate tasks” and ”monitor and evaluate the performance and behaviour of persons” (Annex III, 4 b) [3]. To address this, our system integrates Explainable artificial intelligence (XAI) techniques to provide local explanations for its decisions.

The primary contribution of this research is the development of a fatigue detection system that integrates deep learning methodologies with XAI techniques while operating only on camera data. This work has the potential to enhance the quality of work environments by fostering transparency and trust in AI-driven assistance systems and Industry 4.0.

2 Related work

Fatigue detection has become an area of increasing interest due to its wide-ranging applications, from workplace safety to medical diagnostics. Various techniques have been employed to capture and assess fatigue levels, each offering unique advantages depending on the domain and context of usage. In this section, we explore different approaches to fatigue detection, from conventional methods like eye tracking to more recent advancements involving pose detection and XAI.

One of the widely used methods for fatigue detection is eye tracking, particularly in domains like automotive safety and air traffic control. By measuring parameters such as blink rate, saccadic movement, and gaze patterns, researchers have been able to infer levels of cognitive and physical fatigue. Benedetto et al. [4] demonstrated the correlation between eye blink frequency and driver fatigue in simulated driving environments. Di Stasi et al. [5] leveraged saccadic velocity to evaluate cognitive load and fatigue. Lengenfelder et al. [6] observed mental fatigue from eye tracking while performing interactive image exploitation. Sirois et al. [7] showed that pupil dilation responds to task difficulty and cognitive effort, reinforcing the role of pupil diameter variability (PDV) in fatigue detection. However, these methods, while effective, are often constrained by environmental factors and require specialized, obtrusive hardware, limiting their adaptation and applications.

Opposed to eye trackers, which are highly specialized, nearly any camera can be used for facial recognition and pose detection. Facial recognition techniques leverage the subtle changes in facial expressions and muscle movements that occur as fatigue sets in. For instance, Haque et al. [8] argued that features like drooping eyelids, yawning frequency, and overall facial muscle relaxation can serve as strong indicators of fatigue. In driver monitoring systems, facial recognition has been applied to track drowsiness and fatigue by detecting changes in eye closure duration, blink frequency, and facial muscle slackness, as demonstrated in studies by Bergasa et al. [9], Ji et al. [10] and García et al. [11]. Similarly, Sikander et al. [12] and Liu et al. [13] explored the use of facial landmarks in real-time monitoring systems to detect early signs of cognitive and physical fatigue in drivers.

Pose detection has traditionally been used in fields such as sports science [14] and rehabilitation [15], but its recent application in health monitoring has gained traction. The rise of pose estimation libraries like OpenPose [16] and Mediapipe [2] have made this approach more accessible, enabling the detection of joint coordinates in real-time using just standard cameras. Hawley et al. [17] demonstrated using machine learning that postural sway and joint angle deviations could be used as indicators of physical fatigue in lifting tasks. Similarly, Wang et al. [18] used pose estimation in athletic assistance system by incorporating deep learning methods. Strain and fatigue are detected for risk analysis by Papoutsakis et al. [19] in an industrial environment using pose estimation. This paper aims to provide a feedback system to industry workers on safe and unsafe poses while working. Pose-based methods offer the advantage of being non-invasive and relatively inexpensive making them attractive for broader deployment [20]. There exist many more techniques for fatigue detection like EEG, body-borne sensors of physiological markers. They do however require specialized, wearable hardware for every worker.

XAI's role in fatigue detection is particularly crucial because of the need for trust and validation in AI-driven decisions. Rivera et al. [21] detect mental fatigue using EEG data and deploy XAI techniques to interpret the results. They argue that applying deep learning techniques to detect fatigue levels is of limited use and a thorough XAI technique needs to be implemented. Hussain et al. [22] demonstrated how XAI could be used in cognitive fatigue detection using EEG to highlight the importance of specific brainwave patterns, allowing healthcare professionals to validate the AI's interpretation of EEG signals. The potential for XAI in fatigue detection systems is growing, but research is still in its infancy, with most efforts focused on improving prediction accuracy rather than interpretability. As fatigue detection systems are increasingly integrated into workplaces and healthcare, ensuring that their decisions are explainable will become essential for achieving broader acceptance and fulfilling regulatory requirements.

3 Dataset

We started our work by recording a novel dataset for fatigue detection in assembly tasks. The dataset is multimodal, containing EEG, eye

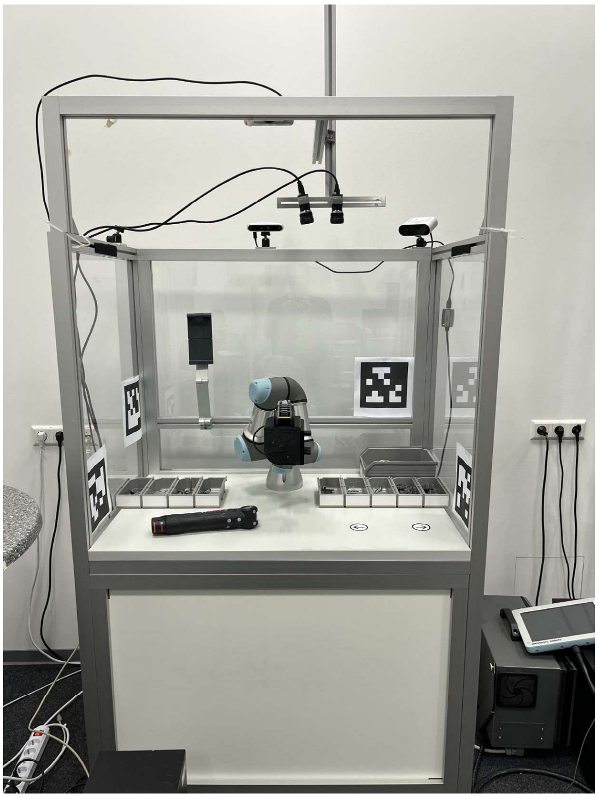

Figure 1: Setup of the assembly table during recording.

tracking, video and depth video data. Additionally, we gathered data from questionnaires that participants had to fill out before and after the experiment. These include the NASA-TLX [23] after the experiment and participants' fatigue on the rating of fatigue (ROF) scale [24] before and after the experiment. The ROF scale is a scale from 0-10 with 0 indicating no fatigue at all and 10 indicating total fatigue and exhaustion.

We invited 30 participants to the recordings. 5 of them participated 3 times each resulting in 40 total recordings. During the experiment, the participants wore an EEG-headset, eye tracking glasses and were recorded with a regular RGB and a stereoscopic depth camera. Pupil Labs Core[1] was used as the eye tracker. To simulate an assembly task participants were asked to first assemble and then disassemble 3 model cars form 3D-printed components. A monitor showed step by step

[1] `https://pupil-labs.com/products/core`

instructions which the participants could control via buttons. Help was also provided by a robot arm that held the partly assembled model cars in place. The setup of the assembly table can be seen in Figure 1.

The dataset will not be publicly available but can be provided to researchers on demand.

4 Explainable Fatigue Detection

Our approach to fatigue detection, once trained, relies only on camera data to predict fatigue levels. It builds on the existing research in fatigue detection and pose estimation, but it introduces a novel combination of these fields using XAI. We begin by extracting key labels from the eye tracking data. While several features are available from eye tracking systems, PDV has been selected as our feature of choice due to its established correlation with cognitive load and fatigue. PDV offers an intuitive measure of how the eye's pupil reacts to changes in focus and brightness, which is often a strong indicator of mental fatigue.

The PDV is calculated using standard algorithms that compute the pupil diameter based on frames obtained from the eye tracker. These frames are timestamped, and the change in pupil size over time is measured to yield the PDV. When labeling the data for fatigue detection, a rolling window approach was utilized, assigning fatigue scores based on a 0-5 scale to reflect varying fatigue intensities.

For extracting human poses, RGBD data from our dataset was used. We used Mediapipe, a state-of-the-art library for pose estimation, which provides 33 3D skeletal keypoints of the participants. Each joint comes with X, Y, Z coordinates (representing spatial location) and a visibility score (indicating how clearly the joint was visible in the frame). As the participants were recorded from the front while standing at an assembly table their legs were not visible. Therefore, we removed joints below hips during preprocessing to avoid noisy data.

Once the pose data was extracted, we aligned them with the previously calculated labels. This allowed us to use supervised learning using Graph Convolutional Networks (GCNs). The GCN consisted of three convolutional layers. It was tasked with predicting the fatigue level of a participant based on their pose. To improve the model's robustness, we experimented with several preprocessing steps, such as balancing the dataset using SMOTEENN, which addressed the issue of class imbalance by combining oversampling of minority classes and under-sampling of majority classes. This technique has proven useful in ensuring that the model does not overfit to the dominant classes while maintaining sufficient samples for the minority classes [25].

To make the model explainable we applied two XAI techniques: saliency maps and Grad-CAM. These methods provided insights into which keypoints (joints) and indirectly which skeletal connections were most influential in determining fatigue levels. These XAI techniques were instrumental in validating that the model was focusing on anatomically relevant areas, aligning with known indicators of physical fatigue [26] [27].

We developed a comprehensive system for detecting fatigue based on video data. By integrating pose estimation, graph-based learning models and leveraging XAI techniques, our method enables better interpretability, which is crucial for identifying key factors contributing to fatigue prediction. Our model lays a strong foundation for future improvements that could enhance its practical applicability with further optimization.

5 Results and discussion

In this section, we will provide a detailed analysis of the outcomes from our experiments, starting with model performance improvements, followed by XAI applications to interpret model predictions using saliency maps and Grad-CAM. Finally, we will demonstrate our XAI techniques in skeletal visualizations with heatmaps, showing how various nodes and skeletal joints contribute to the fatigue prediction.

We started with a baseline model, consisting of three convolution layers, which was trained using a stationary window of 30ms, relying solely on x, y, and z coordinates as features. This model achieved a 54% testing accuracy. To improve performance, we introduced a rolling window of 1.5 seconds, added visibility as a fourth feature, incorporated a dropout layer for regularization and a learning rate scheduler for dynamic optimization. This raised the testing accuracy to 67%.

The next steps involved increasing the convolution layers to five, which further pushed the accuracy to 70%. Early stopping is added to monitor and prevent overfitting and ensure better generalization to test data. Further, we introduced a preprocessing step on the skeleton by removing joints below the hips as they often were often hidden by the assembly table which reduced the skeleton from 33 to 24 joints. When the model was trained on this data, a testing accuracy of 77% was achieved.

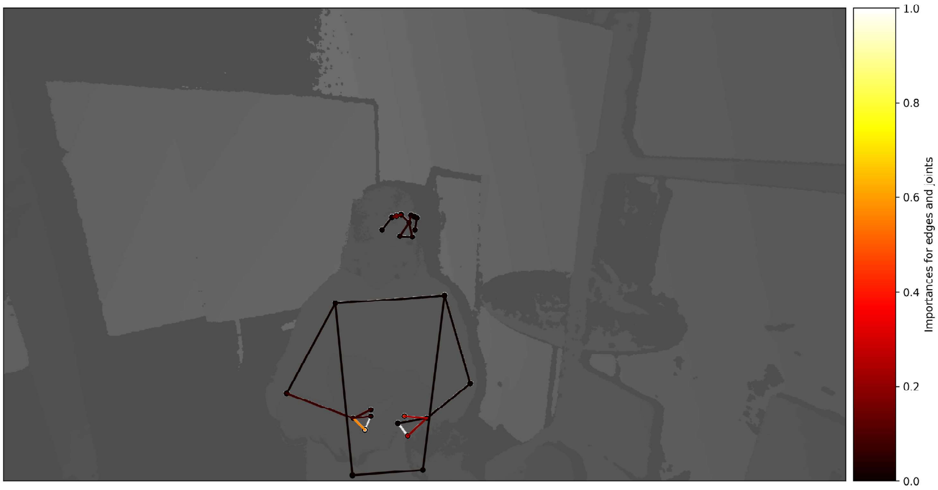

Figure 2: Heatmap of correctly predicted label 0 on depth image.

Figure 3: Heatmap of correctly predicted label 5 on depth image.

We initially trained with ten labels (1-10) but reduced them to 6 (0-5). By reducing the number of fatigue labels from 10 to 6, the model achieved an accuracy improvement from 77% to 80%. By reducing the number of classes, the impact of label noise is reduced, particularly in cases where subjective assessments of fatigue might be inconsistent between adjacent levels. Additionally, simpler categorizations can be more easily understood by non-technical users, leading the user to make better decisions that are more aligned with practical application [28].

As mentioned in the previous chapter we incorporated XAI techniques in the form of saliency maps and Grad-CAM to make our approach more transparent. The saliency maps highlighted the importance of joints such as the shoulders and elbows, which tend to show signs of fatigue during manual tasks. Grad-CAM, on the other hand, visualized the influence of broader skeletal regions, showing how postural deviations in the upper body contributed to the model's predictions. The combined saliency and Grad-CAM visualizations offer a detailed insight into how different parts of the body contribute to fatigue prediction. Figure 2, representing a correctly predicted label of 0 (low fatigue), shows a higher importance around the hands, particularly in the wrist and elbow regions, showing that these regions are indicative of low fatigue levels. Figure 3, which was correctly classified as a 5 (high fatigue), shows a broader spread of important regions, with higher intensity around both the upper body and shoulders, suggesting some reliance on the upper limbs as fatigue increases. However the most important regions are still the hands. In future, we plan to enhance the fatigue prediction model by integrating additional features, such as temporal data from video streams. We also aim to explore advanced explainability techniques to gain deeper insights into the factors influencing fatigue levels.

6 Conclusion

Recognising fatigue in assembly environments is an issue of work safety. We presented an explainable fatigue detection system that works only on image data. Workers do not have to wear any additional devices or sensors hindering them in their work. The video data for our system can stem from cameras that are often already present for assistance systems. Additionally, we incorporated explainability into our system through the use of saliency maps and Grad-CAM. This makes our system more transparent and helps to comply with the European AI Act which demands transparency when monitoring people in work environments. We would like to build on our existing system and develop it into a real-time assistance system.

Acknowledgment

This work was supported by funding from the topic Engineering Secure Systems of the Helmholtz Association (HGF) and by KASTEL Security Research Labs.

References

1. M. A. Boksem and M. Tops, "Mental fatigue: costs and benefits," *Brain research reviews*, vol. 59, no. 1, pp. 125–139, 2008.
2. C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C.-L. Chang, M. G. Yong, J. Lee *et al.*, "Mediapipe: A framework for building perception pipelines," *arXiv preprint arXiv:1906.08172*, 2019.
3. E. Commission, "Regulation (eu) 2024/1689 of the european parliament and of the council of 13 june 2024 laying down harmonised rules on artificial intelligence and amending regulations (ec) no 300/2008, (eu) no 167/2013, (eu) no 168/2013, (eu) 2018/858, (eu) 2018/1139 and (eu) 2019/2144 and directives 2014/90/eu, (eu) 2016/797 and (eu) 2020/1828 (artificial intelligence act)." [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689
4. S. Benedetto, M. Pedrotti, L. Minin, T. Baccino, A. Re, and R. Montanari, "Driver workload and eye blink duration," *Transportation research part F: traffic psychology and behaviour*, vol. 14, no. 3, pp. 199–208, 2011.
5. L. L. Di Stasi, M. B. McCamy, S. L. Macknik, J. A. Mankin, N. Hooft, A. Catena, and S. Martinez-Conde, "Saccadic eye movement metrics reflect surgical residents' fatigue," *Annals of surgery*, vol. 259, no. 4, pp. 824–829, 2014.

6. C. Lengenfelder, J. Hild, M. Voit, and E. Peinsipp-Byma, "Pilot study on gaze-based mental fatigue detection during interactive image exploitation," in *International Conference on Human-Computer Interaction*. Springer, 2023, pp. 109–119.
7. S. Sirois and J. Brisson, "Pupillometry," *Wiley Interdisciplinary Reviews: Cognitive Science*, vol. 5, no. 6, pp. 679–692, 2014.
8. M. A. Haque, R. Irani, K. Nasrollahi, and T. B. Moeslund, "Facial video-based detection of physical fatigue for maximal muscle activity," *IET Computer Vision*, vol. 10, no. 4, pp. 323–330, 2016.
9. L. M. Bergasa and J. Nuevo, "Real-time system for monitoring driver vigilance," in *Proceedings of the IEEE International Symposium on Industrial Electronics, 2005. ISIE 2005.*, vol. 3. IEEE, 2005, pp. 1303–1308.
10. Q. Ji, Z. Zhu, and P. Lan, "Real-time nonintrusive monitoring and prediction of driver fatigue," *IEEE transactions on vehicular technology*, vol. 53, no. 4, pp. 1052–1068, 2004.
11. H. García, A. Salazar, D. Alvarez, and Á. Orozco, "Driving fatigue detection using active shape models," in *Advances in Visual Computing: 6th International Symposium, ISVC 2010, Las Vegas, NV, USA, November 29-December 1, 2010, Proceedings, Part III 6*. Springer, 2010, pp. 171–180.
12. G. Sikander and S. Anwar, "Driver fatigue detection systems: A review," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 6, pp. 2339–2352, 2018.
13. F. Liu, D. Chen, J. Zhou, and F. Xu, "A review of driver fatigue detection and its advances on the use of rgb-d camera and deep learning," *Engineering Applications of Artificial Intelligence*, vol. 116, p. 105399, 2022.
14. Y. Yang, Y. Zeng, L. Yang, Y. Lu, X. Lee, and Y. Enomoto, "Action recognition and sports evaluation of running pose based on pose estimation," *International Journal of Human Movement and Sports Sciences*, vol. 12, pp. 148–163, 2024.
15. P. Picerno, A. Cereatti, and A. Cappozzo, "Joint kinematics estimate using wearable inertial and magnetic sensing modules," *Gait & posture*, vol. 28, no. 4, pp. 588–595, 2008.
16. Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime multi-person 2d pose estimation using part affinity fields," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 7291–7299.
17. S. J. Hawley, A. Hamilton-Wright, and S. L. Fischer, "Detecting subject-specific fatigue-related changes in lifting kinematics using a machine learning approach," *Ergonomics*, vol. 66, no. 1, pp. 113–124, 2023.

18. J. Wang, K. Qiu, H. Peng, J. Fu, and J. Zhu, “Ai coach: Deep human pose estimation and analysis for personalized athletic training assistance,” in *Proceedings of the 27th ACM international conference on multimedia*, 2019, pp. 374–382.

19. K. Papoutsakis, G. Papadopoulos, M. Maniadakis, T. Papadopoulos, M. Lourakis, M. Pateraki, and I. Varlamis, “Detection of physical strain and fatigue in industrial environments using visual and non-visual low-cost sensors,” *Technologies*, vol. 10, no. 2, p. 42, 2022.

20. J. Martinez, R. Hossain, J. Romero, and J. J. Little, “A simple yet effective baseline for 3d human pose estimation,” in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2640–2649.

21. M. M. Rivera, L. Martinez, A. Ochoa, A. N. Zezzatti, J. Rodarte, and N. Lopez, “Prototype interface for detecting mental fatigue with eeg and xai frameworks in industry 4.0,” *Explainable Artificial Intelligence in Medical Decision Support Systems*, vol. 117, 2023.

22. I. Hussain, R. Jany, R. Boyer, A. Azad, S. A. Alyami, S. J. Park, M. M. Hasan, and M. A. Hossain, “An explainable eeg-based human activity recognition model using machine-learning approach and lime,” *Sensors*, vol. 23, no. 17, p. 7452, 2023.

23. S. G. Hart, “Nasa task load index (tlx),” 1986.

24. D. Micklewright, A. St Clair Gibson, V. Gladwell, and A. Al Salman, “Development and validity of the rating-of-fatigue scale,” *Sports Medicine*, vol. 47, pp. 2375–2393, 2017.

25. G. E. Batista, R. C. Prati, and M. C. Monard, “A study of the behavior of several methods for balancing machine learning training data,” *ACM SIGKDD explorations newsletter*, vol. 6, no. 1, pp. 20–29, 2004.

26. R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-cam: Visual explanations from deep networks via gradient-based localization,” in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618–626.

27. K. Simonyan, A. Vedaldi, and A. Zisserman, “Deep inside convolutional networks: Visualising image classification models and saliency maps,” in *Workshop at International Conference on Learning Representations*, 2014.

28. F. Doshi-Velez and B. Kim, “Towards a rigorous science of interpretable machine learning,” *arXiv preprint arXiv:1702.08608*, 2017.