

University of Groningen

Predicting Operators Fatigue in a Human in the AI Loop for Defect Detection in Manufacturing

Rožanec, Jože; Kriznar, Karel; Montini, Elias; Cutrona, Vincenzo; Koehorst, Erik; Fortuna, Blaz; Mladenic, Dunja; Emmanouilidis, Christos

Published in:
 Proceedings of the 2023 IFAC World Congress

DOI:
[10.1016/j.ifacol.2023.10.1157](https://doi.org/10.1016/j.ifacol.2023.10.1157)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
 Publisher's PDF, also known as Version of record

Publication date:
 2023

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Rožanec, J., Kriznar, K., Montini, E., Cutrona, V., Koehorst, E., Fortuna, B., Mladenic, D., & Emmanouilidis, C. (2023). Predicting Operators Fatigue in a Human in the AI Loop for Defect Detection in Manufacturing. In *Proceedings of the 2023 IFAC World Congress* (pp. 7609-7614). (IFAC-PapersOnLine; Vol. 56, No. 2). IFAC-PapersOnLine. <https://doi.org/10.1016/j.ifacol.2023.10.1157>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Predicting Operators' Fatigue in a Human in the Artificial Intelligence Loop for Defect Detection in Manufacturing ^{*}

Jože M. Rožanec ^{*,**,\dagger} Karel Križnar ^{***} Elias Montini ^{****,\S}
Vincenzo Cutrona ^{****} Erik Koehorst [|] Blaž Fortuna [‡]
Dunja Mladenec ^{**} Christos Emmanouilidis [†]
^{*} *Jožef Stefan Institute, Ljubljana, Slovenia, (e-mail: joze.rozanec@ijs.si)*

^{**} *Jožef Stefan International Postgraduate School, Ljubljana, Slovenia (e-mail: dunja.mladenec@ijs.si)*

^{***} *University of Ljubljana, Faculty of mathematics and physics, Ljubljana, Slovenia, (e-mail: kk4876@student.uni-lj.si)*

^{****} *SUPSI (e-mail: elias.montini@supsi.ch, vincenzo.cutrona@supsi.ch).*

[†] *University of Groningen (e-mail: c.emmanouilidis@rug.nl)*

[‡] *Qlector d.o.o., Ljubljana, Slovenia (e-mail: blaz.fortuna@qlector.com).*

[§] *Politecnico di Milano, DEIB, Milan, Italy*

[|] *Philips Consumer Lifestyle BV, Drachten, The Neatherlands (e-mail: erik.koehorst@philips.com)*

Abstract: Quality inspection, typically performed manually by workers in the past, is now rapidly switching to automated solutions, using artificial intelligence (AI)-driven methods. This elevates the job function of the quality inspection team from the physical inspection tasks to tasks related to managing workflows in synergy with AI agents, for example, interpreting inspection outcomes or labeling inspection image data for the AI models. In this context, we have studied how defect inspection can be enhanced, providing defect hints to the operator to ease defect identification. Furthermore, we developed machine learning models to recognize and predict operators' fatigue. By doing so, we can proactively take mitigation actions to enhance the workers' well-being and ensure the highest defect inspection quality standards. We consider such processes to empower human and non-human actors in manufacturing and the sociotechnical production system. The paper first outlines the conceptual approach for integrating the operator in the AI-driven quality inspection process while implementing a fatigue monitoring system to enhance work conditions. Furthermore, it describes how this was implemented by leveraging data and experiments performed for a real-world manufacturing use case.

Copyright © 2023 The Authors. This is an open access article under the CC BY-NC-ND license (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)

Keywords: manufacturing plant control; intelligent manufacturing; human-centric manufacturing; fatigue monitoring

1. INTRODUCTION

Human-centricity is one of the core values of the evolving fifth industrial revolution. New approaches must be devised to ensure that AI-driven solutions developed for manufacturing settings serve human workers and operators better, enhancing their work conditions, while also contributing to operational and enterprise goals. Looking at artificial intelligence (AI) - enabled manufacturing from such a viewpoint, this paper focuses on a key pattern in the emerging transformation of jobs transitioning to functions that involve further cognitive involvement from human workers and operators instead of largely physical routine

activities. A typical example is in manufacturing quality inspection. This was a common manual activity undertaken by workers in the quality control teams but is now served by a range of automated solutions, starting from simple computer vision-based solutions. The job profile of workers in such teams increasingly involves sampling outcomes of the automated visual inspection to confirm the automated classification of parts or larger components. Similarly, it involves curating inspection data, such as inspection images, and labeling them for training a machine learning solution. This requires a cognitive shift in the workers' activities, resulting in higher added-value outcomes. Such outcomes can hardly be achieved by the automated solution alone, without human involvement.

Within this context, the present paper offers a novel view of the human in the loop of AI-driven visual quality

^{*} This work was supported by the Slovenian Research Agency and the European Union's Horizon 2020 program project STAR under grant agreement number H2020-956573.

inspection, with interventions targeting the empowerment of both the human and the AI agent to jointly perform the intended activities in ways that would have been hard for either type of actor alone to achieve. The empowerment of the AI actor is achieved through the integration of the human cognitive capabilities in the AI loop, in the form of labeling instances of data and engaging with the AI process through an Active Learning mechanism. The empowerment of the human actors is twofold: first, the AI-driven automated visual inspection system delivers not only inspection outcomes but offers also hints and explanations as to how this outcome was derived; second, it supports the operator through a fatigue monitoring system, thus enabling more optimal human engagement and triggering fatigue prevention or mitigation actions. The remainder of the paper is structured as follows. The next section outlines related work on automated visual quality inspection, human fatigue monitoring, and human in the AI loop. Section 3 presents the use case that offers the industrial context for the presented research. The active learning methodology and experiment design are provided in Section 4. Results are presented and analyzed in Section 5. A brief discussion and the main conclusions drawn from the research are provided in Section 6.

2. RELATED WORK

A brief overview of literature related to (a) visual quality inspection; (b) humans in the AI loop; and (c) fatigue monitoring is provided here. Visual quality inspection is the main manufacturing operation application focus and serves as a research baseline. The involvement of the human in the AI loop is aimed to achieve dual empowerment: of the worker; and of the AI approach. Therefore key concepts about such involvement are part of the methodology baseline. Finally, human workers, either performing manual inspections or engaged in interacting with AI for achieving a joint quality inspection outcome, are subject to fatigue and relevant monitoring needs to be considered.

2.1 Visual quality inspection

Quality inspection on engineered products may involve simple visual features, such as colors, as well as more complex ones, needed to detect defects, cracks, orientation deficiencies, or other anomalies (Rajan et al. (2020)). Quality assessment based on surface visual inspection relates to products of varying shapes and sizes, and therefore the complexity of the task varies as well (Cao et al. (2018)) (Tsai and Jen (2021)) (Yun et al. (2020)). Production process quality monitoring may also involve considerable complexity, especially in assembly operations. (Frustaci et al. (2022)). Moreover, quality inspection is particularly relevant also for remanufacturing production processes, involving the assessment of individual parts (Saiz et al. (2021)). Whether quality inspection applies to the beginning of life (production), middle of life (operation and maintenance), or end of life (such as in remanufacturing), a commonly encountered problem is that of data sparsity: rarely the available sample images are sufficient for training, a problem which can partly be handled by data augmentation (Yun et al. (2020)). Whether it is in

the management and curation of image data exemplars, the tuning of algorithmic parameters or the labeling of instances for training, the role of the human in such processes is undeniable and yet not sufficiently appreciated.

2.2 Human in the AI Loop

When considering the joint involvement of human and AI actors in collaborative settings, the issue of trust attains paramount importance. Trust in itself depends on (a) AI offering an explanation about its outcomes, related to the concept of explainable AI; (b) humans exercising their own judgment regarding domain-specific decision-making circumstances, related to the concept of interpretable AI; (c) AI outcomes to achieve to narrow the distance between decisions and actions, related to the concept of actionable AI. All these require a careful design for involving the human in the AI loop (Emmanouilidis et al. (2019)) (Lyytinen et al. (2020)).

2.3 Human fatigue monitoring

Research on fatigue estimation produced several contributions in the last decade, with models to estimate human fatigue while doing different kinds of activities, including working, rehabilitation, and sports activities. In the rehabilitation and sport domains, monitored activities are usually predefined, with a desired level of fatigue demand, with involved subjects just repeating an exercise (e.g., *sit-to-stand* exercise (Aguirre et al. (2021)), or shoulders flexion and abduction, and elbow extension (Papakostas et al. (2019))). When coming to work activities, the same activity might be executed in different ways by each operator, possibly showing different fatigue demands. Studies have been conducted to estimate the physical or mental fatigue of operators while monitoring their job routine, including driving and piloting (Hooda et al. (2022)), palletizing and transporting of weighted containers (Maman et al. (2020)), supply carrying (Lambay et al. (2021)), construction (masonry) (Zhang et al. (2019)). Zhang et al. (2014) monitored operators while just walking (involving firefighters and construction workers).

Fatigue estimation is realized in different ways, including biological features, physical features, interactions monitoring, and hybrid approaches (Sikander and Anwar (2018)). While there are many examples of driver fatigue estimation, few can be identified in the manufacturing context. A physiological feature like eye blinking has been applied for fatigue estimation during the execution of push and pull tasks (Biondi et al. (2023)), while biological characteristics, such as heart rate variability and galvanic skin conductance, have been used to estimate fatigue in the execution of human-robot collaboration task (Bettoni et al. (2020)).

The majority of considered contributions investigated physical fatigue (Aguirre et al. (2021); Hooda et al. (2022); Maman et al. (2020); Papakostas et al. (2019); Zhang et al. (2014, 2019); Lambay et al. (2021)), sometimes focusing on specific muscles (Papakostas et al. (2019)). Recently, mental fatigue is getting more attention (Hooda et al. (2022)), even combined with physical fatigue evaluation (Ramos et al. (2020)). Specifically, mental fatigue can be

classified into active (mental depletion caused by active engagement in a task), passive (experienced because of monotonous tasks, causing distractions), and sleep-related fatigue (Sikander and Anwar (2019)).

However, a commonly agreed definition of fatigue is lacking in the literature, with fatigue being considered both a binary (detection problem) and numerical property (prediction problem). The binary definition is the most adopted one, but a few studies tried to predict fatigue on a three-level scale (Aguirre et al. (2021)). Approaches to predict fatigue in larger ranges (e.g., the full Borg CR-10 scale) is still lacking in the literature, possibly because a wider range of values increases the task complexity. Besides, it is hard to collect a balanced dataset with a sufficient amount of samples for each considered value of the scale, especially in non-controlled environments. This aspect prevents the training phase of large machine learning models (as a matter of fact, the largest dataset among the considered studies contains only 660 records (Aguirre et al. (2021))), with some trials conducted in recent years, spreading from traditional machine learning models (e.g., SVM (Aguirre et al. (2021); Hooda et al. (2022); Maman et al. (2020); Papakostas et al. (2019); Ramos et al. (2020); Zhang et al. (2014, 2019)), Random Forest (Aguirre et al. (2021); Maman et al. (2020); Papakostas et al. (2019)), K-nearest neighbors (Aguirre et al. (2021); Hooda et al. (2022); Maman et al. (2020))), to modern deep learning models (e.g., neural networks (Aguirre et al. (2021); Hooda et al. (2022)), Autoencoder (Hooda et al. (2022)), recurrent neural networks (Lambay et al. (2021))). For supervised models, most of the conducted experiments preferred to label collected data by *verbally* asking operators whether they feel tired or not; in other cases, some approaches proposed approaches to infer fatigue from sensors data: Ramos et al. (2020) proposed the Global Fatigue Descriptor (GFD); Zhang et al. (2014) used the maximum voluntary isokinetic exertion (MVE)); various automobile companies rely on the interaction of the driver with vehicle control (e.g., pedals and wheel), or driver's physical features (e.g., eye blinking). Of the drowsiness-detection measures available in the literature, "PERCLOS" represents the most reliable and valid determination of alertness level, which reflects slow eyelid closures ("droops") rather than blinks (Junaedi and Akbar (2018)), however, it does not consider individual differences.

3. USE CASE

The research we present in this paper was performed with data provided by *Philips Consumer Lifestyle BV*, from Drachten, The Netherlands. The manufacturing plant is considered one of the largest development and production centers in Europe Philips has. The use case concerns manual visual inspection of Philips logos printed on shavers. In particular, *Philips Consumer Lifestyle BV* has printing machine setups for various products and logos. Many products whose logos are printed on these machines are manually handled and inspected to determine their visual quality. When a defect is observed, the product is removed from the manufacturing line, leaving only those that comply with the quality standards. Operators usually spend several seconds handling, inspecting, and labeling the products. This process could be enhanced with artificial

intelligence in multiple ways. First, a machine learning model could be developed to partially automate defect identification. The operators would therefore focus only on those cases where the machine could not determine with enough confidence that a certain defect exists. For those cases, artificial intelligence could generate defect hints, thus providing valuable information on where to look for a defect for a given piece. Generative models could be used to mitigate class imbalance and therefore induce a higher degree of attention from the operators and periodically validate the quality of their labeling. Finally, we envision machine learning could be used to predict operators' fatigue and the consequent labeling quality decay.

The original dataset provided by *Philips Consumer Lifestyle BV* has 3.518 images, with three possible categories (see Fig. 1): good printing, double printing, and interrupted printing.



Fig. 1. Examples for the three classes present in the dataset provided by *Philips Consumer Lifestyle BV*. The dataset is imbalanced. Most images correspond to non-defective pieces (good print), while the defective prints (double or interrupted prints) introduce different levels of imbalance.

4. METHODOLOGY

For this research, we developed a web application for defect inspection (see 2). The application required the participants to log in with a username and password and execute a series of experiments (labeling 600 images in each case). The application forced them to execute the experiments sequentially to ensure that the learning path for all participants would be the same. To that end, the following experiment was unlocked only when the current was completed. As a collateral effect, all the participants had an increased learning process between the single experiments and the hinting techniques, which we could not mitigate. For each experiment, we collected data regarding the start and end time of each experiment, the time required to label each image, and the label the participants assigned to the image. Furthermore, we also persisted the participant's ID executing the experiments, but no personal information persisted in the database that could allow associating the ID to a particular person. Finally, we must highlight that for each image, we had complementary data to determine whether it was obtained from the original dataset or generated with a generative model. Furthermore, for each image, we had its ground truth label. We detail the experiments in Table 1. Results on how different hinting techniques affected the participants' labeling accuracy were reported in Rožanec et al. (2022).

We conducted the experiments among two groups of participants: four researchers and students from an artificial intelligence laboratory and three operators working on manual visual inspection at *Philips Consumer Lifestyle*

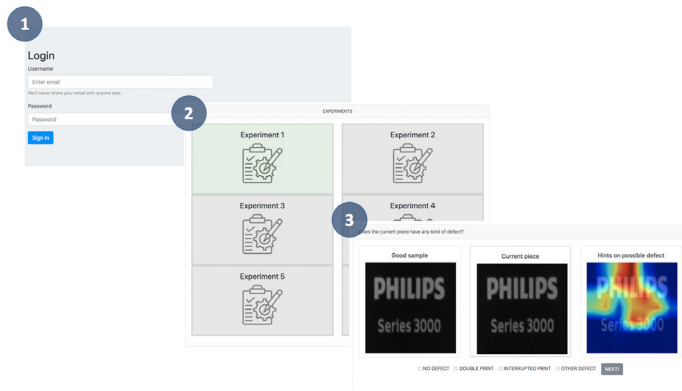


Fig. 2. Three screenshots describing the application we developed to test six labeling scenarios. In the image, we highlight three different screens: (1) the login screen, (2) the experiments menu (to choose the next available experiment), and (3) the particular experimental setup. Among the experiments, when performing a labeling task, we always display a *good* image on the left side, the image to be inspected is shown in the center, and on the right, we eventually provide some defect hinting. In (3), the image corresponds to a defective item (interrupted print), and the defect hint was created with the GradCAM technique (Selvaraju et al., 2017)).

Experiment	Balanced dataset?	Defect hinting technique
1	NO	No defect hinting
2	YES	No defect hinting
3	NO	DRAEM anomaly map
4	YES	DRAEM anomaly map
5	YES	GradCAM heatmap
6	YES	Nearest labeled image, considering SSIM.

Table 1: Experiments description. SSIM is the acronym for Structural Similarity Index Measure.

BV. We asked them to execute the experiments daily to avoid fatigue affecting their performance.

Experiments 1-4 were designed to determine whether the class imbalance affects the quality of labeling, while Experiments 4-6 aimed to examine whether defect hinting could enhance the quality of labeling and which hinting techniques would be most helpful. For this research, we were interested in the quality of labeling dynamics over time and whether a consistent decrease in the quality of labeling could be observed across the experiments. To that end, we computed precision and recall for each participant over a rolling window comprehending the last hundred images. When computing precision and recall, we considered a binary classification (defective vs. non-defective images, with defective samples as the class of interest). We confirmed that decay in precision and recall could be observed for most experiment participants. Furthermore, the highest precision and recall were obtained for Experiment 6. We, therefore, chose the precision and recall time series to develop machine learning models to detect the participants' fatigue (see Fig. 3).

When analyzing the precision and recall time series, we considered that for this research, we would consider the participant fatigued once their performance drops below 0.75, either considering precision or recall, over a rolling

Participant ID	Fatigue point (Image ID)
1	580
2	594
4	242
10	305
11	310
14	133

Table 2: Fatigue points identified for the participants. Participants who did not meet the fatigue criteria were excluded from the table.

time window considering the last hundred images labeled by a given participant within an experiment. In other scenarios, such thresholds could be adjusted based on the use case requirements (e.g., defined acceptance quality levels for each defect). We then checked when such events (we call them *fatigue points*) took place and reported them in Table 2. We also analyzed the time required to label each image and whether participants required more or less time to annotate the images. We found no strong signals in this regard and therefore did not consider labeling time when creating features for a machine-learning model.

We defined fatigue as a performance drop below a 0.75 threshold for precision or recall and decided to solve the problem with two regression models: one to predict the expected precision and one to predict the expected recall. We did so for two-time horizons: one and thirty images ahead. We report the features created for each model in Table 3.

Both models were created with a Linear Regression algorithm, given its simplicity and good results obtained. We trained the models with the first half of the dataset (three hundred images) and predicted the second half. Our decision to do so was mainly guided by the fact that (a) we were interested in identifying the first fatigue point for each user, (b) we had only seven participants, and (c) most fatigue points were located in the second half of the dataset.

5. RESULTS AND ANALYSIS

We present the results in Table 4 and Table 5. From Table 5 we see that the forecasts' performance between an image ahead and thirty images ahead forecast strongly degrades. Nevertheless, in Table 4 we found that the fatigue points were predicted accurately. The fatigue points were predicted perfectly for all relevant cases in a one-image-ahead forecast. For the thirty images ahead, on the other hand, a perfect prediction was issued for participant $id = 2$, and a prediction close to the target for participants $id = 10$ and $id = 11$, predicting a fatigue point three and one images before the actual event occurrence. We consider the model failed to accurately predict the fatigue point for participant $id = 1$, given it predicted a fatigue point 140 images earlier than the actual event. We, therefore, consider this prediction a false positive.

6. DISCUSSION AND CONCLUSION

From multiple experiments we performed for data labeling of manufacturing defects, we observed that most of the participants' quality of labeling (measured as precision and recall) declined over time. Therefore, we hypothesize

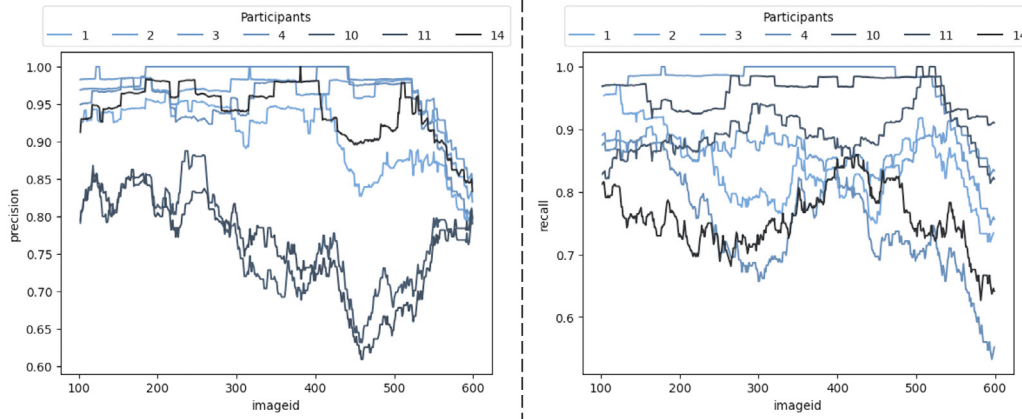


Fig. 3. Precision (left) and recall (right) plots for each of the participants. Participants 1-4 correspond to researchers from an artificial intelligence laboratory, while the rest correspond to operators from *Philips Consumer Lifestyle BV*. A detailed analysis of discrepancies between precision and recall among researchers and operators is presented in Rožanec et al. (2022).

Model	Feature	Explanation
Precision	p-t	Precision value at t, computed over the last 100 images. We considered t=0, 1, 2, 3.
	p-t1/p-t2	Ratio between p-t1 and p-t2.
Recall	r-t	Recall value at t, computed over the last 100 inspected images. We considered t=0, 1, 2, 3.
	r-t1/r-t2	Ratio between r-t1 and r-t2.

Table 3: Features used to predict future precision and recall.

participant id	fatigue point			Precision			Recall		
	target	predicted		target	predicted		target	predicted	
		1 s.a.	30 s.a.		1 s.a.	30 s.a.		1 s.a.	30 s.a.
1	580	580	440	0,8089	0,8117	0,8033	0,7333	0,7450	0,7398
2	594	594	594	0,8261	0,8241	0,8169	0,7500	0,7445	0,7394
10	305	305	302	0,7439	0,7454	0,7309	0,9385	0,9382	0,9347
11	310	310	309	0,7560	0,7464	0,7270	0,9841	0,9840	0,9821

Table 4: Predicted fatigue points and the corresponding predicted precision and recall. *s.a.* abbreviates *steps ahead*, indicating the forecasting horizon.

	One image ahead		Thirty images ahead	
	Precision	Recall	Precision	Recall
RMSE	0,0474	0,0442	0,2622	0,2465
MAE	0,0057	0,0059	0,1079	0,1049
R ²	0,8470	0,8450	0,1433	0,1910

Table 5: Predicted RMSE, MAE, and R² values for both forecasting horizons: one image ahead and thirty images ahead.

such a decline in the quality of labeling is associated with the operators’ fatigue. In this research, we successfully predict operators’ performance in one and thirty images in advance. If operators’ fatigue is measured through their labeling performance, they can decide when to stop the labeling activity based on the expected performance. This could be achieved by generating short streams of labeled, GAN-generated synthetic images on which the labeling accuracy can be measured and the operators’ fatigue determined. Nevertheless, including random streams of synthetic data affects the labeling throughput of unlabeled data. Therefore, this approach could complement predictive models aiming to predict operators’ fatigue based on some sensor data.

We have identified several work limitations based on the experience acquired through this research. While for this

research, we considered the first case when either precision or recall had a value lower than 0.75, more complex rules can be established. For example, we could consider acceptance quality levels for each type of defect or establish the extent to which performance dropping below the established acceptance quality level is tolerated and produce predictions for multiple horizons. This would allow us to determine participants’ fatigue with higher confidence and robustness. Furthermore, while there is a clear decreasing trend in the operators’ labeling performance, such performance fluctuates. Therefore, this must be considered to understand how such fluctuations affect the overall visual inspection performance and what mitigation actions can be taken to prevent such a performance drop when approaching lower bounds of acceptable quality thresholds.

Our future work will be focused on upgrading our existing software to incorporate (a) a machine learning model for eye blinking detection, (b) incorporate information regarding eye blinking into our fatigue detection model, (c) include questionnaires and cognitive tests, to have an objective measurement and baseline understanding of the participants (e.g., regarding fatigue and cognitive abilities) before labeling, (d) complement the application data with data obtained from wearable sensors, and (e) use active learning techniques to improve our fatigue detec-

tion machine learning models. We also plan to conduct experiments with a larger population in a homogeneous environment to enhance the robustness of our findings.

ACKNOWLEDGEMENTS

Philips Consumer Lifestyle B.V. provided industrial cases and data. Credits are due to Jelle Keizer and Paulien Dam for the use case.

REFERENCES

- Aguirre, A., Pinto, M.J., Cifuentes, C.A., Perdomo, O., Diaz, C.A., and Munera, M. (2021). Machine learning approach for fatigue estimation in sit-to-stand exercise. *Sensors*, 21(15), 5006.
- Bettoni, A., Montini, E., Righi, M., Villani, V., Tsvetanov, R., Borgia, S., Secchi, C., and Carpanzano, E. (2020). Mutualistic and adaptive human-machine collaboration based on machine learning in an injection moulding manufacturing line. *Procedia CIRP*, 93, 395–400.
- Biondi, F.N., Saberi, B., Graf, F., Cort, J., Pillai, P., and Balasingam, B. (2023). Distracted worker: using pupil size and blink rate to detect cognitive load during manufacturing tasks. *Applied ergonomics*, 106, 103867.
- Cao, G., Ruan, S., Peng, Y., Huang, S., and Kwok, N. (2018). Large-complex-surface defect detection by hybrid gradient threshold segmentation and image registration. *IEEE Access*, 6, 36235–36246. doi:10.1109/ACCESS.2018.2842028.
- Emmanouilidis, C., Pistofidis, P., Bertonec, L., Katsouros, V., Fournaris, A., Koulamas, C., Ruiz-Carcel, C., C., P.P.B.L.K.V.F.A.K.C.E., and Ruiz-Carcel, C. (2019). Enabling the human in the loop: Linked data and knowledge in industrial cyber-physical systems. *Annual Reviews in Control*, 47, 249–265. doi:10.1016/j.arcontrol.2019.03.004.
- Frustaci, F., Spagnolo, F., Perri, S., Cocorullo, G., and Corsonello, P. (2022). Robust and high-performance machine vision system for automatic quality inspection in assembly processes. *Sensors*, 22, 2839. doi:10.3390/s22082839.
- Hooda, R., Joshi, V., and Shah, M. (2022). A comprehensive review of approaches to detect fatigue using machine learning techniques. *Chronic Diseases and Translational Medicine*, 8(1), 26–35.
- Junaedi, S. and Akbar, H. (2018). Driver drowsiness detection based on face feature and perclos. In *Journal of Physics: Conference Series*, volume 1090, 012037. IOP Publishing.
- Lambay, A., Liu, Y., Morgan, P., and Ji, Z. (2021). A data-driven fatigue prediction using recurrent neural networks. In *2021 3rd International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, 1–6. IEEE.
- Lyytinen, K., Nickerson, J.V., and King, J.L. (2020). Metahuman systems = humans + machines that learn. *Journal of Information Technology*, 0268396220915917. doi:10.1177/0268396220915917.
- Maman, Z.S., Chen, Y.J., Baghdadi, A., Lombardo, S., Cavuoto, L.A., and Megahed, F.M. (2020). A data analytic framework for physical fatigue management using wearable sensors. *Expert Systems with Applications*, 155, 113405.
- Papakostas, M., Kanal, V., Abujelala, M., Tsiakas, K., and Makedon, F. (2019). Physical fatigue detection through emg wearables and subjective user reports: a machine learning approach towards adaptive rehabilitation. In *Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, 475–481.
- Rajan, A.J., Jayakrishna, K., Vignesh, T., Chandradass, J., and Kannan, T.T. (2020). Development of computer vision for inspection of bolt using convolutional neural network. *Materials Today: Proceedings*, 45, 6931–6935. doi:10.1016/j.matpr.2021.01.372.
- Ramos, G., Vaz, J.R., Mendonca, G.V., Pezarat-Correia, P., Rodrigues, J., Alfaras, M., and Gamboa, H. (2020). Fatigue evaluation through machine learning and a global fatigue descriptor. *Journal of healthcare engineering*, 2020.
- Rožanec, J.M., Zajec, P., Keizer, J., Trajkova, E., Fortuna, B., Breclj, B., Šircelj, B., and Mladenčić, D. (2022). Enhancing manual revision in manufacturing with ai-based defect hints. In *Central European Conference on Information and Intelligent Systems*, 357–363. Faculty of Organization and Informatics Varazdin.
- Saiz, F.A., Alfaro, G., and Barandiaran, I. (2021). An inspection and classification system for automotive component remanufacturing industry based on ensemble learning. *Information 2021*, 12(12), 489.
- Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., and Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, 618–626.
- Sikander, G. and Anwar, S. (2018). Driver fatigue detection systems: A review. *IEEE Transactions on Intelligent Transportation Systems*, 20(6), 2339–2352.
- Sikander, G. and Anwar, S. (2019). Driver fatigue detection systems: A review. *IEEE Transactions on Intelligent Transportation Systems*, 20(6), 2339–2352. doi:10.1109/TITS.2018.2868499.
- Tsai, D.M. and Jen, P.H. (2021). Autoencoder-based anomaly detection for surface defect inspection. *Advanced Engineering Informatics*, 48. doi:10.1016/j.aei.2021.101272.
- Yun, J.P., Shin, W.C., Koo, G., Kim, M.S., Lee, C., and Lee, S.J. (2020). Automated defect inspection system for metal surfaces based on deep learning and data augmentation. *Journal of Manufacturing Systems*, 55, 317–324.
- Zhang, J., Lockhart, T.E., and Soangra, R. (2014). Classifying lower extremity muscle fatigue during walking using machine learning and inertial sensors. *Annals of biomedical engineering*, 42(3), 600–612.
- Zhang, L., Diraneyya, M., Ryu, J., Haas, C., and Abdel-Rahman, E. (2019). Automated monitoring of physical fatigue using jerk. In *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction*, volume 36, 989–997. IAARC Publications.