Project Acronym: STAR
Grant Agreement number: 956573 (H2020-ICT-2020-1 – Research and Innovation Action)
Project Full Title: Safe and Trusted Human Centric Artificial Intelligence in Future Manufacturing Lines
Project Coordinator: Netcompany-Intrasoft

Funded by the Horizon 2020 Framework Programme of the European Union

DELIVERABLE

D4.6 – Human Robot Collaboration Knowledge Base

Dissemination level	PU -Public
Type of Document	Report, Prototype
Contractual date of delivery	30/06/2023
Deliverable Leader	JSI
Status - version, date	Final – v1.0, 11/07/2023
WP / Task responsible	WP4
Keywords:	Knowledge Base

Executive Summary

This document provides an overview of the activities and the results achieved in WP4 Safe, Transparent and Reliable Human-Robot Collaboration within Task 4.4 Knowledge Base for Human Robots Interactions.

The main activities carried out during T4.4 are:

1) develop and implement a data collection and preparation process;
2) implement knowledge base prototype based on data collected during experiments;
3) realise and evaluate an application scenario for STAR use cases, in particular for STAR PCL use case, with focus on visual inspection task.

This document describes the developed knowledge base prototypes and approaches applicable to STAR Use case.

The WP4 partners conducted an extensive scientific work presented in the following publications:

- Title: "Adaptive explainable artificial intelligence for visual defect inspection.", Venue: International Conference on Industry 4.0 and Smart Manufacturing (ISM) (submitted).

Deliverable Leader:	JSI
Contributors:	QLE, PCL, DFKI
Reviewers:	PCL, THA
Approved by:	Charalampos Ipektsidis (INTRA)

Document History			
Version	**Date**	**Contributor(s)**	**Description**
V0.1	25/05/2023	JSI	Table of contents
V0.2	12/06/3023	JSI	First draft
V0.3	19/06/2023	JSI	Updated version (Knowledge Based Systems and Technologies, Data Collection and Preparation Process, Knowledge Base Implementation, Applications)
V0.4	20/06/2023	JSI	Finalisation of inputs
V0.5	26/06/2023	THA	First review
V0.6	03/07/2023	JSI	Review Addressed
V1.0	11/07/2023	INTRA	Final QA'ed version submitted

Table of Contents

Table of Figures

Definitions, Acronyms and Abbreviations

Acronym/ Abbreviation	Title
AI	Artificial Intelligence
AL	Active Learning
BoW	Bag of Words
GAN	Generative Adversarial Networks
JSGF	Java Speech Grammar Format
LR	Logistic Regression
MAP	Mean Average Precision
MLP	Multilayer Perceptron
NLP	Natural Language Processing
RF	Random Forest
ROC AUC	Receiver Operating Characteristic, Area Under Curve
SA	Sentiment Analysis
SOTA	State of the Art
STT	Speech-To-Text
SVM	Support Vector Machine
TF-IDF	Term Frequency - Inverse Document Frequency
TTS	Text-to-Speech
WP	Work Package

1 Introduction

Task 4.4 Knowledge Base for Human Robots Interactions targets STAR's objective to explore and implement knowledge bases that can contribute to human robot collaboration and accelerate knowledge acquisition.

Knowledge bases contain domain-specific knowledge obtained from human experts via knowledge acquisition means. For instance, in line with research and implementation work in WP4 Task 4.4, we have developed a specialised web application for conducting experiments and collecting knowledge.

Deliverable D4.6 Human Robot Collaboration Knowledge Base is the last deliverable of STAR WP4. In this deliverable the experimental set up along with data collection process is described in details, the specific attention is dedicated to challenges that occurred in the process of knowledge base development.

The knowledge base development of Task 4.4 has been focused on PCL use case, in particular on quality inspection of Phillips manufactured products. Quality control (QC) is a specific procedure or set of procedures that allow the company to ensure that a manufactured product fulfils all the requirements and quality criteria from the client.

In summary, the main development of Task 4.4 has been focused on developing a web application that allowed us to gather human feedback on explainable artificial intelligence heatmaps and anomaly maps obtained from unsupervised classification models, and understand how these could be helpful to the human when working on defect annotation of product images.

Furthermore, we have used the knowledge bases to train machine learning models and demonstrate how such knowledge bases could be used to personalise the abovementioned heatmaps and anomaly maps based on peoples' perceptions.

As stated in the description of work, Task 4.4 consolidates the knowledge obtained from inputs in the previous STAR WP4 tasks (STAR WP4 deliverables).

The outcomes of Task 4.4 Knowledge Base for Human Robots Interactions are targeted at STAR project partners, as well at the whole research and development community in the area of manufacturing.

The document is structured in the following sections:

- **Section 1** provides an overview of the scope and the structure of this document;
- **Section 2** includes a review of existing technologies related to knowledge bases;
- **Section 3** describes the data collection process required for knowledge bases development;
- **Section 4** provides the implementation activities;
- **Section 5** demonstrated the application of developed knowledge base to STAR use cases;
- **Section 6** summarises the activities and concludes the document.

2 Knowledge Based Systems and Technologies

Knowledge bases are a set of expressions describing knowledge. The knowledge bases contain the knowledge necessary for understanding, formulating and solving problems. They store domain-specific knowledge captured from the human expert via some knowledge acquisition module. A knowledge base contains factual (knowledge of the task domain that is widely shared) and heuristic knowledge (experiential and judgemental knowledge, usually knowledge of good practice, judgement and plausible reasoning of the particular professional field).

Knowledge bases have a rich heritage tracing back to the expert systems of the 1970s. The first knowledge-based systems (systems that represent knowledge explicitly) were developed by artificial intelligence researchers. The advent of machine learning has sparked a resurgence in knowledge bases, as they now play a crucial role in enabling key functionalities of prominent products (e.g., Google Assistant, or Amazon Alexa).

While the current amount of information available surpasses any previous era in human history, from a software point of view, a significant portion of it remains untapped due to its unstructured nature. Knowledge bases present data and information in a structured way, so that it can be used by expert applications. Building a knowledge base remains a challenging task, given it requires to deal with complex input data and many related subtasks such as extracting information, cleaning, linking, and integrating. The advent of deep learning eased some of these tasks allowing to operate on raw data to extract features. Nevertheless, critical decisions must be made on how to design the knowledge base ensuring relevant data is properly linked, can be efficiently ingested and accessed. Furthermore, not all knowledge is digitalised: much knowledge is retained by human experts, and the knowledge retrieval application can be used to gather it and store it into the knowledge base.

Knowledge bases are at the core of knowledge-based systems. Knowledge-based systems can adopt three types of approaches [REF-01]: (a) ontology-based, (b) rule-based, and (c) model-based. Ontology-based approaches leverage ontologies to describe context knowledge through concepts and relationships related to a specific domain and enable and achieve knowledge reuse and sharing between computational entities and reasoning mechanisms. Rule-based approaches encode domain knowledge in the form of rules indicating conditions on certain facts and the possible outcomes affecting the world. It can be used where some experience exists, but there is a lack of details that could enable a modelling approach. Furthermore, while such systems initially aimed to capture human knowledge, experience showed that encoding knowledge in the form of production rules did not scale to address real-world applications. Finally, model-based approaches leverage data and knowledge about a particular phenomenon to simulate or predict a particular behaviour. They can also support multiple data and knowledge types. Nevertheless, model-based approaches may be affected by changes in operating conditions, and therefore, these must be properly considered.

Knowledge bases for the purpose of visual inspection have been designed in the past. [REF-02] reports about a knowledge-based system for automatic defect recognition. To build the knowledge base, several knowledge sources were considered: technical literature, research work, faulty printouts, and human experts (experts in research and development who have a global view, and operators who have a local view of the problem; and divided into different groups based on their expertise and seniority). When considering human input, one of the relevant aspects to be considered is how tacit knowledge can be captured. [REF-03] describes tacit knowledge as an intrinsic understanding of how things work that helps to produce strategies and solutions in new circumstances. Such knowledge, if captured, can be transferred to other operators - a relevant factor when considering an increasing workforce mobility.

3 Data Collection and Preparation Process

For the purpose of research regarding knowledge base systems in the EH H2020 STAR project, we developed a web application that allowed us to carry on multiple experiments. The web application was deployed on a server leveraging Docker for virtualisation and using a MongoDB document database for data persistence. The application required users to login with a username and password (see Fig. 1), so that the experiments performed could be associated to that particular username.

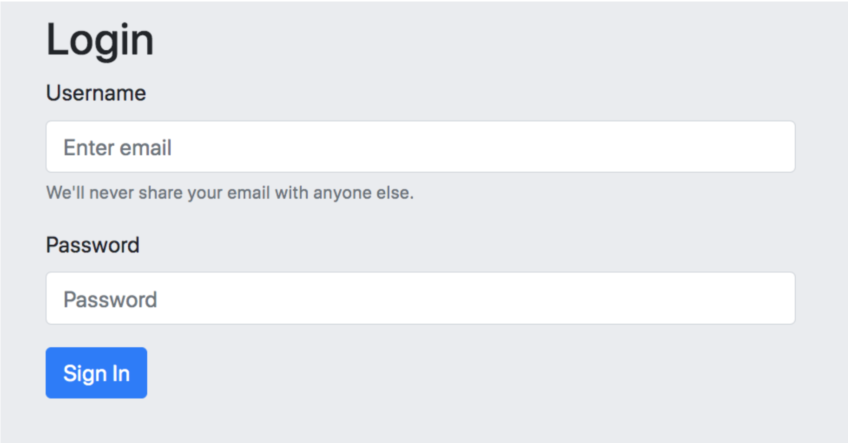

Figure 1: Login screen.

The participants were asked to login with a username and password, so that we could associate responses to a particular username. No personal information was recorded.

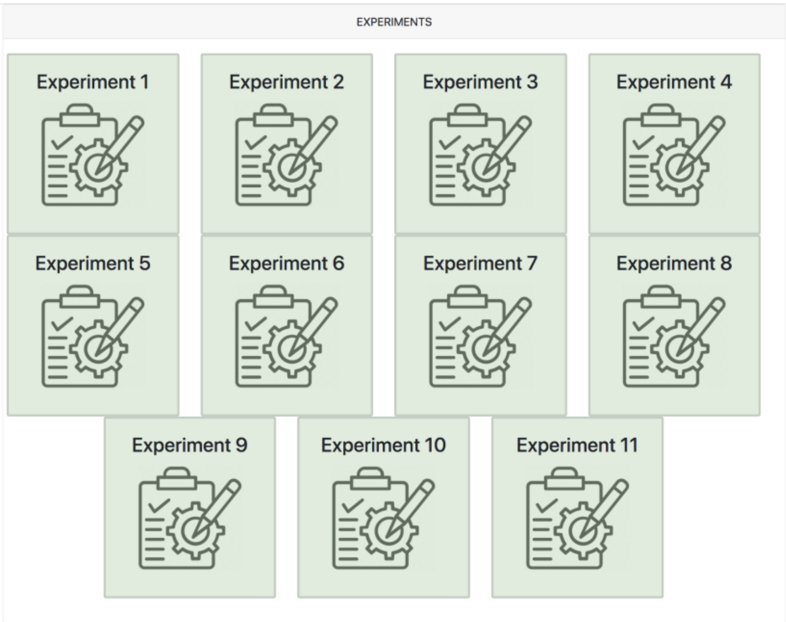

Figure 2: Implemented experiments

We implemented a total of 11 experiments that were used for different purposes of the STAR project. For experiments 9-11, we collected information that was then recorded to a knowledge base and further used to develop new machine learning models.

Three experiments were considered for the development of knowledge bases. For each experiment, we showed 30 images to the participants. A hashing function was implemented to map each user to a given set of images. While each user would label just a few images, the many participants would annotate a wider dataset. Furthermore, we ensured some overlap existed between the labelled images, so that a decision could be made on the final label based on the labelling input from multiple users.

To ensure participants' information was completely anonymised, we had papers written with username, password and the login URL. Each participant could select such a paper from a box, and therefore assignments were completely randomised and no record was kept associating information between the physical participant and the application username. We managed to engage a total of eighteen participants.

Below we provide a detailed description of each experiment, considering the particular goal pursued in them. We rename experiments 9-11 from Fig. 2 to Experiments A-C. The rest of the experiments were performed mostly to collect data regarding explainable artificial intelligence, to evaluate human perception and labelling accuracy. The results were not used so far and are not reported.

Experiment A

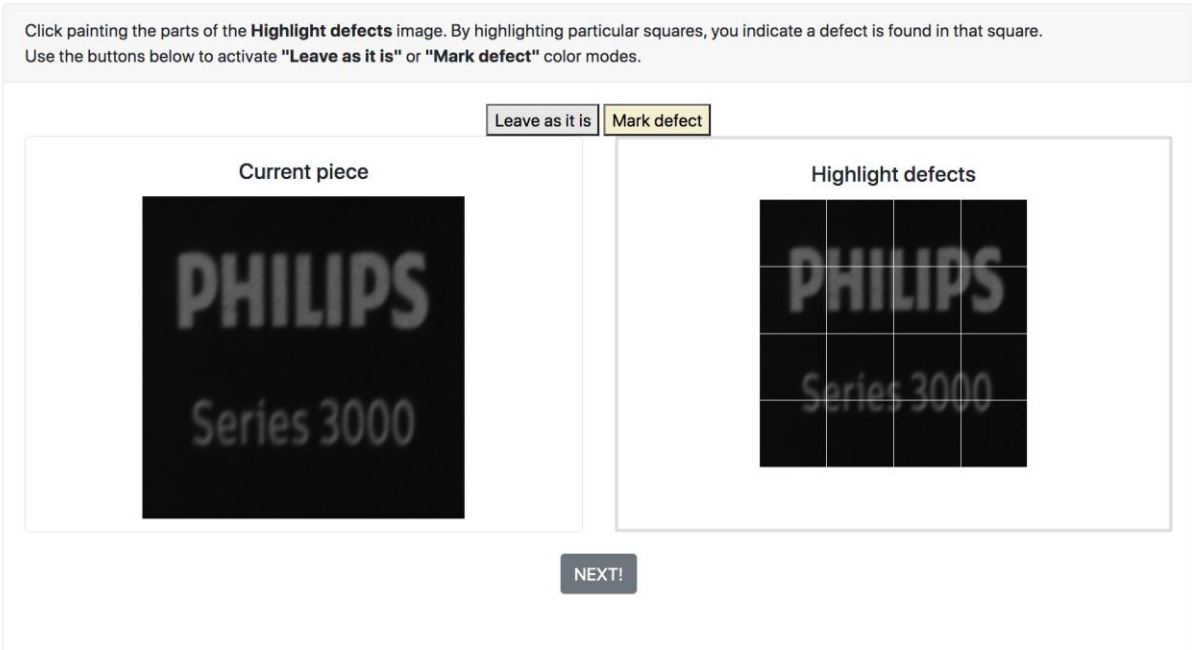

Figure 3: Experiment A

In Experiment A we showed an image twice: without and with a grid. The participant was tasked with marking the squares were some printing defect was observed.

The objective of the experiment was to obtain ground truth labels indicating whether a defect was observed in the squares of the images shown to the participants. We always showed the same image twice: without a grid, and with a grid (see Fig. 3). The defects were to be marked in the image displayed with a grid (see Fig. 4). Nevertheless, the image without the grid was shown to ensure the user could view the whole image without any noise and therefore enhance the quality of labelling.

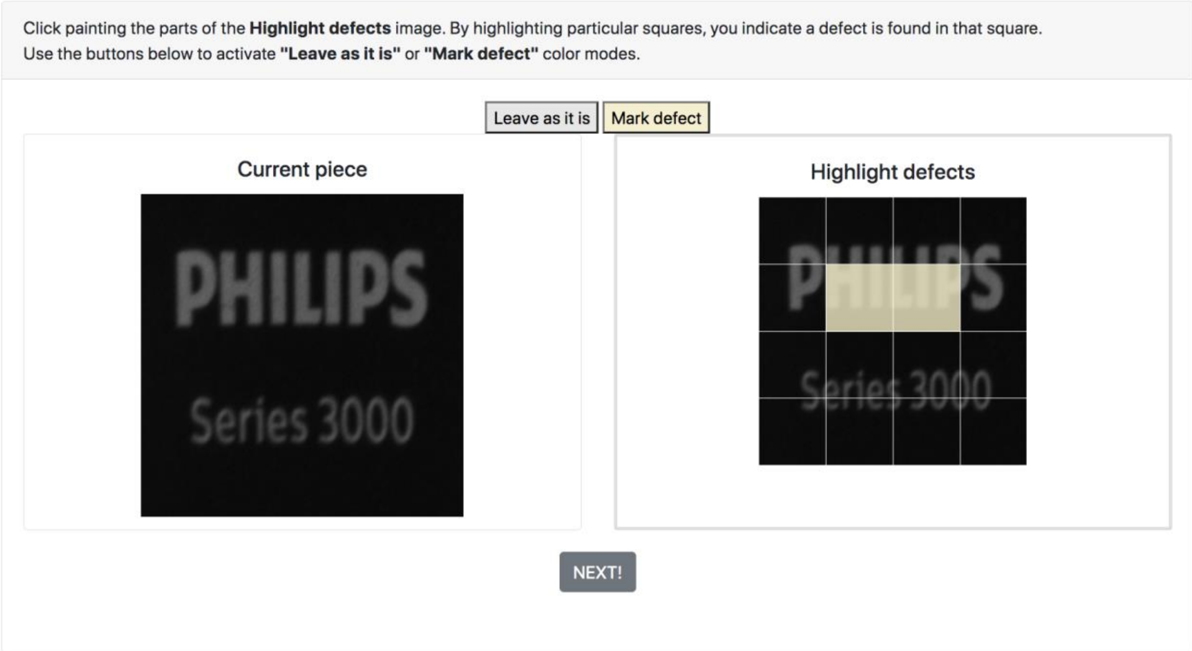

Figure 4: Defects marked in the image

In the image we see two squares marking that a defect was observed in them. The example corresponds to Experiment A.

Experiment B

The objective of the experiment was to obtain ground truth labels indicating whether a defect was observed in the squares of the images shown to the participants. We always showed two images: the image to be inspected and the corresponding DRAEM anomaly map: the image was displayed without a grid, and the anomaly map was displayed with a grid (see Fig. 5). The defects were to be marked in the anomaly map displayed with a grid (see Fig. 6). Nevertheless, the image without the grid was shown to ensure the user could view the whole image without any noise and therefore enhance the quality of labelling. DRAEM (Discriminatively trained Reconstruction Anomaly Embedding Model) is an unsupervised classification technique that aims to learn how images of non-defective products look like, and reconstructs any input image to their non-defective counterpart. By computing the difference between the input and the resulting image, it is capable of finding and highlighting defects for defective products. The DRAEM anomaly map displays regions that deviate from the learned representation for images of non-defective items.

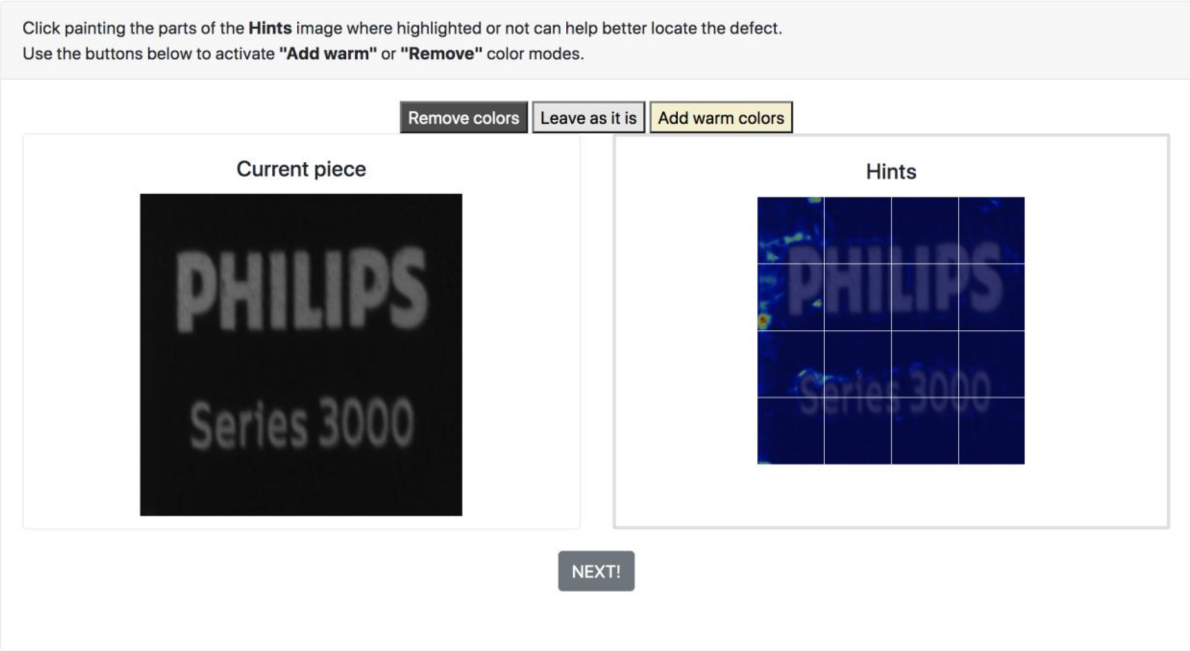

Figure 5: Experiment B with product being inspected and the corresponding DRAEM anomaly map

In Experiment B, we showed an image of the product being inspected and the corresponding DRAEM anomaly map. While the product image was shown without a grid, the DRAEM anomaly map had a grid associated to it, to facilitate anomaly map annotation, considering information displayed on the picture and the image. The participant was tasked with marking the squares and providing feedback whether warm colours should be added or removed.

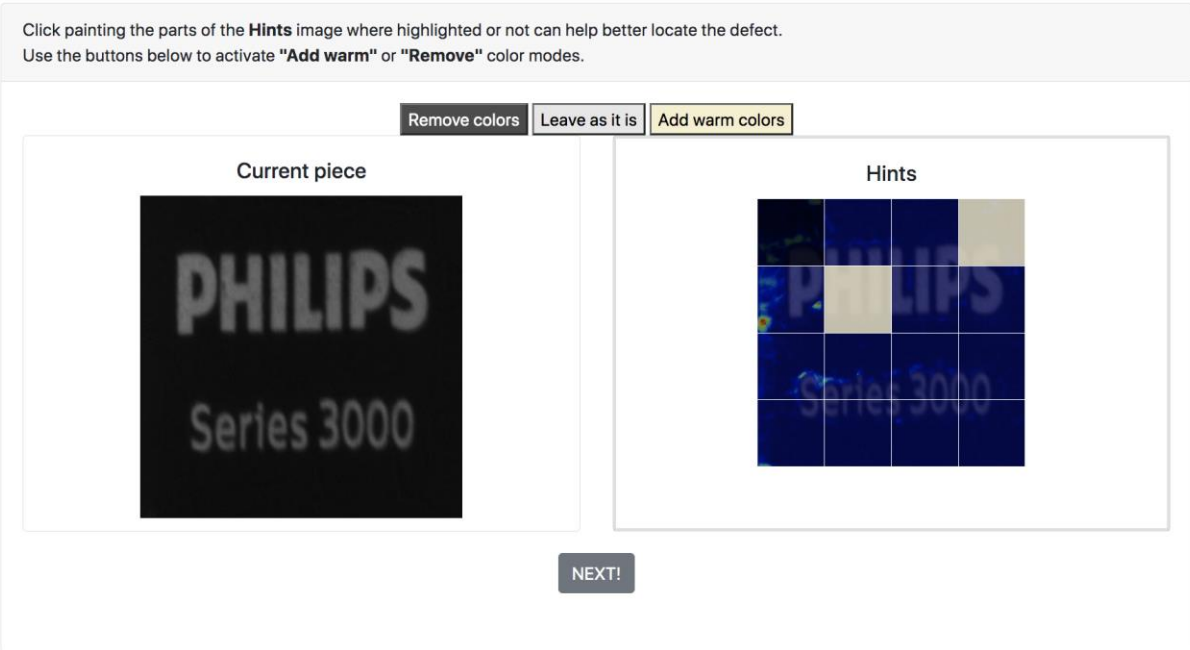

Figure 6: Experiment B with hints marked in the image that the anomaly map should be recoloured

In the image we see three squares marking that the anomaly map should be recoloured. In particular, the participant is asking to remove some spurious warm colours at the top left (marked with a dark *colour) and add some warm colours to another two quadrants at the middle/top right (marked with a light yellow). The example corresponds to Experiment B.*

Experiment C

The objective of the experiment was to obtain ground truth labels indicating whether a defect was observed in the squares of the images shown to the participants. We always showed the two images: the product being inspected (without a grid, and the GradCAM heat map (with a grid (see Fig. 7)). The defects were to be marked in the image displayed with a grid (see Fig. 8). Nevertheless, the image without the grid was shown to ensure the user could view the whole image without any noise and therefore enhance the quality of labelling.

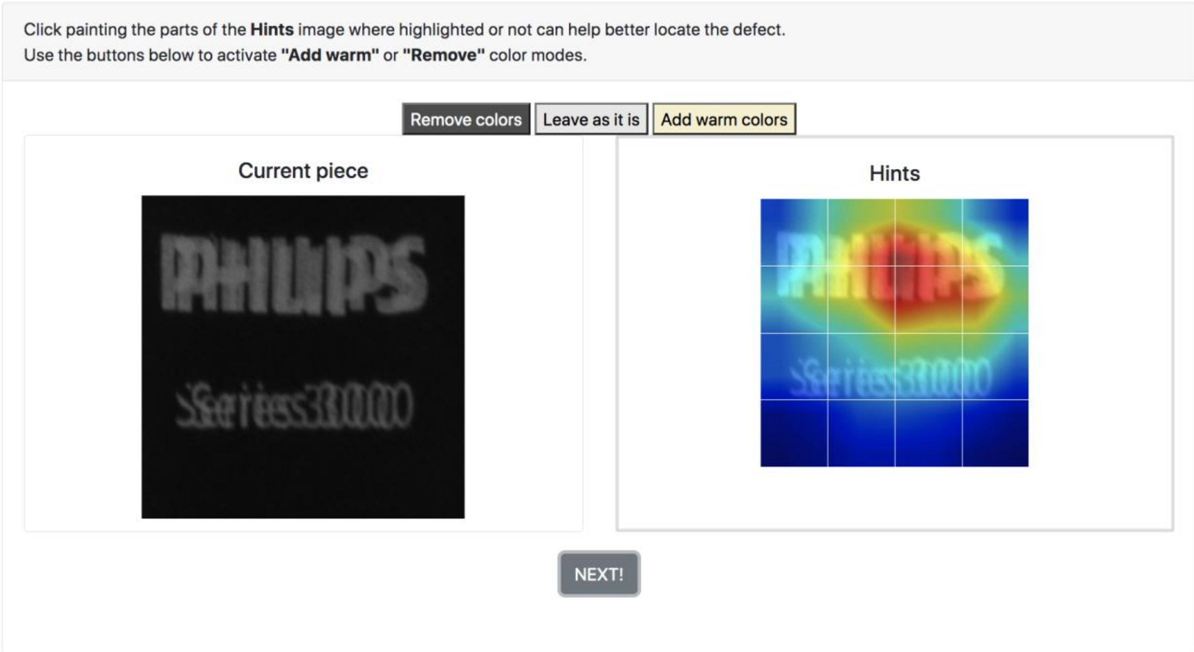

Figure 7: Experiment C with product being inspected and the corresponding GradCAM heatmap

In Experiment C, we showed an image of the product being inspected and the corresponding GradCAM heatmap. While the product image was shown without a grid, the GradCAM heatmap had a grid associated to it, to facilitate anomaly map annotation, considering information displayed on the picture and the image. The participant was tasked with marking the squares and providing feedback whether warm colours should be added or removed.

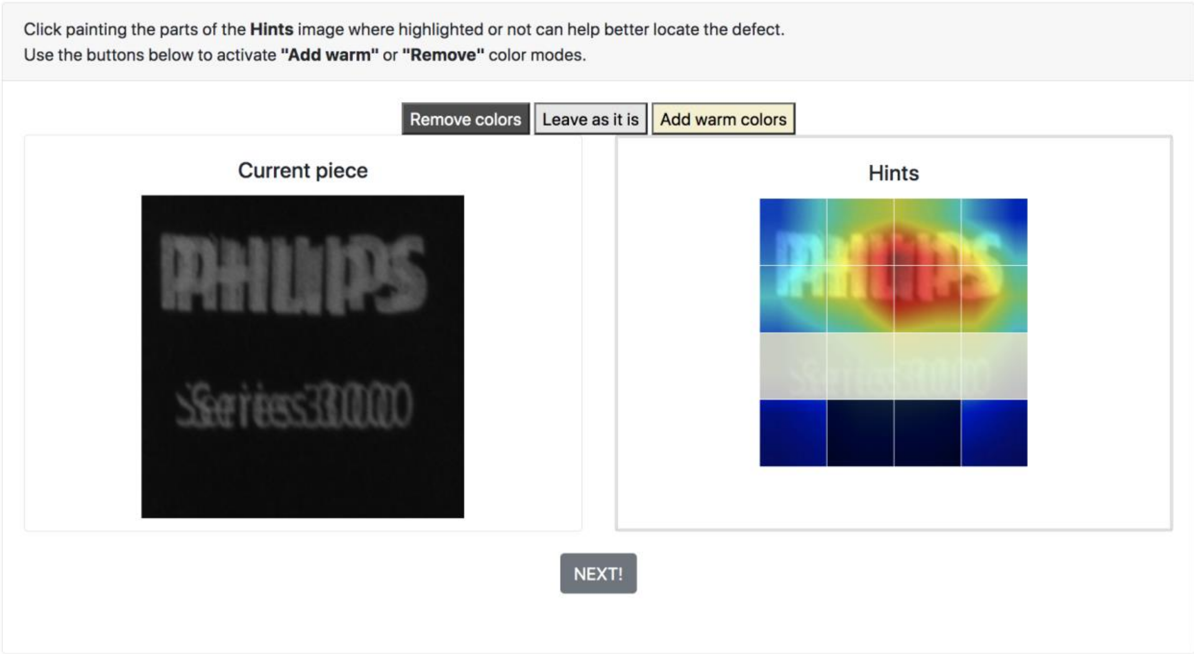

Figure 8: Experiment C with hints marked in the image where the anomaly map should be recoloured

In the image we see six squares marking that the anomaly map should be recoloured. In particular, the participant is asking to remove some spurious warm colours at the middle bottom (marked with a dark colour) and add some warm colours to another four squares at the middle bottom (marked with a light yellow). The example corresponds to Experiment C.

4 Knowledge Base Implementation

The data collected during the experiments was exported in json format, and then converted it into a tabular and graph formats for further processing and to create machine learning models on top of it. While tabular data was used to train machine learning models, we also considered graph information contained in the images. In particular, we characterised each square based on the surrounding squares, as we describe in Fig. 9.

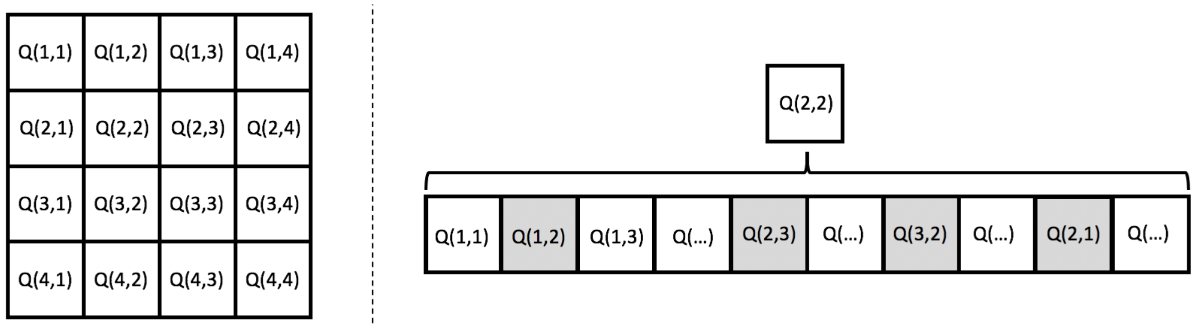

Figure 9: Sample image grid (on the left) and how we translate it to a feature vector (on the right).

When building the graph, we distinguished the following entities:

1. **Image**: represents one of the images labelled during the experiments;
2. **Image Hint**: represents one of the image hints (*GradCAM* or *DRAEM*) provided during the experiments;
3. **Label**: label provided by some participant;
4. **Quadrant**: considering the images and image hints, the quadrant represents one of the squares into which the images or image hints are divided in;
5. **Quadrant Hint**: part of the Image Hint that corresponds to a particular Quadrant;
6. **Quadrant Instance**: part of the Image that corresponds to a particular Quadrant;
7. **Participant**: the person, who took part in the experiments;

The graph was created in the Neo4j graph database, so that could be queried. We ingested a about 20500 nodes, and 3200 relationships between them. We show a diagram describing the relationships between the nodes in Fig. 10, and a screenshot of the graph in Fig. 11.

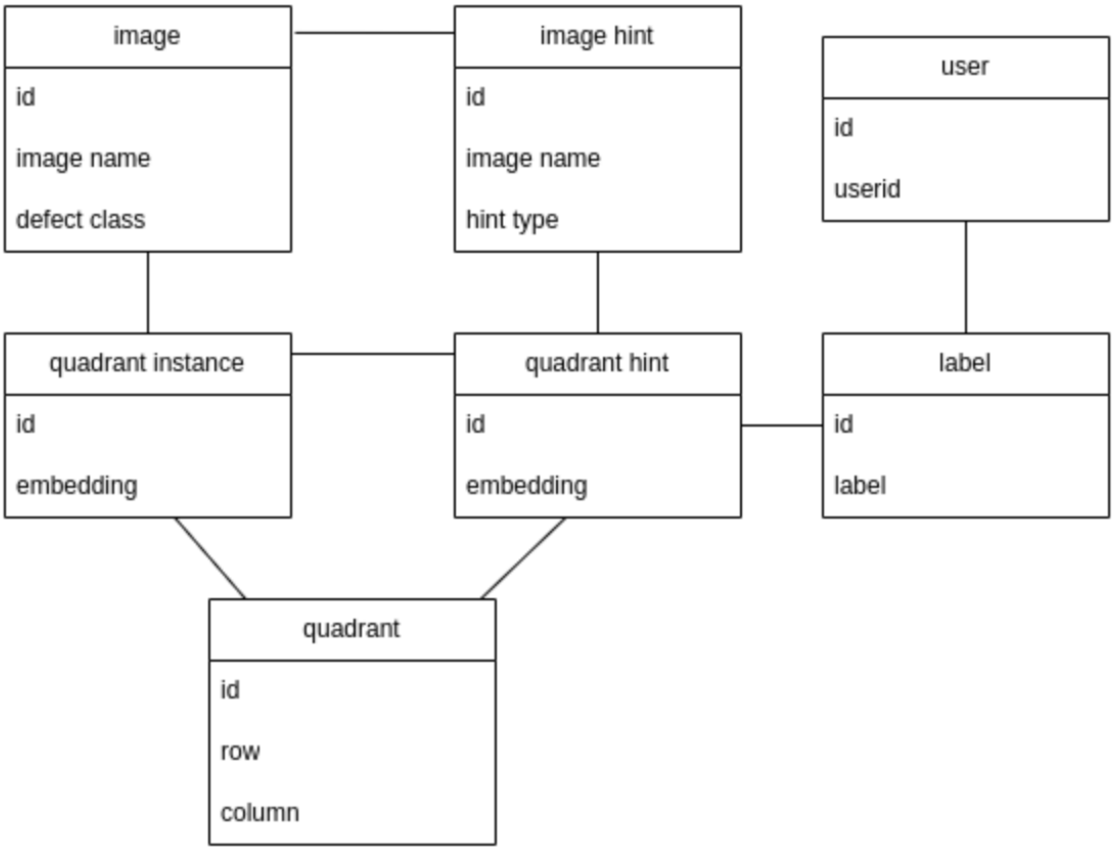

Figure 10: Diagram of entities considered when building a graph.

Challenges we faced

When consolidating the participants' annotations, we noticed that the annotations where heterogeneous. We therefore aimed to use them to build three types of predictive models:

a) models at participant level, considering input from a single participant and predicting future annotations based on their some of the annotated images
b) models at cluster level, we searched whether some participants would be similar enough to train a model based on their annotations, and predict their future annotation behaviour;
c) global models, that considered images from all participants, and aimed to predict future annotations for particular participants.

We observed that some participants tended to label GradCAM heatmaps or DRAEM anomaly maps emphasising only that warm colours should be added or removed. Such a polarisation posed a challenge when evaluating the machine learning models, given two clearly distinct groups could be considered based on this observation. Nevertheless, their behaviour regarding which parts of the images were annotated was much more heterogeneous. Moreover, the fact that not all of the participants annotated the same set of images it posed an additional challenge.

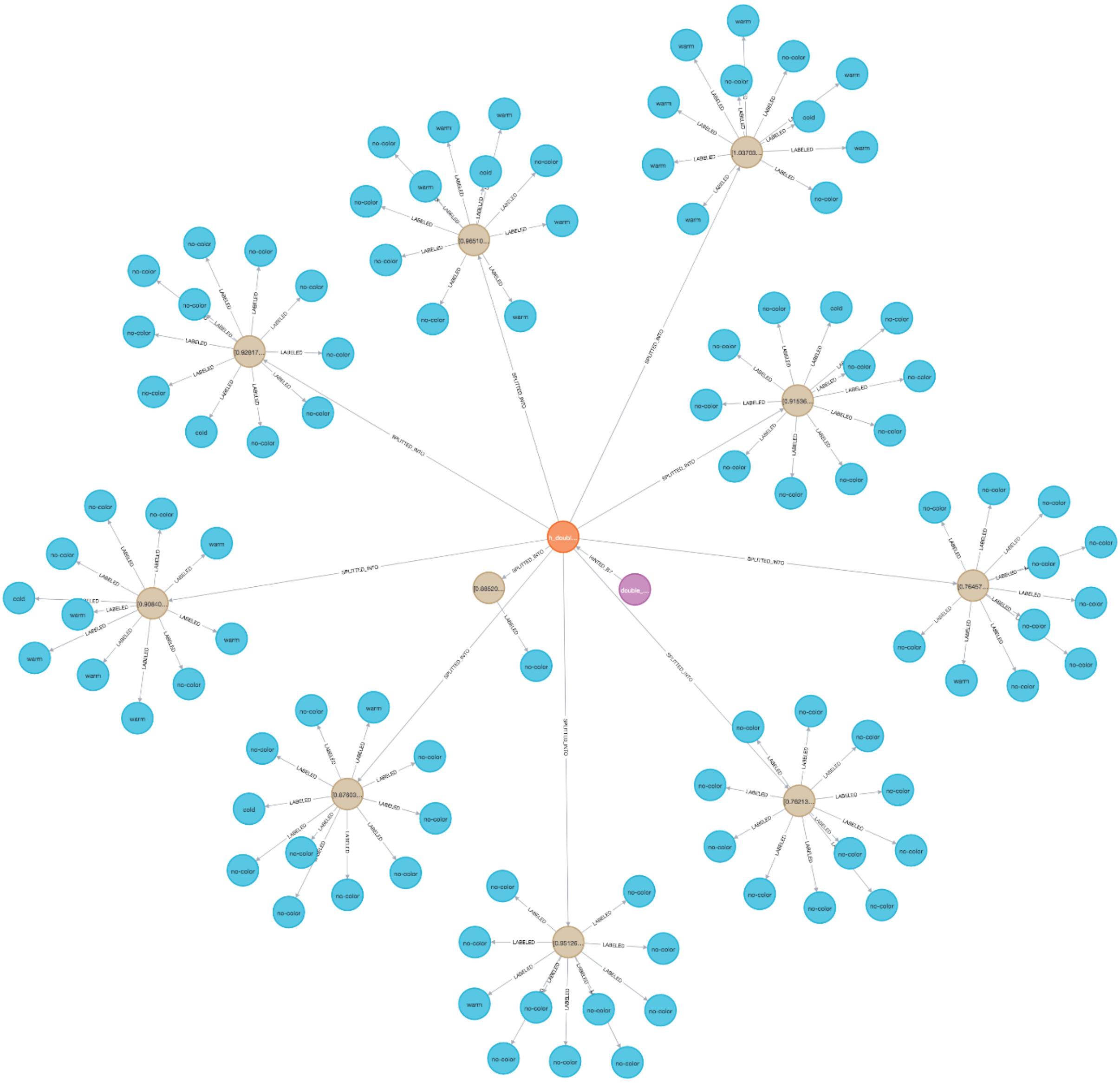

Figure 11: Screenshot of Neo4J graph we created.

5 Application to STAR Use Cases

5.1 PCL Use Case

5.1.1 Application Scenario

This study focuses on the visual inspection task conducted on shavers manufactured by Philips Consumer Lifestyle BV. The objective of this non-destructive quality control testing is to assess the cosmetic aspect of product quality, specifically checking if the company logo of Philips Consumer Lifestyle BV is printed correctly. The inspection process involves the use of various pad-printing setups by the company. With a high volume of daily production, numerous products undergo handling and inspection, and any defective items are removed from the production line. To facilitate this research, a dataset of 3,518 images was provided. The original images were categorised into three labels (good (no defect detected), double printing, and interrupted printing). Nevertheless, the experiments aimed to achieve more detailed labelling, specifically identifying whether any defects were observed at the quadrant level. Therefore, the labels mentioned above were not useful for the experiments conducted in this study. The only exception to this would be Experiment A, where coarse labels could be used to compare whether a defect label was assigned by the participant to a specific square and determine if the corresponding image was defective (fine grained and coarse grained labelling should converge, if annotation was correctly performed).

5.1.2 Evaluation and Results

We assessed the usefulness of the information collected and stored in the knowledge base by developing machine learning models, that were used to predict participants' perception regarding the anomaly or heatmaps showed to them. We developed three model types: (a) per participant, (b) per cluster, and (c) a global model (built with inputs from all of the participants). The idea behind these three types of models was that more data is available and therefore can lead to better results. In particular, to build a model per cluster, we first ran some clustering algorithm to find participants with similar labelling behaviour, and then trained the machine learning classifier on that data. The global model, on the other hand, was trained on all data available.

We considered different sets of features: (i) images and quadrant characterisation as shown in Fig. 9; (b) anomaly map / heat map and quadrant characterisation as shown in Fig. 9; and (c) image, anomaly map / heat map and quadrant characterisation as shown in Fig. 9. The features describing the image and anomaly /heat map quadrants were obtained by using the average pooling layer of a ResNet18 pretrained model. For clustering, we considered a HDBSCAN model [REF-04], computing clusters based on a pairwise similarity. We represented the participants with vectors, which described the participants' rating when considering quadrants for the images displayed to them.

Our results show, that it is possible to successfully forecast participants recolouring choices ahead of time, regardless the model and set of features considered. Further research is required to improve the quality of the predictions, study recolouring strategies and assess how recolouring is then perceived by the participants.

6 Conclusions

In this deliverable we present the recent developments for STAR WP4 Task 4.4 Knowledge Base for Human Robots Interactions. The main development of Task 4.4 has been focused on PCL use case, on quality inspection of Phillips manufactured products (Philips Consumer Lifestyle BV). Quality control allows companies to verify the products' conformance to requirements and specifications and thus build customer satisfaction and the brand's reputation.

The outcomes of Task 4.4 Knowledge Base for Human Robots Interactions are intended to influence not only project partners from STAR use cases but the concepts derived from this research aim to become the basis for future development of knowledge bases in the domain of manufacturing.

While the outcomes are of particular value to the case of visual inspection, we consider they can be generalised to explainable artificial intelligence techniques related to image classification, which can be enhanced considering the concept of personalised, adaptive interfaces.

References

Reference	Name of document
[REF-01]	Ran, Yongyi, et al. "A survey of predictive maintenance: Systems, purposes and approaches." arXiv preprint arXiv:1912.07383 (2019).
[REF-02]	Perner, P. (1994). A knowledge-based image-inspection system for automatic defect recognition, classification, and process diagnosis. Machine Vision and Applications, 7(3), 135–147. doi:10.1007/bf01211659
[REF-03]	Johnson, Teegan L., et al. "How and why we need to capture tacit knowledge in manufacturing: Case studies of visual inspection." Applied ergonomics 74 (2019): 1-9.
[REF-04]	Campello, Ricardo JGB, Davoud Moulavi, and Jörg Sander. "Density-based clustering based on hierarchical density estimates." Advances in Knowledge Discovery and Data Mining: 17th Pacific-Asia Conference, PAKDD 2013, Gold Coast, Australia, April 14-17, 2013, Proceedings, Part II 17. Springer Berlin Heidelberg, 2013.

References to STAR WP4 Documents

Reference	Name of document
D4.1	Library of XAI algorithms-Initial version
D4.2	Library of XAI algorithms-Final version
D4.3	Simulated Reality for Human Robot Collaboration
D4.4	Active Learning Systems and Techniques Initial Version
D4.5	Active Learning Systems and Techniques Final Version