CHALMERS
UNIVERSITY OF TECHNOLOGY

Functional safety approval of an AI-based person-detecting safety product

Bachelor thesis in Electrical Engineering

Carl Ericsson Klein,
Edvin Alverstedt Karlsson

DEPARTMENT OF ELECTRICAL ENGINEERING
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2026
www.chalmers.se

Degree project report 2026

Functional safety approval of an AI-based person-detecting safety product

A conceptual study based on functional safety standards

Carl Ericsson Klein, Edvin Alverstedt Karlsson

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Electrical Engineering
Chalmers University of Technology
Gothenburg, Sweden 2026

Functional safety approval of an AI-based person-detecting safety product
A conceptual study based on functional safety standards
Carl Ericsson Klein, Edvin Alverstedt Karlsson

Supervisor: Johan Maté, RISE
Supervisor: Johan Hedberg, RISE
Supervisor: Kaan Okumus, Chalmers University of Technology
Examiner: Marija Furdek Prekratic, Department of Electrical Engineering

Degree project report 2026
Department of Electrical Engineering
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden
Telephone +46 31 772 1000

Cover: Schematic illustration representing the conceptual development of an AI-based person detection system.

Typeset in LaTeX
Gothenburg, Sweden 2026

Abstract

The use of artificial intelligence (AI) across different sectors has rapidly increased in recent years, yet current functional safety standards are built based on deterministic systems and do not directly align with machine learning (ML). In collaboration with RISE Research Institutes of Sweden AB an investigation was carried out concerning how an AI-based person detection system could be developed and evaluated in accordance with existing safety standards.
A convolutional neural network (CNN) based person identification system was developed in the project according to principles of ISO 12100, ISO 13849-1 and ISO 8800, where traceability, repeatability and documentation was prioritized. The CNN was trained on different datasets consisting of a set of individuals in a static environment. The main goal of the model was to classify each detected individual as authorized, unauthorized, or unknown.
Three different versions of the system were developed, each building upon the previous. The final version achieved consistent performance, with the highest accuracy of 97%. This demonstrated that an AI system can behave predictably under controlled conditions and could potentially be integrated into safety applications.
However, the development also highlighted several limitations of the system. The dataset was limited in size and diversity, and the evaluation was conducted within a confined and controlled environment. This ultimately meant that testing and validation outside of the predefined conditions were lacking, which restricts the generalization of the results.
In conclusion, the study presented a conceptual development of an AI system in accordance with safety standards and that the AI system could potentially support the case of AI integration into safety applications, but to a certain extent. More extensive work regarding expansion and quality of dataset, wider testing and further adaptation of standards is a necessity for a full accommodation of AI in safety systems.

Keywords: Artificial intelligence, Functional safety, Machine learning, ISO 13849-1, ISO 8800, ISO 12100.

Acknowledgements

This degree project has been conducted in 2025 by two Bachelor's mechatronics students at the department of Electrical Engineering at Chalmers University of Technology. The degree project comprises of 15 credits and was carried in collaboration with RISE Research Institutes of Sweden AB.
The degree project was based upon the need for an investigation of how artificial intelligence could be applied in a safety application and what potential risks may be involved.
We are very grateful for the opportunity to have been a part of this project, and we want to thank Johan Maté and Johan Hedberg for assisting us throughout this as supervisors at RISE.

Edvin Alverstedt Karlsson & Carl Klein, Gothenburg, November 2025

List of Acronyms

Below is the list of acronyms that have been used throughout this thesis listed in alphabetical order:

AI	Artificial Intelligence
AISR	AI Safety Requirements
CNN	Convolutional Neural Network
FAR	False Acceptance Rate
FRR	False Rejection Rate
GSN	Goal Structuring Notion
ISO	International Organization for Standardization
KPI	Key Performance Index
ML	Machine Learning
PL	Performance Level
SIS	Svenska Institutet för Standarder
SRS	Safety Requirements Specification
V&V	Verification & Validation

Nomenclature

Below is the nomenclature of input/output space, model, and metrics that have been used throughout this thesis.

Input / output space

x	Input image sample
X	Input space of images
H	Image height in pixels
W	Image width in pixels
C	Number of image channels
y	Class label
Y	Set of class labels

Model

f_θ	Trained neural network with parameters θ
θ	Model parameters
$\mathcal{P}(Y)$	Probability distribution over class labels

Metrics

TP	True positives
TN	True negatives
FP	False positives
FN	False negatives

Contents

List of Acronyms **ix**

Nomenclature **xi**

List of Figures **xv**

List of Tables **xvii**

1 Introduction **1**
1.1 Background . . . 1
1.2 Purpose . . . 1
1.3 Scope and limitations . . . 2
1.4 Research questions . . . 2

2 Theoretical and technical background **3**
2.1 Conceptual system overview . . . 3
2.2 Problem formulation and system model . . . 3
2.3 Deterministic and probabilistic system in functional safety . . . 4
2.4 Functional safety standards and AI-based systems . . . 5

3 Methods **7**
3.1 Product application & concept . . . 7
3.1.1 Product application . . . 7
3.1.2 Risk assessment & reduction measures . . . 8
3.2 AI safety & criteria . . . 9
3.2.1 Assurance argument . . . 9
3.2.2 AI safety refinement . . . 10
3.3 Dataset & database . . . 11
3.3.1 Dataset requirements . . . 11
3.3.2 Implementation of dataset & database . . . 12
3.4 AI system architecture . . . 12
3.5 AI model & system modules . . . 13
3.5.1 AI model - CNN . . . 13
3.5.2 Recognition module . . . 14
3.5.3 Main program integration . . . 14
3.5.4 Learning curve . . . 14

4 System testing and validation **15**
4.1 Testing methodology . . . 15
4.1.1 Experimental setup . . . 15
4.1.2 Evaluation metrics and classification report . . . 16
4.2 Version 1.2 – Baseline . . . 17
4.2.1 Dataset and preprocessing . . . 17
4.2.2 Model training . . . 17
4.2.3 Results and observations . . . 18
4.3 Version 1.3 – Enhanced classic training . . . 18
4.3.1 Improvements . . . 18
4.3.2 Results and analysis . . . 19

5 Discussion **21**
5.1 Overview . . . 21
5.2 Analysis of results . . . 21
5.3 Dataset and preprocessing . . . 22
5.4 System reliability and functional safety context . . . 23
5.5 Limitations . . . 24
5.6 Integration of AI in accordance with safety standards . . . 25
5.7 Future work . . . 25
5.7.1 Dataset . . . 26
5.7.2 Architectural development . . . 26
5.7.3 Verification and validation . . . 26
5.8 Reflection . . . 27

6 Conclusion **29**

Bibliography **31**

A Functional safety documentation **I**
A.1 Risk analysis assessment . . . I
A.2 Risk reduction and residual risk . . . II
A.3 Functional safety concept . . . III
A.4 Safety functional list . . . IV
A.5 Safety requirement specification . . . IV
A.6 AI safety requirement . . . V
A.7 Dataset requirements . . . VI
A.8 Software . . . VII
A.9 Confusion matrix . . . VII
A.10 Assurance argument . . . VII

B Experimental results **XI**
B.1 Training and validation loss curves . . . XI
B.2 Model accuracy . . . XIII
B.3 Confusion matrix . . . XVI
B.4 Per-class metrics . . . XVIII

List of Figures

3.1 Conceptual functional logic of the AI-based access control system, demonstrating input signals, safety-related sub-functions and controlled outputs. 8
3.2 Black-box representation of AI-based access control system, illustrating external inputs, internal processing and safety-related outputs. NO (normally open), NC (normally closed) and COM (common) denote the relay output terminals used to control the gate mechanism. . 8

B.1 Training and validation loss for resized variant model. XI
B.2 Training and validation loss for greyscale variant model. XII
B.3 Training and validation loss for cropped variant model. XII
B.4 Training and validation loss for cropped_greyscale variant model. . . XIII
B.5 Classification accuracy across training epochs for resized variant model.XIII
B.6 Classification accuracy across training epochs for greyscale variant model. XIV
B.7 Classification accuracy across training epochs for cropped variant model. XIV
B.8 Classification accuracy across training epochs for cropped_greyscale variant model. XV
B.9 Average confusion matrix for resized model XVI
B.10 Average confusion matrix for greyscale model XVII
B.11 Average confusion matrix for cropped model XVII
B.12 Average confusion matrix for cropped_greyscale model XVIII
B.13 Average per-class metrics diagram for resized model XVIII
B.14 Average per-class metrics diagram for greyscale model XIX
B.15 Average per-class metrics diagram for cropped model XIX
B.16 Average per-class metrics diagram for cropped_greyscale model . . . XIX

List of Tables

A.1 Initial risk assessment according to ISO 12100, S=Severity, F=Frequency, P=Probability, A=Availability to dodge I
A.2 Risk reduction measures and residual risk evaluation II
A.3 Functional Safety Concept derived from risk analysis (SFC) III
A.4 Safety Functions List (SFL) IV
A.5 Safety Requirements Specification (SRS) IV
A.6 AI Safety Requirements V
A.7 Dataset safety requirements for AI-based access control system VI
A.8 Software implementation and usage VII
A.9 Confusion matrix over version 1.2 and 1.3 with each variant in accuracy order VII
A.10 Assurance argument for AI-based access control system VIII

1
Introduction

1.1 Background

In recent years, artificial intelligence (AI) has evolved from a research topic to a transformative force across global industries. While these advancements present new opportunities, they have also introduced new challenges regarding compliance with existing regulations, laws, and standards. Since there is currently no universal standard that explicitly governs AI-based safety functions in machine-learning (ML) systems, such applications must instead be interpreted through existing functional-safety standards. In this thesis the system is related primarily to ISO 13849-1:2023 – safety of machinery – safety-related parts of control systems – part 1: General principles for design [1] and ISO 12100:2010 – safety of machinery – general principles for design – risk assessment and risk reduction [2]. These standards are fundamentally developed for deterministic control systems, whereas AI-based systems are probabilistic and data-driven, which creates a mismatch when applying them directly.

Although AI-related standardization is actively evolving and several initiatives are underway – for example within SIS TK 421 (Artificial intelligence) – this reinforces the need for continued development for future frameworks that more directly addresses the AI-enabled safety functions.

In parallel, ISO 8800:2025 – road vehicles – safety and artificial intelligence [3]: incorporate elements of AI in the automotive sector. While its scope is limited to road vehicles, certain principles such as structured data management, assurance argument and verification and validation model (V&V) of AI components may be transferable and could be used in combination with ISO 12100 and ISO 13849-1 to assess AI-enabled safety systems in the industrialized sector.

1.2 Purpose

This thesis aims to develop an AI-based system for identifying authorized individuals in a restricted industrial environment using facial recognition, while also investigating how future machine learning safety standards could be structured to support such systems.

Since formal safety standard machine learning are still under development, the work further purposes a conceptual framework intended to align with anticipated requirements for AI-based safety systems.

1.3 Scope and limitations

To maintain the thesis within the given timeframe, certain aspects need to be disregarded. This includes the creation of an extensive dataset, since the manual data collection might pose practical limitations in terms of available time and resources. Certain parts of the Verification & Validation (V&V) method described in ISO 8800, Clause 12 may also need to be partially excluded or simplified due to the limited project duration.

1.4 Research questions

In this thesis, the following questions are considered:

- How can an AI-based person detection system be developed and structured to align with the existing functional safety standards such as ISO 12100, ISO 13849-1 and ISO 8800?
- What challenges arise when applying these frameworks to AI-driven systems?
- Based on the development and evaluation of the system, what conclusions can be drawn regarding the structure and the requirements of future standards for a safe integration of AI?

2

Theoretical and technical background

2.1 Conceptual system overview

The AI-based person detection system under study is intended to support access control to a restricted industrial area, observing individuals approaching and determining whether access should be granted based on visual affirmation and identification.
The visual input received by the system is provided by an external fixed camera monitoring the entrance or hallway leading towards the restricted area. The system processes this information using an ML-based model to classify individuals into predefined categories, such as authorized or unauthorized individuals.
The output from the AI safety system is a probabilistic classification result, which is integrated by the systems safety-related decision logic. The intended application of the safety system is to correctly identify individuals to prevent unauthorized entries into the hazardous zone, while granting access to authorized personnel under defined conditions.
The system under study is evaluated as a conceptual prototype focusing mainly on the AI-based safety decision-making function. Hardware mechanisms and other low-level control logic is outside the scope of this thesis.

2.2 Problem formulation and system model

The objective of the system is to evaluate visual observations of individuals and determine access authorization in accordance with safety-related applications, based on predefined classification categories. The input to the system is represented as:

$$x \in X \subseteq \mathbb{R}^{H \times W \times C} \tag{2.1}$$

The input image x is the representation of the visual observation and where H and W are the pixel height and width, and C is the number of colour channels (three for RGB). The input space X consists of all admissible input images under the defined operating conditions. This type of representation is standard and in accordance with image classification problems [4].
The expected output is label y, representing finite set of possible categorical classes Y. Y in thesis represents the access clearance classes, authorized and unauthorized.

$$y \in Y = \{\text{authorized}, \text{unauthorized}\} \tag{2.2}$$

The objective for the CNN is to map the input images to an estimation of the class label:

$$f_\theta : X \to P(Y) \tag{2.3}$$

Where f_θ is a CNN parameterized model and $P(Y)$ represents probability distribution over the defined set of classes. Standard practice in machine learning is the usage of probabilistic outputs to represent confidence measurement for each set class [4, 5].
The probabilistic output of the model is an intermediate representation of safety actions where the outputs are converted into a discrete system action by a decision rule, $d()$.

$$d\Big(f_\theta(x)\Big) \to \{\text{grant}, \text{deny access}\} \tag{2.4}$$

This separation between classification and decision-making is a fundamental concept and conversion in ML classification systems [4].
When an input is incorrectly classified, resulting in access being granted to an unauthorized individual in a hazardous area, a critical failure mode occurs from a functional safety perspective. This misclassification represents the primary safety risk addressed in this thesis. In contrast, incorrectly denying access to an authorized individual affects system usability rather than safety.
To evaluate the system model in terms of performance and reliability within the scope of a conceptual prototype, the analysis is performed under constrained assumptions, including the camera viewpoint, the physical environment and the limited set of individuals. These assumptions enable focused assessment of the AI-based system's safety-related decision-making function.

2.3 Deterministic and probabilistic system in functional safety

According to the ISO standards 12100 and 13849-1 [1, 2], traditional functional safety systems are based on deterministic behaviour, and assume that under identical conditions, identical inputs will result in reproducible and predictable outputs. This indicates that the system responses for defined operations can be fully analysed and verified.
However, ML-based systems derive their behaviour from statistical learning over data. Established ML literature describes model outputs as probabilistic and a representation of confidence estimations rather than deterministic decisions [4]. Consequently, even identical inputs may result in different outputs, this is particularly when variations arise in data or internal model states.
The combination of both probabilistic and deterministic nature introduces challenges for combined safety-related applications, where the predictability, transparency and

robustness of a system is essential. According to the European commission's high-level expert group (HLEG), AI-based systems may exhibit complications in certain aspects of their functionalities which may impede their usage in safety-critical contexts. The HLEG highlights such systems limitations in generalization and sensitivity to rare or unforeseen inputs [6].
There are fundamental differences between the probabilistic and deterministic system behaviour. Deterministic systems operate based on predefined logical rules, producing identical outputs for identical inputs and allowing full system verification. In contrast, probabilistic AI-based systems generate outputs based on statistical inference and confidence values, introducing uncertainty in prediction outcomes.
When applying traditional functional safety framework directly to AI-based systems, complementary measures are therefore required, such as dataset governance, traceability, V&V methods adapted for data-driven models and controlled input spaces.

2.4 Functional safety standards and AI-based systems

As discussed in previous sections, the probabilistic and data-driven characteristics of AI-based systems pose fundamental challenges when assessed using traditional functional safety standards. These ISO standards assume deterministic and fully specifiable system behaviour, an assumption that does not hold for learning-based systems.
ISO standards such as 12100 and 13849-1 offer limited guidelines on the assessment of safety functions whose behaviour derives from statistical inference and training data. However, they provide established safety validation methods for machinery systems [1, 2]. Consequently, direct application of these standards to AI-based safety functions is insufficient for safety-critical assessments. Instead, compliance must be achieved through interpretation of the standards and the introduction of complementary safety measures.
ISO 8800 introduces AI components within the automotive domain and presents assurance arguments and concepts addressing the safety aspects of AI systems, such as traceability and data management [3]. In parallel, the European Commission's (HLEG) on AI highlight transparency, robustness and accountability as core requirements [6]. Although both initiatives address gaps in safety frameworks, they do not yet constitute a unified AI-based safety standard. However, in conjunction with each other they indicate a shift from deterministic verification towards assurance-based safety framework. This shift motivates the usage of assurance arguments, controlled input space, and traceability in the development and evaluation of the AI-based safety system.

3

Methods

This chapter describes the methodology applied in the development of AI-based detection system. It outlines the processes used to for dataset creation, system design and integration, model development and evaluation procedures as well as methods to align this project with functional safety standards.

3.1 Product application & concept

This section introduces the conceptual foundation of the project and defines the intended product application. It outlines the initial design of how the system could function within an industrial environment as well as how operational purpose and safety role were defined.

3.1.1 Product application

The initial stage of this project was to design and describe an application for intended usage of the product. After discussion and approval by the supervisor at the company the product application was designed as follows: An AI-based safety system consisting of a camera intends to detect and identify individuals seeking access to a restricted gated area. Inside this area, a rotating machine is stationed. The individual with access clearance is identified by the camera outside of the area, after which the gate separating the individual is unlocked. At the same time the machine transitions into a low energy mode that is intended for personnel who are trained and authorized to operate in. Access clearance is evaluated only for a limited and predefined group of approved individuals. However, the system must also be capable of detecting and denying access to all other persons outside of this group. Consequently, while the authorized user group is limited, the semantic input space remains broad and cannot be considered a fully controlled domain.
Figure 3.1 illustrates the conceptual functional logic of the proposed system, highlighting the relationship between inputs, safety-related sub-functions and system outputs.

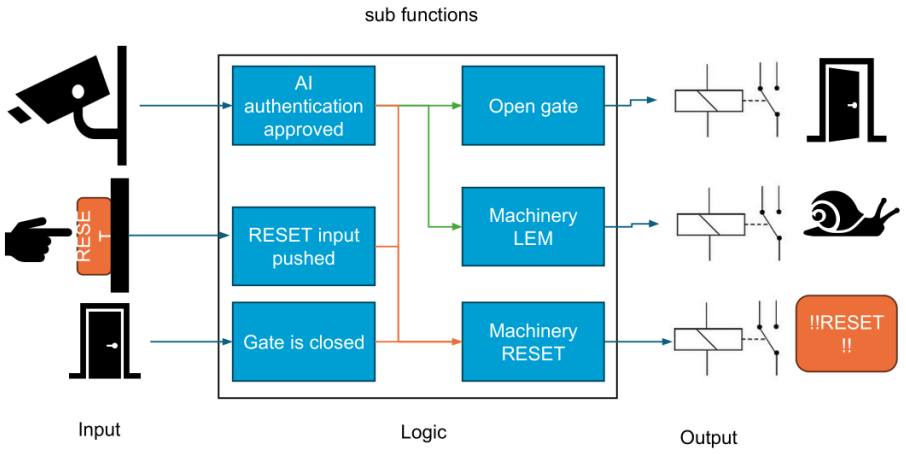

Figure 3.1: Conceptual functional logic of the AI-based access control system, demonstrating input signals, safety-related sub-functions and controlled outputs.

To further clarify the system boundaries and external interfaces, the overall system can be represented as a black-box model.

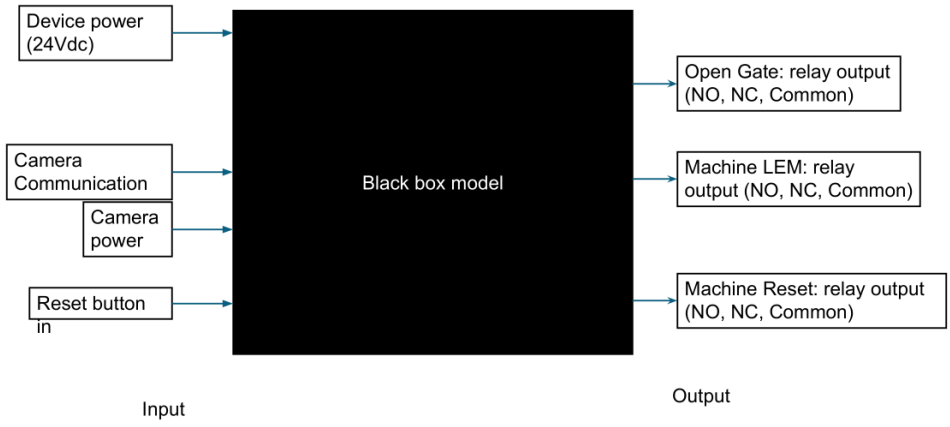

Figure 3.2: Black-box representation of AI-based access control system, illustrating external inputs, internal processing and safety-related outputs. NO (normally open), NC (normally closed) and COM (common) denote the relay output terminals used to control the gate mechanism.

3.1.2 Risk assessment & reduction measures

After completion of the product application description, several principles in accordance with ISO standards 12100, 13849-1 and 8800 were used to evaluate risk/-

hazards and to implement safety measurements and functions for reduction. This included risk analysis, functional safety concept, safety function list and safety requirement specification.

Risk analysis

The risk analysis identifies relevant hazards, operating conditions and potential hazardous events to determine if those can be reduced by appropriate safety measures. Each identified hazard is evaluated and assigned a required performance level (PLr) in accordance with Annex A in ISO 13849-1. The PLr is then used to specify the necessary safety functions and their required level of risk reduction. According to ISO 13849-1, Clause 5.1 "The process for specifying safety functions requires detailed information from the risk assessment performed in accordance with ISO 12100". The complete risk analysis performed in this project is presented in Table A.1–A.2, which shows the identified hazards, operating scenarios, associated PLr values, and the corresponding safety functions derived from the assessment.

Functional safety concept

In the functional safety concept, each function and component was listed along with a description, associated hazard, safety objective, risk reduction measure and standard reference seen in Table A.3. Here, the goal is to accurately define how safety features and functions can be implemented to reduce risks. It is a structured framework based on the first step, the risk analysis according to standard ISO 13849-1. It works by identifying the different safety functions, designing the control system and validating the system. This is used to safeguard the operators and to meet legal regulations regarding the machinery.

Safety function list

According to ISO 13849-1, Clause 4.2, From the risk assessment, the designer shall decide the contribution to the risk reduction provided by each relevant safety function [1]. The Safety function list is a structured overview to identify all the necessary safety functions to reduce the risk to an acceptable level. It ensures risk reduction traceability and allows the designer to link hazards to actions seen in Table A.4.

Safety requirements specification (SRS)

The SRS is a critical link between the safety function list and the system implementations. It defines the safety functions and their functionalities. That includes description of input, logic output, PL level and test method. The SRS ensures that the design can be tested, verifiable and that the safety functions are consistently implemented, see Table A.5. This can later be used as a reference for verification. According to ISO 13849-1 Clause 5.1, "the description shall be linked to hazards identified in the risk assessment and shall state how the function operates to achieve the required safety. The process for specifying safety functions requires detailed information from the risk assessment performed in accordance with ISO 12100" [1, 2].

3.2 AI safety & criteria

3.2.1 Assurance argument

To ensure that the system was developed and evaluated in a structured way, a Safety Assurance Argument was established in accordance with Clause 7 and 8 in ISO 8800

[3].
The assurance argument was represented using the goal structuring notation (GSN) methodology. GSN provides a graphical and logical framework that links safety goals to the context, strategies, assumptions and evidence, creating an overlooking perspective that is useful when developing a system.
The argument was structured such as: A top-level goal (G1) was defined: "The AI-based person-detection system shall be acceptably safe for intended operational design domain (ODD). This goal was decomposed according to a strategy (S1) into sub-goals covering key safety aspects identified during risk analysis:
- (G2) Fail-safe control logic behaviour
- (G3) AI model reliability and consistency
- (G4) Dataset quality and traceability
- (G5) Verification and validation of AI performance
- (G6) Identification of residual risks and improvement plan

Each goal was supported by contexts (C1-C6) describing the environment, operating limits and data conditions.
Assumptions about the system where also listed (A1-A5) covering external dependencies such as hardware integrity.
Finally, evidence (E1-E5) was linked to each goal later at the end of the project, justifying the fulfilment of goals through testing and results.
The full assurance argument can be seen in Table A.10.

3.2.2 AI safety refinement

In this chapter the focus shifts from general safety analysis of the system to the AI system itself. Here each function of the AI is assessed based on risk, trigger conditions, safety requirement, component requirement, acceptance criteria and KPI (Key Performance Indicator)/metrics, see Table A.6. The acceptance criteria and requirements established for each function are based on several factors and assumptions. Each function represents different cases. The following paragraphs discuss each function with respect to the case and applicable factors.
- **Environment** which the system will operate in is considered consistent since the intended usage will be indoors with a stable artificial light source, this however could change considering natural luminescent sources. The test will revolve around different lux scales giving different results around the same values, recommended usage would be preferable somewhere around 600-900 lux.
- **Accuracy** of the system to distinguish an authorized individual from an unauthorized individual will be tested using different and new images on previous subjects, this will also mean that unknown subjects will be tested in the validation. The intended usage of this product would be roughly ten times per year. FAR (false acceptance rate) would preferably result in $\leq$1% of unauthorized individuals to be granted access, resulting in $\leq$1 per ten years. FRR (false rejection rate) would preferably deny $\leq$5%, giving the authorized $\geq$95% accuracy to be granted entry upon attempt; with the intended usage this would result in five denied entries per ten years.

• **Response time** for the entire system and its full complete scan and approval of an authorized individual will be set to roughly 15s. This includes detection of individuals, authentication of said individuals and the signal from which the systems logic lock unlocks.
• **Detection – Authentication** From the point that the system identifies an individual to its decision whether the individual is authorized the time interval will be roughly 15s at the most, if authentication still has not been authorized the individual will be denied entry and new detection will reset the time. From the point of authentication verified to the logic gate being unlocked will be ≥250ms.
Apart from assessing each safety function according to the table, an input space definition was carried out to clarify the conditions under which the AI system is expected to operate. The input space was described at two levels; the semantic level, which defines the meaning and variability of the real-world situations represented in the images (people, poses, movement within the scene), and the syntactic level, which specifies the allowed technical properties of the input data (image format, resolution and structure).
Semantic description of input space
• The system receives a visual input in the form of a video stream from a fixed indoor camera. The camera observes a mostly static indoor environment, such as entrance area, where the background and surroundings remain largely unchanged over long periods of time. The primary dynamic element in this scene is people passing through or appearing in front of the camera within its field of view.
Syntactic description of input space
• The input data consists of Red-Green-Blue (RGB) images represented as three-channel array with a spatial resolution of 160 x 160 pixels and 8-bit colour depth (0-255 per channel). The images are stored in Joint Photographic Experts Group (JPEG) format.
Operational and environmental assumptions
• The system is assumed to operate in an indoor environment with stable and predominantly artificial lightning. The monitored area is considered static, with no significant changes in the physical layout over time. Outdoor environment effects such as rain, fog, snow or weather-related variations are not included within the current scope.

3.3 Dataset & database

3.3.1 Dataset requirements

The dataset was created in accordance with the principles of ISO 8800, Clause 11 which defines requirements for the dataset safety as part of the AI Safety Lifecycle. This clause emphasis on dataset integrity, completeness, representativeness and traceability which are essential for a safety related applicating using AI.
To meet these principles a dataset requirement list was created where each dataset property was analysed according to four aspects:
• safety property – the specific characteristic to ensure dataset suitability
• safety related insufficiency – potential dataset flaws that could negatively affect

the behaviour of the system
- metric – a performance target or acceptance criterion
- countermeasure – technical or procedural measurement implemented to satisfy the property

This approach ensured that dataset creation was not only performance driven but also focused on safety see Table A.7.

3.3.2 Implementation of dataset & database

The implementation of the Dataset_collection.py function was created to easily capture images and label them within the system for accessible traceability and safety. Initially, the dataset was organised as a multi-folder structure to separate data for different processing stages. This structure introduced issues related to file path management. Consequently, the dataset structure was flattened [$dataset/original/ID$]. Each of the images saved was accompanied with metadata information stored in an SQLite database. To maintain traceability, recorded attributes such as timestamp, label, file path, dataset version and SHA-256 checksum. This enables detection of unintended modification for the image in case of corruption during training or during operational usage.
Database management functionality was implemented in a dedicated module (Database.py.) Enabling creation, handling, maintenance and clearance of database records were implemented for easier management an experimentation. The CNN requires a sufficiently varied dataset to prevent overfitting. Therefore, images collected were needed from multiple individuals with different facial orientations, lighting conditions and diversity. In accordance with the guidelines of the industrial partner RISE, Chalmers University of Technology and the ISO safety standards, all individuals included in the dataset had to explicitly consent for image capture and usage of their face. This constraint of ethical and safety requirements created challenges in achieving diversity and optimal model generalisation.

3.4 AI system architecture

The AI-based system architecture defines the interaction between hardware and software components required to achieve safety related functionality previously described in Section 3.1 and in accordance with the safety standards 8800 and 12100, the system follows a traceable pipeline consisting of five integrated subsystems.

1 Data acquisition - A camera-based vision sensor captures visual input, which is processed and stored together with associated with metadata in a local database to ensure full traceability throughout the dataset lifecycle.
2 Dataset management – The collected data is pre-processed and prepared for training and evaluation. Multiple dataset variants are created based on different pre-processing configurations and split into training, validation and test subsets.
3 Model training – The prepared dataset variants are used to train the CNN and create a model. Validation and evaluation are performed in parallel with

training to monitor model performance and detect overfitting.

4 Recognition – The trained model performs person recognition either real-time or on a predefined dataset. Classification is based on probabilistic identification outputs from a SoftMax activation function, and predictions are compared against an authorization list. All recognition results are logged to support traceability.

5 System control – The system control components integrate and connects all subsystem and functions. By handling configuration, user interactions and communication between the recognition output and the simulated safety logic.

3.5 AI model & system modules

3.5.1 AI model - CNN

A CNN is an artificial neural network architecture optimised for analysing visual data, images and videos. The requirements to create a CNN model in this project was to meet the four principal requirements, non-bias in the decision-making, full traceability through the model development, compliance with the safety guidelines of the ISO standards to maintain the safety related properties based on the safety analysis and robustness in the operational performance.

The dataset was divided into training, validation and test subsets, ensuring proper data separation, preventing unintended data leakage, and maintaining traceability throughout the development process. The model training component of the CNN handles training, validation, performance monitoring, and storage of the resulting Keras model for later use in the recognition phase.

The images were pre-processed and augmented before training to implement transformations such as zooming, shifting and flipping. These measures reduce possibility of overfitting as well as improving the system abilities making it more robust. This is done by the function ImageDataGenerator which is a Keras utility.

The CNN architecture is comprised of three convolutional blocks, containing 32, 64 and 128 feature maps respectively. Each convolutional layer uses 3x3 kernels, followed by batch normalization and max-pooling, enabling feature extraction while stabilizing the intermediate activations.

The first convolutional block focuses on extracting low-level features such as edges and simple textures, the second block captures more complex spatial patterns, facial contours, while the third block is responsible for higher-level and more abstract features relevant to facial representation.

Batch normalization is applied after the convolutional layers to stabilize the training process by normalizing the activations. After the convolutional feature extraction, the resulting feature maps are flattened into a one-dimensional vector. A dropout layer is applied before the final classification stage to reduce overfitting by randomly deactivating neurons during training.

The final dense layer uses a SoftMax activation function to convert the output into a probability distribution over the defined classes, ensuring that the predicted class probabilities sum to one.

3.5.2 Recognition module

The recognition module loads the status and class labels of the individuals defined in the dataset. Continuous real-time monitoring is performed to identify authorized individuals by comparing the output to the predefined accuracy criteria. During operation the trained model is loaded, and the live feed is analysed. The extracted values are processed by the model and converted into probabilistic values representing the likelihood that the individual standing in frame belongs to one of the defined classes.
Initially, the system logged all recognition attempts to the SQLite database. However, considering the high frequency of recognition attempts (upwards of ten a second) the database grew too rapidly, leading to an overload in data. The policy surrounding the logging was therefore modified to only store recognition attempts classified as "authorized", resulting in reduced database storage and improved efficiency without removing traceability.
Once the main program calls the recognition function, the trained CNN model is loaded and initialized to begin the recognition. An initial inference pass is performed to prepare the model for real-time prediction. The display image follows the same structure as in the data collection phase, showing the predicted label, confidence score and status. Based on predefined criteria, the classification confidence score must be $\geq 95\%$ for admittance.

3.5.3 Main program integration

The main control module served as the central interface through the entire system. It provides a user-friendly environment for the execution of the modules, dataset collection, model training testing and recognition tasks. It also includes an authorization-list management function, allowing the operator to add, remove or modify registered individuals and their associated access status.

3.5.4 Learning curve

The performance monitoring module visualized training, validation, and test metrics to support evaluation of model behaviour. The results were plotted using python-based visualization tools, enabling clear interpretation and serving as a diagnostic tool to assess the model performance and identify potential overfitting or underfitting behaviour of the CNN.

4

System testing and validation

Following the design and development described in chapter 3, this chapter presents the systematic testing, validation and results of the AI person-detection system developed in this project. The primary objective was to evaluate the accuracy, robustness, and the overall performance of the CNN model based on previously mentioned metrics such as FAR and FRR. Also, evaluate how the results may inform future standardization efforts of AI integrated machine learning safety systems in accordance with safety and traceability requirements in ISO 12100, 13849-1 and 8800. The analysis of the incremental changes in the dataset, training pipeline and preprocessing influence on the systems behaviour. The improvement process was structured across three main system versions (v1.2-v1.4). Version 1.1 served as an initial real-time prototype used to explore model behaviour and dataset feasibility.

- Version 1.1 – Initial prototype using raw, unprocessed input data without filtering or preprocessing.
- Version 1.2 – Basic implementation.
- Version 1.3 – Enhanced classic training pipeline.
- Version 1.4 – AI-augmented dataset with stratified splitting of the dataset into training, validation and test subsets, preserving class distribution to prevent class imbalance between individuals.

Each version builds upon the previous one, introducing gradual changes based on ideas and results conducted from previous versions tests. This also includes improvement of specific identified limitations to enhance performance, safety and reliability.

4.1 Testing methodology

The primary objective for the testing process is to evaluate the ability of the system to identify and classify individuals within a controlled environment. The evaluation process for testing the system structure focusing both on overall system performance and per class behaviour, using standard classification metrics.

4.1.1 Experimental setup

All images were collected in accordance with the safety and ethical guidelines specified in ISO 12100 and ISO 8800. Every version was based on a custom created dataset collected and curated specifically for this project by utilizing the collection function with each metadata and labelling infused. The v1.1 dataset was comprised of 1800 images with each image labelled into four different classes, representing au-

thorized and unauthorized individuals as well as unknown such as static background, objects and noise:

- Carl (650 images)
- Edvin (650 images)
- Jakob (200 images)
- Unknown (300 images)

To create a versatile dataset the environment around each class was altered and manipulated during creation. This included light, noise, orientation, positioning, introducing objects and alter static background conditions.
To preserve dataset integrity and ensure traceability, the v1.1 dataset was copied and split into three subsets: training (70%), validation (20%) and test (10%). This hold-out validation strategy was used to ensure unbiased performance evaluation and to separate model training from validation and testing, while creating a backup copy of the dataset reduced the risk of unintended corruption of the original images.
Based on the results from the first version, mainly focusing on the performance of the systems continuous recognition, these four variants were selected to study the outcome of the preprocessing effect on accuracy and model performance based on the isolated effect of colour information and background complexity:

- Resized – images scaled to 160x160 for a fixed resolution
- Greyscale – images scaled to 160x160, greyscale
- Cropped – images cropped with face only in frame
- Cropped greyscale – images cropped with face only in frame, greyscale

4.1.2 Evaluation metrics and classification report

The reliability and performance of the identification system was assessed by a set of well-established evaluation metrics. These metrics provide complementary perspective on both overall system performance and per class effectiveness. All metric definitions and formulas are based on ISO 8800, Annex H.
Accuracy represents the proportion of correctly classified predictions relative to total number of samples, calculated using formula:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.1}$$

Where TP is True positive, TN True negatives, FP false positives, and FN false negatives. Accuracy provides an overall measurement of the systematic performance, however in certain cases like imbalanced dataset over overfitting, accuracy could be misleading representation of safety related behaviour.
Precision is defined as the ratio of correctly predicted positive observations.

$$Precision = \frac{TP}{TP + FP} \tag{4.2}$$

High precision usually indicates low false-positive rate.
Recall measures the proportion of correctly identified positive instances:

$$Recall = \frac{TP}{TP + FN} \tag{4.3}$$

High recall indicates the system efficiently detects intended class without missing true positives.
The F1-score measurement combines the precision and recall measurement into a single metric.

$$F1 - score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{4.4}$$

This metric is useful when the dataset is imbalanced.
The confusion matrix shown in Figure B.9 visualizes the relationship between predicted and true classifications across the defined classes (Carl, Edvin, Jakob and Unknown). Diagonal cells represent the correct classification, while the off diagonal represent the misclassification between the classes and how they were classified.
The values represent images correctly classified in a series of 5 runs of 65 images, thus representing an average value of correctly classified images.
For example, the cell (TRUE = Carl, predicted = Carl) = 66.0 indicates that 64 images were correctly classified, while 0.8 and 0.2 were mistakenly identified as the individuals Edvin and Jakob.
The loss function represents the error between the models predicted output and the true labels during training and acts like feedback for the parameter optimization. Categorical cross-entropy loss was used to monitor model convergence. Monitoring training and validation loss over time provides insight into the model performance, improvement, stagnation, and potential overfitting.

4.2 Version 1.2 – Baseline

The first system iteration v1.2 served as baseline assessment for the performance of the CNN model using the v1.1 dataset to find flaws within the code and dataset. The test function was implemented to verify the distinction between the variants and the models.

4.2.1 Dataset and preprocessing

V1.1 dataset comprised of 1800 images divided between four classes were copied and split into four variants previously mentioned in section 4.1.1. This direct approach was to ensure diversity in training of each model, creating a random split between each variant's dataset. Test set represented approximately 10% of the total dataset, corresponding to around 180 images per variant but introduced potential bias.

4.2.2 Model training

The lightweight CNN was trained for 10 epochs with batch size of 16, using basic data augmentation with ImageDataGenerator which is a Keras utility, utilizing augmentation techniques such as rescaling, rotation, shifting, zooming and horizontal flipping. Each variant model was trained separately with these settings to determine independent impact on detection accuracy and establishing robustness.

4.2.3 Results and observations

Based on the observations and results, significant variation arose across all dataset variants. Both the Cropped and Cropped + Grayscale variant achieved high accuracy upwards 95-96% which aligns with the required metric. This demonstrates that reduced background noise and more facial features greatly improve performance. Greyscale variant on the other hand greatly dropped in accuracy in comparison, down towards 76% indicating that colour information is an important role in determining the individual.

The two key indications taken from this are that all the splits were derived from the same v1.1 dataset which most likely impacted the systems performance and could reduce robustness against unseen data. This version also revealed the difference random splitting could make as the greyscale variant was re-runed based on a technicality and increased from 76

4.3 Version 1.3 – Enhanced classic training

The second system iteration v.1.3 addressed the identified limitations present in v. 1.2, the consistency, reliability and robustness. Minor changes were initialized within the code, the more notable changes were the increase in epochs, enhanced data augmentation to increase robustness and the average of multiple runs to reduce the inconsistency.

4.3.1 Improvements

The three key improvements that were introduced in version 1.3:

1. Increased epochs and early stopping: Epochs were increased from 10 to 30 to allow deeper convergence. An early stopping function terminates the training when validation loss stopped improving. This reduces and prevents overfitting, as this could pose a risk of incorrectly identification posing safety concerns. To ensure the complete training of each model variant a minimum epoch is required.
2. Consistent experimental conditions: To improve reproducibility and enable controlled comparison between experimental runs, a fixed random seed was introduced. This was primarily used to control randomness arising from data shuffling, mini-batch sampling and training operations such as dropout and data augmentation, rather than weight initialization. The fixed seed was applied to ensure consistent experimental conditions during evaluation. The performance was still assessed over multiple runs to mitigate potential bias.
3. Averaging multiple training runs: Each variant was trained and evaluated over five independent test runs. Results from variants are averaged to provide a more decisive performance estimation. Reducing minor fluctuations and the effect of outline decisions.

4.3.2 Results and analysis

The methodology changed in v1.3 greatly improved both the accuracy of the evaluation and the reliability. The results disclosed that averaging over multiple runs achieves increased stability and robustness of variants and their metrics. The resized variant had an average accuracy of approximately 97.5% over five sequential runs, compared to the single run results with fluctuations of 2-3% observed in v1.2. The classification report remains consistent across classes, with f1-score close to or above 0.97 for all categories.
The resized variant average results achieved an accuracy of 97.49% with a loss of 0.3432. Greyscale slightly outperformed with an accuracy of 97.6%, and a loss of 0.3295. The cropped variant also remained acceptable accuracy 96.72%, loss 0.2605. Cropped greyscale however saw a notable performance decline resulting in an unacceptable accuracy at 86,45%, loss: 4.994 indicating the removal of background and colour servery limits models' ability to distinguish features.
Additionally, as demonstrated by the confusion matrix seen in Table A.9, the balance between classes saw an improvement and a reduction in misclassification could be observed, especially for the "unknown" category, which represents individuals without access authorization. Misclassification of this category corresponds to the safety-critical failure mode, where unauthorized individuals are incorrectly granted access. The improved generalization ability of the model was contributed to the deeper convergence and enhanced pipeline.
While overall accuracy provides an indication of functional performance (design accuracy), it does not fully reflect the safety-related behaviour of the system. From a safety perspective, the critical requirement is that individuals without authorization are not granted access. Therefore, safety-critical accuracy is defined as the proportion of "unknown" instances correctly classified as non-authorized. Based on the confusion matrix, notable differences can be observed between variants when evaluated under this criterion, highlighting that models with similar overall accuracy may exhibit substantially different safety-critical performance [7].
Despite these gains, the evaluation remains limited using the training, validation and test sets derived from the same dataset version, which may introduce optimistic performance estimates. More importantly, the dataset does not fully represent the semantic diversity of real-world unauthorized individuals (e.g., visitors, children and contractors), which constitutes a known limitation in AI-based safety systems. Addressing dataset completeness and dataset independence was beyond of the scope of this prototype and is therefore treated as an acknowledged safety limitation rather than a mitigated or resolved risk. To address this, future work v1.4 will expand class distribution, include stratified splitting and incorporate AI augmented data images to strengthen the systems robustness and increase divergency.

5
Discussion

5.1 Overview

This chapter interprets and discusses the results obtained from the development and evaluation of the AI-based person detection system in relation to functional safety principles. The discussion focuses on the reliability, robustness, dataset design and system behaviour, and evaluates how the results align with the intent of ISO standards ISO 8800, ISO 12100 and ISO 13849-1. Rather than introducing new experimental results, this chapter critically examines the strengths and limitations of the findings in chapter 4, possible areas of future improvements and alterations. Emphasis on safety-related behaviour and applicability of existing standards to data-driven AI systems.

5.2 Analysis of results

The results obtained from version 1.2 and version 1.3 demonstrate clear improvement in classification performance and training stability, reflected by higher accuracy and reduced variability between runs. The first version of the baseline model, v1.2 showed a performance variation between runs with each preprocess variations tested, accuracy fluctuating between roughly 76-96%. These inconsistencies were primarily caused by random factors such as dataset shuffling which caused inconsistent conditions for the experiment [4,5]. Using a fixed number of epochs caused training to continue regardless of improvement by the model and could cause tendency for overfitting [5, 8].
V1.3 enhanced the training procedure by enabling a multiple run average to counter the fluctuations, fixed random seed to maintain control over the random factors, early stopping to negate the loss and in accordance increment the epochs to increase the accuracy while maintaining the ability to stop when overfitting could be considered an issue. The resized and greyscale variants achieved accuracies above 97%, maintaining high precision, recall, f1-scores across all other results, indicating stable performance.
The variant-based results revealed that environmental and visual factors significantly influenced classification performance, where cropping reduced the background noise and improved accuracy, while the removal of the colour information (greyscale) had a limited effect. However, the combined variant cropped greyscale displayed a significant drop in accuracy across both versions, implying that excessive removal and preprocessing of the dataset can negatively impact the identification by removing

distinguishing features and reliable image information, reducing the model's discriminative capability [9].
Despite the achieved accuracy values indicating improved robustness and the overall performance; the absence of an independent external dataset causes these values and results to be considered biased and restrict the conclusion [3, 4]. Consequently, although high performance was achieved under controlled conditions, the results do not necessarily reflect realistic operational conditions and should be interpreted accordingly.

5.3 Dataset and preprocessing

The dataset used in this project was manually collected using a Logitech webcam. Each image was captured under controlled indoor conditions to reduce environmental noise and background interference. Each collected image was labelled and stored with corresponding metadata in a local SQLite database to ensure traceability and data integrity as required in ISO 8800 Clause 11. To adhere to the standards general data protection and ethical standards all participants consented, and data management procedure were maintained.
However, while this approach ensured compliance and consistency, it also imposed constraints on the dataset's variety. Limiting the number of individuals, lighting and environmental conditions, these constraints resulted in restricting the dataset generalization to real-world conditions.
To improve dataset alterations and evaluate dataset impacts, preprocessing techniques such as augmentation and transformation were applied. Creating variations resizing, greyscale conversions and cropping. The use of ImageDataGenerator facilitated dataset balancing, improve robustness and reduce overfitting [10, 11]. In accordance with the results, a moderated amount of preprocessing demonstrated by variants resized and cropping enhanced the performance by reducing background noise and compelling the model to focus only on the relevant features. Even so, excessive preprocessing such as cropped greyscale resulted in decline in accuracy. This highlights the importance of key features, noise reduction and environmental aspects to a certain degree.
The dataset splitting approach introduced potential bias due to uneven class distribution and individual representation [4, 12], causing minority classes such as "Unknown" to create an underfitting towards those categories. This was reduced in version 1.3 by averaging the results over multiple runs and using consistent randomization. However, further improvements and applications could be enhanced by utilizing features such as stratified splitting or class-weight training which could possibly improve future robustness, increased accuracy, cause fairer distribution and reduce overfitting.
In addition, v1.4 under consideration included the planned integration of AI-generated images and stratified splitting to diversify the dataset and compensate the class imbalance. This would allow the synthetic generated images to introduce extended demographic and environmental variations, potentially creating a more flexible and robust dataset. Although, considering that this approach would offer a means to overcome the dataset limitations it introduces new challenges concerning the trace-

ability, compliance and provenance with the functional safety and ethical requirements yet to be explored [6].

5.4 System reliability and functional safety context

The developed system was designed with focus on reliability, traceability and functional safety following the principles outlined in the standards ISO 8800, ISO 12100, and ISO 13849-1. This section discusses how the achieved results and observed system behaviours aligns with these standards and where further improvements would be necessary for compliance in real industrial context.
Across repeated training runs with fixed initialization, the model exhibited low variance in performance metrics, indicating high repeatability under controlled conditions, which supports the evaluation process rather than statistical generalization. The dataset was separated into training, validation, and test subsets. Using strict data management, each image appeared only in one of these subsets, ensuring full separability between the different subsets. Fixed random seeds were used to ensure repeatable splits and training behaviour, which is essential for reproducibility and evidence-based safety assessment in accordance with assurance-based safety principles [3].
The dataset was collected in a controlled environment with limited variability in subjects and conditions. As a result, while the model performs consistently within the fixed domain, the ability to maintain performance under different conditions or unseen individuals remains uncertain [7]. Limited image augmentation was applied to reduce sensitivity to minor input variations, supporting robustness under controlled deviations. However, these augmentations do not represents remain limited full operating variability.
Data management and traceability in this project were addressed by logging every collected image in structured database with metadata, file path, label, timestamp, dataset version, and SHA-256 checksum. This enables every training sample to be uniquely identified and verified, ensuring reproducibility and protection against data corruption, aligning with traceability principles [3].
During runtime operation, the system enforces a strict confidence threshold of 0.95. Classification outputs from the CNN are interpreted by an independent control logic, the model enforces decision rule:

- If confidence ≥ 0.95 and individual is authorized, state is unlocked
- If confidence < 0.95 resulting in class unknown or the individual is unauthorized, state is locked

This design ensured a fail-safe default behaviour in ambiguous situations and reflect functional safety principles [2].
The architectural separation between the AI model and the safety-related control logic reduces the likelihood that a malfunction directly results to a hazardous system action. The logging mechanism during runtime recognition events further supports traceability and subsequently the analysis of the system behaviour. However, the system implementation lacks redundancy, fault diagnostics or independent monitor-

ing. Consistently, the system does not operate with fully comprehensive functional safety.
Addressing these limitations is necessary to adhere to the functional and safety-critical industrial applications.

5.5 Limitations

Despite the overall improvements achieved throughout the entire development process especially between versions 1.2 to 1.3, several limitations were identified. These limitations constrained the evolution of the system and in turn restricted the results. They were primarily concerns with the dataset composition and technical framework.
The primary limitation is the constrained dataset size and diversity. The relative minor dataset with 1800 images representing three individuals with similar demographic characteristics, lighting and background conditions, causing potential overfitting. The accuracy reported by the system may be inflated rather than represent real-world conditions a well-known effect in ML systems trained on limited datasets [12].
Furthermore, the absence of external validation or test dataset constraints the evaluation. All variant datasets were derived from the v1.1 dataset internally split; each training, validation and test sets have the same controlled environment. This concludes that each image might reflect similarities between them causing the model to memorize patterns or similarities rather than learn them. This condition introduces potential bias towards classes causing the accuracy metrics to inflate, a risk that has been problematic in safety-critical AI systems [7]. Implementing a separated, independent dataset for testing would provide more reliable measurement, including unseen data.
The hardware and resources constraints also influenced the outcomes. Due to the limited experience, resources at hand, hardware and time, the structural architecture and training parameters were simplified in comparison towards industrial or automotive-grade AI-systems. With lower computational power, certain training parameters could not be invoked, these constraints limited the exploration of advanced alternatives, extensive hyperparameters and larger convolutional networks which could significantly enhanced performance. Hardware and resources also constrained training time, dataset versatility, extensive preprocessing variations and higher resolution.
Collectively, these limitations highlight that this projects representation an early-stage prototype rather than a complete system, focusing primarily on identifying the role of AI integration in machine learning with respect to new standardization, rather than a safety-certified product. Addressing the dataset expansion, environmental variability, model complexity and compliance-oriented evaluation is essential to advance towards discovering the standardization frameworks of a certifiable machine learning AI-based safety system.
Lastly, given the limited dataset and simplified architecture, the presented system should be regarded as a conceptual demonstration rather than a certifiable safety product.

5.6 Integration of AI in accordance with safety standards

The integration of AI within safety related control systems introduces new challenges compared to conventional deterministic systems. Traditional standards such as ISO 12100 and ISO 13849-1 define systematic approaches for risk reduction, functional safety functions and performance levels. These frameworks assume that the system behaviour is predictable, testable and reproducible under defined conditions [1, 2]. In contrast AI systems derive their functionality from data and statistical inference, making their behaviour probabilistic and dependent on the quality and representativeness on training data [7].
ISO 8800 on the other hand provides guidance to accommodate AI into safety related systems. Instead of focusing solely on component reliability, it introduces the concept of an AI safety lifecycle [3]. This lifecycle formalises how safety can be achieved through concept definition, data management, AI development, verification and validation (V&V) and operation.
Although the present system is conceptual and experimental in a way that it investigates how to incorporate principles of safety standards, the development process can be mapped to many of the principles of ISO 8800 as discussed in chapter 3 [3, 6]. Applying this framework provided a systematic and structured approach of developing the system. Helping clarify each development phase such as risk analysis, data collection, model training and testing. For instance, the dataset lifecycle brought the important property of "completeness". This led to the implementation of detailed dataset metadata to support traceability and transparency throughout the development process, both for the developer and the end user, while acknowledging that full semantic completeness of the dataset was beyond the scope of this project.
Another key property of ISO 8800 was the use of an assurance argument to structure safety reasoning [3, 13]. By formulating a GSN argument, the safety claims of the system could be explicitly linked to context, strategies and evidence. This helped visualising the safety of the system from a greater perspective and creating a path with goals, context, strategies and eventually evidence to support each safety case.
ISO 8800 served as a valuable framework for an organized and transparent development process. The principles encouraged structured documentation, safety reasoning and clear traceability between risks, requirements and evidence. By adapting these principles, the project gained a clearer understanding of critical properties of AI driven systems and how these could be implemented and evaluated.

5.7 Future work

For advancing the development of the AI-based person detection system building upon the limitations and requirements identified throughout this project while maintaining the established safety critical aspects. The identified key areas comprise of the evolution of the dataset expansion and AI integration, architectural and algorithms improvements and the usage of the V&V framework.

5.7.1 Dataset

Primary direction for the continued development concerns the improvement and expansion of the dataset. The current limited dataset constraints the system's ability to generalize the diversity and mange new unseen data [4, 12]. To improve the model's diversity, enhancing the robustness and accuracy the dataset should increase in diversity, general size, alternative demographical characteristics and diversified environmental settings to visualize real-world scenarios.
One discussed approach towards this would be the implementation of AI-generated images without the extensive manual collection This strategy of incorporating synthetic data could give alterations to real-world conditions. For example, using generative AI techniques for synthetic image creation could expand the diversity and environmental settings of the dataset [10]. However, the risk of using the synthetic generated human images is the ethical, validation and traceability concerns [6]. It requires meticulous management in accordance with the ISO standards 12100 and 8800. The focus for future research should therefore be to evaluate the synthetical data implications of safety critical AI systems.

5.7.2 Architectural development

The current lightweight CNN framework is not considered sufficient for the usage in a modern AI-based safety system. The SoftMax-based classification architecture presents limitations in safety-critical evaluation, as it forces a closed-set decision and may assign high confidence to unknown individuals [7]. As a potential alternative, a cosine-based embeddings representation could be explored, enabling the usage of high-dimensional feature vectors and distance-based decision thresholds rather than direct class probability approach [14]. This approach could increase the systems robustness, enable scalability and the management of more complex environmental settings across larger dataset. Nevertheless, this creates the requirements of the dataset expansion, as the heavier CNN would require much larger and wider dataset to complement.
Mentioned in the discussion, further improvements would be the implementation of training methods, such as stratified data splitting and weight classification.

5.7.3 Verification and validation

The development and implementation of a formal Verification and validation (V&V) framework would be a critical future direction. The V&V method grounded in the ISO standards 8800, 12100 and 13849-1 is the baseline for the development of AI-base systems. ISO 8800 introduces and provides a framework for AI-based systems in safety related applications by defining the lifecycle requirements.
The V&V model is a functional aspect into the development of an AI-based system and future work should establish a tailored process based on the framework in 8800. Integrating failure analysis, performance metrics and uncertainty estimation to enable a systematic assessment of the model [4, 7]. Focusing on the safety aspects, model reliability, robustness and the continuous learning within the lifecycle. This

is a necessity to estimate the evolving standardization of AI-based safety systems in machine learning.

5.8 Reflection

The development of AI-based detection systems provided valuable insight into how artificial intelligence can be combined with functional safety. Throughout the project, it became evident that there is a difference between deterministic control logic, on which the existing standards are based and probabilistic nature of AI-systems which rely on statistical patterns learned from data. This explains why traditional safety standards such as ISO 12100 and 13849-1 can be difficult to apply directly to AI-based applications.
Through the development and testing of the system, it became apparent that maintaining transparency and traceability throughout the dataset and model lifecycle is as critical as the algorithm performance itself. The project also highlighted the importance of structured documentation collected to support the future assurance argument as described in ISO 8800.
While the achieved results demonstrate that AI can achieve a high performance under controlled conditions, the study also revealed the current limitations of applying such a system within certified safety applications. The project should therefore be viewed as an early-stage conceptual study that explores the interaction between machine learning and functional safety standards, rather than a verified industrial product. Future research should continue to bridge the gap between the safety systems and adaptive AI systems through larger datasets, formal V&V frameworks and further adaptation of standards for accommodation of the evolution of AI.

6
Conclusion

This thesis explored how the development and creation of an AI-based person detection system can be designed and evaluated in relation to the functional safety standards such as ISO 12100, 8800, 13849-1. The project combined the technical implementation of creating a CNN with the principles required for a traceable, safe and reproducible development process.
According to the results, the system achieved high accuracy and consistent performance under controlled testing conditions. Improvements implemented in v1.3 included early stopping and performance averaging across multiple runs to reduce randomness which led to reproducible results. This shows that even though AI-based models are probabilistic, it is possible to make them behave predictably and that they could be evaluated within a safety framework.
However, several limitations were identified. The dataset used was limited in both size and diversity. The system also lacked independent validation, redundancy and diagnostic monitoring. These factors restrict the achievable safety integrity level and that additional safety measures are required before such systems can be deployed in certified industrial applications. Nevertheless, this project presents a conceptual and possible approach for how AI-based systems can be integrated into safety-related applications with the use of standard frameworks.
In conclusion, this work highlights both the potential and the remaining challenges of applying AI-based systems in functional safety domains. It also shows that ISO 8800 can serve as a valuable guidance and foundation for a structured AI development. Future work should emphasize on a broader dataset coverage to increase size and variety, enhancing verification and validation methods and finally refining existing standards to better support the evolving safety requirements of AI driven safety systems.

Bibliography

[1] ISO 13849-1:2023, *Safety of machinery – Safety-related parts of control systems – Part 1: General principles for design*, International Organization for Standardization, 2023.

[2] ISO 12100:2010, *Safety of machinery – General principles for design – Risk assessment and risk reduction*, International Organization for Standardization, 2010.

[3] ISO 8800:2025, *Road vehicles – Safety and artificial intelligence*, International Organization for Standardization, 2025.

[4] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.

[5] J. Heaton, "Ian Goodfellow, Y. Bengio, and A. Courville: Deep Learning," *Genetic Programming and Evolvable Machines*, vol. 19, no. 1–2, pp. 305–307, 2018, doi:10.1007/s10710-017-9314-z.

[6] European Commission, *Ethics Guidelines for Trustworthy Artificial Intelligence*, High-Level Expert Group on Artificial Intelligence, 2019.

[7] D. Amodei et al., "Concrete Problems in AI Safety," arXiv preprint arXiv:1606.06565, 2016.

[8] F. Chollet et al., *Keras Documentation – EarlyStopping Callback*. Available: `https://keras.io/api/callbacks/early_stopping/`(Accessed: Jan. 2026)

[9] R. Szeliski, *Computer Vision: Algorithms and Applications*, Springer, 2011.

[10] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 60, 2019.

[11] Keras, *ImageDataGenerator API*. Available: `https://keras.io/api/data_loading/image/`(Accessed: Jan. 2026)

[12] A. Geron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, O'Reilly Media, 2019.

[13] R. Hawkins, T. Kelly, J. Knight, and P. Graydon, "Assurance cases and the safety of software-intensive systems," *IEEE Computer*, vol. 44, no. 11, pp. 21–29, 2011.

[14] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A Unified Embedding for Face Recognition and Clustering," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 815–823, doi:10.1109/CVPR.2015.7298682.

A

Functional safety documentation

A.1 Risk analysis assessment

Table A.1: Initial risk assessment according to ISO 12100, S=Severity, F=Frequency, P=Probability, A=Availability to dodge

Phase	Ref.	Hazard (ISO 12100 TABELL B.1)	Hazardous action	Injury	S	F	P	A	Risk index	Risk level	PLr
Operational machinery	1.1	Operational machinery	Person gains access to machinery	Collision between individual and machine	2	2	2	2	5	High	PLe
Operational machinery	1.2	Operational machinery	Non-instructed individual gains access to moving machinery	Collision between individual and machine	2	1	1	2	2	Low	PLd
Operational machinery	1.3	Operational machinery	Instructed individual gains access to moving machinery	Collision between individual and machine	2	1	2	2	3	Medium	PLd
Operational system	1.4	Operational system	Machine restarts while individuals are in restricted area	Collision between individual and machine	2	1	2	2	3	High	PLd

A.2 Risk reduction and residual risk

Table A.2: Risk reduction measures and residual risk evaluation

Ref.	Risk reduction measures	S	F	P	A	Risk index	Risk level	Achieved PL	Tolerable risk	Comments
1.1	Machinery shall not be accessible by ordinary persons during operation. Instructed personnel may operate machinery in low-energy mode.	2	2	2	2	5	High	PLe	No	See 1.2 & 1.3 for solutions
1.2	Restricted access area using fence and electronically controlled gate with AI-based identification system. Unauthorized individuals are denied access.	2	1	1	1	2	Low	PLc	Yes	The restriction of access using fence, gate and AI access system ensures that individuals without training and approval to work with the machine are blocked from access
1.3	AI-based access system places machinery in safe or low-energy mode when instructed individuals are present.	1	1	1	1	1	Low	PLa	Yes	Even though the instructed trained induvuduals are qualified and approved for access to the machinery. When operating att max capacity this is still deemed to dangerous. Thus the machine needs to be put in a Safe low energy mode by the AI system. Still only instructed induviduals are allowed to work with the machinery in this state. This will put a reliability quality reqiurment on the AI system to correctly identify the induviduals for access as otherwise ordinary persons could end upp geting injured
1.4	Restart only possible when gate is closed, restart button pressed, and authorization is verified.	2	1	1	1	2	Low	PLc	Yes	Reduces the risk of restarting whilst individuals are inside restricted area, does not account for human error

A.3 Functional safety concept

Table A.3: Functional Safety Concept derived from risk analysis (SFC)

Function / Component	Description	Associated Hazard	Safety Objective	Risk Reduction Measure	Standard Reference
Rotating machine, full energy mode (FE)	Operates when entering restricted area	Entanglement, impact	Prevent contact with machine during operation	Restrict all access to area and transition to low energy mode	ISO 12100, 5.4
Rotating machine, low energy mode (LE)	Operates inside restricted area	Entanglement, impact	Prevent contact with machine during operation	Restrict access to authorized and instructed individuals approved for work in LE mode	ISO 12100, 5.4
AI system for person detection	Identifies individual presence at entry point	Unintended individual entry	Trigger risk reduction when presence is detected	AI-enabled camera triggers control signal to reduce machine speed	ISO 12100, 6.3.2.2
Reset operations to full energy mode	Ensures no individuals remain in the machine zone	Premature restart of machine	Only resume full speed when area is confirmed clear	Control signal holds machine in LE mode until reset is confirmed, gate is closed, and authorization is verified	ISO 12100, 6.3.3

A.4 Safety functional list

Table A.4: Safety Functions List (SFL)

ID	Safety Function	Hazard	Trigger Condition	Requirement	PL	Comments
SF-01	Detection of authorized access	Unauthorized access	Authorized individual detected	Machine enters LE mode	PLc	Only authorized individuals allowed
SF-02	Gate opening control	Unauthorized access	Detection of authorized individual	Gate unlocks	PLc	Gate remains locked otherwise
SF-03	Low energy mode	Risk of injury from moving parts	Authorized individual detected + gate unlocks	Machine operates in LE mode	PLa	Automatic speed reduction
SF-04	Reset control	Restart with individual inside area	Authorized individual detected + gate closed + reset pressed	Machine resumes normal operation	PLc	All reset conditions must be fulfilled

A.5 Safety requirement specification

Table A.5: Safety Requirements Specification (SRS)

Sub-function	Safety Goal	Functional Requirement	Input	Logic	Output	PL	Test Method
Authentication (SF-01)	Prevent unauthorized access	Gate remains locked; machine continues normal operation	Camera interface (AI component)	Safety CPU/GPU detection logic		PLc	Test with authorized / unauthorized individuals
Unlock gate (SF-02)	Allow access only when authorized	Gate unlocks only if detection is reliable	Camera interface	Safety CPU/GPU	Relay	PLc	Detection verification
Machine LEM (SF-03)	Prevent injury during maintenance	Machine switches to low energy mode	Authorization + gate state	Safety CPU/GPU	Relay	PLa	Entry simulation
Gate closed (SF-04)	Prevent premature restart	Gate must be closed	Safety switch signal	Safety CPU/GPU		PLa	Signal verification
Reset button (SF-04)	Prevent premature restart	Reset button pressed	Safety switch signal	Safety CPU/GPU		PLa	Signal verification
Machine reset (SF-04)	Controlled restart	Gate closed + reset pressed + authentication approved	Combined safety signals	Safety CPU/GPU	Relay	PLc	Full functional test

A.6 AI safety requirement

Table A.6: AI Safety Requirements

Function	Risk	Trigger conditions	Safety requirements	Component requirements	Acceptance criteria	KPI / metrics
Environmental	Misidentification of authorized or unauthorized individual	Light pollution, shadowing	System shall operate only under suitable indoor lighting conditions	High-resolution indoor-rated camera with stable lighting for image processing	Identification at 600–900 lux	
Accuracy (False Acceptance Rate)	Unauthorized access accepted	Unauthorized individual attempts entry	False acceptance rate shall not exceed defined limit	Sufficient and representative training dataset	Unauthorized individual denied entry in tests	FAR $\leq 1\%$
Accuracy (False Rejection Rate)	Authorized access denied	Authorized individual attempts entry	False rejection rate shall not exceed defined limit	Sufficient training on authorized dataset	Authorized individual granted access in tests	FRR $\leq 5\%$
Detection and authentication time	Operational inefficiency due to delay	Deficiency in code, model training, or sensor input	Detection and authentication shall complete within required timeframe	Efficient code execution and clear sensor input	Detection and authentication time ≤ 15 s	
Authorization signal generation	Delay in gate opening	Hardware or execution inefficiency	Authorization signal shall be generated within defined time	Efficient hardware and signal handling logic	signal time ≤ 250 ms	
Gate unlock control	Unauthorized gate unlock	False positive authorization	Gate shall unlock only after verified authorization	Relay controlled by locked logic signal		
Fail-safe behavior	System failure permits access	System or component failure	Gate shall remain locked during system failure	Relay defaults to locked state	100% locked state on system failure	

A.7 Dataset requirements

Table A.7: Dataset safety requirements for AI-based access control system

Safety properties	Definition	Safety related insufficiencies	Metrics	Countermeasures
Accuracy	The data corresponds to their source with respect to semantic representation and interpretation based on previously mentioned subjects	Resolution of the camera images is not sufficient for subject detection	Accuracy ≥95%	Selection of hardware capable of functioning according to input space. Inspection of manually labelled data
Completeness	The data elements (including metadata) are populated and the data have defined coverage of the input space, safety-relevant cases and plausible data perturbations	Perturbations like noise, brightening, darkening, vibration, obstacle and blurring interference are not reflected in the dataset	Input space coverage ≥30%	Investigation of general use cases. Verification that the data covers the input space and that the collection of data matches the environmental reality
Correctness (or fidelity)	The data correspond to the phenomenon they intend to capture and include features and metadata which help to characterize the phenomenon	No distinction has been made between facial accessories in an image label, though this is relevant for the task that the AI system performs	Correct label assignment ≥99%	Characterization of the essential features of the target
Independence of datasets	The datasets sufficiently avoid leakage of information amongst themselves with respect to data sources and the methods used to capture, gather, generate and process the data	All datasets used to develop an AI model come from the same exact environment (e.g. same camera position, lighting, shadowing and noise). All datasets are collected relying upon the same means. Dataset creation is based on the same technique and conducted by the same person.	No overlap of image frames between train, validation and test datasets	Use of different sources of data. Use of different technical means for data capturing, e.g. different sensors, vendors and brands. Deployment of different processes or methods for dataset creation
Integrity	The data are not altered by natural phenomenon or intentional action	Errors introduced by corrupted hardware or failure in database memory. Untrained or overtrained data introduces inconsistent data elements	Dataset corruption rate ≤1%	Analysis of robustness to adversarial attacks on the dataset. Integrity checks in databases
Representativeness	The distribution of data corresponds to the information in the environment of the phenomenon to be captured and is free of biases	Synthetic data do not capture certain real-world features such as shadows. Uneven gender distribution in the monitored environment	Balanced class distribution ≥20%	Analysis and comparison of theoretical and experimental distributions
Temporality	The data gives sufficient consideration to time-based characteristics	Monitoring system does not consider changes in styles or necessary accessories such as facemasks	Temporal coverage ≥2 months	Metadata containing time of creation and validity
Traceability	The derivation of the data from their origin is demonstrated	Images lacking source information are integrated into the dataset	100% metadata coverage	Use of data management system. Creation and inclusion of sufficient metadata
Verifiability	The data include sufficient features to be amenable for verification	Unsafe outcomes cannot be reproduced. No checksum, CRC or hash mechanisms are used	Reproducibility ≥99%	Set and log random seeds for all randomized processes. Manual analysis. Integrity verification mechanisms

A.8 Software

Table A.8: Software implementation and usage

Software/Library	Usage
Keras	was employed as a high-level Application Programming Interface (API) for TensorFlow. It provided with an integrated data augmentation utility (ImageDataGenerator) enabling efficient image pre-processing improving model
OpenCV	was utilised for its real-time image preprocessing functions and for its robust webcam interfacing capabilities.
SQLite	was selected as the solution regarding the metadata and result storage. This was based on the lightweight architecture and lack of dependency of external database servers.
JSON	was adopted for storing system configurations and status data of individuals in a readable format, facilitating external editing.
SoftMax	classification was implemented as the model output layer in preference to embeddings-based (e.g. FaceNet with cosine similarity). Embedding generally generates superior generalisation and robustness with larger-scale datasets. SoftMax requires less architectural complexity, is suitable for relative small-scale datasets available in this project and is more computationally efficient.

A.9 Confusion matrix

Table A.9: Confusion matrix over version 1.2 and 1.3 with each variant in accuracy order

Variant	Version	Accuracy [%]	Loss	Precision	Recall	F1-Score
Greyscale	1.3	97.60	0.3295	0.981	0.973	0.976
Resized	1.3	97.49	0.3432	0.977	0.981	0.978
Cropped	1.3	96.72	0.2605	0.972	0.969	0.969
Cropped	1.2	95.88	0.4122	0.958	0.949	0.953
Resized	1.2	92.31	0.5421	0.912	0.902	0.907
Cropped Greyscale	1.2	89.02	0.6355	0.892	0.887	0.889
Cropped Greyscale	1.3	86.45	4.9945	0.879	0.859	0.849
Greyscale	1.2	76.02	0.7440	0.752	0.738	0.745

A.10 Assurance argument

Table A.10: Assurance argument for AI-based access control system

ID	Type	Statement / Claim	Evidence
G1	Goal	The AI based person detecting system is acceptable safe for intended usage within ODD (Operational Design Domain).	Risk analysis, Functional Safety Concept, Safety Function List & SRS. Report: Chapter 5.4 & 5.6
C1	Context	Intended use: access control "light gate" supervising entry to restricted area.	System description & Functional Safety Concept. Report: Chapter 1–3
C2	Context	Defined operation domain: indoor, single fixed camera, subject distance/pose and lightning comparable to training conditions.	Dataset & capture setup. Report: Chapter 3.3
A1	Assumption	Camera and computer remain operational (no HW faults/errors).	Out of scope, acknowledged in report. Report: Chapter 5.4 & 5.6
A2	Assumption	Authorized roster (AUTH_JSON) is accurate and maintained.	Code: AUTH_JSON use. Report: Chapter 3
A3	Assumption	No external manipulation is attempted during operational use.	Out of scope. Report: Chapter 5
S1	Strategy	Fail-safe control logic, AI reliability in ODD, dataset safety & traceability.	ISO 8800:2025 Annex B
G2	Goal	Fail-safe control logic ensures safe state during uncertainty or unauthorized cases.	Functional safety. Code: Recognize_face.py. Report: Chapter 5.4
C3	Context	CNN is advisory, deterministic control logic enforces final action.	Code structure. Report: Chapter 5.4
E1	Evidence	Decision rules implemented: if conf < 0.95 or label = "Unknown" or "Unauthorized" $\rightarrow$ locked state. If conf ≥ 0.95 and label = authorized $\rightarrow$ unlocked state.	Code: Recognition_face.py, Config.py
G3	Goal	AI reliability within ODD meets stated acceptance criteria.	Report: Chapter 4. AI Safety Requirements
C4	Context	Acceptance criteria (requirements), e.g. FAR $\leq 1\%$, FRR $\leq 5\%$.	AI Safety Requirements
E2	Evidence	Validation and test accuracy results. Repeated runs show $<1\%$ variation.	Code: Train_and_Test.py, averages JSON. Report: Chapter 4
A4	Assumption	The threshold of 0.95 was set on validation data representative of the ODD.	Report: Chapter 3.3, 3.5 & 5.4
G4	Goal	Dataset safety and traceability meet ISO 8800:2025 properties.	Report: Chapter 3.3.1. Dataset Requirements

ID	Type	Statement / Claim	Evidence
C5	Context	Dataset splits 70/20/10 with no leakage between sets.	Code: Split_dataset.py
E3	Evidence	SQLite image metadata including path, label, person_name, timestamp, dataset_version, SHA-256 and augmentation.	Code: Database.py, Train.py. Dataset Requirements
A5	Assumption	Single camera, controlled environment, limited subjects and diversity are adequate for intended use.	Report: Chapter 3.3.1 & 5.4
G5	Goal	Verification and validation confirm consistent AI behaviour and traceability to performance and safety objectives.	Report: Chapter 4 & 5.4
E4	Evidence	Learning curves, confusion matrices and test results support performance stability.	Code: Learning_curve.py, classification results. Report: Chapter 5.4
C6	Context	Runtime logging records label and confidence on state change, no HW monitoring.	Report: Chapter 5.4 Logging
G6	Goal	Residual risks identified and bounded, documented with improvement plan.	Report: Chapter 5.6
E5	Evidence	Risk analysis, Functional Safety Concept, Safety Function List, SRS and discussion.	Report: Chapter 3 & 5

B

Experimental results

B.1 Training and validation loss curves

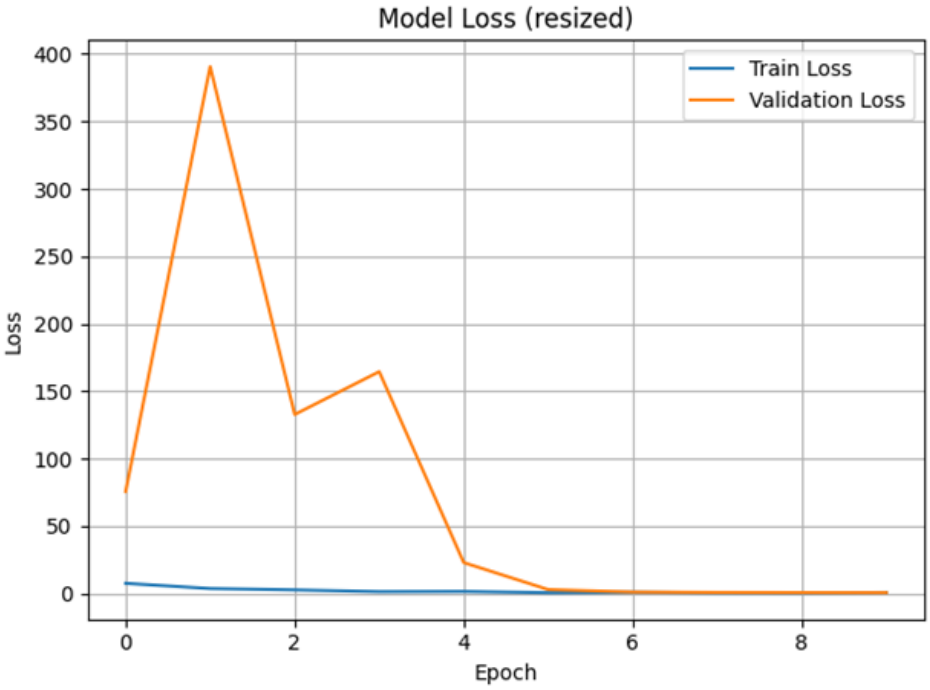

Figure B.1: Training and validation loss for resized variant model.

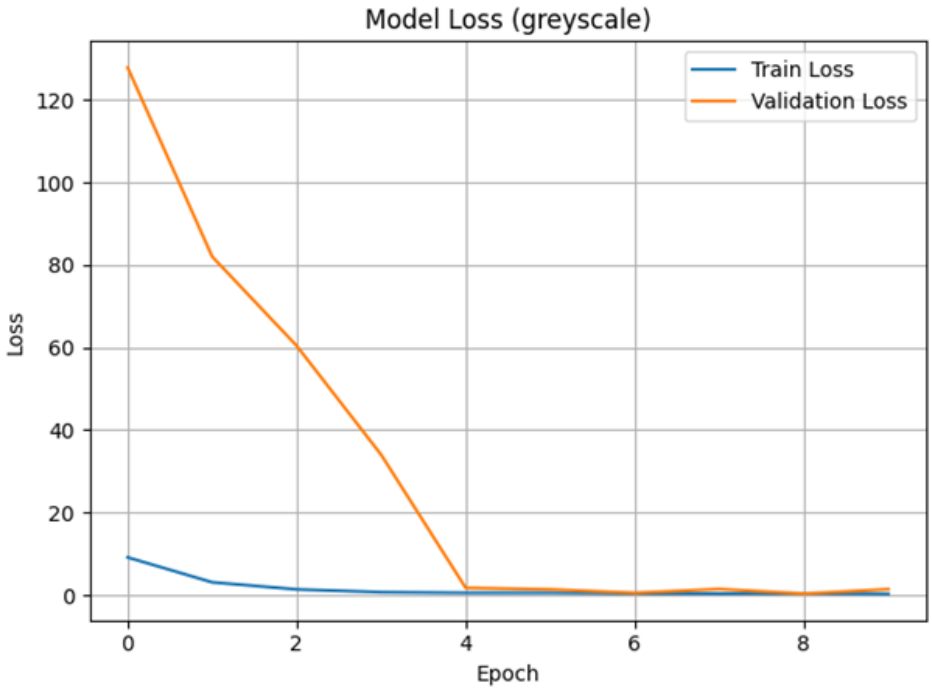

Figure B.2: Training and validation loss for greyscale variant model.

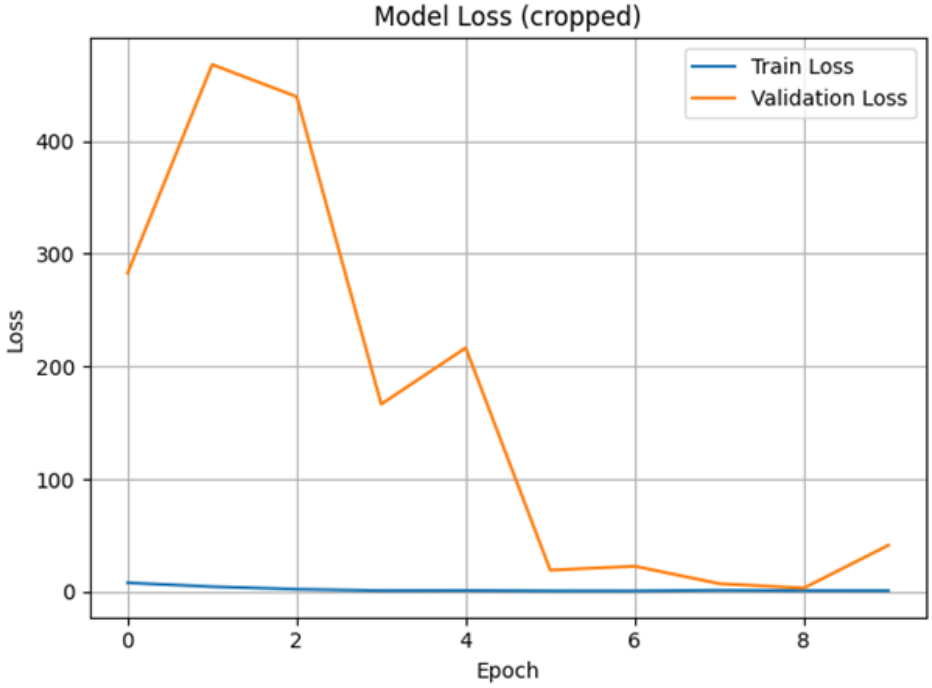

Figure B.3: Training and validation loss for cropped variant model.

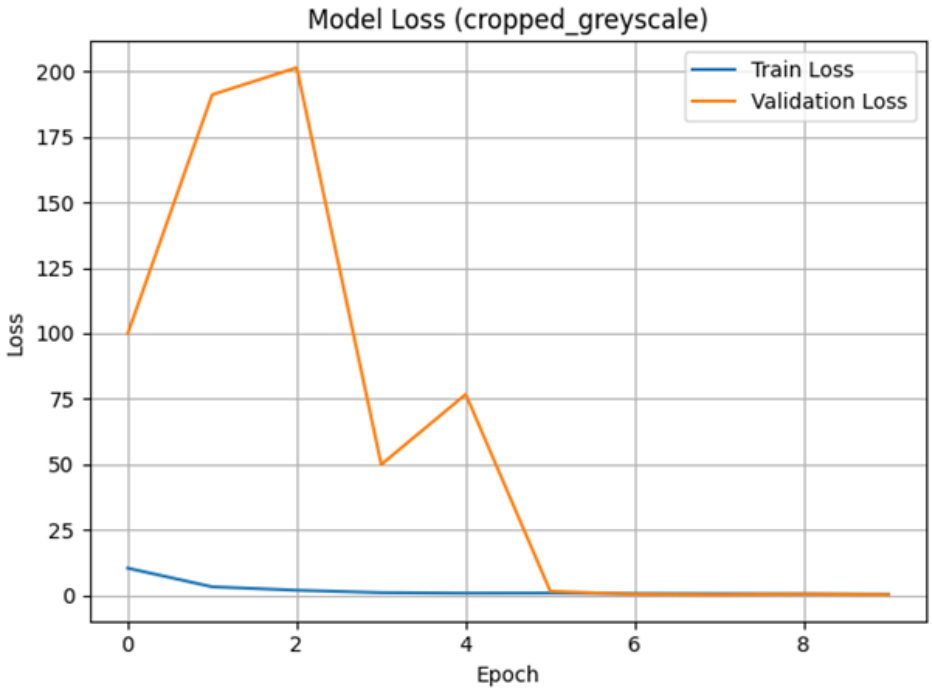

Figure B.4: Training and validation loss for cropped_greyscale variant model.

B.2 Model accuracy

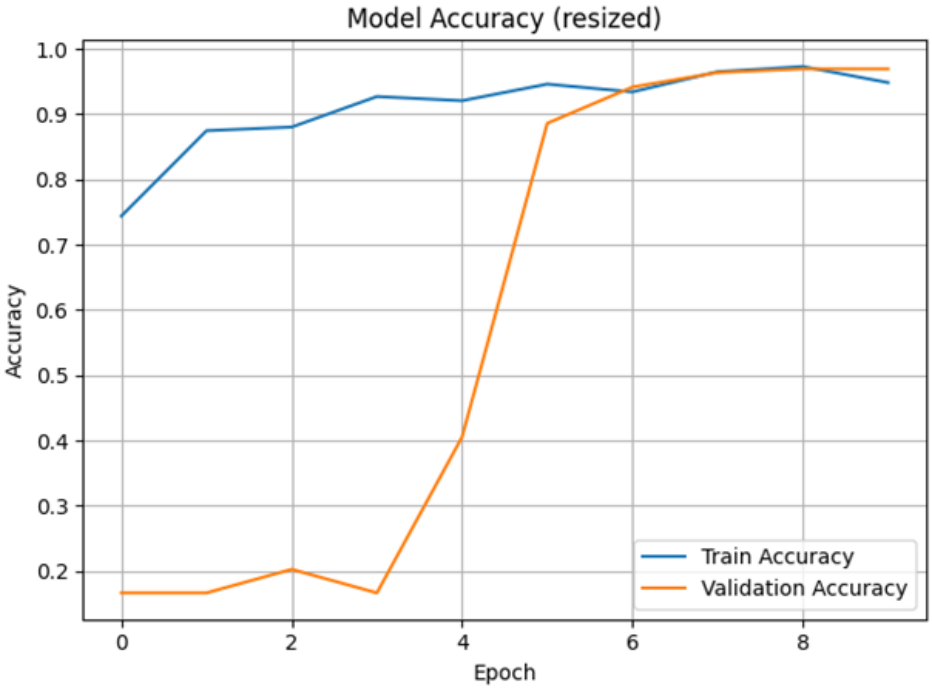

Figure B.5: Classification accuracy across training epochs for resized variant model.

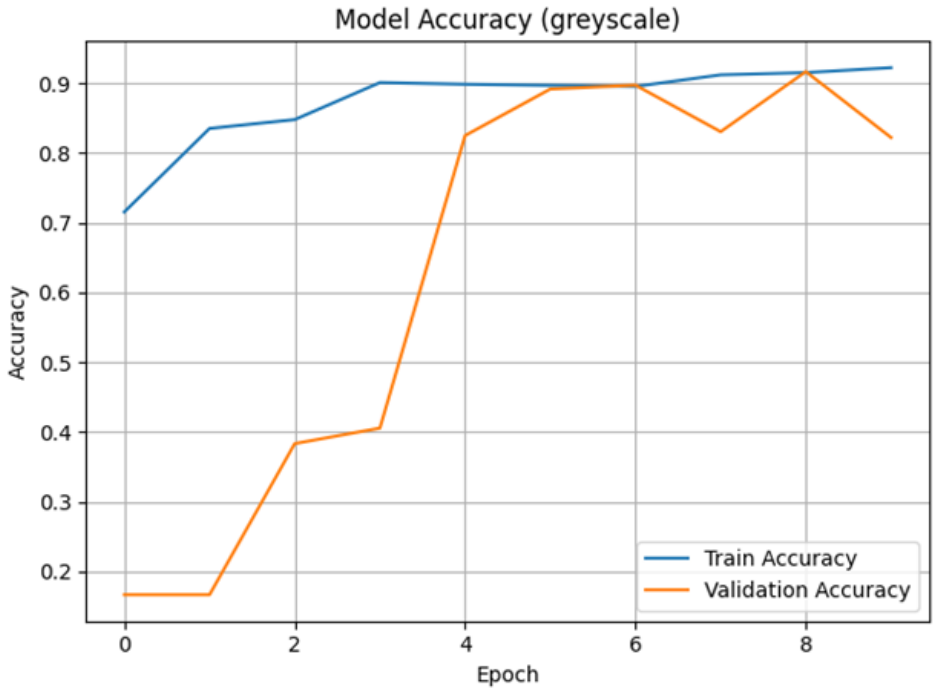

Figure B.6: Classification accuracy across training epochs for greyscale variant model.

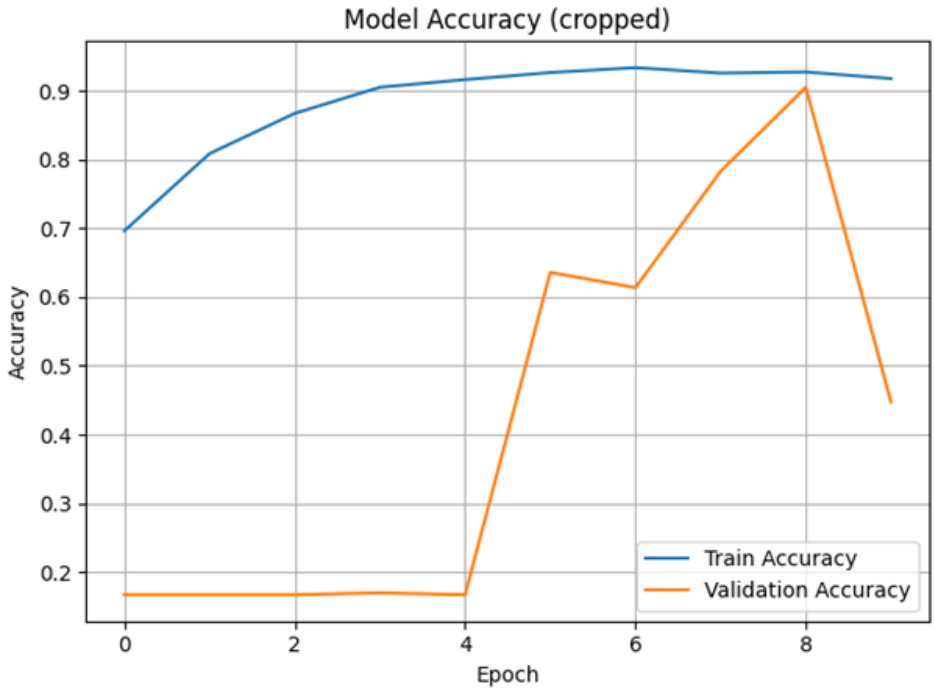

Figure B.7: Classification accuracy across training epochs for cropped variant model.

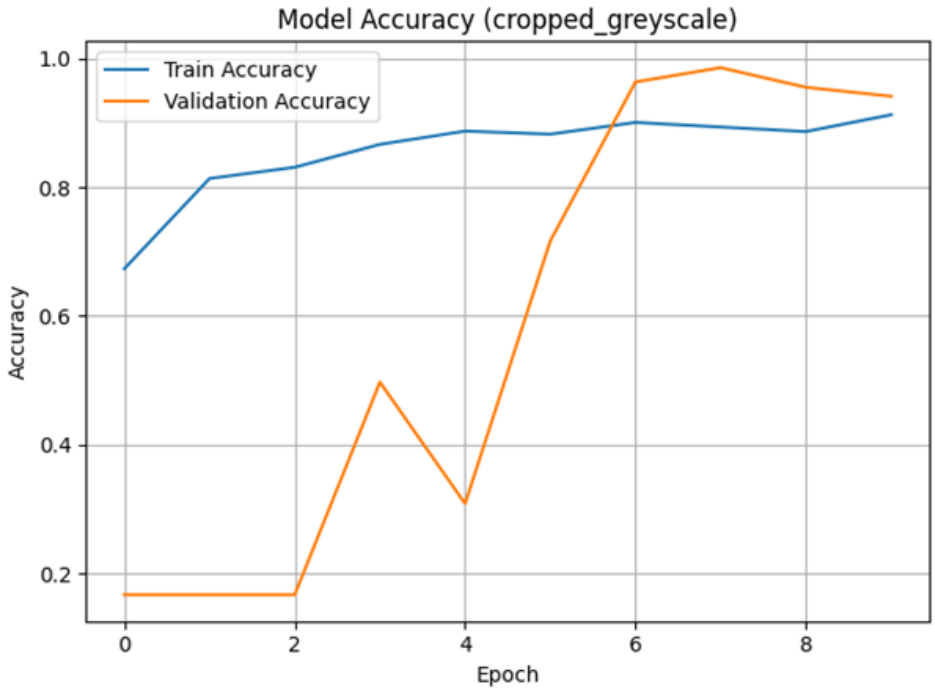

Figure B.8: Classification accuracy across training epochs for cropped_greyscale variant model.

B.3 Confusion matrix

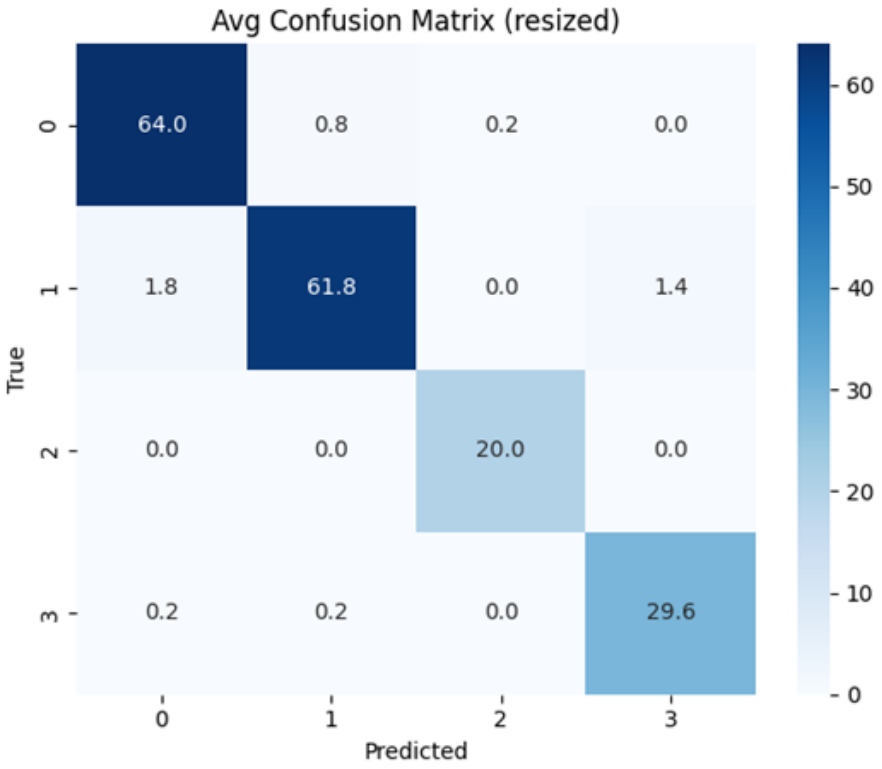

Figure B.9: Average confusion matrix for resized model

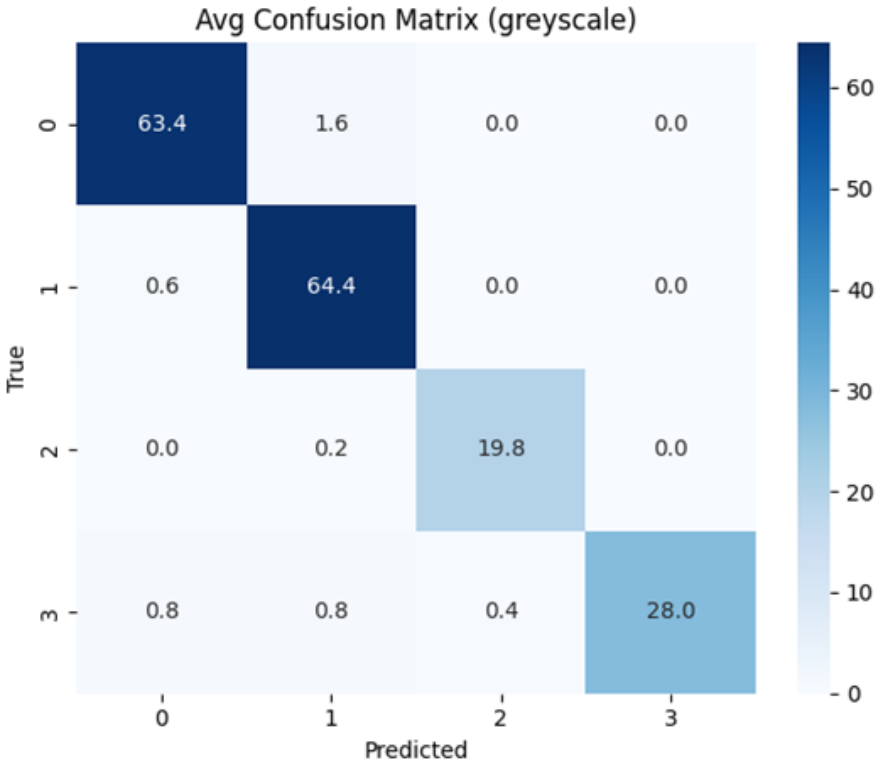

Figure B.10: Average confusion matrix for greyscale model

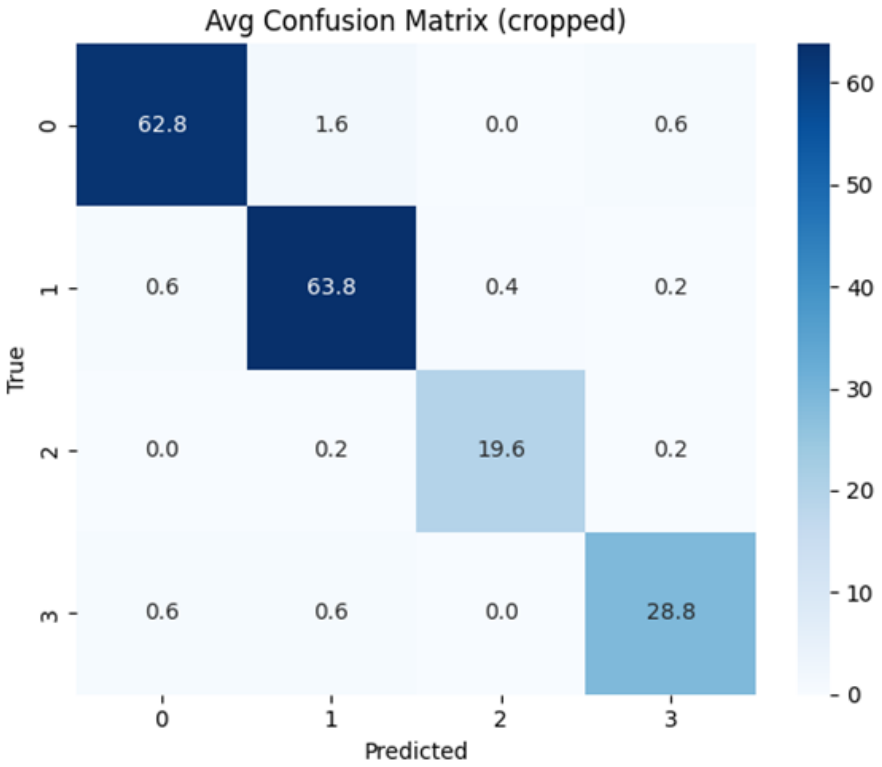

Figure B.11: Average confusion matrix for cropped model

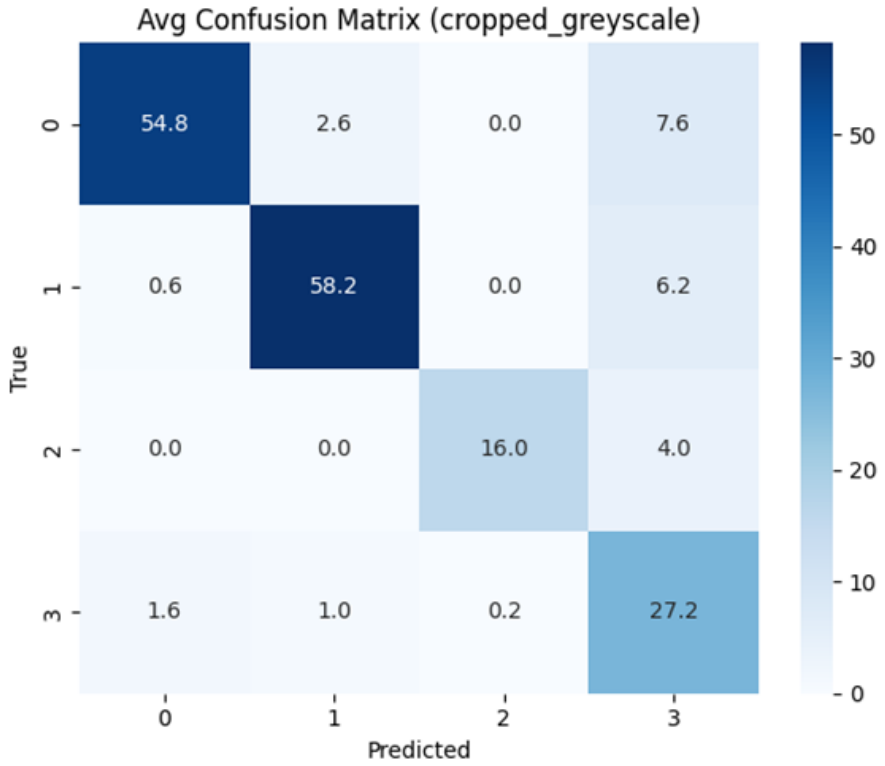

Figure B.12: Average confusion matrix for cropped_greyscale model

B.4 Per-class metrics

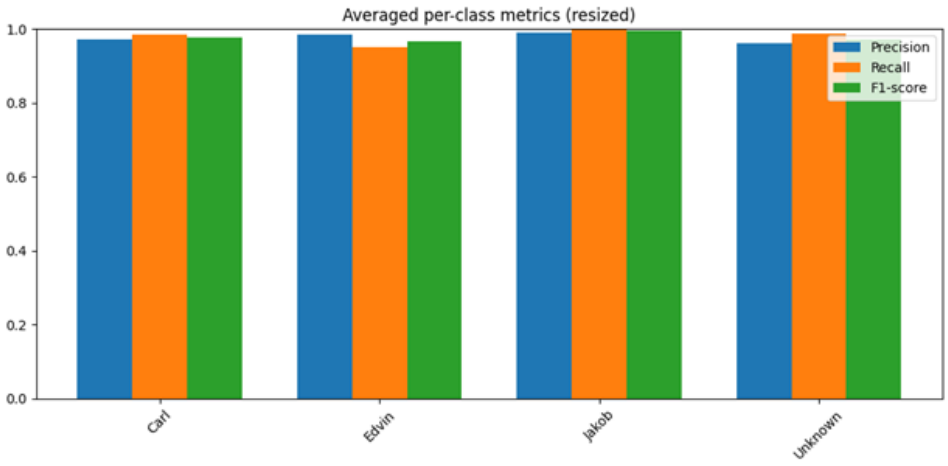

Figure B.13: Average per-class metrics diagram for resized model

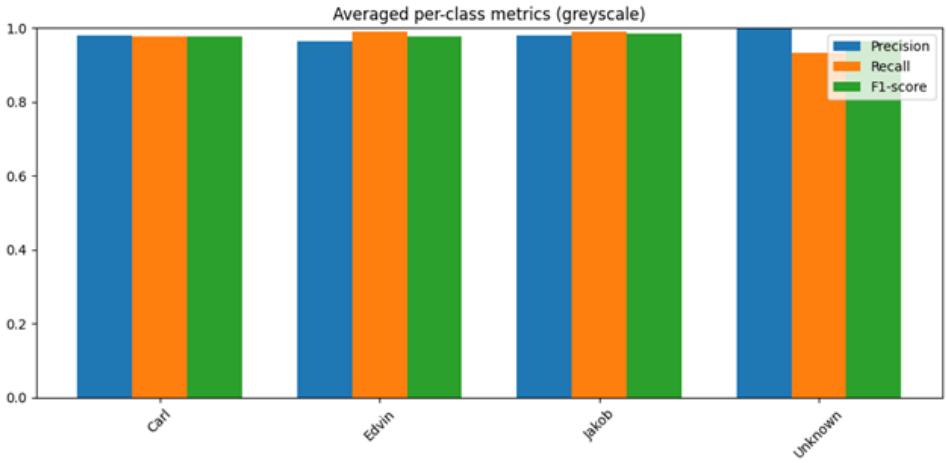

Figure B.14: Average per-class metrics diagram for greyscale model

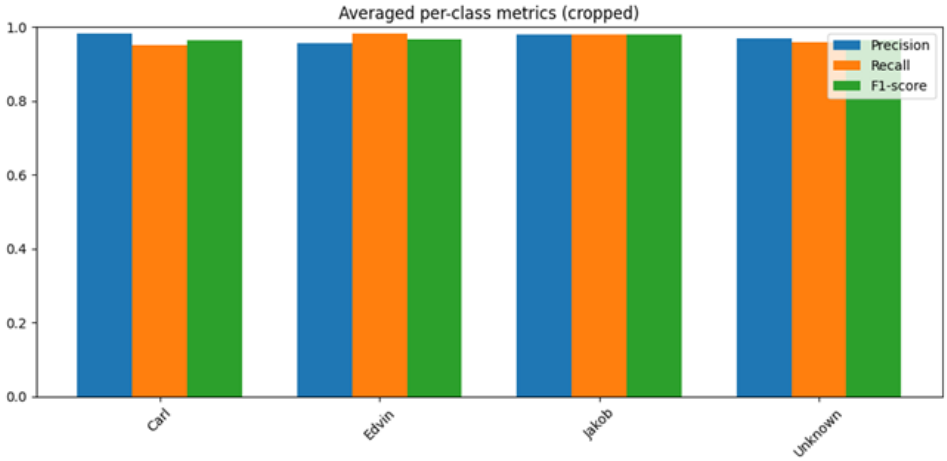

Figure B.15: Average per-class metrics diagram for cropped model

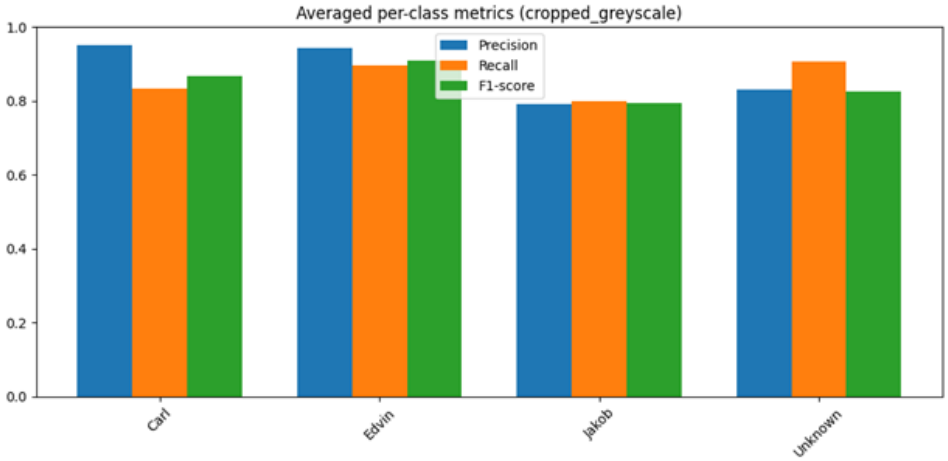

Figure B.16: Average per-class metrics diagram for cropped_greyscale model

DEPARTMENT OF ELECTRICAL ENGINEERING
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY