INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Applying ISO/ICAO Standards in an Institutional Photo using Deep Learning

João Pedro Pinheiro Bicho

Master's in Data Science

Supervisor:
PhD João Pedro Oliveira, Assistant Professor,
ISCTE-IUL

Supervisor:
PhD Maurício Breternitz, Associate Professor,
ISCTE-IUL

September, 2025

Department of Quantitative Methods for Management and Economics

Department of Information Science and Technology

Applying ISO/ICAO Standards in an Institutional Photo using Deep Learning

João Pedro Pinheiro Bicho

Master's in Data Science

Supervisor:
PhD João Pedro Oliveira, Assistant Professor,
ISCTE-IUL

Supervisor:
PhD Maurício Breternitz, Associate Professor,
ISCTE-IUL

September, 2025

Em memória do Professor Maurício

Acknowledgment

I would like to express my deepest gratitude to all those who, in various ways, contributed to the realization of this dissertation.

To my parents and my sister, for your unconditional love, constant support, and for always believing in me, even in moments when I doubted myself. Your presence, patience, and encouragement were the foundation upon which I could build this journey.

To my supervisors, for your guidance, availability, and for sharing your knowledge, which helped me grow not only as a researcher but also as a professional. Your high standards and continuous support were essential to the quality of this work.

To Matilde, for all the understanding, for the motivation during the most challenging times, and for making me believe that every effort would be worthwhile.

To my friends, for the companionship, the conversations that relieved the pressure, and the moments of relaxation that helped me maintain balance throughout this journey.

I would like to extend my gratitude to the Instituto de Telecomunicações (IT) and Iscte - Instituto Universitário de Lisboa, for their support and for providing the computational and infrastructural conditions for the development of this thesis.

To all of you, my sincere thanks. Without you, this milestone would not have the same meaning.

Resumo

A verificação de fotografias em instituições académicas é, atualmente, um processo subjetivo e inconsistente. Ao contrário do controlo de passaportes, que segue normas internacionais rigorosas (ISO/ICAO), os ambientes institucionais carecem de uma definição universal para fotografias aceitáveis, o que resulta em critérios variáveis e numa dependência da aprovação manual. Esta dissertação aborda tal ineficiência através do estabelecimento de uma classificação formal de requisitos visuais para a identificação académica e do desenvolvimento de um sistema de classificação automatizado, baseado em deep learning, para assegurar o seu cumprimento.

Neste estudo utilizaram-se dois conjuntos de dados distintos: uma base de dados privada de fotos utilizadas em ambiente institucional, e o conjunto de dados público de imagens usadas para treino de modelos de identificação de caracteristicas em imagens de passaporte. Foram realizadas experiências comparativas entre Weightless Neural Networks (WNN) e modelos YOLO (v8 e v11). Os resultados identificaram inequivocamente os modelos YOLO como superiores, superando consistentemente as WNN e modelos CNN utilizados para obter métricas base. Adicionalmente, uma análise crítica à qualidade dos dados revelou que os modelos treinados no conjunto de dados público, classificado por especialistas, alcançaram uma melhor generalização do que os treinados em dados institucionais, demonstrando que a precisão das classificações e o balanceamneto dos dados é mais valiosa para este estudo e obtem melhores resultados que o treino com o dataset de imagens institucuinais.

A contribuição final consiste numa API que implementa diretamente estas conclusões empíricas. O sistema integra o modelo melhor YOLO específico e a estratégia de pré-processamento (por exemplo, recorte versus imagem completa) adaptada a cada requisito individual. A validação numa base de dados universitária real demonstrou que apenas 12% das imagens atualmente aceites estavam em total conformidade, sublinhando a gravidade do problema da verificação manual e a eficácia da solução otimizada proposta neste trabalho.

Palavras Chave: Fotografias Institucionais, Validação de Requisitos em Fotos, ISO/ICAO, YOLO

Abstract

The verification of photographs in academic institutions is currently a subjective and inconsistent process. Unlike passport control, which adheres to strict international standards (ISO/ICAO), institutional environments lack a universal definition for acceptable photographs, resulting in variable criteria and a reliance on manual approval. This dissertation addresses this inefficiency by establishing a formal classification of visual requirements for academic identification and developing an automated classification system, based on deep learning, to ensure compliance.

Two distinct datasets were used in this study: a private database of photos used in an institutional environment, and a public dataset of images that are usually used to train models for feature identification in passport images. Comparative experiments were conducted between Weightless Neural Networks (WNN) and YOLO models (v8 and v11). The results unequivocally identified YOLO models as superior, consistently outperforming WNNs and the CNN models used to obtain baseline metrics. Additionally, a critical analysis of data quality revealed that models trained on the expert-labeled public dataset achieved better generalization than those trained on institutional data. This demonstrates that label accuracy and data balance are more valuable for this study, yielding better results than training with the specific institutional image dataset.

The final contribution consists of an API that directly implements these empirical findings. The system integrates the optimal specific YOLO model and a pre-processing strategy (e.g., cropping versus full image) adapted to each individual requirement. Validation on a real university database demonstrated that only 12% of currently accepted images were in full compliance, underscoring the severity of the manual verification problem and the effectiveness of the optimized solution proposed in this work.

Keywords: Institutional Photo, Photo Requirements Validation, ISO/ICAO, YOLO

Contents

Acknowledgment iii

Resumo v

Abstract vii

List of Tables xiii

List of Figures xv

Glossary of Abbreviations and Acronyms xvii

Chapter 1. Introduction 1
1.1. Context 1
1.2. Motivation 1
1.3. Objectives 2
1.4. Research Questions 2
1.5. Dissertation Contributions 3
1.6. Dissertation Structure 3

Chapter 2. Background Concepts and Related Work 5
2.1. Methodology 5
2.2. Technologic Background 5
2.2.1. ISO/ICAO Standards 5
2.2.2. Gamification 8
2.3. Previous Implementations 8
2.3.1. Parente et al. [**1**] 10
2.3.2. HMAX [**2**] 11
2.3.3. ICAONet [**3**] 13
2.4. Models Tested in This Dissertation 14
2.4.1. VGG and ResNet Models 14
2.4.2. WNN (Weightless Neural Network) 15
2.4.3. YOLO (You Only Look Once) 16
2.5. Discussion and Conclusions 16

Chapter 3. Methodology 17
3.1. Approach to the problem 17
3.2. Institutional Photo Definition 17

3.2.1. Characteristics of an Institutional Photo 17
3.2.2. Requirements to define an Institutional Photo 17
3.2.3. Comparation with ISO/ICAO requirements 19
3.3. Private Dataset 20
3.3.1. Image Labeling 21
3.3.2. Train and Validation Dataset 25
3.4. Public Dataset 26
3.5. Comparison of Datasets 27
3.6. Modeling 28
3.6.1. Pre-processing 28
3.6.2. Baseline Models 31
3.6.3. WNN 32
3.6.4. YOLO 33
3.7. Trained Models 33
3.8. Training Average Time 35
3.9. Experiments 35
3.9.1. Experiment sections and subsections 36
3.9.2. Description of Experiments Performed 38
3.9.3. Experiments Summary 39

Chapter 4. Results and Discussion 41
4.1. Criteria for Choosing the Best Metric per Requirement 41
4.2. Section A- Intra-Dataset Evaluation 42
4.2.1. Experiment A.1. 42
4.2.2. Experiment A.2. 44
4.2.3. Experiment A.3. 46
4.3. Section B- Cross-Validation Dataset 47
4.3.1. Experiment B.1. 47
4.3.2. Experiment B.2. 49
4.3.3. Experiment B.3. 50
4.4. Summary of Best Results 53
4.5. Average Image Prediction Time 54

Chapter 5. API Development 57
5.1. API Purpose 57
5.2. Implementation Baseline 57
5.3. Application Architecture 58
5.3.1. Programmatic Interface (API) 58
5.3.2. Graphical User Interface (GUI) 59
5.4. Considerations 60

Chapter 6. Conclusions and Future Work 61

6.1. The Critical Role of Data Quality 61
6.2. Model Performance and Architecture 61
6.2.1. Deep Learning Dominance (RQ3) 61
6.2.2. Limitations of Lightweight Models (RQ2) 62
6.3. Requirement-Specific Optimization 62
6.4. Methodological Limitations 62
6.5. Future Work 63

Bibliography 65

Appendix A. Merged Datasets Distribution 69
A.1. Merged Train Private Validation 69
A.2. Merged Train Public Validation 69

Appendix B. Best Results by Model for Section A Experiments 71
B.1. Experiment A.1. 71
B.1.1. Baseline Models- A.1. 71
B.1.2. WNN Models- A.1. 71
B.1.3. Yolov8 Models- A.1. 72
B.1.4. Yolo 11 Models- A.1. 72
B.2. Experiment A.2. 72
B.2.1. Baseline Models- A.2. 72
B.2.2. WNN Models- A.2. 73
B.2.3. Yolov8 Models- A.2. 73
B.2.4. Yolo 11 Models- A.2. 73
B.3. Experiment A.3. 74
B.3.1. Baseline Models- A.3. 74
B.3.2. WNN Models- A.3. 74
B.3.3. Yolov8 Models- A.3. 74
B.3.4. Yolo 11 Models- A.3. 75

Appendix C. Best Results by Model for Section B Experiments 77
C.1. Experiment B.1. 77
C.1.1. Yolov8 Models- B.1. 77
C.1.2. Yolo 11 Models- B.1. 77
C.2. Experiment B.2. 78
C.2.1. Yolov8 Models- B.2. 78
C.2.2. Yolo 11 Models- B.2. 78
C.3. Experiment B.3. 78
C.3.1. Yolov8 Models- B.3. 78
C.3.2. Yolo 11 Models- B.3. 79

List of Tables

1 "Face Image ISO Compliance Verification" benchmark requirements [**4**]. 7

2 ISO/ICAO standards tests baseline according to [**3**]. 9

3 Different HMAX layers for brain Visual pathway simulating [**5**]. 12

4 Institutional photograph requirements. 18

5 Distribution of Images with final classification by requirement. 23

6 Distribution of images with 2 different labels by requirement. 24

7 Distribution of labeled images by requirement. 24

8 Final distribution of Images by requirement. 25

9 Private dataset division in Train and Validation datasets. 26

10 Public Dataset requirements distribution for Train and Validation datasets. 27

11 Private dataset 6 most common dimensions. 28

12 Number of images that suffered Cropping preprocessment. 31

13 Description of the Datasets used for Modeling. 34

14 Models Average Training Time. 35

15 Summary of Experiments of Section A. 40

16 Summary of Experiments of Section B. 40

17 Results obtained by ICAONet network validation dataset. 42

18 Experiment A.1. best models by requirement. 43

19 Comparation of YOLO models metrics for Public dataset results without train dataset changes. 43

20 Comparation between ICAONet and this studies metrics for the same training and validation datasets. 44

21 Experiment A.2. best models by requirement. 44

22 Comparation of metrics between best models by requirement of A.1. and A.2. Experiment. 45

23 Experiment A.3. best models by requirement. 46

24 Experiment B.1. best models by requirement. 47

25 Comparation of YOLO models metrics for models of Experiment B.1.. 48

26 Experiment B.2. best models by requirement. 49

27 Comparation of YOLO models metrics for models of Experiment B.2.. 49
28 Comparation of metrics between best models by requirement of B.1. and B.2. Experiment 50
29 Experiment B.3. best models by requirement. 51
30 Comparation of YOLO models metrics for models of Experiment B.3.. 51
31 Best models per requirement using Private dataset training images only and Public dataset validation images 52
32 Comparison of the best results from models trained on the Private dataset only versus the best models from Experiment B.3, evaluated on the Public validation set 52
33 General best model per requirement that are in the final API. 53

List of Figures

1 "Face Image ISO Compliance Verification" benchmark requirements representation [**4**]. 7

2 Parent et al. implementation for an image with hair across the face. 10

3 Parent et al. implementation for an image without hair across the face. 10

4 Proposed method for Facial Image Assessment using HMAX model [**2**]. 12

5 HMAX Model. 12

6 ICAONet Architecture. 13

7 Example of question and interface of the tool. 23

8 Original size (194x259). 29

9 Resize without cropping. 29

10 Cropped Image. 29

11 Original Image. 30

12 Image with cropped preprocessing applied. 30

13 Average Prediction time results for YOLOv8 (requirement 2). 54

14 Average Prediction time results for YOLO11 (requirement 2). 54

15 Example of final GUI classifying an image. 59

Glossary of Abbreviations and Acronyms

AI - Artificial Intelligence
ANN - Artificial Neural Network
API - Application Programming Interface
BPI - Bits per Input
CNN - Convolutional Neural Network
DL - Deep Learning
DPO - Data Protection Office
ERR - Equal Error Rate
GDPR - General Data Protection Regulation
GUI - Graphical User Interface
IAM - Identity and Access Management
ICAO - International Civil Aviation Organization
IEC - International Electrotechnical Commission
ISO - International Standard Organization
ROI - Region of Interest
SSO - Single Sign On
SVM - Support Vector Machine
WNN - Weightless Neural Network
XOR - Exclusive OR
YOLO - You Only Look Once

CHAPTER 1

Introduction

1.1. Context

In education institutions, any new student or employee is required to go through a registration process, which involves the collection of their personal data. A crucial requirement of this collection is a photograph that identifies the individual. This photo must be validated to ensure it maintains the decorum and visual consistency required by the institutional environment. As thousands of people enrol in higher education at the start of each academic year, thousands of photographs need to be verified.

While strict international standards exist for identification documents such as passports (e.g., ISO/IEC 19794-5), there is no single, unique definition for an "institutional" or "academic" photograph. Instead, each institution establishes its own specific criteria and acceptable boundaries. For instance, a university may permit a smiling student, whereas a passport control system would strictly reject such an image. This variability implies that a "one-size-fits-all" automated solution is rarely applicable across different organizations.

Consequently, this lack of a standardized definition impacts the availability of suitable software. Since existing automated tools are typically designed for rigid passport compliance, they lack the flexibility required for institutional contexts. As a result, many institutions rely on manual verification to handle their specific, heterogeneous criteria.

Therefore, there is a need for a classification approach that moves beyond exclusive manual verification but allows for more flexibility than strict passport control algorithms.

1.2. Motivation

One of the main purposes of using photographs in an academic environment, or even in companies, is to identify the person themselves. To this end, when a person enrols at university or takes up a new job, depending on the context in question, they are asked for a photograph that identifies them.

This photo will then be analyzed to check whether it really identifies the person, and whether the photo itself has the necessary characteristics to belong to an institutional environment. This verification is the responsibility of the university/employment, and the form normally used is a purely manual verification for approval. As seen above, without the definition of criteria that characterize what an institutional photograph is, this classification is subjective and sometimes not very effective, accepting photos that make no sense.

Thus, the main motivation for this study lies in the need to create a more automated method that no longer relies solely on human approval to validate photos. A further motivation is the absence of a universal standard for institutional photographs. This study addresses this challenge by identifying the core characteristics that effectively define such images within an academic context, thereby establishing a baseline for automated classification.

1.3. Objectives

The primary objective of this study is to identify the characteristics that define an institutional photograph and to determine a realistic approach for its classification.

Once these criteria are established, the subsequent objective is to develop an algorithm trained to verify whether a given set of pass-type photographs meets these specific requirements.

Finally, this study aims to implement an API to assist in the verification process. This tool is intended to allow administrative staff to upload photographs and receive an automated analysis of each characteristic, along with an indicator of compliance.

1.4. Research Questions

In order to achieve the objectives set out in this study, it will be necessary to understand the characteristics that defines an institutional photo in order to be able to classify it. Initially, it is essential to identify fundamental characteristics—features that may seem intuitive but require formal definition to serve as a basis for automation. For instance, an institutional photograph typically depicts a single subject, facing the camera, set against an appropriate background. However, vague descriptors such as "appropriate background" necessitate precise technical specification.

Once we have a general idea of what an institutional photo is, we need to define specific criteria that clearly define it. One standard that can serve as a basis for defining these criteria are the ISO/ICAO standards [**6**], which define passport photographs. An institutional photo can have several associated passport photo bases, but it is necessary to understand how rigid the definition needs to be. For example, in a passport photo, the person cannot smile or have their mouth open. Is this a criterion that makes sense in institutional photos? Examples like these need to be discussed and conclusions drawn.

In light of these complexities, the research questions guiding this study are as follows:

- **RQ1:** What are the characteristics and requirements that define an institutional photo?
- **RQ2:** To what extent can computationally lightweight models, such as Weightless Neural Networks (WNN), achieve competitive performance when compared to more complex models?
- **RQ3:** Which model best identifies features of institutional photographs?

1.5. Dissertation Contributions

This dissertation presents several contributions in the domain of image classification, with a particular focus on institutional photographs. The main contributions are outlined below:

- **Dataset Classification and Optimization:** A comprehensive classification of an image dataset was conducted, focusing on identifying and labeling specific visual characteristics relevant to institutional photographs. Additionally, the datasets were optimized to enhance their usability and improve model training performance.
- **Comparative Model Evaluation:** Extensive experiments were carried out using a variety of machine learning and deep learning models to assess their effectiveness in classifying images based on visual features. This involved benchmarking models across different metrics to identify the most suitable approach for the specific classification task.
- **API Development for Institutional Use:** An API was developed to support the university, providing a tool to assist in the verification and quality assessment of submitted photographs. This API is designed to be integrated into existing workflows and aims to automate part of the manual review process.
- **Scientific Dissemination:** As part of the dissemination of the work, a scientific article has been written.

1.6. Dissertation Structure

This dissertation is structured into six chapters, each addressing a critical aspect of the research and development process.

- **Chapter 1** provides a brief introduction to the dissertation topic, outlining the motivations behind its selection and the primary goals of the project. It sets the stage for the work by contextualising the research focus.
- **Chapter 2** presents the literature review, introducing key concepts essential for understanding the study. It examines findings from similar works and positions the dissertation within the existing body of research.
- **Chapter 3** describes the methodology employed in the study. It outlines the problem being addressed, details the dataset used, explains the preprocessing steps, and introduces the models applied—justifying their relevance to the research. Explica também todos os Experiemnts efetuados no etsudo
- **Chapter 4** presents the results of each experiment and analyzes the outcomes, discussing the findings and drawing conclusions regarding model effectiveness. It evaluates the performance of each approach and identifies the most suitable models for the problem at hand.

- **Chapter 5** describes the API developed as part of the project. It explains its functionality, how to access it, and highlights its practical utility within the context of the overall system.
- **Chapter 6** reflects on the research questions, considering what could have been done differently, and provides a critical reflection based on the conclusions drawn.

CHAPTER 2

Background Concepts and Related Work

2.1. Methodology

With the exponential growth of AI, the number of papers related to image processing and classification has increased significantly. Despite this growth, research specifically addressing the validation of "institutional" photographs—as distinct from strict passport standards—appears less frequently in standardized scientific venues. This is likely not due to a lack of interest, but rather because there is no single, unique definition of an institutional photograph; each institution establishes its own specific criteria. Consequently, while systems with similar objectives may exist within specific organizational contexts, they do not adhere to a universal standard, resulting in a fragmented body of literature compared to the rigorous ISO/ICAO passport studies.

To identify the most relevant work in this area, a multi-stage approach was adopted. Initially, three keywords were defined: 'ICAO Standard,' 'Passport photo classification,' and 'Passport feature extraction.' The IEEE Xplore database was utilized to retrieve results, applying exclusion criteria that required studies to specifically address ICAO standards and photograph classification.

Based on the results obtained through the previous criteria, Google Scholar was also used. This platform proved valuable for verifying the references of the selected studies. Additionally, the feature that shows who has cited a particular study was particularly useful, as it helped identify further relevant research and expand the scope of the literature review.

2.2. Technologic Background

Before delving deeper into this study, it's important to explore and understand some related topics that are crucial for a comprehensive understanding of the dissertation.

Some of the important topics we will discuss include:

- ISO/ICAO standards [**6**];
- Gamification;

2.2.1. ISO/ICAO Standards

The image that each person has in their passport follows defined criteria, which have been previously thought out and discussed. The quality of an image for documents such as a passport is determined on the basis of international standards that can be found in ISO/IEC 19794 [**7**] and ICAO9303 [**8**].

Starting with the standard defined by International Civil Aviation Organization (ICAO) [**8**], a specific working group was set up with the aim of identifying physical characteristics that would allow a person's identity to be identified using a machine. The physical characteristics discussed for the main identifier were the face, iris and fingerprints, with the face being defined as the main one. Specific standards necessary for identifying a person's identity were then defined [**6**]. The ICAO standards were first published in 2003 and have been updated several times over the years, the most recent being published in 2021.

ISO19794-5 [**7**], is a proposed standard for the facial photograph to be used in electronic passports. It was defined by the International Standards Organization (ISO), in conjunction with the International Electrotechnical Commission (IEC). Based on the guidelines defined by the ICAO standards, this standard defines a set of conditions that must be met in order to identify an individual.

According to the work of [**6**], both the ICAO standards and the ISO19794-5 standard have criteria that are not self-explanatory, and are not as clear as possible. In order to eliminate this ambiguity, the authors presented 30 well-defined characteristics based on the ISO/ICAO standards to determine the quality of an image [**9**].

The same authors also provided a benchmark, that can be found on the FVC-onGoing website [**4**]. The FVC-onGoing is an ongoing online competition of fingerprint recognition algorithms, where participants can test their algorithms on the provided benchmarks, and publish their results. Within the available benchmarks, there is the "Face Image ISO Compliance Verification" benchmark. This benchmark provides 24 tests and each test evaluates a different compliance requirement. These 24 requirements are a subset of the initial 30 characteristics defined by the authors. It is important to note that of the 24 defined requirements, the competition mainly tests the last 23, given that the first requirement is not a characteristic but rather the detection of the eye center location."

The "Face Image ISO Compliance Verification" benchmark is defined by 24 requirements, as the following Table 1 presents:

Nº	Description of the test
Feature extraction accuracy tests	
1	Eye center location accuracy
Photographic and pose-specific tests	
2	Blurred
3	Looking away
4	Ink marked/ creased
5	Unnatural skin tone
6	Too dark/ light
7	Washed out
8	Pixelation
9	Hair across eyes
10	Eyes closed
11	Varied background
12	Roll/pitch/yaw rotations greater than a predefined thresholds
13	Flash reflection on skin
14	Red eyes
15	Shadows behind head
16	Shadows across face
17	Dark tinted lenses
18	Flash reflection on lenses
19	Frames too heavy
20	Frames covering eyes
21	Hat/cap
22	Veil over face
23	Mouth open
24	Presence of other faces or toys too close to face

TABLE 1. "Face Image ISO Compliance Verification" benchmark requirements [**4**].



FIGURE 1. "Face Image ISO Compliance Verification" benchmark requirements representation [**4**].

These 24 requirements, shown in Figure 1, will be an important baseline for our study, and will be useful for cementing the criteria for an institutional photo, checking whether

a given criterion is too rigid for a photo of that style. It also allows us to see if there are specific characteristics of institutional images that are not weighted in these criteria.

2.2.2. Gamification

In this study, once the characteristics that define an institutional photograph have been defined, it is necessary to use a database of passport photos that have already been properly classified in terms of each of the requirements defined, since this is a supervised learning problem [**10**].

The database that will be used belongs to a higher education institution with around 100,000 students and ex-students. Since the database is not properly classified according to the characteristics that will be defined, and since this task is quite time-consuming to do individually, a Gamification strategy will be used. Gamification itself is based on the use of gaming techniques, such as a level system, points for actions, or a badge acquisition system, to motivate users to collaborate with the study [**11**].

The aim of this technique is to help assign labels to the dataset images according to the requirements, with the possibility of having more than one person involved in the classification. As the users who classify are not real experts in the requirements of institutional photographs, and their motivation and concentration tend to diminish over time, it is not advisable to rely solely on the classification of a single user.

2.3. Previous Implementations

As mentioned earlier, the ISO/ICAO standards, redefined in [**6**], can be instrumental in the task of defining the requirements for an institutional photo.It was also noted that the authors of that study organized a competition, namely the FVC-onGoing [**4**], where participants can assess their algorithms using provided benchmarks and publish their results.

The competition includes a tool that compares participants models with the best-performing models for each requirement and provides comprehensive results. The latest published results associated with the competition are from [**3**]. The requirements are evaluated individually in terms of Equal Error Rate (ERR). The ERR is the threshold at which the system is equally likely to wrongly accept a non matching individual as it is to wrongly reject a matching individual. EER is an important metric in biometric security systems as it provides a balance between security and convenience. By analysing the results obtained from their algorithm, ICAONet, with other top-performing methods, the current baseline for the 23 requirements - all applied to the same database - is presented in Table 2.

Requirement	Method	EER	Rejection
2	BioPass Face	1.60	3.30
3	ICAONet	5.40	0.00
4	BioLab	3.40	1.20
5	ICAONet	1.70	0.00
6	ICAONet	1.20	0.00
7	BioPass Face	0.00	0.00
8	SDK 2	0.00	0.00
9	Parente et al.	11.90	3.40
10	id3	0.20	1.00
11	BioTest	3.70	7.90
12	ICAONet	4.60	0.20
13	BioLab	0.60	0.00
14	id3	1.00	2.00
15	BioLab	2.30	0.20
16	ICAONet	3.30	0.00
17	ICAONet	0.40	0.00
18	ICAONet	0.80	0.00
19	HMAX	0.00	0.00
20	HMAX	0.00	0.10
21	ICAONet	5.80	0.00
22	ICAONet	0.00	0.00
23	id3	0.60	0.40
24	BioPass Face	1.20	2.70

TABLE 2. ISO/ICAO standards tests baseline according to [**3**].

From this baseline, it is possible to understand that there are eight different approaches that yield the best results in at least one of the 23 requirements. These approaches are:

- BioPass Face
- BioLab
- BioTest
- SDK 2
- id3
- HMAX
- Parente et al.
- ICAONet

Given that the ISO/ICAO standards align with the research questions and serve as the foundation for the study, our objective is to comprehend each of the best implementations to determine which one best suit our problem.

As can be seen from the results presented in Table 2, it is possible to see that although the ISO/ICAO standards were defined more than 10 years ago, and the competition is open and makes it possible to constantly submit new studies to improve the results, there are still requirements that have an ERR of more than 10%, which shows that there is still a significant margin for innovation on the subject.

In addition, of the 8 approaches that obtain at least one requirement with the best results, 4 of them belong to private companies: BioTest, BioPass Face, id3 and ICAO SDK. This privatization means that the results are not public, which does not contribute to faster progress in the area [**3**].

Of the 4 remaining approaches, BioLab's approach turns out not to be very specific. Although the implementation achieves the best results in the Flash reflection on skin (13)

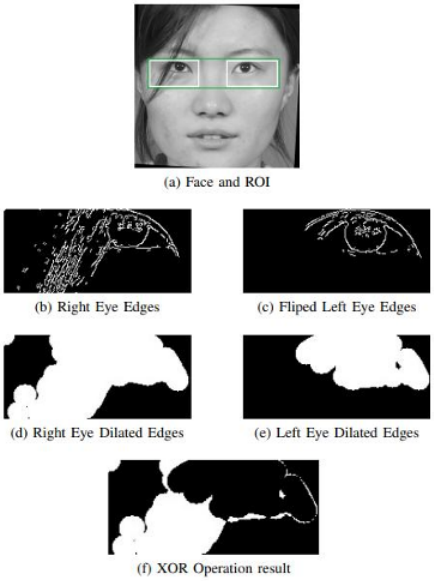


FIGURE 2. Parent et al. implementation for an image with hair across the face.

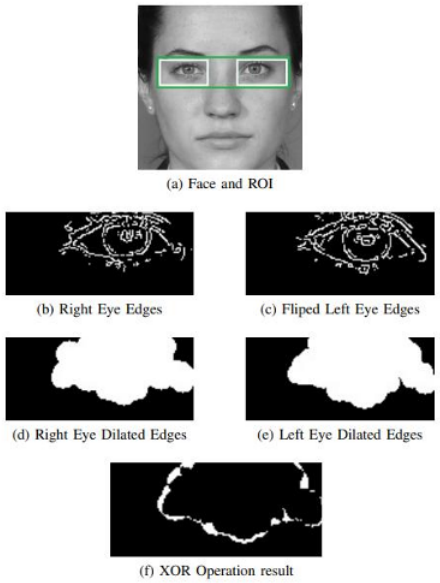


FIGURE 3. Parent et al. implementation for an image without hair across the face.

requirement, the implementation is not clear, which does not provide much in the way of conclusions or inspiration for this study.

As for the other 3 implementations, Parente et al. [**1**], HMAX [**2**] and ICAONet [**3**], the different approaches and techniques they used to obtain their results are notable.

2.3.1. Parente et al. [1]

Giving priority to the number of requirements whose result is the best, the first implementation to be addressed is that of Parente et al. [**1**]. This study, carried out in 2016, decided to create a specific method for 4 requirements of the ISO/ICAO standards. These were: Image Pixelation, presence of hair across eyes, veil over the face and mouth opened. Currently, only the implementation made to check if the person has hair in front of their face gives the best results, so only this implementation will be analyzed further. The aim of the requirement is to ensure that people's eyes are fully visible, guaranteeing that there is no hair in front of the eyes.

The algorithm developed in [**1**], was implemented taking into account that the human face is symmetrical. This symmetry means that if one of the eyes is inverted and overlaid on the other, the images look very similar. If, when this overlay is made, one of the eyes has hair in front of it, the difference between the images will be noticeable and it will be possible to detect whether there is hair in front of the eyes in the image.

Figures 2 and 3 represent cases of images with and without hair, and how they are approached by the algorithm. This implementation itself is based on older image processing techniques, which can be studied in depth at [**12**, **13**].

In a nutshell, we start by removing the slope between the eyes and then locate the pupils. With this location, the surrounding area that covers the entire eye is defined. Then, the LogAbout filter [**14**], whose function is to reduce the influence of luminosity and shadows projected onto the face, is applied. Then, the Canny edge detection filter, proposed by [**15**], is run to check the boundaries of the eyes and the possible presence of hair in the image.

Finally, the image of the left eye is inverted and overlaid on the right eye. Thus, as represented in figures 2 and 3, in this case the f points of each of the figures, an XOR (Exclusive OR) [**16**] is performed, and if there are differences between the 2 images, it is possible to detect the presence of hair or not. The HMAX and ICAONet implementations differ in that they are based on DL (Deep Learning) [**17**, **18**]. Both implementations use Artificial Neural Networks (ANN) [**19**, **20**]), more specifically Convolutional Neural Networks (CNN) [**21**–**23**], as the subject is related to image processing and classification.

2.3.2. HMAX [2]

Following the logic of priority, the next implementation to be explored is HMAX [**2**], which has the best results in 2 of the 23 requirements of the ISO/ICAO standards which are frames too heavy (19) and frames covering eyes (20) (Table 2). In this implementation, they decided to only use 9 requirements of the 23, which were Looking Away (03), Unnatural Skin Tone (05), Hair Across Eyes (09), Head Rotation (12), Red Eyes (14), Shadow Across Face (16), Frame Too Heavy (19), Frame Across Eyes (20) and Mouth Open (23) (Table 2). This decision is based on the accuracy of each requirement, with the ones chosen being more challenging to achieve good results.

Figure 4 is a representation of the proposed method for Facial Image Assessment using HMAX model. The implementation itself is divided in different phases.

An initial phase, where the input images are pre-processed, fundamental image processing steps are executed to generate a suitable image for further analysis. In the next phase, the process consists of acquiring facial semantic patches for both key and non-key components. The human face generally has similar elements located in the same places, such as the eyes, nose, mouth and so on. With this ease in predicting the location of these elements, divisions are made into 8 components, 4 key components, (the 2 eyes, the mouth and the nose), and 4 non-key components (the cheeks, the forehead and the chin). The size of each division is adjusted according to the distance between the 2 eyes. The Viola-Jones algorithm [**24**] is also used in the process [**2**].

Once the key and non-key components have been obtained from the images, they will be the inputs to the HMAX model, as shown in Figure 4.

This model expresses a hierarchy of brain regions through which object recognition is performed in the cerebral cortex. The purpose of this model is the cognitive phenomena describing in terms of simple and complex computational processes in an acceptable physiological model, as illustrated in Table 3.

In Figure 5, there´s a representation of how the model operates.

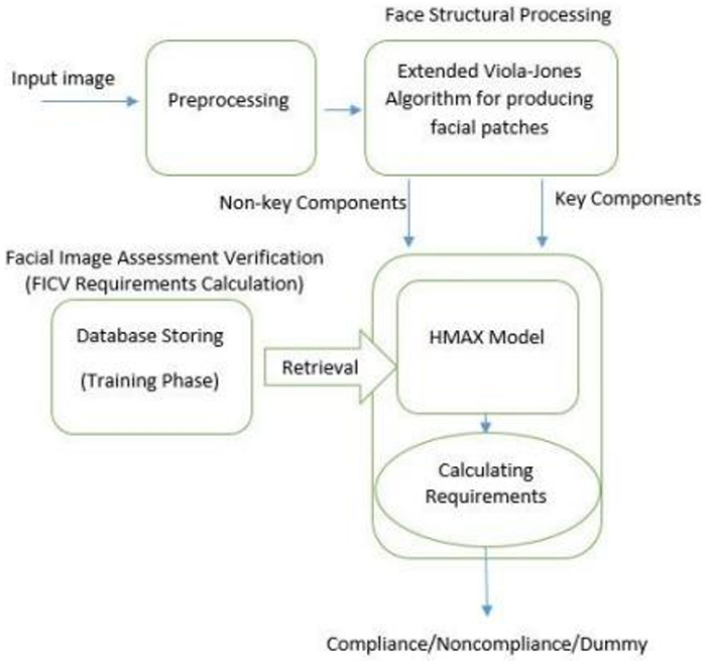


FIGURE 4. Proposed method for Facial Image Assessment using HMAX model [**2**].

Layers of the HMAX model	Ventral Visual Pathway of the Cerebral Cortex
S1 (The first Simple layer)	V1 (The primary Visual Cortex)
C1 (The first Complicated layer)	V2 (The second Visual cortex)
S2 (The Second Simple layer)	V4 (The fourth Visual cortex)
C2 ((The Second Complicated layer)	IT (The InferoTemporal cortex)

TABLE 3. Different HMAX layers for brain Visual pathway simulating [**5**].

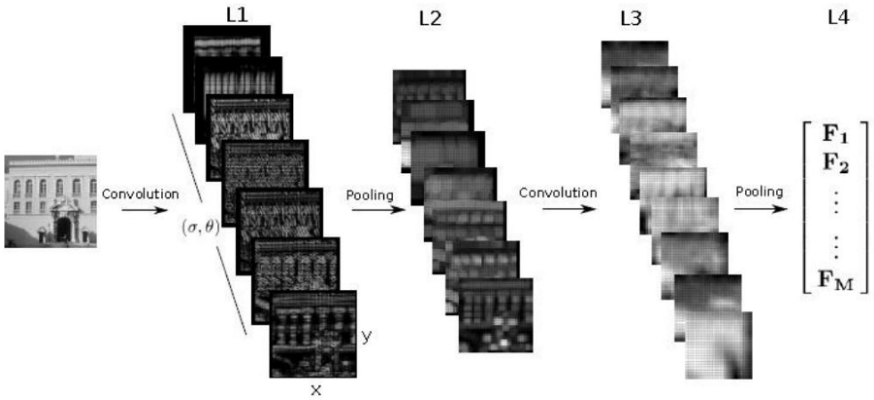


FIGURE 5. HMAX Model.

The model is divided into 4 layers, 2 simple layers and 2 complex layers. In the first layer (First Simple Layer), the patches defined in last phase are the inputs, and to these are applied a set of Gabor filters [**25**]. These filters are used to simple cell activation in the V1 region of the visual cortex modeling. In this second layer (First Complex Layer), each feature of this layer is a dimension reduction of the previous layer. The dimension

reductions was obtained by applying the process of Max Pooling [**26**]. The third Layer (Second Simple Layer) is the result of applying HL filters to the output of the second layer. The fourth and last layer (Second Complex Layer) produces a vector of approximately 1000 dimensions, determining each coefficient for the maximum output of each HL filter for every scale in every location. All this process it represented in Figure 5

Looking back to Figure 4, it's possible to understand that there´s a database storing retrieving to the HMAX Model, used in the training phase. In this phase, each labeled photo is associated with nine labels, and each label is stored in the corresponding requirement subspace.

To classify an image that wasn´t labeled during the training phase/storing, the HMAX model compares the new photo with the requirement stored to make a decision. With this purpose, nine binary SVMs classifiers [**27**], each one for each of the nine requirement, classify each requirement of the new photo, with 1 for compliant, 0 for non-compliant and -1 for dummy classes. For 1 and 0 classifications, there´s a score in range of zero to 100.

2.3.3. ICAONet [3]

The last implementation to be explored is ICAONet, from [**3**]. This implementation obtained the best results in 9 of the 23 requirements which are Looking away (03), Unnatural skin tone (05), Too dark/ light (06), Roll/pitch/yaw rotations greater than a predefined thresholds (12), Shadows across the face (16), Dark tinted lenses (17), Flash reflection lenses (18), Hat/cap (21) and Veil over face (22).

The overall architecture of ICAONet is composed of an Autoencoder [**28**] in combination with a branch with a dense network.

In an initial phase, the input images are initially processed in BGR colors. Facial detection is performed using a single-shot multi-box detector (SSD). However, the bounding box generated by this detector only covers the region from the forehead to the chin. To capture a more comprehensive view, the bounding box is expanded by 1.5 times, ensuring that the background is also included in the detection area.

Getting into the architecture of the ICAONet network itself [**3**], it is made up of 3 main components, a shared network, an unsupervised network and a supervised network.

As we can see in Figure 6, there´s the ICAONet architecture.

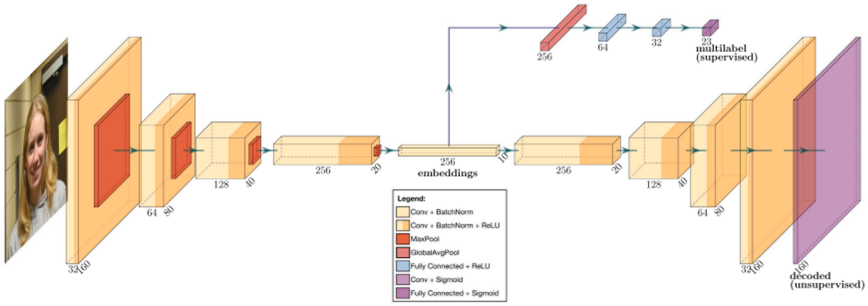


FIGURE 6. ICAONet Architecture.

The initial part of the network is the shared network part, which is responsible for learning the embeddings shared by the unsupervised and supervised branches. These embeddings can be interpreted as a proper input representation Besides being used for dimensionality reduction, the shared embeddings are also used for feature learning [**3**]. In this layer, 4 convolutional layers composed of 3x3 convolutions with batch normalization filters [**29**] and the ReLU activation function [**30**] are applied to the input images.

At the end of each layer, a Max-pooling dimension reduction layer [**26**] is applied. The fifth layer is also composed of 3x3 convolutions with batch normalization. However, *tanh* [**30**] is used as a nonlinear activation function instead of ReLU to normalize the embeddings into values between 1 and 1.

The unsupervised branch works as the decoder of an Autoencoder [**28**], and the main goal is to get the result of the shared network and to get the original inputs again. To achieve this purpose, the first four layers of the shared network will be the same in this branch, except the Max-pooling, that is replaced by 2D Transposed Convolutional layers [**31**]. The last layer activation function is of sigmoid type [**30**] since the input image pixels are already normalized into the [0,1] range.

The last network to be discussed is the supervised branch. Their inputs are the embeddings, that are the result from the shared network. The goal of this network is to perform multi-label classification. The first layer of this branch performs a GlobalAveragePooling [**32**] to the shared embeddings

After this layer, there are two consecutive dense layers, with 64 and 32 neurons. Dropout layer [**33**] are included between each layer to prevent overfitting. The final layer, is a 23 neurons, each one corresponds to the 23 requirements. Each one of the neurons outputs an normalized score (0-1).

2.4. Models Tested in This Dissertation

Although several models have been previously used for similar purposes, such as identifying features in images, this dissertation focuses on testing a set of alternative implementations. These models were selected due to their relevance and applicability to the problem of institutional image classification. The models tested are:

- VGG
- ResNet
- WNN (Weightless Neural Network)
- YOLO (You Only Look Once)

2.4.1. VGG and ResNet Models

The VGG and ResNet models were used in this dissertation as Baseline Models. The goal with these models was to obtain quick results without the need to modify the architecture, allowing for the achievement of initial metrics in a straightforward manner.

In the case of VGG, the pre-trained model used was specifically VGG16. VGG-16 is a deep convolutional neural network architecture comprising 13 convolutional layers followed

by 3 fully connected layers, all using small 3×3 filters and 2×2 max-pooling operations, which collectively enable deep hierarchical feature learning with architectural simplicity [**34**–**36**]. This design captures fine-grained visual characteristics—edges and textures in early layers and more complex patterns in deeper layers—making VGG-16 particularly effective for image characteristic identification. The VGG16 model was ultimately chosen over VGG19 because it is smaller and faster, which aligns with the objectives of this work.

For ResNet, the pre-trained model used was specifically ResNet-50. ResNet-50 is a 50-layer convolutional neural network introduced by He et al. [**37**] that pioneered the use of residual connections to train much deeper models without vanishing-gradient issues [**37**]. In this architecture, each convolutional block learns a residual function by adding its input to its output, allowing gradients to propagate through dozens of layers and enabling effective feature learning at great depth. ResNet-50's design led to remarkable image feature extraction performance and state-of-the-art recognition accuracy – the ResNet family (which won the 2015 ImageNet competition) achieved much deeper networks (up to 152 layers) with lower complexity than prior CNNs like VGG [**37**, **38**].

This model was ultimately selected because it offers the best balance between performance and operational cost. It meets the intended requirements of being a fast model capable of producing good results without the need for architectural modifications, allowing for the efficient acquisition of initial metrics.

2.4.2. WNN (Weightless Neural Network)

At the time this dissertation is being written, WNN networks had not yet been used for the purpose of identifying features in institutional images. This model is particularly interesting due to its low computational footprint. In the case of positive results, it is an effective model for delivering quick responses when called via API. The model's low computational overhead makes it particularly well-suited for mobile device implementation, which is a key reason for our interest in it.

More specifically, Weightless Neural Networks (WNNs) are neural models that use random-access memory (RAM) devices in place of weighted synapses, storing binary input patterns as addresses in memory rather than adjusting numeric weights [**39**]. Originally introduced in the 1980s (e.g. the WiSARD system by Aleksander et al.) [**40**], a WNN neuron "learns" by recording which input bit patterns (tuples) map to a 1 or 0 output in a RAM-based lookup table [**39**]. This architecture yields extremely fast training and inference, since learning is essentially a one-pass update of a truth table and inference is a simple table lookup – no iterative weight tuning or costly matrix multiplications are required [**41**]. The absence of weight adjustments confers several advantages: WNN models are computationally efficient, with low latency and energy use, and they can be trained in one shot with new data without retraining the entire network [**39**].

2.4.3. YOLO (You Only Look Once)

YOLO is a viral and widely used algorithm [**42**]. YOLO is famous for its object detection characteristics. In 2015, Redmon et al. gave the introduction of the first YOLO version [**43**]. In the past years, scholars have published several YOLO subsequent versions [**44**–**47**]. At the time of writing, the most recent version of YOLO with classification mode is YOLO 11 [**46**, **48**]. For this study, it was decided to check the performance differences of the same datasets for 2 different YOLO models. Thus, the models trained and evaluated in this dissertation are the YOLOv8 and YOLO11 models.

2.5. Discussion and Conclusions

This study addresses the challenge of reducing the manual and subjective classification of institutional photographs by proposing a more automated solution: developing an algorithm capable of determining whether images meet the requirements of an institutional photograph. To achieve this, it is essential to first define what constitutes an institutional image and then establish the specific requirements that best characterize it.

To support this effort, previous implementations of the ISO/ICAO standards were analyzed, serving as benchmarks for effective approaches. Table 2 presents 23 ISO/ICAO requirements, and only three public implementations were identified: Parente et al., HMAX, and ICAONet. The Parente et al. approach focused on just four requirements, achieving the best result in only one. HMAX covered 2 of the 23 applicable requirements and led in two. ICAONet emerged as the most effective, obtaining the best results in 9 of the 23 requirements. These results highlight the potential of deep learning (DL)-based models, especially ICAONet, to guide future work.

This comparative analysis not only reinforces the relevance of DL approaches but also helps frame the methodological direction of this study. A key decision moving forward will be whether to adopt a modular approach—treating each requirement individually, as done by Parente et al.—or to pursue an integrated model, as demonstrated by HMAX and ICAONet. Additionally, it may be valuable to explore the potential of Weightless Neural Networks (WNN) and YOLO, to assess their performance in relation to the newly defined institutional requirements.

Ultimately, the goal is to implement an algorithm and an API capable of automatically classifying institutional photographs, indicating which requirements are met and which are not. The lack of a standardized framework in the literature—stemming from the variability of institutional definitions—underscores the importance of this work in proposing a flexible, automated solution.

CHAPTER 3

Methodology

This chapter will describe the proposed methodology that preceded the implementation of the problem, together with the description and analysis of the implementation itself, its problems, reflections and decisions.

3.1. Approach to the problem

As mentioned in Chapter I, the main objective of this dissertation is to define a set of requirements for an institutional photo, so that one or more models can be trained to automatically identify whether or not the photographs meet the requirements in question.

In the same chapter, 3 research questions were raised, whose answer helps to solve one of the study's main objectives. They were:

- **RQ1:** 1. What are the characteristics and requirements that define an institutional photo?
- **RQ2:** 2. To what extent can computationally lightweight models, such as Weightless Neural Networks (WNN), achieve competitive performance when compared to more complex models?
- **RQ3:** 3. Which model best identifies features of institutional photographs?

3.2. Institutional Photo Definition

3.2.1. Characteristics of an Institutional Photo

When analyzing and attempting to answer the first research question, it becomes clear that an institutional photograph is primarily a photo that depicts only a person's face in a passport-style format. This means that only the face and a small portion of the person's upper torso are visible, and the person is in the correct posture. In such a photo, the entire face must be visible and perceptible, including the person's hair, so that it is possible to identify and verify their identity. Extra care must also be taken with the environment in which the photo is taken, and professionalism is required in the background of the image.

3.2.2. Requirements to define an Institutional Photo

With the general characteristics defined, it is important to highlight specific requirements that define what institutional photography is in the best way, as asked in the first research question. Given that only one person's face appears, in the correct proportions, the entire face must be visible and perceptible, these characteristics require specific requirements to be defined.

Therefore, it is mandatory that the photo is not pixelated or blurry and has a good-quality resolution. In addition to resolution, the colors of the photo need to be as natural

as possible. Photos that are too bright, too dark, with washed-out colors, unnatural skin tones or with filters are not suitable for institutional photography.

The eyes are also a very important factor. Since one of the characteristics is that the entire face is visible, the eyes need to be open, with no object obstructing them. Sunglasses must not be present in an institutional photo. People with long hair should ensure that it is not covering any part of the eye, such as significant parts of the face. Wearing glasses in a photo is fine as long as the frames do not cover any part of the eye, and the lenses do not reflect light in a way that obstructs any part of the person's face and makes it difficult to identify them. Still related to the eyes, the gaze of the individual in the photo must be directed towards the camera, in order to facilitate verification of the person's identity.

Since the entire face must be visible, the use of accessories such as hats or caps does not comply with the characteristics of an institutional photograph. In addition to obstructing part of the person's body, which makes the task of verifying identity more difficult, in an institutional environment the individual will not be wearing this type of accessory, so it makes no sense for it to be present in the photo that identifies them. The exception to this requirement is the possibility of the veil being in the photograph, for religious reasons. This use has limitations, as the face must be fully visible.

The posture of the person taking the photo is very important. It is also necessary for the person to look directly at the camera. The body posture itself must be appropriate, and the person must be upright and centered since appearing with the body turned, even when looking at the camera, or arched is not appropriate for an institutional environment.

As for the last characteristic described, related to the background of the image, it is necessary that the focus is on the person's face, and that the background is not distracting or makes it difficult to identify the person. For example, a photo taken with a background on a beach or in a garden is not suitable for an institutional context. Likewise, if the background has several colors or several words. The right background is a flat, one-color background.

In Table 4, we can see the requirements that were choose to define an institutional photograph.

Requirement	Definition
1	Good image quality
2	Person looking directly at the camera
3	Photography with natural colors
4	No hair in front of the eyes
5	Eyes wide open
6	Not having a diversified fund
7	Do not wear sunglasses
8	Glasses lenses don't reflect
9	Do not wear a hat/cap
10	Person must have the correct posture
11	Wearing the veil appropriately

TABLE 4. Institutional photograph requirements.

3.2.3. Comparation with ISO/ICAO requirements

Comparation

The problem of defining characteristics to guarantee the quality of a photo for a professional environment such as an institutional one is not something new, as has already been mentioned. Characteristics have been defined through the ISO/ICAO requirements [**4**], which define what a passport photo should look like.

The decision to define each of the requirements for an institutional photo was based on the characteristics discussed above, along with some of the requirements already defined by the ISO/ICAO requirements, given the similarities between the 2 types. The main difference between a passport photo and an institutional photo lies in their level of strictness. A passport photo must ensure that the individual is identified in the most secure, standardized and professional way possible. An institutional photo shares the same core purpose, identifying the person, but its requirements do not need to be as rigorous or detailed. In an institutional environment, the photo only needs to be professional and restrictive enough to guarantee that the individual can be readily recognized. If a formal identity check is necessary, it is performed using the individual's institutional ID, not only the photo. This means that certain aspects, such as allowing the person to smile, are acceptable in an institutional photo, as the context does not demand the same level of security as, for instance, border control at an airport.

A comparison between the ISO/ICAO requirements and the 11 previously suggested requirements helps clarify the inspirations, differences, and rationale behind each choice.

The first suggested requirement, "Good image quality", is a requirement inspired by requirements 2 and 8 of the ISO/ICAO requirements, more specifically the Blurred and Pixelation requirements.

The third suggested requirement is inspired by the combination of several requirements from the ISO/ICAO Standards, which results in a more simplified criterion. These requirements are 5, 6, 7, 13, 15 and 16, more specifically "Unnatural skin tone", "Too dark/light", "Washed out", "Flash reflection on skin", "Shadows behind head" and "Shadows across face" more respectively. What they all have in common is that they preserve the naturalness of the face color to ensure that the face is identifiable. The 5 requirements turn out to be too specific and it was felt that not so much detail was needed to guarantee something that can be guaranteed with only a single requirement. Although there is the subjectivity factor, since questions can arise when classifying photos as to what natural color really is, all the requirements listed in the ISO/ICAO standards end up having their own degree of subjectivity individually. Basically, this suggestion aims to generalize all the listed requirements into one. If the image is too light, too dark, with washed up colors or with shadows that interfere with the identification of the person's identity, these images are not suitable to be institutional photographs.

The fifth suggested requirement is inspired by requirement 10 of the ISO/ICAO standards, but seeks to evaluate a different concept. While in requirement 10 the aim is to determine whether the eyes are open, and if they are closed the image is considered to be Non-Compliant, our requirement aims to classify as a Non-Compliant image all images whose eyes are not completely open, whether they are half-closed or completely closed.

The remaining requirements defined for this dissertation were inspired by the requirements of the ISO/ICAO standards. Specifically, the second requirement was inspired by requirement 3. The sixth requirement was inspired by requirement 11. The seventh requirement was inspired by requirement 17. The eighth requirement was inspired by requirement 18. The ninth requirement was inspired by requirement 21. The tenth requirement was inspired by requirement 12 but no angle limitations. Only what is strictly necessary, i.e. the person must be facing the camera and not leaning away from it. Finally, the eleventh requirement was inspired by requirement 22.

Conclusions

So, taking a closer look at the list of our requirements, represented in Table 4 in comparison with the ISO/ICAO Requirements, it is noticeable that some of the requirements ended up neither serving as inspiration nor being the basis for some of the requirements altogether. The excluded requirements were requirements 4, 14, 19, 20, 23 and 24, more specifically, Ink Marked/ creased, Red Eye, Frames too heavy, Frames covering eyes, Mouth open e Presence of other faces or toys too close to the face. This exclusion is ultimately justified by the fact that the requirements are too specific, demanding a level of security that is not necessary for the institutional context, or simply do not fit the requirements for institutional photography. For example, in the case of requirement 4, Ink Marked/ creased, nowadays images are very rarely defined based on a real photograph, so the risks of a blur are practically zero. Another example is the case of requirement 23, Mouth open, where the people themselves can be smiling in the photo, showing their teeth, and that's not a problem.

3.3. Private Dataset

Once the requirements that define an institutional photograph, it's time to start the implementation stage. In order to reach the final goal and obtain a model that verifies that one or more characteristics are met, it is necessary to train models. This training phase is only possible using a properly labeled dataset of images. This dataset will be divided into a training set and a validation set.

Starting with the training set, this dataset does not include any images from existing public datasets. Instead, it consists of 4,898 photographs provided specifically for this project through the university's platform. All images are official passport-style photos of students and staff from the institution.

All individuals represented in the dataset have given their consent for their images to be used, in full compliance with General Data Protection Regulation (GDPR) regulations.

The entire process of image collection and data handling was carried out in collaboration with the institution's Data Protection Office (DPO).

3.3.1. Image Labeling

Since there are 4898 photos, and each photo must be labeled for each of the 11 requirements, 53 878 labels need to be made. Making all these classifications by just one individual would make the process excessively long, without tolerance for failure and influenced in a certain way, since there would only be one person labeling, so the assigned label would be solely and exclusively according to that person's vision. To combat these limitations and increase the speed of labeling, it was decided to use a gamification technique.

Gamification techniques usually incorporate elements of games, such as levels or rewards, for tasks that are not games in themselves. In this case, each person was asked to rate 100 photos in total, 50 per requirement for 2 different requirements. This classification involved 50 randomly selected photographs for a random requirement. For each of these photographs, there were two buttons, YES and NO, which people chose to label the image. The total number of photos per classifier was discussed. Initially, the use of 200 photos per person was considered, 100 for each requirement, but this number meant that the total labeling time was around 12 minutes, and the attention and predisposition of the classifiers, being people who don't really have a specialized background in the area of photo analysis and classification, was significantly reduced, increasing the number of errors and deteriorating the classifications. The choice of 100 photos ensures this attention and guarantees better results, although it requires more people to label. As highlighted above, this implementation is intended to be flexible and fault-tolerant. So if one of the people doing the annotation makes a mistake, it doesn't mean that the image will be misclassified. This is because each photo needs 2 equal ratings out of a possible 3 to be given a final rating. This way, the labeling is optimized. When the first 2 ratings of a given photo are equal, a third rating is not necessary, as it no longer changes the final label result. This means that 107756 mandatory labels are required, assuming that there are no different opinions and that all photos have the same top 2 classifications.

In the specific phase of implementing the photo classifier, Django and SQL were used. Django was used for the graphics and for developing the algorithm that managed the photographs. The algorithm was connected to an SQL database, with which the interactions were made and the classifications of each photograph were placed for the appropriate requirement.

Explaining the structure of the SQL database used, 4 tables were created called `Requirement`, `Photo`, `Evaluation` and `Timestamp`. The Timestamp table is formed by 4 columns, `id_timestamp`, `start_time`, `first_requirement_time` and `second_requirement_time`. The `id_timestamp` is the table's primary key. The `start_time` is the time the classifier started the classification. The `first_requirement_time` is the time at which the classifier finished classifying the first 50 images and the

`second_requirement_time` is the time at which it finished classifying the second 50 images.

The Photo table has 2 columns, `id_photo` and `url_photo`. This table contains the number of the photo in the database and its name, since the photos themselves have not been saved in SQL.

The Requirement table contains the `id_requirement`, `requirement_question`, `reverse_binary`, `right_way`, `wrong_way` and `requirement_def` columns. This table, as the name suggests, is used to store the different requirements that define an institutional photograph; the `id_requirement` indicates the number of the requirement in the table. The `requirement_question` has the question that will be presented to the classifier for it to understand whether the image satisfies the question or not. The `reverse_binary` column is a column for binary values, which helps to place all the requirements with the same logic in terms of whether the image meets the requirement or not. For example, for requirement 1, "Does the image have good quality?", the YES button means that the requirement is complied with, but for requirement 7, "Is the person wearing sunglasses?", the same YES button means that the image is not complied with. As such, this binary value sets all the requirements to follow the same logic and correctly classify the images. The `right_way` and `wrong_way` columns contain the url for images that exemplify what should be labeled as YES and NO, respectively, for each requirement. Finally, the `requirement_def` column contains the definition of the requirement itself, to be subsequently placed in the graphical part presented to the classifiers.

The Evaluation table is a combination of the Photo and Requirement tables. It is made up of 7 columns, the primary key id, two columns that are foreign keys, the `photo_id` and `requirement_id` columns, referring to the `id_photo` columns of the Photo table and the `id_requirement` column of the Requirement table respectively. Finally, the `FirstEvaluation`, `SecondEvaluation`, `ThirdEvaluation` and `FinalEvaluation` columns. All of these are initialized to -1. This table combines all the photos in the Photo table with each of the requirements defined in the Requirement table.

It is worth noting that in the implementation, it was always taken into account that only photos with a `FinalEvaluation` column of -1 would appear to the classifiers, and when the number of unclassified photos for a given requirement is less than 50, only the remaining ones are shown. The same logic applies to requirements: when all the photos of a requirement have already been labeled, this requirement will no longer appear to be classified.

Once the code development phase was complete, it was time to make the tool open to classifiers. This process was carried out on a controlled network, where only ISCTE students could be classifiers, requiring authentication through the faculty's centralized Identity and Access Management (IAM) system, via Single Sign-On (SSO), to ensure that only those belonging to ISCTE could classify, since the images are private. Below, in Figure 7, is an example of what a classifier could see and be asked to classify:

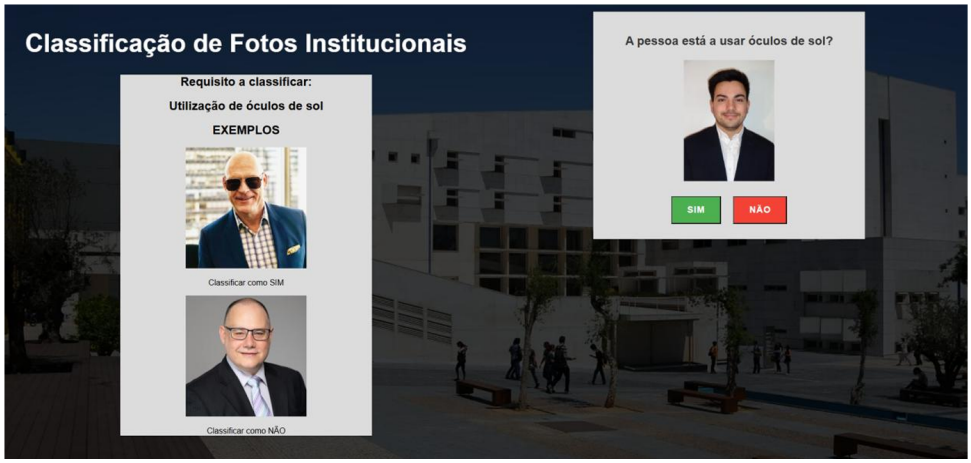


FIGURE 7. Example of question and interface of the tool.

The sorter was only available to the public for 2 days. In that time, a total of 369 rankings were completed (only get the classification record if all 100 photos have been labeled, if any classifiers have left their labeling halfway through, they don't end up in this record). Of all these classifications, there were 262 authentications, i.e. 262 different people took part in the labeling.

The statistics from this labeling show that 27312 images received at least one classification. Of these images, 18655 received only a first classification. 7938 images got their final label. Of these final images, the distribution by requirement are represented in Table 5:

Requirement	Final Classification
1	605
2	683
3	431
4	751
5	499
6	799
7	1174
8	923
9	729
10	528
11	816

TABLE 5. Distribution of Images with final classification by requirement.

Another interesting piece of data taken from the classifications is the images that had a mixed opinion, where the first and second labels were different. There were a total of 331 such ratings. The distribution by requirement can be seen in Table 6.

Requirement	Final Classification
1	23
2	42
3	40
4	30
5	34
6	43
7	7
8	17
9	17
10	56
11	0

TABLE 6. Distribution of images with 2 different labels by requirement.

In the end, no significant conclusions can be drawn from these labels, as there were very few instances of indecision in the labeling and all of them ended up being relatively balanced. Of particular note are requirements 7, 8 and 9, which are easier to identify and therefore cause less indecision. Requirement 11, as will be seen below, turns out not to be very present in the dataset, so there were no cases of indecision.

It turns out that there are 50 034 images remaining to be labeled. These classifications were made by a smaller group of classifiers. Although this is quite a large number of classifications, it was necessary due to the way it was implemented. The final results obtained can be seen in Table 7.

Requirement	Complient	Non-Complient
1	2524	2375
2	4521	378
3	3814	1085
4	4782	117
5	4228	671
6	2466	2433
7	4890	9
8	4779	120
9	4890	9
10	4105	794
11	0	4899

TABLE 7. Distribution of labeled images by requirement.

It is quite clear that requirements 1, 3, and 6 are the most balanced requirements, with a large unbalance in the rest. Requirements 7, 9 and 11 have a fairly small number of cases in which the requirement is non-compliant, which makes sense given that these images were taken from the university's database and, although it was known that some images didn't make much sense in an institutional context, it was expected that there would be a small number. Requirement 11 needs special attention, as there is only one image in which a person is wearing a burqa, and in this case they are wearing it correctly. It is important that this requirement has the presence of people wearing the burqa correctly

Requirement	Compliant	Non-Compliant
1	2778	2375
2	4630	523
3	4055	1098
4	5033	120
5	4380	773
6	2629	2524
7	5057	96
8	5002	151
9	5048	105
10	4209	944
11	65	5088

TABLE 8. Final distribution of Images by requirement.

and incorrectly to realize that the requirement is not whether or not to wear the burqa, but how the burqa is worn.

So, in order to make the above-mentioned requirements more balanced, it was decided to add some images focused on the most lacking cases of these requirements. A total of 255 photos were added, all of them labeled for the 11 requirements according to the implementation logic of the previous photos. The dataset final distribution can be seen in Table 8.

After completing the labeling and beginning the analysis of the dataset to divide it into training and validation sets, it was determined that the question defining requirement 11 was poorly formulated. In the ISO/ICAO standards, the relevant question is phrased as "Veil over face," meaning that if there is a veil covering the face, the image does not meet the requirement; otherwise, it does. However, for the dissertation, the question was framed as "Wearing the veil appropriately," which led to a different interpretation: images were only considered compliant if the person was wearing the veil correctly. If the person was not wearing a veil or was wearing it incorrectly, the image was marked as non-compliant. This inconsistency created confusion and led to unreliable image classification. As a result, it was decided to remove this requirement from the analysis.

3.3.2. Train and Validation Dataset

Once the training set had been defined, it was decided to separate it and create a validation set. The split was 80%/20%, but not with the data being chosen at random. It was taken into account that each of the requirements needs to have both labels in the appropriate proportions in the validation set. Since there are requirements with a very small number of occurrences, the separation could not be random. Thus, the division between the training set and the validation is represented in Table 9.

Private Dataset				
Requirement	Train		Validation	
	Compliant	Non-Compliant	Compliant	Non-Compliant
1	2220	1898	558	477
2	3704	414	926	109
3	3224	894	831	204
4	4022	96	1011	24
5	3502	616	878	157
6	2096	2022	533	502
7	4045	73	1012	23
8	4002	116	1000	35
9	4023	95	1025	10
10	3355	763	854	181

TABLE 9. Private dataset division in Train and Validation datasets.

3.4. Public Dataset

After defining our training and validation dataset, we requested access to the training and validation datasets used to develop ICAONet [**3**]. For the purpose of this study, this dataset was named the Public dataset, an adaptation for the ICAONet network name. This dataset is based on the built-in image dataset for benchmarking of the ISO/IEC 19794-5 standard from the FICV competition, but with some modifications. The dataset was thus composed of a combination of public datasets, specifically the AFW [**49**], AR, FRGC, and PUT databases. In addition, images were added that were taken with a focus on ISO/ICAO requirements. These new additions to the dataset came from requests to hundreds of volunteers to reproduce characteristics that do not comply with the requirements. These images were also annotated for label assignment by three specialized individuals to complete the task. Thus, the dataset is composed of 5763 images distributed as follows:

- 22 images from AFW database (size from 362 × 362 to 1984 × 1984 pixels);
- 1368 images from AR database;
- 50 images from PUT database;
- 1772 images from FRGC database;
- 2551 newly acquired images of size (976 × 1301 to 4272 × 2848 pixels)

A distinctive feature of this dataset is that its images are labeled with 3 different values:

- 0 when the requirement is not met
- 1 when the requirement is met
- -1 for cases where, in the image being analyzed, the requirement being evaluated is not explicit about whether it is being met or not. In these cases, -1 represents a Dummy classification

The dataset we received consists of 5211 training images and 571 validation images. While the original dataset was labeled according to all 23 ISO/ICAO requirements [**4**], our specific dataset only covers 10 of these requirements. Furthermore, only 7 of these 10

Public Dataset						
Requirement	Train			Validation		
	Compliant	Non-Compliant	Dummy	Compliant	Non-Compliant	Dummy
2	3644	588	979	433	31	106
4	4481	648	82	523	12	36
6	3740	1465	6	335	155	81
7	4594	617	0	538	31	2
8	4262	945	4	478	84	9
9	4441	768	2	538	32	1
10	3690	1425	96	476	4	91

TABLE 10. Public Dataset requirements distribution for Train and Validation datasets.

requirements aim to classify the same aspects as those described in the original 23 requirements. These correspond to our requirements 2, 4, 6, 7, 8, 9, and 10. The distribution of images in the received dataset is represented in Table 10.

Therefore, we decided to use only these 7 matching requirements, including only images labeled as 1 or 0, discarding cases with -1 label per requirement.

3.5. Comparison of Datasets

This section aims to highlight the differences between the two datasets used in this study and to analyze how their distinct characteristics may influence the results obtained.

Private dataset

The Private Dataset utilizes images sourced from a real-world usage context, specifically from a university database. This dataset is characterized by a high degree of heterogeneity, as the images were not carefully curated or captured with the ISO/ICAO requirements in mind. Consequently, it is also imbalanced, with a very low percentage of non-compliant images for certain requirements, which is an expected outcome given the collection methodology. The dataset's heterogeneity is a key feature, as images may have additional characteristics beyond the 23 official ISO/ICAO requirements.

Furthermore, this dataset was labeled by students, professors, and university staff through gamification. It is worth noting that these individuals were not specialists in the subject. The classification labels used were exclusively Compliant and Non-Compliant. As a result, images with ambiguous or unclear characteristics were included in the training and validation phases, which could potentially impact model performance.

Public dataset

In contrast, the Public dataset possesses distinct and more carefully curated characteristics. It was created by combining existing datasets specifically designed for training models to detect these requirements. It also includes images captured specifically for this purpose, with an emphasis on collecting both compliant images and those that intentionally violate the criteria.

For the new images added by the study's members [**3**], labeling was performed by subject-matter experts. The labels assigned were Compliant, Non-Compliant, and Dummy, though for our study, only the Compliant and Non-Compliant images for each requirement were used. This approach ensured that the training dataset only included images where the characteristics were clearly met or not met, effectively removing any ambiguity and excluding images where the presence of a characteristic was uncertain.

3.6. Modeling

Once the image dataset has been defined with the appropriate labels, it's time to start the modeling phase, where the models will be trained in order to identify the previously defined requirements and obtain the best possible metrics.

3.6.1. Pre-processing

Resize of the images

Given the significant diversity in image dimensions across both the training and validation sets, it was determined that dataset preprocessing would be necessary. The combined datasets contain a total of 1399 different dimensions. The 6 most common dimensions are represented in Table 11.

Dimensions	Number of images
100x100	448
3200x2400	382
640x480	333
180x180	260
75x100	189
160x200	117

TABLE 11. Private dataset 6 most common dimensions.

As evident from these 6 most common dimensions, there is a mix of square images, images with greater length than width, and images with greater width than length. The YOLO model, one of the models that will be used for modeling, only processes images with equal length and width. For this reason, it was decided that all models would follow the same image preprocessing approach. Two dimensions were chosen for image resizing: 256x256 pixels and 640x640 pixels.

The initial approach considered for this adjustment was to perform a standard resize operation, setting both length and width to the desired dimensions. However, this method



FIGURE 8. Original size (194x259).



FIGURE 9. Resize without cropping.



FIGURE 10. Cropped Image.

would distort the image quality and potentially result in loss of important content. In Figures 8 and 9 can be seen a representation of the distortion that an original image would suffer.

Instead, a weighted cropping approach was implemented. For square images, a simple resize operation was sufficient, but for images with unequal dimensions, a different resizing method was applied. The smaller dimension was resized to the target size, while the larger dimension was proportionally resized. From this larger dimension, the excess size relative to the target dimension was calculated, divided by 2, and cropped equally from both sides. For example, consider an image of 100x50 pixels that needs to be resized to 200x200 (using round numbers for easier explanation). The smaller dimension is 50 pixels, which is scaled up to 200. To maintain proportions, the 100-pixel dimension scales up to 400. At this point, we have an image of 400x200 pixels. The difference between the current width (400) and the target width (200) is 200 pixels. Therefore, 100 pixels are cropped from both the right and left sides of the image. By using this approach, the central portion of

the image is preserved, which in passport-type photos is typically where the person's face is located - the most crucial element for identification purposes. In Figure 10, there is a representation of the final results of the same image at 256x256 pixels.

Cropped images

Another notable factor was that in some images, people were positioned too far from the camera, with their entire body visible. In reality, the torso, arms, hands, and other body parts do not contribute to person identification. Moreover, when the face is too distant, certain identification requirements become compromised and more difficult to assess. Therefore, it was decided to crop the images to primarily retain the facial region, discarding other less informative body parts.

To achieve this, OpenCV was utilized, implementing code previously used in another dissertation [**9**], to make the necessary modifications to the dataset images. The code implements face detection algorithms, followed by eye detection and center point calculation. Finally, it calculates the distance between the two eye centers.

Minimum and maximum thresholds were established for the inter-eye distance, more specifically, 70 pixels and 150 pixels. In cases where this distance was very small, indicating that the person was too far away, the image was cropped to include only the face and its immediate border, effectively zooming in on the facial region. Conversely, when the inter-eye distance exceeded the maximum threshold, it indicated that the person was too close to the camera. An example of the procedure applied to an image is represented in figures 11 and 12.



FIGURE 11. Original Image.



FIGURE 12. Image with cropped preprocessing applied.

This change was applied to both the images in the Private dataset and the images in the Public Dataset, and these images were then used for the remaining 2 datasets, explained above. The number of changed images is represented in Table 12.

The technique was applied to all images, and in some cases, eye detection was unsuccessful. This occurred for various reasons: the person in the image had closed eyes, was wearing sunglasses, or their eyes were too narrowed to be detected. These images were retained since they represent other requirements being evaluated. There were also cases where the distance exceeded the threshold, though these were few enough to allow

Dataset	Number of images whose distance was greater than the threshold	Number of images whose distance was less than the threshold
Private Dataset	227	0
Public Dataset	337	23

TABLE 12. Number of images that suffered Cropping preprocessment.

individual assessment of each case to determine whether the person's identity was still recognizable despite their proximity to the camera, and consequently decide whether to keep them in the dataset or not. This process led to a new iteration and a new dataset, which was applicable to all 4 cases from the previously described datasets. In essence, this created 4 new datasets for training purposes.

3.6.2. Baseline Models

In order to find baseline metrics using our dataset, it was decided that it would be interesting to train with pre-trained models in order to understand the model's performance in comparison with the best available results, defining a starting point. From this point, we will have a comparison for future models, better understanding the possible evolutions obtained. Two pre-trained models were chosen to serve as a baseline, VGG-16 model and ResNet50 model, as previously mentioned.

One of the questions that arose was the choice of how to train the models. The issue centers on the fact that the study that ends up having the closest relationship with what is intended to be done, the ICAONet network [**3**], gives an output of 23 requirements at once. Thus, the question is whether to choose to train 10 requirements in a single model or to train 10 different models, one per requirement. The decision made was to train one model per requirement, justified by the fact that the other models to be used in the study, WNN and YOLO, can only be trained for 1 model per requirement, so it is the approach that makes the most sense for making the comparison.

Moreover, by choosing this approach, each feature's performance is expected to be better than if all features were trained together in a single model. This enables the creation of more specialized models focused on each characteristic, which can be individually improved if necessary. Furthermore, this method allows new features to be added in the future without impacting the results of the already trained models.

So, one model by requirement was trained for each of the pre-trained models. In both models, Early Stopping [**50**] was used with a patience of 5, so that the model didn't go into overfitting [**51**] or underfitting. For both problems, models were trained with 256x256 pixel and 640x640 pixel images. In addition to the Private dataset, the 3 datasets described previously were also trained, both for the original images and with the preprocessing of cropping and face close-ups when applicable, for both sizes.

For the moment of making image predictions using the trained models, both the VGG model and the ResNet model only return a confidence level for the image. Thus, it was necessary to find a value that would serve as a reference to know from which value it

should be considered as label 0, and which should be label 1. For this, the EER metric was used. With this value found and used as a threshold, it was possible to assign labels to the images, thus creating the confusion matrix and the respective metrics.

All the metrics obtained, whether from the baseline models or the models that will be addressed, will be presented in the Results section. As mentioned in chapter 2.4, the aim is to understand the behavior of models such as Weightless Neural Networks (WNN) and YOLO.

The choice of these models is intended to understand their performance in detecting facial features compared to the best results already obtained by other companies or researchers since no records were found of the use of any of the models in this context. For both, one model per requirement is trained, then the results will be compared with the baseline metrics of the second problem.

3.6.3. WNN

In addition to the model training process itself, one of the most important phases in modeling the WNNs was the preprocessing of the images. This pre-processing was based on images from the MNIST dataset, the dataset used as the basis for the study [**41**]. Pre-processing consisted of loading the grayscale of each of the images from both the training set and the validation set, resizing the images 256x256, using the crop method, as explained before.

The preprocessing pipeline applied to the grayscale images employs a sophisticated approach known as thermometer encoding [**52**], specifically tailored for weightless neural network training. Initially, each image undergoes conversion to grayscale, normalization to the range [0,1], and flattening into a one-dimensional array of 65,536 features. The system then calculates mean and standard deviation across all training images, using these statistics to standardize both training and validation sets, ensuring consistent data distribution. At the core of this process lies the thermometer encoding technique, which creates multiple binary thresholds based on Gaussian distribution principles, generating either 8, 12, or 16 separate binary vectors for each pixel in the original image. We use these 3 values in order to understand how they influence the results. These different bit-per-input configurations represent varying approaches to the dimensionality-representation power trade-off. With 8 bits per input, the process generates 524,288 binary features, offering a computationally efficient but potentially less expressive representation. The 12-bit configuration produces 786,432 features, striking a balance between computational requirements and representation capacity. Meanwhile, the 16-bit approach creates over one million binary features (1,048,576), providing the highest representation power at the cost of increased computational demands.

After completing all these necessary preprocessing steps, the 4 previously described datasets were trained and validated. Both the original photos and the cropped image datasets were processed.

In the phase of using the trained models to make predictions of the images, this model has 3 possible outputs, the labels 0, 1 and '0 and 1'. This means that in cases where the model predicts '0 and 1', it was unable to predict any of the labels, therefore the image is considered as undefined.

3.6.4. YOLO

Both YOLOv8 and YOLO11 models were used in classification mode for all requirements. All trained models used Early Stopping [**50**], with a patience of 5 epochs. Models were trained for both 256x256 pixel and 640x640 pixel dimensions.

In addition to normal training, the models were also tuned for both dimensions in order to see the effect on the metrics. The dropout technique [**33**] was applied, one with a dropout of 25% of the training dataset and the other with a dropout of 50%. This technique was chosen because the training dataset of many requirements ends up being quite unbalanced. Training was conducted on all previously defined datasets using three different dropout configurations: no dropout, 25% dropout, and 50% dropout. The same datasets were also trained, for both dimensions, with the three different dropout configurations, but for the datasets with preprocessing of cropping and face close-ups of images when applicable.

For the moment of using the trained models to predict the images, the YOLO model returns a confidence level for each of the labels 0 or 1. Thus, it was decided to consider the prediction with the highest confidence level as the predicted label.

3.7. Trained Models

This section presents all the experiments conducted during the modeling phase. To facilitate understanding of what was intended to be tested, an overview of all relevant information is provided.

Starting with the datasets, it is helpful to assign specific names to each one:

- **Private Dataset:** This dataset consists images provided by the university. These images were later labeled using the developed Image Labeling Tool. This dataset includes 10 requirements.
- **Public Dataset:** This dataset is composed for all the images used for training the ICAONet network [**3**]. This dataset includes 7 requirements.

For each dataset, the images were resized to two different resolutions:

- 256x256 pixels
- 640x640 pixels

For both image sizes across each dataset, a preprocessing step was applied to adjust the position of faces that were too far from to the camera, as this could compromise identity recognition. In this way, the models were trained using two different types of images: images without adjusting the proximity of the face, and images with the proximity of the face adjusted to the camera.

- **Original Images:** Dataset with images that were not processed in any way.
- **Cropped Images:** Dataset in which images where the person is too far from the camera were processed.

For all these variations, we have 4 different models that we want to train for this study:

- **Baseline Models:** Composed for **VGG-16** and **ResNet50**
- **WNN**
- **YOLOv8**
- **YOLO11**

For the YOLO models, in addition to all the variations described previously, further tuning was performed, specifically on the dropout parameter of the training set. As a result, there are YOLO models with all the previously described characteristics but using different dropout values. The distinguishing characteristics were therefore as follows:

- **No Dropout:** The training set was used as-is, without applying any dropout technique.
- **25% Dropout:** Randomly, 25% of the neurons in each classification layer were deactivated during each training batch.
- **50% Dropout:** Randomly, 50% of the neurons in each classification layer were deactivated during each training batch.

In addition to all these models and their combinations, it is important to note that Early Stopping [**50**] with a patience of 5 epochs was applied to all trained models. Beyond examining the different model architectures, this study aimed to identify which dataset can achieve the best results in a production environment, by detecting relevant characteristics in institutional photographs. Furthermore, it aimed to evaluate whether combining the two datasets into a single, unified set would lead to improved results. To this end, the models were trained using **four different datasets**, as shown in Table 13.

Dataset	Description	Requirements
Private Dataset	The training and validation sets composed of images provided by the university, which were classified using a tool developed specifically for this dissertation.	10
Public Dataset	The training and validation sets consisted of the images used to train the ICAONet network.	7
Merged Train with Private Validation	The training set combines the training sets from both the Private and Public datasets, along with the validation set from the Private dataset.	7
Merged Train with Public Validation	The training set combines the training sets from both the Private and Public datasets, along with the validation set from the Public dataset.	7

TABLE 13. Description of the Datasets used for Modeling.

All four datasets were used to train each of the previously described models, across both resolutions, using both preprocessing methods, and with all types of Dropout configurations for the YOLO models.

3.8. Training Average Time

This section aims to present the average time required to train each model. Since multiple combinations of configurations and preprocessing steps were applied to each model, the reported times correspond to the models themselves, without delving into the specific details of each combination. To account for the influence of image size on training speed, the results were first separated by image dimension, as models trained with 640×640-pixel images require more time than those trained with 256×256-pixel images. Accordingly, the approximate training time for each model was determined separately for each image dimension. These averages are presented in Table 14.

Model		Average Training Time	
		256x256	640x640
VGG-16		≈ 1 hour 30 minutes	≈ 4 hours
ResNet50		≈ 1 hour 10 minutes	≈ 10 hours
WNN	8 BPI	≈ 15 mins	-
	12 BPI	≈ 35 mins	
	16 BPI	≈ 1 hour	
YOLOv8		≈ 2 mins	≈ 8 mins
YOLO11		≈ 2 mins	≈ 8 mins

Table 14. Models Average Training Time.

All models were trained using an NVIDIA A100 GPU. Among them, YOLO models were the only ones for which multiprocessing was successfully implemented, which also explains their shorter training times compared to the other models. In theory, WNN architectures should have been the fastest to train. However, due to the unsuccessful implementation of multiprocessing, this expectation was not confirmed in practice.

3.9. Experiments

Once the models were trained, it was necessary to obtain performance metrics in order to draw several conclusions. These include:

- Assessing whether the models used in this study are competitive compared to those presented in other research;
- Comparing the performance of the Baseline Models with the YOLO models and the WNN models;

- Understanding whether computationally lighter models, such as the WNNs, can be competitive compared to other models **(RQ 2)**;
- Evaluating whether the image dataset provided and labeled by the university performs better in identifying features in institutional photographs than the Public dataset when applied to the same model;
- Determining whether combining both training sets into a single dataset leads to improved results.

To address these questions, several experiments were carried out, each with its own specific goal. These experiments will perform predictions on different sets of images, depending on the datasets and combinations used, as will be described later. The experiments are organized into sections and subsections:

- Each **section** defines the type of image used for prediction;
- Each **subsection** specifies the dataset and the training configuration that the models were subjected to within that particular experiment;

3.9.1. Experiment sections and subsections

Setup

To provide brief context on how the experiments were set up, they combine each of the sections to be described with all of their subsections, resulting in experiments A.1, A.2, A.3, B.1, and so on.

Section A- Intra-Dataset Evaluation

For Section A, the predicted images correspond to the validation set used during the training of the respective model. If the model was trained with a validation set from the Public dataset, the predicted images will be from the Public validation set. If it was trained with a validation set from the Private dataset, the predicted images will be from the Private validation set.

Predictions were made on the validation sets used during training to understand whether the models developed in this study achieve performance levels comparable to the reported ones. This approach enables a similar comparison with previous studies, particularly ICAONet [**3**], for which both the training/validation sets were made available, alongside with the results obtained from the prediction of the validation images.

Since our models were designed to assess only 10 requirements — of which just 7 align with the official ISO/ICAO standards — they could not be submitted to the competition nor evaluated using the official test set, which is restricted to models that implement all 23 requirements. Therefore, evaluating on the validation set remains the most viable strategy for performance comparison.

Section B- Cross-Dataset Evaluation

For Section B, the predicted images are those from the validation set **opposite** to the one used during the training of the corresponding model. If the model was trained using a validation set from the Public dataset, then the predicted images will be from the validation set of the Private dataset. Conversely, if the model was trained with a validation set from the Private dataset, the predicted images will be from the validation set of the Public dataset.

One of the objectives of this section is to evaluate whether the models exhibit signs of overfitting [**51**]. Models trained on the validation set of the Private Dataset will make predictions on the validation set of the Public Dataset, and models trained on the validation set of the Public Dataset will make predictions on the validation set of the Private Dataset. Since the predicted images were not used during training, this allows us to evaluate the generalization capability of the models.

Another objective is to identify the models that will be integrated into the API. The models achieving the best performance in experiments using this section will be selected — with the exception of requirements 1, 3, and 5, which are exclusive to the Private Dataset and will therefore be defined based on the best-performing models from Section A.

For this section, only the **YOLOv8** and **YOLO11** models were used. The rationale behind this choice will be explained later.

Subsection 1

The experiments in Subsection 1 focus exclusively on predicting images using models trained with the **Public Dataset**. These experiments include both original and cropped images, resized to 256×256 and 640×640 pixels. All four models are predicting, except in the case of the YOLO models, where only images from models trained with **No Dropout** are used for prediction.

Subsection 2

The experiments in Subsection 2 focus on predicting images using models trained on the **Public Dataset** and on the **Merged Train with Public Validation Dataset**. For the Public Dataset models, predictions are made using YOLO models with **25% Dropout** and **50% Dropout**, considering both preprocessing types and both image dimensions. For the Merged Train with Public Validation dataset, predictions are performed for all four models, also covering both preprocessing types and both image dimensions.

Subsection 3

The experiments in Subsection 3 focus on predicting images using models trained with the Private Validation Set, the **Private Dataset**, and the **Merged Train with Private Validation Dataset**. For both datasets, predictions are performed across both preprocessing types and both image dimensions, using all four models.

3.9.2. Description of Experiments Performed

This subsection outlines the objectives pursued in each of the experiments, as well as the datasets and models used in each case. A summary of the models employed in each experiment will be presented later.

A.1.

The goal of this experiment is to compare the performance of these models with the results obtained for the validation set of the ICAONet network [**3**]. Models are predicting **Public Validation Dataset images** in models trained on the Public Dataset. Predictions are made for both original and cropped images, resized to 256×256 and 640×640 pixels. All four models are predicting, except in the case of the YOLO models, where only images from models trained with No Dropout are used for prediction.

A.2.

This experiment aims to evaluate the influence of dropout on YOLO model performance, as well as whether merging the Private Dataset improves the results. Models are predicting **Public Validation Dataset images** in models trained on the Public Dataset and on the Merged Train with Public Validation Dataset. For the Public Dataset models, predictions are made using YOLO models with 25% Dropout and 50% Dropout, considering both preprocessing types and both image dimensions. For the Merged Train with Public Validation dataset, predictions are performed for all four models, also covering both preprocessing types and both image dimensions.

A.3.

The goal of this experiment is to analyze the metrics on a different validation set and to determine whether the trends observed in experiments A.1 and A.2 also apply to this dataset. Furthermore, the best-performing models for requirements 1, 3, and 5 (which are exclusive to the Private Dataset) will be the ones integrated into the final API. Models are predicting **Private Validation Dataset images** in models trained with the Private dataset and the Merged Train with Private Validation Dataset. For both datasets, predictions are performed across both preprocessing types and both image dimensions, using all four models.

B.1.

The goal of this experiment is to establish baseline results for the model trained with the Public training and validation datasets, without any modifications to the Private validation dataset. This will allow for meaningful comparisons in subsequent experiments and help identify the best approach to achieve optimal results. Models are predicting **Public Validation Dataset images** in models trained on the Private Dataset. Predictions are made for both original and cropped images, resized to 256×256 and 640×640 pixels. All

four models are predicting, except in the case of the YOLO models, where only images from models trained with **No Dropout** are used for prediction.

B.2.

The goal is to compare the results with those obtained in experiment B.1, in order to understand whether merging the Private Dataset improves or worsens the results, as well as to assess the influence of dropout in YOLO model performance. Models are predicting **Private Validation Dataset images** in models trained on the Public Dataset and on the Merged Train with Public Validation Dataset. For the Public Dataset models, predictions are made using YOLO models with 25% Dropout and 50% Dropout, considering both preprocessing types and both image dimensions. For the Merged Train with Public Validation dataset, predictions are performed for all four models, also covering both preprocessing types and both image dimensions.

B.3.

This experiment aims to compare models trained only on the Private Dataset with those trained on the merged dataset in order to understand how Public dataset images influence the results. Also to evaluate the influence of image dimensions, preprocessing, and dropout on the final performance of YOLO models. Models are predicting **Public Validation Dataset images** in models trained with the Private Dataset and the Merged Train with Private Validation Dataset. For both datasets, predictions are performed across both preprocessing types and both image dimensions, using all four models.

3.9.3. Experiments Summary

This section aims to summarize each of the experiments conducted, in order to help clarify which datasets were used in each case. For the experiments, three training sets were used:

- **Train Private** - Images from the training set of the Private Dataset
- **Train Public** - Images from the training set of the Public Dataset
- **Train Merged** - Combination of images from the training set of the Private Dataset and the Public Dataset

And two validation datasets:

- **Validation Private** - Images from the validation set of the Private Dataset
- **Validation Public** - Images from the validation set of the Public Dataset

A summary of the Experiments of Section A- Intra-Dataset Evaluation- is represented in Table 15.

Experiment	Train Dataset	Validation Dataset	Predicted Images	Models Used	Image Size and Preprocessing
A.1.	Train Public	Validation Public	Validation Public	Baseline Models; WNN; Both YOLO models without Dropout	Sizes: 256x256 640x640 Preprocessing: Original Cropped
A.2.	Train Public	Validation Public	Validation Public	Both YOLO models with 25% and 50% Dropout;	Sizes: 256x256 640x640 Preprocessing: Original Cropped
	Train Merged	Validation Public	Validation Public	Baseline Models; WNN; Both YOLO models;	
A.3.	Train Private	Validation Private	Validation Private	Baseline Models; WNN; Both YOLO models;	Sizes: 256x256 640x640 Preprocessing: Original Cropped
	Train Merged	Validation Private	Validation Private		

TABLE 15. Summary of Experiments of Section A.

And a summary of the Experiments of Section B- Cross-Dataset Evaluation- is represented in Table 16.

Experiment	Train Dataset	Validation Dataset	Predicted Images	Models Used	Image Size and Preprocessing
B.1.	Train Public	Validation Public	Validation Public	Best Models of A Experiments	Sizes: 256x256 640x640 Preprocessing: Original Cropped
B.2.	Train Public	Validation Public	Validation Public	Best Models of A Experiments	Sizes: 256x256 640x640 Preprocessing: Original Cropped
	Train Merged	Validation Public	Validation Public		
B.3.	Train Private	Validation Private	Validation Private	Best Models of A Experiments	Sizes: 256x256 640x640 Preprocessing: Original Cropped
	Train Merged	Validation Private	Validation Private		

TABLE 16. Summary of Experiments of Section B.

CHAPTER 4

Results and Discussion

This chapter will present the results obtained from each Experiment. Due to the high number of trained models, metrics will be presented only for the most relevant ones. All the models trained as part of this dissertation can be found in https://tinyurl.com/Institutional-photo, with specific information on the training phase of each model.

It is important to note that throughout this chapter, as well as in the Appendices, some tables include the symbol "–". This symbol indicates that, for the corresponding column, different model configurations achieved the same metric value. For instance, in the Preprocessing column, the symbol "–" means that both models trained with the original images and those trained with cropped images reached the same results. Similarly, in the Size column, it signifies that models trained with both 256-pixel and 640-pixel inputs obtained identical metrics.

4.1. Criteria for Choosing the Best Metric per Requirement

Since it is necessary to choose the best model for each requirement, it is necessary to select a metric to make this definition. Thus, the metrics chosen to define the performance of the models were Specificity and F1-Score.

Specificity was chosen because it is very important that the models, in addition to properly detecting cases where images meet a certain requirement, also detect when that requirement is not being met. With requirements datasets that end up being quite unbalanced, with many more cases of images that meet the requirement than those that don't, it is important that these non-compliant cases are detected. Additionally, these are the key cases that should not be approved in the university, since the problem is that images that should not be approved are being approved.

The F1-Score was chosen because it provides a balance between precision and recall, combining both metrics into a single value. This is particularly valuable in imbalanced datasets, as it takes into account both false positives and false negatives. Since the requirements have varying levels of imbalance, the F1-Score helps ensure that models are not just achieving high accuracy by always predicting the majority class, but are actually performing well at identifying both compliant and non-compliant images.

In cases where, in order to determine the best model for a given requirement, one model achieved higher Specificity while another achieved a higher F1-Score, the absolute selection metric was defined as the average of both. Although these metrics assess different aspects of performance, they are equally important, which justifies the adoption of this

criterion. This approach helps identify the model(s) that offer the best overall balance between the two metrics.

4.2. Section A- Intra-Dataset Evaluation

As previously described, the validation set results are important for comparing the performance between the ICAONet model study [**3**] and this study, to understand whether they are performing in similar ways at this stage or not. We can see the results of the best models for the Public validation dataset without changes in the training set, Public validation dataset with changes in the training set, and Private validation dataset.

4.2.1. Experiment A.1.

The goal of this experiment is to compare the model's metrics when predicting on the validation set of the Public dataset, without any processing applied to this set except for cropping the images as needed, with the results obtained for the validation set of the ICAONet network [**3**]. The results that the ICAONet network obtained for the validation set are presented in Table 17.

ICAONet Network validation set metrics	
Metric	Value (%)
Accuracy	94.27
Precision	94.53
Recall	97.89
F-measure	96.15
F-beta	97.14
NPV	91.67
Specificity	81.69
MCC	82.78

TABLE 17. Results obtained by ICAONet network validation dataset.

This table was obtained by predicting each image for all 23 requirements and creating a confusion matrix with all these predicted values and their respective labels. The main difference is that these results address all 23 requirements, while our study only covers 7 of them.

Since it is not possible to understand directly, using the same test set as the one used for the metrics of the ICAONet network test set [**3**], the aim was to understand the performance of the models, and to determine if they were capable of obtaining metrics similar to those obtained by the study. Therefore, the same images from the study were used. The results of the best model per requirement for Experiment A.1. are represented in Table 18:

For this experiment, the baseline models, the WNN networks, and the YOLO models (YOLOv8 and YOLO11) were used. The detailed metrics obtained for each model can be found in Appendix B.

Requirement	Model	Pre-processing	Size	Specificity	F1-Measure
2	YOLOv8	Cropped	640	0.906	0.996
4	YOLOv8	Crooped	640	0.833	0.995
6	YOLO 11	Original	640	1.000	0.993
7	YOLOs and Baseline	-	-	0.967	0.999
8	YOLOv8	Original	256	0.964	0.989
9	YOLOv8	Cropped	640	1.000	1.000
10	-	-	-	0.500	0.998

TABLE 18. Experiment A.1. best models by requirement.

The best result of the Baseline Models demonstrated significant difficulties with requirement 2, where the specificity was 0, meaning it was unable to identify whether a person was looking directly at the camera or not. Requirements 4 and 10, which identify whether a person has hair in front of their eyes and whether the person's posture is correct, also achieved poor metrics. The remaining requirements obtained interesting results for a baseline model, but fell short of the best results. It is worth noting that for requirement 7, which identifies whether a person is wearing sunglasses or not, achieved results equal to those obtained by the best model for this requirement.

The WNN model, despite modeling with different BPIs and using the bleaching technique that yielded the best results during the training phase, failed to demonstrate any specificity when applied to the validation dataset. Its predictions consistently indicated that all images met the given requirement, or it predicted '0 and 1', indicating that it was unable to assign a proper label to the image.

Regarding the YOLO models, both YOLOv8 and YOLO11 achieved quite satisfactory results in their predictions on the validation set. In Table 19, a comparison of the metrics from both models was made to understand the differences between them.

YOLO models difference						
Requirement	Specificity			F1-Measure		
	YOLOv8	YOLO 11	Difference	YOLOv8	YOLO 11	Difference
2	**0.906**	0.813	0.093	**0.996**	0.989	0.007
4	**0.833**	0.417	0.416	**0.995**	0.992	0.003
6	0.994	**1.000**	0.006	**0.995**	0.993	0.002
7	**0.967**	**0.967**	0.000	**0.999**	**0.999**	0.000
8	**0.964**	0.952	0.012	0.989	**0.994**	0.005
9	**1.000**	**1.000**	0.000	**1.000**	0.999	0.001
10	**0.500**	**0.500**	0.000	**0.998**	**0.998**	0.000

TABLE 19. Comparation of YOLO models metrics for Public dataset results without train dataset changes.

With the results obtained, it is evident that both YOLO models demonstrate remarkably similar performance. The requirement where the most significant difference in metrics is observed is requirement 4, which verifies whether a person has hair covering their eyes or not. There appears to be no clear justification for poorer performance specifically for

this requirement when using the same training and validation images. Therefore, with the best models per requirement for the Public training and validation dataset represented in Table 18, it is important to make a comparison to understand whether the best models per requirement are at the same level as the ICAONet network or not. This comparison is presented in Table 20.

Metric	Model	
	ICAONet	Study best models by req
Accuracy	94.27	**99.26**
Precision	94.53	**99.27**
Recall	97.89	**99.26**
F-measure	96.15	**99.27**
F-beta	97.14	**99.27**
NPV	91.67	**95.49**
Specificity	81.69	**96.86**
MCC	82.78	**95.77**

TABLE 20. Comparation between ICAONet and this studies metrics for the same training and validation datasets.

Despite these results not being directly comparable, since one model predicts 23 requirements while the other has a specific model for only 7 requirements, the findings indicate that the best models from this study perform at the same level or even better when predicting validation images. With such impressive results, concerns arise regarding whether these images are in an overfitting state [**51**] or not. This question is clarified when testing the models with a test dataset.

4.2.2. Experiment A.2.

This Experiment aims to understand the influence, in terms of metrics, of using a dataset subjected to techniques different from the original ones, such as a dataset composed of more images, or the use of dropout technique on the initial images. Prections were made for the same Public validation set, but using different training sets. The results of the best model per requirement for Experiment A.2. are represented in Table 21:

Requirement	Model	Dataset	Pre-processing	Size	Dropout (%)	Specificity	F1-Measure
2	YOLOv8	Merged	Original	640	0	0.906	0.994
4	YOLOv8	Public	Original	256	50	0.583	0.987
6	YOLOv8	Public	Both	640	50	0.994	0.995
7	YOLOs and Baseline	-	-	-	-	0.967	0.999
8	Both YOLO versions	-	-	-	-	0.976	0.995
9	YOLOv8	-	Cropped	-	-	1.000	1.000
10	YOLOv8	Merged	Original	256	0	1.000	0.995

TABLE 21. Experiment A.2. best models by requirement.

For this experiment, the baseline models, the WNN networks, and the YOLO models (YOLOv8 and YOLO11) were used. The detailed metrics obtained for each model can be found in Appendix B.

Looking at the Baseline Models, for both the original dataset and the modified dataset, the model that achieved better results between VGG and ResNet was the ResNet model, which predominated with the exception of requirement 7 in the modified training dataset, where the VGG model obtained better results. No significant changes were observed between the use of different datasets. The most notable difference was in terms of Specificity for requirement 1, but despite the improvement with the modified training dataset, the metric remained poor. It is worth highlighting that requirement 7 achieved the best recorded result for this requirement among all models for both datasets.

The results of the WNN network continued to be quite negative, with the predictions still unable to detect cases where images do not meet the requirement, which is ultimately a very important metric for this study. It is concluded that the issue is not with the dataset itself, but rather with parameterization that may not have been correct. For this study, the model proves to be inadequate for identifying characteristics in passport-type images.

The predictions of the YOLOv8 and YOLO11 models using a different training dataset are the models that obtained the best results among the models that made predictions for the Public validation datasets with alterations in the training set, as was previously observed in the predictions for the original dataset. The difference in metrics when using a training dataset with alterations or not ultimately does not have much impact, with quite similar metrics, except for 2 requirements: requirement 4 for the YOLOv8 model and requirement 10 for both YOLO models. This same difference in metrics is evident in Table 22, with the comparison of the best overall results between both datasets.

Best Models by requirement comparation						
Requirement	Specificity			F1-Measure		
	A.1.	A.2.	Difference	A.1.	A.2.	Difference
2	**0.906**	**0.906**	0.000	**0.996**	0.994	0.002
4	**0.833**	0.583	0.250	**0.995**	0.987	0.008
6	**1.000**	0.994	0.006	0.993	**0.994**	0.001
7	**0.967**	**0.967**	0.000	**0.999**	**0.999**	0.000
8	0.964	**0.976**	0.012	0.989	**0.995**	0.006
9	**1.000**	**1.000**	0.000	**1.000**	**1.000**	0.000
10	0.500	**1.000**	0.500	**0.998**	0.995	0.003

TABLE 22. Comparation of metrics between best models by requirement of A.1. and A.2. Experiment.

The models that made the predictions of the best models were all predicted using the YOLOv8 model, as demonstrated in Table 21, which obtained slightly superior metrics compared to the YOLO11 model, and where the metrics were equal for requirements 7 and 8.

The main conclusions that can be drawn from the metrics of using an original dataset compared to a dataset with applied alterations were that, for requirement 4, the model with the original dataset ended up having a better metric, with an exact value of 0.250. In this case, the model trained with the altered training dataset that obtained the best results was using the same dataset, but with the 50% dropout technique applied to it.

The other requirement where differences in metrics are noticeable, more specifically in Specificity, was in requirement 10, for both YOLOv8 and YOLO11. Both YOLO models obtained a Specificity of 0.500 with the model trained on the original dataset, whereas using the altered training dataset, the Specificity for YOLO11 was 0.750 and for YOLOv8 it was 1.000. The dataset that obtained these metrics for both YOLO models was the dataset that combines training sets from the Private and Public datasets. With this information, it can be concluded that, for the validation dataset, the images from the Private dataset help to better understand whether a person's posture is correct or not.

4.2.3. Experiment A.3.

This Experiment proves to be important, as it will make it possible to understand the results obtained in requirements 1, 3, and 5, which were not evaluated in the previous results sections, as they are unique requirements of this study, and to understand whether the validation dataset of the Private dataset responds well to images from the training set of both processed datasets.

The results of the best model per requirement for Experiment A.3. are represented in Table 23:

Requirement	Model	Dataset	Pre-processing	Size	Dropout (%)	Specificity	F1-Measure
1	YOLOv8	Private	Original	256	25	0.943	0.965
2	YOLOv8	Private	Original	256	50	0.339	0.958
3	YOLOv8	Private	Cropped	256	0	0.569	0.926
4	YOLOv8	Private	Cropped	256	50	0.292	0.989
5	YOLO 11	Private	Original	256	50	0.701	0.945
6	YOLOv8	Private	Cropped	256	0	0.964	0.959
7	YOLOv8	Private	Cropped	256	25	1.000	0.997
8	YOLOv8	Merged	Original	640	50	0.429	0.988
9	YOLO 11	Merged	Original	640	50	0.600	0.997
10	YOLOv8	Private	Original	256	0	0.691	0.946

TABLE 23. Experiment A.3. best models by requirement.

For this experiment, the baseline models, the WNN networks, and the YOLO models (YOLOv8 and YOLO11) were used. The detailed metrics obtained for each model can be found in Appendix B. From what can be seen in Table 23, the YOLOv8 model ends up being the model that obtains better metrics compared to the YOLO11 model, which only managed to be better for requirements 5 and 9. Furthermore, for the vast majority of requirements, the dataset that obtains the best results is the Private dataset alone, and in only 2 requirements, 8 and 9, the dataset with better results was the one combining the training datasets from the Private and Public datasets.

4.3. Section B- Cross-Validation Dataset

In this section, all models validated with the Public validation dataset will make predictions on the Private validation dataset and vice versa. This approach will make it possible to understand how the models behave when predicting images that were not included in their validation set, and to determine whether the best models from the validation phase will be the same as those from the test set phase. Additionally, it will help identify which models demonstrate the best overall performance to determine which models will be featured in the API. Regarding requirements 1, 3, and 5, the best results in the validation phase will be the models transferred to the final API, as there is no dataset capable of testing them.

It is also important to note that predictions for the Experiments of this Section were only made for **YOLO models**. The baseline models served their purpose by providing reference results and helping to understand how pretrained models without any tuning performed, serving as a point of comparison for both the WNN and YOLO models. Additionally, the fact that both the VGG and ResNet models consistently underperformed compared to other models contributed to this decision. As for the WNN models, they were excluded because they failed to make reliable predictions, even on images from their own validation set — a scenario where overfitting is actually expected. This suggested that WNNs were not suitable for identifying characteristics in images.

Furthermore, YOLO models not only achieved better performance but were also the fastest to train. Considering that Section B focuses on models intended for integration into the API, training efficiency was also taken into account, further reinforcing the choice to prioritise YOLO models in these experiments.

4.3.1. Experiment B.1.

The goal of this experiment is to establish baseline results for the model trained with the Public training and validation datasets, without any modifications to the Private validation dataset. The results of the best model per requirement for Experiment B.1. are represented in Table 24:

Best of both YOLO models					
Requirement	Model	Pre-processing	Size	Specificity	F1-Measure
2	YOLOv8	Original	640	0.266	0.827
4	YOLOv8	Cropped	640	0.500	0.938
6	YOLOv8	Cropped	256	0.841	0.889
7	YOLOv8	Original	640	0.826	0.997
8	YOLOv8	Original	256	0.743	0.960
9	YOLOv8	Cropped	640	0.300	0.996
10	YOLO11	Original	256	0.425	0.874

Table 24. Experiment B.1. best models by requirement.

For this experiment, the baseline models, the WNN networks, and the YOLO models (YOLOv8 and YOLO11) were used. The detailed metrics obtained for each model can be found in Appendix C. Before analyzing the obtained results, it is important to understand the differences between the YOLOv8 and YOLO11 metrics for Experiment B.1., represented in Table 25.

YOLO models difference						
Requirement	Specificity			F1-Measure		
	YOLOv8	YOLO 11	Difference	YOLOv8	YOLO 11	Difference
2	**0.266**	0.184	0.082	0.827	**0.887**	0.060
4	**0.500**	0.292	0.208	0.938	**0.976**	0.038
6	0.841	**0.883**	0.042	**0.889**	0.825	0.064
7	**0.826**	0.696	0.130	**0.997**	0.996	0.001
8	**0.743**	0.514	0.229	**0.960**	0.959	0.001
9	**0.300**	**0.300**	0.000	**0.996**	0.992	0.004
10	0.293	**0.425**	0.132	**0.896**	0.874	0.022

TABLE 25. Comparation of YOLO models metrics for models of Experiment B.1..

Analysis of the updated comparative data between YOLOv8 and YOLO 11 models reveals distinct performance patterns across the requirements. YOLOv8 demonstrated superior specificity in requirements 2, 4, 7, 8, and tied in requirement 9, with particularly substantial advantages in requirements 4, 7, and 8 (differences of 0.208, 0.130, and 0.229 respectively). YOLO 11 exhibited better specificity only in requirements 6 and 10, with requirement 10 showing the more notable difference (0.132). When analyzing these values alongside those from the following experiments, it becomes evident that low Specificity, even when associated with a high F1-Score, indicates that the models are predicting all images as Compliant. Since the classes Compliant and Non-Compliant are unbalanced, most predictions turn out to be correct, as the majority of images belong to the Compliant class. However, in the few cases where the requirement under evaluation was not met, the models were unable to correctly identify this situation.

Regarding F1-Measure, YOLOv8 outperformed YOLO 11 across all requirements except requirement 7, where YOLO 11 marginally excelled. The F1-Measure differences, while generally small, were most pronounced in requirement 6 (0.064) and requirement 2 (0.060). Both models maintained strong F1-Measure performance. Overall, YOLOv8 appears to deliver more consistent and generally superior results when considering both metrics holistically, with its advantages in specificity being particularly noteworthy across multiple requirements.

Regarding the values themselves, they primarily demonstrate much lower specificities compared to the values obtained in the validation set, which indicates overfitting of the model.

4.3.2. Experiment B.2.

This Experiment goal is to compare the results with those obtained in experiment B.1, in order to understand whether merging the Private Dataset improves or worsens the results, as well as to assess the influence of dropout in YOLO model performance. The results of the best model per requirement for Experiment B.1. are represented in Table 26:

Best of both YOLO models							
Requirement	Model	Dataset	Pre-processing	Size	Dropout (%)	Specificity	F1-Measure
2	YOLOv8	Merged	Original	640	25	0.312	0.947
4	YOLOv8	Public	Cropped	256	50	0.375	0.963
6	YOLO11	Public	Cropped	256	50	0.958	0.955
7	YOLO11	Public	Cropped	256	50	0.913	0.997
8	YOLOv8	Public	Cropped	640	50	0.543	0.968
9	YOLO11	Public	Cropped	640	25	0.600	0.986
10	YOLOv8	Merged	Original	256	0	0.613	0.929

TABLE 26. Experiment B.2. best models by requirement.

For this experiment, the baseline models, the WNN networks, and the YOLO models (YOLOv8 and YOLO11) were used. The detailed metrics obtained for each model can be found in Appendix C. Before analyzing the obtained results, it is important to understand the differences between the YOLOv8 and YOLO11 metrics, represented in Table 27.

YOLO models difference						
Requirement	Specificity			F1-Measure		
	YOLOv8	YOLO 11	Difference	YOLOv8	YOLO 11	Difference
2	**0.312**	0.257	0.055	**0.947**	**0.947**	0.000
4	**0.375**	0.250	0.125	0.963	**0.988**	0.025
6	0.952	**0.958**	0.006	**0.958**	0.955	0.003
7	0.869	**0.913**	0.044	**0.998**	0.997	0.001
8	**0.543**	0.457	0.086	0.968	**0.981**	0.013
9	**0.600**	**0.600**	0.000	0.977	**0.986**	0.009
10	**0.613**	0.585	0.028	0.929	**0.931**	0.002

TABLE 27. Comparation of YOLO models metrics for models of Experiment B.2..

Comparing the performance of YOLOv8 and YOLO 11 models across the requirements reveals distinct patterns in their effectiveness. YOLOv8 demonstrated superior specificity in requirements 2, 4, 7, and 10, with the most notable advantages appearing in requirements 4 and 2 (differences of 0.125 and 0.055 respectively). Conversely, YOLO 11 exhibited better specificity in requirement 8 (0.457 versus 0.343) while maintaining identical performance for requirement 9. Regarding F1-Measure, the differences between models were minimal across all requirements, with the largest gap being only 0.025 in requirement 4. Both models achieved consistently high F1-Measure scores exceeding 0.90 in most cases, indicating strong overall classification performance. In general terms, YOLOv8 appears to offer marginally better results when considering both metrics collectively.

Looking at Table 26 with the best results of the models trained with the Public validation dataset that predicted the Test Set, it was expected that models trained with the combination of training sets from both the Private and Public dataset would achieve the best results, since the test set consists of images similar to those in the Private dataset. In reality, the approach that generated the best results for the modified training datasets was using only images from the Public dataset with some type of dropout, with 5 of the 7 requirements falling under these conditions.

From the perspective of metrics between the best models by requirement without modification of the training dataset and models with changes in the training dataset, these can be observed in Table 28.

Best Models by requirement comparation						
Requirement	Specificity			F1-Measure		
	B.1.	B.2.	Difference	B.1.	B.2.	Difference
2	0.266	**0.312**	0.046	0.827	**0.947**	0.120
4	**0.500**	0.375	0.125	0.938	**0.963**	0.025
6	0.841	**0.958**	0.117	0.889	**0.955**	0.066
7	0.826	**0.913**	0.087	**0.997**	**0.997**	0.000
8	**0.743**	0.543	0.200	0.960	**0.968**	0.008
9	0.300	**0.600**	0.300	**0.996**	0.986	0.010
10	0.425	**0.613**	0.188	0.874	**0.929**	0.055

TABLE 28. Comparation of metrics between best models by requirement of B.1. and B.2. Experiment

Analyzing the table, an improvement in the Specificity metric is observed in 5 of the 7 requirements. The most notable improvements are observed in requirements 9 and 10. On the other hand, using the dataset without any type of alterations, the Specificity results for both requirement 4 and requirement 8—more specifically, in verifying whether a person has hair covering their eyes or not and verifying whether a person's eyeglass lenses are causing reflection or not, respectively—were significantly better than in the dataset with alterations. Regarding the F1-Measure metric, it is observed that all requirements except for 9, and considering that 7 has equal metrics, show an improvement in the metric with the use of a dataset.

In this way, it is concluded that the Private dataset ultimately contains images that do not add the necessary information to improve the metrics. Only 2 of the 7 requirements improve their metrics when using the combination of the 2 training datasets.

4.3.3. Experiment B.3.

This experiment aims to compare models trained only on the Private Dataset with those trained on the merged dataset in order to understand how Public images influence the results. Also to evaluate the influence of image dimensions, preprocessing, and dropout on the final performance of YOLO models.

The results of the best model per requirement for Experiment B.1. are represented in Table 29:

Best of both YOLO models							
Requirement	Model	Dataset	Pre-processing	Size	Dropout (%)	Specificity	F1-Measure
2	YOLOv8	Merged	Original	640	50	0.656	0.987
4	YOLOv8	Merged	Cropped	256	25	0.417	0.991
6	YOLO11	Merged	Cropped	256	50	1.000	0.974
7	-	-	-	-	-	0.967	0.999
8	YOLOv8	Merged	Cropped	640	50	0.952	0.993
9	YOLOv8	Merged	Original	640	25	1.000	1.000
10	YOLO11	Merged	Original	640	0	0.750	0.998

TABLE 29. Experiment B.3. best models by requirement.

For this experiment, the baseline models, the WNN networks, and the YOLO models (YOLOv8 and YOLO11) were used. The detailed metrics obtained for each model can be found in Appendix C. Before analyzing the obtained results, it is important to understand the differences between the YOLOv8 and YOLO11 metrics, represented in Table 30.

YOLO models difference						
Requirement	Specificity			F1-Measure		
	YOLOv8	YOLO 11	Difference	YOLOv8	YOLO 11	Difference
2	**0.656**	**0.656**	0.000	**0.987**	0.986	0.001
4	**0.417**	0.083	0.334	**0.991**	0.986	0.005
6	0.987	**1.000**	0.013	**0.983**	0.974	0.009
7	**0.967**	**0.967**	0.000	**0.999**	**0.999**	0.000
8	**0.952**	0.941	0.011	**0.993**	**0.993**	0.000
9	**1.000**	0.968	0.032	**1.000**	0.997	0.003
10	**0.750**	**0.750**	0.000	0.996	**0.998**	0.002

TABLE 30. Comparation of YOLO models metrics for models of Experiment B.3..

Based on the comparative analysis of YOLOv8 and YOLO11 models across different requirements, several performance patterns emerge. YOLOv8 demonstrated superior specificity in requirements 4 and 9, with a particularly notable difference of 0.334 in requirement 4. Conversely, YOLO11 achieved better specificity in requirement 6. For requirements 2, 7, 8, and 10, both models exhibited identical or nearly identical specificity values, indicating comparable performance in these aspects.

Regarding F1-Measure, the differences between the two models are minimal across all requirements, with the largest difference being only 0.009 in requirement 6. This suggests that both models maintain similar overall classification performance when considering both precision and recall. YOLOv8 shows marginally better F1-Measure values in most requirements, though these differences are not substantial enough to indicate a clear advantage.

Observing the results, extremely satisfactory metrics are verified for requirements 6, 7, 8, 9, and 10, especially for the first 4, with 2 of them even reaching a Specificity of 1.0.

On the other hand, analyzing Table 29, which demonstrates the best models by requirement among models trained with the Private dataset with and without alterations, and whose models made predictions on the Public validation dataset, it is noted that

all the best models result from a training dataset that consists of a combination of the Private dataset training set with the Public dataset training set.

This fact ends up coinciding with the results obtained in Experiment B.2.. It is understood that the images from the Private dataset do not have sufficient quality, and that they ultimately require the images from the Public dataset to obtain quality metrics, while conversely, the images from the Public dataset end up achieving better results by themselves, and do not improve metrics when trained with the Private dataset. Furthermore, with the test dataset containing images similar to those in the Public dataset, the influence that the Public images had on the results was verified, while the opposite was not observed.

In addition, an experiment was conducted to investigate whether the metrics obtained in Experiment B.3 were influenced by the images from the Public dataset. For this purpose, predictions were generated on the Public validation set using models trained exclusively with images from the Private dataset. The results are presented in Table 31.

Best Initial results without ICAO dataset						
Requirement	Model	Pre-processing	Size	Dropout (%)	Specificity	F1-Measure
2	YOLO11	Cropped	640	25	0.125	0.966
4	YOLOv8	Original	256	25	0.083	0.987
6	YOLO11	Cropped	256	25	0.819	0.947
7	YOLOv8	Original	256	0	0.967	0.999
8	YOLOv8	Original	256	25	0.547	0.961
9	YOLOv8	Original	256	50	0.344	0.952
10	YOLOv8	Original	640	25	0.500	0.988

TABLE 31. Best models per requirement using Private dataset training images only and Public dataset validation images

We can observe the comparison of the values obtained in Table 31 with the best results presented in Table 29 in Table 32.

Requirement	Specificity			F1-Measure		
	Initial	B.3.	Difference	Initial	B.3.	Difference
2	0.125	**0.656**	0.531	0.966	**0.987**	0.021
4	0.083	**0.417**	0.334	0.987	**0.991**	0.004
6	0.819	**1.000**	0.181	0.947	**0.974**	0.027
7	**0.967**	**0.967**	0.000	**0.999**	**0.999**	0.000
8	0.547	**0.952**	0.405	0.961	**0.993**	0.032
9	0.344	**1.000**	0.656	0.952	**1.000**	0.048
10	0.500	**0.750**	0.250	0.988	**0.998**	0.010

TABLE 32. Comparison of the best results from models trained on the Private dataset only versus the best models from Experiment B.3, evaluated on the Public validation set

As suspected, all requirements achieved better results with the presence of images from the Public dataset. Significant differences in Specificity, such as those for requirement 9

of 0.656, or requirement 2 of 0.531, are representative of the impact that the presence or absence of the dataset makes on the metrics.

Therefore, Experiment B.3. is particularly significant, as it underscores the influence of the differing characteristics of the two datasets. The Public dataset consists of more homogeneous images, each specifically meeting the requirements of the study. It is a carefully prepared dataset designed to reduce features that are difficult to identify and optimized for the study. In contrast, the Private dataset consists of a more heterogeneous collection of real-world images, captured without a study-specific focus. This comparison indicates that the Public dataset is more suitable for model training and feature learning, while the Private dataset, despite having assigned labels, presents greater variability that may hinder the learning process.

4.4. Summary of Best Results

Considering the results and discussion, it was concluded that, within the context of this study, the more homogeneous and carefully prepared dataset—containing images specifically tailored to the requirements under investigation—was more effective for learning the targeted features than the more heterogeneous dataset, which consisted of images from an institutional context but lacked careful attention to requirement identification. Consequently, it is concluded that the models to be integrated into the final API should be those trained with the Public dataset's validation set, namely the best model by requirement from Experiments B.1 or B.2. This approach ensures greater generalization for images that will be analyzed in the future. An exception is made for requirements 1, 3, and 5, which were only present in the experiments of Section A. For these three specific requirements, the best-performing model is obtained in Experiment A.3. Consequently, these models are derived from those trained and validated using the Private dataset. The models that will be implemented in the API are represented in Table 33 and are the following:

Requirement	Model	Dataset	Pre-processing	Size	Dropout (%)
1	YOLOv8	Private	Original	256	25
2	YOLOv8	Merged	Original	640	25
3	YOLOv8	Private	Cropped	256	0
4	YOLOv8	Public	Cropped	640	0
5	YOLO11	Private	Original	256	50
6	YOLO11	Public	Cropped	256	50
7	YOLO11	Public	Cropped	256	50
8	YOLOv8	Public	Original	256	0
9	YOLO11	Public	Cropped	640	25
10	YOLOv8	Merged	Original	256	0

TABLE 33. General best model per requirement that are in the final API.

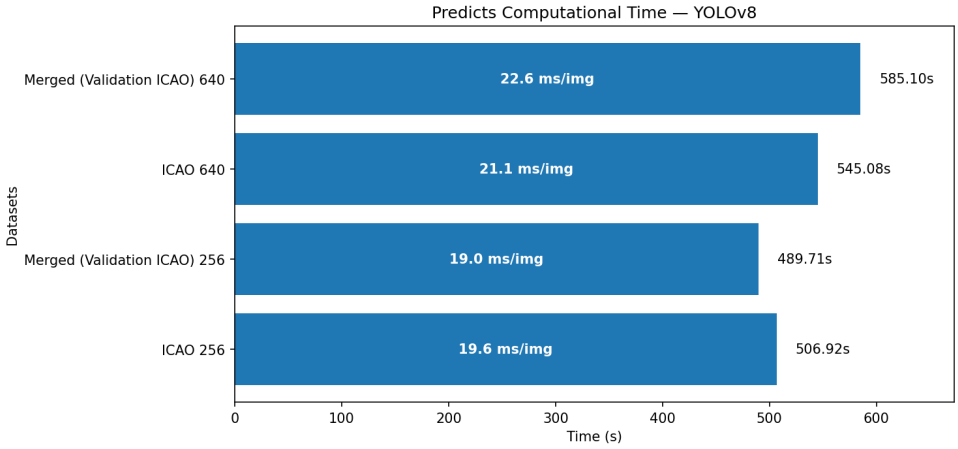


FIGURE 13. Average Prediction time results for YOLOv8 (requirement 2).

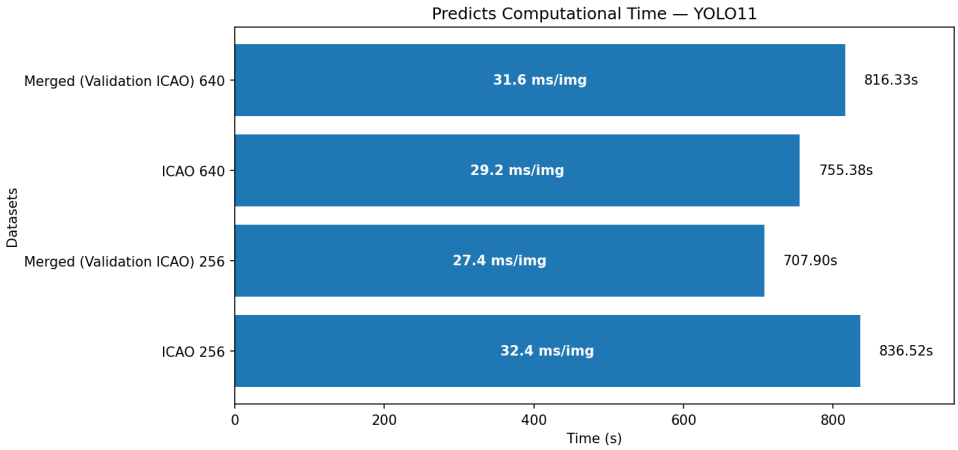


FIGURE 14. Average Prediction time results for YOLO11 (requirement 2).

4.5. Average Image Prediction Time

After analyzing all the experiments and selecting the models to be integrated into the API, an important question arises: are there significant differences in using images of 256×256 pixels instead of 640×640 pixels? If predicting a requirement from 256-pixel images is considerably faster, and given that the API is expected to provide quick responses, would it be more appropriate to adopt models that offer better computational performance?

To address this, we evaluated both the YOLOv8 and YOLO11 models using both image resolutions. The tests were carried out on models trained with the Public dataset as well as on models trained with the Merged dataset (consisting of the original training set merged with ICAO's training set) and validated against the Public validation set. The goal was to measure not only the total prediction time across all images but also the average time required to predict a single image.

For this purpose, the Public training dataset was expanded 5 times, resulting in a total of 25 835 images. These images were predicted three times in order to mitigate external factors such as caching effects and to ensure results more consistent with a production environment. The correctness of the predictions was not relevant here, as the

performance metrics of the best models had already been established. The sole objective was to determine how long the models take to generate predictions and whether image size significantly impacts API performance. The results reported correspond to the average of the three runs for each model. For all models, requirement 2 was used as the prediction target.

Figures 13 and 14 present the average total prediction time for the 25 835 images, as well as the average time per image, for the YOLOv8 and YOLO11 models, respectively.

In summary, the results show that the choice between 256×256 and 640×640 input resolutions does not lead to a significant impact on computational performance. Although there are slight differences in prediction times (typically within 2–3 ms per image), such variations are negligible in practice and would not noticeably affect the API's response time. Therefore, reducing all images to 256×256 pixels is not justified from a computational perspective, and the decision regarding image size should instead be guided by accuracy and compliance considerations rather than speed.

CHAPTER 5

API Development

In this chapter, we will explore the system developed for the university, which aims to verify whether the characteristics that define an institutional image, as established in this dissertation, are present or not. This tool serves as an aid to those responsible for validating these images, simplifying the verification process and ensuring consistency in image validation.

5.1. API Purpose

The university welcomes thousands of new students each year, and each student needs to submit a photo for institutional purposes. The traditional approach of having these photos manually validated by human operators is not only time-consuming but also lacks standardized criteria. This system has been developed to evaluate photos according to the institutional image criteria defined in this dissertation. When the system identifies that a particular criterion is not met, it allows the human validator to focus their attention specifically on that aspect, making the verification process more efficient and targeted. This approach doesn't eliminate human oversight but rather enhances it by directing attention to potential issues that require careful consideration.

To serve different needs, a single, unified application was built using the Django framework. This application provides two distinct interfaces to access the validation logic: a programmatic API for automated systems and a web-based graphical user interface (GUI) for human operators.

5.2. Implementation Baseline

In both interfaces, the image is passed through the same core logic and the appropriate prediction is made for each of the requirements. Just as for obtaining metrics for the validation set and for the test set, the YOLO models have an output that initially returns the requirement that they consider should be the final label, along with the degree of confidence of the prediction, and then delivers the other label, with the degree of confidence of that same one. Thus, this first label is considered the predicted label and is used as the result.

It is important to emphasize that the system includes a requirement that was not present in the training phase in the same way as the other requirements. A Requirement 11 was added, which checks the proportions of the person's face relative to the rest of the image. In cases where the image is too far from or too close to the camera, it will be detected and indicated as a requirement not met. This requirement was used for preprocessing cropping, with code already efficiently implemented by OpenCV, and did

not require conventional training like the other requirements. Thus, when an image is analysed, it is first determined whether the face is at a correct distance, and then the original image is passed to be predicted by each of the models to obtain the final result.

5.3. Application Architecture

The entire application was developed using the Django framework. This architecture allows a single, robust backend to serve both programmatic and interactive user needs. At its core is a processing pipeline that is triggered regardless of whether a request comes from the API or the GUI.

The processing pipeline involves several steps:

- **Decoding:** For API requests, the Base64-encoded image is decoded and converted into an image object that can be manipulated in memory. For GUI requests, the uploaded image file is read directly.
- **Resizing:** Two resized versions of the image are generated—one with a resolution of 256x256 pixels, and another with 640x640 pixels. These versions serve as inputs for the best model by requirement that have been trained with corresponding image sizes.
- **Face Analysis (Requirement 11):** A preliminary check is performed to determine whether the face in the image is at an appropriate distance. Based on the face's size relative to the entire image, it is determined whether cropping is needed. If the face is too close or too distant, this step is flagged as unmet (req 11 = 0). Otherwise, the original image is considered adequate (req 11 = 1).
- **Model Inference:** Each of the ten trained YOLO models (for requirements 1 through 10) is loaded and applied to the appropriate resized image. The model performs object detection, and the first detected label is used as the predicted result for that requirement. The label is extracted directly from the prediction output and transformed into binary format: 1 if the criterion is met, 0 otherwise.

5.3.1. Programmatic Interface (API)

For automated integration, the Django application exposes a dedicated HTTP endpoint at /api/. This interface was designed to receive image data and return the classification results in a machine-readable format. The API expects requests to be made via the POST method and accepts a JSON payload containing a single field named `image_base64`, which must contain the image encoded in Base64 format [**53**].

Upon receiving a valid request, the image is passed through the processing pipeline described above. Once all predictions are completed, the API returns a structured JSON object with the classification results. Each field (req 1 to req 11) contains a binary value (0 or 1) indicating whether the image satisfies the corresponding institutional requirement.

An example of a response from the API is as follows:

```
{"req_1": 1,"req_2": 0,"req_3": 1,"req_4": 1,"req_5": 1,"req_6": 0,
"req_7": 1,"req_8": 1,"req_9": 1,"req_10": 1,"req_11": 1}
```

This JSON format facilitates easy integration with other systems, including the university's web platform or any internal validation interface, enabling automated pre-screening of submitted images.

5.3.2. Graphical User Interface (GUI)

For human validators, the Django application provides a user-friendly and visually structured web interface designed to make the validation process intuitive and efficient.

The interface consists of three main elements: a form for uploading the image, a left-hand panel listing all the requirements being assessed, and a right-hand panel that displays the classification results. The user is initially prompted to select an image file and click the "Evaluate" button. During the evaluation, a dynamic loading animation is shown to provide visual feedback.

When an image is uploaded via the GUI, it is sent to the backend and processed by the exact same pipeline used by the API. Once the evaluation is complete, the system renders the results on the web page, presenting the uploaded image alongside the outcome for each of the eleven requirements. Visual feedback is provided through intuitive symbols—a green check mark (✓) for fulfilled requirements and a red cross (✗) for those not met—enhancing clarity and immediate comprehension.

At the end of the evaluation, users can perform a new assessment by uploading another image without needing to reload the page. This seamless interaction supports the continuous validation of multiple photographs in a single session, bridging the gap between automatic analysis and manual review.

In summary, the graphical interface complements the back-end processing by providing an accessible and informative visualization of the classification results, bridging the gap between automatic analysis and manual review in a highly efficient way.

We can observe an example of the GUI classifying the 11 requirements for an image in Figure 15.

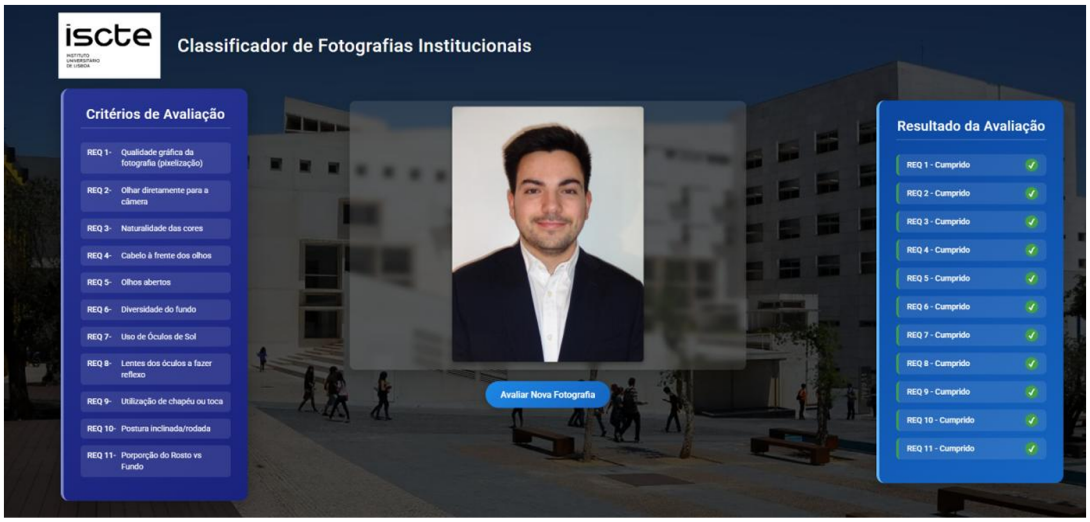


FIGURE 15. Example of final GUI classifying an image.

5.4. Considerations

This section outlines the key measures that should be implemented before deploying the developed GUI and API into a production environment.

Before deployment, it is crucial to address specific aspects related to security and access control. The following measures are recommended:

- Ensure that all communication protocols use HTTPS rather than HTTP, to guarantee encrypted data transmission;
- Implement authentication mechanisms based on the university's access credentials for both the GUI and the API.

Adopting these measures will strengthen the overall security of the system, safeguarding data integrity and protecting access to both the API and the GUI during operational use.

CHAPTER 6

Conclusions and Future Work

This dissertation addressed the persistent challenge of automated verification of institutional photo requirements in academic environments. Unlike the standardized domain of passport control (ISO/ICAO), institutional photography lacks a universal definition, leading to subjective manual verification processes that are prone to inconsistency.

By establishing a specific classification of visual requirements and implementing a modular deep learning architecture, this study demonstrated that automated validation is not only feasible but essential. The experimental results confirm that while "institutional" criteria vary, they can be rigorously defined and effectively monitored using modern computer vision techniques. The following sections synthesize the key findings regarding data quality, model architecture, and the necessity of requirement-specific optimization.

6.1. The Critical Role of Data Quality

A pivotal finding of this research challenges the common assumption that training on "real-world" data always yields the best deployment performance. The comparative analysis between models trained on the Private Dataset (real-world, heterogeneous, noisy labels) and the Public Dataset (expert-curated, balanced, homogeneous) revealed a decisive conclusion.

Models trained on the Public dataset consistently achieved higher generalization performance, even when tested against the validation set of the Private dataset. The strict adherence to balanced class distributions and expert-verified labels in the Public set eliminated the "noise" found in the university's raw database. This demonstrates that for institutional photo verification, data quality and label precision are more critical than domain proximity. Training a model on a cleaner and strictly defined dataset produces a more robust classifier than training on a noisy dataset derived from the actual production environment. Consequently, the final API prioritizes models trained on the ICAO standard to ensure consistent, objective decision-making.

6.2. Model Performance and Architecture

The comparative analysis of computational models provided decisive answers to the research questions regarding performance and efficiency.

6.2.1. Deep Learning Dominance (RQ3)

The experimental results definitively identify the YOLO architecture as the superior approach for this domain. Both YOLOv8 and YOLO11 consistently outperformed baseline CNNs and Weightless Neural Networks (WNNs). While YOLOv11 is the more recent

iteration, it demonstrated no significant performance advantage over YOLOv8. Additionally, computational latency tests revealed that increasing input resolution from 256×256 to 640×640 results in a negligible processing penalty (2–3 ms), validating the feasibility of high-resolution analysis in real-time.

6.2.2. Limitations of Lightweight Models (RQ2)

A critical finding is the inadequacy of Weightless Neural Networks (WNNs) for this specific application. Despite their theoretical advantages in low-resource environments, WNNs failed to achieve competitive accuracy compared to CNNs. This suggests that the nuanced aesthetic criteria of institutional photos (e.g., specific posture, background uniformity) require the complex feature extraction capabilities inherent in deep learning models, rendering WNNs unsuitable for this specific task under current constraints.

6.3. Requirement-Specific Optimization

One of the most significant conclusions of this study is that there is no "one-size-fits-all" configuration for the verification of institutional photo requirements. The results demonstrate that optimal training parameters are highly dependent on the specific visual requirement being analyzed.

- **Preprocessing Strategy:** The choice between "cropped" (face-centric) and "original" (full image) preprocessing varies logically. Requirements focusing on facial features (e.g., Requirement 2, 7, 8) performed best with cropped images, while context-dependent requirements (e.g., background uniformity) required the full original image.
- **Regularization (Dropout):** The application of Dropout yielded inconclusive results across different requirements. While some requirements benefited from Dropout during training, showing improved performance, others demonstrated better results without it. No clear pattern emerged to predict when Dropout would be beneficial, suggesting that its effectiveness may be highly dependent on the specific characteristics of each individual requirement's data distribution and model architecture.
- **Input Resolution:** Similarly, the choice between 256×256 and 640×640 pixels yielded different results per requirement, with no single resolution dominating the leaderboard.

This finding validates the modular architecture proposed in this dissertation. Rather than deploying a monolithic single model, the final system integrates a specific, optimized model for each individual requirement—selecting the exact combination of resolution, preprocessing, and architecture that yielded the highest metric for that specific task.

6.4. Methodological Limitations

While the proposed solution successfully met the study's objectives, certain limitations must be acknowledged. The reliance on non-expert annotators for the Private dataset

introduced noise into the training labels. The labeling tool, while functional, lacked a batch-processing feature, which led to an uneven distribution of labels across certain requirements. This randomness meant that some images were heavily reviewed while others received fewer annotations, potentially affecting the "ground truth" reliability for ambiguous cases. Additionally, the aggressive preprocessing required to standardize the highly heterogeneous real-world data may have introduced artifacts in edge cases.

6.5. Future Work

Building upon the architecture established in this dissertation, several avenues for future research are proposed:

- **Threshold Optimization:** The current API accepts predictions based on the highest confidence score. Future work should investigate dynamic confidence thresholds (e.g., implementing a "gray zone" between 0.45 and 0.55) to flag ambiguous images for human review rather than forcing a binary automated decision.
- **Integration of Generative AI Agents:** A significant advancement would be moving beyond binary classification to semantic interaction. Integrating Large Language Models (LLMs) as reasoning agents could allow the system to explain *why* a photo was rejected (e.g., "The background contains cluttered objects"). Such an agent could interpret the numeric outputs of the YOLO models and generate user-friendly feedback, significantly improving the user experience.
- **Continuous Learning Pipelines:** Given the parity between YOLOv8 and YOLOv11, future iterations should focus less on model architecture and more on automated MLOps workflows. These pipelines could periodically retrain models as new "corner cases" are discovered in the university's intake process, ensuring the system evolves alongside institutional policies.

Bibliography

[1] R. Parente, L. V. Batista, I. L. Andrezza, E. V. Borges, and R. A. Mota, "Assessing facial image accordance to iso/icao requirements," in *2016 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, pp. 180–187, IEEE, 2016.

[2] A. Nourbakhsh, M.-S. Moin, and A. Sharifi, "Facial images quality assessment based on iso/icao standard compliance estimation by hmax model," *Journal of Information Systems and Telecommunication*, vol. 7, no. 27, pp. 225–237, 2009.

[3] A. G. d. A. e Silva, H. M. Gomes, and L. V. Batista, "A collaborative deep multitask learning network for face image compliance to iso/iec 19794-5 standard," *Expert Systems with Applications*, vol. 198, p. 116756, 2022.

[4] Biometric System Laboratory, University of Bologna, "Benchmark area: Face image iso compliance verification." Website, 2023. Accessed via BioLab.CSR.Unibo.it.

[5] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio, "Robust object recognition with cortex-like mechanisms," *IEEE transactions on pattern analysis and machine intelligence*, vol. 29, no. 3, pp. 411–426, 2007.

[6] M. Ferrara, A. Franco, D. Maio, and D. Maltoni, "Face image conformance to iso/icao standards in machine readable travel documents," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 4, pp. 1204–1213, 2012.

[7] "Information technology - biometric data interchange formats - part 5: Face image data / amendment 1: Conditions for taking photographs or face image data," Technical Report ISO/IEC 19794-5:2005/Amd 1, International Organization for Standardization (ISO), 2007. Retrieved from the Biometrics Institute.

[8] "Machine readable travel documents (mrtds) - part 1: General specifications," Technical Report Doc 9303, Part 1, International Civil Aviation Organization (ICAO), 2021. Retrieved from ICAO.

[9] J. M. C. Ruivo, "Deep neural networks for image quality: a comparison study for identification photos," Master's thesis, ISCTE-Instituto Universitario de Lisboa (Portugal), 2018.

[10] P. Cunningham, M. Cord, and S. J. Delany, "Supervised learning," in *Machine learning techniques for multimedia: case studies on organization and retrieval*, pp. 21–49, Springer, 2008.

[11] E. D. Mekler, F. Brühlmann, K. Opwis, and A. N. Tuch, "Disassembling gamification: the effects of points and meaning on user motivation and performance," in *CHI'13 extended abstracts on human factors in computing systems*, pp. 1137–1142, 2013.

[12] R. C. Gonzales and P. Wintz, *Digital image processing*. Addison-Wesley Longman Publishing Co., Inc., 1987.

[13] A. K. Jain, *Fundamentals of digital image processing*. Prentice-Hall, Inc., 1989.

[14] H. Liu, W. Gao, J. Miao, and J. Li, "A novel method to compensate variety of illumination in face detection.," in *JCIS*, pp. 692–695, 2002.

[15] J. Canny, "A computational approach to edge detection," *IEEE Transactions on pattern analysis and machine intelligence*, no. 6, pp. 679–698, 1986.

[16] J. F. Groote, R. Morel, J. Schmaltz, and A. Watkins, *Logic Gates, Circuits, Processors, Compilers and Computers*. Springer, 2021.

[17] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*, vol. 1. MIT press Cambridge, 2016.

[18] A. Zhang, Z. C. Lipton, M. Li, and A. J. Smola, *Dive into deep learning*. Cambridge University Press, 2023.

[19] Y.-c. Wu and J.-w. Feng, "Development and application of artificial neural network," *Wireless Personal Communications*, vol. 102, pp. 1645–1656, 2018.

[20] D. Svozil, V. Kvasnicka, and J. Pospichal, "Introduction to multi-layer feed-forward neural networks," *Chemometrics and intelligent laboratory systems*, vol. 39, no. 1, pp. 43–62, 1997.

[21] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A survey of convolutional neural networks: analysis, applications, and prospects," *IEEE transactions on neural networks and learning systems*, vol. 33, no. 12, pp. 6999–7019, 2021.

[22] M. Krichen, "Convolutional neural networks: A survey," *Computers*, vol. 12, no. 8, p. 151, 2023.

[23] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai, *et al.*, "Recent advances in convolutional neural networks," *Pattern recognition*, vol. 77, pp. 354–377, 2018.

[24] M. Jones and P. Viola, "Face recognition using boosted local features," in *IEEE international conference on computer vision*, vol. 278, 2003.

[25] J. R. Movellan, "Tutorial on gabor filters," *Open source document*, vol. 40, pp. 1–23, 2002.

[26] V. Christlein, L. Spranger, M. Seuret, A. Nicolaou, P. Král, and A. Maier, "Deep generalized max pooling," in *2019 International conference on document analysis and recognition (ICDAR)*, pp. 1090–1096, IEEE, 2019.

[27] M. A. Chandra and S. Bedi, "Survey on svm and their application in image classification," *International Journal of Information Technology*, vol. 13, no. 5, pp. 1–11, 2021.

[28] D. Bank, N. Koenigstein, and R. Giryes, "Autoencoders," *Machine learning for data science handbook: data mining and knowledge discovery handbook*, pp. 353–374, 2023.

[29] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International conference on machine learning*, pp. 448–456, pmlr, 2015.

[30] S. Sharma, S. Sharma, and A. Athaiya, "Activation functions in neural networks," *Towards Data Sci*, vol. 6, no. 12, pp. 310–316, 2017.

[31] H. Gao, H. Yuan, Z. Wang, and S. Ji, "Pixel transposed convolutional networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 42, no. 5, pp. 1218–1227, 2019.

[32] D. Yu, H. Wang, P. Chen, and Z. Wei, "Mixed pooling for convolutional neural networks," in *Rough Sets and Knowledge Technology: 9th International Conference, RSKT 2014, Shanghai, China, October 24-26, 2014, Proceedings 9*, pp. 364–375, Springer, 2014.

[33] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[34] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[35] D. Albashish, R. Al-Sayyed, A. Abdullah, M. H. Ryalat, and N. A. Almansour, "Deep cnn model based on vgg16 for breast cancer classification," in *2021 International conference on information technology (ICIT)*, pp. 805–810, IEEE, 2021.

[36] M. A. Raza, H. U. R. Siddiqui, A. A. Saleem, K. Zafar, A. Zafar, S. Dudley, and M. Iqbal, "Advanced feature extraction for cervical cancer image classification: Integrating neural feature extraction and autoint models," *Sensors*, vol. 25, no. 9, p. 2826, 2025.

[37] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[38] M. Talele and R. Jain, "A comparative analysis of cnns and resnet50 for facial emotion recognition," *Engineering, Technology & Applied Science Research*, vol. 15, no. 2, pp. 20693–20701, 2025.

[39] N. V. Sridharan, J. V. Joseph, S. Vaithiyanathan, and M. Aghaei, "Weightless neural network-based detection and diagnosis of visual faults in photovoltaic modules," *Energies*, vol. 16, no. 15, p. 5824, 2023.

[40] I. Aleksander, W. Thomas, and P. Bowden, "Wisard· a radical step forward in image recognition," *Sensor review*, vol. 4, no. 3, pp. 120–124, 1984.

[41] Z. Susskind, A. Arora, I. D. Miranda, L. A. Villon, R. F. Katopodis, L. S. De Araújo, D. L. Dutra, P. M. Lima, F. M. França, M. Breternitz Jr, *et al.*, "Weightless neural networks for efficient edge inference," in *Proceedings of the International Conference on Parallel Architectures and Compilation Techniques*, pp. 279–290, 2022.

[42] F. Sultana, A. Sufian, and P. Dutta, "A review of object detection models based on convolutional neural network," *Intelligent computing: image processing based applications*, pp. 1–16, 2020.

[43] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779–788, 2016.

[44] P. Jiang, D. Ergu, F. Liu, Y. Cai, and B. Ma, "A review of yolo algorithm developments," *Procedia computer science*, vol. 199, pp. 1066–1073, 2022.

[45] J. Terven, D.-M. Córdova-Esparza, and J.-A. Romero-González, "A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas," *Machine learning and knowledge extraction*, vol. 5, no. 4, pp. 1680–1716, 2023.

[46] N. Jegham, C. Y. Koh, M. Abdelatti, and A. Hendawi, "Evaluating the evolution of yolo (you only look once) models: A comprehensive benchmark study of yolo11 and its predecessors," *arXiv preprint arXiv:2411.00201*, 2024.

[47] J. E. Gallagher and E. J. Oughton, "Surveying you only look once (yolo) multispectral object detection advancements, applications and challenges," *IEEE Access*, 2025.

[48] M. Sohan, T. Sai Ram, and C. V. Rami Reddy, "A review on yolov8 and its advancements," in *International Conference on Data Intelligence and Cognitive Informatics*, pp. 529–545, Springer, 2024.

[49] X. Zhu and D. Ramanan, "Face detection, pose estimation, and landmark localization in the wild," in *2012 IEEE conference on computer vision and pattern recognition*, pp. 2879–2886, IEEE, 2012.

[50] L. Prechelt, "Early stopping-but when?," in *Neural Networks: Tricks of the trade*, pp. 55–69, Springer, 2002.

[51] X. Ying, "An overview of overfitting and its solutions," in *Journal of physics: Conference series*, vol. 1168, p. 022022, IOP Publishing, 2019.

[52] O. O. Napoli, A. M. de Almeida, E. Borin, and M. Breternitz Jr, "Memory-efficient drasiw models," *Neurocomputing*, vol. 610, p. 128443, 2024.

[53] S. Wen and W. Dang, "Research on base64 encoding algorithm and php implementation," in *2018 26th International Conference on Geoinformatics*, pp. 1–5, IEEE, 2018.

APPENDIX A

Merged Datasets Distribution

A.1. Merged Train Private Validation

Merged Train Private Validation				
Requirement	Train		Validation	
	Compliant	Non-Compliant	Compliant	Non-Compliant
2	7348	1002	926	109
4	8503	744	1011	24
6	5836	3487	533	502
7	8639	690	1012	23
8	8264	1061	1000	35
9	8464	863	1025	10
10	7045	2188	854	181

Merged Train Private Validation distribution for Train and Validation datasets

A.2. Merged Train Public Validation

Merged Train Public Validation				
Requirement	Train		Validation	
	Compliant	Non-Compliant	Compliant	Non-Compliant
2	7348	1002	433	31
4	8503	744	523	12
6	5836	3487	335	155
7	8639	690	538	31
8	8264	1061	478	84
9	8464	863	538	32
10	7045	2188	476	4

Merged Train Public Validation distribution for Train and Validation datasets

APPENDIX B

Best Results by Model for Section A Experiments

B.1. Experiment A.1.

B.1.1. Baseline Models- A.1.

Requirement	Pre-processing	Size	Model	Specificity	F1-Measure
2	-	-	-	0.000	0.964
4	Original	256	ResNet	0.250	0.987
6	Original	256	ResNet	0.819	0.959
7	-	-	-	0.967	0.999
8	Original	256	ResNet	0.797	0.981
9	Cropped	256	ResNet	0.781	0.993
10	Original	256	ResNet	0.250	0.996

Best Baseline Model by requirement- Experiment A.1.

B.1.2. WNN Models- A.1.

Requirement	Pre-processing	BPI	Specificity	F1-Measure
2	-	-	0.000	0.000
4	-	-	0.000	0.000
6	-	-	0.000	0.000
7	-	-	0.226	0.000
8	-	-	0.000	0.000
9	-	-	0.000	0.000
10	-	-	0.000	0.000

Best WNN Model by requirement- Experiment A.1.

B.1.3. Yolov8 Models- A.1.

Requirement	Pre-processing	Size	Specificity	F1-Measure
2	Cropped	640	0.906	0.996
4	Cropped	640	0.833	0.995
6	Both	256	0.994	0.995
7	-	-	0.967	0.999
8	Original	256	0.964	0.989
9	Cropped	640	1.000	1.000
10	-	-	0.500	0.998

Best Yolov8 Model by requirement- Experiment A.1.

B.1.4. Yolo 11 Models- A.1.

Requirement	Pre-processing	Size	Specificity	F1-Measure
2	Cropped	640	0.813	0.989
4	Original	256	0.417	0.992
6	Original	640	1.000	0.993
7	-	-	0.967	0.999
8	Original	640	0.952	0.994
9	Original	640	1.000	0.999
10	-	-	0.500	0.998

Best Yolo 11 Model by requirement- Experiment A.1.

B.2. Experiment A.2.

B.2.1. Baseline Models- A.2.

Requirement	Pre-processing	Size	Model	Specificity	F1-Measure
2	Cropped	256	ResNet	0.156	0.968
4	Original	256	ResNet	0.250	0.992
6	Cropped	640	ResNet	0.832	0.957
7	Both	640	VGG	0.967	0.999
8	Original	256	ResNet	0.774	0.979
9	Cropped	256	ResNet	0.844	0.994
10	Original	256	Both	0.250	0.996

Best Baseline Model by requirement- Experiment A.2.

B.2.2. WNN Models- A.2.

Requirement	Pre-processing	BPI	Specificity	F1-Measure
2	-	-	0.000	0.000
4	-	-	0.000	0.000
6	-	-	0.000	0.000
7	Original	16	0.172	0.000
8	-	-	0.000	0.000
9	-	-	0.000	0.000
10	-	-	0.000	0.000

Best WNN Model by requirement- Experiment A.2.

B.2.3. Yolov8 Models- A.2.

Requirement	Dataset	Pre-processing	Size	Dropout (%)	Specificity	F1-Measure
2	Merged	Original	640	0	0.906	0.994
4	Public	Original	256	50	0.583	0.987
6	Public	Both	640	50	0.994	0.995
7	-	-	-	-	0.967	0.999
8	Merged	Cropped	640	50	0.976	0.995
9	-	-	-	-	1.000	1.000
10	Merged	Original	256	0	1.000	0.995

Best Yolov8 Model by requirement- Experiment A.2.

B.2.4. Yolo 11 Models- A.2.

Requirement	Dataset	Pre-processing	Size	Dropout (%)	Specificity	F1-Measure
2	Merged	Cropped	640	0	0.844	0.994
4	Merged	Cropped	640	0	0.333	0.992
6	Public	Cropped	640	25	0.994	0.994
7	-	-	-	-	0.967	0.999
8	-	-	640	0 and 25	0.976	0.995
9	-	-	-	-	0.968	0.999
10	-	-	640	0 and 25	0.750	0.998

Best Yolo 11 Model by requirement- Experiment A.2.

B.3. Experiment A.3.

B.3.1. Baseline Models- A.3.

Requirement	Dataset	Pre-processing	Size	Model	Specificity	F1-Measure
1	Private	Original	256	ResNet	0.689	0.880
2	Merged	Original	256	ResNet	0.697	0.491
3	Private	Original	640	ResNet	0.299	0.908
4	Merged	Original	256	ResNet	0.083	0.988
5	Private	Cropped	256	ResNet	0.172	0.927
6	Private	Cropped	640	VGG	0.898	0.932
7	-	-	-	-	0.652	0.996
8	Merged	Cropped	640	ResNet	0.314	0.964
9	Merged	Original	640	ResNet	0.500	0.915
10	Merged	Cropped	640	ResNet	0.629	0.662

Best Baseline Model by requirement- Experiment A.3.

B.3.2. WNN Models- A.3.

Requirement	Dataset	Pre-processing	Size	BPI	Specificity	F1-Measure
1	-	-	-	-	0.000	0.000
2	-	-	-	-	0.000	0.000
3	-	-	-	-	0.000	0.000
4	-	-	-	-	0.000	0.000
5	-	-	-	-	0.000	0.000
6	-	-	-	-	0.000	0.000
7	-	-	-	-	0.000	0.000
8	-	-	-	-	0.000	0.000
9	-	-	-	-	0.000	0.000
10	-	-	-	-	0.000	0.000

Best WNN Model by requirement- Experiment A.3.

B.3.3. Yolov8 Models- A.3.

Requirement	Dataset	Pre-processing	Size	Dropout (%)	Specificity	F1-Measure
1	Private	Original	256	25	0.943	0.965
2	Private	Original	256	50	0.339	0.958
3	Private	Cropped	256	0	0.569	0.926
4	Private	Cropped	256	50	0.292	0.989
5	Private	Original	256	50	0.682	0.947
6	Private	Cropped	256	0	0.964	0.959
7	Merged	Cropped	640	0	1.000	0.999
8	Merged	Original	640	50	0.429	0.988
9	Merged	Cropped	256	25	0.500	0.997
10	Private	Original	256	0	0.691	0.946

Best Yolov8 Model by requirement- Experiment A.3.

B.3.4. Yolo 11 Models- A.3.

Requirement	Dataset	Pre-processing	Size	Dropout (%)	Specificity	F1-Measure
1	Private	Cropped	256	25	0.941	0.964
2	Private	Cropped	256	0	0.284	0.956
3	Private	Original	640	0	0.519	0.924
4	Merged	Cropped	640	25	0.125	0.987
5	Private	Original	256	50	0.701	0.945
6	Private	Original	256	25	0.956	0.957
7	Private	Cropped	256	25	1.000	0.997
8	Merged	Original	640	50	0.429	0.988
9	Merged	Original	640	50	0.600	0.997
10	Private	Original	640	25	0.619	0.921

Best Yolo 11 Model by requirement- Experiment A.3.

APPENDIX C

Best Results by Model for Section B Experiments

C.1. Experiment B.1.

C.1.1. Yolov8 Models- B.1.

Requirement	Pre-processing	Size	Specificity	F1-Measure
2	Original	640	0.266	0.827
4	Cropped	640	0.500	0.938
6	Cropped	256	0.841	0.889
7	Original	640	0.826	0.997
8	Original	256	0.743	0.960
9	Cropped	640	0.300	0.996
10	Original	640	0.293	0.896

Best Yolov8 Model by requirement- Experiment B.1.

C.1.2. Yolo 11 Models- B.1.

Requirement	Pre-processing	Size	Specificity	F1-Measure
2	Original	640	0.184	0.887
4	Original	256	0.292	0.976
6	Cropped	640	0.883	0.825
7	Original	Both	0.696	0.996
8	Cropped	640	0.514	0.959
9	Original	256	0.300	0.992
10	Original	256	0.425	0.874

Best Yolo 11 Model by requirement- Experiment B.1.

C.2. Experiment B.2.

C.2.1. Yolov8 Models- B.2.

Requirement	Dataset	Pre-processing	Size	Dropout (%)	Specificity	F1-Measure
2	Merged	Original	640	25	0.312	0.947
4	Public	Cropped	256	50	0.375	0.963
6	Merged	Cropped	256	25	0.952	0.958
7	Merged	Original	640	25	0.869	0.998
8	Public	Cropped	640	50	0.543	0.968
9	Public	Cropped	640	50	0.600	0.977
10	Merged	Original	256	0	0.613	0.929

Best Yolov8 Model by requirement- Experiment B.2.

C.2.2. Yolo 11 Models- B.2.

Requirement	Dataset	Pre-processing	Size	Dropout (%)	Specificity	F1-Measure
2	Merged	Cropped	256	0	0.257	0.947
4	Merged	Cropped	640	0	0.250	0.988
6	Public	Cropped	256	50	0.958	0.955
7	Public	Cropped	256	50	0.913	0.997
8	Merged	Cropped	256	25	0.457	0.981
9	Public	Cropped	640	25	0.600	0.986
10	Public	Cropped	640	50	0.585	0.931

Best Yolo 11 Model by requirement- Experiment B.2.

C.3. Experiment B.3.

C.3.1. Yolov8 Models- B.3.

Requirement	Dataset	Pre-processing	Size	Dropout (%)	Specificity	F1-Measure
2	Merged	Original	640	50	0.656	0.987
4	Merged	Cropped	256	25	0.417	0.991
6	Merged	-	640	-	0.987	0.983
7	-	-	-	-	0.967	0.999
8	Merged	Cropped	640	50	0.952	0.993
9	Merged	Original	640	25	1.000	1.000
10	-	-	-	-	0.750	0.996

Best Yolov8 Model by requirement- Experiment B.3.

C.3.2. Yolo 11 Models- B.3.

Requirement	Dataset	Pre-processing	Size	Dropout (%)	Specificity	F1-Measure
2	Merged	Cropped	640	0	0.656	0.986
4	Merged	Original	256	25	0.083	0.986
6	Merged	Cropped	256	50	1.000	0.974
7	-	-	-	-	0.967	0.999
8	Merged	Original	640	50	0.941	0.993
9	Merged	Cropped	640	25	0.968	0.997
10	Merged	Original	640	0	0.750	0.998

Best Yolo 11 Model by requirement- Experiment B.3.