Article

Automatic Classification of Hospital Settings through Artificial Intelligence

Ernesto Iadanza [1,2,*,†] **, Giovanni Benincasa** [1,†] **, Isabel Ventisette** [1,†] **and Monica Gherardelli** [1]

1 Department of Information Engineering, University of Florence, 50139 Firenze, Italy; giovanni.benincasa@stud.unifi.it (G.B.); isabel.ventisette@stud.unifi.it (I.V.); monica.gherardelli@unifi.it (M.G.)
2 Department of Medical Biotechnologies, University of Siena, 53100 Siena, Italy
* Correspondence: ernesto.iadanza@unisi.it; Tel.: +39-0577-232584
† These authors contributed equally to this work.

Abstract: Modern hospitals have to meet requirements from national and international institutions in order to ensure hygiene, quality and organisational standards. Moreover, a hospital must be flexible and adaptable to new delivery models for healthcare services. Various hospital monitoring tools have been developed over the years, which allow for a detailed picture of the effectiveness and efficiency of the hospital itself. Many of these systems are based on database management systems (DBMSs), building information modelling (BIM) or geographic information systems (GISs). This work presents an automatic recognition system for hospital settings that integrates these tools. Three alternative proposals were analysed in terms of the construction of the system: the first was based on the use of general models that are present on the cloud for the classification of images; the second consisted of the creation of a customised model and referred to the Clarifai Custom Model service; the third used an object recognition software that was developed by Facebook AI Research combined with a random forest classifier. The obtained results were promising. The customised model almost always classified the photos according to the correct intended use, resulting in a high percentage of confidence of up to 96%. Classification using the third tool was excellent when considering a limited number of hospital settings, with a peak accuracy of higher than 99% and an area under the ROC curve (AUC) of one for specific classes. As expected, increasing the number of room typologies to be discerned negatively affected performance.

Keywords: artificial intelligence; CNN; image classification; hospital settings

1. Introduction

This work aims to provide a method for the automatic classification and labelling of hospital rooms based on their typologies. The need for such a method stems from the fact that many computer-aided facility management (CAFM) systems are being used in hospitals nowadays, but their value and usefulness is tightly linked to the correlation between the data they provide in terms of room use and performed activities and the real situation. Looking at the current situation, in which the updating of these data is delegated to inspectors who manually assign the use of rooms based on inspections and surveys, improving the level of automation in updating this information is paramount. Each hospital is a very complex structure that provides a multitude of services. This complexity keeps growing because modern technology increases the range of diagnostic capabilities and expands the number of treatment options [1]. A combination of medical research, engineering and biotechnology has resulted in a multitude of new treatments and instrumentation, which often require specialised training and facilities for their use. Therefore, hospitals have become more expensive to run and healthcare managers are increasingly interested in quality, cost, effectiveness and efficiency issues, leading to the need to develop new technical tools that allow hospital monitoring through measuring

quantitative, architectural, technological and people-related parameters [2–4]. From these reflections, the idea of this project was born and we aimed to present solutions for the automatic classification of hospital settings from images of hospital spaces in order to manage them more quickly and efficiently. The pervasive presence of autonomous mobile robots (AMRs) in hospitals [5], which are often provided with video cameras, is likely to increase, as testified by many EU-funded projects, such as "Robotics4EU" [6] and "Odin Smart Hospitals" [7]. These robots continuously move around hospitals and can acquire photos and videos of the hospital rooms. The method suggested in this article is a novel supplement to such technologies, allowing for the extraction of as much information as possible from these valuable sources and leveraging their presence to also provide decision-makers with knowledge of the real usage of hospital spaces. With regard to the Italian healthcare system, it is necessary to refer to the Decree of the President of the Italian Republic, issued on 14 January 1997 [8], which states that in order to carry out healthcare activities in the national territory, it is necessary to comply with specific accreditation requirements. This document is the first legislative reference with a national nature, which identifies the minimum, general and specific requirements for authorising the exercise of public and private health activities. Within established terms, regions can integrate these requirements for authorisation and define additional requirements for the accreditation of already authorised structures. Consequently, since 1997, regions have followed different transposition paths and issued different requirements for authorisation and accreditation. All requirements can be grouped according to their type: organisational, structural, plant and technological [9]. In Tuscany, the healthcare system is governed by the Regional Law of 24 February 2005, no. 40 [10], and by its subsequent amendments and additions. Within these documents, the different types of requirements for different healthcare settings can be found. Thanks to these legislative documents, it is possible to identify the characteristics of different types of hospital settings. Clearly, all wards, operating rooms, intensive care units and the many other spaces that constitute modern hospitals have very different characteristics. For the design and implementation of a system that performs the automatic classification of such settings, it is important to identify the structural and technological elements that distinguish rooms that are used for different purposes, together with their specific plant elements.

1.1. Related Works

This subsection presents related works that address the problems of developing technical tools to improve hospital facility management (FM). Providing healthcare facility management professionals with enhanced decision-making support systems would have a positive impact on the productivity and success of these structures. Irizarry et al. [11] proposed a conceptual ambient intelligent environment for enhancing the decision-making process of facility managers. This environment uses building information modelling (BIM) and mobile augmented reality (MAR) as the technological bases for the human–computer interfaces and uses aerial drones as technological tools. The BIM approach is becoming very common for designing and managing hospitals. Spatial and structural functional data could be obtained using this approach, but implementing a complete BIM model for a complex scenario, such as healthcare structures, requires many resources. Wanigarathna et al. [12] investigated how BIM can be used to integrate a wide range of information and improve built asset management (BAM) decision-making during the in-use phase of hospital buildings. In parallel, many authors [13–16] have proposed systems that are based on the applications of data management in the Internet of Things (IoT) in order to better manage hospital organisation. In particular, healthcare computer-aided facility management (CAFM) and healthcare space management activities are strategic for establishing a dialogue between information and stakeholders. They extrapolate the elements characterising the functions of the management process from the heterogeneity of data and users. CAFM techniques have the aim of defining expert tools for the control of the information that is associated with assets. This is carried out through integrated systems of

graphical and numerical databases. Luschi et al. [4] illustrated the methodology and tools used by a multidisciplinary research team, which was composed of architects and computer engineers who supported the requalification project for the Careggi University Hospital of Florence. The authors described a tool that was developed by the team: SACS (System for the Analysis of Hospital Equipment), a custom software that guides AutoCAD to manage and analyse digital floor plans of buildings that are encoded on specific levels. The software maps the departments and related operating units, uses, healthcare technologies and environmental comfort by grouping the information into single room and homogeneous areas, thereby providing quantitative and qualitative results [8]. However, the labelling of rooms is performed manually, room by room, and no automatic classification system was described. In [17], an integrated workplace management system (WMS) tool was introduced. It produces key performance indicators (KPIs) and quantitative parameters that are typical of CAFM systems. Such systems allow for the assessment of an entire building or technological estate and can also prioritise the assignment of the most urgent interventions. The system imports plain 2D maps to offer a central management cockpit that deals not only with structural and constructional data, but also technologies, assets and medical equipment.

Over the years, some papers on automatic room classification have been produced. A system that extracts both structural and semantic information from given floor plans was proposed in 2012 [18]. In 2018, Brucker [19] presented an approach to automatically assign semantic labels to rooms that are reconstructed from 3D RGB maps of apartments. Evidence for the room types is generated using state-of-the-art deep learning techniques for scene classification and object detection based on automatically generated virtual RGB views, as well as geometric analyses of the mapped 3D structures. Recently, a new article proposed a floor plan information retrieval algorithm, which is based on shape extraction and room identification. A classification model based on a regression model was also proposed to classify rooms according to their function [20].

In particular, there have been some recent studies dedicated to room categorisation and semantic mapping. Sünderhauf et al. [21] introduced transferable and expandable place categorisation and semantic mapping using a robot without environment-specific training. Mancini et al. [22] realised a work that focused on the problems of semantic place categorisation using visual data. They presented a deep learning model for addressing domain generalisation (DG) in the context of semantic place categorisation. In a 2019 study, Pal et al. [23] designed five models for room labelling that combined object detection and scene recognition algorithms. In 2020, Li et al. [24] presented a regional semantic learning method based on convolutional neural networks (CNNs) and conditional random fields. The method combines global information that is obtained by a scene classification network with local object information that is obtained by an object detection network to train a CRF scene recognition model. In 2021, Jin et al. [25] proposed a deep learning-based novel feature fusion method for indoor scene classification, which combines object detection and enriched semantic information. Finally, Liu et al. [26] proposed a vision-based cognitive system to support the independence of visually impaired people. A 3D indoor semantic map is first constructed with a hand-held RGB-D sensor and is then deployed for indoor topological localisation. CNNs are used for both semantic information extraction and location inference. The semantic information is then used to further verify the localisation results and eliminate errors.

1.2. The Role of Artificial Intelligence

The project presented in this article aimed to implement a system for the automatic classification of hospital settings through tools based on artificial intelligence (AI) [27,28]. AI is a field of computer science that includes several branches, among which is machine learning (ML). ML encompasses a range of methods and algorithms that make a program able to identify patterns from data or improve learning. hlDeep learning (DL), a class of ML algorithms, creates learning models at multiple levels [29]. In the specific case of our

project, the aim was image classification. In this context, ML allows the manual selection of features and provides a classifier for sorting the images. The features are then used to create a model for assigning categories to objects in images. In DL workflows, the significant features are automatically extracted from images. In addition, DL performs end-to-end learning through a network that automatically learns to process raw data and carry out an activity, for example, a classification. Another key difference is that DL algorithms scale data, while superficial learning uses convergence. By superficial learning, we mean ML methods that do not allow for further development once a certain level of performance has been reached, even when further training examples and data are added to the network. A key benefit of DL networks is the possibility to improve performance as data formats increase. The optimal approach clearly depends on the problem at hand and the tools that are available for that purpose. As far as image classification and object recognition are concerned, ML can be an effective technique in many cases, especially when the image characteristics (features) that are best suited to differentiating classes of objects are known. For applications of *object recognition* and *image classification*, DL has become the best tool thanks to *convolutional neural networks* (CNNs) [30–32]. A CNN consists of tens or hundreds of layers, each of which learns to detect different image features. Indeed, each level hosts a "feature map", which is the specific characteristic that each node is looking for. For this purpose, filters are applied to each image at different resolutions and the output of each processed image is used as the input for the next layer. Filters can initially consist of very simple features, such as brightness and edges, and can then gradually take on more complex shapes that uniquely define the object. As with other neural networks, a CNN is composed of an input layer (which is the set of all images taken from the dataset), several hidden layers and an output [33–35]. The image classification performance of CNNs has been improving steadily since 2015 [36–38]. This performance is mainly due to training, which is a human-like process. All of the main technological companies within the medical field are studying AI applications. These applications are mainly for archiving medical records [39,40] and for medical diagnostics [41–43], with important applications in oncology [44,45]. The use of AI has recently been extended to cardiovascular imaging techniques [46], diagnosis of pulmonary/respiratory diseases [47,48], as well as to hepatology [49], and ocular diseases [50,51]. An interesting future research direction would be the study of AI applications for neurocritical care, especially the design of a system that can evaluate better strategies for neurocritical care patients [52].

Although the BIM-based approach is interesting and looks promising for the future, the availability of hospital 3D BIM models is very limited at the moment. The approaches based on CAFM software are extremely time-consuming and rely on the continuous manual updating of data from manual surveys. To the best of our knowledge, none of the works based on automatic room categorisation have been specific for healthcare settings, which is a challenging task compared to general purpose classifiers. Finally, the key issues of systems based on RGB-D cameras are that they need specific hardware and cannot exploit large volumes of available images and videos that come from widely spread RGB cameras, nor can they use available databases of images for training. Our research aimed to fill the gaps mentioned above by means of a system that can achieve a high performance for automated hospital-specific room categorisation and requires nothing but simple, widely available and medium-quality RGB images. The proposed system does not require a manual labelling step, which is required in many existing works, nor does it require 3D BIM models or manual data entry.

1.3. Novelty of the Proposed Approach

The goal of our project was to develop a mechanism for automatically classifying hospital facilities to be used for the continuous updating of the CAFM systems that are currently utilised in hospitals, whose worth and utility is directly tied to the correlation between the data they offer and the actual reality. A schematic representation of the system is shown in Figure 1. The system solves the problem of having to carry out

continuous inspections to manually update CAFM systems. Images of rooms that are taken by robots, surveillance cameras or other sources are interpreted by the designed classifier and labelled with a specific use. Hospital CAFM systems are then continuously updated with this information.
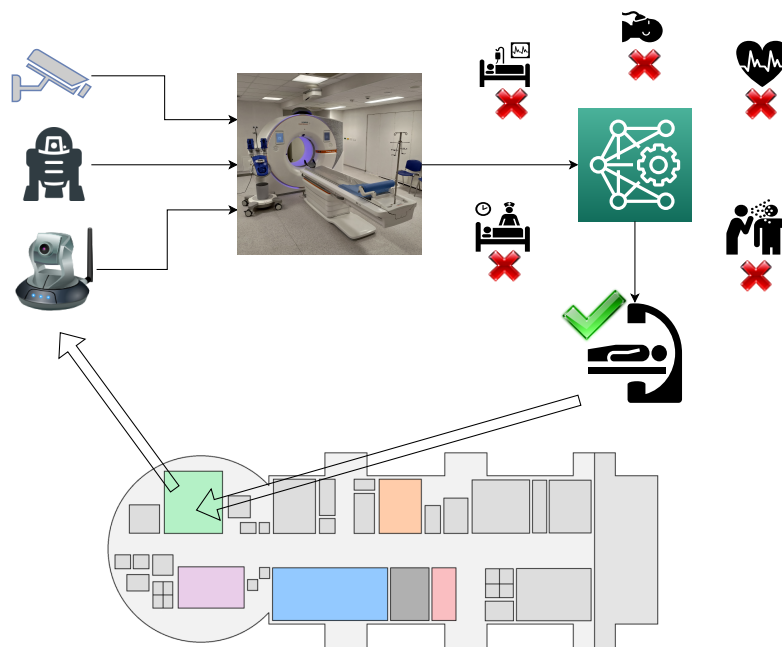
Figure 1. Schematic representation of the proposed system. Images of rooms that are taken by robots, surveillance cameras or other sources are interpreted by the designed classifier and labelled with a specific use. Hospital CAFM systems are then continuously updated with this information.

The usefulness of the proposed approach is related to huge time savings compared to the current process of updating information about the use of hospital rooms. In fact, this process currently requires manual surveys from inspectors who then manually update the hospital CAFM systems. An automated process based on artificial intelligence that is able to classify and label rooms by just analysing pictures would be a tremendous gain for both saved time and update frequency. We addressed different operational solutions:

- We analysed and compared three general models for environment classification: Google Vision API, Microsoft Azure Cognitive Services and the Clarifai General Model;
- We then created a customised model that was specifically trained for our needs using the Clarifai Custom Model;
- Finally, we carried out one last type of classification using Detectron2, an object detection software that works in combination with an RF classifier for image recognition.

In the next section, we describe the different methodologies that were used to analyse the AI-based image classification methods in our project. Section 3.1 reports the results that were obtained by the different classification techniques. A discussion of the obtained results is developed in Section 4.

2. Materials and Methods

Three alternative methods for image classification are being proposed. The first relies on the use of cloud-based image understanding services that are offered by IT service providers, such as *Amazon, Google, Microsoft* and *Clarifai*. These service providers offer *application programming interfaces* (APIs), which enable the classification of images without requiring the large amounts of training data and long training times that are standard for DL. The second option comprises the independent development of an on-site software that is based on CNN models, which requires configuration and training [53–55]. There are

different ways to approach customised recognition using DL, namely using a pre-trained model or training a model from scratch. In our work, the preferred solution was to refine a pre-trained network using transfer learning (TL). This approach transfers knowledge from one or more related tasks to boost learning in the target task [56]. It is generally much faster and simpler than training a model from scratch since it requires a minimal amount of data and computing resources [57]. The third alternative consists of using an object recognition software (Detectron2, which was developed by *Facebook AI Research* (FAIR) [58]) combined with a *random forest* (RF) classification algorithm. The RF algorithm classifies *intended uses* (IUs) based on the results from Detectron2.

2.1. Datasets

2.1.1. First Dataset

In order to test the three general models and create a customised model to compare them to, we approached the matter of selecting a set of images of the objects of interest. The quality of a dataset is crucial for implementing a user model. The larger the dataset, the higher the quality of the resulting model. Several datasets of different sizes can be considered. Among these, one of the best-known datasets within the image recognition community is *ImageNet* [59,60], which currently contains 14.2 million images. On the *WordNet Structure* web page, it is possible to identify the types of images of interest. Our objective was to classify hospital images into four categories: "hospitalisation", "acceptance", "surgery" and "diagnostic and therapeutic radiology". We selected a dataset that included 80 photographs, which were acquired from Google Images [61] and belonged to four different IUs, as shown in Appendix A:

- 20 "surgery" images (from 1 to 20);
- 20 images of "diagnostic and therapeutic radiology" (from 21 to 40);
- 20 "hospitalisation" images (from 41 to 60);
- 20 "acceptance" images (from 61 to 80).

The dataset characteristics are described in Appendix C, in terms of size and quality. The general models were tested using all 80 photographs. Afterwards, since it was necessary to train and test our customised model, we split the photographs of this dataset into two distinct groups:

- The *training set*, which was only used during the model training phase. This set was composed of images that were divided into two groups:

  - Positive examples, i.e., photographs for each of the four classes that were introduced as positive benchmark examples;
  - Negative examples, i.e., photographs of negative examples that were imported for each of the four classes from the remaining IUs.

- The *test set*, which was used in the model performance verification phase. This was made up of 40 images from the four chosen IUs.

  Two versions of the customised model were produced:

- The first version comprised 10 positive examples and 18 negative examples for each IU (6 images for each incorrect IU);
- The second version comprised 20 positive examples and 18 negative examples for each IU (6 images for each incorrect IU).

The 80 photographs that were selected for the training and test sets of the two versions of the model are shown in Appendix A. An additional 10 images were used for each IU in the second version of the model. These 40 images, which were different from the previous images, increased the training sets of the four IUs. They are not included in this manuscript for brevity, but they are available from the corresponding authors. Their description is presented in Appendix C, in terms of dimensions and quality. The following criterion was used to select the negative examples: for each IU, the first six elements that were used as

positive examples for the other three IUs were selected as negative examples. For example, for the "surgery" IU, the negative examples were represented by the following images:

- Items 21, 22, 23, 24, 25 and 26 (from the "diagnostic and therapeutic radiology" IU);
- Items 41, 42, 43, 44, 45 and 46 (from the "hospitalisation" IU);
- Items 61, 62, 63, 64, 65 and 66 (from the "acceptance" IU).

The selection of the training set, along with its division into positive and negative examples, and the test set for the first version of the customised model is shown in Table 1. The same selection for the second version of the customised model is shown in Table 2.

Table 1. Selection of the training set, along with its division into positive and negative examples, and the test set for the first version of the customised model.

IU	Positive Training Examples	Negative Training Examples	Test Set
Surgery	1, 2, 3, 4, 5, 6, 7, 8, 9, 10	21, 22, 23, 24, 25, 26, 41, 42, 43, 44, 45, 46, 61, 62, 63, 64, 65, 66	11, 12, 13, 14, 15, 16, 17, 18, 19, 20
Radiology	21, 22, 23, 24, 25, 26, 27, 28, 29, 30	1, 2, 3, 4, 5, 6, 41, 42, 43, 43, 44, 45, 46, 61, 62, 63, 64, 65, 66	31, 32, 33, 34, 35, 36, 37, 38, 39, 40
Hospitalisation	41, 42, 43, 44, 45, 46, 47, 48, 49, 50	1, 2, 3, 4, 5, 6, 21, 22, 23, 24, 25, 26, 61, 62, 63, 64, 65, 66	51, 52, 53, 54, 55, 56, 57, 58, 59, 60
Acceptance	61, 62, 63, 64, 65, 66, 67, 68, 69, 70	1, 2, 3, 4, 5, 6, 21, 22, 23, 24, 25, 26, 41, 42, 43, 44, 45, 46	71, 72, 73, 74, 75, 76, 77, 78, 79, 80

Table 2. Selection of the training set, along with its division into positive and negative examples, and the test set for the second version of the customised model.

IU	Positive Training Examples	Negative Training Examples	Test Set
Surgery	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90	21, 22, 23, 24, 25, 26, 41, 42, 43, 44, 45, 46, 61, 62, 63, 64, 65, 66	11, 12, 13, 14, 15, 16, 17, 18, 19, 20
Radiology	21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100	1, 2, 3, 4, 5, 6, 41, 42, 43, 43, 44, 45, 46, 61, 62, 63, 64, 65, 66	31, 32, 33, 34, 35, 36, 37, 38, 39, 40
Hospitalisation	41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110	1, 2, 3, 4, 5, 6, 21, 22, 23, 24, 25, 26, 61, 62, 63, 64, 65, 66	51, 52, 53, 54, 55, 56, 57, 58, 59, 60
Acceptance	61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120	1, 2, 3, 4, 5, 6, 21, 22, 23, 24, 25, 26, 41, 42, 43, 44, 45, 46	71, 72, 73, 74, 75, 76, 77, 78, 79, 80

2.1.2. Second Dataset

This section describes the dataset that was used for developing and testing two models that were based on the Detectron2 object recognition software. Two datasets were built in order to compare the results obtained from the first model, which was trained using the first dataset considering only three IUs, to those obtained from the second model, which was trained using the second dataset considering nine IUs:

- The first model examined "hospitalisation", "radiology" and "surgery" rooms, for which 40 images per room were acquired from Google Images [61] using the corresponding keywords for a total of 120 images in the first dataset;

- The second model included six more IUs ("ambulance", "analysis laboratory", "intensive therapy", "medical clinic", "rehabilitation and physiotherapy" and "toilet") for a total of nine hospital settings, for which 40 images per room were selected from Google Images [61] using the corresponding keywords for a total of 360 images in the second dataset.

A full description can be found in Appendix C for both datasets (Tables A7 and A8), in terms of image quality and size.

To train the object recognition algorithm, the two datasets were divided into three separate sets, as shown in Tables 3 and 4:

- The *training Set*, which was composed of 25 images per IU and was used to train the algorithm to recognise the objects of interest;
- The *validation Set*, which was composed of 10 images per IU and was used to refine the hyperparameters of the model during training;
- The *test Set*, which was composed of the remaining 5 images per IU and was used at the end of the training to produce a final evaluation of the model.

Two versions of each model were considered: the first used the original dataset and the second used a dataset that was modified by data augmentation changes.

Table 3. Selection of the training set, validation set and test set for the first version of the Detectron2 model (3 IUs).

IU	Training Set	Validation Set	Test Set
Hospitalisation	11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35	1, 2, 3, 4, 5, 6, 7, 8, 9, 10	36, 37, 38, 39, 40
Radiology	11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35	1, 2, 3, 4, 5, 6, 7, 8, 9, 10	36, 37, 38, 39, 40
Surgery	11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35	1, 2, 3, 4, 5, 6, 7, 8, 9, 10	36, 37, 38, 39, 40

Table 4. Selection of the training set, validation set and test set for the second version of the Detectron2 model (9 IUs).

IU	Training Set	Validation Set	Test Set
Ambulance	11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35	1, 2, 3, 4, 5, 6, 7, 8, 9, 10	36, 37, 38, 39, 40
Analysis Laboratory	11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35	1, 2, 3, 4, 5, 6, 7, 8, 9, 10	36, 37, 38, 39, 40
Hospitalisation	11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35	1, 2, 3, 4, 5, 6, 7, 8, 9, 10	36, 37, 38, 39, 40
Intensive Therapy	11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35	1, 2, 3, 4, 5, 6, 7, 8, 9, 10	36, 37, 38, 39, 40
Medical Clinic	11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35	1, 2, 3, 4, 5, 6, 7, 8, 9, 10	36, 37, 38, 39, 40

Table 4. *Cont.*

IU	Training Set	Validation Set	Test Set
Radiology	11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35	1, 2, 3, 4, 5, 6, 7, 8, 9, 10	36, 37, 38, 39, 40
Rehabilitation and Physiotherapy	11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35	1, 2, 3, 4, 5, 6, 7, 8, 9, 10	36, 37, 38, 39, 40
Surgery	11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35	1, 2, 3, 4, 5, 6, 7, 8, 9, 10	36, 37, 38, 39, 40
Toilet	11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35	1, 2, 3, 4, 5, 6, 7, 8, 9, 10	36, 37, 38, 39, 40

2.2. Models Based on Image Understanding Services: General Classification Models

Many service providers offer general image classification models, including *Google Vision API* [62], *Amazon Rekognition* [63], *Microsoft Azure Cognitive Services* [64] and the *Clarifai General Model* [65]. All of these models provide similar features, such as object labelling, face detection, text extraction (optical character recognition, OCR), image attribute statistics, etc. Information about the models underlying these services is, for the most part, not available in the public domain, except for Clarifai, which has made the CNN model it uses public. Nevertheless, since CNNs are now considered to be state of the art within the field of image recognition, it is very likely that most services exploit this approach [28,57].

An image-based cognitive API receives an image from an external application, extracts specific information from it and then returns the information, usually in *JavaScript Object Notation* (JSON) format. This information usually contains a set of words called "tags" or "labels", which are objects and concepts that the API has recognised within the given image. Some examples of tags that may be returned by an API include "living room", "indoors" or "classroom". The labels are also accompanied by a confidence percentage value, which denotes how well the model recognises those specific objects or concepts in the image. It is not possible to go into any meaningful detail about the documentation concerning how cloud services are trained. The manufacturers simply state that their models undergo continuous training using images from the web. Brief descriptions of the services offered by Google, Microsoft and Clarifai are provided below:

Google Cloud Vision (API Vision): Google offers two AI-based computer vision products for image understanding: AutoML Vision and API Vision [62,66,67]. AutoML Vision allows users to build their own customised models through TL, whilst the second product is based on "ready-to-use" models. API Vision labels images to quickly classify them into millions of predefined categories and can detect objects and faces by determining their position and number. In order to test API Vision, Google launched a demonstration website through which the API issues labels, identifies and reads texts and detects faces in each selected image [68–70].

Clarifai General Model: The Clarifai and Microsoft service providers also offer services that are similar to Google's. In particular, Clarifai offers an open-image API called Clarifai Predict [65], whose operation is similar to Google's API in that once an image is entered, a list of labels and corresponding probability levels is generated. In this case, a generic image classification model, such as the Clarifai General Model, or a customised model can be applied [71,72].

Microsoft Azure Cognitive Services: Microsoft Azure Cognitive Services [64] enable visual data processing in order to label content (from objects to concepts), extract printed and handwritten text and recognise familiar objects, such as trademarks and places of interest.

2.3. Models Customised through Transfer Learning

As already pointed out, it is often preferable to create a customised model through TL. Many service providers offer such solutions, which make it extremely easy to develop individualised models. These include *Google AutoML Vision*, which allows the automation of custom model training so that images can be classified using the labels that were selected by users, based on their own specific requirements. Users can simply upload their images and train their models using a specific graphical user interface (GUI). Then, they can export the images to on-site devices or cloud-based applications. Another similar tool is *Amazon Rekognition Custom Labels*. Again, the user needs a small number of training images (usually a few hundred or less) that are specific to their use. Even *IBM Watson Visual Recognition*, which runs in the cloud or on iOS devices, enables users to train custom image models and develop their own image classifiers using specific image collections by leveraging TL. Finally, Clarifai also allows users to build their own models from a model that has been pre-trained through the *Clarifai Custom Model* service [73]. This service works similarly to the previous solutions, thereby allowing users to employ their own images and label them with the concepts that they need.

Implementation of the Customised Models

To create a customised model that was trained specifically for our needs, we selected the above-mentioned Clarifai Custom Model, which allowed us to create our model using a free *community plan* that includes a limited number of monthly operations and inputs. First, we selected the dataset, as shown in Section 1. Once the collection and organisation of the set of images was complete, we moved on to the implementation of the model on the Clarifai platform. It was then necessary to create an application. Inside the application, we introduced the concepts of interest, i.e., the four IUs: "surgery", "diagnostic and therapeutic radiology", "hospitalisation" and "acceptance". These were the four outputs that we wanted to obtain from the model. The model was then trained using the *Custom Model* section of Clarifai. Starting from a predefined model offered by Clarifai, we could implement our own classification model by splitting the training images into positive and negative examples for each considered concept. We only loaded the images that we needed to obtain the first version of the model; afterwards, we introduced the additional photos that were needed for the development of the second version of the model. The Clarifai starting model that we chose to train the custom model was the *edit context-based classifier model*, which is the most suitable option for image classification. Once the model had been trained, we moved on to creating two workflows, one for each version of the model. Each workflow was a calculation graph in which the output from one model could be used as the input for the next model. We introduced our custom model into the workflow (the first and the second versions into the two different workflows) as the output from another model of the *Clarifai visual embedder* type. Indeed, the custom model that we created had "embeddings" as inputs and returned "concepts" as outputs. Therefore, a Clarifai model was used that returned the appropriate outputs (embeddings) from the images. We used the test sets to evaluate the performances of the two versions of the customised model, after they had been built and trained. For each version, the four groups of ten images relating to the four considered IUs were analysed, as specified above. The data were collected as already described for the first approach, building a total of eight tables (four tables for each version of the model). Each column was specific to one of the images for that particular IU, while each row corresponded to one of the four "trained" concepts. The confidence percentage with which our model assigned each label (in rows) to each image (in columns) was found at the intersection of each row and column. Positive classification outcomes were highlighted in green, while negative results were in red.

2.4. Combined Use of Detectron2 and an RF Classification Algorithm

The classification took place using Detectron2, an object detection software [58] whose outputs are in the form of a dataframe that trains an RF classifier for image recognition. Detectron2 is an open-source software system, developed by FAIR, which implements state-of-the-art computer vision algorithms. This software is implemented on PyTorch, an open-source ML framework [74], and is capable of providing fast training using single or multiple graphic processing units (GPUs). Detectron2 includes the implementation of state-of-the-art detection and segmentation algorithms. RF is a scheme for building a classification ensemble with a set of decision trees that grow in randomly selected subspaces [75,76]. It leverages several decision tree classifiers in different subsamples of the dataset and uses the average to improve its predictive accuracy and control overfitting. Decision trees are non-parametric supervised training methods that are used for classification and regression [77]. The goal is to create a model that predicts the value of a target variable by learning simple decision rules, which are deduced from data features. In this project, we used the RF classifier model from the scikit-learn library [78].

2.4.1. Dataset Pre-Processing

The open-source *LabelMe* software was used to annotate the dataset images [79]. LabelMe was developed by the *Computer Science and Artificial Intelligence Laboratory* (CSAIL) at the Massachusetts Institute of Technology (MIT). It is a software for building image databases that are to be used within computer vision or datasets that are already annotated and ready for use. LabelMe creates *JavaScript Object Notation* (JSON) files for each image. These text files contain a lot of information, such as the "label", which identifies the annotated object, the "points", which are the coordinates of the points that describe the perimeter of the object, and other information that is necessary to solve the object detection problem. In the first version of our model, we annotated 120 images by identifying the main objects within the three selected hospital rooms. The following eight labels were used: "bed", "cabinet", "chair", "monitor", "operating table", "RMN machine", "surgical light" and "window". In total, 240 images were annotated in the second version and we also added the following labels: "ball", "bidet", "bicycle", "desk", "examination bed", "grab bar", "IVD", "mirror", "sink", "stool", "surgical instrument table", "toilet" and "wall bars". Overall, 21 objects were considered. The datasets of both versions were uploaded to Roboflow, a framework for computer vision developers that helps to collect and organise data that are to be pre-processed [80]. Roboflow has public datasets that are readily available for users and offers the opportunity to upload your own custom data. There is also the possibility to customise your dataset during pre-processing. Roboflow allows you to automatically split the images into three types of datasets: training, validation and test datasets. The validation dataset is used to refine the hyperparameters of the customised model. The available images of each hospital setting were divided into the three datasets: 25 training images, 10 validation images and 5 test images. Roboflow also allows you to edit training images by adding features such as orientation and data augmentation. It is recommended to apply these characteristics to make the model more precise and invariant to the photo angle, the brightness of the room and blur. We considered two variations of both datasets:

- The first version contained the dataset without modifications: 75 training images, 30 validation images and 15 test images for the first model; 225 training images, 90 validation images and 45 test images for the second model;
- The second version contained the modified dataset, with an image rotation of up to ±45° and a blur of up to 1 pixel were applied (this choice was motivated by the size of the images that were downloaded from Google): 224 training images, 30 validation images and 15 test images for the first model; 671 training images, 90 validation images and 45 test images for the second model.

These data were converted into the COCO format [81] used by the Facebook API and Detectron2 for training.

2.4.2. Parameter Selection and Model Calibration

The complete dataset (after annotation and parting into the training, validation and test datasets) was uploaded to the cloud using the appropriate Roboflow web service. We then proceeded to register the dataset on Detectron2 in its standard format.

Detectron2: Parameters and Calibration

We selected the "faster_rcnn_X_101_32 × 8d_FPN_3x" model because of its highest average precision (AP = 43.0), according to the tests carried out by the developers using *Big Basin*, which is a new generation GPU server. However, this was at the expense of long training times (0.638 s/iter) and a large memory consumption (6.7 GB). The name, definition and set value of each hyperparameter are listed in Table 5.

The last hyperparameter referred to the test configuration and the others referred to the training configuration.

Once the training was configured and carried out, we evaluated our model's performance through the average precision, average recall and total loss metrics:

- Average precision (AP) is the ratio between the true positives (correct answers) and the sum of the true positives and false positives (incorrect answers that are considered correct by the model). It indicates the percentage with which the model identifies an object. In the results, six types of average precision were considered, whose meaning is described in Table 6. Three APs were based on the intersection over union (IoU), which represents the overlap between the "predicted" and real bounding boxes. A bounding box is a box that is outlined around the object of interest in order to locate it within the image. The IoU is calculated as the intersection area of the union area of these two cited bounding boxes. A value of 1 represents a perfect overlap.
- Average recall is the ratio between the true positives and the sum of the true positives and false negatives (correct answers but considered wrong by the model). It indicates the percentage with which the model correctly identifies an object.
- Total loss evaluates the model's behaviour with the datasets: the lower the value, the better the behaviour. It is calculated during the training and validation phases.

Table 5. Name, definition and set value of each hyperparameter of the Detectron2 model.

Name	Definition	Set Value
cfg.DATALOADER.NUM_WORKERS	Number of data loading threads	2
cfg.SOLVER.IMS_PER_BATCH	Number of images per batch on all machines (GPU) and number of training images per iteration	2
cfg.SOLVER.BASE_LR	Learning rate controlling how quickly the model adapts to the problem (less than 1.0)	0.00025
cfg.SOLVER.MAX_ITER	Number of iterations during training (variable)	Mod 1 ver 1:2500 1 ver 2:5000 Mod 2:5000
cfg.MODEL.ROI_HEADS.BATCH_SIZE_PER_IMAGE	Number of regions per image used to train the region proposal network (RPN)	128
cfg.MODEL.ROI_HEADS.NUM_CLASSES	Number of classes/objects noted in the dataset (the number of classes + 1)	9 with 3 hospital settings 22 with 9 settings
cfg.MODEL.ROI_HEADS.SCORE_THRESH_TEST	Threshold for object identification: the object is not taken into account when its confidence percentage is lower than this threshold	80%

Table 6. The metrics used to evaluate the model (average precision and average recall). AP is the average precision of the intersection over union (IoU) in steps of 0.05, from 0.5 to 0.95. AP (IoU = 0.50) and AP (IoU = 0.75) correspond to APs with IoUs of 0.50 and 0.75, respectively. AR describes the doubled area under the recall–IoU curve.

Average Precision (AP)	
AP	AP at IoU = 0.50:0.05:0.95 (primary challenge metric)
AP (IoU = 0.50)	AP at IoU = 0.50 (PASCAL VOC metric)
AP (IoU = 0.75)	AP at IoU = 0.75 (strict metric)
AP Across Scales	
AP Small	AP for small objects: area < 32^2 px
AP Medium	AP for medium objects: 32^2 px < area < 96^2 px
AP Large	AP for large objects: area > 96^2 px
Average Recall (AR)	
AR (max = 1)	AR given 1 detection per image

RF Classifier: Parameters and Calibration

The RF algorithm used a dataframe as the input, which is a two-dimensional structure within which data is stored. Two pieces of information were needed: the features identifying the characteristics of the object to be classified and the target, which is the label of the object to be classified to which the features correspond. The dataframe was generated from each dataset. The datasets were obtained from the Detectron2 outputs, more specifically from the pred_classes values of each image. The pred_classes output was a vector consisting of all objects recognised by Detectron2 within an image, with object encoded with a number. For instance, "cabinet" was encoded with 4, "examination bed" was encoded with 8, etc. In each dataset, the values corresponding to the features of the dataframe were all objects that were used to train Detectron2 (9 features in the first model; 21 features in the second model), while the target elements were the rooms to which the identified features corresponded. Each feature equalled the number of identical objects identified in the image. Consider the following example: 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 2, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0. Among the twenty-two elements in this vector, twenty-one were features, while the last one was the target (0 was "ambulance"). We chose to count the number of objects instead of only checking for their presence because some rooms were almost identical in terms of the objects inside them but different in the number of hosted objects. The training was carried out using the default hyperparameter values. The parameters used for the evaluation of the model's performance are listed below, where TP means true positive, FP means false positive, TN means true negative and FN means false negative:

- Accuracy is the ratio of correctly predicted observations to the total observations, i.e., (TP + TN)/(TP + TN + FP + FN);
- Score is the harmonic mean between precision and recall (which is usually more useful than accuracy, especially for non-symmetrical datasets and when the costs of false positives and false negatives are very different), i.e., 2 ∗ ((precision ∗ recall)/(precision + recall));
- Precision is the ratio of correctly predicted positive observations to the total predicted positive observations (the higher the value, the lower the number of false positives), i.e., TP/(TP + FP);
- Recall or TPR is the ratio of correctly predicted positive observations to all truly positive observations, i.e., TP/(TP + FN);
- Specificity is the ratio of correctly predicted negative observations to the total negative observations, i.e., TN/(TN + FP);
- The receiver operating characteristic curve (ROC curve) is a graph showing the diagnostic capability of a binary classification system as a function of its discrimination thresholds, which plots the true positive rate (TPR) versus the false positive rate (FPR = 1 − TPR) at different threshold settings;

- The area under the ROC curve (ROC AUC score) is the area under the ROC curve, which equals 1 when the classifier works perfectly.

3. Results

3.1. Comparison of the General Models of Cloud Services

We analysed and compared the three general models that were previously introduced: Google Vision API, Microsoft Azure Cognitive Services and the Clarifai General Model. The key objectives were to verify whether these models are suitable for the classification of hospital settings and figure out which of them produces the best results. For this purpose, each of the models was applied to the same test dataset, which was composed of the 80 photographs described in Section 1 and shown in Appendix A.

Data were organised into twelve tables: three tables for each of the four IUs (one for each examined cloud service). Each column of the tables was dedicated to one of the 20 images for that specific IU, while each row corresponded to the label that was assigned by the model. At the intersection of each row and column, the confidence percentage with which the model recognised the label of that row when examining an image from that column was displayed. A selection of these tables is reported in Appendix B.

3.2. Results Obtained with the Clarifai Custom Model

For a better visualisation of the differences between the two versions of the customised model, the results tables were sorted by IU. The tables relating to the results obtained with the first version of the model were placed first, then those relating to the second version.

- Tables 7 and 8 refer to the results obtained for the "surgery" IU;
- Tables 9 and 10 refer to the results obtained for the "radiology" IU;
- Tables 11 and 12 refer to the results obtained for the "hospitalisation" IU;
- Tables 13 and 14 refer to the results obtained for the "acceptance" IU.

Table 7. Results obtained with the first version of the customised model for the "surgery" IU. Each column refers to one of the test images (11 to 20) and each row shows the labels that were returned by the model. The row–column intersection shows the success rate in recognising the concept stated in the row for the image in that column.

	Image 11	Image 12	Image 13	Image 14	Image 15	Image 16	Image 17	Image 18	Image 19	Image 20
Surgery	27%	11%	19%	84%	5%	67%	2%	10%	70%	66%
Acceptance	0%	0%	0%	0%	0%	0%	0%	4%	0%	0%
Hospitalisation	5%	2%	0%	2%	5%	0%	0%	0%	1%	0%
Radiology	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%

Table 8. Results obtained with the second version of the customised model for the "surgery" IU. Each column refers to one of the test images (11 to 20) and each row shows the labels that were returned by the model. The row–column intersection shows the success rate in recognising the concept stated in the row for the image in that column.

	Image 11	Image 12	Image 13	Image 14	Image 15	Image 16	Image 17	Image 18	Image 19	Image 20
Surgery	85%	48%	37%	93%	68%	79%	10%	3%	91%	93%
Acceptance	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Hospitalisation	17%	3%	1%	18%	21%	4%	1%	0%	1%	2%
Radiology	0%	0%	0%	0%	0%	0%	0%	9%	0%	0%

Table 9. Results obtained with the first version of the customised model for the "radiology" IU. Each column refers to one of the test images (31 to 40) and each row shows the labels that were returned by the model. The row–column intersection shows the success rate in recognising the concept stated in the row for the image in that column.

	Image 31	Image 32	Image 33	Image 34	Image 35	Image 36	Image 37	Image 38	Image 39	Image 40
Surgery	0%	0%	0%	44%	0%	0%	0%	0%	0%	0%
Acceptance	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Hospitalisation	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%
Radiology	39%	68%	13%	1%	8%	16%	16%	96%	51%	7%

Table 10. Results obtained with the second version of the customised model for the "radiology" IU. Each column refers to one of the test images (31 to 40) and each row shows the labels that were returned by the model. The row–column intersection shows the success rate in recognising the concept stated in the row for the image in that column.

	Image 31	Image 32	Image 33	Image 34	Image 35	Image 36	Image 37	Image 38	Image 39	Image 40
Surgery	0%	0%	0%	19%	0%	0%	0%	0%	0%	0%
Acceptance	0%	0%	0%	0%	0%	0%	1%	0%	0%	0%
Hospitalisation	0%	0%	1%	1%	0%	0%	0%	0%	0%	0%
Radiology	63%	93%	30%	11%	41%	64%	49%	89%	59%	8%

Table 11. Results obtained with the first version of the customised model for the "hospitalisation" IU. Each column refers to one of the test images (51 to 60) and each row shows the labels that were returned by the model. The row–column intersection shows the success rate in recognising the concept stated in the row for the image in that column.

	Image 51	Image 52	Image 53	Image 54	Image 55	Image 56	Image 57	Image 58	Image 59	Image 60
Surgery	0%	0%	1%	0%	0%	0%	0%	0%	5%	1%
Acceptance	0%	0%	0%	0%	0%	2%	0%	0%	0%	0%
Hospitalisation	28%	62%	86%	72%	87%	82%	77%	61%	32%	47%
Radiology	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%

Table 12. Results obtained with the second version of the customised model for the "hospitalisation" IU. Each column refers to one of the test images (51 to 60) and each row shows the labels that were returned by the model. The row–column intersection shows the success rate in recognising the concept stated in the row for the image in that column.

	Image 51	Image 52	Image 53	Image 54	Image 55	Image 56	Image 57	Image 58	Image 59	Image 60
Surgery	0%	1%	1%	4%	2%	0%	3%	0%	2%	14%
Acceptance	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%
Hospitalisation	31%	74%	90%	68%	90%	56%	65%	61%	58%	64%
Radiology	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%

Table 13. Results obtained with the first version of the customised model for the "acceptance" IU. Each column refers to one of the test images (71 to 80) and each row shows the labels that were returned by the model. The row–column intersection shows the success rate in recognising the concept stated in the row for the image in that column.

	Image 71	Image 72	Image 73	Image 74	Image 75	Image 76	Image 77	Image 78	Image 79	Image 80
Surgery	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Acceptance	27%	36%	71%	19%	34%	44%	46%	66%	70%	60%
Hospitalisation	1%	0%	0%	0%	1%	2%	0%	0%	0%	0%
Radiology	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%

Table 14. Results obtained with the second version of the customised model for the "acceptance" IU. Each column refers to one of the test images (71 to 80) and each row shows the labels that were returned by the model. The row–column intersection shows the success rate in recognising the concept stated in the row for the image in that column.

	Image 71	Image 72	Image 73	Image 74	Image 75	Image 76	Image 77	Image 78	Image 79	Image 80
Surgery	0%	1%	0%	0%	14%	0%	0%	0%	0%	0%
Acceptance	52%	85%	71%	91%	52%	47%	94%	73%	85%	67%
Hospitalisation	2%	0%	0%	0%	0%	1%	0%	0%	0%	0%
Radiology	0%	0%	0%	0%	0%	1%	1%	0%	1%	1%

3.3. Results from the Combined Use of Detectron2 and the RF Classification Algorithm

At first, only the results obtained with both versions of the first model, which refers to only three hospital environments, were compared and analysed. The second model was considered later.

3.3.1. Performance Obtained with the First Model

Detectron2 has an internal performance evaluator with many metrics for performance evaluation, the most significant of which are listed and explained in Section 2.4.2. Table 15 shows the values obtained for the different performance metrics (listed in the first row) when applying the first and second versions of this model.

Table 15. Detectron2 metrics obtained with the first model. Each column refers to a different metric, as defined in Section 2.4.2. Each row refers to a different version of the model.

Versions	AP	AP50	AP75	APs	APm	APl	AR	Total Loss (×100)
Version 1 (0 AUG)	48.976	69.375	53.721	38.026	36.337	67.599	46.8	14.01
Version 2 (2 AUG)	46.761	74.425	49.907	35.6	38.292	63.738	45.7	17.4

Regarding the RF classifier's performance, the synthesis results obtained for the two versions are shown in Table 16, which was built using the scikit-learn functions for the calculation of the metrics. It should be noted that the algorithm rarely classified the hospital environment incorrectly. The results for each hospital environment also indicated that the values were almost optimal for all environments. The ROC curve completely shifted towards the upper left corner, which represents the condition of optimal results. This was also confirmed by the AUC, which was maximal for the three hospital environments.

Table 16. Metrics of the RF classification algorithm obtained with the first model. Each column refers to a different metric, as defined in Section 2.4.2. Each row refers to a different version of the model.

Versions	Accuracy	F1 Score	Precision	Recall
Version 1 (0 AUG)	0.97777	0.97775	0.97916	0.97777
Version 2 (2 AUG)	0.97777	0.97775	0.97916	0.97777

3.3.2. Performance Obtained with the Second Model

Table 17 shows the metrics that were obtained automatically by Detectron2 when considering the second model. With reference to the performance of the RF classifier, the results obtained for the second model using the scikit-learn functions are reported in Table 18.

Table 17. Detectron2 metrics obtained with the second model. Each column refers to a different metric, as defined in Section 2.4.2. Each row refers to a different version of the model.

Versions	AP	AP50	AP75	APs	APm	APl	AR	Total Loss ($\times$100)
Version 1 (0 AUG)	44.479	65.268	49.875	34.436	42.344	50.831	43.8	19.71
Version 2 (2 AUG)	34.477	66.193	29.352	29.68	31.59	37.886	36.5	39.42

Table 18. Metrics of the RF classification algorithm obtained with the second model. Each column refers to a different metric, as defined in Section 2.4.2. Each row refers to a different version of the model.

Versions	Accuracy	F1 Score	Precision	Recall
Version 1 (0 AUG)	0.7555	0.75104	0.7645	0.7555
Version 2 (2 AUG)	0.7037	0.7044	0.7194	0.7037

Since the algorithm produced different performances for the different examined hospital settings, Tables 19 and 20 report the detailed results for each setting for both versions of the model. The ROC curves, shown in Figure 2, did not drop below 0.87 for the AUC value in either version of the model.

Table 19. Metrics of the RF classification algorithm obtained with the first version of the second model. Each column refers to a different metric, as defined in Section 2.4.2. Each row refers to a different IU.

Rooms	Accuracy	F1 Score	Precision	Recall	Specifity
Ambulance	0.903703704	0.580645161	0.6	0.5625	0.949579832
Analysis Laboratory	0.948148148	0.740740741	0.666666667	0.833333333	0.959349593
Hospitalisation	0.911111111	0.647058824	0.733333333	0.578947368	0.965517241
Intensive Therapy	0.940740741	0.714285714	0.666666667	0.769230769	0.959016393
Medical Clinic	0.911111111	0.5	0.4	0.666666667	0.928571429
Radiology	0.985185185	0.9375	1	0.882352941	1
Rehabilitation and Physiotherapy	0.933333333	0.742857143	0.866666667	0.65	0.982608696
Surgery	0.985185185	0.928571429	0.866666667	1	0.983606557
Toilet	0.992592593	0.967741935	1	0.9375	1

Table 20. Metrics of the RF classification algorithm obtained with the second version of the second model. Each column refers to a different metric, as defined in Section 2.4.2. Each row refers to a different IU.

Rooms	Accuracy	F1 Score	Precision	Recall	Specifity
Ambulance	0.896296296	0.461538462	0.4	0.545454545	0.927419355
Analysis Laboratory	0.933333333	0.689655172	0.666666667	0.714285714	0.958677686
Hospitalisation	0.933333333	0.727272727	0.8	0.666666667	0.974358974
Intensive Therapy	0.918518519	0.64516129	0.666666667	0.625	0.957983193
Medical Clinic	0.896296296	0.588235294	0.666666667	0.526315789	0.956896552
Radiology	0.977777778	0.888888889	0.8	1	0.975609756
Rehabilitation and Physiotherapy	0.911111111	0.647058824	0.733333333	0.578947368	0.965517241
Surgery	0.940740741	0.0692307692	0.6	0.818181818	0.951612903
Toilet	1	1	1	1	1
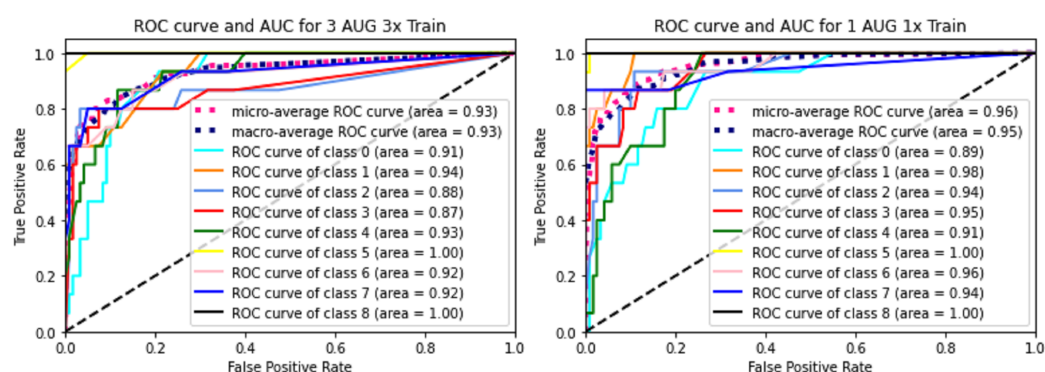
Figure 2. ROC and ROC AUC curves obtained for the second model (first version on the left; second version on the right).

4. Discussion of Results

4.1. Discussion of Results Obtained with General Models and the Clarifai Custom Model

The tables described in Section 3.1, which are partially shown in Appendix B, allowed for some comments on the collected data. Starting from the first examined IU, namely the "surgery" IU, we could conclude that Google Vision API and Clarifai General Model were able to correctly classify images (except in a single case) because they assigned labels, such as "operating theatre, "operating room" and "surgery", with good accuracy (Google: never less than 55%; Clarifai: never less than 94.2%). Conversely, the same photos produced very different results when analysed by Microsoft Azure Cognitive Services. Indeed, the Microsoft model almost always recognised the general scope, namely the hospital environment by assigning labels such as "hospital" and "hospital room", but it only classified an image as "operating theatre" in one case. On the other hand, these results strictly depended on the training dataset that was used and the labels that were introduced during the training of the models, as well as the different architecture of the involved networks. The three tables relating to the "surgery" IU led us to think that the classification models by Google and Clarifai were trained with a good number of images for this IU. Conversely, Microsoft likely used a poorer set of images when training its model. This model rarely proved capable of recognising a specific IU, despite possessing a specific label for it. With reference to the second IU, namely "diagnostic and therapeutic radiology", the results confirmed the performance of the Microsoft model, which was capable of recognising the general medical–health field but even in this case, did not correctly assign the IU label. The Google model, on the other hand, maintained a good, but not better, performance compared to the previous IU. It always assigned labels such as "radiology" and "radiography" with high levels of confidence (except in one case). On the contrary, the Clarifai General Model returned inadequate labels and percentages for this second IU. In fact, it very often recognised "surgery" settings in the analysed images with too much

confidence and only sometimes assigned the correct labels. This outcome highlighted that the recognition of a hospital environment was not essential for the proposed classification task. It was much more important for the task to identify characteristic features of the IU in order to correctly assign labels that entail the correct classification of the image. Concerning the "hospitalisation" IU, the collected data were satisfactory compared to the previous two IUs. Probably, none of the three considered services had a specific label for this IU. In any case, the models by both Google and Microsoft recognised the general context for most of the photos (as in the previous IUs), assigning labels such as "hospital" or "medical equipment". They were rarely wrong because of using, for instance, using surgery-related labels. A specific label, "hospital room", is highlighted in green in Table A3, which refers to the classification of the "hospitalisation" IU by the Microsoft model. This label was the closest to the definition of hospitalisation and was sometimes actually assigned by the API. The Clarifai model maintained the same behaviour of the previous IUs and still recognised the hospital environment. It often assigned misleading labels ("surgery" and "emergency") with very high confidence rates. Finally, it was interesting to note that in the three cases, some completely wrong labels, such as "bathroom", "living room", "classroom" and several others, were assigned to some images.

Regarding the "acceptance" IU, as expected, the cloud services rarely classified the image as relating to a hospital (an element that was not fundamental to our purpose, as already specified). In addition to this, labels such as "waiting room" or "reception", which would have led to the correct classification of the image, were rarely assigned.

In conclusion, the obtained results and the consequent observations showed that the examined models generally produced good performances. These systems were able to attribute a great and varied number of correct labels to different levels of taxonomy. Google Vision API, Microsoft Azure Cognitive Services and the Clarifai General Model were able to identify high-level concepts as "indoors" and most of the mid-level concepts as "hospital". However, they behaved differently according to the specific application. Regarding the specific objective of this study, namely the classification of hospital settings, the three interfaces showed different and not completely satisfactory outputs overall. In particular, the obtained results suggested that Google's Vision API would be the best choice for directly classifying a hospital room. However, it should be noted that only four IUs and twenty images for each of model were examined. Therefore, the chosen images could have favoured one system over another and the systems could have produced completely different outputs with another test dataset. This study highlighted that these models would not be the best choice for classifying hospital environments. In fact, the proposed objective was extremely specific, while the used general models were trained with millions of images that are different and have many labels. It would be advisable to develop a model that only returns the outputs of interest, i.e., the IU in this case. Finally, the APIs did not know all of the labels needed to recognise each IU. For all of these reasons, a customised model was then developed. The obtained results relating to both versions of the customised model, as shown in the tables in Section 3.2, highlighted better performances than those of the Clarifai General Model. In fact, both systems almost always classified the photos according to the correct IU and attributed the highest percentage of confidence. For both versions, an image was not classified correctly in only two cases with very low confidence values. In fact, a major problem with the Clarifai General Model was the overconfidence in assigning wrong labels. On the other hand, when the results from the first version of our model were considered, the confidence values attributed to the correct labelling were also quite low. This was probably due to the very small training sets. In fact, the second version of our customised model generally produced much higher confidence percentages in an overwhelming majority of cases. The results were therefore promising. The results relating to the "surgery" IU were an exception to this positive trend. Indeed, in this specific case, the general model performed better than the customised model. This was not surprising since the general model, as illustrated above, associated labels such as "surgery" to many hospital images with very high confidence values.

4.2. Discussion of Results Obtained with the Combined Use of Detectron2 and the RF Classification Algorithm

Model 1: Detectron2 Performance. A very small image dataset was available in this case, with images that were characterised by very low resolutions. Even though the dataset was inappropriate for an object detection problem, the results obtained were satisfactory since both versions of the model achieved greater than 45% for the average precision metric. A model is considered good when it scores around 70. The difference between the two versions is worth noting. The first version scored higher than the second version in almost all metrics. We could justify this unexpected result with the inappropriate structure of the dataset. The changes introduced to the dataset did not lead to an improvement in the model's performance. We also considered the impact of the number of iterations on the obtained results. For the first version, it was appropriate to increase the number of iterations to improve the model accuracy. For the second version, the problem was probably the data augmentation type that was applied because such a small dataset could not support these changes. It is necessary to report that the algorithm recognised some objects better than others due to their shape. In fact, the correctness of the recognition also depended on the reference images with which the training was carried out. A change that would certainly lead to an improvement in the performance of this model is the expansion of the starting dataset by selecting images with better resolutions.

Model 1: RF Classifier Performance. The results obtained were excellent for both versions of the model. Indeed, the algorithm rarely classified the hospital environments incorrectly. The excellent results were confirmed by the ROC curve. In fact, the curve completely shifted towards the upper left corner, which represents the condition of optimal results. This was also confirmed by the AUC, which was maximal for all three hospital environments. This model thus produced very promising results despite the small dataset size and the low resolutions of the images. These results prompted us to test the limits of this type of project by proposing the second model.

Model 2: Dtectron2 Performance. The second model did not perform as well as the first. In fact, it achieved lower values than the first model in all metrics, including the total loss metric, which was very high (especially for the second version of this model). This was only due to the increased number of objects that Detectron2 should have identified. Indeed, we did not increase the size of the dataset or the number of iterations. For this model, as for the previous model, the number of iterations also played a fundamental role. The limit of the number of iterations was relevant as both versions of the second model exhibited an increasing behaviour, even at Iteration 5000. It could be deduced that for both versions, but especially the second version, we could increase the number of iterations to improve the model accuracy. In addition to the number of iterations, the dataset also had a lot of influence. As mentioned for the first model, we would have gained better results for this part of the object detection problem by using a larger dataset with higher quality images, if we had increased the number of iterations.

Model 2: RF Classifier Performance. In this case, we obtained worse results for the classification than those obtained with the first model. This was due to the greater number of hospital environments that were examined: increasing this number led the algorithm to be more prone to making mistakes. Some hospital environments had very similar characteristics, for example, "hospitalisation" and "intensive therapy" or "ambulance" and "medical clinic". The worst results were obtained for critical hospital settings, such as "medical clinic" and "ambulance".

As anticipated, many authors have conducted studies that relate to that presented in this article. The automatic method developed by Brucker et al. [19] for assigning semantic labels to rooms from RGB-D data reported an average accuracy of around 67%. Mewada et al. [20] achieved an average room detection accuracy of 85.71% and a room recognition accuracy of 88% with their algorithm, which is based on shape extraction and room identification. The system for automatic room detection and room labelling from architectural floor plans proposed by Ahmed et al. [18] was able to correctly label

around 80% of the analysed rooms. Sünderhauf et al. [21] obtained an average accuracy that did not exceed 67.7% with their transferable and expandable place categorisation and semantic mapping system. The DL model for addressing domain generalisation proposed by Mancini et al. [22] reached an average accuracy of no more than 56.5%. The five models proposed by Pal et al. [23] for place categorisation achieved, in the best case, a 70.1% average accuracy. The regional semantic learning method developed by Li et al., which is based on CNNs and conditional random fields, was able to obtain an average accuracy of 77.6%. Finally, the feature fusion method for indoor scene classification proposed by Jin et al. obtained an average accuracy rate of 66%.

A comparison of the results from the literature to those from the present study allows us to be optimistic. In fact, the room recognition accuracy obtained with the second version of the model developed with the Clarifai General Model was 95%. This model successfully classified 38 out of 40 images with various levels of certainty, which increased with the number of images in the training dataset. The first model that was made with the use of Detectron2 and an RF classification algorithm also reported an average accuracy of over 97%. The second model produced a worse performance due to the greater number of examined hospital facilities. However, we are convinced that a higher number and quality of images in the dataset could produce equally positive results. Table 21 shows the comparison between the performances of the models in the literature and those of our work, in terms of average accuracy.

Table 21. Comparison of model performances in terms of average accuracy.

Model	Average Accuracy
Brucker et al.	67%
Mewada et al.	85.71%
Ahmed et al.	80%
Sünderhauf et al.	67.7%
Mancini et al.	56.5%
Pal et al.	70.1%
Li et al.	77.6%
Jin et al.	66%
Second version of our model, developed with the Clarifai General Model	**95%**
First version of our model, developed with Detectron2 and the RF classification algorithm	**97.78%**

The results discussed above show that a novel approach for the automatic classification of hospital spaces based on computer vision is possible. The increasing presence of autonomous mobile robots (AMRs) in hospitals, which are exploited for many tasks, from disinfection to telemedicine, and are often provided with cameras [5], is providing an endless source of updated images of hospital premises. The approach proposed in this work is a novel complement to these pervasive technologies in order to extract as much information as possible from these precious sources.

5. Conclusions

This paper presented a project that aimed to implement a system for the automatic classification of hospital settings using tools based on AI. For this purpose, three alternatives were proposed: the first was based on the use of general cloud models for image classification; the second consisted of a customised model, which was implemented through the personalisation services offered by the same service providers; the last exploited the combined use of Detectron2, an open-source software system developed by FAIR, and an RF classification algorithm. In order to evaluate the effectiveness of the first solution, three cloud services were tested and compared: Google Vision API, the Clarifai General Model and Microsoft Azure Cognitive Services. The interfaces offered by the service providers are based on general models that have been trained with many images of different types.

These models returned labels that were sometimes not suitable for the IU of the analysed images. In fact, these models offered a general recognition of the image environment and objects, i.e., they were not specialised for the specific environments of hospital settings. Google Vision API proved to be the most reliable system in the classification task overall. It rarely assigned misleading labels and could recognise elements that actually characterised the IUs. Even though the Clarifai General Model was excellent in the classification of surgery images, it encountered much more difficulty in the classification of the other IUs, almost always identifying "surgery" elements both for hospitalisation and radiology rooms. Finally, the API offered by Microsoft rarely succeeded in labelling rooms according to their use. We then moved on to the implementation of a custom model using the Clarifai Custom Model service. It was possible to develop this model with much more specific images through TL. The model, which was created in a very simple way, almost always labelled the images correctly. When the number of training images was increased, the confidence percentage of IU recognition also increased. This suggests that it would be possible to develop an extremely precise model by using a suitable training dataset. For the third alternative, two models were proposed: the first was general and the second was more specific. For the first model, the system correctly classified almost all hospital settings. The implementation of the second model followed the same steps as the first; however, it obtained worse results (although still acceptable). The limitation of this model lay in the construction of the dataset, which consisted of images from Google Images with very low resolutions. One more limitation was the small size of the dataset, which was very small for effective object detection and image recognition. To improve this model, it is necessary to improve the dataset by increasing the size and choosing images of higher quality. Therefore, it is particularly important to take care of the size and quality of the images in the training set, both for the second and third alternatives proposed in this project.

Author Contributions: Conceptualisation, E.I.; methodology, E.I.; software, I.V. and G.B.; validation, I.V. and G.B.; formal analysis, E.I., I.V. and G.B.; investigation, E.I., I.V. and G.B.; resources, E.I. and M.G.; data curation, E.I., I.V. and G.B.; writing—original draft preparation, M.G.; writing—review and editing, M.G. and E.I.; visualisation, I.V. and G.B.; supervision, E.I.; project administration, M.G. All authors have read and agreed to the published version of the manuscript.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
AP	Average Precision
API	Applicant Program Interface
BIM	Building Information Modelling
CAFM	Computer-Aided Facility Management
CNN	Convolutional Neural Network
CSAIL	Computer Science and Artificial Intelligence Laboratory
DBMS	Database Management System
DL	Deep Learning
FAIR	Facebook Artificial Intelligence Research
GIS	Graphical User Interface
GPU	Graphics Processing Unit
IoU	Intersection Over Union
IT	Information Technology
IU	Intended Use
JSON	JavaScript Object Notation
ML	Machine Learning

OCR	Optical Character Recognition
RF	Random Forest
ROC Curve	Receiver Operating Characteristic Curve
ROC AUC Score	Area Under the ROC Curve Score
TL	Transfer Learning

Appendix A

The photographs included in Figures A1–A3 show the training and test sets that were used for the two versions of the model described in Section 2.3. We labelled the photographs with progressive numbers to facilitate the comparison and description of the results. The number of images was enough for the first version of the model. The photographs were distributed as follows:

- Photographs 1–10 in Figure A1: positive examples of the training set used for both versions of the custom model for the "surgery" IU;
- Photographs 11–20 in Figure A1: test set used for both versions of the custom model for the "surgery" IU;
- Photographs 21–30 in Figure A1: positive examples of the training set used for both versions of the custom model for the "diagnostic and therapeutic radiology" IU;
- Photographs 31–40 in Figure A2: test set used for both versions of the custom model for the "diagnostic and therapeutic radiology" IU;
- Photographs 41–50 in Figure A2: positive examples of the training set used for both versions of the custom model for the "hospitalisation" IU;
- Photographs 51–60 in Figure A2: test set used for both versions of the custom model for the "hospitalisation" IU;
- Photographs 61–70 in Figure A3: positive examples of the training set used for both versions of the custom model for the "acceptance" IU;
- Photographs 71–80 in Figure A3: test set used for both versions of the custom model for the "acceptance" IU.

Figure A1. The images used in the training sets (1 to 10) and test sets (11 to 20) of both custom models for the "surgery" IU and the images used in the training sets (21 to 30) of both custom models for the "diagnostic and therapeutic radiology" IU (from Google Images).
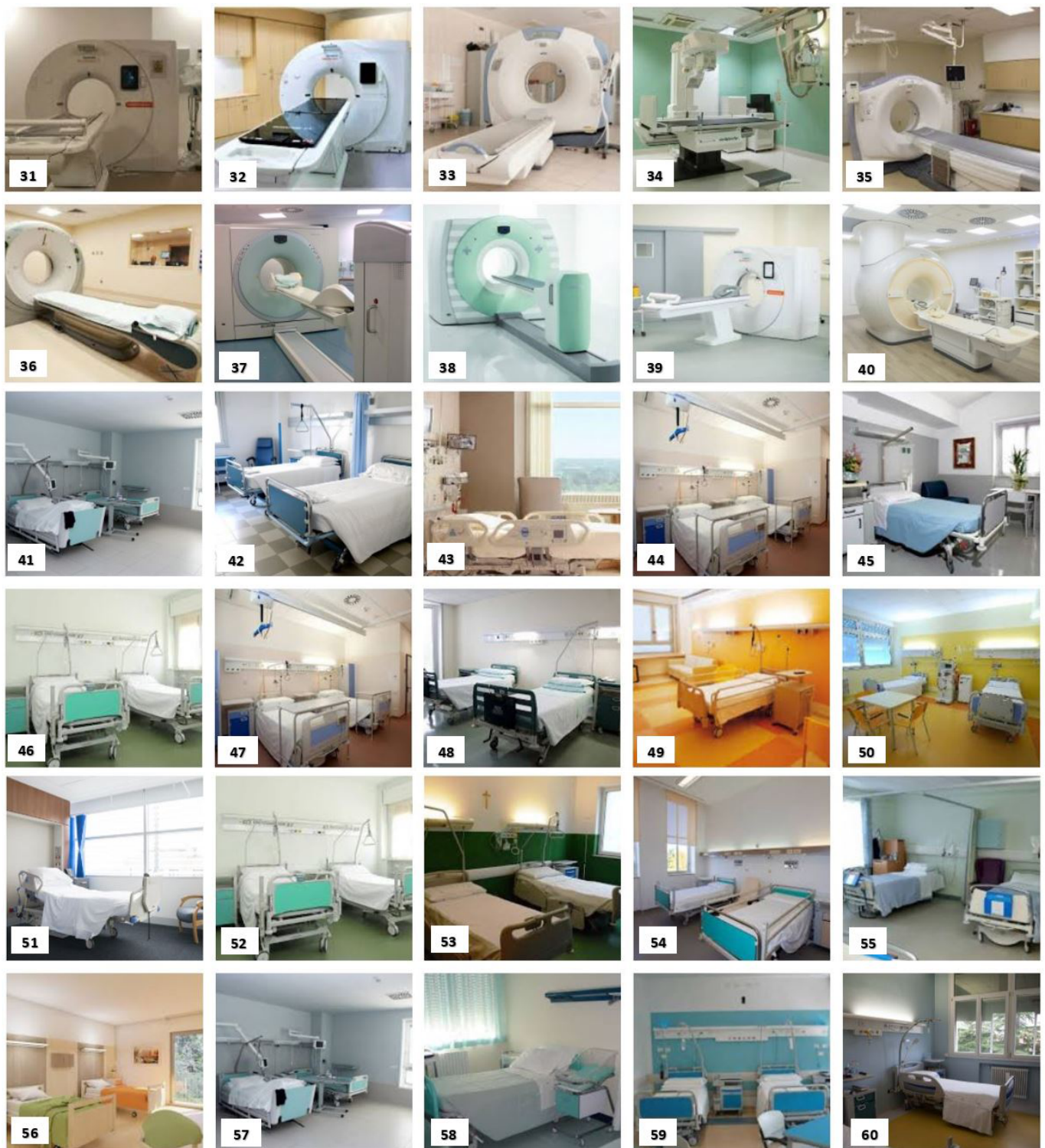
Figure A2. The images used in the test sets (31 to 40) of both custom models for the "diagnostic and therapeutic radiology" IU and the images used in the training sets (41 to 50) and test sets (51 to 60) of both custom models for the "hospitalisation" IU (from Google Images).
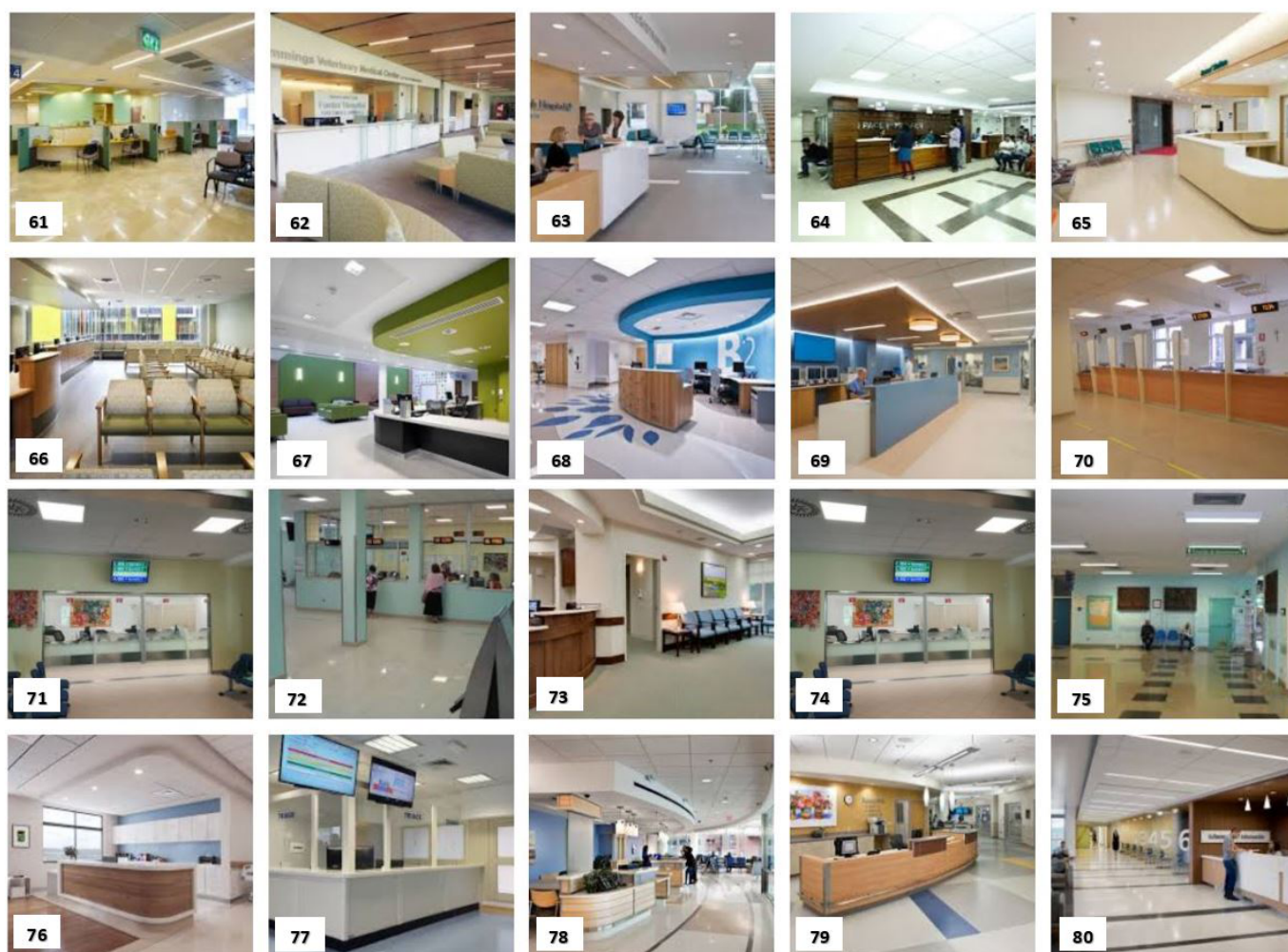
Figure A3. The images used in the training sets (61 to 70) and test sets (71 to 80) of both custom models for the "acceptance" IU (from Goole Images).

Appendix B

Here, we show a selection of the tables described in Section 3.1. Specifically, we selected the table that refers to the best results for each IU:

- Table A1 refers to the "surgery" IU and the results that were obtained with Google Vision API;
- Table A2 refers to the "radiology" IU and the results that were obtained with the Clarifai General Model;
- Table A3 refers to the "hospitalisation" IU and the results that were obtained with Microsoft Azure Cognitive Services;
- Table A4 refers to the "acceptance" IU and the results that were obtained with Google Vision API.

The tables only show the labels returned by each API that were significant for the recognition of the hospital setting under consideration. For example, generic labels, such as "indoor" or "place" (certainly correct for each image provided, but not needed to classify the environment), are not present in the tables. Furthermore, in this context, the "objects" in the images that were identified by the systems were not considered. Indeed, in some cases, they were returned separately (Google); in other cases, they were merged with other labels (Clarifai and Microsoft). The labels that correctly classified the hospital setting or contributed to a correct classification are highlighted in green. The labels that led to an incorrect recognition are highlighted in red.

Table A1. Percentage of confidence obtained with Google Vision API for the "surgery" IU. Each column corresponds to Image 1 to 20.

	Im. 1	Im. 2	Im. 3	Im. 4	Im. 5	Im. 6	Im. 7	Im. 8	Im. 9	Im. 10	Im. 11	Im. 12	Im. 13	Im. 14	Im. 15	Im. 16	Im. 17	Im. 18	Im. 19	Im. 20
Hospital	93%	97%	/	98%	79%	95%	94%	95%	96%	97%	97%	97%	97%	91%	89%	96%	92%	92%	86%	96%
Medical Equipment	91%	97%	81%	94%	/	97%	96%	98%	92%	98%	97%	97%	98%	95%	94%	96%	96%	/	89%	98%
Room	93%	91%	83%	95%	92%	84%	94%	94%	89%	93%	95%	93%	92%	92%	85%	96%	86%	81%	89%	92%
Operating Theater	96%	90%	/	96%	79%	63%	98%	98%	87%	96%	77%	78%	88%	57%	70%	98%	55%	83%	89%	90%
Medical	64%	80%	/	86%	/	92%	85%	94%	76%	96%	88%	90%	95%	59%	/	88%	79%	96%	77%	93%

Table A2. Percentage of confidence obtained with the Clarifai General Model for the "radiology" IU. Each column corresponds to Image 21 to 40.

	Im. 21	Im. 22	Im. 23	Im. 24	Im. 25	Im. 26	Im. 27	Im. 28	Im. 29	Im. 30	Im. 31	Im. 32	Im. 33	Im. 34	Im. 35	Im. 36	Im. 37	Im. 38	Im. 39	Im. 40
Hospital	99.5%	99.2%	94.3%	97.5%	98.8%	99.4%	93.8%	99.2%	98.7%	97.9%	99.4%	96.3%	99.3%	99.8%	98.4%	98.8%	98.1%	85.8%	99.4%	99.0%
Medicine	99.3%	99.1%	96.6%	98.0%	98.8%	99.4%	97.6%	99.0%	95.7%	98.5%	99.5%	98.0%	99.0%	99.7%	97.5%	98.1%	98.4%	96.7%	99.5%	98.9%
Equipment	98.9%	/	93.9%	94.9%	98.4%	97.5%	95.7%	95.0%	89.8%	94.3%	98.5%	97.4%	94.9%	97.7%	90.9%	93.3%	/	97.0%	97.1%	/
Clinic	98.8%	/	85.8%	94.6%	97.0%	99.0%	92.3%	98.1%	96.2%	97.3%	98.4%	93.6%	98.5%	98.3%	94.5%	97.7%	/	81.6%	98.7%	96.3%
Surgery	98.2%	98.9%	94.9%	95.9%	98.3%	97.2%	/	98.2%	97.4%	93.2%	98.0%	93.4%	95.2%	99.8%	95.3%	/	94.4%	86.0%	97.6%	97.0%
Room	96.4%	96.7%	/	/	/	94.5%	95.2%	94.0%	89.6%	96.6%	95.1%	/	98.4%	98.0%	99.1%	97.9%	97.7%	/	98.1%	98.5%
Scrutiny	/	91.3%	/	91.8%	92.5%	97.3%	/	96.7%	97.0%	95.9%	97.2%	/	/	97.9%	93.0%	/	/	/	95.7%	94.1%
Radiography	/	91.7%	/	/	/	96.5%	/	/	/	/	98.9%	/	/	94.5%	/	/	91.5%	/	95.6%	/
Radiology	/	90.8%	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
Diagnosis	/	93.7%	91.0%	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	85.0%
Treatment	/	88.5%	86.0%	91.6%	91.6%	/	/	/	/	/	95.7%	/	/	/	/	/	/	82.7%	/	90.2%
Emergency	/	/	88.3%	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
Operating Room	/	/	/	/	/	/	/	/	/	/	/	/	/	98.8%	/	/	/	/	/	/

Table A3. Percentage of confidence obtained with Microsoft Azure Cognitive Services for the "hospitalisation" IU. Each column corresponds to Image 41 to 60 and each row is related to a different label that was returned by the model.

	Im. 41	Im. 42	Im. 43	Im. 44	Im. 45	Im. 46	Im. 47	Im. 48	Im. 49	Im. 50	Im. 51	Im. 52	Im. 53	Im. 54	Im. 55	Im. 56	Im. 57	Im. 58	Im. 59	Im. 60
Medical Equipment	96.2%	95.7%	/	92.6%	55.5%	97.3%	92.6%	94.6%	/	/	86.6%	97.8%	92.7%	74.0%	87.2%	/	92.7%	84.7%	95.0%	66.1%
Furniture	18.0%	92.4%	92.9%	17.6%	40.7%	33.2%	17.6%	22.4%	91.4%	23.0%	29.1%	35.6%	88.3%	28.3%	41.0%	69.1%	33.8%	97.0%	36.3%	93.8%
Bedroom	64.7%	48.5%	/	/	58.4%	/	/	/	/	/	70.6%	/	54.3%	47.8%	53.8%	57.5%	39.1%	39.5%	/	48.1%
Clinic	51.5%	/	/	/	/	63.9%	/	54.4%	/	/	/	68.9%	52.4%	/	/	/	/	/	/	/
Hospital	79.2%	77.1%	/	75.9%	/	86.0%	75.9%	82.2%	/	/	56.3%	88.5%	80.3%	/	65.9%	/	74.3%	56.2%	70.1%	/
Room	76.4%	55.4%	/	72.0%	96.2%	76.0%	72.0%	80.4%	73.2%	84.0%	90.3%	43.5%	77.5%	41.0%	92.2%	93.3%	78.7%	53.5%	77.1%	81.8%
Hotel	/	/	/	/	76.6%	/	/	/	68.5%	/	/	/	71.8%	82.2%	/	95.2%	/	/	/	71.7%
Plumbing Fixture	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	72.2%
Bathroom	/	/	/	56.7%	/	/	56.7%	/	54.6%	/	/	/	/	/	/	79.9%	/	/	/	86.3%
House	/	/	/	60.6%	89.4%	/	60.6%	/	/	53.6%	/	/	/	70.9%	/	89.6%	/	/	/	75.1%
Hospital Room	/	/	/	77.0%	/	/	77.0%	60.4%	/	/	/	/	/	/	/	/	82.1%	/	/	/
Office Building	/	/	/	/	/	/	/	66.3%	/	76.4%	/	/	/	/	/	/	/	/	69.7%	/
Operating Theatre	/	/	/	/	/	54.6%	/	/	/	/	/	61.0%	/	/	/	/	/	/	/	/

Table A4. Percentage of confidence obtained with Google Vision API for the "acceptance" IU. Each column corresponds to Image 61 to 80 and each row is related to a different label that was returned by the model.

	Im. 61	Im. 62	Im. 63	Im. 64	Im. 65	Im. 66	Im. 67	Im. 68	Im. 69	Im. 70	Im. 71	Im. 72	Im. 73	Im. 74	Im. 75	Im. 76	Im. 77	Im. 78	Im. 79	Im. 80
Room	80%	92%	74%	71%	93%	92%	89%	89%	81%	79%	80%	66%	94%	80%	83%	92%	79%	80%	88%	74%
Waiting Room	66%	/	/	/	76%	/	/	/	/	82%	/	/	89%	65%	/	/	/	/	/	/
Office	56%	87%	89%	61%	71%	59%	88%	92%	92%	51%	60%	/	84%	56%	/	90%	60%	89%	/	88%
Hospital	/	/	66%	/	51%	/	/	/	/	/	86%	62%	/	70%	50%	/	/	/	/	/
Reception	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	60%	/

Appendix C

The tables in this appendix describe the datasets that were used in this study:

- Table A5 shows the dataset employed to test the three general models and the images are those that were used to train and test the two versions of the model described in Section 2.3;
- Table A6 describes the additional 40 images that were used for the second version of the model described in Section 2.3;
- Table A7 shows the dataset used for the first version of the Detectron2 model;
- Table A8 shows the dataset used for the second version of the Detectron2 model.

For each table, the first column refers to the IU, the second refers to the name of the image and the third refers to the size of the photograph (in pixels). The fourth and fifth columns refer to the horizontal and vertical resolutions of the images (in dpi), respectively, and the last column reports the bit depth for each image.

Table A5. Dataset employed to test the three general models and to train and test the two versions of the model described in Section 2.3.

IU	Name	Size (Pixel)	Horizontal Resolution (dpi)	Vertical Resolution (dpi)	Bit Depth
Surgery	Image 1	640 × 427	72	72	24
	Image 2	275 × 183	96	96	24
	Image 3	118 × 510	96	96	24
	Image 4	880 × 586	96	96	24
	Image 5	258 × 195	96	96	24
	Image 6	275 × 183	96	96	24
	Image 7	259 × 194	96	96	24
	Image 8	487 × 325	72	72	24
	Image 9	475 × 316	96	96	24
	Image 10	275 × 183	96	96	24
	Image 11	225 × 225	96	96	24
	Image 12	273 × 185	96	96	24
	Image 13	275 × 183	96	96	24
	Image 14	225 × 225	96	96	24
	Image 15	850 × 510	96	96	24
	Image 16	550 × 413	96	96	24
	Image 17	275 × 183	96	96	24
	Image 18	341 × 148	96	96	24
	Image 19	275 × 183	96	96	24
	Image 20	304 × 166	96	96	24
Radiology	Image 21	275 × 183	96	96	24
	Image 22	2254 × 2056	72	72	24
	Image 23	800 × 533	96	96	24
	Image 24	267 × 189	96	96	24
	Image 25	274 × 184	96	96	24
	Image 26	270 × 187	96	96	24
	Image 27	276 × 183	96	96	24
	Image 28	259 × 194	96	96	24
	Image 29	243 × 207	96	96	24
	Image 30	275 × 183	96	96	24
	Image 31	300 × 168	96	96	24
	Image 32	281 × 180	96	96	24
	Image 33	276 × 183	96	96	24
	Image 34	225 × 225	96	96	24
	Image 35	275 × 183	96	96	24
	Image 36	286 × 176	96	96	24
	Image 37	2048 × 1536	300	300	24
	Image 38	245 × 206	96	96	24
	Image 39	259 × 194	96	96	24
	Image 40	870 × 575	96	96	24

Table A5. *Cont.*

IU	Name	Size (Pixel)	Horizontal Resolution (dpi)	Vertical Resolution (dpi)	Bit Depth
Hospitalisation	Image 41	800 × 600	96	96	24
	Image 42	700 × 525	72	72	24
	Image 43	301 × 167	96	96	24
	Image 44	275 × 183	96	96	24
	Image 45	275 × 183	96	96	24
	Image 46	299 × 168	96	96	24
	Image 47	275 × 183	96	96	24
	Image 48	275 × 183	96	96	24
	Image 49	307 × 164	96	96	24
	Image 50	275 × 183	96	96	24
Hospitalisation	Image 51	1000 × 667	96	96	24
	Image 52	299 × 168	96	96	24
	Image 53	275 × 183	96	96	24
	Image 54	275 × 183	96	96	24
	Image 55	284 × 177	96	96	24
	Image 56	270 × 187	96	96	24
	Image 57	259 × 232	96	96	24
	Image 58	259 × 194	96	96	24
	Image 59	217 × 232	96	96	24
	Image 60	1200 × 800	96	96	24
Acceptance	Image 61	275 × 183	96	96	24
	Image 62	259 × 194	96	96	24
	Image 63	275 × 183	96	96	24
	Image 64	275 × 183	96	96	24
	Image 65	275 × 183	96	96	24
	Image 66	270 × 187	96	96	24
	Image 67	230 × 219	96	96	24
	Image 68	274 × 184	96	96	24
	Image 69	276 × 182	96	96	24
	Image 70	252 × 200	96	96	24
	Image 71	259 × 194	96	96	24
	Image 72	259 × 194	96	96	24
	Image 73	275 × 183	96	96	24
	Image 74	512 × 384	96	96	24
	Image 75	319 × 158	96	96	24
	Image 76	290 × 174	96	96	24
	Image 77	300 × 168	96	96	24
	Image 78	255 × 197	96	96	24
	Image 79	275 × 183	96	96	24
	Image 80	275 × 183	96	96	24

Table A6. Additional 40 images used for the second version of the model described in Section 2.3.

IU	Name	Size (Pixel)	Horizontal Resolution (dpi)	Vertical Resolution (dpi)	Bit Depth
Surgery	Image 81	864 × 534	96	96	24
	Image 82	1600 × 1077	96	96	24
	Image 83	288 × 175	96	96	24
	Image 84	299 × 168	96	96	24
	Image 85	275 × 183	96	96	24
	Image 86	275 × 183	96	96	24
	Image 87	289 × 175	96	96	24
	Image 88	1800 × 1200	300	300	24
	Image 89	261 × 193	96	96	24
	Image 90	921 × 617	96	96	24
Radiology	Image 91	1024 × 576	72	72	24
	Image 92	800 × 450	72	72	24
	Image 93	751 × 401	96	96	24
	Image 94	1000 × 665	180	180	24
	Image 95	259 × 194	96	96	24
	Image 96	251 × 201	96	96	24
	Image 97	225 × 225	96	96	24
	Image 98	275 × 183	96	96	24
	Image 99	300 × 168	96	96	24
	Image 100	260 × 194	96	96	24
Hospitalisation	Image 101	600 × 338	72	72	24
	Image 102	986 × 657	96	96	24
	Image 103	901 × 568	96	96	24
	Image 104	1779 × 1192	300	300	24
	Image 105	283 × 178	96	96	24
	Image 106	1024 × 768	96	96	24
	Image 107	667 × 500	96	96	24
	Image 108	840 × 480	96	96	24
	Image 109	259 × 194	96	96	24
	Image 110	312 × 161	96	96	24
Acceptance	Image 111	194 × 259	96	96	24
	Image 112	279 × 180	96	96	24
	Image 113	275 × 183	96	96	24
	Image 114	276 × 183	96	96	24
	Image 115	300 × 168	96	96	24
	Image 116	374 × 135	96	96	24
	Image 117	259 × 194	96	96	24
	Image 118	301 × 168	96	96	24
	Image 119	260 × 194	96	96	24
	Image 120	259 × 194	96	96	24

Table A7. Dataset used for the first version of the Detectron2 model (3 IUs).

IU	Name	Size (Pixel)	Horizontal Resolution (dpi)	Vertical Resolution (dpi)	Bit Depth
Hospitalisation	Image 1	297 × 170	96	96	24
	Image 2	264 × 191	96	96	24
	Image 3	270 × 187	96	96	24
	Image 4	275 × 183	96	96	24
	Image 5	300 × 168	96	96	24
	Image 6	276 × 183	96	96	24
	Image 7	259 × 194	96	96	24
	Image 8	310 × 163	96	96	24
	Image 9	300 × 168	96	96	24
	Image 10	285 × 177	96	96	24
	Image 11	259 × 194	96	96	24
	Image 12	275 × 183	96	96	24
	Image 13	299 × 168	96	96	24
	Image 14	275 × 183	96	96	24
	Image 15	340 × 148	96	96	24
	Image 16	314 × 160	96	96	24
	Image 17	307 × 164	96	96	24
	Image 18	194 × 259	96	96	24
	Image 19	325 × 155	96	96	24
	Image 20	259 × 194	96	96	24
	Image 21	275 × 183	96	96	24
	Image 22	361 × 140	96	96	24
	Image 23	314 × 161	96	96	24
	Image 24	275 × 183	96	96	24
	Image 25	275 × 183	96	96	24
	Image 26	276 × 183	96	96	24
	Image 27	275 × 183	96	96	24
	Image 28	259 × 194	96	96	24
	Image 29	275 × 183	96	96	24
	Image 30	275 × 183	96	96	24
	Image 31	275 × 183	96	96	24
	Image 32	260 × 194	96	96	24
	Image 33	261 × 193	96	96	24
	Image 34	275 × 183	96	96	24
	Image 35	275 × 183	96	96	24
	Image 36	274 × 184	96	96	24
	Image 37	275 × 183	96	96	24
	Image 38	300 × 168	96	96	24
	Image 39	275 × 183	96	96	24
	Image 40	275 × 183	96	96	24
Radiology	Image 1	257 × 196	96	96	24
	Image 2	275 × 183	96	96	24
	Image 3	261 × 193	96	96	24
	Image 4	233 × 216	96	96	24
	Image 5	259 × 194	96	96	24
	Image 6	300 × 168	96	96	24
	Image 7	292 × 173	96	96	24
	Image 8	311 × 162	96	96	24
	Image 9	299 × 168	96	96	24
	Image 10	273 × 185	96	96	24
	Image 11	290 × 174	96	96	24
	Image 12	275 × 183	96	96	24
	Image 13	300 × 168	96	96	24
	Image 14	259 × 194	96	96	24
	Image 15	301 × 168	96	96	24
	Image 16	270 × 187	96	96	24
	Image 17	183 × 275	96	96	24

Table A7. *Cont.*

IU	Name	Size (Pixel)	Horizontal Resolution (dpi)	Vertical Resolution (dpi)	Bit Depth
Radiology	Image 18	299 × 168	96	96	24
	Image 19	356 × 141	96	96	24
	Image 20	270 × 186	96	96	24
	Image 21	300 × 168	96	96	24
	Image 22	308 × 164	96	96	24
	Image 23	304 × 166	96	96	24
	Image 24	275 × 183	96	96	24
	Image 25	275 × 183	96	96	24
	Image 26	300 × 168	96	96	24
	Image 27	251 × 201	96	96	24
	Image 28	283 × 178	96	96	24
	Image 29	259 × 194	96	96	24
	Image 30	301 × 168	96	96	24
	Image 31	300 × 168	96	96	24
	Image 32	299 × 168	96	96	24
	Image 33	271 × 186	96	96	24
	Image 34	248 × 203	96	96	24
	Image 35	244 × 206	96	96	24
	Image 36	244 × 206	96	96	24
	Image 37	276 × 183	96	96	24
	Image 38	299 × 168	96	96	24
	Image 39	288 × 175	96	96	24
	Image 40	302 × 167	96	96	24
Surgery	Image 1	275 × 183	96	96	24
	Image 2	275 × 183	96	96	24
	Image 3	168 × 188	96	96	24
	Image 4	292 × 173	96	96	24
	Image 5	240 × 210	96	96	24
	Image 6	275 × 183	96	96	24
	Image 7	291 × 173	96	96	24
	Image 8	275 × 183	96	96	24
	Image 9	318 × 159	96	96	24
	Image 10	194 × 259	96	96	24
	Image 11	259 × 194	96	96	24
	Image 12	269 × 187	96	96	24
	Image 13	256 × 197	96	96	24
	Image 14	300 × 168	96	96	24
	Image 15	254 × 198	96	96	24
	Image 16	324 × 155	96	96	24
	Image 17	259 × 194	96	96	24
	Image 18	258 × 195	96	96	24
	Image 19	318 × 159	96	96	24
	Image 20	259 × 194	96	96	24
	Image 21	275 × 183	96	96	24
	Image 22	286 × 176	96	96	24
	Image 23	275 × 183	96	96	24
	Image 24	258 × 195	96	96	24
	Image 25	300 × 168	96	96	24
	Image 26	264 × 191	96	96	24
	Image 27	299 × 168	96	96	24
	Image 28	295 × 171	96	96	24
	Image 29	259 × 194	96	96	24
	Image 31	340 × 148	96	96	24
	Image 32	274 × 184	96	96	24
	Image 33	275 × 183	96	96	24
	Image 34	329 × 153	96	96	24
	Image 35	275 × 183	96	96	24

<center>**Table A7.** *Cont.*</center>

IU	Name	Size (Pixel)	Horizontal Resolution (dpi)	Vertical Resolution (dpi)	Bit Depth
Surgery	Image 36	275 × 183	96	96	24
	Image 37	259 × 194	96	96	24
	Image 38	259 × 194	96	96	24
	Image 39	251 × 201	96	96	24
Ambulance	Image 1	275 × 183	96	96	24
	Image 2	229 × 220	96	96	24
	Image 3	275 × 183	96	96	24
	Image 4	276 × 183	96	96	24
	Image 5	300 × 168	96	96	24
	Image 6	194 × 259	96	96	24
	Image 7	244 × 206	96	96	24
	Image 8	274 × 184	96	96	24
	Image 9	276 × 183	96	96	24
	Image 10	259 × 194	96	96	24
	Image 11	325 × 155	96	96	24
	Image 12	260 × 194	96	96	24
	Image 13	274 × 184	96	96	24
	Image 14	275 × 183	96	96	24
	Image 15	259 × 194	96	96	24
	Image 16	260 × 194	96	96	24
	Image 17	347 × 145	96	96	24
	Image 18	275 × 183	96	96	24
	Image 19	275 × 183	96	96	24
	Image 20	225 × 225	96	96	24
	Image 21	300 × 168	96	96	24
	Image 22	268 × 188	96	96	24
	Image 23	358 × 141	96	96	24
	Image 24	278 × 181	96	96	24
	Image 25	290 × 174	96	96	24
	Image 26	275 × 183	96	96	24
	Image 27	319 × 158	96	96	24
	Image 28	275 × 183	96	96	24
	Image 29	318 × 159	96	96	24
	Image 30	275 × 183	96	96	24
	Image 31	276 × 183	96	96	24
	Image 32	272 × 185	96	96	24
	Image 33	268 × 188	96	96	24
	Image 34	259 × 194	96	96	24
	Image 35	254 × 198	96	96	24
	Image 36	274 × 184	96	96	24
	Image 37	225 × 225	96	96	24
	Image 38	301 × 168	96	96	24
	Image 39	259 × 194	96	96	24
	Image 40	356 × 141	96	96	24
Analysis Laboratory	Image 1	250 × 167	96	96	24
	Image 2	331 × 152	96	96	24
	Image 3	318 × 159	96	96	24
	Image 4	274 × 184	96	96	24
	Image 5	200 × 150	96	96	24
	Image 6	267 × 189	96	96	24
	Image 7	299 × 168	96	96	24
	Image 8	320 × 158	96	96	24
	Image 9	275 × 183	96	96	24
	Image 10	300 × 168	96	96	24
	Image 11	271 × 186	96	96	24
	Image 12	240 × 200	96	96	24

Table A7. *Cont.*

IU	Name	Size (Pixel)	Horizontal Resolution (dpi)	Vertical Resolution (dpi)	Bit Depth
Analysis Laboratory	Image 13	313 × 161	96	96	24
	Image 14	259 × 194	96	96	24
	Image 15	259 × 194	96	96	24
	Image 16	268 × 188	96	96	24
	Image 17	319 × 158	96	96	24
	Image 18	275 × 183	96	96	24
	Image 19	276 × 183	96	96	24
	Image 20	275 × 183	96	96	24
	Image 21	275 × 183	96	96	24
	Image 22	264 × 191	96	96	24
	Image 23	276 × 183	96	96	24
	Image 24	259 × 194	96	96	24
	Image 25	305 × 165	96	96	24
	Image 26	370 × 136	96	96	24
	Image 27	382 × 132	96	96	24
	Image 28	321 × 157	96	96	24
	Image 29	300 × 168	96	96	24
	Image 30	263 × 192	96	96	24
	Image 31	330 × 153	96	96	24
	Image 32	300 × 168	96	96	24
	Image 33	322 × 156	96	96	24
	Image 34	250 × 202	96	96	24
	Image 35	299 × 169	96	96	24
	Image 36	402 × 125	96	96	24
	Image 37	262 × 193	96	96	24
	Image 38	284 × 177	96	96	24
	Image 39	304 × 166	96	96	24
	Image 40	259 × 194	96	96	24
Hospitalisation	Image 1	297 × 170	96	96	24
	Image 2	264 × 191	96	96	24
	Image 3	270 × 187	96	96	24
	Image 4	275 × 183	96	96	24
	Image 5	300 × 168	96	96	24
	Image 6	276 × 183	96	96	24
	Image 7	259 × 194	96	96	24
	Image 8	310 × 163	96	96	24
	Image 9	300 × 168	96	96	24
	Image 10	285 × 177	96	96	24
	Image 11	259 × 194	96	96	24
	Image 12	275 × 183	96	96	24
	Image 13	299 × 168	96	96	24
	Image 14	275 × 183	96	96	24
	Image 15	340 × 148	96	96	24
	Image 16	314 × 160	96	96	24
	Image 17	307 × 164	96	96	24
	Image 18	194 × 259	96	96	24
	Image 19	325 × 155	96	96	24
	Image 20	259 × 194	96	96	24
	Image 21	275 × 183	96	96	24
	Image 22	361 × 140	96	96	24
	Image 23	314 × 161	96	96	24
	Image 24	275 × 183	96	96	24
	Image 25	275 × 183	96	96	24
	Image 26	276 × 183	96	96	24
	Image 27	275 × 183	96	96	24
	Image 28	259 × 194	96	96	24
	Image 29	275 × 183	96	96	24

Table A7. *Cont.*

IU	Name	Size (Pixel)	Horizontal Resolution (dpi)	Vertical Resolution (dpi)	Bit Depth
Hospitalisation	Image 30	275 × 183	96	96	24
	Image 31	275 × 183	96	96	24
	Image 32	260 × 194	96	96	24
	Image 33	261 × 193	96	96	24
	Image 34	275 × 183	96	96	24
	Image 35	275 × 183	96	96	24
	Image 36	274 × 184	96	96	24
	Image 37	275 × 183	96	96	24
	Image 38	300 × 168	96	96	24
	Image 39	275 × 183	96	96	24
	Image 40	275 × 183	96	96	24
Intensive Therapy	Image 1	300 × 168	96	96	24
	Image 2	301 × 168	96	96	24
	Image 3	275 × 183	96	96	24
	Image 4	263 × 192	96	96	24
	Image 5	259 × 194	96	96	24
	Image 6	303 × 166	96	96	24
	Image 7	275 × 183	96	96	24
	Image 8	259 × 194	96	96	24
	Image 9	259 × 194	96	96	24
	Image 10	259 × 194	96	96	24
	Image 11	300 × 168	96	96	24
	Image 12	259 × 194	96	96	24
	Image 13	299 × 168	96	96	24
	Image 14	299 × 168	96	96	24
	Image 15	259 × 194	96	96	24
	Image 16	275 × 183	96	96	24
	Image 17	275 × 183	96	96	24
	Image 18	299 × 168	96	96	24
	Image 19	276 × 183	96	96	24
	Image 20	335 × 150	96	96	24
	Image 21	275 × 183	96	96	24
	Image 22	300 × 168	96	96	24
	Image 23	318 × 159	96	96	24
	Image 24	268 × 188	96	96	24
	Image 25	299 × 168	96	96	24
	Image 26	299 × 168	96	96	24
	Image 27	305 × 165	96	96	24
	Image 28	275 × 183	96	96	24
	Image 29	275 × 183	96	96	24
	Image 30	301 × 168	96	96	24
	Image 31	275 × 183	96	96	24
	Image 32	259 × 194	96	96	24
	Image 33	299 × 168	96	96	24
	Image 34	259 × 194	96	96	24
	Image 35	256 × 197	96	96	24
	Image 36	268 × 188	96	96	24
	Image 37	278 × 181	96	96	24
	Image 38	275 × 183	96	96	24
	Image 39	275 × 183	96	96	24
	Image 40	300 × 168	96	96	24
Medical Clinic	Image 1	286 × 176	96	96	24
	Image 2	273 × 185	96	96	24
	Image 3	259 × 195	96	96	24

<div align="center">

Table A7. *Cont.*

</div>

IU	Name	Size (Pixel)	Horizontal Resolution (dpi)	Vertical Resolution (dpi)	Bit Depth
Medical Clinic	Image 4	301 × 167	96	96	24
	Image 5	360 × 140	96	96	24
	Image 6	275 × 183	96	96	24
	Image 7	286 × 176	96	96	24
	Image 8	275 × 183	96	96	24
	Image 9	277 × 182	96	96	24
	Image 10	275 × 183	96	96	24
	Image 11	275 × 183	96	96	24
	Image 12	275 × 183	96	96	24
	Image 13	301 × 167	96	96	24
	Image 14	275 × 183	96	96	24
	Image 15	383 × 132	96	96	24
	Image 16	275 × 183	96	96	24
	Image 17	259 × 194	96	96	24
	Image 18	275 × 183	96	96	24
	Image 19	275 × 183	96	96	24
	Image 20	194 × 259	96	96	24
	Image 21	259 × 194	96	96	24
	Image 22	274 × 184	96	96	24
	Image 23	259 × 194	96	96	24
	Image 24	275 × 183	96	96	24
	Image 25	330 × 153	96	96	24
	Image 26	259 × 194	96	96	24
	Image 27	306 × 165	96	96	24
	Image 28	300 × 168	96	96	24
	Image 29	194 × 259	96	96	24
	Image 30	259 × 194	96	96	24
	Image 31	183 × 276	96	96	24
	Image 32	275 × 183	96	96	24
	Image 33	259 × 194	96	96	24
	Image 34	259 × 194	96	96	24
	Image 35	247 × 204	96	96	24
	Image 36	275 × 183	96	96	24
	Image 37	194 × 259	96	96	24
	Image 38	275 × 183	96	96	24
	Image 39	273 × 185	96	96	24
	Image 40	316 × 160	96	96	24
Radiology	Image 1	257 × 196	96	96	24
	Image 2	275 × 183	96	96	24
	Image 3	261 × 193	96	96	24
	Image 4	233 × 216	96	96	24
	Image 5	259 × 194	96	96	24
	Image 6	300 × 168	96	96	24
	Image 7	292 × 173	96	96	24
	Image 8	311 × 162	96	96	24
	Image 9	299 × 168	96	96	24
	Image 10	273 × 185	96	96	24
	Image 11	290 × 174	96	96	24
	Image 12	275 × 183	96	96	24
	Image 13	300 × 168	96	96	24
	Image 14	259 × 194	96	96	24
	Image 15	301 × 168	96	96	24
	Image 16	270 × 187	96	96	24
	Image 17	183 × 275	96	96	24

Table A7. *Cont.*

IU	Name	Size (Pixel)	Horizontal Resolution (dpi)	Vertical Resolution (dpi)	Bit Depth
Radiology	Image 18	299 × 168	96	96	24
	Image 19	356 × 141	96	96	24
	Image 20	270 × 186	96	96	24
	Image 21	300 × 168	96	96	24
	Image 22	308 × 164	96	96	24
	Image 23	304 × 166	96	96	24
	Image 24	275 × 183	96	96	24
	Image 25	275 × 183	96	96	24
	Image 26	300 × 168	96	96	24
	Image 27	251 × 201	96	96	24
	Image 28	283 × 178	96	96	24
	Image 29	259 × 194	96	96	24
	Image 30	301 × 168	96	96	24
	Image 31	300 × 168	96	96	24
	Image 32	299 × 168	96	96	24
	Image 33	271 × 186	96	96	24
	Image 34	248 × 203	96	96	24
	Image 35	244 × 206	96	96	24
	Image 36	244 × 206	96	96	24
	Image 37	276 × 183	96	96	24
	Image 38	299 × 168	96	96	24
	Image 39	288 × 175	96	96	24
	Image 40	302 × 167	96	96	24
Rehabilitation and Physiotherapy	Image 1	348 × 145	96	96	24
	Image 2	259 × 194	96	96	24
	Image 3	259 × 194	96	96	24
	Image 4	275 × 183	96	96	24
	Image 5	277 × 182	96	96	24
	Image 6	300 × 168	96	96	24
	Image 7	259 × 194	96	96	24
	Image 8	275 × 183	96	96	24
	Image 9	259 × 194	96	96	24
	Image 10	275 × 183	96	96	24
	Image 11	297 × 170	96	96	24
	Image 12	243 × 208	96	96	24
	Image 13	259 × 194	96	96	24
	Image 14	275 × 183	96	96	24
	Image 15	294 × 171	96	96	24
	Image 16	300 × 168	96	96	24
	Image 17	259 × 194	96	96	24
	Image 18	248 × 203	96	96	24
	Image 19	275 × 183	96	96	24
	Image 20	259 × 194	96	96	24
	Image 21	300 × 168	96	96	24
	Image 22	329 × 153	96	96	24
	Image 23	300 × 168	96	96	24
	Image 24	248 × 203	96	96	24
	Image 25	259 × 194	96	96	24
	Image 26	259 × 194	96	96	24
	Image 27	321 × 157	96	96	24
	Image 28	194 × 259	96	96	24
	Image 29	275 × 183	96	96	24
	Image 30	372 × 135	96	96	24
	Image 31	259 × 194	96	96	24

Table A7. *Cont.*

IU	Name	Size (Pixel)	Horizontal Resolution (dpi)	Vertical Resolution (dpi)	Bit Depth
Rehabilitation and Physiotherapy	Image 32	259 × 194	96	96	24
	Image 33	259 × 194	96	96	24
	Image 34	316 × 159	96	96	24
	Image 35	300 × 168	96	96	24
	Image 36	225 × 225	96	96	24
	Image 37	259 × 194	96	96	24
	Image 38	275 × 183	96	96	24
	Image 39	275 × 184	96	96	24
	Image 40	275 × 183	96	96	24
Surgery	Image 1	275 × 183	96	96	24
	Image 2	275 × 183	96	96	24
	Image 3	168 × 188	96	96	24
	Image 4	292 × 173	96	96	24
	Image 5	240 × 210	96	96	24
	Image 6	275 × 183	96	96	24
	Image 7	291 × 173	96	96	24
	Image 8	275 × 183	96	96	24
	Image 9	318 × 159	96	96	24
	Image 10	194 × 259	96	96	24
	Image 11	259 × 194	96	96	24
	Image 12	269 × 187	96	96	24
	Image 13	256 × 197	96	96	24
	Image 14	300 × 168	96	96	24
	Image 15	254 × 198	96	96	24
	Image 16	324 × 155	96	96	24
	Image 17	259 × 194	96	96	24
	Image 18	258 × 195	96	96	24
	Image 19	318 × 159	96	96	24
	Image 20	259 × 194	96	96	24
	Image 21	275 × 183	96	96	24
	Image 22	286 × 176	96	96	24
	Image 23	275 × 183	96	96	24
	Image 24	258 × 195	96	96	24
	Image 25	300 × 168	96	96	24
	Image 26	264 × 191	96	96	24
	Image 27	299 × 168	96	96	24
	Image 28	295 × 171	96	96	24
	Image 29	259 × 194	96	96	24
	Image 30	275 × 183	96	96	24
	Image 31	340 × 148	96	96	24
	Image 32	274 × 184	96	96	24
	Image 33	275 × 183	96	96	24
	Image 34	329 × 153	96	96	24
	Image 35	275 × 183	96	96	24
	Image 36	275 × 183	96	96	24
	Image 37	259 × 194	96	96	24
	Image 38	259 × 194	96	96	24
	Image 39	251 × 201	96	96	24
	Image 40	343 × 147	96	96	24
Toilet	Image 1	194 × 259	96	96	24
	Image 2	194 × 259	96	96	24
	Image 3	276 × 183	96	96	24
	Image 4	194 × 259	96	96	24
	Image 5	286 × 176	96	96	24

Table A8. Dataset used for the second version of the Detectron2 model (9 IUs).

IU	Name	Size (Pixel)	Horizontal Resolution (dpi)	Vertical Resolution (dpi)	Bit Depth
Toilet	Image 6	268 × 188	96	96	24
	Image 7	286 × 176	96	96	24
	Image 8	242 × 208	96	96	24
	Image 9	259 × 194	96	96	24
	Image 10	259 × 194	96	96	24
	Image 11	286 × 176	96	96	24
	Image 12	290 × 174	96	96	24
	Image 13	194 × 259	96	96	24
	Image 14	259 × 194	96	96	24
	Image 15	194 × 259	96	96	24
	Image 16	225 × 225	96	96	24
	Image 17	285 × 177	96	96	24
	Image 18	275 × 183	96	96	24
	Image 19	300 × 168	96	96	24
	Image 20	275 × 183	96	96	24
	Image 21	259 × 194	96	96	24
	Image 22	259 × 194	96	96	24
	Image 23	276 × 183	96	96	24
	Image 24	286 × 176	96	96	24
	Image 25	275 × 183	96	96	24
	Image 26	225 × 224	96	96	24
	Image 27	259 × 194	96	96	24
	Image 28	183 × 275	96	96	24
	Image 29	225 × 225	96	96	24
	Image 30	262 × 193	96	96	24
	Image 31	183 × 275	96	96	24
	Image 32	177 × 284	96	96	24
	Image 33	264 × 191	96	96	24
	Image 34	194 × 259	96	96	24
	Image 35	262 × 192	96	96	24
	Image 36	278 × 181	96	96	24
	Image 37	259 × 194	96	96	24
	Image 38	259 × 194	96	96	24
	Image 39	194 × 259	96	96	24
	Image 40	300 × 168	96	96	24

References

1. Encyclopædia Britannica. Available online: https://www.britannica.com/ (accessed on 23 May 2022).
2. Associazione Italiana Ingegneri Clinici. AIIC Website. 2020. Available online: https://www.aiic.it/ (accessed on 16 March 2021). (In Italian)
3. Iadanza, E.; Luschi, A. Computer-aided facilities management in health care. In *Clinical Engineering Handbook*; Elsevier: Amsterdam, The Netherlands, 2020; pp. 42–51.
4. Luschi, A.; Marzi, L.; Miniati, R.; Iadanza, E. A custom decision-support information system for structural and technological analysis in healthcare. In Proceedings of the XIII Mediterranean Conference on Medical and Biological Engineering and Computing 2013, Seville, Spain, 25–28 September 2013; pp. 1350–1353.
5. Fragapane, G.; Hvolby, H.H.; Sgarbossa, F.; Strandhagen, J.O. Autonomous mobile robots in hospital logistics. In Proceedings of the IFIP International Conference on Advances in Production Management Systems, Novi Sad, Serbia, 30 August–3 September 2020; pp. 672–679.
6. Robotics4EU Project. 2021. Available online: https://www.robotics4eu.eu/ (accessed on 23 May 2022).
7. Odin is a European Multi-centre Pilot Study Focused on the Enhancement of Hospital Safety, Productivity and Quality. Available online: https://www.odin-smarthospitals.eu/ (accessed on 23 May 2022).
8. President of the Italian Republic. DPR 14 Gennaio 1997. 1997. Available online: https://www.gazzettaufficiale.it/eli/gu/1997/02/20/42/so/37/sg/pdf (accessed on 16 March 2021). (In Italian)
9. Cicchetti, A. *L'organizzazione Dell'ospedale. Fra Tradizione e Strategie per il Futuro*; Vita e Pensiero: Milan, Italy, 2020, Volume 3. (In Italian)

10. Government of the Tuscany Region . LR 24 Febbraio 2005, n. 40. 2005. Available online: http://raccoltanormativa.consiglio.regione.toscana.it/articolo?urndoc=urn:nir:regione.toscana:legge:2005-02-24;40 (accessed on 16 March 2021). (In Italian)

11. Irizarry, J.; Gheisari, M.; Williams, G.; Roper, K. Ambient intelligence environments for accessing building information: A healthcare facility management scenario. *Facilities* **2014**, *32*, 120–138. [CrossRef]

12. Wanigarathna, N.; Jones, K.; Bell, A.; Kapogiannis, G. Building information modelling to support maintenance management of healthcare built assets. *Facilities* **2019**, *37*, 415–434. [CrossRef]

13. Singla, K.; Arora, R.; Kaushal, S. An approach towards IoT-based healthcare management system. In Proceedings of the Sixth International Conference on Mathematics and Computing, Online Event, 14–18 September 2020; pp. 345–356.

14. Noueihed, J.; Diemer, R.; Chakraborty, S.; Biala, S. Comparing Bluetooth HDP and SPP for mobile health devices. In Proceedings of the 2010 International Conference on Body Sensor Networks, Singapore, 7–9 June 2010; pp. 222–227.

15. Peng, S.; Su, G.; Chen, J.; Du, P. Design of an IoT-BIM-GIS based risk management system for hospital basic operation. In Proceedings of the 2017 IEEE Symposium on Service-Oriented System Engineering (SOSE), San Francisco, CA, USA, 6–9 April 2017; pp. 69–74.

16. Thangaraj, M.; Ponmalar, P.P.; Anuradha, S. Internet Of Things (IOT) enabled smart autonomous hospital management system—A real world health care use case with the technology drivers. In Proceedings of the 2015 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Madurai, India, 10–12 December 2015; pp. 1–8.

17. Iadanza, E.; Luschi, A. An integrated custom decision-support computer aided facility management informative system for healthcare facilities and analysis. *Health Technol.* **2020**, *10*, 135–145. [CrossRef]

18. Ahmed, S.; Liwicki, M.; Weber, M.; Dengel, A. Automatic room detection and room labeling from architectural floor plans. In Proceedings of the 2012 10th IAPR International Workshop on Document Analysis Systems, Gold Coast, QLD, Australia, 27–29 March 2012; pp. 339–343.

19. Brucker, M.; Durner, M.; Ambruş, R.; Márton, Z.C.; Wendt, A.; Jensfelt, P.; Arras, K.O.; Triebel, R. Semantic labeling of indoor environments from 3d rgb maps. In Proceedings of the 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, QLD, Australia, 21–25 May 2018; pp. 1871–1878.

20. Mewada, H.K.; Patel, A.V.; Chaudhari, J.; Mahant, K.; Vala, A. Automatic room information retrieval and classification from floor plan using linear regression model. *Int. J. Doc. Anal. Recognit. (IJDAR)* **2020**, *23*, 253–266. [CrossRef]

21. Sünderhauf, N.; Dayoub, F.; McMahon, S.; Talbot, B.; Schulz, R.; Corke, P.; Wyeth, G.; Upcroft, B.; Milford, M. Place categorization and semantic mapping on a mobile robot. In Proceedings of the 2016 IEEE international conference on robotics and automation (ICRA), Stockholm, Sweden, 16–21 May 2016; pp. 5729–5736.

22. Mancini, M.; Bulo, S.R.; Caputo, B.; Ricci, E. Robust place categorization with deep domain generalization. *IEEE Robot. Autom. Lett.* **2018**, *3*, 2093–2100. [CrossRef]

23. Pal, A.; Nieto-Granda, C.; Christensen, H.I. Deduce: Diverse scene detection methods in unseen challenging environments. In Proceedings of the 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Macau, China, 3–8 November 2019; pp. 4198–4204.

24. Li, K.; Qian, K.; Liu, R.; Fang, F.; Yu, H. Regional Semantic Learning and Mapping Based on Convolutional Neural Network and Conditional Random Field. In Proceedings of the 2020 IEEE International Conference on Real-Time Computing and Robotics (RCAR), Asahikawa, Japan, 28–29 September 2020; pp. 14–19.

25. Jin, C.; Elibol, A.; Zhu, P.; Chong, N.Y. Semantic Mapping Based on Image Feature Fusion in Indoor Environments. In Proceedings of the 2021 21st International Conference on Control, Automation and Systems (ICCAS), Jeju, Korea, 12–15 October 2021; pp. 693–698.

26. Liu, Q.; Li, R.; Hu, H.; Gu, D. Indoor topological localization based on a novel deep learning technique. *Cogn. Comput.* **2020**, *12*, 528–541. [CrossRef]

27. Kok, J.N.; Boers, E.J.; Kosters, W.A.; Van der Putten, P.; Poel, M. Artificial intelligence: Definition, trends, techniques, and cases. *Artif. Intell.* **2009**, *1*, 270–299.

28. Russell, S.; Norvig, P. *Künstliche Intelligenz*; Pearson Studium: München, Germany, 2012; Volume 2.

29. Affonso, C.; Rossi, A.L.D.; Vieira, F.H.A.; de Leon Ferreira, A.C.P.; others. Deep learning for biological image classification. *Expert Syst. Appl.* **2017**, *85*, 114–122. [CrossRef]

30. MathWorks. MATLAB per il Deep Learning. 2021. Available online: https://mathworks.com/solutions/deep-learning.html (accessed on 16 March 2021).

31. Sun, Y.; Xue, B.; Zhang, M.; Yen, G.G. Evolving deep convolutional neural networks for image classification. *IEEE Trans. Evol. Comput.* **2019**, *24*, 394–407. [CrossRef]

32. Izadinia, H.; Shan, Q.; Seitz, S.M. Im2cad. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 5134–5143.

33. Sun, Y.; Xue, B.; Zhang, M.; Yen, G.G.; Lv, J. Automatically designing CNN architectures using the genetic algorithm for image classification. *IEEE Trans. Cybern.* **2020**, *50*, 3840–3854. [CrossRef]

34. Wu, J. Introduction to convolutional neural networks. *Natl. Key Lab Nov. Softw. Technol. Nanjing Univ. China* **2017**, *5*, 495.

35. O'Shea, K.; Nash, R. An introduction to convolutional neural networks. *arXiv* **2015**, arXiv:1511.08458.

36. Han, X.; Laga, H.; Bennamoun, M. Image-based 3D object reconstruction: State-of-the-art and trends in the deep learning era. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *43*, 1578–1604. [CrossRef] [PubMed]

37.	He, T.; Zhang, Z.; Zhang, H.; Zhang, Z.; Xie, J.; Li, M. Bag of tricks for image classification with convolutional neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 558–567.

38.	Srinivas, S.; Sarvadevabhatla, R.K.; Mopuri, K.R.; Prabhu, N.; Kruthiventi, S.S.; Babu, R.V. A taxonomy of deep convolutional neural nets for computer vision. *Front. Robot. AI* **2016**, *2*, 36. [CrossRef]

39.	Hosny, A.; Parmar, C.; Quackenbush, J.; Schwartz, L.H.; Aerts, H.J. Artificial intelligence in radiology. *Nat. Rev. Cancer* **2018**, *18*, 500–510. [CrossRef] [PubMed]

40.	Niazi, M.K.K.; Parwani, A.V.; Gurcan, M.N. Digital pathology and artificial intelligence. *Lancet Oncol.* **2019**, *20*, e253–e261. [CrossRef]

41.	Mirbabaie, M.; Stieglitz, S.; Frick, N.R. Artificial intelligence in disease diagnostics: A critical review and classification on the current state of research guiding future direction. *Health Technol.* **2021**, *11*, 693–731. [CrossRef]

42.	Jiang, F.; Jiang, Y.; Zhi, H.; Dong, Y.; Li, H.; Ma, S.; Wang, Y.; Dong, Q.; Shen, H.; Wang, Y. Artificial intelligence in healthcare: Past, present and future. *Stroke Vasc. Neurol.* **2017**, *2*, 230–239. Available online: https://svn.bmj.com/content/svnbmj/2/4/230.full.pdf (accessed on 23 May 2022). [CrossRef]

43.	Rong, G.; Mendez, A.; Assi, E.B.; Zhao, B.; Sawan, M. Artificial intelligence in healthcare: Review and prediction case studies. *Engineering* **2020**, *6*, 291–301. [CrossRef]

44.	Rudie, J.D.; Rauschecker, A.M.; Bryan, R.N.; Davatzikos, C.; Mohan, S. Emerging applications of artificial intelligence in neuro-oncology. *Radiology* **2019**, *290*, 607–618. [CrossRef]

45.	Bera, K.; Schalper, K.A.; Rimm, D.L.; Velcheti, V.; Madabhushi, A. Artificial intelligence in digital pathology—New tools for diagnosis and precision oncology. *Nat. Rev. Clin. Oncol.* **2019**, *16*, 703–715. [CrossRef] [PubMed]

46.	Popescu, C.; Laudicella, R.; Baldari, S.; Alongi, P.; Burger, I.; Comelli, A.; Caobelli, F. PET-based artificial intelligence applications in cardiac nuclear medicine. *Swiss Med. Wkly.* **2022**, *152*, 1–4. Available online: https://smw.ch/article/doi/smw.2022.w30123 (accessed on 23 May 2022).

47.	Tran, D.; Kwo, E.; Nguyen, E. Current state and future potential of AI in occupational respiratory medicine. *Curr. Opin. Pulm. Med.* **2022**, *28*, 139–143. [CrossRef] [PubMed]

48.	Ijaz, A.; Nabeel, M.; Masood, U.; Mahmood, T.; Hashmi, M.S.; Posokhova, I.; Rizwan, A.; Imran, A. Towards using cough for respiratory disease diagnosis by leveraging Artificial Intelligence: A survey. *Inform. Med. Unlocked* **2022**, *29*, 100832. [CrossRef]

49.	Su, T.H.; Wu, C.H.; Kao, J.H. Artificial intelligence in precision medicine in hepatology. *J. Gastroenterol. Hepatol.* **2021**, *36*, 569–580. [CrossRef] [PubMed]

50.	Hogarty, D.T.; Mackey, D.A.; Hewitt, A.W. Current state and future prospects of artificial intelligence in ophthalmology: A review. *Clin. Exp. Ophthalmol.* **2019**, *47*, 128–139. [CrossRef] [PubMed]

51.	Kapoor, R.; Walters, S.P.; Al-Aswad, L.A. The current state of artificial intelligence in ophthalmology. *Surv. Ophthalmol.* **2019**, *64*, 233–240. [CrossRef]

52.	Citerio, G. Big Data and Artificial Intelligence for Precision Medicine in the Neuro-ICU: Bla, Bla, Bla. *Neurocritical Care* **2022**. [CrossRef]

53.	Zhou, B.; Lapedriza, A.; Torralba, A.; Oliva, A. Places: An image database for deep scene understanding. *J. Vis.* **2017**, *17*, 1–9. [CrossRef]

54.	Heller, M. What Is Computer Vision? AI for Images and Video. 2020 Available online: https://infoworld.com/article/3572553/what-is-computer-vision-ai-for-images-and-video.html (accessed on 16 March 2021).

55.	Al-Saffar, A.A.M.; Tao, H.; Talab, M.A. Review of deep convolution neural network in image classification. 2017 International Conference on Radar, Antenna, Microwave, Electronics, and Telecommunications (ICRAMET), Jakarta, Indonesia, 23–24 October 2017; pp. 26–31.

56.	Torrey, L.; Shavlik, J. Transfer learning. In *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques*; IGI Global: Hershey, PA, USA, 2010; pp. 242–264.

57.	Nilsson, K.; Jönsson, H.E. A Comparison of Image and Object Level Annotation Performance of Image Recognition Cloud Services and Custom Convolutional Neural Network Models. 2019. Available online: https://www.diva-portal.org/smash/get/diva2:1327682/FULLTEXT01.pdf (accessed on 23 May 2022).

58.	Wu, Y.; Kirillov, A.; Massa, F.; Lo, W.Y.; Girshick, R. Detectron2: A PyTorch-Based Modular Object Detection Library. 2019. Available online: https://ai.facebook.com/blog/-detectron2-a-pytorch-based-modular-object-detection-library-/ (accessed on 10 December 2021).

59.	Fei-Fei, L.; Deng, J.; Russakovsky, O.; Berg, A.; Li, K. ImageNet. 2021. Available online: http://image-net.org/ (accessed on 16 March 2021).

60.	Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]

61.	Google. Goole Images. Available online: https://www.google.com/imghp?hl=en_en&tbm=isch&gws_rd=ssl (accessed on 23 May 2022).

62.	Google Cloud. Vision AI | Use Machine Learning to Understand Your Images with Industry-Leading Prediction Accuracy. 2020. Available online: https://cloud.google.com/vision (accessed on 31 January 2022).

63. Amazon. Amazon Rekognition—Automate Your Image and Video Analysis with Machine Learning. 2022. Available online: https://aws.amazon.com/rekognition/?nc1=h_ls (accessed on 4 February 2022).

64. Microsoft Azure. Computer Vision. 2021. Available online: https://azure.microsoft.com/en-us/services/cognitive-services/computer-vision/ (accessed on 16 March 2021).

65. Clarifai. General Image Recognition AI Model For Visual Search. 2020. Available online: https://www.clarifai.com/models/general-image-recognition (accessed on 16 March 2021).

66. Bisong, E. *Building Machine Learning and Deep Learning Models on Google Cloud Platform*; Springer: Berlin, Germany, 2019.

67. Chen, S.H.; Chen, Y.H. A content-based image retrieval method based on the google cloud vision api and wordnet. In Proceedings of the Asian Conference on Intelligent Information and Database Systems, Kanazawa, Japan, 3–5 April 2017; pp. 651–662.

68. Mulfari, D.; Celesti, A.; Fazio, M.; Villari, M.; Puliafito, A. Using Google Cloud Vision in assistive technology scenarios. In Proceedings of the 2016 IEEE Symposium on Computers and Communication (ISCC), Messina, Italy, 27–30 June 2016; pp. 214–219.

69. Li, X.; Ji, S.; Han, M.; Ji, J.; Ren, Z.; Liu, Y.; Wu, C. Adversarial examples versus cloud-based detectors: A black-box empirical study. *IEEE Trans. Dependable Secur. Comput.* **2019**, *18*, 1933–1949. [CrossRef]

70. Hosseini, H.; Xiao, B.; Poovendran, R. Google's cloud vision api is not robust to noise. In Proceedings of the 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA), Cancun, Mexico, 18–21 December 2017; pp. 101–105.

71. Lazic, M.; Eder, F. Using Random Forest Model to Predict Image Engagement Rate. 2018. Available online: https://www.diva-portal.org/smash/get/diva2:1215409/FULLTEXT01.pdf (accessed on 16 March 2021).

72. Araujo, T.; Lock, I.; van de Velde, B. Automated Visual Content Analysis (AVCA) in Communication Research: A Protocol for Large Scale Image Classification with Pre-Trained Computer Vision Models. *Commun. Methods Meas.* **2020**, *14*, 239–265. [CrossRef]

73. Clarifai. Enlight ModelForce: Custom AI Model Building Services From Clarifai. 2020. Available online: https://www.clarifai.com/custom-model-building (accessed on 16 March 2021).

74. PyTorch developer community. From Research to Production. 2021. Available online: https://pytorch.org/ (accessed on 22 December 2021).

75. Cutler, A.; Cutler, D.R.; Stevens, J.R. Random forests. In *Ensemble Machine Learning*; Springer: Berlin, Germany, 2012; pp. 157–175.

76. Guidi, G.; Pettenati, M.C.; Miniati, R.; Iadanza, E. Random forest for automatic assessment of heart failure severity in a telemonitoring scenario. In Proceedings of the 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Osaka, Japan, 3–7 July 2013; pp. 3230–3233.

77. Rokach, L.; Maimon, O. Decision trees. In *Data Mining and Knowledge Discovery Handbook*; Springer: Berlin, Germany, 2005; pp. 165–192.

78. Scikit-learn Team. Scikit-learn - Machine Learning in Python. 2021. Available online: https://scikit-learn.org/stable/ (accessed on 10 December 2021).

79. MIT, Computer Science and Artificial Intelligence Laboratory. LabelMe Welcome Page. 2021. Available online: http://labelme.csail.mit.edu/Release3.0/ (accessed on 10 December 2021).

80. Roboflow Team. Give Your Software the Sense of Sight. 2021. Available online: https://roboflow.com/ (accessed on 22 December 2021).

81. COCO Consortium. COCO—Common Objects in Context. 2022. Available online: https://cocodataset.org/#home (accessed on 31 January 2022).