This project has received funding from
The European Union's Horizon 2020 research and innovation programme under
Grant agreement No 101017562

Finding Endometriosis using Machine Learning
FEMaLe

Call/Topic: Digital transformation in Health and Care
Type of action**:** RIA

Date: 04.07.2022

DELIVERABLE NUMBER	D6.2
DELIVERABLE TITLE	Dataset of laparoscopic semantic annotated images
RESPONSIBLE AUTHOR	Julie DESTERNES (SurgAR), Abir KHEMIRI (SurgAR)

GRANT AGREEMENT No.	101017562
DOCUMENT TYPE	Report
WORKPACKAGE N. | TITLE	WP6 | DIAGOSIS: machine learning automatic surgical phenotyping
LEAD CONTRACTOR	AUH
AUTHOR(S)	Julie DESTERNES, Abir KHEMIRI
PLANNED DELIVERY DATE	June 30st, 2022
ACTUAL DELIVERY DATE	July 04th, 2022
DISSEMINATION LEVEL	Confidential
STATUS	Completed
VERSION	Final version (1.5)
REVIEWED BY	FEMaLE PMO

Document history

Version	Date [1]	Comment	Author	Status [2]
1.1	24-05-2022	First draft created	AUH	Drafted
1.2	14-06-2022	Second draft created, including recommendations from WP participants.	SurgAR	Drafted
1.3	28-06-2022	Third draft prepared for FEMaLe Review Panel.	SurgAR	Drafted
1.4	04-07-2022	Final draft created, based on FEMaLe Review Panel feedback.	SurgAR	Completed
1.5	05-07-2022	Final version ready for submission, quality checked by FEMaLe PMO.	AU	Validated

[1] As per the project's cloud storage or per email date if applicable.
[2] Drafted, completed or validated as per the project's cloud storage or per email date if applicable.

TABLE OF CONTENTS

Disclaimer

Copyright notice

Acknowledgement

The project 'Finding Endometriosis using Machine Learning' (FEMaLe) has received funding from the European Union Horizon 2020 programme under grant number 101000640.

Citation

Be so kind as to cite this work as: Finding Endometriosis using Machine Learning, 2022: *Dataset of laparoscopic semantic annotated images* under the supervision of the Project's Coordinator.

Legislation

Legislation H2020 Framework Programme – Regulation (EU) No 1291/2013 of the European Parliament and of the Council of 11 December 2013 establishing Horizon 2020 - The Framework Programme for Research and Innovation (2014-2020) (OJ 347, 20.12.2013, p. 104).

H2020 Specific Programme – Council Decision 2013/743/EU of 3 December 2013 establishing the Specific Programme Implementing Horizon 2020 - The Framework Programme for Research and Innovation (2014-2020) (OJ L 347, 20.12.2013, p. 965).

Rules for Participation (RfP) – Regulation (EU) No 1290/2013 of the European Parliament and of the Council of 11 of December 2013 laying down the rules for the participation and dissemination in Horizon 2020 – the Framework Programme for Research and Innovation (2014-2020) (OJ L 347, 20.12.2013, p.81).

Financial Regulation (FR) – Regulation (EC, Euratom) No 966/2012 of the European Parliament and of the Council of 25 October 2012 on the financial rules applicable to the general budget of the European Union (OJ L 298, 26.10.2012, p.1).

Rules of Application (RAP) – Commission Regulation (EC, Euratom) No 1268/2012 of 29 October 2012 on the rules of application of l Regulation (EC, Euratom) No 966/2012 of the European Parliament and of the Council on the financial rules applicable to the general budget of the Union (OJ L 298, 26.10.2012, p.1).

Finding Endometriosis using Machine Learning: FEMaLe

1. Introduction

WP	6	Lead	Goal	Content	Gantt (M3 - M38)
	6.1	SWU	Collecting videos with metadata	200 surgeries	M3 - M9*
	6.2	AUH	Annotation of laparoscopic videos	Between 50,000-100,000 images or short videos will be annotated.	M9 - M18
	6.3	RTU	Developing deep learning algorithm to automatic r-ASRM/ENZIAN classification	Deep learning algorithm to automatically determine the stage of endometriosis. Machine learning technology to automate the evaluation of the endometriosis specific score.	M18 - M26
	6.4	SURGAR	Integration to the software and real time assessment of algorithm	Integration to the software and real time asComputer vision-based software for real time assessment of endometriosis. The tool allows an automatic and objective confirmation of the diagnosis and assessment of the stage and extension of the disease evaluated during laparoscopy.sessment of algorithm	M26- M32
	6.5	SURGAR	Clinical validation of the algorithm		M32-M38

Fig.1: Tasks and contents of Work Package 6 (*extended to M12)

The main objective of this document is to deliver a general report regarding the second phase of the WP6 'Collecting videos with metadata'. Firstly, it is important to state that the previous phase (task 6.1) is still and will remain in progress in parallel with the second phase. This is done to provide as much high-quality data as possible for both training and testing the deep learning algorithm, which will be developed in the next task. It is therefore more relevant to start by providing an update on the progress of the task 6.1 before proceeding with 6.2.

1.1. Constitution of the FEMaLe video dataset

Currently, more than 200 raw endometriosis surgeries (videos and metadata) are collected, compared to 132 in March. These surgeries are collected from healthcare centers with which a legal agreement has been established: *Hungary* (Semmelweis University Hospital), *France* (Clermont-Ferrand University Hospital) and more recently *Brazil* (BP, Beneficência Portuguesa).

Please bear in mind that not all the surgeries are considered relevant when creating the FEMaLe database due to their complexity, however, they might be included in the future. In addition, other new surgeries have recently been shared on the SurgAR HDP certified storage and therefore not yet processed. These results will be reported in later deliverables.

The evolution of the number of surgeries used in the constitution of the dataset between March and June as well as their distribution according to the different healthcare centers is presented in Fig. 2 below. Here, it is shown how the number of used surgeries has been significantly enriched in only a few months in terms of both quantity and quality; data source heterogeneity. In March, we counted 64 used surgeries to create the dataset; today, we are counting 104 surgeries with more to come.

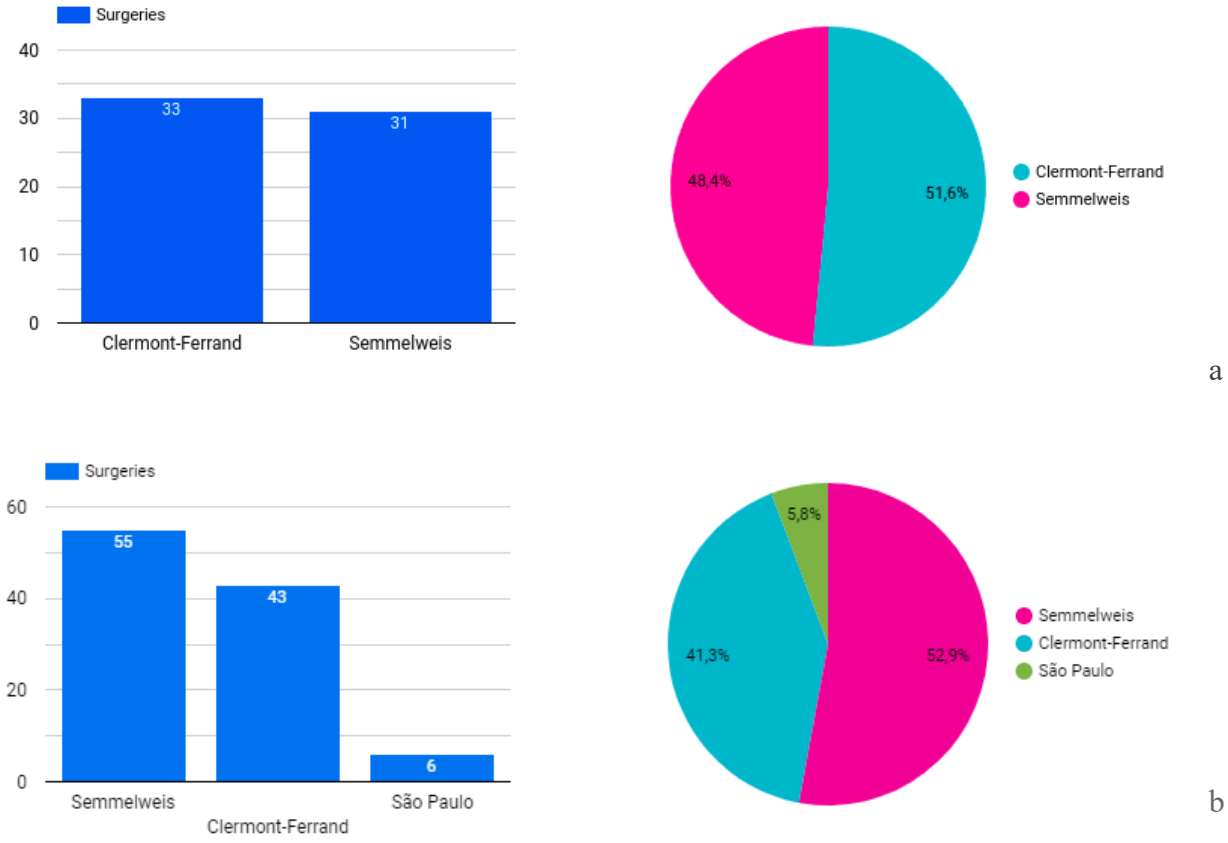


Fig.2: Dataset distribution according to source of surgeries in March 2022 (Fig.2.a) and June 2022 (Fig.2.b).

From these videos, short sequences were extracted to enrich the dataset (please consult section 1.2), which will then be annotated. We currently count 223 video sequences compared to 114 in March.

The distribution of these sequences according to their duration can be seen in Fig.3 below:

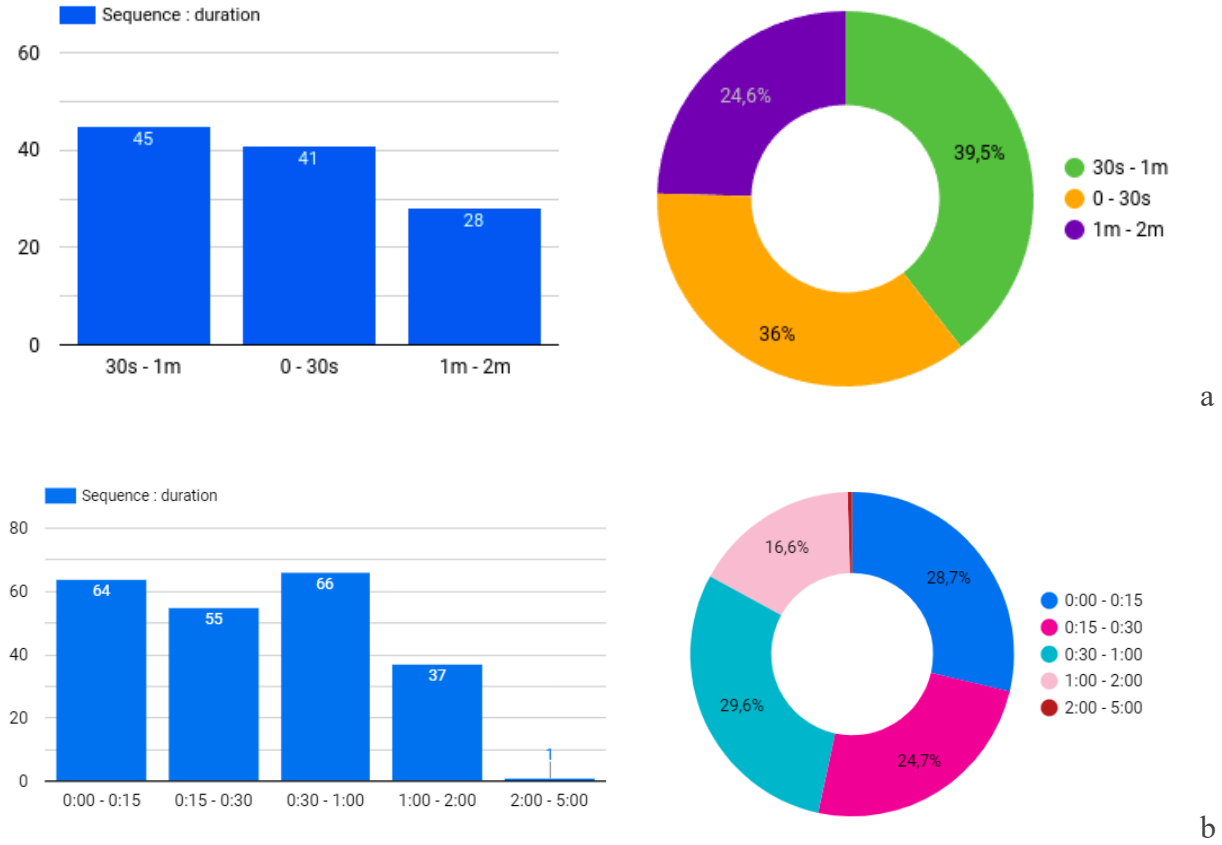


Fig.3: Duration of video sequences in March 2022 (Fig.3.a) and in June 2022 (Fig.3.b).

1.2. Video sequences extraction criteria

The dataset to be annotated is constituted of short video sequences. Unlike image annotation, video annotation provides a context for the labelers, which allows a better understanding of the selected frame, provides additional useful information, and helps reduce annotation errors.

These video sequences are identified on the laparoscopic videos by doctors with expertise in endometriosis and are then extracted according to several criteria:

- **Duration**: Video sequences should in general be short (only >2 minutes in rare cases).
- **Quality**: In these sequences, the endometriosis lesions must be clearly visible and relatively detectable by the surgeon.
- **Surgery phase**: The sequences should be extracted at certain key moments, particularly at the beginning of the surgery, when the abdominal cavity is explored.

As the next step, other video sequences of gynecological surgeries on healthy patients (without endometriosis) or surgeries related to other pathologies (carcinomatosis, myomas …) will be extracted. These video sequences will be used by RIGAS TEHNISKA UNIVERSITATE (P7) as negative cases for the training of the machine/deep learning algorithms in WP6.

1.3. Main goals

The aim of Task 6.2 is the annotation of 50,000 to 100,000 extracted frames or short videos by junior and expert surgeons according to the ontology presented in Fig.4. T

he role of ontology is to set up a common vocabulary and to classify data in order to extract relevant information with a clinical impact from laparoscopic videos.

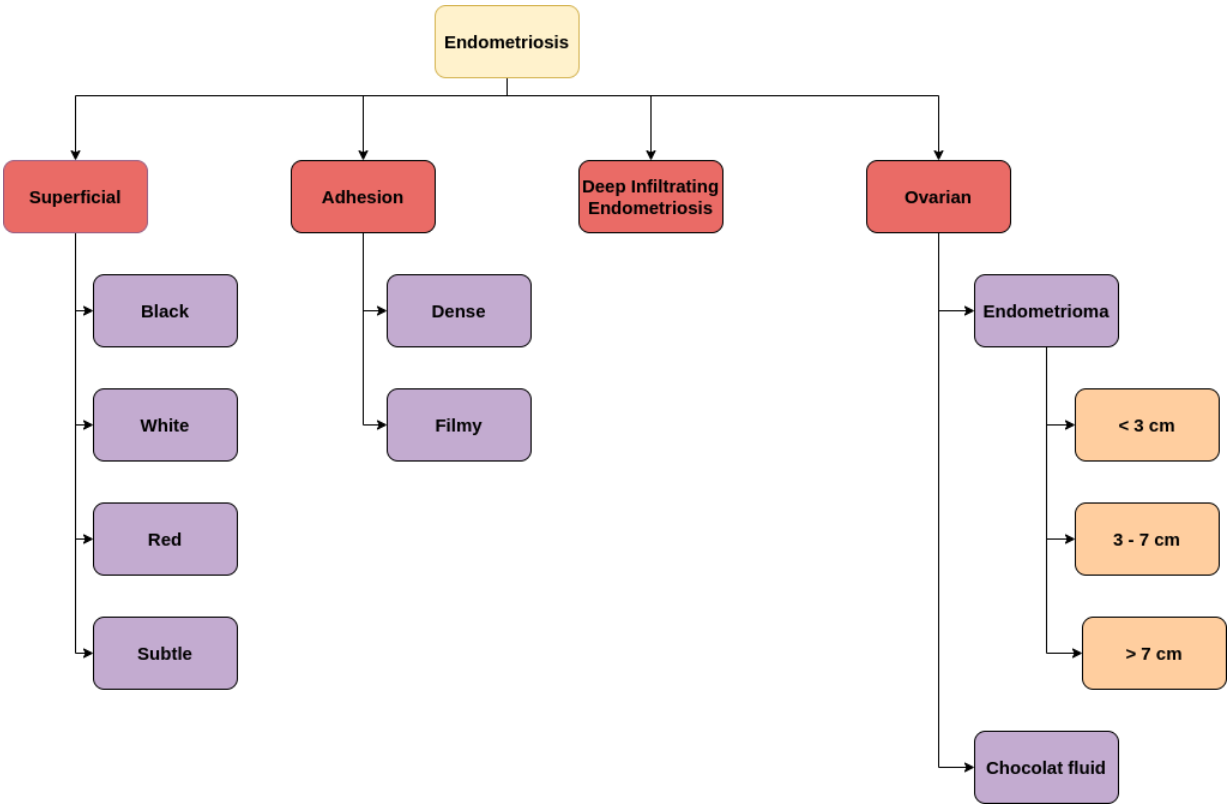


Fig.4: Endometriosis ontology

2. What is the most appropriate annotation method?

In the early stages of the Tasks 6.2 'Semantic annotation of laparoscopic videos', we set objectives which, with the benefit of hindsight and after several tests with several surgeons, proved to be very difficult to achieve.

2.1. Basic solution

Initially, annotators were asked to segment on static frames, the delimitation of each endometriosis lesion and to indicate its size and type based on a very large and *detailed* ontology. The objective of this very accurate classification is to be able to calculate a score associated with each patient.

The main problematics of this approach are:

- The complexity of the ontology.
- The complexity of identifying the location of a lesion in the abdominal cavity on a single image.
- The impossibility to evaluate the size of lesion on static frames.
- The lack of a complete overview of the abdominal cavity.
- The absence of MRI images.
- The unclear delimitations (huge inter-annotator variance): not accurate and very time-consuming annotation method (Fig.5).

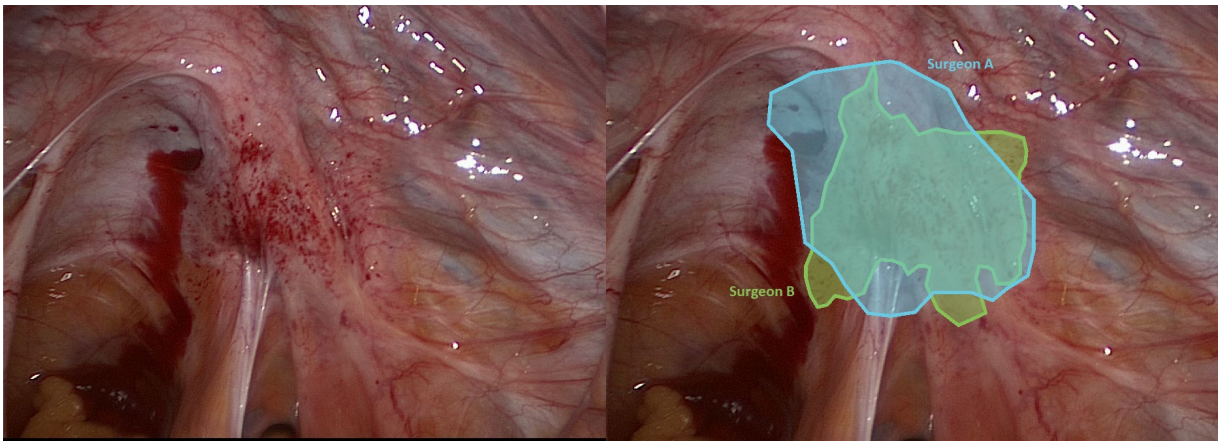


Fig.5: Different segmentations of the same endometriosis lesion performed by two different surgeons

2.2. Delphi method

As a solution, we opted for the strategy of using the Delphi method. This method is usually used to aggregate opinions from a diverse set of experts through several rounds of questionnaires (open-ended questions at first and increasingly closed-ended questions later). The collected responses are aggregated and shared with the experts after each round, so that their further responses are adjusted to the group response in order to reach a general consensus.

For this task, the Delphi method is selected to establish consensus between medical experts on an optimal ontology and the right method to annotate endometriosis lesions which are highly variable and different.

The questionnaires are designed to be as time efficient as possible (10 to 15 minutes) for experts with an already busy schedule. The development of this new annotation method is carried out by SurgAR in collaboration with surgeons (Nicolas Bourdel and Antoine Netter) and residents (Fanny Duchateau and Henrique Abrao) from two French university hospitals (Clermont-Ferrand and Marseille). Our Delphi strategy is described in detail in Fig.6:

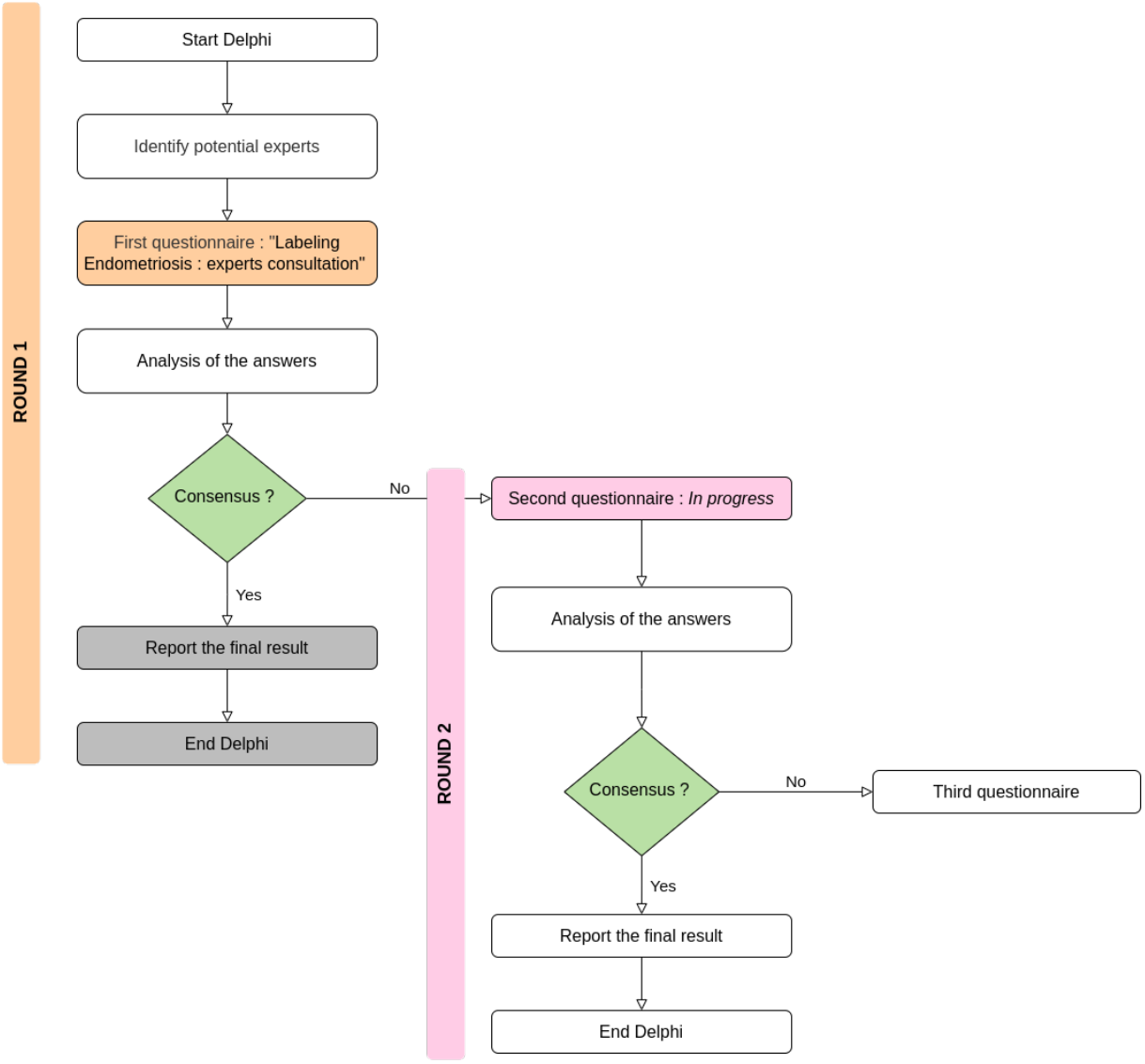


Fig.6: Delphi survey flowchart

Round 1

A first questionnaire was sent to experts in December 2021 with a title of: *Labelling Endometriosis: Expert's consultation (Delphi Round One)*, and was designed by A. Netter, H. Abrao, and N. Bourdel.

This round consists of open-ended questions to identify disagreements between experts on the ontological classification and annotation of endometriosis lesions. Reaching a consensus on this part provides a solid basis to effectively initiate the annotation task and to provide higher quality data for the training of machine learning algorithms.

This preliminary questionnaire completed by 14 participants (surgeons and expert surgeons) from 7 countries (France, Germany, Denmark, Greece, Hungary, Austria, and Brazil) revealed the difficulty of segmenting endometriosis lesions and challenged the annotation method initially selected. The difficulty to establish a consensus on all the issues was significantly observable. A detailed analysis of the results of this round was elaborated and then shared with all participants.

Round 2

Based on the answers of the first Delphi questionnaire, a second questionnaire entitled: *Labelling Endometriosis: Expert's consultation (Delphi Round Two)* was created. It was composed of a set of closed-ended questions regarding the appropriate method of annotating the endometriosis lesions.

The answers are crucial to:

- Validate the ontology.
- Obtain a more consensual annotation method for endometriosis lesions.

The second round was sent in May 2022 to the 14 experts from 7 countries, who participated in the first round. Currently, we have received 7 replies from the 14 experts (50%), who received the first questionnaire. Based on the initial results, a third round will most probably be required to eliminate the minor points of disagreement.

3. The type of annotation

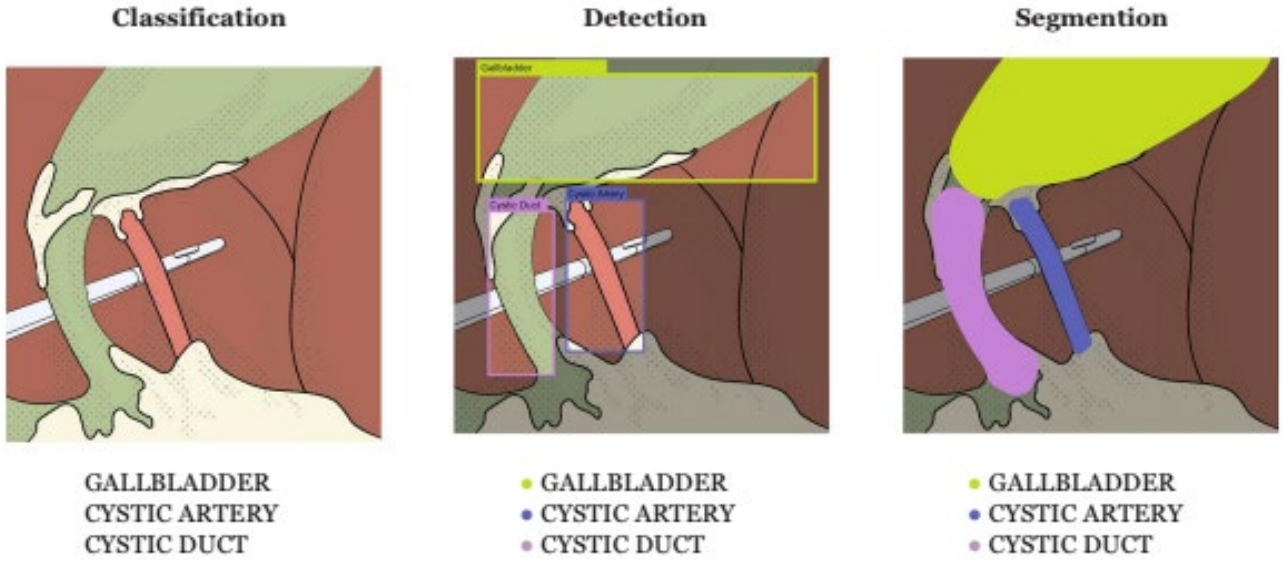


Fig.7: Example of image analysis techniques from Anteby et al., 2021[3] in the case of cholecystectomy.

The most common annotation methods are classification, detection, and segmentation, as illustrated in Fig.7:

- **Classification** consists in assigning a specific class(es) to an entire image.
- **Detection** introduces a notion of localization by identifying the location of an object of interest (lesion, anatomical structure, instrument, etc.) by using a bounding box.
- **Segmentation** allows defining the precise pixel-wise boundaries of an object of interest (lesion, anatomical structure, instrument, etc.) [4]

Results of the first Delphi round and preliminary results of the second Delphi round suggest that the data annotation for WP6 needs to be performed using *bounding boxes* (object detection) on video sequence using the ontology defined with the Delphi method. The video sequence annotation will be performed on a few key frames (a maximum of ten: five after and five before) of the video.

Bounding boxes have been selected as they offer fast annotation, less variability between annotators and a 'raw estimation' of the lesion position on images. Contrary to accurate segmentations of lesions which induce a high inter-annotator and intra-annotator variability, object classification corresponds to concepts that might be easier to learn, which could decrease the difficulties in the learning process.

[3] Anteby, R., Horesh, N., Soffer, S., Zager, Y., Barash, Y., Amiel, I., Rosin, D., Gutman, M., & Klang, E. (2021). Deep learning visual analysis in laparoscopic surgery : a systematic review and diagnostic test accuracy meta-analysis. *Surgical Endoscopy*, *35*(4), 1521-1533. https://doi.org/10.1007/s00464-020-08168-1

Considering the different feedback from experts, we have included an exception to this annotation method regarding two classes in particular: 'Endometrioma' and 'Chocolate fluid'. Due to the very distinct contour of these elements, an accurate segmentation is more easily possible and performed as presented in Fig.8.

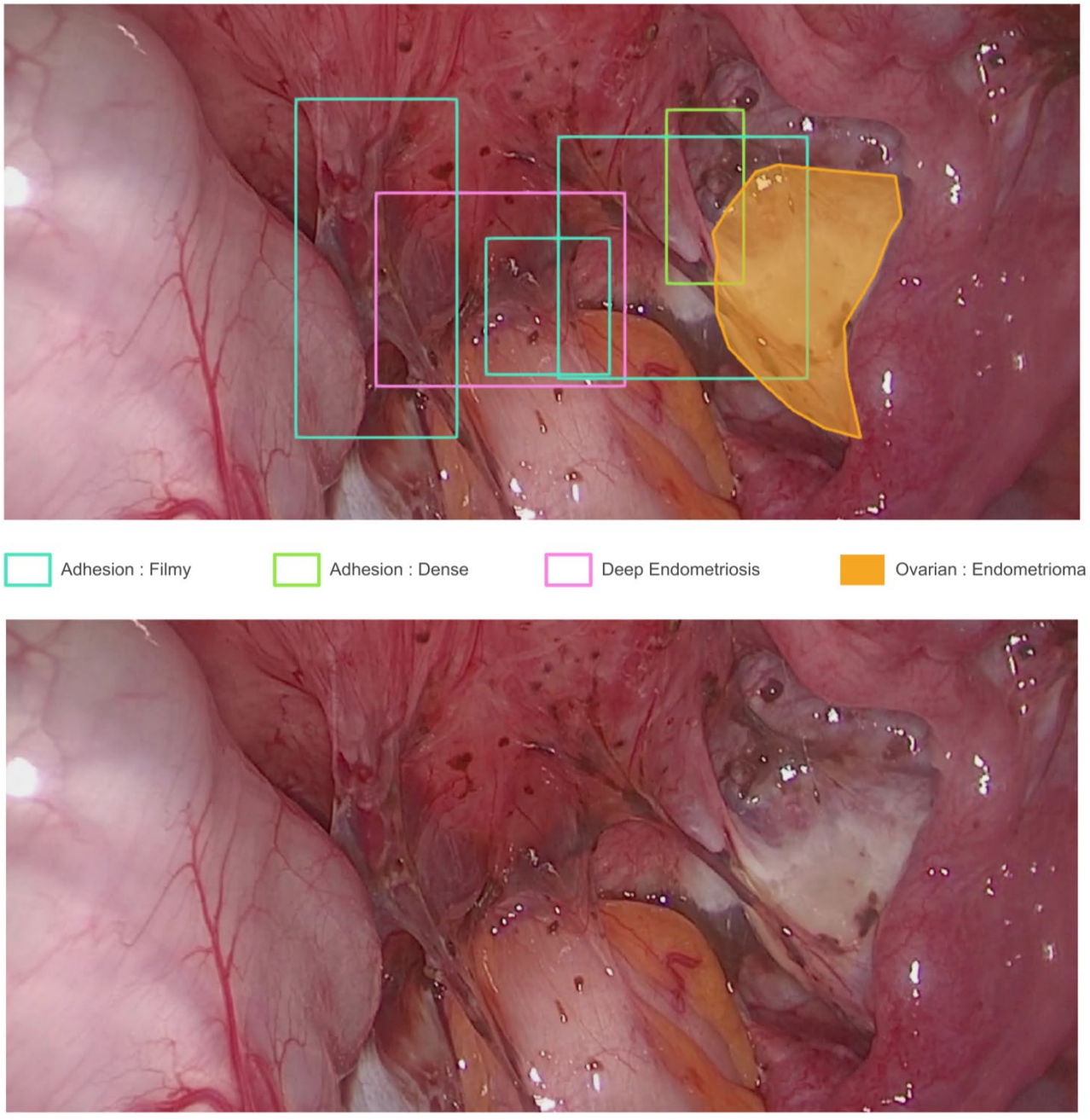


Fig.8: a.) Frame with different endometriosis lesions (see Fig.4 for the ontology). b.) Bounding box annotation and pixel-wise segmentation of different endometriosis lesions.

4. The annotation pipeline

The annotation process is conducted by the annotator team on the Supervisely[4] platform. The certification process of the annotators is carried out as follows:

- The annotator must justify his/her medical background by sending his/her CV, which will be reviewed by the experts.
- Each new annotator will then follow an annotation workshop to understand the global procedure, the guidebooks which are prepared by the experts, the ontology, and the annotation platform.
- He/she will then validate a general exam, based on the usability of the annotation platform.
- If the annotators pass the exam by gaining 80/100 in less than three attempts, then they must pass the specialized exam on the detection of different endometriosis lesions. If they don't pass the exam, they are able to attend a review workshop and then retake the exam.
- The specialized exam will be verified by comparing it to a consensus agreement by experts.

The annotators come from around the world to respect the diversity of ideas: Marseille (France), CHU Clermont-Ferrand (France) and new ones are to be recruited, e.g., from Aarhus University Hospital in Denmark. This variability can also be observed through the profiles of the annotators ranging from junior (intern) to expert.

All these preliminary steps are useful to justify the data quality. After the data is uploaded to SurgAR servers, the sequences are extracted, and they are exported to Supervisely platform for the annotation process. The annotated videos are also reviewed and corrected by experts. If the annotation is not satisfactory, the necessary modifications will be made during the review stage until the final annotations are qualitatively adequate. The pipeline is depicted in Fig.9:

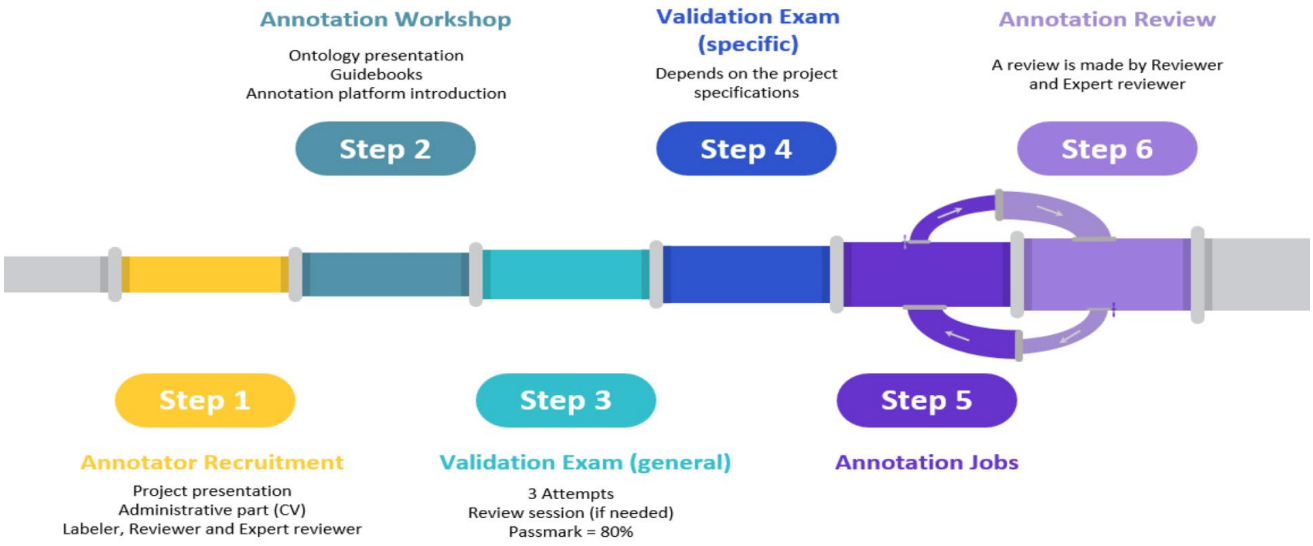


Fig.9: The annotation pipeline.

[4] Supervisely: unified OS for computer vision

5. Annotation state of progress at M18

The annotations of video sequences are constructively in progress. The annotations are done in two-day sessions (batches) during which the team of technical experts and annotators do their best to annotate the data according to the defined pipelines. Up to this moment, 26,869 endometriosis lesions are annotated in the videos. The details of the annotated lesions are shown in Fig.10:

TOTAL	26869
Adhesions.Dense	6856
Deep Endometriosis	2247
Superficial.Red	4793
Superficial.Black	400
Adhesions.Filmy	339
Superficial.White	4493
Ovarian.Endometrioma	3761
Superficial.Subtle	1738
Ovarian.Chocolate Fluid	2242

Fig.10: The total number of lesions which are annotated in the dataset.

We are owning three batches of annotated data thus far. The corresponding dates of the batches and the number of lesions for each batch are shown in Fig.11:

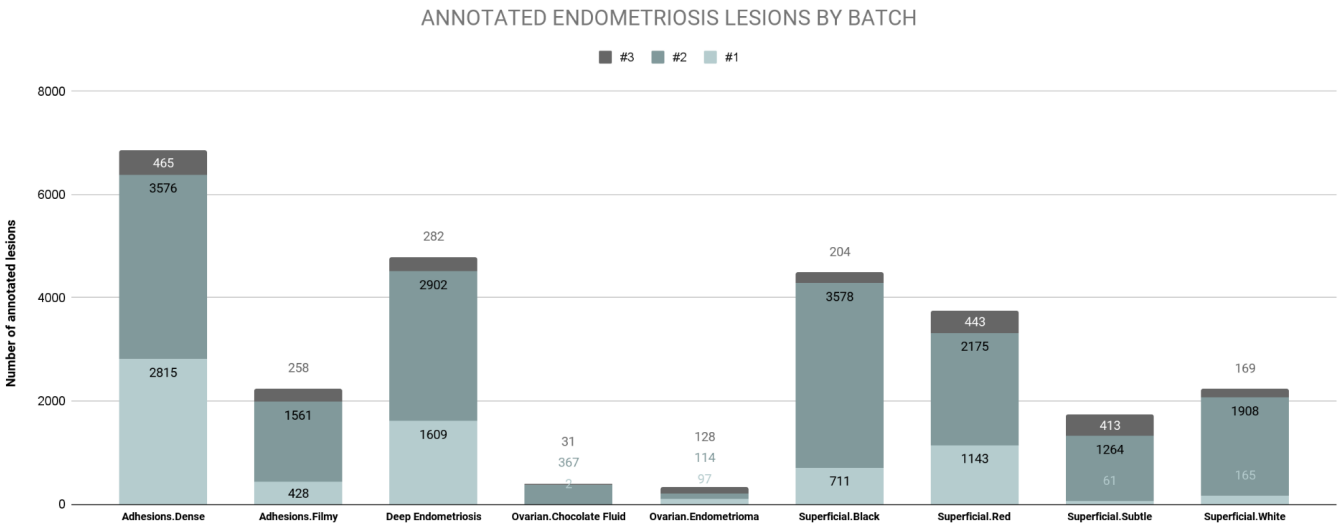
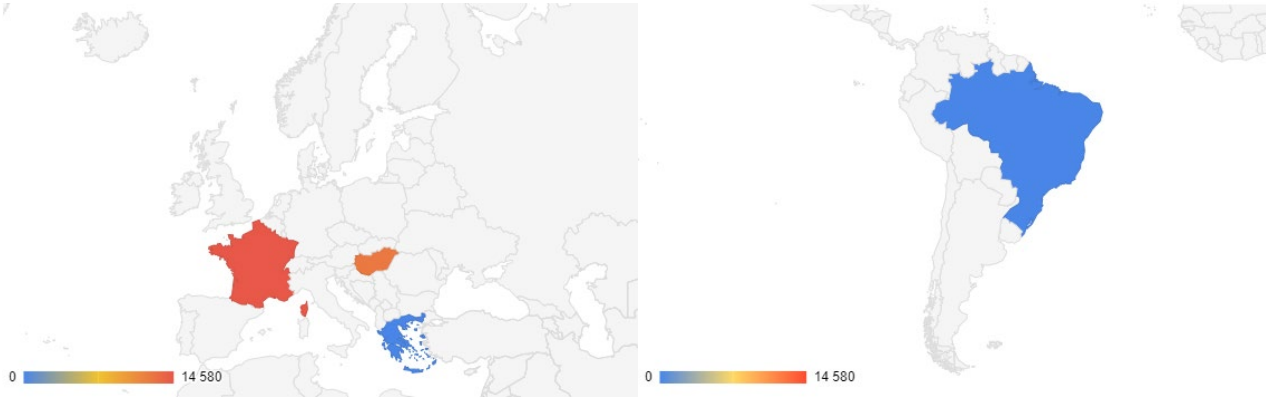


Fig.11: The number of annotated lesions shown for each batch.

In Fig.12 it is shown how the annotated data are from different countries around the world:



CITY, COUNTRY	ANNOTATED ENDOMETRIOSIS LESIONS
Clermont-Ferrand, France	14580
Budapest, Hungary	12289
Sao Paulo, Brazil	0
Athens, Greece	0

Fig.12: The geographical variation of the annotated data. The red and orange shows the places from which we owe annotated data right now, and the blue shows the origin of the data that will be annotated.

Currently, the annotated dataset contains frames from Clermont-Ferrand, France and Budapest, Hungary in comparable amounts. Another major challenge will be to diversify the origin of the annotated data. The next annotation workshop will include data from Sao Paulo, Brazil and Athens, Greece.

The job of annotating the videos is progressively being done. The two next workshops are planned, and the dates are reserved in July 2022, and the next ones will be planned until December 2022. New annotators will be joining the team for whom the next workshops are already planned.

The first two batches are already sent to RIGAS TEHNISKA UNIVERSITATE (P7), who will be developing a deep learning algorithm for the detection of lesions in the framework of Task 6.3. The transfer of the data is still in progress.

After the first three annotation batches, we have come to realize that some endometriosis lesions are still under-represented, namely ovarian endometriomas and chocolate fluid (please see Fig.11). Balancing the size of different classes is an important element to take into consideration during the next annotation workshops.

6. Conclusions and perspectives

To wrap up, the annotation procedure was explained in this deliverable to shed light on the road for the future tasks of FEMaLe project.

The annotation process is done in the framework of WP6, Task 6.2). Here, the annotation pipeline, and the contribution of technical and medical experts are explained. The dataset that is being created and structured is also explained.

It is important to note that the annotation project has already been done for more than 25k of different endometriosis lesions according to the defined ontology. The annotated data is from two countries (France and Hungry), but the dataset from Brazil and Greece is also ready to be annotated according to the described pipeline.

The annotated data will then be used by Riga Technical University for the design of a deep learning method to detect and classify different endometriosis lesions.

We aim to reach the target number of annotations by December 2022 through regular annotation workshops. In order not to lose time for the next task, these annotations will be regularly sent to Riga Technical University for the good progress of the next Task 6.3 'Developing deep learning algorithm'.