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: 09.03.2023

DELIVERABLE NUMBER	D7.2
DELIVERABLE TITLE	Post-Operative Dataset of laparoscopic semantic annotated images
RESPONSIBLE AUTHOR	SEMMELWEIS

GRANT AGREEMENT No.	101017562
DOCUMENT TYPE	Website
WORKPACKAGE N. | TITLE	WP7 | VISUAL: augmented reality to improve laparoscopic surgery
LEAD CONTRACTOR	SEMMELWEIS,
AUTHOR(S)	SURGAR Saman Noorzadeh, Julie Desternes
PLANNED DELIVERY DATE	31 January 2023
ACTUAL DELIVERY DATE	March 09th, 2023
DISSEMINATION LEVEL	Public
STATUS	Completed
VERSION	Final version (1.5)
REVIEWED BY	FEMaLe PMO

Document history

Version	Date [1]	Comment	Author	Status [2]
1.1	02-01-2023	First draft created	SURG	Drafted
1.2	25-01-2023	Second draft created, including recommendations from WP participants.	SURG	Drafted
1.3	20-02-2023	Third draft prepared for FEMaLe Review Panel.	SURG	Drafted
1.4	02-03-2023	Final draft created, based on FEMaLe Review Panel feedback.	SURG	Completed
1.5	08-03-2023	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, 2023: Post-Operative 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).

Euratom Research and Training Programme (2014-2018) – Council Regulation (Euratom) No 1314/2013 of 16 December 2013 on the Research and Training Programme of the European Atomic Energy Community (2014-2018) complementing the Horizon 2020 – The Framework Programme for Research and Innovation (OJ L 347, 20.12.2013, p. 948).

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).

1. Introduction

WP7 aims to create an augmented reality (AR) tool to aid surgeons in the laparoscopic operation of endometriosis. This tool will visualise the division plane for the surgeon, a task that needs years of experience. Beside expert surgeons, this tool will also allow young surgeons to have better endometriosis surgeries. In other words, this AR tool aims at improving the outcome of surgeries both for expert and junior surgeons, for the sake of better patient care. Such a tool can also provide a standardisation of operation in an international resolution. This task can be achieved by training surgeons through visualisation platforms which is suggested by such software.

This objective can only be achieved through deep learning approaches which are based on artificial neural networks. For deep learning to succeed, good data is needed in the first place. This data has to have certain characteristics like the minimum quantity, the quality, and an unavoidable property: the annotation. This means all the collected laparoscopy video data which is trained by the algorithm needs to be annotated by surgeons so that the division plane is marked on them. Through this large amount of annotated data, the machine can be trained and learn to suggest the division area in this regard. The algorithm will finally be integrated into the SURGAR (P7) software to allow a real-time AR guidance tool to assist the surgeon in finding the division planes around the endometriotic lesion.

This deliverable (D7.2) is about the annotation of such data. The dataset is already explained in detail in D7.1, and now the annotation of this data will be fully covered in this report. Before going through the rest of the report, some of AI trustworthiness concepts, and how they are tackled in this wp are explained.

2. Trustworthy AI

According to the EC guidelines of April 2019[3], trustworthy AI should be *lawful, ethical, and robust* – and several requirements for AI system should be met. We comply with the requirements this way:

1. **Human agency and oversight**: The endometriosis operation suggestion device will empower surgeons and allow them to make informed decisions. At the same time, the ground truth decisions upon which the model is built are made by expert surgeons, and it is the surgeon who finally takes the decision about operation after the machine suggestion (we stick to the word 'suggest' for the machine to show that the main surgeon finally takes the decision)
2. **Technical Robustness and safety**: SURGAR is a company whose task is to develop such an AI based platform. All the security, robustness, and safety instructions are qualified according to the national and international regulations.
3. **Privacy and data governance**: Since, the collected data are sensitive, this data is made sure to be anonymized before use. The European and GDPRS regulations are applied on the data.
4. **Transparency**: The data, system, and strategies are made clear both in official documentations at SURGAR and in FEMaLe reports.
5. **Diversity, non-discrimination, and fairness**: We consider the diversity of data to ensure optimal performance of the AI model. For this purpose, the data is tried to be collected from different centres from around Europe and the world.
6. **Societal and environmental well-being**: The AI system being developed here is beneficial to improve the surgeries and impacts the help of patients with endometriosis.
7. **Accountability**: This point will be included in D7.3, which is dealing with the AI algorithm.

[3] https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

3. Dataset

The dataset and its characteristics are well explained in D7.1. However, to clarify this report and to update the latest status that was reported in D7.1, it will also be summarised in this section. According to trustworthy AI (as described above), if an AI model is trained without accounting for the varying data across different scenarios, its performance may be greatly and negatively impacted. It is thus important to consider the diversity of data to ensure optimal performance of the AI model. Currently, more than 370 (see Figure 1 below) raw endometriosis surgeries (videos and metadata) are collected. This number has increased by about 238 since the last report in M14. To date, SURGAR has contracts with five healthcare centres for the collection of data, compared to the two centres mentioned in the last report. These centres are listed as follows:

- University Hospital of Clermont-Ferrand (France).
- University Hospital of Semmelweis (Hungary).
- Beneficência Portugues Hospital (Brazil).
- Euroclinic Group of Hospitals - Athens (Greece).
- University Hospital of Bologna (Italy).

These data from almost around the world can assure a good variability in the dataset to ensure an output as general as possible. The data-share partnership is based on legal agreement which ensures full compliance with the General Data Protection Regulation (Regulation (EU) 2016/679). All our data are stored in the EU and anonymized or pseudonymized format to meet the safety requirements. It should be remembered that not all the collected data are being used in the dataset because of the quality, and complexity of some of the data, and the lack of expert annotators at the moment.

Short video sequences are extracted from the laparoscopic videos at certain key moments, particularly at the beginning of the surgery when the abdominal cavity is explored and when the endometriosis lesions are clearly visible. To date, 152 short videos are used in annotation, however, this number is increasing daily. The annotation will be performed on these short video sequences extracted from laparoscopic surgeries. The labeller will be asked to annotate a certain number of frames on the selected video sequence. These frames are selected and validated by surgeons. Unlike image annotation, video annotation provides a context for the labellers. This context allows a better understanding of the selected frame, provides additional useful information and can reduce annotation errors. In the rest of this document the necessary background will be first explained to understand the importance and necessity of annotations, then the complete procedure of annotating and the results are presented. Finally, the conclusions are discussed.

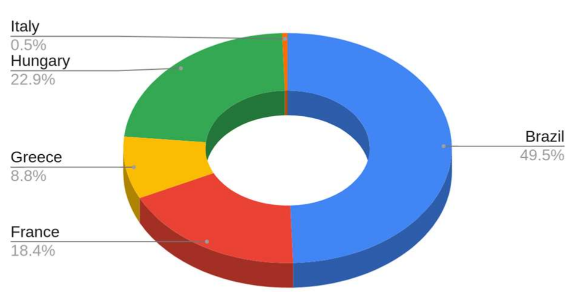


Figure 1: The distribution of collected data from around the world.

4. Theoretical Background

4.1. Deep Learning

Deep learning is considered as machine learning, and more generally AI methods, which uses the artificial neural networks for technology advancement, like in computer vision, medical image analysis and a wider domain of applications. Deep learning and in general neural networks are inspired from the biological nervous system and the way they communicate and process information. These methods let machines gain knowledge, experience, and information or be 'trained' in a more scientific vocabulary. The training of such models needs a lot of data. The machine learns on the input data which is provided to it. This learning can be supervised, non-supervised or semi-supervised.

4.2. Supervised learning and Annotation

The application of deep learning in supervised learning considers the fact that the input data is labelled. The labels provide a ground truth for the machine to learn the patterns out of the data. Annotation is the task that we do to label the input data. Therefore, the annotation task, refers to (annotate) the division boundaries on laparoscopic images, what we expect from the machine to finally predict for us. We can conclude that to use supervised learning in deep learning we need to annotate the images. An ideal annotation is the one which results in high quality, and trustable labels. To measure that, we consider expert surgeons in all the annotation processes. A good discussion is always planned to ensure the variability and a more precise annotation.

4.3. Semantic Segmentation

To annotate the division boundaries, these regions should be marked on the frames of the videos. In other words, these regions should be segmented on the image. Semantic segmentation classifies every pixel of an image into two or more classes. In the next sections the principles of such segmentation, and the number of classes, and their characteristics are explained.

5. Annotation Process

5.1. The Annotation Platform

The annotations are done on a web-app platform called Supervisely[4]. SURGAR buys the licence with features dedicated to such annotations from Supervisely and provides the annotators with private accounts to log in to the platform and do the annotations.

[4] https://supervise.ly/

5.2. Annotators

The annotators should have gynaecological surgery background. This can include expert surgeons or junior surgeons in gynaecology. The annotators are certified before starting the annotation job. The general process that an annotator goes through is depicted in Figure 2:

Figure 2: The general process for an annotator to start the job.

When a new annotator enters the project, the following process should pass so that he/she can be certified for starting the annotation job:

- The CV of the medical expert annotator is verified by the experts.
- A session is set in which the project, its objective, the motivations, and process are explained.
- The general workshop is made in which the annotator learns the basics of how to annotate and how to work with the platform.
- The annotator takes the general exam. The pass mark is set as 80%. If he/she does not pass, he/she must take another general exam.
- When the general exam is passed, the annotator must follow a specialised workshop in which the ontology and the incision boundary annotation is explained.
- The annotator takes the specialised exam. The pass mark is set as 75%. He has two attempts for each exam. If he/she fails, he has another two attempts for another specialised exam.
- When the annotator passes, he is certified for annotations and can start the job of videos.

5.3.Ontology and Guidelines

The initial basis for annotation is to have a common vocabulary for all the annotators which must be clearly defined. Since the objective of this WP is completely innovative, no previous ontologies are existing in this regard. Therefore, all the ontology, procedure and guidelines must be defined inside this WP. The complexity of such annotation that must be considered are the following:

- The complexity of annotation of division plane since it requires 3D (in depth) information,
- The high inter-variability of surgeon's approaches to operate the endometriosis lesions,
- The complexity of the definition of a common ontology adapted to both AI and surgical operations,
- The uncertain accuracy of junior surgeon's annotation due to their lack of experience,
- The uncertain accuracy of expert surgeon's annotation due to their lack of time,
- The lack of pre-operative information (e.g., patient's age, MRI, etc.),
- The lack of a standardisation of operation between surgeons.

The annotation guideline has tried to be as accurate as possible and to consider all the above.

It is of great importance to firstly remember the ontology which was defined for WP6 through a consensus Delphi study. This is depicted in figure 3.

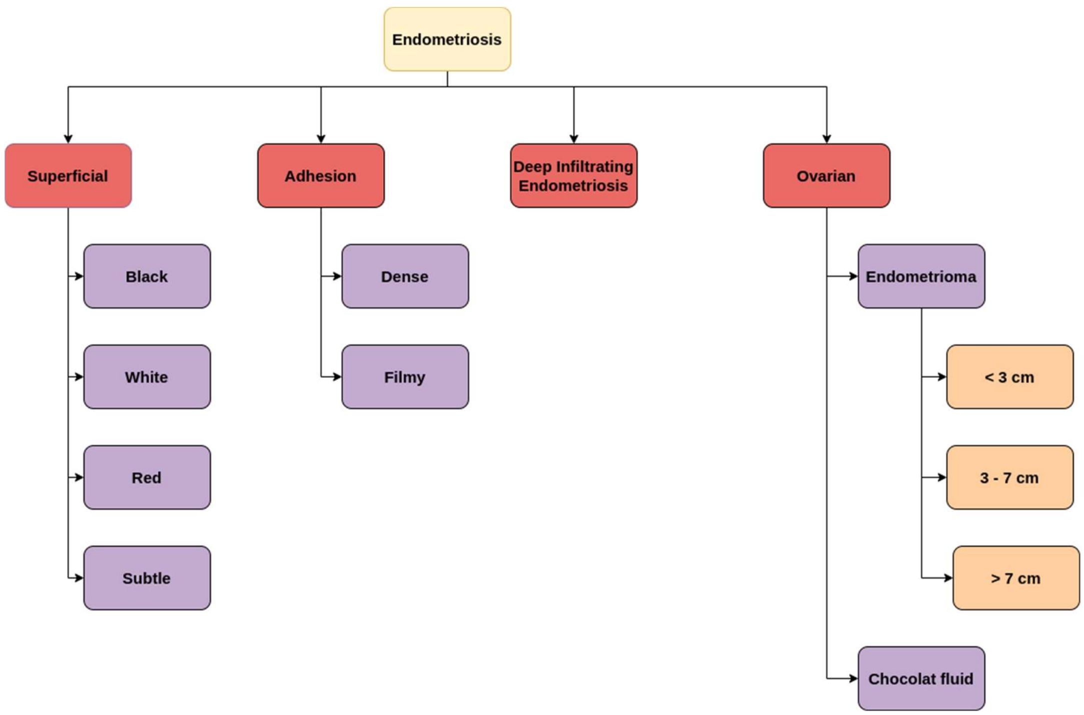


Figure 3: The Ontology for Endometriosis Classification.

The Lesions

Here in WP7 the annotations will be on the incision area and not the whole division plane. This is because in-depth annotations will have a lot of complexities. The including and excluding lesions to be annotate are as follows:

- Including lesions:
    - Superficials (Black, White, Red, Subtle).
    - Deep Endometriosis.

- Excluding lesions:
    - Adhesions (Dense & Filmy).
    - Ovarian (Endometrioma & Ch. fluid).
    - Adenomyosis and uterine superficial lesions.

The Annotation Shapes and Classes

The incision area is marked by polygons. They should be as accurate as possible. If an instrument crosses a zone of disease, we exclude it from the annotated area.

We have 2 classes of annotated area:

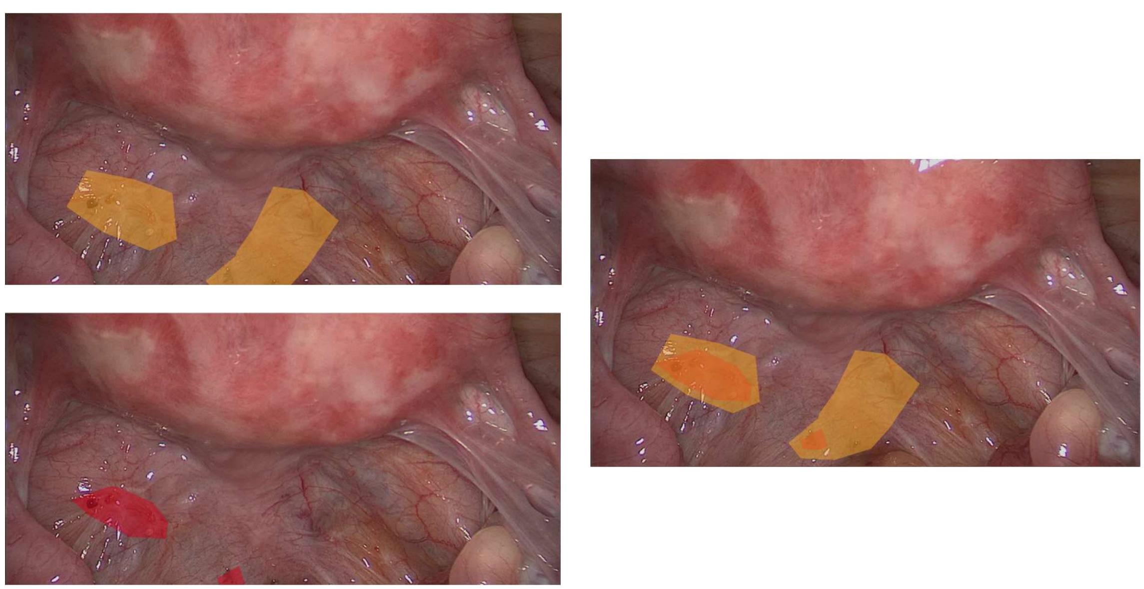
- *Resect* zone: The zone that must be treated. The purpose is to indicate to the surgeon the lesions that must be treated, regardless of the technique chosen by the surgeon (surgical excision, coagulation, laser...).

- *Check* zone: The surgeon should consider him/herself in a realistic situation. This annotation has a double purpose.
  - Firstly, the *check* zone helps the operator to define an adequate safety margin around the *resect* zone, to guarantee the complete removal of the endometriotic lesion and any perilesional fibrosis.

  - Secondly, to increase the sensitivity of this tool, the *check* zone allows the surgeon to indicate any areas which, in the analysed frame, cannot be classified as certainly healthy or certainly site of endometriosis: areas which, therefore, must at least be carefully analysed by the surgeon. Therefore, in case of doubt (for example if the frame is not high quality), signalling an area as a *check* zone allows the increasing of the sensitivity of the tool, and it ensures complete recognition of the lesions.

This consensus has been made according to hours of discussion and a considerable number of annotations done by surgeons (It will be explained in Section 5.4). Due to the high complexity of such a task, and the high variability of different surgeon annotations, some other changes might affect the presented ontology for improvement. The described classes are shown in Figure 4:

Figure 4: The *resect* area is annotated in red. The *check* area is annotated in yellow. The final annotation of this frame is depicted on the right.



5.4. Pipeline and Procedures

Since the annotation of surgical incision boundaries is a sensitive subject which directly affects the health of the patients, the annotations are trustable if they are annotated by expert surgeons. However, in practice we face some complexities:

- experts have less time and motivation to do annotations,
- they put less concentration, resulting in inaccurate annotations.

Therefore, the strategy taken in this regard was to design a plan whose objective is to take the least possible time from experts, and the most qualitative annotations. This is done by having both expert and junior surgeons and planning to reach a consensus for the annotations. So, we have a study plan (see figure 5) for junior surgeons with the following motivations and objectives:

- train them for surgeries by practising with annotation,
- discuss their annotations with experts to reach a consensus on the medical aspect.

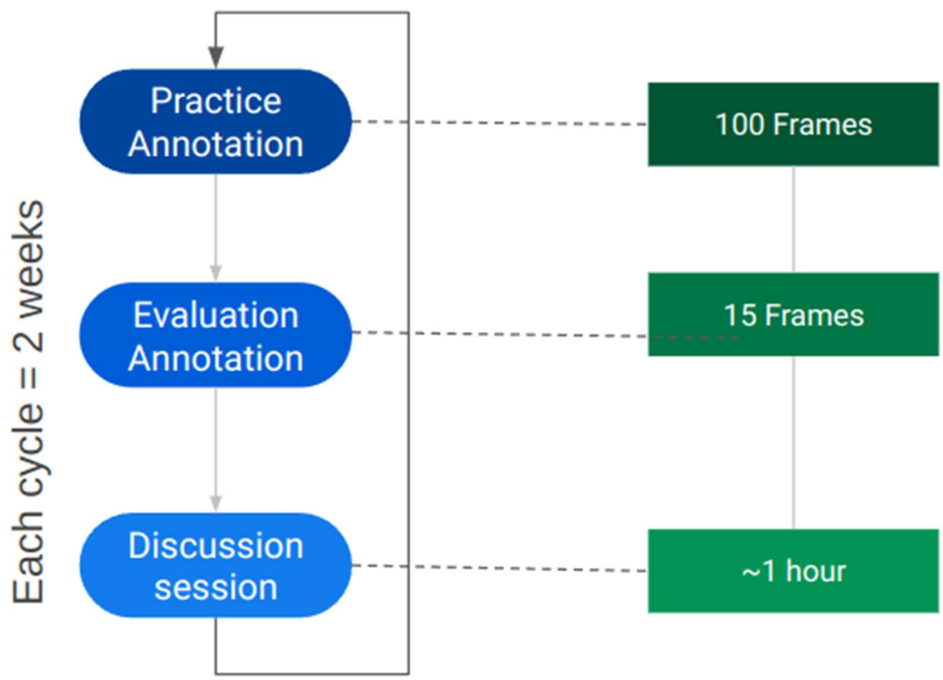


Figure 5: The design for junior surgeon annotations.

This study consists of three phases which are repeated every two weeks.

1. Practice: In this phase the junior surgeons are given 100 different video frames to annotate. So that they can practise while annotating. It should be noted that the annotators are encouraged to watch the entire short video before annotating the selected frames.
2. Evaluation: In this phase, 15 frames (from 15 different surgeries) are given to the junior surgeons for being annotated so that they can be evaluated according to these frames' annotations. The expert surgeons also annotate the same 15 frames.
3. Discussion session: The data manager will prepare the annotations in a readable format for this session in which all the experts and junior surgeons gather, and they all see each other's annotations and discuss the challenges, the consensus, and the best annotation.

6. Results

6.1. Statistics

To date, more that 19 Person-Hour is dedicated to only **discussions** between experts to reach a consensus for the annotations, the zones, and the procedures. People participating in such discussions were surgeons, including expert and junior surgeons and ML scientists. The amount of time (which is enormous) dedicated to do the annotation job itself is not considered in these 19 Person-Hour calculations. The number of annotations for each defined class is summarised in Table 1:

TOTAL

472 Frames

CLASS	FIGURES
Total	**3053**
To Resect	1554
To Check	1499

Table 1: The number of annotations on incision boundaries.

6.2.Variabilities

As already mentioned, there is a high variability between the annotator's ideas in annotating incision boundaries. The amount of agreement between the two expert surgeons is depicted in figure 6.

The merged zone is considered as the union of *resect* and *check* zones. The bars in the figures are calculated as agreement rate which is calculated as the following formula for any two annotators noted as Ann 1 and Ann 2, on each zone:

$$Agreement\ Rate = \frac{Intersection\ (Ann\ 1\ , Ann\ 2)}{Union\ (Ann\ 1, Ann\ 2)}$$

Each cycle takes two weeks and is already defined in the previous section. The cycles are time indicators. This figure shows that through time, doing more annotations and more discussions dedicated to this task, even the two expert surgeons agree more to each other on the annotations.

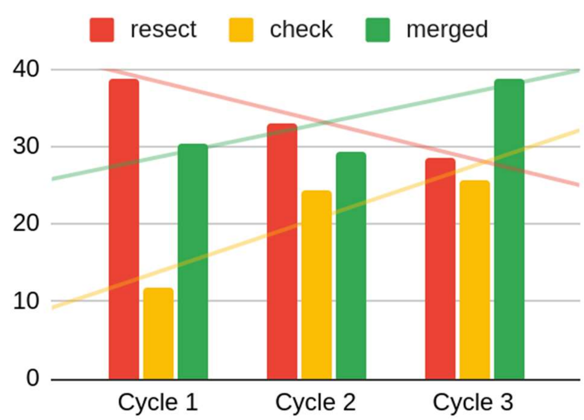


Figure 6: The amount of agreement between two experts with time.

To reach the objective of training the junior surgeon to annotate better, Figure 7 and Figure 8 are shown which depict the progress of the two junior surgeons in the study. Both figures show the average increase in the progress of junior surgeons. Of course, a trendline with a sharper increase is expected after next cycles.

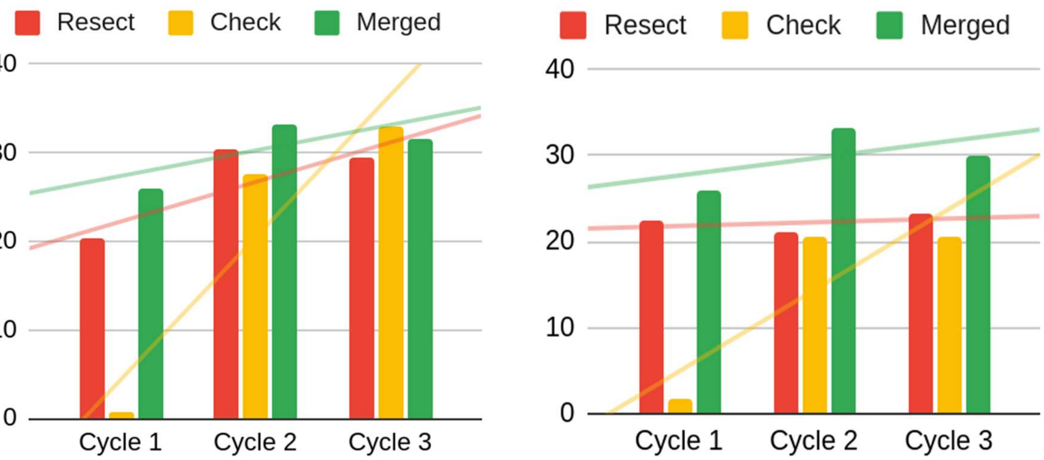


Figure 7: The agreement rate of annotator, 1 with expert surgeon 1 (on Left), and with expert surgeon 2 (on Right).

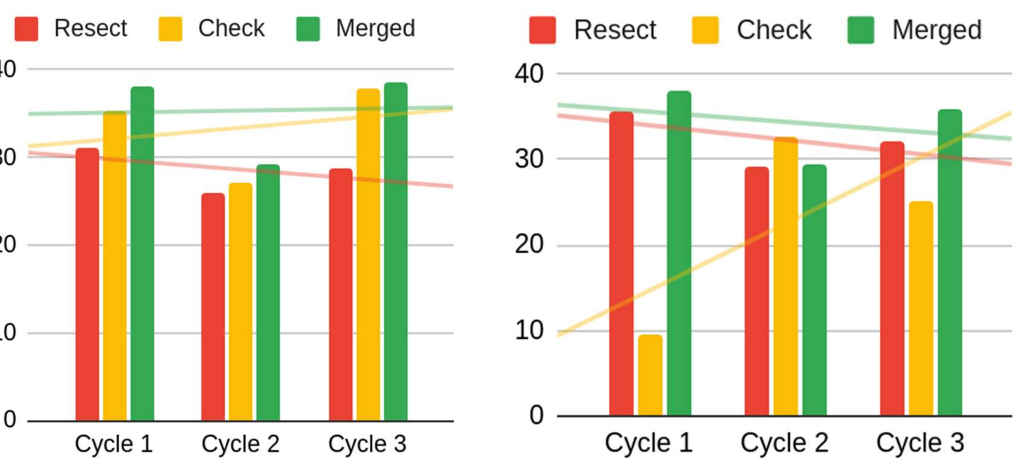


Figure 8: The agreement rate of Annotator 2 with expert surgeon 1 (on Left), and expert with surgeon 2 (on Right).

To state the complexity of variability and emphasise on the need to standardise such annotations to reduce such variability, some of the laparoscopic frames and the different annotations are depicted in Figure 9. In these figures two image frames are depicted (in the left and right column). On top, and bottom the annotation of *resect*, and *check* zones are depicted, respectively for all the 4 annotators. The colour bar shows the percentage of agreement between the 4 annotators for the segmented pixels.

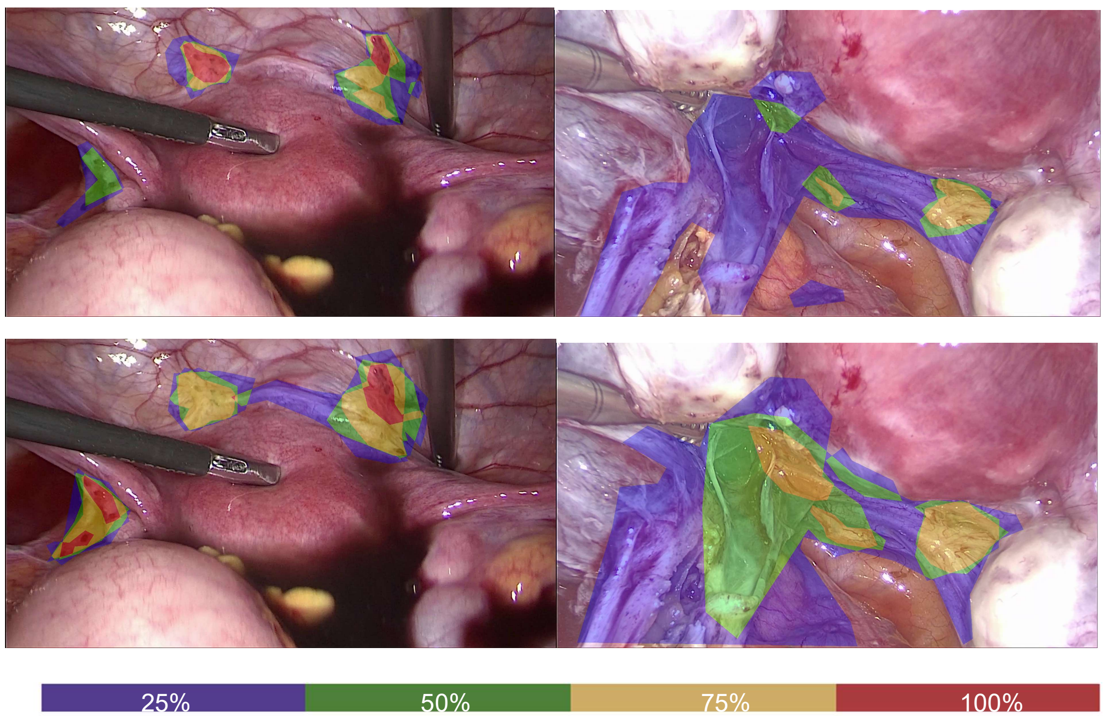


Figure 9: The inter-surgeon variabilities of annotations among 4 surgeons (on top *Resect*, and on bottom *Check* zones).

7. Conclusions and Discussions

WP7 is dedicated to an AR tool to suggest the surgeons the division planes in endometriosis surgeries. Since the division plane is a general word which needs in depth recognition of the videos, the incision boundaries are defined as the first objective to be recognized and be suggested to the surgeons. Big data from around the world (by considering the variability for trustworthy AI) is collected and they should be annotated for the task of supervised learning in deep learning.

The annotation of the laparoscopic videos is explained. This annotation can be done for two different objectives: First, they can help to standardise the endometriosis operations, which does not exist among the surgeons, now. If a good consensus is achieved, it can be generalised to train the surgeons in a more spatial resolution (even world-wide).

Second, it can result in a rich data for the task of machine learning and automatization of such surgeries. The annotations can be considered as ground truth, and they can be learned by machines through neural networks, and finally a platform can be developed to give surgeons suggestions about the incision areas. As the long-term objective, this can be considered as the first step in full robotic surgeries.

To achieve such tasks, the annotations are being done in this task in the FEMaLe project. There are complexities in this regard which were discussed in this report. One of the main ones is the inter-surgeon variability which leads to large variability in annotations and a noisy annotation. That is why the leaders of WP7 (SURGAR) have a study plan to manage such annotations and to reach a consensus in this regard. This can not only result in less variability and thus more trustable data labels, but also in a more standardised surgery which can have a large impact in the medical field, as well as in the technology aspects.

Finally, we have to note that the task of annotating the data is still ongoing to generate an acceptable quantity of annotations. The delay in providing the expected quantity at this moment is because of the considerable attention taken to the quality to provide more trustable data. This is one of the aspects of trustworthiness of AI which does not have to be neglected.