Prospective multi-centre validation study on ovarian cancer risk prediction using AI-models applied to ultrasound images. (OV-AID, Phase I)

Purpose and effect

This project aims to improve the diagnosis and management of women with ovarian tumours – through artificial intelligence (AI) using deep neural networks (DNN) applied to ultrasound images.

We hypothesize that DNN model performance from our previous studies[1] generalizes in a prospective setting, with images acquired and assessed by examiners of various levels of expertise, with different ultrasound systems, and patient cohorts.

We anticipate that DNN models can be used in the triage of women with ovarian tumours, aiding and improving clinical decision-making. In the case of non-expert examiners, we expect an AI-driven support tool to reduce the need for second-opinion referrals and unnecessary surgery, potentially detect cancer at an earlier stage, and reduce patient anxiety; thus, resulting in a more cost-effective utilization of healthcare resources and improved wellbeing and quality of life among patients.

State-of-the-art

Ovarian cancer is a rare gynaecological malignancy with a poor prognosis. While benign ovarian tumours occur in 1 in 10 women, they are often asymptomatic and hence incidentally detected. This puts great demand on the healthcare system for triaging women to the optimal management. Benign masses can be managed conservatively with ultrasound follow-up or with minimal invasive surgery, while reducing morbidity and avoiding unnecessary fertility loss. Women with suspected ovarian cancer should be referred directly to a gynae-oncology centre, as surgical treatment of such patients by gynaecological oncologists is associated with higher likelihood of complete tumour removal and improved survival rate[2].

There is currently a shortage of examiners with experience and competence in accurately discriminating benign from malignant ovarian tumours, particularly in developing countries, but also in high-income countries as Sweden. Most gynaecologists see only a few patients with complex ovarian tumours annually, and therefore, have difficulties increasing their diagnostic skills. We estimate that less than 10% of gynaecologists have expert competence in assessing ovarian tumours. Studies have shown that assessment by less experienced examiners results in low diagnostic accuracy[3, 4], and thus suboptimal use of healthcare resources, with unnecessary surgery and delayed cancer diagnosis. We estimate that $> 25\%$ of ovarian surgeries could be avoided if expert ultrasound assessment was available, generating a yearly saving of one billion SEK. Thus, we see a definite need to improve the diagnostic

accuracy in differentiating benign from malignant ovarian lesions among non-expert examiners.

It is not advisable to take a biopsy from ovarian tumours as this might spread the tumour in case of malignancy, worsening the prognosis[5]. Tumour markers have been used for decades to improve diagnostic accuracy. However, all known biomarker have weaknesses, including the most commonly used cancer antigen 125 (CA125), which is increased only in $< 50\%$ of patients with epithelial ovarian cancer, stage I [6] and non-epithelial ovarian cancer. Moreover, CA125 may be elevated also in women with benign conditions, such as endometriosis, fibromas, inflammatory processes, ascites from non-malignant causes, and pregnancy[7, 8].

Recent advances in artificial intelligence (AI), using deep neural networks (DNN), have shown promising results in discriminating between benign and malignant tumours with performance on par with expert radiologists in other domains[9-11]. However, until recently, there has been no data on the use of DNN models for assessing ovarian tumours. Our previous retrospective internal study, including > 3000 images from 750 women, was the first to indicate that DNN models can differentiate benign and malignant tumours, with a diagnostic accuracy similar to that of an expert examiner[12]. A recent Chinese study showed promising results in a retrospective multi-centre validation cohort[13]. Furthermore, our own [unpublished] multi-centre validation study (20 centres), including over 17 thousand images from 3652 cases, shows a robust model performance across centres with a median AUC of 0.93[preliminary], outperforming the vast majority of experts ($n = 25$) and non-experts ($n = 41$).

However, these results need to be prospectively validated in a multi-centre setting to demonstrate ensure robustness in a clinical setting.

Project description

Aim:

To compare the accuracy in differentiating benign from malignant masses using DNN models as compared to subjective assessment, on-line or off-line, using patter recognition or the IOTA-ADNEX model[14], by examiners of different levels of expertise.

Variables and measures:

Prospective study including ≥ 800 patients (incl. ≥ 300 malignant) with ovarian tumours, assessed by examiners with varying expertise (including at least $\geq$ 450 assessments by non-experts, and 350 by experts). Subjective assessment by the examining doctor using pattern recognition and the IOTA-ADNEX model score will be compared to DNN-model assessment. DNN-model performance will also be compared to

external off-line expert ($n \geq 5$) and non-expert assessment (n ≥ 5). Histological outcome from surgery or ultrasound follow-up (after a minmum of 9 months, i. e. with follow − up exams at 3 and 9 months from inclusion, with no suspicion of malignancy) serves as gold standard. We will use a final DNN model from our own validation-study for the analysis.

Inclusion criteria and case selection:

Consecutive women with a newly (<4 months) detected adnexal lesion, planned for surgery or ultrasound follow-up. Patients > 15 years, given their informed consent to participate.

Estimated sample size:

We will have 80% power to detect the superiority of the DNN model compared to non-expert examiners (79% vs. 86%) based on the validation study https://doi.org/10.1186/ISRCTN51927471 . We will use a two-sided conditional logistic regression (Mantel-Haenszel test) and assume a baseline probability of 0.8. To achieve this power, we need to include 450 assessments by non-expert examiners.

Additionally, we will have a power of over 80% to detect non-inferiority to expert examiners with a difference of 0.05. Assuming a baseline probability of 0.9, we will include 350 assessments by expert examiners.

Statistical analysis:

We will calculate and compare F1, accuracy, sensitivity and specificity, and area under the receiver operating characteristic curve (ROC-AUC) for IOTA subjective assessment using pattern recognition, and IOTA ADNEX model with that of the DNN model. Sub-analysis will be performed for expert examiners and non-expert examiners, respectively.

Status and timeline:

We have already included 490 cases at Södersjukhuset, and the 16 external centers started inclusion in April 2023. Inclusion will be finalized once the estimated sample size has been reached.

Ethical approval:

Ethical approval has been obtained from Swedish Ethical Review Authority https://etikprovningsmyndigheten.se/en 2020-07200, with addendum 2021-04549, Dnr 2021-06357-02, and Dnr 2023-01834-02.

Study registration (ISRCTN) for phase I: https://doi.org/10.1186/ISRCTN88222986

Study CRF

We will use the electronic data management system REDCap (Research Electronic Data Capture) as it offers numerous benefits as a database for a clinical multicenter study (robust and secure database for clinical multicenter studies, providing efficient data management, strong data security, collaboration capabilities, data quality control, support for longitudinal studies, data export options, and project management tools).

Required image acquisition:

For every case ≥ 4 grayscale, ≥ 2 power Doppler still images (preferentially without callipers) and ≥ 2 slow (5–10 seconds long) video-clips scanning through the lesion from one side to the other (at least one with and one without power Doppler) should be collected. In case the lesion is large it is desirable to also include transabdominal images and video-clips. Make sure to optimize the image before saving – so that the lesion is filling approximately 75% of the image (Figure 1). Do **not** use the split images – save only single images. Do **not** include images with biopsy callipers, as these might confer a risk of bias (Figure 2). In case where there are bilateral lesions, each lesion shall present a separate case, if both lesions are included.

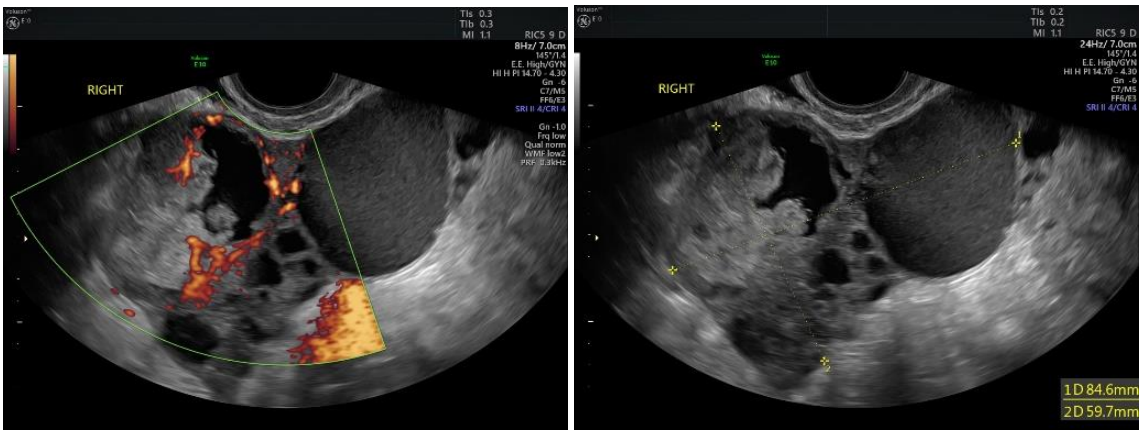

Figure 1. Example of adequate images; whole lesion seen, lesion borders seen, adequate resolution. Top blacked out or de-identified.

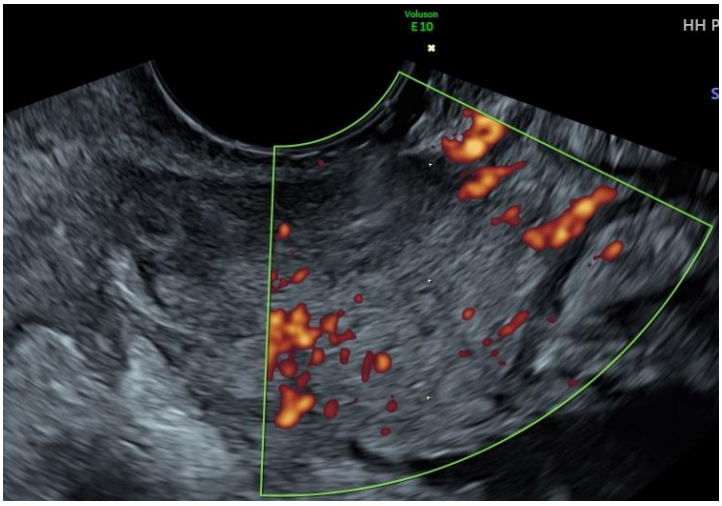

Figure 2: Do **not** include images with biopsy callipers (white dotted, vertical line).

File management:

When the study is finalized, we request centres to prepare and send images according to the instructions below.

Image anonymization:

De-identify images by "blacking out the top" – Do **not** crop images! (Figure 3A how to anonymize and export from Viewpoint and Figure 3B how to anonymize directly in GE ultrasound system). In case you use other ultrasound systems or workstations, please contact the manufacturer for instructions. Lastly, put images in a *case folder* with the same name as the ***CASE-ID*** as written in the file;
`ViewPoint_variables_other_centres_20230424.xls`

Anonymizing and exporting images from Viewpoint

Selected images to export

Push – *exportera bilder*

A) Make sure image top is correctly blacked out
B) Mark images with the patients CASE-ID
C) Push Exportera

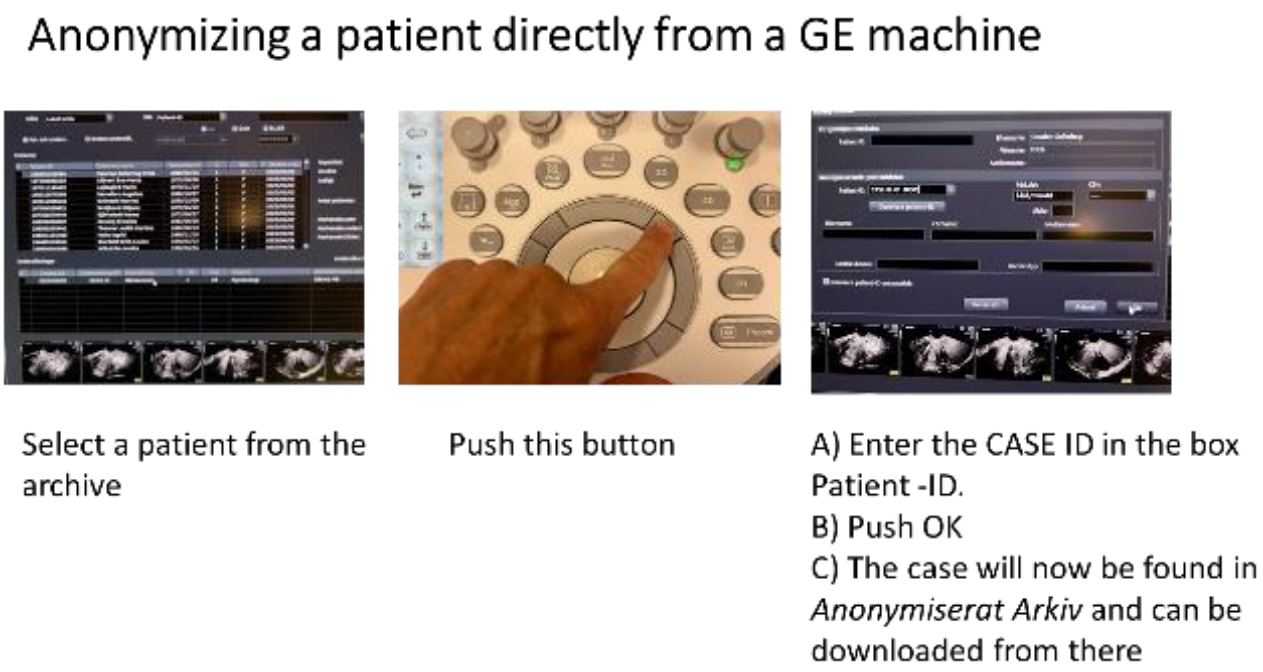

Figure 3. A) How to anonymize and export from Viewpoint, B) How to anonymize directly from GE´s ultrasound systems

Summary

Per case:

- Still images: ≥ 4 grayscale, ≥ 2 power Doppler
- Video-clips: ≥ 1 grayscale, ≥ 1 power Doppler

Important:

- Do ***not*** use the split images. Save only single images.
- Do ***not*** include images with biopsy callipers (Figure 2)
- De-identify images by "blacking out the top" – Do ***not*** crop images!
- ***Avoid*** callipers when possible.
- Optimize the image before saving so that the lesion is filling approximately 75% of the image (Figure 1).
- Video: ***Slowly*** and ***steadily*** scan through the lesion from one side to the other, preferentially 5-10 sec long.
- In case the lesion is large it is desirable to also include transabdominal images and video-clips.

Image format:

If possible, send images as JPEG (still images) and MP4 for videoclips; however, other image formats are also accepted, such as DICOM.

Sending images:

The de-identified images shall be sent to us via https://send.tresorit.com, when the study is closed, or prior to that in case we decide to do an interim analysis. All images for a given case shall be put in a ***case folder*** with the same name as the ***CASE-ID*** as written in the file `ViewPoint_variables_other_centres_2023-04-24.xls`. All benign cases shall then be put in the ***benign folder*** and the malignant in the ***malignant folder***. The ***benign folder*** and the ***malignant folder***, together with the file `ViewPoint_variables_other_centres_2023-04-24.xls`, shall finally be put in the ***CENTRE-ID folder*** (Figure 4). The ***CENTRE-ID folder***, containing all material, shall be compressed/zipped (Figure 5) and then be sent via https://send.tresorit.com. By following the simple instructions given on the webpage you will be given a link which shall be sent by email to filipchr@kth.se and elisabeth.epstein@ki.se.

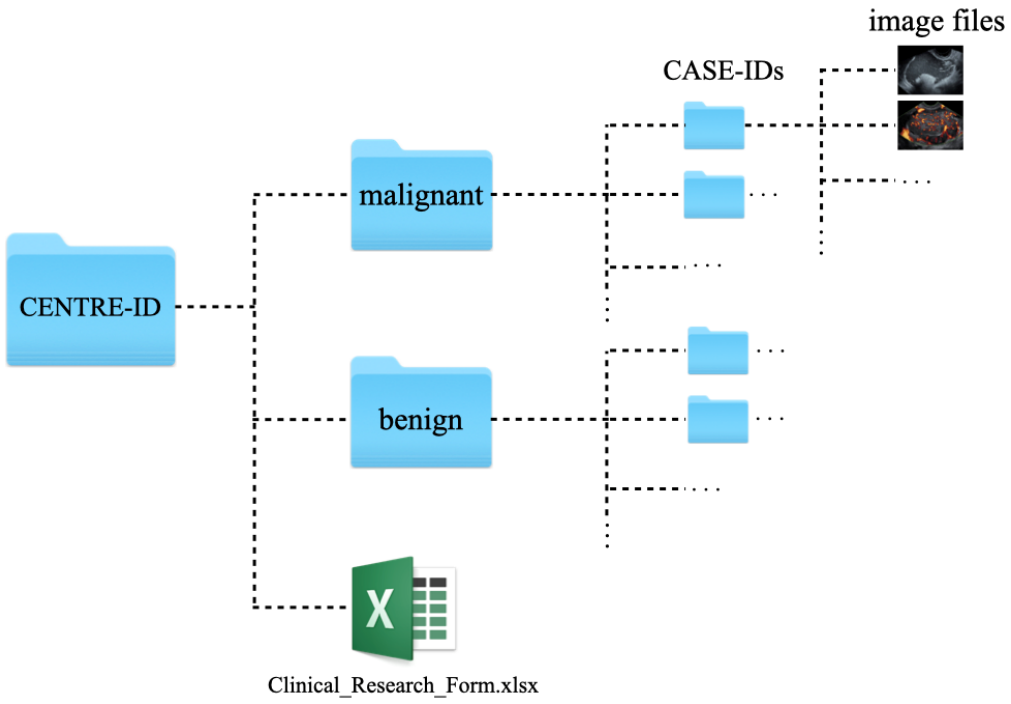

Figure 4. Arrangement of images and folders when uploading to Tresorit.

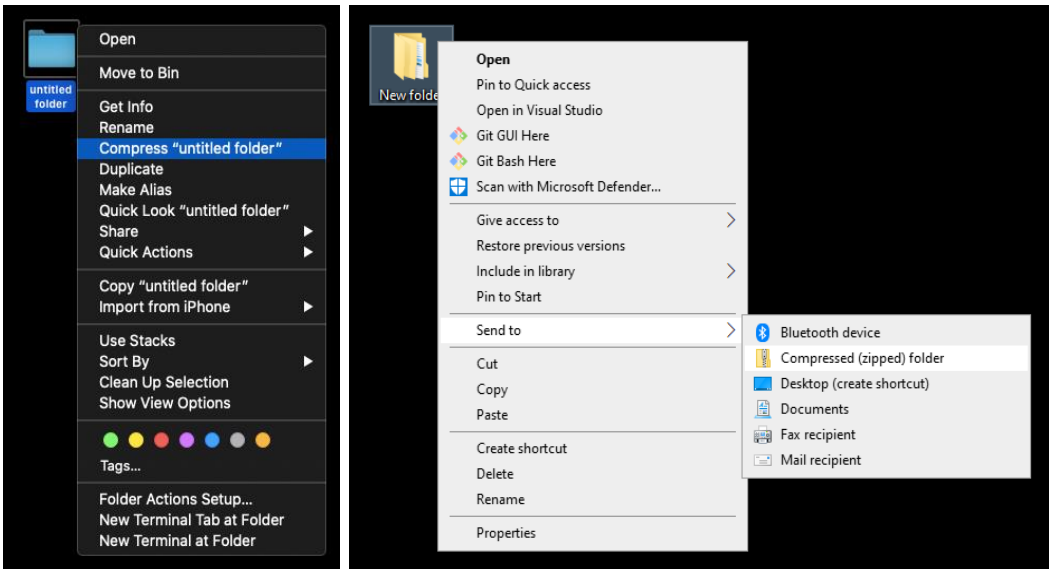

Figure 5. How to compress/zip a folder on MacOS (left) and Widows 10 (right).

Intellectual property:

The Department of Clinical Science and Education/Karolinska Institutet, Stockholm, Sweden KI owns the raw data/original information that originates from this study, that is handled and stored within KI. Intelligyn, has the right to use de-identified data originating from the KI data-set, including, the aggregated dataset and the trained and validated DNN models in their product Intelligyn AI, for the purpose of continuously improving its functionality. Intelligyn´s AI platform will be used for the prospective randomized study, the next step towards clinical implementation. All participating centres retain full and unrestricted rights to the use of their own data.

Publication policy:

The steering committee is responsible for publication of the data in scientific journals. Principal investigators from each collaborating centre are considered for co-authorship, after a minimum 20 included cases, on condition that they contribute to writing the papers, read and approve the final version, and agree to be accountable for all aspects of the work, as defined by the International Committee of Medical Journal Editors and in accordance with the requirements of the respective medical journal. In addition, co-authorship is prioritized by the number of cases that investigators contribute to the study, given that there is a journal restriction.

Study coordinators:

Erica Smedberg, *Department of Clinical Science and Education, Karolinska Institutet, Stockholm, Sweden. Mail:* erica.smedberg@regionstockholm.se

Elisabeth Epstein, *Department of Clinical Science and Education, Karolinska Institutet, Stockholm, Sweden. Mail:* elisabeth.epstein@ki.se

Filip Christiansen, *School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Stockholm, Sweden. Mail:* filipchr@kth.se

Steering committee

Elisabeth Epstein, *Department of Clinical Science and Education, Karolinska Institutet, Stockholm, Sweden*

Filip Christiansen, *School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Stockholm, Sweden*

Kevin Smith, *Science for Life Laboratory, School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Stockholm, Sweden*

Collaborating centres:

The study was initiated at *Stockholm South General Hospital* (*Södersjukhuset*) in March 2021. Since April 2023, the following 16 additional Swedish centres have been including cases:

1. Aleris UltraGyn, Sabbatsberg Hospital *(Sabbatsbergs sjukhus), Stockholm*
2. Central Hospital Karlstad *(Centralsjukhuset Karlstad)*
3. Central Hospital Växjö *(Centrallasarettet Växjö)*
4. Danderyd Hospital *(Danderyds sjukhus)*
5. GynStockholm, Cevita Care, *Stockholm*
6. Hallands Hospital Halmstad *(Hallands sjukhus Halmstad)*
7. Huddinge Hospital *(Karolinska Universitetssjukhuset Huddinge)*
8. Linköping University Hospital *(Universitetssjukhuset i Linköping)*
9. North Älvsborg County Hospital *(Norra Älvsborgs Länssjukhus), Trollhättan*
10. Nyköpings Lasarett, *Nyköping*
11. Sahlgrenska University Hospital *(Sahlgrenska Universitetssjukhuset), Gothenburg*
12. Skåne University Hospital *(Skånes Universitetssjukhus), Lund*
13. Skåne University Hospital *(Skånes Universitetssjukhus), Malmö*
14. Uppsala University Hospital *(Akademiska sjukhuset)*
15. Örebro University Hospital *(Universitetssjukhuset Örebro)*
16. Masthugget Gynecology outpatient unit, *Gothenburg (Masthugget gynekologi och obstetrik mottagning, Gothenburg)*

Other potential new centers

1. Sygehus Sønderjylland, Denmark, (Ervin Kallfa)

Furthermore, there are pending invitations to the following 18 international collaborating centres from our retrospective multi-centre validation study:

1. Charles University and General University Hospital in Prague, *Prague, Czech Republic* (Lucia Haak)
2. Clínica Universidad de Navarra, *Pamplona, Spain* (Juan Luis Alcázar)
3. European Institute of Oncology IRCCS, *Milan, Italy* (Dorella Franchi)
4. First Faculty of Medicine, Charles University and General University Hospital in Prague, *Prague, Czech Republic* (Daniela Fischerová and Petra Šašková)
5. Fondazione Poliambulanza Istituto Ospedaliero, *Brescia, Italy* (Elisa Mor)
6. Hospital Universitario Dexeus, *Barcelona, Spain* (Maria Àngela Pascual)
7. Institute for Maternal and Child Health, IRCCS "Burlo Garofolo", *Trieste, Italy* (Francesca Buonomo)
8. Lithuanian University of Health Sciences, *Kaunas, Lithuania* (Adrius Gaurilcikas)
9. Luigi Sacco University Hospital, *Milan, Italy* (Francesco Leone)
10. Mater Olbia Hospital, *Olbia, Italy* (Debora Verri)
11. Medical University of Lublin, *Lublin, Poland* (Artur Czekierdowski)
12. Medical University of Silesia, *Sosnowiec, Poland* (Marek Kudla)
13. National and Kapodistrian University of Athens, *Athens, Greece* (Ekaterini Domali)
14. Ospedale "G.Salesi", *Ancona, Italy* (Nina Montik)
15. Policlinico Universitario Duilio Casula, *Monserrato, Cagliari, Italy* (Stefano Guerriero)
16. Policlinico Sant'Orsola-Malpighi, *Bologna, Italy* (Luca Savelli and Maria Munaretto)
17. Rizal Medical Center, *Manila, Philippines* (Nelinda Catherine Pangilinan)
18. San Gerardo Hospital, *Monza, Italy* (Robert Fruscio)

References

1. Christiansen F, Epstein E, Smedberg E, Akerlund M, Smith K, Epstein E. Ultrasound Image Analysis using Deep Neural Networks to Discriminate Benign and Malignant Ovarian Tumours - A Comparison to Subjective Expert Assessment. *Ultrasound Obstet Gynecol* 2020.
2. Bristow RE, Tomacruz RS, Armstrong DK, Trimble EL, Montz FJ. Survival effect of maximal cytoreductive surgery for advanced ovarian carcinoma during the platinum era: a meta-analysis. *J Clin Oncol* 2002; **20**: 1248-1259.
3. Van Holsbeke C, Daemen A, Yazbek J, Holland TK, Bourne T, Mesens T, Lannoo L, Boes AS, Joos A, Van De Vijver A, Roggen N, de Moor B, et al. Ultrasound experience substantially impacts on diagnostic performance and confidence when adnexal masses are classified using pattern recognition. *Gynecol Obstet Invest*; **69**: 160-168.

4. Van Holsbeke C, Daemen A, Yazbek J, Holland TK, Bourne T, Mesens T, Lannoo L, De Moor B, De Jonge E, Testa AC, Valentin L, Jurkovic D, et al. Ultrasound methods to distinguish between malignant and benign adnexal masses in the hands of examiners with different levels of experience. *Ultrasound Obstet Gynecol* 2009; **34**: 454-461.
5. Vergote I, De Brabanter J, Fyles A, Bertelsen K, Einhorn N, Sevelda P, Gore ME, Kaern J, Verrelst H, Sjovall K, Timmerman D, Vandewalle J, et al. Prognostic importance of degree of differentiation and cyst rupture in stage I invasive epithelial ovarian carcinoma. *Lancet* 2001; **357**: 176-182.
6. Bast RC, Jr., Badgwell D, Lu Z, Marquez R, Rosen D, Liu J, Baggerly KA, Atkinson EN, Skates S, Zhang Z, Lokshin A, Menon U, et al. New tumor markers: CA125 and beyond. *Int J Gynecol Cancer* 2005; **15 Suppl 3**: 274-281.
7. Niloff JM, Klug TL, Schaetzl E, Zurawski VR, Jr., Knapp RC, Bast RC, Jr. Elevation of serum CA125 in carcinomas of the fallopian tube, endometrium, and endocervix. *Am J Obstet Gynecol* 1984; **148**: 1057-1058.
8. Park Y, Lee JH, Hong DJ, Lee EY, Kim HS. Diagnostic performances of HE4 and CA125 for the detection of ovarian cancer from patients with various gynecologic and non-gynecologic diseases. *Clin Biochem* 2011; **44**: 884-888.
9. McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafian H, Back T, Chesus M, Corrado GC, Darzi A, Etemadi M, Garcia-Vicente F, et al. International evaluation of an AI system for breast cancer screening. *Nature* 2020; **577**: 89-94.
10. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017; **542**: 115-118.
11. Ardila D, Kiraly AP, Bharadwaj S, Choi B, Reicher JJ, Peng L, Tse D, Etemadi M, Ye W, Corrado G, Naidich DP, Shetty S. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat Med* 2019; **25**: 954-961.
12. Christiansen F, Epstein EL, Smedberg E, Åkerlund M, Smith K, Epstein E. Ultrasound image analysis using deep neural networks for discriminating between benign and malignant ovarian tumors: comparison with expert subjective assessment. *Ultrasound in obstetrics & gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology* 2021; **57**: 155-163.
13. Gao Y, Zeng S, Xu X, Li H, Yao S, Song K, Li X, Chen L, Tang J, Xing H, Yu Z, Zhang Q, et al. Deep learning-enabled pelvic ultrasound images for accurate diagnosis of ovarian cancer in China: a retrospective, multicentre, diagnostic study. *Lancet Digit Health* 2022; **4**: e179-e187.
14. VanCalster B, VanHoorde K, Valentin L, Testa AC, Fischerova D, VanHolsbeke C, Savelli L, Franchi D, Epstein E, Kaijser J, VanBelle V, Czekierdowski A, et al. Evaluating the risk of ovarian cancer before surgery using the ADNEX model to differentiate between benign, borderline, early and advanced stage invasive, and secondary metastatic tumours: prospective multicentre diagnostic study. *BMJ (Clinical research ed.)* 2014; **349**: g5920.