ARTICLE TYPE

Diagnostic accuracy of transvaginal ultrasound (TVUS) and diffused weighted imaging (DWI) in female patient with ovarian cancer: a systematic review and meta-analysis

Ainul Shafiqah Zulkifli, Mohd Zulfadli Adenan

Centre of Medical Imaging, Faculty of Health Sciences, Universiti Teknologi MARA Cawangan Selangor Kampus Puncak Alam, 42300 Bandar Puncak Alam, Selangor, Malaysia

***Corresponding Author**

Mohd Zulfadli Adenan
Email: mohdzulfadli@uitm.edu.my

Abstract:

This review is aimed to compare the diagnostic accuracy of transvaginal ultrasound (TVUS) and diffused weighted image (DWI) in assessing ovarian cancer. An extensive search was performed in Dimension, Google Scholar and Wiley Online Library for studies published from January 2011 to April 2020. Following the screening, the eligibility of the studies was checked. Then, the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) checklist was adapted to assess the quality of each of the study being reviewed. The information on the diagnostic performance of TVUS and DWI were extracted from each of the studies. The heterogeneity of the studies was explored. Overall, pooled sensitivity, specificity, LR+, and LR− of TVUS for detecting ovarian cancer were 75% (95% confidence interval [CI]=57%–87%), 92% (95% CI=80%–97%), 8.97 (95% CI=3.21–25.08), and 0.28 (95% CI=0.14–0.52), respectively. High heterogeneity was established for sensitivity (I^2=92.28%; Cochran Q=38.85; p=0.00). Furthermore, high heterogeneity was found for specificity (I^2=88.54%; Cochran Q=26.17; p=0.00). TVUS has high specificity for detection of ovarian cancer. TVUS should be considered as good enough for being used in clinical settings with limited resources.

Keywords: Diffused weighted imaging, transvaginal ultrasound, ovarian cancer

1. INTRODUCTION

Cancer incidence and mortality are rapidly growing worldwide. Ovarian cancer is one of the most common gynecologic cancers that rank third after cervical and uterine cancer. It also has the worst prognosis and the highest mortality rate [1]. Although ovarian cancer has a lower prevalence in comparison with breast cancer, it is three times more lethal [2]. The high mortality rate of ovarian cancer is caused by asymptomatic and secret growth of the tumor, delayed onset of symptoms, and lack of proper screening that result in its diagnosis in the advanced stages [1]. In 2015, ovarian cancer was diagnosed in 21, 290 women in the United States, and more than 14, 000 died from this disease. Effective early detection of ovarian cancer through regularly repeated screening tests may have a real impact on survival and, potentially, on mortality from the disease [3].

Type I ovarian cancers are generally large, unilateral, cystic tumors at diagnosis with indolent behavior. They are thought to usually develop from extraovarian benign lesions that embedded in the ovary and subsequently undergo a series of mutations resulting in malignant transformation. In this way, low- grade serous carcinomas are thought to originate from benign deposits of fallopian tube epithelium in the ovaries (endosalpingiosis); endometrioid and clear cell carcinomas from benign foci of endometrial tissue in the ovaries (endometriosis); and most mucinous carcinomas from benign foci of transitional epithelium from the tuboperitoneal junction [4]. Type I ovarian cancers are considered low grade, except for clear cell carcinomas, and account for only a small fraction of ovarian cancer deaths [5].

Type II epithelial cancers are high grade and characterized by involvement of both ovaries, aggressive behavior, late stage at diagnosis, and low survival. They are thought to originate as fallopian tube fimbriae carcinomas that spread to the ovaries and/or peritoneum. Women with these cancers often present with extensive extraovarian disease and ascites. Type II cancers are primarily high- grade serous carcinomas, the most common epithelial subtype, but also include carcinosarcomas and undifferentiated carcinomas. It is notable that although tumor grade is important in clinical practice, it is not a robust independent prognostic indicator [5].

Studies on miRNA in formalin fixed placenta tissues are limited [2] especially in Malaysia thus this study aimed at detecting the presence of mir-210 in formalin fixed normal placenta tissues in order to provide essential information for

analysis of the variation in miRNA levels in archived placentas.

Nonepithelial cancers are typically less aggressive than epithelial malignancies. Germ cell and sex cord-stromal tumors make up the majority of nonepithelial cancers but account for only 3% and 2%, respectively, of all ovarian cancers. Sex cord-stromal tumors arise from various connective tissue cell types, including granulosa, Sertoli, and/or Leydig cells. Other nonepithelial ovarian cancers include small cell carcinoma (hypercalcemic and non-hypercalcemic types) and ovarian sarcoma [4].

Transvaginal ultrasound (TVUS) is the most practical modality for assessment because it is widely available, well accepted by patients, non-invasive, low cost and not use ionizing radiation [6]. The pooled sensibility and specificity were respectively 77.0% and 83.0% for TVUS [6].

DWI is a newly developed magnetic resonance functional imaging technique based on water molecules movement rather than structure. Malignant tumors are composed of randomly organized tumor cells and the free movement of water molecules inside malignant dense mass is hindered. The inhibited diffusion of water is attributed to hypercellularity, thus DWI could provide unique information of tissue structure by tissue cellularity evaluation [7]. [7] The pooled SEN and SPE were 0.86 (95% confidence interval [CI], 0.83–0.89) and 0.81 (95%CI, 0.77–0.84), respectively.

2. MATERIALS AND METHODS

2.1 Protocol and registration

Overview of systematic reviews of diagnostic test accuracy will be conducted. The systematic review will be performed according to Preferred Reporting Items for Systematic reviews and Meta-Analyses: the PRISMA statement [8] http://www.primastatement.org/statement.htm. The search was conducted in Google Scholar, Dimension and Wiley Online Library databases that published in January 2011 until April 2020. All methods for inclusion/ exclusion criteria, data extraction and quality assessment will be specified in advance. The protocol will be not registered.

2.2 Data sources and search

Studies were identified by searching electronic databases and the limit were applied for language which is only English. This search was applied to Google scholar (2011- Present), Dimensions (2011- Present) and Wiley Online Library (2011-Present). The search strategy will include database subject headings and text words as follows: “ovarian cancer”, “ultrasound”, “transvaginal ultrasonography”, “TVUS” or “TVS”, “diffused weighted image” and “magnetic resonance imaging”. All the identified studies will be retrieved, and their references will also be checked for other relevant publications.

2.3 Study selection and data collection

Two assessors screened all abstracts and checked relevant full texts independently. Studies were enrolled in the meta-analysis if they satisfied the eligibility criteria, including types of studies, types of participants, types of interventions, types of outcomes, language and year publication. The studies will first be screened by their titles and abstracts to exclude evidently irrelevant article. After that, the full text retrieved for further clarification.

In order to avoid inclusion of duplicate cohorts in the meta-analysis in the case of 2 studies from the same authors, the study period of each study was examined; if the data overlapped, we chose the latest study according to the publication date, considering that patients from the first study was also included in the latest one. We used “snowball” strategy to identify potential interesting papers by reading reference list of those papers selected for full text reading. No attempts were made to contact the authors.

- Type of studies: Prospective or retrospective articles that will be chosen should have evaluated the diagnostic accuracy of TVUS or DWI alone or combination with other techniques in diagnosing ovarian cancer, regardless of the applied protocol of acquisition.
- Type of participants: Participant will be adult women undergoing histopathology of biopsy or surgery and no limitations to nationality. Patients with any stage of the disease will be included.
- Type of intervention: TVUS and DWI will be regarded as index tests because these tests are usually used to detect ovarian cancer.
- Types of outcome: The primary outcomes will be sensitivity (SEN), specificity (SPE), positive predictive value, negative predictive value, area under the curve, and their respective 95% confidence intervals.
- Language and year publication: Articles were published in English in January 2011 until present.

The Patients, Intervention, Comparator, Outcomes, Study Design (PICOS) criteria were used for describing the studies include. The diagnostic performance results as well as the supplementary useful information on procedures and patients were retrieved from selected main studies independently by the same reviewers. Disagreements were resolved peacefully by discussion between reviewers. The study period of each study was screened to avoid inclusion of duplicate cohorts in the meta-analysis in the case of two studies from the same authors. The latest study according to the publication date was taken if dates overlapped. No attempts were made to contact the authors for further information regarding the studies.

2.4 Risk of bias in individual studies

Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) is a tool that used to assess the quality of the studies and this tool had been adapted in this study. There are four domains in the QUADAS-2 format: 1) patient selection,

2) index test, 3) reference standard, 4) flow and timing. The possibility of bias and concerns about applicability were evaluated in each domain (the latter not applying to the 4th domain) and rated as low, high, or unclear risk. The outcomes of quality assessment were used for descriptive purposes to explore potential sources of heterogeneity and to provide an assessment of the overall quality of the included studies.

The quality of methodology was evaluated independently by the two reviewers, using a standard form with quality assessment criteria and a flow diagram; discussion had been done to resolve the disagreements.

2.5 Statistical analysis

We extracted or derived information on diagnostic performance of TVS and MRI. A random effects model was used to determine overall pooled sensitivity, specificity, positive likelihood ratio (LR+) and negative likelihood ratio (LR−). Positive and negative likelihood ratios (LRs) were used to characterize the clinical utility of a test and to estimate the post-test probability of disease. With TP, TN, FP, FN from extracted 2×2 contingency tables, we quantified the pooled SEN, SPE, LR, and DOR with 95% confidence intervals (95%CI) to evaluate DWI diagnosis accuracy for ovarian cancer. P value based on the likelihood ratio test were provided (α=0.05, two-sided). LR+ and LR- were applied to characterize the clinical utility of test and to assess the post-test probability of disease. A LR of 0.2–5.0 provides weak evidence for either ruling out or confirming the disease. A LR of 5.0–10.0 and 0.1–0.2 provides moderate evidence to either confirm or rule out the disease. A LR >10 or <0.1 provides strong evidence to either confirm or rules out the disease.

Heterogeneity for sensitivity and specificity were explored graphically thru constructing forest plots of sensitivity and specificity of each primary study. Then, they were plotted in the hierarchical summary receiver operating characteristic (HSROC) space, the latter to identify whether any heterogeneity could be attributable to an implicit threshold effect. Besides, HSROC curves for each technique were plotted to illustrate relationship between sensitivity and specificity. Means of a test on the Q statistic and I2 index were calculated to assess the presence of heterogeneity. A p-value <0.1 points to heterogeneity. The I2 index was measured to define the percentage of overall variation across studies that are due to heterogeneity rather than chance. The I2 value of 25%, 50% and 75% would be considered to specify low, moderate, and high heterogeneity, respectively.

All analysis was performed using Meta-analytical Integration of Diagnostic Accuracy Studies (MIDAS) and (METANDI) commands in STATA version 13.0.

3. RESULTS AND DISCUSSION

3.1 Search result

The search of Google Scholar, Dimension and Wiley Online Library databases provided a total 355 citations. After removal of duplicate records, 307 citations were remained. From these, 296 were discarded because it was clear from the title and abstract that they did not meet the criteria. The remaining 11 papers were examined. Finally, 5 [9][10][11][12][13] studies were excluded because these articles did not meet inclusion criteria and the remaining 6 [14][15][16][17][18][19] were included in the review and meta-analysis. There were no additional related studies were found from references cited in the articles included in the review.

3.2 Characteristics of included studies

A total of 6 studies published between January 2012 to April 2020 reporting on 1627 patients were included in the final analyses. Among these 1627 women, 974 had ovarian cancer. Two of the studies were retrospective studies and the remaining were prospective studies. Three studies used TVUS to assess the ovarian cancer and another two studies used DWI. Only one study assesses ovarian cancer using both TVUS and DWI on the same patients. Figure 1 below is a flowchart summarizing literature identification and selection.

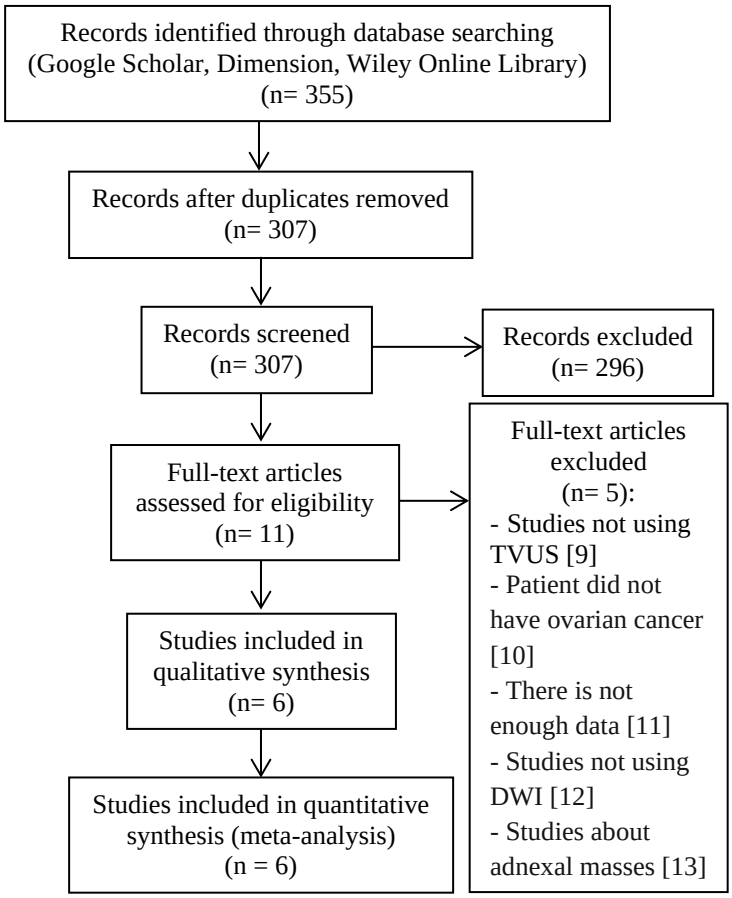

Figure 1 Flow chart showing studies selection

Table 1 Characteristic of included studies in this systematic review according to PICOS criteria

Study	Study design	No. of patient	Age (years)		Observers TVUS	Observers DWI
			Mean	Min-max		
(Semelka et al., 2014)	Pro-	100	44	16-65	NA	Multiple
(Zikan et al., 2016)	Pro-	191	59	39-79	Multiple	NA
(Fischerova et al., 2017)	Pro-	393	59	40-76	Multiple	NA
(Testa et al., 2012)	Pro-	115	59	31-85	Multiple	NA
(Mohammed et al., 2020)	Retro-	44	46	21-78	NA	Single
(Li et al., 2012)	Retro-	131	NA	NA	NA	Single

Pro-, Prospective; Retro-, Retrospective

Table 2 The technical aspects of MR protocols

Study	Tesla	Mark b value	Thickness (mm)	FOV (mm)	Matrix	T2-fused
(Semelka et al., 2014)	1.5	1000	-	-	-	Yes
(Li et al., 2012)	1.5	1000	6	320-420	256 x 256	Yes
(Mohammed et al., 2020)	1.5	1000	-	-	-	Yes

Table 1 shows PICOS features of studies included. All studies were based on the radiologist's or physician's impression. The technical aspects of MR protocols are described in Table 2. All studies used 1.5T and were used surface coil. For TVUS, all studies followed the standard protocol.

3.3 Methodological quality of included studies

Regarding risk of bias and the domain patient selection, one study was considered high risk and the other five studies were considered low. Concerning the domain index test regarding DWI, three studies adequately describe the method of index test as well as how it was performed and interpreted. The other three studies not applicable because it is for TVUS. Regarding TVUS, four studies adequately described the method of index test as well as how it was performed and interpreted. For the domain reference standard, all studies were likely interpreted the reference standard results with knowledge of the results of the index test. Regarding the domain flow and timing, the time elapsed between the index test and reference standard were unclear in three studies.

Regarding applicability, for the domain patient selection, five studies were deemed to include patients that matched the review question. For the domain index test, regarding DWI, all studies were considered as having low concerns for applicability as the index test was described well enough for study replication as was the reference standard domain. However, for TVUS, the domain index test was high.

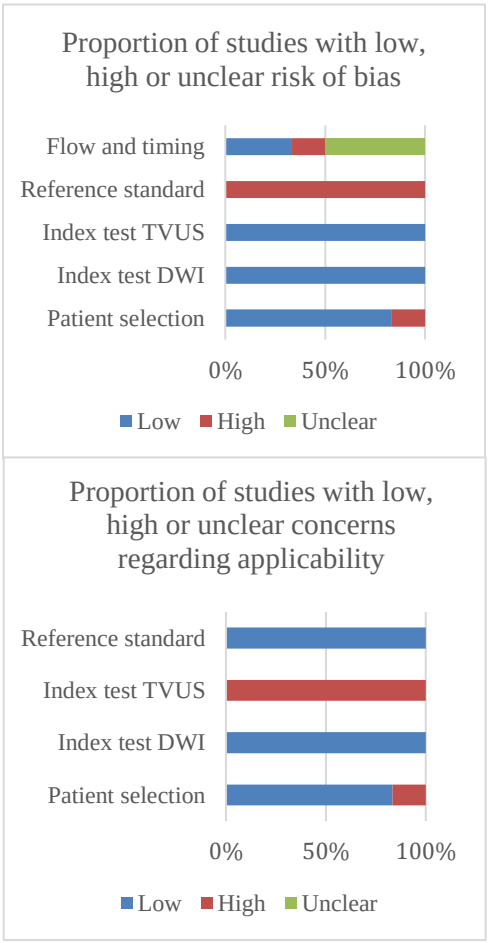

Figure 2 Histogram plot showing quality assessment (risk of bias and concerns about applicability) for all studies included in the meta-analysis

3.4 Diagnostic performance of TVUS and DWI in detecting ovarian cancer

Individual results of TVUS are shown on the forest plots in Figure 3. Overall, pooled sensitivity, specificity, LR+, and LR− of TVUS for detecting ovarian cancer were 75% (95% confidence interval [CI]=57%–87%), 92% (95% CI=80%–97%), 8.97 (95% CI=3.21–25.08), and 0.28 (95% CI=0.14–0.52), respectively. High heterogeneity was established for sensitivity (I^2=92.28%; Cochran Q=38.85; p=0.00). Furthermore, high heterogeneity was found for specificity (I^2=88.54%; Cochran Q=26.17; p=0.00).

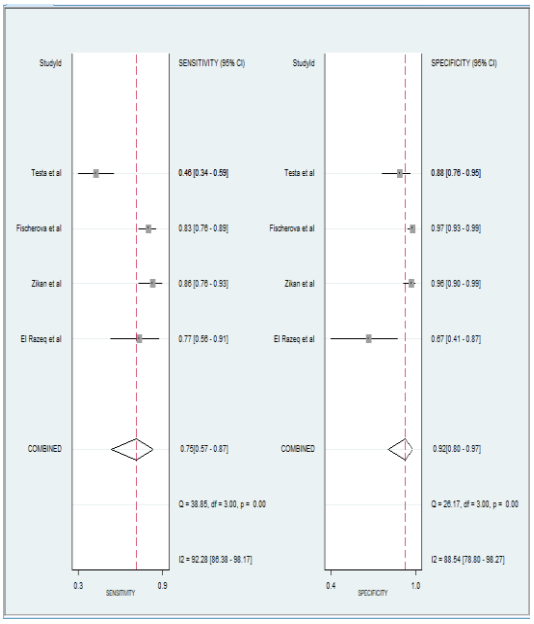

Figure 3 Forest plot for sensitivity and specificity for TVUS

On the other hand, pooled sensitivity, specificity, LR+, and LR− of DW-MRI for detecting ovarian cancer not applicable. This is because for STATA version 13.0. to be working, all the value for true-positive, false-positive, false-negative and true-negative must not have value 0.

HSROC curves are shown in Figure 4. The figure shows the summary point with a 95% prediction region and 95% confidence region. It can be observed that TVUS techniques has wider prediction contour than the confidence contour, respectively. The region of confidence is plotted from the CI around the summary point and shows that, the 'real value' would be estimated to be inside the region 95% of the time. The region of prediction around the summary point specifies the region where the results from a new research in the future are expected to lie. Furthermore, the prediction region is wider compared to the region of confidence because it goes more than the improbability in the presented data.

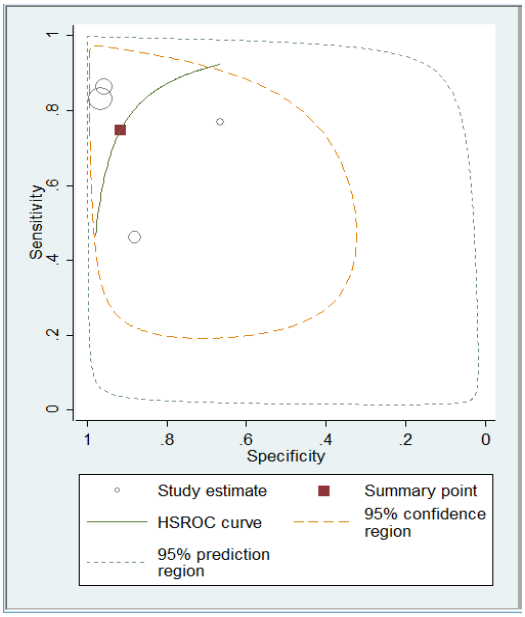

Figure 4 HSROC curve for TVUS

Fagan nomogram (Figure 5). show that a positive test for TVUS significantly increases the pretest probability ovarian cancer, from 12% to 55% in case of TVUS, while a negative test significantly decreases the pretest probability, from 12% to 4%. There is no apparent sign of publication bias in the funnel plots presented in Figure 6 and the Deeks adaptation for funnel plot asymmetry was significant for both TVUS (p=0.20).

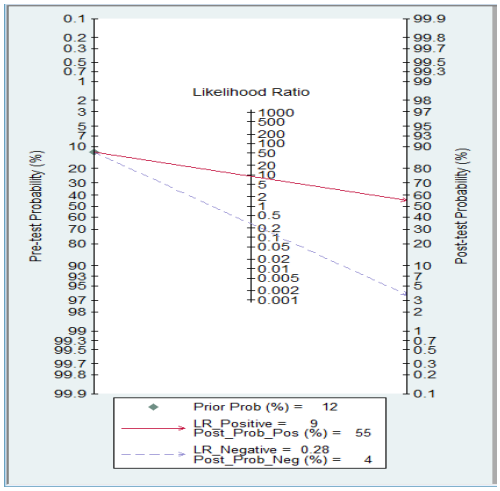

Figure 5 Fagan nomogram showing how pre-test probability change when the test is performed (post-test probability) depending on a positive or negative result for TVUS

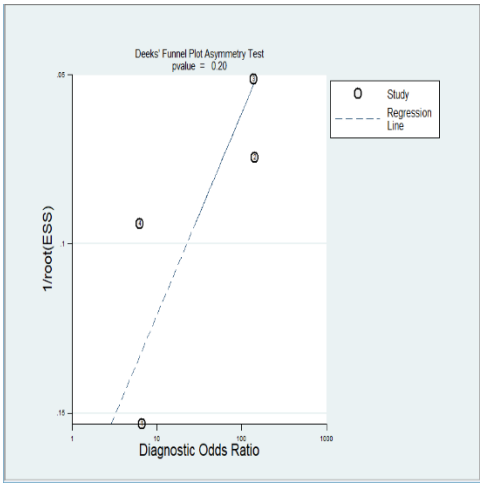

Figure 6 Funnel plot according to Deeks for graphical exploration of publication bias

Table 3 Pooled results of meta-analysis for TVUS and DWI

Analysis	Sensitivity (95% CI)	Specificity (95% CI)	LR+ (95% CI)	LR- (95% CI)	DOR (95% CI)
DWI	NA	NA	NA	NA	NA
TVUS	0.75 (0.57, 0.87)	0.92 (0.80, 0.97)	8.97 (3.21, 25.08)	0.28 (0.14, 0.52)	32.59 (6.90, 153.9)

4. CONCLUSION

Overall, the evidence is not sufficiently robust to determine the diagnostic accuracy between TVUS and DWI to detect the ovarian cancer. This is because for STATA version 13.0. to be working, all the value for true-positive, false-positive, false-negative and true-negative must not have value 0. Due to that, the diagnostic accuracy for DWI cannot be determine. Sensitivity, specificity, LR+, and LR− of TVUS for detecting ovarian cancer were 75% (95% confidence interval [CI]=57%–87%), 92% (95% CI=80%–97%), 8.97 (95% CI=3.21–25.08), and 0.28 (95% CI=0.14–0.52), respectively. High heterogeneity was established for sensitivity (I^2=92.28%; Cochran Q=38.85; p=0.00). Furthermore, high heterogeneity was found for specificity (I^2=88.54%; Cochran Q=26.17; p=0.00).

Notwithstanding, some limitations of the meta-analysis also should be acknowledged. First, only a small number of studies were included in the final meta-analysis because many studies were excluded based on eligibility criteria and may not be qualified to evaluate the diagnostic accuracy. All included studies were published in English which may have negated some of the gray literature.

In conclusion, TVUS has high specificity for detection ovarian cancer. TVUS should be considered as good enough for being used in clinical settings with limited resources.

REFERENCES

[1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA. Cancer J. Clin.*, vol. 68, no. 6, pp. 394–424, 2018.

[2] S. Coburn, F. Bray, M. Sherman, and B. Trabert, "International patterns and trends in ovarian cancer incidence, overall and by histologic subtype," *Int J Cancer*, vol. 140, no. 11, pp. 2451–2460, 2017.

[3] A. Coovadia and J. A. Eggert, "Future Perspectives in Cancer Screening and Early Detection," *Semin. Oncol. Nurs.*, vol. 33, no. 2, pp. 219–222, 2017.

[4] L. A. Torre *et al.*, "Ovarian cancer statistics, 2018," *CA. Cancer J. Clin.*, vol. 68, no. 4, pp. 284–296, 2018.

[5] R. J. Kurman and I. M. Shih, "The dualistic model of ovarian carcinogenesis revisited, revised, and expanded," *Am. J. Pathol.*, vol. 186, no. 4, pp. 733–747, 2016.

[6] G. M. Borrelli, L. A. de Mattos, M. de P. Andres, M. O. Gonçalves, R. M. Kho, and M. S. Abrão, "Role of Imaging Tools for the Diagnosis of Borderline Ovarian Tumors: A Systematic Review and Meta-Analysis," *J. Minim. Invasive Gynecol.*, vol. 24, no. 3, pp. 353–363, 2017.

[7] X. Yuan, L. Guo, W. Du, F. Mo, and M. Liu, "Diagnostic accuracy of DWI in patients with ovarian cancer," *Med. (United States)*, vol. 96, no. 19, pp. 1–7, 2017.

[8] D. Moher, A. Liberati, J. Tetzlaff, D. G. Altman, and T. P. Group, "Preferred Reporting Items for Systematic Reviews and Meta-Analyses : The PRISMA Statement," vol. 6, no. 7, 2009.

[9] Z. Liang, J. S. B, D. Zhang, and L. Gao, *Small Object Detection Using Deep*, vol. 1, no. September 2018. Springer International Publishing, 2018.

[10] J. Kaijser *et al.*, "Imaging techniques for the pre-surgical diagnosis of adnexal tumours," *Best Pract. Res. Clin. Obstet. Gynaecol.*, vol. 28, no. 5, pp. 683–695, 2014.

[11] P. Yadav, "Magnetic resonance imaging in the evaluation of female pelvis," *Med. J. Dr. D.Y. Patil Univ.*, vol. 9, no. 5, p. 627, 2016.

[12] S. M. Mansour, S. Saraya, and Y. El-Faissal, "Semi-quantitative contrast-enhanced MR analysis of indeterminate ovarian tumours: When to say malignancy?," *Br. J. Radiol.*, vol. 88, no. 1053, 2015.

[13] K. Shimada *et al.*, "Ultrasound-based logistic regression model LR2 versus magnetic resonance imaging for discriminating between benign and malignant adnexal masses: a prospective study," *Int. J. Clin. Oncol.*, vol. 23, no. 3, pp. 514–521, 2018.

[14] D. Fischerova *et al.*, "Ultrasound in preoperative assessment of pelvic and abdominal spread in patients with ovarian cancer: a prospective study," *Ultrasound Obstet. Gynecol.*, vol. 49, no. 2, pp. 263–274, 2017.

[15] W. Li, C. Chu, Y. Cui, P. Zhang, and M. Zhu, "Diffusion-weighted MRI : a useful technique to discriminate benign versus malignant ovarian surface epithelial tumors with solid and cystic components," no. October 2011, pp. 897–903, 2012.

[16] G. Mohammed, A. El, and M. A. M. Ahmed, "The Usefulness of Diffusion Weighted and Contrast Enhanced Magnetic Resonance Imaging in Characterization of Inconclusive Ovarian Mass," pp. 24–33, 2020.

[17] R. Semelka, K. Shawky, S. AbdelMoniem, and H. Tantawy, "Role of Mri and Diffusion Mri in Evaluation of Pelvic Masses of Gynaecological Origin," *Zagazig Univ. Med. J.*, vol. 20, no. 5, pp. 1–14, 2014.

[18] A. C. Testa *et al.*, "Ultrasound evaluation of intra-abdominal sites of disease to predict likelihood of suboptimal cytoreduction in advanced ovarian cancer: A prospective study," *Ultrasound Obstet. Gynecol.*, vol. 39, no. 1, pp. 99–105, 2012.

[19] M. Zikan *et al.*, "A prospective evaluation of ultrasound accuracy in the prediction of rectosigmoid infiltration in patients with epithelial ovarian cancer," p. 368, 2016.