Prospective comparison of one-step and two-step models for the classification of adnexal masses as benign or malignant

[1]A. Daemen, [2,3]C. Van Holsbeke, [1]O. Gevaert, [4]R. Fruscio, [5]S. Guerriero, [6]A. Czekierdowski, [7]L. Valentin, [8]L. Savelli, [9]A. Testa, [10]D. Fischerova, [2,11]T. Bourne, [2]I. Vergote, [1]B. De Moor, [2]D. Timmerman

[1]Department of Electrical Engineering ESAT/SCD, Katholieke Universiteit Leuven, Leuven, Belgium; [2]Department of Obstetrics and Gynecology, University Hospitals Leuven, campus Gasthuisberg, Katholieke Universiteit Leuven, Leuven, Belgium; [3]Ziekenhuis Oost-Limburg, Genk, Belgium; [4]San Gerardo Hospital, Monza, Italy; [5]Ospedale San Giovanni di Dio, Cagliari, Italy; [6]Medical University, Lublin, Poland; [7]University Hospital, Malmö, Sweden; [8]Reproductive Medicine Unit, Bologna, Italy; [9]Università Cattolica del Sacro Cuore, Rome, Italy; [10]First Medical Faculty of Charles University, Prague, Czech Republic; [11]St Georges Hospital Medical School, London, UK

Running title: One- vs. two-step models

Correspondence:

A. Daemen

Department of Electrical Engineering

Katholieke Universiteit Leuven

Kasteelpark Arenberg 10 bus 2446

3001 Leuven

Belgium

e-mail: anneleen.daemen@esat.kuleuven.be

Key words:

Abstract

Objective

In the first phase of a multicenter study by the International Ovarian Tumor Analysis (IOTA) group, 11 mathematical models were developed to distinguish between benign and malignant adnexal masses: 2 logistic regression models (LR1 and LR2), 3 Least-Square Support Vector Machines (LS-SVM), 3 Relevance Vector Machines (RVM), 2 Bayesian Multi-Layer Perceptrons (BMLP) and 1 Bayesian Perceptron model (BPER). In phase 2 of the IOTA study we investigated the optimal number of patients – i.e., those having uncertain classification by mathematical models or by an ultrasound expert – that might benefit from second stage testing.

Methods

The outcome of all models is a probability, increasing the uncertainty of diagnosis when positioned closer to the decision boundary. Each ultrasound expert classified masses as certainly or probably benign, uncertain, or probably or certainly malignant. The AUC, sensitivity, and specificity were calculated for each method after different percentages of patients had been removed either because the calculated probability of malignancy lay close to the decision boundary or because the ultrasound expert was uncertain about the diagnosis.

Results

19 centers participated in this study and contributed 1938 new cases. When 0% to 25% of the patients were removed because of uncertain diagnosis, a monotone increase in AUC was observed from 0.934 to maximum 0.974. According to the subjective assessment of experts, 5.9% of patients were classified as 'uncertain' and 33.9% as 'probably' benign or malignant. Although a higher AUC was achieved when patients were removed based on pattern

recognition by an expert, the models obtained a higher sensitivity and specificity at the optimal cut-off levels.

Conclusions

For the models to be applicable in all centers, independent of the experience of the ultrasound examiners, we suggest using a first stage model to decide which subgroup of patients requires a two-step model for classification. According to global performance and sensitivity, logistic regression model LR1 is superior to LS-SVM, RVM, and BMLP models, while the LS-SVM models have a higher specificity. Because second stage tests introduce an extra cost, the number of patients referred to second stage testing should be kept to a minimum. The optimal percentage of patients to refer to second line testing needs to be determined in combination with the performance of each individual second stage test.

Introduction

In the first phase of the study of the International Ovarian Tumor Analysis (IOTA) group, 1066 patients were gathered from 9 centers in 5 countries. Eleven mathematical models were developed as first stage tests to preoperatively distinguish benign from malignant adnexal masses based on ultrasound variables. The output of these models is a probability on malignancy, for which the cut-off to distinguish benign from malignant was determined on the training set of 746 patients (i.e. 70% of 1066). An overview of these models is given in Table 1. Table 2 shows the ultrasound variables included in each of the models.

In phase 2, 1938 new cases were gathered from 19 centers in 8 countries. This data set contains 542 malignant masses (28%) and 1396 benign masses (72%). Eight centers were already involved in phase 1 whereas 12 centers were new.

In IOTA phase 1, one-step models were considered to classify all masses as benign or malignant. When the outcome of the developed mathematical models however lies too close to the cut-off probability of the model for distinguishing benign from malignant, this model is not sufficient enough to classify a mass as benign or malignant. Also the ultrasound experts had to classify masses as certainly benign or malignant, probably benign or malignant, or uncertain. Hundred-fifteen masses (5.9%) were classified as uncertain and 657 masses (33.9%) as uncertain or probably benign or malignant. This subgroup of difficult masses which is based on uncertain classification by mathematical models or by ultrasound experts and for which no successful model has been constructed yet require two-step models. In such models, masses are referred to a reliable second stage test to help even experienced ultrasound experts. An example of a possible second stage test is proteomics with which each mass is characterized on molecular level. Such testing has the potential to be more reliable than the diagnoses performed by experts or by models based on ultrasound observations. A second

example can be research to combinations of new tumor markers that improve classification of difficult tumors contrary to the value of serum CA 125.

The goal of this study is to investigate the best criterion for referring patients to second stage testing to obtain good performing two-step models.

Methods

The mathematical models assign a probability between 0 and 1 to each mass for being malignant. Using the trained model cut-offs tabulated in Table 1, this continuous outcome can be transformed into a binary label 'benign' or 'malignant'. The uncertainty on this label increases when the outcome for a mass lies closer to the decision boundary (i.e. cut-off) of the model. For the logistic regression models for example, the optimal cut-off to distinguish benign from malignant was found to be a probability of 0.10. The closer the outcome of a mass to the decision boundary of 0.10, the higher the uncertainty on its label. Therefore, the rule for the order of patients to be removed from the data set will be based on the distance of the outcome to the decision boundary on a scale from 0 to 1. The decision boundaries of the models are asymmetrically located between 0.10 and 0.20, and the prevalence of benign is higher in this data set. This results in much more patients lying between 0 and the decision boundary x (0.10 to 0.20) than between x and 1. To account for this asymmetry, the steps along the probability axis with which patients are removed from the data set need to be $(1-x)/x$ times larger in the upper part of the axis above the decision boundary compared to the lower part. This results in an equal number of steps from x to 1 and x to 0. When the decision boundary is for example equal to 0.1, the upper part is 9 times larger than the lower part and therefore the steps should be 9 times larger. The AUC values (area under the ROC curve) are calculated for percentages of patients ranging from 0 to 50%, removed from the data set according to the uncertainty of the models.

The ultrasound experts on the other hand indicated their degree of confidence when classifying a mass. This lead to a subjective likelihood of diagnosis given by the experts, defined as certainly benign or malignant, probably benign or malignant, and uncertain. The AUC for all models is calculated when removing patients classified as 'uncertain' by the expert (5.9%) and when including only patients which were classified by the expert as 'certainly' benign or malignant (33.9%).

Results

In this study we focus on determining the best criterion for referring patients to second stage tests as a trade-off between performance and number of patients to be referred. Patients can be removed according to two criteria: based on the uncertainty of the model (i.e. the distance of the probability defined by the model to the decision boundary of the model) and by subjective assessment of the expert. We left out different percentages of patients from the data set according to these two criteria. The area under the ROC curves (AUC), the sensitivity, and specificity were calculated and compared for all the models on these reduced data sets. For the sensitivity and specificity, the optimized cut-off levels shown in Table 1 were used.

In Table 3 and Figure 1, the AUC for the models are shown against different percentages of patients referred to second stage tests based on the uncertainty of the models. Figure 1 shows a larger increase in AUC for small percentages of patients referred to second stage testing compared to percentages over 30%. Table 4 compares the AUC values at percentages 5.9% and 33.9% when referral of patients is based on subjective assessment of the experts with referral based on the models' uncertainty. When classifying for example all masses with the logistic regression model LR1, this model has an AUC of 0.951. When removing 115 masses (5.9%) for which experts were uncertain, the AUC increases to 0.959.

When removing the same amount of masses for which the model was most uncertain, the AUC increases to 0.957. When removing all masses classified by the experts as uncertain or as probably benign or malignant (33.9%) and applying LR1 on the remaining patients, the AUC increases to 0.981. When removing the same amount of masses based on the outcome of the model, the AUC increases to 0.978.

Tables 5, 6, and Figure 2 show the results for the sensitivity, while the results for the specificity are summarized in Tables 7, 8, and Figure 3. Both sensitivity and specificity increase when removing a rising number of patients, although the increase in specificity is larger. For all models the sensitivity is better when removing all masses for which the experienced ultrasound examiners were uncertain. However, when removing also the masses that were classified by the experts as probably benign or malignant, the sensitivity was better when the models' outcomes were used as criterion to remove 33.9% of masses, except for the models LR2, RVM 2 and BMLP 2. For the specificity, referring patients based on the uncertainty of the models outperformed subjective assessment for all models at both percentages 5.9 and 33.9.

The optimal number of patients to be referred to second stage testing depends on the performance of these second stage tests. Due to the current lack on knowledge about second stage performances, Figure 4 shows the performance of logistic regression model LR1 for possible, but currently unknown AUC values for the second stage tests, ranging from 0.84 to 1. The percentage of patients that need to be classified in first line increases when the performance of second line testing decreases. For second stage performances between 0.84 and 1, the optimal percentage of patients that should be referred to second stage testing can be determined to obtain globally the best performance.

Finally, we compared the masses for which the experts were uncertain with the ones referred to second stage testing by the models to reveal the difficulty of each pathology. Table

9 shows the number and percentage of benign, borderline, and malignant masses that were classified by the experts as uncertain as well as the characteristics of the 115 masses for which the models were most uncertain. Table 10 shows the results when considering the masses for which the experts were uncertain or probably certain, corresponding to 657 masses. Tables 11 and 12 show the overlap and discrepancy for the 115 and 657 masses, respectively, considered as difficult by the ultrasound examiners versus the models LR1, LS-SVM 1, RVM 1, and BMLP 1. For the 115 masses experienced as difficult by the ultrasound examiners, there is a large discrepancy in masses with the mathematical models (Table 11). For only 9.6 to 18% of these masses, the models are as uncertain as the experts. Table 13 shows more in detail the uncertainty of the models on the 115 masses for which the ultrasound experts were uncertain. Around half of these masses (43 to 63) appear in the 20% most difficult masses based on the models' uncertainty. For the 657 masses for which the ultrasound examiners were uncertain or probably certain, slightly more than half of the masses appeared in the top 33.9% of most difficult masses according to the models (Table 12). We also compared the 115 most difficult masses of LR1 with the ones of the other models (for simplification, only the comparison with LR2, LS-SVM 1, RVM 1, and BMLP 1 are shown in Table 14). From this table, it seems that rather different masses are considered as difficult by the models. Therefore, a combination of multiple models that encounter other subsets of masses as difficult can improve the global performance for distinguishing malignant from benign masses.

Discussion

We can conclude that the performance of the models increases when leaving out the most difficult tumors determined by the models or by subjective assessment of the ultrasound

experts. Classification of these tumors becomes possible with one of the proposed second stage tests.

For the majority of models, referral of patients based on subjective assessment lead to a larger AUC value at percentages 5.9 and 33.9. Also at the cut-off levels for which the models performed best in the training set of IOTA phase 1, the sensitivity was better at 5.9% using subjective assessment. However when removing patients based on the models' outcome, the sensitivity was better for the majority of the models at 33.9% and the specificity was better for all models at both percentages. Gynecologists with a high to excellent level of experience were operative in the centers involved in this study. When these models would be applied in centers with less experienced ultrasound examiners, the percentage of masses classified as uncertain would increase. Although we do not have data of less experienced centers, we believe that, at percentages above 5.9%, not only the specificity but also the sensitivity will be better when referring patients based on the uncertainty of the models. Only counting on a model has as extra advantage the ability to choose freely the number of patients to refer to a second stage test, for example the percentage of patients that would lead to the best performance combining classification by the first stage model and a second stage test. Therefore, we suggest using a first stage model to decide which subgroup of patients requires a two-step model for classification.

To know which masses are most difficult according to the ultrasound examiners and the models, we compared the masses referred to second stage testing by the ultrasound examiners with the ones referred by the models. Clinical interpretation tables 9 to 14?

In the next phase of the IOTA study we will select up to 20% of patients for second stage testing. Not more patients will be referred because this will never become the standard in hospitals due to the extra costs associated with these tests. The second stage tests that will be investigated are intravenous contrast agents, proteomics, new tumor markers, 3D color

Doppler and new models for difficult tumors. We will determine the global performance of the first stage models together with the second stage tests to finally end up with the second stage tests that provide reliable classifications in cases where the presently developed first stage models result in uncertain diagnoses as well as with an uncertainty level for the first stage models above which patients should be referred to these second stage tests.

Literature

1. Timmerman D, Testa A, Bourne T, Ferrazzi E, Ameye L, Konstantinovic M, Van Calster B, Collins W, Vergote I, Van Huffel S, Valentin L. Logistic regression model to distinguish between the benign and malignant adnexal mass before surgery: a multicenter study by the international ovarian tumor analysis group. Journal of Clinical Oncology 2005;23(34):8794-8801.

2. Van Calster B, Timmerman D, Nabney I, Valentin L, Van Holsbeke C, Van Huffel S. Classifying ovarian tumors using Bayesian Multi-Layer Perceptrons and Automatic Relevance Determination: a multi-center study. Proc. of the 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC2006), New York City, USA, 2006;5342-45.

3. Van Calster B, Timmerman D, Lu C, Suykens J, Valentin L, Van Holsbeke C, Amant F, Vergote I, Van Huffel S. Preoperative diagnosis of ovarian tumors using Bayesian kernel-based methods. Ultrasound Obstet Gynecol 2007;29:496-504.

Table 1: Eleven mathematical models developed in IOTA phase 1

Model	Description	Number of variables	Cut-off probability	Publication
LR1	Logistic regression model	12	0.10	Timmerman *et al.*, 2005
LR2	Logistic regression model	6	0.10	Timmerman *et al.*, 2005
LS-SVM 1	Bayesian LS-SVM with linear kernel	12	0.15	Van Calster *et al.*, 2007
LS-SVM 2	Bayesian LS-SVM with RBF kernel	12	0.12	Van Calster *et al.*, 2007
LS-SVM 3	Bayesian LS-SVM with additive RBF kernel	12	0.12	Van Calster *et al.*, 2007
RVM 1	RVM with linear kernel	12	0.20	Van Calster *et al.*, 2007
RVM 2	RVM with RBF kernel	12	0.15	Van Calster *et al.*, 2007
RVM 3	RVM with additive RBF kernel	12	0.15	Van Calster *et al.*, 2007
BMLP 1	Bayesian Multi-Layer Perceptron model; ARD with 10 hidden neurons	11	0.15	Van Calster *et al.*, 2006
BMLP 2	Bayesian Multi-Layer Perceptron model; ARD with 2 hidden neurons	11	0.15	Van Calster *et al.*, 2006
BPER	Bayesian Perceptron model	11	0.15	Van Calster *et al.*, 2006

LS-SVM, Least Squares Support Vector Machine; RBF, Radial Basis Function; RVM, Relevance Vector Machine; ARD, Automatic Relevance Determination

Table 2: Overview of the ultrasound variables included in the mathematical models

Variables	LR1	LR2	LS-SVM 1, 2 and 3	RVM 1, 2 and 3	BMLP 1	BMLP 2	BPER
Age (years)	X	X	X	X	X	X	X
Ascites (0-1)	X	X	X	X	X	X	X
Blood flow within papillary projection (0-1)	X	X	X	X	X	X	X
Max. diameter of solid component (mm)	X	X	X	X	X	X	X
Irregular internal cyst walls (0-1)	X	X	X	X	X	X	X
Acoustic shadows (0-1)	X	X	X	X		X	X
Personal history of ovarian cancer (0-1)	X		X	X		X	X
Hormonal therapy (0-1)	X		X	X	X		
Max. diameter of the lesion (mm)	X				X	X	X
Purely solid tumor (0-1)	X				X	X	X
Color score of intratumoral blood flow (1-4)	X				X		X
Pelvic pain during examination (0-1)	X					X	
Multilocular tumor + solid component (0-1)			X	X	X		
Max. diameter of the ovary (mm)			X	X			
Very strong intratumoral blood flow (0-1)			X	X			
Presumed ovarian origin of tumor (0-1)			X	X			
Unilocular tumor (0-1)						X	X
Number of papillary projections (0-4)					X		

Table 3: Performance (AUC) of the mathematical models against different percentages of patients referred to second stage tests based on uncertainty of the models

% patients removed	LR1	LR2	LS-SVM 1	LS-SVM 2	LS-SVM 3	RVM 1	RVM 2	RVM 3	BMLP 1	BMLP 2	BPER
0	0.951	0.934	0.943	0.946	0.935	0.943	0.944	0.937	0.945	0.946	0.950
2.5	0.953	0.936	0.946	0.949	0.939	0.946	0.947	0.940	0.948	0.949	0.953
5	0.956	0.939	0.949	0.952	0.942	0.949	0.950	0.944	0.952	0.951	0.956
7.5	0.959	0.941	0.952	0.955	0.945	0.952	0.953	0.947	0.955	0.955	0.958
10	0.961	0.944	0.954	0.958	0.948	0.955	0.956	0.950	0.958	0.958	0.961
12.5	0.963	0.947	0.956	0.960	0.951	0.957	0.958	0.953	0.961	0.960	0.964
15	0.966	0.949	0.958	0.962	0.954	0.959	0.961	0.955	0.964	0.963	0.966
17.5	0.968	0.951	0.960	0.964	0.956	0.962	0.963	0.957	0.966	0.965	0.968
20	0.970	0.953	0.962	0.966	0.958	0.964	0.966	0.959	0.969	0.968	0.970
22.5	0.972	0.955	0.963	0.968	0.961	0.966	0.968	0.961	0.971	0.970	0.972
25	0.974	0.956	0.965	0.969	0.963	0.968	0.970	0.963	0.973	0.972	0.973
27.5	0.975	0.958	0.967	0.971	0.966	0.970	0.972	0.965	0.974	0.974	0.974
30	0.976	0.958	0.968	0.972	0.967	0.971	0.973	0.967	0.976	0.975	0.975
32.5	0.977	0.959	0.969	0.973	0.969	0.973	0.974	0.968	0.978	0.977	0.975
35	0.978	0.960	0.970	0.974	0.971	0.974	0.975	0.970	0.979	0.978	0.975
37.5	0.979	0.960	0.971	0.975	0.973	0.975	0.975	0.972	0.980	0.979	0.975
40	0.980	0.960	0.971	0.975	0.975	0.976	0.976	0.974	0.980	0.980	0.975
42.5	0.980	0.960	0.972	0.976	0.976	0.976	0.976	0.976	0.981	0.981	0.975
45	0.980	0.959	0.973	0.977	0.977	0.977	0.977	0.976	0.981	0.982	0.975
47.5	0.981	0.960	0.973	0.977	0.979	0.977	0.978	0.977	0.982	0.983	0.974
50	0.981	0.961	0.974	0.978	0.980	0.978	0.977	0.978	0.982	0.984	0.973

Table 4: Comparison of the performance (AUC) of the mathematical models when leaving out masses based on the expert's or model's uncertainty

Model	Subj. assessment Uncertain	Model 5.9%	Subj. assessment Uncertain / Probably benign or malignant	Model 33.9%
LR1	0.959	0.957	0.981	0.978
LR2	0.945	0.940	0.971	0.959
LS-SVM 1	0.952	0.950	0.976	0.969
LS-SVM 2	0.955	0.953	0.978	0.974
LS-SVM 3	0.946	0.943	0.969	0.970
RVM 1	0.952	0.950	0.975	0.973
RVM 2	0.953	0.951	0.976	0.974
RVM 3	0.947	0.945	0.970	0.969
BMLP 1	0.954	0.953	0.974	0.978
BMLP 2	0.957	0.953	0.982	0.977
BPER	0.958	0.957	0.981	0.975

Table 5: Sensitivity of the mathematical models against different percentages of patients referred to second stage tests based on uncertainty of the models

% patients removed	LR1	LR2	LS-SVM 1	LS-SVM 2	LS-SVM 3	RVM 1	RVM 2	RVM 3	BMLP 1	BMLP 2	BPER
0	92.44	90.41	86.90	88.75	85.79	89.30	89.67	86.90	87.82	91.14	90.22
2.5	92.32	90.40	87.19	88.78	85.80	89.33	89.64	86.93	87.83	91.34	90.57
5	92.50	90.54	87.35	88.63	86.10	89.22	90.33	86.74	87.60	91.19	90.58
7.5	92.53	90.35	88.04	88.46	86.00	89.61	90.42	86.80	88.20	91.88	90.48
10	92.73	90.61	88.72	89.38	86.49	90.06	90.74	87.19	87.97	92.01	90.89
12.5	92.98	90.85	88.96	90.27	86.49	89.74	90.77	87.66	87.97	91.95	90.67
15	93.52	90.69	89.10	90.39	87.27	89.62	90.65	88.41	88.30	92.39	90.93
17.5	93.85	90.57	90.17	90.50	87.98	89.91	90.68	88.28	89.49	92.08	91.50
20	94.29	91.00	90.21	90.30	88.38	89.95	90.46	88.72	89.77	92.31	91.09
22.5	94.24	90.84	90.13	90.70	88.68	91.00	91.03	88.95	90.44	92.01	91.94
25	94.97	91.41	90.63	91.40	89.13	90.88	92.23	89.07	91.59	92.10	92.20
27.5	95.10	91.62	91.09	93.20	89.80	91.57	92.54	89.60	91.79	91.57	91.61
30	95.13	92.26	92.35	93.77	91.92	91.69	93.57	90.88	92.28	92.59	92.31
32.5	95.47	92.15	92.70	94.30	91.98	92.79	93.77	91.46	92.86	94.10	92.36
35	95.57	92.38	92.93	94.48	92.88	92.69	93.29	91.33	93.17	94.04	92.80
37.5	96.62	92.28	93.31	95.24	93.00	94.70	93.48	91.90	93.31	93.58	91.78
40	96.76	92.53	94.46	95.02	93.75	94.78	93.82	92.81	93.97	93.72	92.00
42.5	96.93	92.28	94.96	95.19	94.95	94.96	94.09	93.31	94.47	94.20	91.26
45	97.11	92.53	95.22	95.45	95.52	95.02	94.22	93.17	95.00	94.29	91.88
47.5	96.98	92.11	95.44	96.09	95.72	95.91	94.66	93.10	95.11	94.30	92.14
50	96.83	91.78	95.24	95.94	97.13	95.61	94.59	93.46	95.86	94.89	91.87

Table 6: Comparison of the sensitivity of the mathematical models when leaving out masses based on the expert's or model's uncertainty

Model	Subj. assessment Uncertain	Model 5.9%	Subj. assessment Uncertain / Probably benign or malignant	Model 33.9%
LR1	92.48	92.46	93.87	95.64
LR2	90.89	90.52	93.87	92.21
LS-SVM 1	88.12	88.00	90.80	93.09
LS-SVM 2	89.31	88.49	91.72	94.27
LS-SVM 3	86.53	85.91	89.57	92.36
RVM 1	90.10	89.04	92.02	92.56
RVM 2	90.50	90.25	93.56	93.46
RVM 3	87.92	86.64	91.10	91.56
BMLP 1	88.71	88.06	90.49	93.01
BMLP 2	92.08	91.67	94.17	94.22
BPER	90.89	90.49	93.87	92.64

Table 7: Specificity of the mathematical models against different percentages of patients referred to second stage tests based on uncertainty of the models

% patients removed	LR1	LR2	LS-SVM 1	LS-SVM 2	LS-SVM 3	RVM 1	RVM 2	RVM 3	BMLP 1	BMLP 2	BPER
0	83.74	82.88	88.83	87.89	86.46	87.03	85.17	86.25	86.25	83.24	85.89
2.5	85.85	84.78	90.35	89.67	88.19	88.36	87.13	88.04	88.06	85.15	87.73
5	87.52	86.63	91.69	91.37	90.11	90.17	88.98	89.69	89.59	86.97	89.64
7.5	89.26	88.65	92.82	92.54	91.58	91.10	90.33	91.27	91.19	88.44	91.32
10	91.04	90.20	93.61	93.41	92.56	92.37	91.58	92.70	92.79	90.14	92.71
12.5	92.42	91.12	94.49	94.22	93.13	92.74	92.36	93.52	93.31	91.43	93.61
15	93.25	91.98	95.11	95.05	93.87	94.27	93.41	94.04	94.47	92.76	94.78
17.5	93.97	93.05	95.81	95.84	94.17	95.14	94.19	94.59	95.13	93.44	95.29
20	94.43	93.80	96.13	96.34	94.81	95.45	95.10	95.31	95.95	94.27	95.94
22.5	95.02	94.63	96.63	96.68	95.81	95.60	95.60	96.26	96.48	95.25	96.82
25	95.63	95.33	97.07	96.77	96.22	96.39	96.21	96.60	96.75	96.04	97.29
27.5	96.44	95.95	97.35	97.25	96.39	96.67	97.03	96.89	97.12	96.99	97.79
30	96.83	96.43	97.28	97.26	96.29	96.96	97.44	97.18	97.61	97.39	98.58
32.5	97.03	96.83	97.48	97.28	96.55	97.27	97.67	97.48	98.13	97.51	98.65
35	97.25	97.15	97.61	97.90	96.74	97.81	97.92	97.71	98.37	97.85	98.71
37.5	97.71	97.81	98.28	98.15	96.93	98.28	98.29	97.63	98.64	98.10	98.89
40	97.86	98.07	98.54	98.30	97.03	98.33	98.56	97.86	98.82	98.49	99.17
42.5	97.89	98.13	98.48	98.46	97.26	98.25	98.63	98.35	99.11	98.65	99.14
45	97.94	98.43	98.41	98.51	97.12	98.43	98.57	98.90	99.08	98.83	99.45
47.5	98.22	98.48	98.33	98.43	97.11	98.50	98.77	98.86	99.28	98.91	99.43
50	98.40	98.67	98.38	98.62	97.66	98.56	99.24	98.81	99.25	98.87	99.41

Table 8: Comparison of the specificity of the mathematical models when leaving out masses based on the expert's or model's uncertainty

Model	Subj. assessment Uncertain	Model 5.9%	Subj. assessment Uncertain / Probably benign or malignant	Model 33.9%
LR1	86.72	88.45	92.67	97.09
LR2	85.74	87.53	91.94	97.09
LS-SVM 1	90.90	92.22	94.66	97.55
LS-SVM 2	90.36	91.74	94.24	97.52
LS-SVM 3	88.85	90.48	92.57	96.80
RVM 1	89.45	90.54	93.40	97.53
RVM 2	87.71	89.78	92.04	97.64
RVM 3	88.62	90.27	93.19	97.54
BMLP 1	88.47	90.33	93.40	98.39
BMLP 2	86.19	87.61	92.15	97.77
BPER	88.54	90.22	94.14	98.63

Table 9: Characteristics of the 5.9% (115) masses diagnosed as uncertain by the ultrasound examiner or by the models

		Subjective assessment		LR1		LR2		LS-SVM 1		LS-SVM 2		LS-SVM 3	
	N	N	%	N	%	N	%	N	%	N	%	N	%
Benign	***1396***	***78***	***5.6***	***90***	***6.4***	***90***	***6.4***	***73***	***5.2***	***77***	***5.5***	***83***	***5.9***
Endometrioma	400	10	2.5	13	3.3	17	4.3	13	3.3	12	3.0	10	2.5
Teratoma	226	5	2.2	6	2.7	8	3.5	7	3.1	7	3.1	11	4.9
Simple cyst + parasalpingeal cyst	131	3	2.3	5	3.8	4	3.1	3	2.3	4	3.1	1	0.8
Functional cyst	77	1	1.3	5	6.5	5	6.5	4	5.2	3	3.9	6	7.8
Hydrosalpinx + Salpingitis	49	3	6.1	6	12.2	4	8.2	0	0.0	0	0.0	1	2.0
Peritoneal Pseudocyst	11	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
Abscess	24	2	8.3	3	12.5	1	4.2	6	25.0	4	16.7	2	8.3
Fibroma	81	10	12.3	9	11.1	9	11.1	12	14.8	12	14.8	8	9.9
Serous Cystadenoma	236	27	11.4	23	9.7	29	12.3	13	5.5	18	7.6	26	11.0
Mucinous Cystadenoma	138	14	10.1	19	13.8	10	7.2	12	8.7	14	10.1	16	11.6
Rare benign	18	3	16.7	1	5.6	3	16.7	2	11.1	2	11.1	2	11.1
Uterine Fibroid	5	0	0.0	0	0.0	0	0.0	1	20.0	1	20.0	0	0.0
Borderline	***111***	***20***	***18.0***	***13***	***11.7***	***9***	***8.1***	***18***	***16.2***	***18***	***16.2***	***10***	***9.0***
Borderline serous stage I	55	9	16.4	8	14.5	4	7.3	11	20.0	9	16.4	4	7.3
Borderline serous stage II	3	1	33.3	1	33.3	1	33.3	1	33.3	1	33.3	0	0.0
Borderline serous stage III	8	0	0.0	0	0.0	2	25.0	0	0.0	1	12.5	0	0.0
Borderline mucinous stage I	41	9	22.0	3	7.3	2	4.9	5	12.2	6	14.6	6	14.6
Borderline mucinous stage IV	1	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
Borderline Endometroid stage I	1	0	0.0	1	100	0	0.0	1	100	1	100	0	0.0
Rare Borderline	2	1	50.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
Primary invasive	***373***	***17***	***4.6***	***9***	***2.4***	***11***	***2.9***	***19***	***5.1***	***18***	***4.8***	***17***	***4.6***
Primary invasive stage I	70	3	4.3	4	5.7	4	5.7	3	4.3	3	4.3	3	4.3
Primary invasive stage II	30	1	3.3	2	6.7	2	6.7	3	10.0	3	10.0	2	6.7
Primary invasive stage III	202	7	3.5	2	1.0	3	1.5	9	4.5	8	4.0	9	4.5
Primary invasive stage IV	30	0	0.0	1	3.3	0	0.0	2	6.7	2	6.7	1	3.3
Rare primary Invasive	41	2	4.9	0	0.0	2	4.9	2	4.9	2	4.9	2	4.9
Metastatic	***58***	***4***	***6.9***	***3***	***5.2***	***5***	***8.6***	***5***	***8.6***	***2***	***3.4***	***5***	***8.6***

Table 9 (cont.): Characteristics of the 5.9% (115) masses diagnosed as uncertain by the ultrasound examiner or by the models

		RVM 1		RVM 2		RVM 3		BMLP 1		BMLP 2		BPER	
	N	N	%	N	%	N	%	N	%	N	%	N	%
Benign	***1396***	***75***	***5.4***	***85***	***6.1***	***82***	***5.9***	***84***	***6.0***	***89***	***6.4***	***88***	***6.3***
Endometrioma	400	12	3.0	12	3.0	11	2.8	8	2.0	8	2.0	13	3.3
Teratoma	226	12	5.3	7	3.1	11	4.9	10	4.4	6	2.7	3	1.3
Simple cyst + parasalpingeal cyst	131	3	2.3	7	5.3	2	1.5	2	1.5	3	2.3	5	3.8
Functional cyst	77	3	3.9	2	2.6	5	6.5	4	5.2	5	6.5	4	5.2
Hydrosalpinx + Salpingitis	49	1	2.0	0	0.0	2	4.1	1	2.0	2	4.1	6	12.2
Peritoneal Pseudocyst	11	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
Abscess	24	1	4.2	3	12.5	2	8.3	4	16.7	3	12.5	3	12.5
Fibroma	81	12	14.8	10	12.3	6	7.4	14	17.3	8	9.9	14	17.3
Serous Cystadenoma	236	22	9.3	27	11.4	26	11.0	20	8.5	34	14.4	23	9.7
Mucinous Cystadenoma	138	7	5.1	13	9.4	16	11.6	17	12.3	20	14.5	16	11.6
Rare benign	18	1	5.6	3	16.7	1	5.6	3	16.7	0	0.0	1	5.6
Uterine Fibroid	5	1	20.0	1	20.0	0	0.0	1	20.0	0	0.0	0	0.0
Borderline	***111***	***15***	***13.5***	***10***	***9.0***	***10***	***9.0***	***16***	***14.4***	***12***	***10.8***	***6***	***5.4***
Borderline serous stage I	55	9	16.4	6	10.9	4	7.3	7	12.7	6	10.9	4	7.3
Borderline serous stage II	3	1	33.3	1	33.3	0	0.0	0	0.0	1	33.3	0	0.0
Borderline serous stage III	8	0	0.0	1	12.5	0	0.0	1	12.5	0	0.0	0	0.0
Borderline mucinous stage I	41	5	12.2	2	4.9	6	14.6	6	14.6	4	9.8	2	4.9
Borderline mucinous stage IV	1	0	0.0	0	0.0	0	0.0	1	100	0	0.0	0	0.0
Borderline Endometroid stage I	1	0	0.0	0	0.0	0	0.0	1	100	1	100	0	0.0
Rare Borderline	2	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
Primary invasive	***373***	***19***	***5.1***	***17***	***4.6***	***19***	***5.1***	***12***	***3.2***	***13***	***3.5***	***17***	***4.6***
Primary invasive stage I	70	3	4.3	2	2.9	3	4.3	4	5.7	5	7.1	9	12.9
Primary invasive stage II	30	3	10.0	2	6.7	2	6.7	0	0.0	0	0.0	0	0.0
Primary invasive stage III	202	7	3.5	8	4.0	10	5.0	6	3.0	7	3.5	6	3.0
Primary invasive stage IV	30	2	6.7	1	3.3	1	3.3	0	0.0	1	3.3	0	0.0
Rare primary Invasive	41	4	9.8	4	9.8	3	7.3	2	4.9	0	0.0	2	4.9
Metastatic	***58***	***6***	***10.3***	***3***	***5.2***	***4***	***6.9***	***3***	***5.2***	***1***	***1.7***	***4***	***6.9***

Table 10: Characteristics of the 33.9% (657) masses classified by the ultrasound examiner as uncertain or probably certain or diagnosed as most uncertain by the models

		Subjective assessment		LR1		LR2		LS-SVM 1		LS-SVM 2		LS-SVM 3	
	N	N	%	N	%	N	%	N	%	N	%	N	%
Benign	***1396***	441	31.6	436	31.2	436	31.2	419	30.0	428	30.7	429	30.7
Endometrioma	400	84	21.0	70	17.5	69	17.3	63	15.8	65	16.3	72	18.0
Teratoma	226	53	23.5	30	13.3	31	13.7	41	18.1	42	18.6	47	20.8
Simple cyst + parasalpingeal cyst	131	21	16.0	30	22.9	44	33.6	31	23.7	33	25.2	27	20.6
Functional cyst	77	23	29.9	20	26.0	18	23.4	18	23.4	20	26.0	20	26.0
Hydrosalpinx + Salpingitis	49	13	26.5	19	38.8	15	30.6	9	18.4	8	16.3	8	16.3
Peritoneal Pseudocyst	11	3	27.3	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
Abscess	24	14	58.3	19	79.2	16	66.7	12	50.0	13	54.2	12	50.0
Fibroma	81	44	54.3	54	66.7	56	69.1	57	70.4	56	69.1	56	69.1
Serous Cystadenoma	236	101	42.8	117	49.6	125	53.0	120	50.8	120	50.8	106	44.9
Mucinous Cystadenoma	138	73	52.9	64	46.4	50	36.2	57	41.3	60	43.5	67	48.6
Rare benign	18	11	61.1	10	55.6	8	44.4	8	44.4	8	44.4	9	50.0
Uterine Fibroid	5	1	20.0	3	60.0	4	80.0	3	60.0	3	60.0	5	100.0
Borderline	***111***	75	67.6	66	59.5	59	53.2	66	59.5	64	57.7	59	53.2
Borderline serous stage I	55	37	67.3	33	60.0	34	61.8	36	65.5	31	56.4	28	50.9
Borderline serous stage II	3	1	33.3	3	100	2	66.7	3	100	3	100	1	33.3
Borderline serous stage III	8	3	37.5	4	50.0	4	50.0	4	50.0	4	50.0	4	50.0
Borderline mucinous stage I	41	31	75.6	24	58.5	16	39.0	21	51.2	24	58.5	23	56.1
Borderline mucinous stage IV	1	1	100	1	100	1	100	1	100	1	100	1	100
Borderline Endometroid stage I	1	1	100	1	100	1	100	1	100	1	100	1	100
Rare Borderline	2	1	50.0	0	0.0	1	50.0	0	0.0	0	0.0	1	50.0
Primary invasive	***373***	141	37.8	129	34.6	133	35.7	144	38.6	136	36.5	137	36.7
Primary invasive stage I	70	35	50.0	29	41.4	29	41.4	29	41.4	28	40.0	28	40.0
Primary invasive stage II	30	14	46.7	12	40.0	11	36.7	15	50.0	14	46.7	11	36.7
Primary invasive stage III	202	48	23.8	61	30.2	63	31.2	72	35.6	66	32.7	67	33.2
Primary invasive stage IV	30	5	16.7	8	26.7	8	26.7	7	23.3	7	23.3	8	26.7
Rare primary invasive	41	21	51.2	19	46.3	22	53.7	21	51.2	21	51.2	23	56.1
Metastatic	*58*	18	31.0	26	44.8	29	50	28	48.3	29	50.0	32	55.2

Table 10 (cont.): Characteristics of the 33.9% (657) masses classified by the ultrasound examiner as uncertain or probably certain or diagnosed as most uncertain by the models

		RVM 1		RVM 2		RVM 3		BMLP 1		BMLP 2		BPER	
	N	N	%	N	%	N	%	N	%	N	%	N	%
Benign	***1396***	***424***	***30.4***	***421***	***30.2***	***422***	***30.2***	***401***	***28.7***	***409***	***29.3***	***373***	***26.7***
Endometrioma	400	64	16.0	63	15.8	71	17.8	47	11.8	54	13.5	61	15.3
Teratoma	226	39	17.3	39	17.3	40	17.7	47	20.8	35	15.5	17	7.5
Simple cyst + parasalpingeal cyst	131	32	24.4	36	27.5	31	23.7	26	19.8	26	19.8	23	17.6
Functional cyst	77	19	24.7	18	23.4	20	26.0	12	15.6	15	19.5	14	18.2
Hydrosalpinx + salpingitis	49	8	16.3	7	14.3	8	16.3	13	26.5	10	20.4	17	34.7
Peritoneal pseudocyst	11	0	0.0	0	0.0	0	0.0	0	0.0	2	18.2	0	0.0
Abscess	24	12	50.0	12	50.0	13	54.2	17	70.8	11	45.8	16	66.7
Fibroma	81	56	69.1	56	69.1	56	69.1	68	84.0	61	75.3	53	65.4
Serous cystadenoma	236	123	52.1	119	50.4	112	47.5	98	41.5	116	49.2	99	41.9
Mucinous cystadenoma	138	58	42.0	60	43.5	60	43.5	61	44.2	68	49.3	61	44.2
Rare benign	18	9	50.0	8	44.4	7	38.9	9	50.0	6	33.3	8	44.4
Uterine Fibroid	5	4	80.0	3	60.0	4	80.0	3	60.0	5	100. 0	4	80.0
Borderline	***111***	***64***	***57.7***	***71***	***64.0***	***59***	***53.2***	***71***	***64.0***	***55***	***49.5***	***81***	***73.0***
Borderline serous stage I	55	36	65.5	39	70.9	32	58.2	37	67.3	23	41.8	43	78.2
Borderline serous stage II	3	3	100	3	100	2	66.7	2	66.7	3	100	3	100
Borderline serous stage III	8	4	50.0	4	50.0	4	50.0	3	37.5	4	50.0	5	62.5
Borderline mucinous stage I	41	19	46.3	23	56.1	19	46.3	27	65.9	23	56.1	27	65.9
Borderline mucinous stage IV	1	1	100	1	100	1	100	1	100	0	0.0	1	100
Borderline Endometroid stage I	1	1	100	1	100	1	100	1	100	1	100	1	100
Rare Borderline	2	0	0.0	0	0.0	0	0.0	0	0.0	1	50.0	1	50.0
Primary invasive	***373***	***141***	***37.8***	***137***	***36.7***	***146***	***39.1***	***159***	***42.6***	***162***	***43.4***	***171***	***45.8***
Primary invasive stage I	70	32	45.7	30	42.9	31	44.3	32	45.7	39	55.7	43	61.4
Primary invasive stage II	30	14	46.7	14	46.7	13	43.3	17	56.7	13	43.3	17	56.7
Primary invasive stage III	202	71	35.1	66	32.7	72	35.6	82	40.6	77	38.1	83	41.1
Primary invasive stage IV	30	7	23.3	8	26.7	8	26.7	9	30.0	11	36.7	9	30.0
Rare primary invasive	41	17	41.5	19	46.3	22	53.7	19	46.3	22	53.7	19	46.3
Metastatic	*58*	*28*	*48.3*	*28*	*48.3*	*30*	*51.7*	*26*	*44.8*	*31*	*53.4*	*32*	*55.2*

Table 11: Comparison of the 115 (5.9%) most difficult masses between the ultrasound examiners and the models LR1, LS-SVM 1, RVM 1, and BMLP 1

		LR 1			LS-SVM 1			RVM 1			BMLP 1		
	N total	N overlap*	N model£	N subj ass$	N overlap*	N model£	N subj ass$	N overlap*	N model£	N subj ass$	N overlap*	N model£	N subj ass$
Benign	***1396***	***18***	***72***	***60***	***12***	***61***	***66***	***13***	***62***	***65***	***14***	***70***	***64***
Endometrioma	400	1	12	9	1	12	9	2	10	8	2	6	8
Teratoma	226	1	5	4	0	7	5	1	11	4	0	10	5
Simple cyst + parasalpingeal cyst	131	1	4	2	2	1	1	2	1	1	0	2	3
Functional cyst	77	1	4	0	0	4	1	0	3	1	1	3	0
Hydrosalpinx + Salpingitis	49	2	4	1	0	0	3	1	0	2	0	1	3
Peritoneal Pseudocyst	11	0	0	0	0	0	0	0	0	0	0	0	0
Abscess	24	1	2	1	0	6	2	0	1	2	0	4	2
Fibroma	81	1	8	9	2	10	8	2	10	8	2	12	8
Serous Cystadenoma	236	7	16	20	4	9	23	3	19	24	5	15	22
Mucinous Cystadenoma	138	3	16	11	3	9	11	2	5	12	4	13	10
Rare benign	18	0	1	3	0	2	3	0	1	3	0	3	3
Uterine Fibroid	5	0	0	0	0	1	0	0	1	0	0	1	0
Borderline	***111***	***3***	***10***	***17***	***4***	***14***	***16***	***4***	***11***	***16***	***3***	***13***	***17***
Borderline serous stage I	55	3	5	6	2	9	7	2	7	7	1	6	8
Borderline serous stage II	3	0	1	1	1	0	0	1	0	0	0	0	1
Borderline serous stage III	8	0	0	0	0	0	0	0	0	0	0	1	0
Borderline mucinous stage I	41	0	3	9	1	4	8	1	4	8	2	4	7
Borderline mucinous stage IV	1	0	0	0	0	0	0	0	0	0	0	1	0
Borderline Endometroid stage I	1	0	1	0	0	1	0	0	0	0	0	1	0
Rare Borderline	2	0	0	1	0	0	1	0	0	1	0	0	1
Primary invasive	***373***	***0***	***9***	***13***	***0***	***19***	***13***	***1***	***18***	***12***	***0***	***12***	***13***
Primary invasive stage I	70	0	4	3	0	3	3	0	3	3	0	4	3
Primary invasive stage II	30	0	2	1	0	3	1	0	3	1	0	0	1
Primary invasive stage III	202	0	2	7	0	9	7	1	6	6	0	6	7
Primary invasive stage IV	30	0	1	0	0	2	0	0	2	0	0	0	0
Rare primary Invasive	41	0	0	2	0	2	2	0	4	2	0	2	2
Metastatic	***58***	***0***	***3***	***4***	***0***	***5***	***4***	***0***	***6***	***4***	***0***	***3***	***4***
TOTAL	***1938***	***21***	***94***	***94***	***16***	***99***	***99***	***18***	***97***	***97***	***17***	***98***	***98***

* number of overlapping masses diagnosed as uncertain by the ultrasound examiners and the mathematical models, [£] number of masses for which only the models are uncertain, [$] number of masses for which only the ultrasound examiners are uncertain

Table 12: Comparison of the 657 (33.9%) most difficult masses according to the ultrasound examiners and the models LR1, LS-SVM 1, RVM 1, and BMLP 1

		LR 1			LS-SVM 1			RVM 1			BMLP 1		
	N total	N overlap*	N model[f]	N subj ass[$]	N overlap*	N model[f]	N subj ass[$]	N overlap*	N model[f]	N subj ass[$]	N overlap*	N model[f]	N subj ass[$]
Benign	***1396***	***239***	***197***	***202***	***220***	***199***	***221***	***227***	***197***	***214***	***226***	***175***	***215***
Endometrioma	400	37	33	47	31	32	53	32	32	52	28	19	56
Teratoma	226	13	17	40	15	26	38	16	23	37	21	26	32
Simple cyst + parasalpingeal cyst	131	10	20	11	7	24	14	8	24	13	11	15	10
Functional cyst	77	14	6	9	11	7	12	11	8	12	9	3	14
Hydrosalpinx + Salpingitis	49	7	12	6	5	4	8	5	3	8	6	7	7
Peritoneal Pseudocyst	11	0	0	3	0	0	3	0	0	3	0	0	3
Abscess	24	11	8	3	8	4	6	8	4	6	12	5	2
Fibroma	81	31	23	13	35	22	9	34	22	10	36	32	8
Serous Cystadenoma	236	70	47	31	69	51	32	71	52	30	58	40	43
Mucinous Cystadenoma	138	40	24	33	34	23	39	35	23	38	38	23	35
Rare benign	18	6	4	5	5	3	6	6	3	5	6	3	5
Uterine Fibroid	5	0	3	1	0	3	1	1	3	0	1	2	0
Borderline	***111***	***46***	***20***	***29***	***45***	***21***	***30***	***44***	***20***	***31***	***51***	***20***	***24***
Borderline serous stage I	55	22	11	15	25	11	12	24	12	13	26	11	11
Borderline serous stage II	3	1	2	0	1	2	0	1	2	0	1	1	0
Borderline serous stage III	8	2	2	1	1	3	2	1	3	2	0	3	3
Borderline mucinous stage I	41	19	5	12	16	5	15	16	3	15	22	5	9
Borderline mucinous stage IV	1	1	0	0	1	0	0	1	0	0	1	0	0
Borderline Endometroid stage I	1	1	0	0	1	0	0	1	0	0	1	0	0
Rare Borderline	2	0	0	1	0	0	1	0	0	1	0	0	1
Primary invasive	***373***	***65***	***64***	***58***	***66***	***78***	***57***	***66***	***75***	***57***	***74***	***85***	***49***
Primary invasive stage I	70	21	8	14	20	9	15	22	10	13	23	9	12
Primary invasive stage II	30	9	3	5	11	4	3	11	3	3	12	5	2
Primary invasive stage III	202	20	41	28	20	52	28	19	52	29	26	56	22
Primary invasive stage IV	30	3	5	2	3	4	2	3	4	2	2	7	3
Rare primary Invasive	41	12	7	9	12	9	9	11	6	10	11	8	10
Metastatic	*58*	*8*	*18*	*10*	*11*	*17*	*7*	*11*	*17*	*7*	*9*	*17*	*9*
TOTAL	*1938*	*358*	*299*	*299*	*342*	*315*	*315*	*348*	*309*	*309*	*360*	*297*	*297*

* number of overlapping masses diagnosed as uncertain or probably certain by the ultrasound examiners and the mathematical models, £ number of masses for which only the models are uncertain, $ number of masses for which only the ultrasound examiners are uncertain or probably certain

Table 13: Fraction of 115 masses – classified as uncertain by ultrasound experts – that appear in the top x% of difficult masses according to the models

	LR1	LR2	LS-SVM 1	LS-SVM 2	LS-SVM 3	RVM 1	RVM 2	RVM 3	BMLP 1	BMLP 2	BPER
5.9%	21	12	16	18	11	18	19	12	17	11	21
10%	30	21	26	28	20	29	27	28	23	24	35
15%	46	35	48	44	35	45	40	39	37	38	48
20%	59	50	54	56	43	55	53	55	48	53	63
30%	76	68	68	68	69	68	70	69	69	70	86
40%	87	86	86	83	74	83	88	82	86	83	98
50%	95	94	98	97	83	94	99	94	96	94	104
60%	101	99	102	99	103	101	108	102	102	105	108
70%	104	105	103	104	104	109	110	105	107	111	110
80%	110	110	106	109	105	111	113	108	112	112	112
90%	113	112	111	111	110	114	114	113	113	113	112
100%	115	115	115	115	115	115	115	115	115	115	115

Table 13: Comparison of the 115 most difficult masses between LR1 and the models LR2, LS-SVM 1, RVM 1, and BMLP 1

		LR 2			LS-SVM 1			RVM 1			BMLP 1		
	N total	N overlap*	N model£	N LR1$	N overlap*	N model£	N LR1$	N overlap*	N model£	N LR1$	N overlap*	N model£	N LR1$
Benign	***1396***	***31***	***59***	***59***	***20***	***53***	***70***	***22***	***53***	***68***	***29***	***55***	***61***
Endometrioma	400	6	11	7	4	9	9	4	8	9	3	5	10
Teratoma	226	1	7	5	0	7	6	1	11	5	1	9	5
Simple cyst + parasalpingeal cyst	131	0	4	5	1	2	4	1	2	4	1	1	4
Functional cyst	77	2	3	3	0	4	5	0	3	5	1	3	4
Hydrosalpinx + Salpingitis	49	3	1	3	0	0	6	0	1	6	0	1	6
Peritoneal Pseudocyst	11	0	0	0	0	0	0	0	0	0	0	0	0
Abscess	24	1	0	2	1	5	2	1	0	2	1	3	2
Fibroma	81	4	5	5	2	10	7	4	8	5	2	12	7
Serous Cystadenoma	236	11	18	12	6	7	17	7	15	16	9	11	14
Mucinous Cystadenoma	138	2	8	17	6	6	13	4	3	15	10	7	9
Rare benign	18	1	2	0	0	2	1	0	1	1	1	2	0
Uterine Fibroid	5	0	0	0	0	1	0	0	1	0	0	1	0
Borderline	***111***	***2***	***7***	***11***	***4***	***14***	***9***	***3***	***12***	***10***	***6***	***10***	***7***
Borderline serous stage I	55	1	3	7	1	10	7	1	8	7	3	4	5
Borderline serous stage II	3	1	0	0	0	1	1	0	1	1	0	0	1
Borderline serous stage III	8	0	2	0	0	0	0	0	0	0	0	1	0
Borderline mucinous stage I	41	0	2	3	2	3	1	2	3	1	2	4	1
Borderline mucinous stage IV	1	0	0	0	0	0	0	0	0	0	0	1	0
Borderline Endometroid stage I	1	0	0	1	1	0	0	0	0	1	1	0	0
Rare Borderline	2	0	0	0	0	0	0	0	0	0	0	0	0
Primary invasive	***373***	***4***	***7***	***5***	***4***	***15***	***5***	***5***	***14***	***4***	***3***	***9***	***6***
Primary invasive stage I	70	3	1	1	2	1	2	2	1	2	2	2	2
Primary invasive stage II	30	1	1	1	1	2	1	2	1	0	0	0	2
Primary invasive stage III	202	0	3	2	1	8	1	1	6	1	1	5	1
Primary invasive stage IV	30	0	0	1	0	2	1	0	2	1	0	0	1
Rare primary Invasive	41	0	2	0	0	2	0	0	4	0	0	2	0
Metastatic	***58***	***2***	***3***	***1***	***0***	***5***	***3***	***1***	***5***	***2***	***0***	***3***	***3***
TOTAL	***1938***	***39***	***76***	***76***	***28***	***87***	***87***	***31***	***84***	***84***	***38***	***77***	***77***

* number of overlapping masses diagnosed as uncertain by LR1 and the other mathematical models, £ number of masses for which the other mathematical models are uncertain, $ number of masses for which only LR1 is uncertain

Figure 1: Performance (AUC) of the mathematical models against different percentages of patients referred to second stage tests based on uncertainty of the models

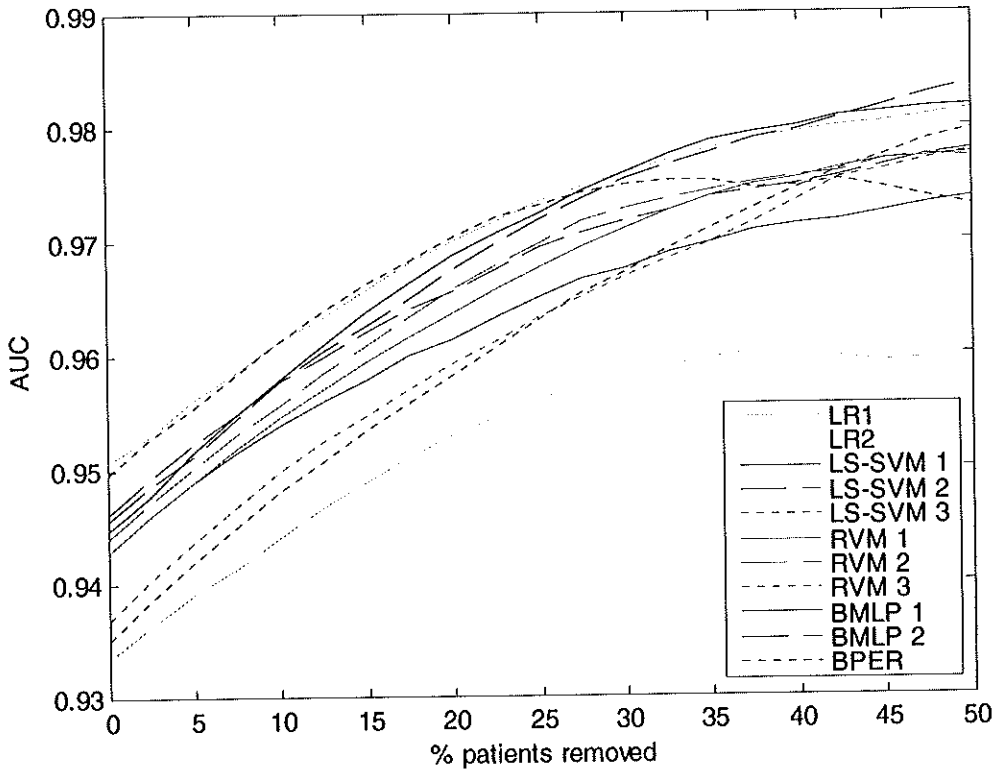

Figure 2: Sensitivity of the mathematical models against different percentages of patients referred to second stage tests based on uncertainty of the models

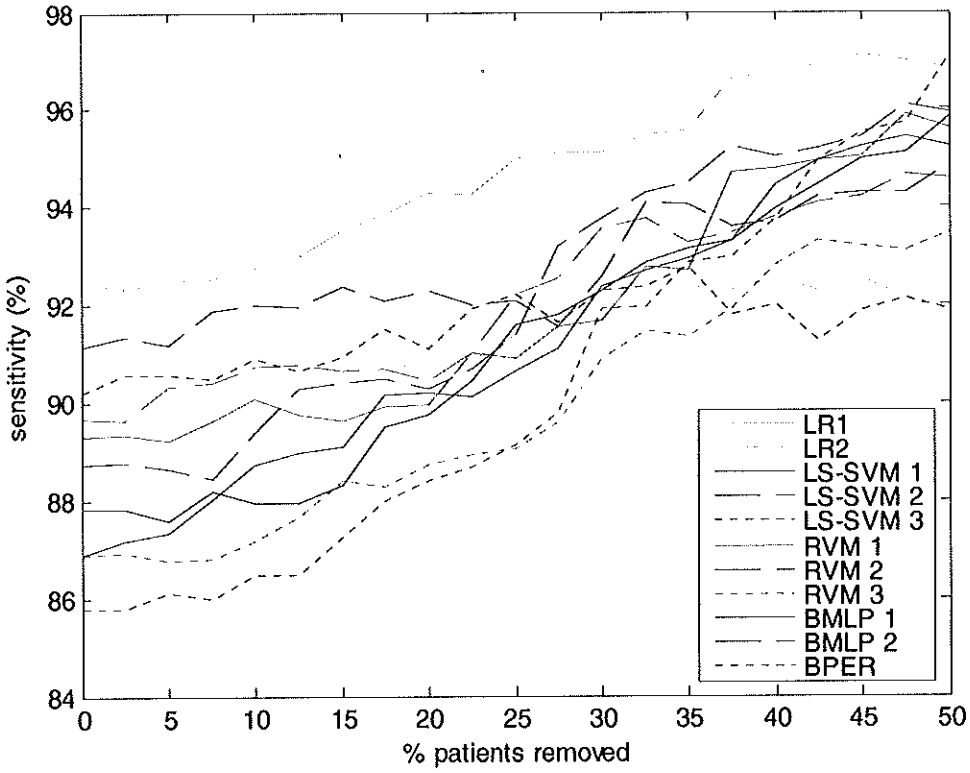

Figure 3: Specificity of the mathematical models against different percentages of patients referred to second stage tests based on uncertainty of the models

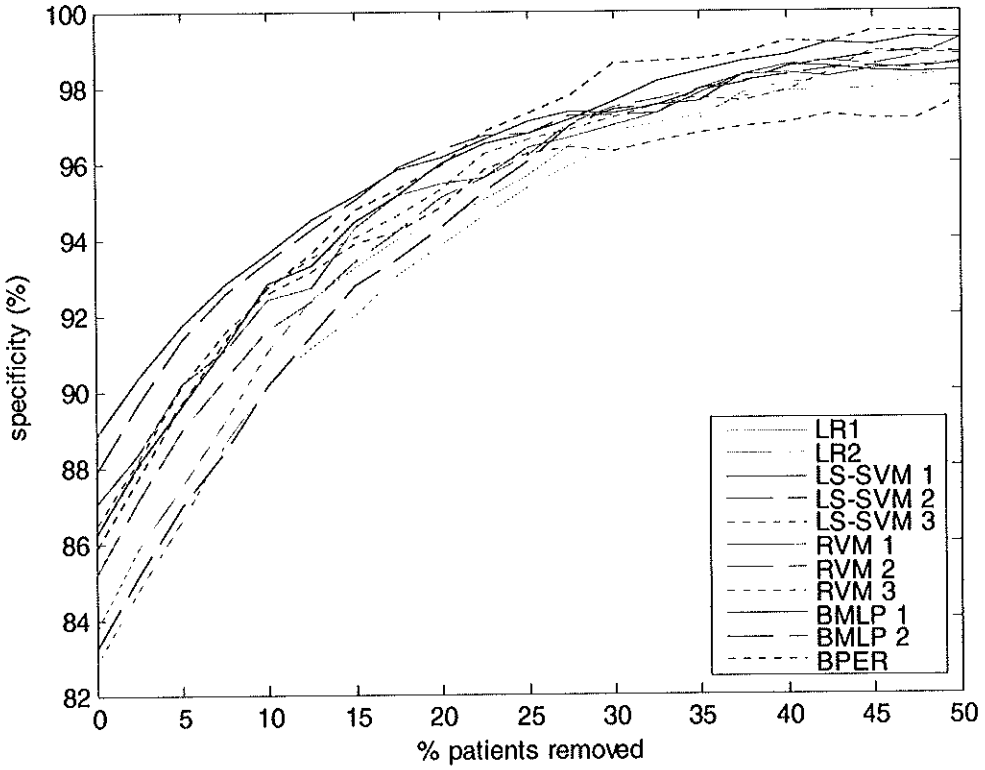

Figure 4: Performance (AUC) of LR1 against different percentages of patients referred to second stage tests based on the model's uncertainty, for possible but unknown performances (AUC) of second stage testing, ranging from 0.84 to 1