New models to predict depth of infiltration in endometrial carcinoma based on transvaginal sonography

F. DE SMET*†, J. DE BRABANTER*‡, T. VAN DEN BOSCH§, N. POCHET*, F. AMANT§,
C. VAN HOLSBEKE§, P. MOERMAN¶, B. DE MOOR*, I. VERGOTE§ and D. TIMMERMAN§

*Departments of *Electrical Engineering ESAT-SCD, §Obstetrics and Gynecology and ¶Pathology, Katholieke Universiteit Leuven, Leuven, †Medical Direction, Alliance of Christian Sickness Funds, Brussels and ‡Sint-Lieven Hogeschool, Departement Industrieel Ingenieur (Associatie K.U. Leuven), Gent, Belgium*

ABSTRACT

Objectives *Preoperative knowledge of the depth of myometrial infiltration is important in patients with endometrial carcinoma. This study aimed at assessing the value of histopathological parameters obtained from an endometrial biopsy (Pipelle® de Cornier; results available preoperatively) and ultrasound measurements obtained after transvaginal sonography with color Doppler imaging in the preoperative prediction of the depth of myometrial invasion, as determined by the final histopathological examination of the hysterectomy specimen (the gold standard).*

Methods *We first collected ultrasound and histopathological data from 97 consecutive women with endometrial carcinoma and divided them into two groups according to surgical stage (Stages Ia and Ib vs. Stages Ic and higher). The areas (AUC) under the receiver–operating characteristics curves of the subjective assessment of depth of invasion by an experienced gynecologist and of the individual ultrasound parameters were calculated. Subsequently, we used these variables to train a logistic regression model and least squares support vector machines (LS-SVM) with linear and RBF (radial basis function) kernels. Finally, these models were validated prospectively on data from 76 new patients in order to make a preoperative prediction of the depth of invasion.*

Results *Of all ultrasound parameters, the ratio of the endometrial and uterine volumes had the largest AUC (78%), while that of the subjective assessment was 79%. The AUCs of the blood flow indices were low (range, 51–64%). Stepwise logistic regression selected the degree of differentiation, the number of fibroids, the endometrial thickness and the volume of the tumor. Compared with the AUC of the subjective assessment (72%), prospective evaluation of the mathematical models resulted in a higher AUC for the LS-SVM model with an RBF kernel (77%), but this difference was not significant.*

Conclusions *Single morphological parameters do not improve the predictive power when compared with the subjective assessment of depth of myometrial invasion of endometrial cancer, and blood flow indices do not contribute to the prediction of stage. In this study an LS-SVM model with an RBF kernel gave the best prediction; while this might be more reliable than subjective assessment, confirmation by larger prospective studies is required. Copyright © 2006 ISUOG. Published by John Wiley & Sons, Ltd.*

INTRODUCTION

Carcinoma of the endometrium is the most common female pelvic malignancy[1]. Initial preoperative evaluation of patients suspected of having a carcinoma of the endometrium includes transvaginal sonography with or without color Doppler imaging and endometrial biopsy.

The distinction between FIGO surgical Stages Ib and Ic[2] endometrial carcinoma (assessed postoperatively) is determined by the degree of myometrial invasion (Stage Ib is less and Stage Ic is more than 50% invasion)[3]. This is an important prognostic factor[4] and in many institutions it determines the treatment protocol. The accurate preoperative distinction between patients with Stages Ia or Ib carcinoma and patients with Stages Ic or higher would allow identification of high-risk patients who might need

ORIGINAL PAPER

pelvic lymphadenectomy. The importance of this is that in many countries, patients who will need lymphadenectomy are referred to a gynecological oncologist, while patients not requiring lymphadenectomy are operated on by a general gynecologist or surgeon.

Several techniques are used to estimate the depth of myometrial invasion, but all have specific limitations. Intraoperative gross visual inspection or frozen section do not allow preoperative planning of the surgical procedure. Franchi *et al.*[5] reported an accuracy of 85.3% in predicting the degree of myometrial invasion in a series of 403 patients using intraoperative gross visual inspection, whereas Kucera *et al.*[6] reported an accuracy of 88% using frozen section in a combined set of 624 patients. Contrast-enhanced magnetic resonance imaging (MRI) is the most reliable method. In a meta-analysis, Kinkel *et al.*[7] reported an area (AUC) under the receiver–operating characteristics (ROC) curve of 91% with respect to the prediction of myometrial invasion. However, MRI is costly, has limited availability and is not appropriate for all patients (e.g. those with claustrophobia, obesity and contrast allergies). Different groups[8–18] have studied the value of transvaginal sonography and color Doppler imaging using different morphological or color Doppler parameters, with considerable variation in the results. Arko and Takac[19] published one of the largest series that investigated the use of transvaginal sonography to estimate the depth of myometrial invasion in 120 patients, reporting an accuracy of 73% in predicting myometrial invasion.

In our study on patients with endometrial carcinoma, we analyzed ultrasound measurements obtained from transvaginal sonography with color Doppler imaging and histopathological data, obtained from preoperative endometrial biopsy (Pipelle® de Cornier)). We then explored whether they contributed to the prediction of myometrial invasion as assessed postoperatively by the final histopathological examination (gold standard). Moreover, we aimed to construct models to predict the presence of deep myometrial invasion, which could help the clinician to identify preoperatively patients that might need more extensive surgery.

METHODS

We first collected data from 97 consecutive patients with endometrial carcinoma, who underwent sonography between September 1994 and February 2000 by a single operator (D.T.)[20]. Here we refer to these patients as the 'training set'. Their mean age was 65.9 (range, 45–83) years, with 88 women being postmenopausal. The distribution of the different surgical FIGO stages was as follows: 24 Stage Ia, 35 Stage Ib, 12 Stage Ic, eight Stage II, 13 Stage III and five Stage IV. The histopathological subtypes were: 76 endometrioid adenocarcinoma, three serous papillary and 18 mixed type (five of which had a clear cell and three a serous papillary component). Fifty-four tumors were differentiated highly, 18 moderately and 25 poorly. Tumors with a serous papillary or a clear cell component were considered to be poorly differentiated.

All patients gave informed consent and underwent a preoperative ultrasound examination with transvaginal sonography and color Doppler imaging in the department of Obstetrics and Gynecology (University Hospitals Leuven) using the same protocol. The uterus was assessed both in sagittal and coronal planes with an Acuson Sequoia (Siemens-Acuson Inc., Mountain View, CA, USA) ultrasound system, equipped with highly sensitive color Doppler imaging capability and a MultiHertz intravaginal probe with a field of view of 140°. The color Doppler imaging examination always included measurements of flow indices from both uterine arteries and subendometrial blood vessels. High-quality transparent color copies (Agfa Drystar, Agfa Gevaert, Mortsel, Belgium) and schematic hand-made drawings of the sonographic findings were obtained for every patient.

Histopathology was assessed preoperatively by endometrial biopsy using a Pipelle de Cornier, which has been shown to reflect accurately histopathological parameters[21–23]. The patients were divided into two groups as determined by the final histopathological examination of the hysterectomy specimen: those with surgical Stages Ia or Ib and those with surgical Stages Ic or higher.

Several morphological parameters visualized by grayscale transvaginal sonography are available for univariate analysis (endometrial (ET) and myometrial (MT) thickness; endometrial (EV) and uterine (UV) volume; ET/uterine anteroposterior diameter (AP); EV/UV; MT/AP; endometrial echogenicity (EE: homogeneous or heterogeneous); endometrial lining (EL: regular or irregular)). EV and UV (expressed in mL) were calculated from three measurements of the endometrium or the uterus in two perpendicular planes and the volume was calculated according to the formula for a prolate ellipsoid: $\pi/6 \times D1 \times D2 \times D3$ (where D1, D2, and D3 represent the three diameters of the structure). Blood flow indices obtained using color and spectral Doppler ultrasound included intratumoral peak systolic velocity (PSV), time-averaged maximum mean velocity (TAMXV), resistance index (RI) and pulsatility index (PI). Furthermore, uterine artery PSV, TAMXV (maximum of the values measured at both the left and right uterine arteries, i.e. the worst case), RI and PI (minimum of the values measured at both the left and right uterine arteries) were measured. The subjective assessment by the gynecologist of the depth of myometrial invasion (using a four-value scoring system: 0 = Stage Ia; 1 = Stage Ib; 2 = Stage Ic; 3 = Stage II or higher) was also recorded. The gynecologist was not blinded to the histological results and tumor grading but he based his assessment mainly on the volume of the tumor and the myometrium remaining between tumor and serosa.

Univariate analysis

Univariate analysis was performed using the SAS software package (Release 8.01; SAS Institute Inc., Cary, NC, USA).

We used the Wilcoxon rank-sum test (for continuous data) and Fisher's exact test (for categorical data) to calculate *P*-values that reflected whether there was a significant difference for a certain variable between patients with surgical Stages Ia or Ib and patients with surgical Stages Ic or higher[24]. In addition, the ROC curves and the AUCs were estimated[25] and compared[26] for the individual parameters using custom scripts written in MATLAB (Version 6.5 Release 13; The Mathworks, Inc., Natick, MA, USA. See also Epstein *et al.*[27] in which the same scripts were applied). The optimal cut-off point on the ROC curve was defined as the point that obtained the best trade-off between sensitivity and specificity (point at which the tangent to the ROC curve had a slope of 1, for which it could be proven that it maximized the sum of the sensitivity and specificity). The resulting sensitivity and specificity values were also calculated. For all hypothesis tests, two-sided tests were used and $P < 0.05$ was used as the level of significance.

Multivariate analysis

We trained three models (i.e. used the patients of the training set to determine the coefficients of a model in order to optimize its ability to differentiate between patients with and without deep myometrial invasion) based on a set of variables selected after stepwise logistic regression analysis. Subsequently, these models were validated prospectively on a new and independent set of patients. A schematic overview of the multivariate analysis procedure is given in Figure 1.

Variable selection

With multivariate stepwise logistic regression analysis (using stepwise selection in the LOGISTIC procedure from SAS) we aimed to select the variables that contributed significantly in a standard logistic regression model that predicted deep myometrial invasion. We considered the following variables for inclusion in the model: the ultrasound parameters discussed above, the number of fibroids detected during ultrasound examination (NF; range 0–2; this parameter has been reported to be a potential factor disturbing sonographic prediction, leading to overestimation of invasion[28]), the degree of differentiation of the cancer, the presence of a clear cell component and the presence of a serous papillary component. Note that the latter three (histopathological) variables were assessed by endometrial biopsy preoperatively (using Pipelle de Cornier). In the model, obtained at the end of the stepwise logistic regression analysis, only variables having a coefficient significantly different from zero (*P*-value < 0.05; Wald chi-square statistic) were allowed[29]. Note that only 74 of the 97 patients from the training set could be used for the stepwise logistic regression analysis because of missing values in some of the variables considered.
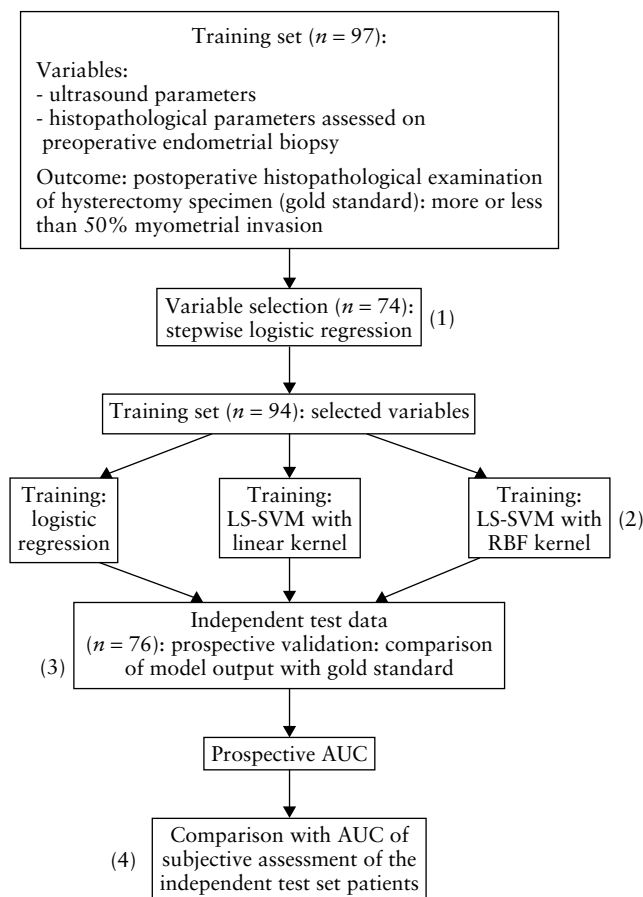
Figure 1 Schematic overview of multivariate analysis and model-building. (1) Variable selection step: using stepwise logistic regression analysis and the training set, the variables that contributed significantly in a standard logistic regression model (that aims to predict the degree (more or less than 50%) of myometrial invasion as assessed by the final histopathological examination) were selected. Note that the values for all variables that were considered for inclusion in the logistic regression model were known preoperatively and could therefore be used to make a (preoperative) prediction of the result of the final (and postoperative) histopathological examination of the degree of myometrial invasion. (2) Model training (determination of the coefficients of a model in order to optimize its classification performance using the training set): a standard logistic regression model and least squares support vector machines (LS-SVM) models with linear and radial basis function (RBF) kernels (also aiming to predict the result of the final histopathological assessment) were fitted to the training data. The variables used in these models were restricted to the variables selected in Step 1. Model training also involved the determination of an optimal cut-off level with the best trade-off between sensitivity and specificity as assessed on a receiver–operating characteristics (ROC) curve. Patients with a model output larger than the cut-off were predicted to have an endometrial cancer of Stage Ic or higher. Because the calculations in Steps 1 and 2 were based on the patients without missing values in any of the variables, the number of patients used in Step 2 (in which only a subset of the variables was taken into account) could be larger than that in Step 1. (3) Prospective validation: the models trained in Step 2 were applied subsequently on an independent set of new patients that had not been used in model training. ROC curves (and the associated areas under the curve (AUCs)) were constructed by comparing the model output with the final histopathological assessment of the degree of myometrial invasion. (4) Finally, the model AUCs were compared with the AUC of the expert subjective assessment of the same independent test set patients.

Model-building

The variables selected after the stepwise logistic regression analysis were used subsequently to fit a standard logistic regression model and least squares support vector machine (LS-SVM) models[30] with linear and radial basis function (RBF) kernels to the training set.

Support vector machines are a relatively new method for solving classification problems and have already been used extensively for various applications, including medical ones[31] (for more details, see the Opinion published in the same issue of this Journal[32]).

Since the models in this section were based on only a subset of the variables used during variable selection and since only the patients without missing values in any of the variables could be taken into account, the number of patients in the model-building step (94) was larger than the number of patients used in the variable selection step (74). As described above, the single valued output of the models could also be analyzed and compared using the Wilcoxon rank-sum test and ROC curves, and could also be used to estimate an optimal cut-off point or threshold for these models. Patients with a model output larger than this cut-off were then predicted to have deep myometrial invasion.

The standard logistic regression model was fitted with the LOGISTIC procedure from SAS. The class labels for patients with Stages Ia or Ib were 0, and they were 1 for patients with Stage Ic or higher. The Wald chi-square statistic was used to assess the significance of the coefficient of a certain variable in the fitted model.

Using LS-SVMlab version 1.5[30,33] for MATLAB we trained two LS-SVM models using a linear and an RBF kernel. It is possible to write an LS-SVM with a linear kernel as a simple linear equation in its variables. An LS-SVM with an RBF kernel has a more complex form, (in this case it was a sum with 95 terms) which is why it is not stated explicitly in this manuscript.

Prospective validation

In the previous section, the AUCs of the mathematical models were estimated using the same collection of patients that was used to fit or train these models. This could have led to results that were too optimistic. Therefore, we validated prospectively our results using independent data from 78 consecutive new patients. Here we refer to these patients as the 'independent test set', which became available after the derivation of these models (collected prospectively). The mean age of the patients in the test set was 64.1 years (range, 31–89 years) and 72 of them were postmenopausal. They were assessed using the same protocol as that used for the patients of the training set. The distribution of their FIGO stages was: 14 Stage Ia, 36 Stage Ib, 16 Stage Ic, one Stage II, nine Stage III and two Stage IV. The following histopathological subtypes were present: 59 endometrioid adenocarcinoma, one mucinous, two serous papillary, 15 mixed type (of which nine had a serous papillary and four

a clear cell component) and one endometrial tumor with unspecified histopathological subtype. Forty tumors were differentiated highly, 14 moderately and 24 poorly. Using these independent test data, we calculated the AUCs of the three models discussed above and compared them with the AUC of the subjective assessment of the expert. We also evaluated the performance of our models at the optimal cut-off points obtained after the ROC analysis of the training set. We used the method described by Hanley and McNeil[25,26] to estimate the sample size needed to reach statistical significance.

RESULTS

The results (based on the training set) of the univariate analysis of the ultrasound parameters and the subjective assessment are presented in Table 1. Of all the ultrasound parameters, EV/UV had the largest AUC (78%), comparable to that of the subjective assessment (79%; difference not statistically significant). Also, there was no significant difference between the AUC of EV/UV and the AUCs of ET, MT, EV, ET/AP and MT/AP. Compared to these morphological parameters, the AUCs of the blood flow indices were low. Uterine artery RI and PI were higher in Stages Ia−Ib compared with Stages Ic or higher (differences were significant but *P*-values were close to 5%).

Multivariate stepwise logistic regression selected the degree of differentiation, NF, ET and EV as variables that contributed significantly in a standard logistic regression model aiming to discriminate between patients with and without deep myometrial invasion on the final histopathological assessment. None of the blood flow indices was selected.

The resulting logistic regression model fitted to the training data was given by:

$$y = \frac{\exp(\beta_0 + \beta_1.DD1 + \beta_2.DD2 + \beta_3.NF + \beta_4.ET + \beta_5.EV)}{1 + \exp(\beta_0 + \beta_1.DD1 + \beta_2.DD2 + \beta_3.NF + \beta_4.ET + \beta_5.EV)},$$

where DD1 and DD2 equal 1 if, respectively, the tumor is moderately and poorly differentiated, and 0 in other cases, and where *y* is the model output, which is a number on a continuous scale between 0 and 1 (note that since we had to take only the missing variables in the four selected variables into account, 94 patients could be used to fit the three models, which is more than the number of patients (74) that was used for variable selection). A patient was predicted to have a tumor of Stage Ia or Ib if $y \leq a$ certain cut-off level and was predicted to have a tumor of Stage Ic or higher if $y >$ this cut-off level. The coefficients (rounded to two decimal places) were: $\beta_0 = -3.70$ (95% CI, −5.53 to −1.86, $P < 0.0001$), $\beta_1 = 2.36$ (95% CI, 0.82 to 3.91, $P = 0.0027$), $\beta_2 = 2.42$ (95% CI, 1.00 to 3.84, $P = 0.0008$), $\beta_3 = -2.45$ (95% CI, −4.23 to −0.67, $P = 0.0070$), $\beta_4 = 0.20$ (95% CI, 0.07 to 0.32, $P = 0.0021$) and $\beta_5 = -0.11$ (95% CI, −0.19 to −0.03,

Table 1 Univariate analysis of the ultrasound parameters, the subjective assessment, the standard logistic regression model and the least squares support vector machines (LS-SVM) models with a linear and radial basis function (RBF) kernel (training set, $n = 97$)

	Range	AUC [95% CI]	Optimal cut-off value*	Sensitivity (%)	Specificity (%)	Mean or proportion in Stage Ia or Stage Ib	Mean or proportion in Stage Ic or higher	P
Endometrial thickness (ET) (mm)	2–65	0.76 [0.66, 0.86]	14	81	64	15	25	< 0.0001
Myometrial thickness (MT) (mm)	2–18	0.71 [0.59, 0.82]	8	74	61	8.8	6.4	0.001
Endometrial volume (EV) (mL)	0–84	0.76 [0.66, 0.86]	4.9	71	69	8.2	18	< 0.0001
Uterine volume (UV) (mL)	16–1075	0.61 [0.49, 0.72]	89	58	69	91	147	0.08
ET/uterine anteroposterior diameter (AP)	0.07–1.5	0.75 [0.65, 0.86]	0.43	72	71	0.37	0.54	< 0.0001
EV/UV	< 0.0001–0.75	0.78 [0.68, 0.87]	0.09	69	80	0.07	0.15	< 0.0001
MT/AP	0.04–0.44	0.75 [0.64, 0.85]	0.17	74	75	0.24	0.15	< 0.0001
Endometrial echogenicity (EE) (% heterogeneous)	—	0.60 [0.49, 0.72]	—	65	56	44%	65%	0.06
Endometrial lining (EL) (% irregular)	—	0.61 [0.50, 0.73]	—	78	44	56%	78%	0.03
Intratumoral								
PSV (cm/s)	0–0.96	0.61 [0.49, 0.73]	0.13	59	64	0.14	0.21	0.09
TAMXV (cm/s)	0–0.77	0.61 [0.49, 0.73]	0.06	82	46	0.09	0.14	0.09
RI	0.05–1	0.62 [0.48, 0.75]	0.5	50	78	0.62	0.54	0.08
PI	0.23–6.0	0.61 [0.48, 0.74]	0.61	38	88	1.4	1.1	0.10
Uterine artery								
Peak systolic velocity (PSV) (cm/s)	0.09–2.1	0.51 [0.39, 0.65]	0.62	31	84	0.49	0.53	0.81
TAMXV (cm/s)	0.04–0.75	0.57 [0.45, 0.70]	0.25	37	80	0.20	0.24	0.27
Resistance index (RI)	0.41–1.2	0.64 [0.52, 0.76]	0.71	49	78	0.78	0.71	0.03
Pulsatility index (PI)	0.16–6.0	0.64 [0.52, 0.76]	1.3	49	78	1.9	1.5	0.04
Subjective assessment (Stage Ia: 0; Stage Ib: 1; Stage Ic: 2; Stage II or higher: 3)	0–3	0.79 [0.69, 0.88]	1	61	86	0:51% 1:36% 2:12% 3:2%	0:13% 1:26% 2:39% 3:21%	< 0.0001
Standard logistic regression	0–1	0.89 [0.83, 0.96]	0.45	77	86	0.21	0.65	< 0.0001
LS-SVM with linear kernel	−1.5 to 1.4	0.88 [0.81, 0.95]	−0.31	91	73	−0.52	0.20	< 0.0001
LS-SVM with RBF kernel	−1.2 to 0.93	0.99 [0.97, 1]	−0.30	97	100	−0.74	0.56	< 0.0001

*The optimal cut-off point was defined as the point that obtained the best trade-off between sensitivity and specificity. *P*-values show the statistical significance of any differences between results shown in the previous two columns. AUC, area under the receiver–operating characteristics curve; TAMXV, time-averaged maximum mean velocity

$P = 0.0054$). These coefficients indicate that the predicted probability of deep myometrial invasion increased when the degree of differentiation and the ET increased and that the predicted probability of deep myometrial invasion decreased when the NF and the EV increased. The negative influence of the EV was unexpected, but can be seen as a non-linear effect of the ET (since EV $\sim$ ET^3). The performance of the standard logistic regression model on the training data and the optimal cut-off level are also summarized in Table 1.

The resulting LS-SVM model with a linear kernel fitted to the training data was given by:

$$y = \beta_0 + \beta_1.DD + \beta_2.NF + \beta_3.ET + \beta_4.EV,$$

where DD equals 1, 2 and 3 if the degree of differentiation is highly, moderately and poorly differentiated, respectively and where y is the model output, which is a number on a continuous scale. Again, a patient was predicted to have a tumor of Stage Ia or Ib if $y \leq$ a certain cut-off level

and was predicted to have a tumor of Stage Ic or higher if $y >$ this cut-off level. The coefficients (rounded to two decimal places) were: $\beta_0 = -1.44$, $\beta_1 = 0.37$, $\beta_2 = -0.37$, $\beta_3 = 0.05$ and $\beta_4 = -0.03$. According to the sign of these coefficients, the influence of the different variables was the same qualitatively as that in the logistic regression model.

As mentioned previously, the LS-SVM model with an RBF kernel could not be written in a simplified form and is therefore not stated explicitly here. However, it could be implemented easily in for example, Microsoft Excel. The model output was a single and continuous number that had to be compared with a certain cut-off level. The performance of the LS-SVM models with a linear and RBF kernel on the training data and the optimal cut-off levels are also described in Table 1.

Evaluated on the training set, the standard logistic regression and the LS-SVM models with a linear and RBF kernel had a larger AUC than did the subjective

Table 2 Prospective validation: performance of the standard logistic regression model and the least squares support vector machines (LS-SVM) models with linear and radial basis function (RBF) kernels for the patients of the independent test set; comparison with the ultrasound parameter (endometrial/uterine volume (EV/UV)) from Table 1 with the best discriminatory potential and the subjective assessment ($n = 78$ for the subjective assessment and $n = 76$ for EV/UV and the mathematical models)

	AUC [95% CI]	Optimal cut-off value*	Sensitivity (%)	Specificity (%)	LR+	LR−
EV/UV	0.70 [0.58, 0.82]	0.085	57	72	2.1	0.59
Subjective assessment	0.72 [0.59, 0.84]	1	61	80	3.0	0.49
Standard logistic regression	0.66 [0.53, 0.79]	0.45	50	75	2.0	0.67
LS-SVM with linear kernel	0.72 [0.59, 0.84]	−0.31	75	69	2.4	0.36
LS-SVM with RBF kernel	0.77 [0.66, 0.87]	−0.30	79	67	2.4	0.32

*The optimal cut-off values were taken from Table 1 as evaluated on the training set. LR+, positive likelihood ratio; LR−, negative likelihood ratio.

assessment. This difference was only significant for the LS-SVM with an RBF kernel ($P < 0.0001$) and had borderline significance for the standard logistic regression model ($P = 0.0595$).

The results of the prospective validation, which was only possible in 76 (of 78) test-set patients because of missing values in EV, are presented in Table 2 and Figure 2. From these results we can conclude that prospective evaluation on the independent test set resulted in a higher AUC only for the LS-SVM model with a RBF kernel (difference not significant) and in an equally good AUC for the LS-SVM model with a linear kernel when compared with the AUC of the subjective assessment. The performance of the standard logistic regression model was poor. For the optimal cut-off value, the positive likelihood ratio for a positive result (positive likelihood ratio, LR+) of the subjective assessment was better compared with that of the LS-SVM models. The opposite was true for the negative likelihood ratio (LR−). This means that, at the chosen cut-off level, the LS-SVM models were better at ruling out deep myometrial invasion than they were at ruling it in, when compared with the subjective assessment.

DISCUSSION

Our study indicates that single morphological parameters do not improve the predictive power when compared with subjective assessment, and that spectral Doppler analysis does not contribute to the prediction of the degree of myometrial invasion in endometrial cancer. Combining the degree of differentiation, ET, EV and NF in an LS-SVM model with a linear or RBF kernel might deliver predictions that are as reliable as is the subjective impression of an experienced sonologist. Assuming that a real difference exists between the true AUC of the LS-SVM model with an RBF kernel and the true AUC of the subjective assessment, the number of patients in the independent test set, however, was not sufficient to reach statistical significance in a prospective evaluation. If the values in Table 2 represent the true AUCs (i.e. those that would be achieved by infinite populations), one would need a sample size of approximately 919 patients to be able to detect, with 80% power, the difference between these AUCs as being statistically significant[34]. Confirmation of the performance of LS-SVM models with an RBF kernel in larger prospective studies is therefore necessary.

As could be expected and as is explained in the Opinion of this issue[32], the performance on the test set or level of generalization of the LS-SVM model with a linear kernel was better than was the performance of the standard logistic regression model. Evaluation on the training set (Table 1) gave the opposite order of performance, although the difference was small. The LS-SVM model with an RBF kernel had the best overall performance, both on the training set and on the independent test set. This is an indication that non-linear effects might play a role in the distinction between patients with and those without deep myometrial invasion. The better sensitivity for deep invasion of the LS-SVM model could be helpful in selecting patients who might benefit from a pelvic lymphadenectomy by an experienced surgeon.
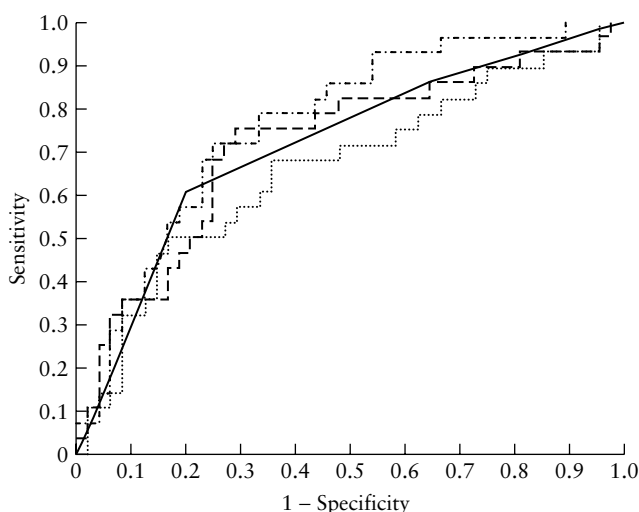
Figure 2 Comparison of the receiver–operating characteristics (ROC) curves for the subjective assessment (———), the standard logistic regression model (··········), and the least squares support vector machines (LS-SVM) models with a linear (– – – –) and RBF (–··–··–) kernel for the patients of the independent test set ($n = 78$ for the subjective assessment and $n = 76$ for endometrial/uterine volume and the mathematical models).

It is important to emphasize that the models described in this study might not be ready to be implemented in routine clinical practice. First of all, the measurements that were considered in our study all originated from the same sonologist. Because of differences that might exist between different centers, or even individual sonologists (who might, for example, use different ultrasound equipment), the models discussed here should be tested on multicenter prospective data using a stringent and detailed protocol; we have planned this multicenter prospective study. Moreover, the techniques used by the same expert might undergo subtle changes with time, causing a drop in model performance when the model is applied on new patients. These comments also apply to the evaluation of the degree of differentiation, a variable that was also included in our models. This parameter is, at least partially, a subjective measure that can differ between centers, between pathologists and in time. There is also the possibility of change in the characteristics of the population of patients, causing new patients to be drawn from a distribution different from the one that was used to derive the models. This again might cause a drop in model performance when applied to new data.

Despite these possible limitations, we believe that the proposed models could represent a simple and inexpensive method that might contribute to the preoperative distinction between low- and high-risk patients, allowing for better preoperative allocation of patients with endometrial carcinoma. Further research is therefore needed in this area.

REFERENCES

1. Young RC. Gynecologic malignancies. In *Harrison's Principles of Internal Medicine* (14th edn), Fauci AC, Braunwald E, Isselbacher KJ, Wilson JD, Martin JB, Kasper DL, Hauser SL, Longo DL (eds). McGraw-Hill: New York, 1998; 605–611.

2. Creasman WT, Odicino F, Maisonneuve P, Beller U, Benedet JL, Heintz APM, Ngan HYS, Pecorelli S. Carcinoma of the corpus uteri. *Int J Gynaecol Obstet* 2003; **83**: 79–118.

3. Levine DA, Hoskins WJ. Update in the management of endometrial cancer. *Cancer J* 2002; **8** (Suppl 1): S31–S40.

4. Ludwig H. Prognostic factors in endometrial cancer. *Int J Gynecol Obstet* 1995; **49** (Suppl): S1–S7.

5. Franchi M, Ghezzi F, Melpignano M, Cherchi PL, Scarabelli C, Apolloni C, Zanaboni F. Clinical value of intraoperative gross examination in endometrial cancer. *Gynecol Oncol* 2000; **76**: 357–361.

6. Kucera E, Kainz C, Reinthaller A, Sliutz G, Leodolter S, Kucera H, Breitenecker G. Accuracy of intraoperative frozen-section diagnosis in stage I endometrial adenocarcinoma. *Gynecol Obstet Invest* 2000; **49**: 62–66.

7. Kinkel K, Kaji Y, Yu KK, Segal MR, Lu Y, Powell CB, Hricak H. Radiologic staging in patients with endometrial cancer: a meta-analysis. *Radiology* 1999; **212**: 711–718.

8. Artner A, Bosze P, Gonda G. The value of ultrasound in preoperative assessment of the myometrial and cervical invasion in endometrial carcinoma. *Gynecol Oncol* 1994; **54**: 147–151.

9. Cacciatore B, Lehtovirta P, Wahlstrom T, Ylostalo P. Preoperative sonographic evaluation of endometrial cancer. *Am J Obstet Gynecol* 1989; **160**: 133–137.

10. Develioglu OH, Bilgin T, Yalcin OT, Ozalp S, Ozan H. Adjunctive use of the uterine artery resistance index in the preoperative prediction of myometrial invasion in endometrial carcinoma. *Gynecol Oncol* 1999; **72**: 26–31.

11. Fishman A, Altaras M, Bernheim J, Cohen I, Beyth Y, Tepper R. The value of transvaginal sonography in the preoperative assessment of myometrial invasion in high and low grade endometrial cancer and in comparison to frozen section in grade 1 disease. *Eur J Gynaecol Oncol* 2000; **21**: 128–130.

12. Gordon AN, Fleischer AC, Reed GW. Depth of myometrial invasion in endometrial cancer: preoperative assessment by transvaginal ultrasonography. *Gynecol Oncol* 1990; **39**: 321–327.

13. Karlsson B, Norstrom A, Granberg S, Wikland M. The use of endovaginal ultrasound to diagnose invasion of endometrial carcinoma. *Ultrasound Obstet Gynecol* 1992; **2**: 35–39.

14. Olaya FJ, Dualde D, Garcia E, Vidal P, Labrador T, Martinez F, Gordo G. Transvaginal sonography in endometrial carcinoma: preoperative assessment of the depth of myometrial invasion in 50 cases. *Eur J Radiol* 1998; **26**: 274–279.

15. Prompeler HJ, Madjar H, du Bois A, Lattermann U, Wilhelm C, Kommoss F, Pfleiderer A. Transvaginal sonography of myometrial invasion depth in endometrial cancer. *Acta Obstet Gynecol Scand* 1994; **73**: 343–346.

16. Weber G, Merz E, Bahlmann F, Mitze M, Weikel W, Knapstein PG. Assessment of myometrial infiltration and preoperative staging by transvaginal ultrasound in patients with endometrial carcinoma. *Ultrasound Obstet Gynecol* 1995; **6**: 362–367.

17. Sahakian V, Syrop C, Turner D. Endometrial carcinoma: transvaginal ultrasonography prediction of depth of myometrial invasion. *Gynecol Oncol* 1991; **43**: 217–219.

18. Szantho A, Szabo I, Csapo ZS, Balega J, Demeter A, Papp Z. Assessment of myometrial and cervical invasion of endometrial cancer by transvaginal sonography. *Eur J Gynaecol Oncol* 2001; **22**: 209–212.

19. Arko D, Takac I. High frequency transvaginal ultrasonography in preoperative assessment of myometrial invasion in endometrial cancer. *J Ultrasound Med* 2000; **19**: 639–643.

20. Timmerman D, De Smet F, De Brabanter J, Thijs I, Moreau Y, De Moor B, Vergote I. Pre-operative prediction of depth of myometrial invasion in patients with endometrial cancer: evaluation of ultrasound parameters and development of a

new logistic regression model. *Int J Gynecol Cancer* 2003; **13** (Suppl 1): S26.

21. Dijkhuizen FP, Mol BW, Brolmann HA, Heintz AP. The accuracy of endometrial sampling in the diagnosis of patients with endometrial carcinoma and hyperplasia: a meta-analysis. *Cancer* 2000; **89**: 1765–1772.

22. Clark TJ, Mann CH, Shah N, Khan KS, Song F, Gupta JK. Accuracy of outpatient endometrial biopsy in the diagnosis of endometrial cancer: a systematic quantitative review. *BJOG* 2002; **109**: 313–321.

23. Amant F, Moerman P, Neven P, Timmerman D, Van Limbergen E, Vergote I. Endometrial cancer. *Lancet* 2005; **366**: 491–505.

24. Dawson-Saunders B, Trapp RG. *Basic & Clinical Biostatistics* (2nd edn). Appleton & Lange: Connecticut, 1994.

25. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982; **143**: 29–36.

26. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* 1983; **148**: 839–843.

27. Epstein E, Skoog L, Isberg PE, De Smet F, De Moor B, Olofsson PA, Gudmundsson S, Valentin L. An algorithm including results of gray-scale and power Doppler ultrasound examination to predict endometrial malignancy in women with postmenopausal bleeding. *Ultrasound Obstet Gynecol* 2002; **20**: 370–376.

28. Weber G, Merz E, Bahlmann F, Mitze M, Weikel W, Knapstein PG. Assessment of myometrial infiltration and preoperative staging by transvaginal ultrasound in patients with endometrial carcinoma. *Ultrasound Obstet Gynecol* 1995; **6**: 362–367.

29. Hosmer DW, Lemeshow S. *Applied Logistic Regression*. John Wiley & Sons: New York, 1989.

30. Suykens JAK, Van Gestel T, De Brabanter J, De Moor B, Vandewalle J. *Least Squares Support Vector Machines*. World Scientific: Singapore, 2002.

31. Van Gestel T, Suykens J, Baesens B, Viaene S, Vanthienen J, Dedene G, De Moor B, Vandewalle J. Benchmarking least squares support vector machine classifiers. *Machine Learning* 2004; **54**: 5–32.

32. Pochet NLMM, Suykens JAK. Support vector machines versus logistic regression: improving prospective performance in clinical decision making. *Ultrasound Obstet Gynecol* 2006; **27**: 607–608.

33. LS-SVMlab version 1.5. http://www.esat.kuleuven.ac.be/sista/lssvmlab/ [Accessed 21 April 2006].

34. Jaeschke R, Guyatt G, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. A. Are the results of the study valid? B. What are the results and will they help me in caring for my patients? Evidence-Based Medicine Working Group. *JAMA* 1994; **271**: 389–391 and 703–707.