A new method for modelling preoperative diagnosis of ovarian tumours

Viktoriya Stalbovskaya, Emmanuel C. Ifeachor*, *Member, IEEE,* Sabine Van Huffel, *Senior Member, IEEE,* and Dirk Timmerman

Abstract

In this paper we present a sequential non-uniform procedure, an inference method which combines feature selection based on the Kullback information gain and a step-wise classification procedure to produce a reliable, interpretable and robust model. We applied the model to an ovarian tumour data set to distinguish between malignant and benign tumours. The performance of the model was assessed using ROC analysis and gave an overall accuracy over 85%, AUC 0.887 which compares well with existing methods. The method presented here is significant because of its ability to handle missing values and it only uses a small number of variables which are graded according to their discriminative relevance. This, together with the fact that the resulting model is interpretable and has good performance, is likely to lead to widespread clinical acceptance of the method. The method is also generic and can be readily adapted for other classifications problems in biomedicine.

Index Terms

inference system, sequential non-uniform procedure, Kullback Leibler divergence, cancer classification, ovarian tumour, medical diagnosis, ultrasound

Manuscript received July 28, 2006. This work was supported by the European Union Network of Excellence BIOPATTERN (Contract No. FP6-2002-IST 508803). *Asterisk indicates corresponding author.*

V. Stalbovskaya is with the Signal Processing and Multimedia Communications Group, School of Computing, Communications and Electronics, University of Plymouth, Drake Circus, Plymouth, PL4 8AA, UK (e-mail: vstalbovskaya@plymouth.ac.uk).

E. C. Ifeachor is with the Signal Processing and Multimedia Communications Group, School of Computing, Communications and Electronics, University of Plymouth, Drake Circus, Plymouth, PL4 8AA, UK (phone: +44(0)1752-232574; fax: +44(0)1752-232583; e-mail: e.ifeachor@plymouth.ac.uk).

S. Van Huffel is with the Department of Electrical Engineering ESAT-SCD (SISTA), Katholieke Universiteit Leuven, 3001 Leuven, Belgium (e-mail: Sabine.VanHuffel@esat.kuleuven.be).

D. Timmerman is with the Department of Obstetrics and Gynaecology, University Hospitals, Katholieke Universiteit Leuven, Herestraat 49, 3000 Leuven, Belgium (e-mail: dirk.timmerman@uz.kuleuven.ac.be).

I. Introduction

Ovarian tumours are common among women. In Europe and North America the age-adjusted standardised incidence rate of ovarian cancer is over 10 per 100,000 women [1]. According to the EUROCARE data [2], five year survival rate across Europe is 35%. The high mortality is caused by late diagnosis when cancer is detected on the already advanced stage. The main reason for this is a lack of pathognomonic signs of early stages of the neoplastic process.

Preoperative diagnosis is very important, because it can prevent unnecessary surgery for benign functional cysts or in the case of benign neoplastic lesions only minimal surgical intervention would be required. On the other hand, patients with malignant forms of tumour require not only surgical operation but also an appropriate pre-, peri- and postoperative management. A number of papers have been devoted to the development of algorithms and systems for preoperative prediction of ovarian tumours. These include scoring systems, such as the risk of malignancy index [3], logistic regression models [4], [5], artificial neural networks [6], support vector machines [7], and neuro-fuzzy models [8]. A number of these show high performance and some are used in a clinical practice, but are not readily interpretable.

For clinical acceptance, a predictive model should satisfy the following requirements: the model should

(i) have reasonably high sensitivity and specificity levels, typically 90% and 75%, respectively [9],
(ii) be interpretable, and
(iii) use as few diagnostic techniques/parameters as possible.

In relation to (iii), the range of laboratory and instrumental diagnostic techniques for ovarian cancer is wide and includes transvaginal and transabdominal ultrasonography, serum tumour markers, laparoscopy, computer tomography and magnetic resonance imaging. The problem is in the choice of necessary procedures taking into account their diagnostic value, cost and invasiveness.

The aim of this paper is to present a method, based on the Sequential Non-uniform Procedure (SNuP), which meets the requirements above. SNuP is based on the Naïve Bayes classification, but with additional restrictions. In particular, consecutive multiplication of likelihood ratios of input variables, $\frac{P(x_i|A_1)}{P(x_i|A_2)}$ is interrupted when one of the diagnostic thresholds [10] is reached. Values of thresholds are specified according to an acceptable level of the diagnostic errors.

The method presented here is a variant of Wald's Sequential Probability Ratio Test (SPRT) which is widely used for hypothesis testing in areas such as fault diagnosis in control and sensor systems [11], [12].

In our approach, in contrast to SPRT, the SNuP operates sequentially on the variables (features) as

the cases (observations) are accumulated. This is important because it makes it possible to personalise differential diagnosis. This is achieved by varying the number of attributes used, ranking the variables according to their discriminative relevance and the specified confidence level.

The remainder of the paper is organised as follows. In Section II, the new method for data analysis for ovarian tumours is described. In Section III, an application of the method to ovarian tumour data is given followed by a performance evaluation. Finally, we present the conclusion and future work.

II. A NEW APPROACH FOR DATA ANALYSIS OF OVARIAN TUMOURS

A. Problem outline

The key issue in preoperative diagnosis is to determine whether a given patient belongs to one of two groups: benign or malignant tumour, given the symptoms and laboratory data. The task can be viewed as a two-class classification (A_k, where $k = 1, 2$) problem, given a vector of input variables, $\mathbf{x}$.

B. Methods of analysis

Let us denote $P(A_k)$ - prior probability of class k, $k = 1 \dots n$ - number of classes (groups), $P(x_i|A_k)$ - conditional probability of x_i given A_k, i.e. probability of presence of symptom x_i in the group A_k, $P(x_i)$ - prior probability of symptom x_i. So the posterior probability of the patient to belong to the group A_k having symptom x_i can be defined using Bayes' theorem:

$$P(A_k|x_i) = \frac{P(A_k)P(x_i|A_k)}{\sum\limits_k P(A_k)P(x_i|A_k)} \tag{1}$$

Sequential non-uniform procedure produces a model of classification into two groups. The ratio of conditional probabilities of the groups is equal to the ratio of symptom's occurrences in the two groups.

$$\frac{P(A_1|x_i)}{P(A_2|x_i)} = \frac{P(x_i|A_1)}{P(x_i|A_2)} \tag{2}$$

where $\frac{P(A_1|x_i)}{P(A_2|x_i)}$ is a likelihood ratio of probability of a group given symptom x_i, $\frac{P(x_i|A_1)}{P(x_i|A_2)}$ is a likelihood ratio of the probability of the symptom x_i given groups A_k.

Accumulation of the diagnostic information given the presence of independent features/symptoms $x_1, x_2, ..., x_n$ is performed as

$$\frac{P(A_1|x_1, x_2, ..., x_n)}{P(A_2|x_1, x_2, ..., x_n)} = \prod_{i=1}^{n} \frac{P(x_i|A_1)}{P(x_i|A_2)} \tag{3}$$

The *inference process* uses two types of errors to determine the thresholds for the ratio in (3) to make a decision - α and β. In terms of 'disease-health' classification, α specifies the probability of false

assignment of a diseased patient into a healthy group, and β specifies the probability of false assignment of a healthy patient to a diseased group. In terms of classification into groups A_1 and A_2, α is the rate of misclassification in group A_1, β is the rate of misclassification in group A_2.

The threshold for a diagnostic hypothesis is the minimum acceptable rate of correct diagnoses over incorrect ones. Let's denote A^+ as correct diagnosis and A^- as incorrect diagnosis. The probabilities of correct and incorrect diagnoses in the groups are $P(A_1^+)$, $P(A_1^-)$, $P(A_2^+)$, $P(A_2^-)$. So the decision rule for group 1 is $\frac{P(A_1|x_1,x_2,...)}{P(A_2|x_1,x_2,...)} \geq \frac{P(A_1^+)}{P(A_1^-)}$, for group 2 is $\frac{P(A_1|x_1,x_2,...)}{P(A_2|x_1,x_2,...)} \leq \frac{P(A_2^-)}{P(A_2^+)}$, where $\frac{P(A_1^+)}{P(A_1^-)}$ and $\frac{P(A_2^-)}{P(A_2^+)}$ are the levels of acceptable classification errors. Using types I and II errors $P(A_1^+) = 1-\alpha$, $P(A_1^-) = \beta$. The ratio of correct to incorrect diagnoses in group 1 is

$$\frac{P(A_1^+)}{P(A_1^-)} = \frac{1-\alpha}{\beta} \tag{4}$$

Similarly for group 2, $P(A_2^-) = \alpha$, $P(A_2^+) = 1-\beta$ and the ratio of correct to incorrect diagnoses is

$$\frac{P(A_2^-)}{P(A_2^+)} = \frac{\alpha}{1-\beta} \tag{5}$$

The threshold values for different levels of α and β are listed in Table I.

Considering the task of assignment of the input vector $\mathbf{x}$ to one of the groups A_1 or A_2 the inference rules for the SNuP are as follows:

- If $\frac{P(A_1|x_1,x_2,...)}{P(A_2|x_1,x_2,...)} \geq \frac{1-\alpha}{\beta}$
  then the decision is "$\mathbf{x} \in$ Group A_1".

- If $\frac{P(A_1|x_1,x_2,...)}{P(A_2|x_1,x_2,...)} \leq \frac{\alpha}{1-\beta}$
  then the decision is "$\mathbf{x} \in$ Group A_2".

- If $\frac{\alpha}{1-\beta} < \frac{P(A_1|x_1,x_2,...)}{P(A_2|x_1,x_2,...)} < \frac{1-\alpha}{\beta}$
  then additional information is required to assign $\mathbf{x}$ to one of the groups.

- If $\frac{\alpha}{1-\beta} < \frac{P(A_1|x_1,x_2,...)}{P(A_2|x_1,x_2,...)} < \frac{1-\alpha}{\beta}$
  and no more features are available then the decision is "membership of $\mathbf{x}$ is undefined".

In order to remove the multiplication operation in the right part of (3), we can transform it to a summation by taking a logarithm. A diagnostic coefficient DC_i of symptom x_i is a score value which

is defined as

$$DC_i = 10 \log_{10} \frac{P(x_i|A_1)}{P(x_i|A_2)} \tag{6}$$

When the probability of the symptom x_i is higher in group A_1 than in group A_2 the value of DC_i is greater than 0. When the probability of the symptom x_i is higher in the group A_2 the value of DC_i is less than 0.

Accumulation of the diagnostic information using the diagnostic coefficients is performed as a sum:

$$\sum DC(x_i) = DC(x_1) + DC(x_2) + ... + DC(x_n) \tag{7}$$

Information about *a priori* probability of classes $P(A_i)$ can be taken into account by adding $\frac{P(A_1)}{P(A_2)}$ to (3) and as an independent diagnostic coefficient $DC(x_0) = 10 \log_{10} \frac{P(A_1)}{P(A_2)}$ in (7).

Thresholds for the sum of the diagnostic coefficients are defined as

$$DC_{th}(A_1) = 10 \log_{10} \frac{1-\alpha}{\beta}$$
$$DC_{th}(A_2) = 10 \log_{10} \frac{\alpha}{1-\beta}$$

Values of the thresholds for different levels of acceptable errors are presented in the last two columns of Table I. The SNuP using diagnostic coefficients is performed until the following inequality is true:

$$DC_{th}(A_2) < \sum_i DC(x_i) < DC_{th}(A_1) \tag{8}$$

When the inequality is broken the diagnostic decision is made.

The feature selection process and ranking of input variables/symptoms is based on the calculation of symmetrised Kullback's information measure between two distributions, P and Q

$$J(P,Q) = \frac{\mathrm{D}(P\|Q) + \mathrm{D}(Q\|P)}{2} \tag{9}$$

where D is the Kullback's divergence so that $\mathrm{D}(P\|Q) = \sum_j P_j \log \frac{P_j}{Q_j}$ and $\mathrm{Q}(P\|P) = \sum_j Q_j \log \frac{Q_j}{P_j}$.

Adapting the equation above the information measure of the *distinct value* of the variable can be defined as

$$J(x_{ij}) = \frac{P(x_{ij}|A_1) - P(x_{ij}|A_2)}{2} 10 \log_{10} \frac{P(x_{ij}|A_1)}{P(x_{ij}|A_2)} \tag{10}$$

The information measure of the *variable* is the sum of the information measures of all its distinct values:

$$J(x_i) = \sum_j J(x_{ij}) \tag{11}$$

where m is the number of distinct values of variable (e.g. 'low', 'normal', 'high').

In order to obtain the most informative features the information measures were calculated for all variables and then sorted in a descending order according to the J-value.

An algorithm for building the decision rule for differential diagnosis involves the following steps.

1) calculate values of diagnostic coefficients DC_i for all symptoms x_i using (6).
2) calculate information measures for all x_i using (11) and sort values in a descending order.
3) specify an acceptable level of errors, α and β, and calculate the thresholds for diagnostic coefficients from (4) and (5).
4) start the process of accumulation of diagnostic information according to (7).
5) end the process when the inequality (8) is broken or there are no variables left.

Figure 1 demonstrates the inference process for four examples showing one-step SNuP (cases 1 and 2) and multi-step SNuP (cases 3 and 4).

The quality of the model was assessed by calculating the parameters of receiver operating characteristics: overall accuracy (Acc), sensitivity (Se), specificity (Sp), predictive positive value (PPV), predictive negative value (PNV) according to Table II and the equations given below

$$\begin{aligned} \mathrm{Acc} &= \frac{A+D}{A+B+C+D} \\ \mathrm{Se} &= \frac{A}{A+C} \\ \mathrm{Sp} &= \frac{D}{B+D} \\ \mathrm{PPV} &= \frac{A}{A+B} \\ \mathrm{PNV} &= \frac{D}{C+D} \end{aligned}$$

Evaluation of the model is performed by applying a 3-fold cross validation. The initial data set was split randomly into a training set and a test set with the proportion of malignant to benign cases equal to 1:2. The results were summarised as mean and standard deviation for Acc, Se, Sp, PPV, and PNV.

III. Application to ovarian tumour data

We applied the methodology described above to an ovarian data set. The data set was obtained in a previous study [4].

A. Ovarian tumour data

The study included 525 patients admitted to the Department of Obstetrics and Gynecology at the University Hospitals Katholieke Universiteit Leuven. All the patients underwent a transvaginal ultrasonography

with B-mode and colour Doppler imaging. The level of serum oncomarker CA125 was measured for 432 patients. A summary of the data set is given in Table III (a detailed description of the data acquisition process can be found in [4]). After excluding missing cases the data set consisted of 425 cases.

Transformation of the data. As part of the ultrasound examination the amount of blood flow was assessed within the septa, cyst walls, solid tumor areas, or ovarian stroma. Depending on wherever the amount of the blood flow was rather strong or very strong two new binary variables were added - 'Col3' and 'Col4'. The variable CA125 was transformed to binary values, 0 and 1, depending on a threshold value of 30 U/ml [3]. A value of 0 was assigned if CA125$\leq$30, a value of 1 otherwise.

The Risk of Malignancy Index (RMI) was used as a benchmark during the performance evaluation. RMI values were calculated according to the formula RMI=Jacobs$\times$Meno$\times$CA125 [3]. The ultrasound score (Morph) was calculated as the sum of scores for the presence of multilocular cyst, evidence of solid areas, evidence of metastases, presence of ascites and bilateral lesions. Jacobs' index was assigned a value of 0 if Morph=0, a value 1 if Morph=1 and a value 3 if Morph$>$1. The menopause state (Meno) was equal to 1 if premenopausal and equal to 3 if postmenopausal.

B. Results

The calculated diagnostic coefficients and information measures for all nominal input variables are presented in Table IV. The last column shows the rank of the symptom. The sequential non-uniform procedure for preoperative differential diagnosis between benign and malignant forms of adnexal tumour is recommended to start from the most informative variables, i.e. variables with the highest rank (e.g. smooth internal wall, strong blood flow, presence of unilocular cyst, level of serum CA125 above 30 U/ml, presence of ascites, etc).

In this paper we considered only binary variables. A large value of DC means a high discriminative ability of the variable and the information measures gives an indication of how reliable this is. Features with positive DC values correspond to malignancy, negative values are assigned to the benign group. Accumulation of the diagnostic information was carried out by summation of the diagnostic coefficients and comparison of the sum with a specified threshold.

The number of variables used for classification during SNuP is presented in Table V. The first column gives the levels of diagnostic errors, α and β. For every pair of error levels we calculated the minimum, median and maximum number of variables which were used for classification. The corresponding percentage of classifications made is given in brackets. As can be seen from the table, at least in half the cases the number of variables required to make a decision with 95% confidence ($\alpha = 0.05$, $\beta = 0.05$)

was no more than three (smoothness of internal wall, presence of the strong blood flow, and presence of unilocular cyst). Thus, for 50% of cases measurement of the level of serum CA125 was not necessary for the inference process.

C. Examples of classification

Let us consider two cases of ovarian tumour to illustrate how the algorithm works. A graphical representation of the inference process for these two cases is presented in Figure 2.

Case 1. A woman with benign adnexal mass, age 31, pre-menopausal, strong blood flow, CA125 is not raised (9 U/ml), no ascites, with unilocular ovarian cyst, smooth internal wall, mixed echogenicity (patient N 3). Acceptable levels of errors are: $\alpha = 0.05$, $\beta = 0.05$, i.e. assuming 95% confidence for both decisions (benign and malignant). From Table I the thresholds for $Sum(DC)$ are 12.8 for malignant and -12.8 for benign. To make the inference process more comprehensible it is presented in steps.

Step 1:Variable 'Smooth'=1, $DC(Smooth = 1) = -10$, $Sum(DC) = -10$. Thresholds are not reached. Conclusion: continue procedure.

Step 2:Variable 'Col4'=1, $DC(Col4 = 1) = 11.1$, $Sum(DC) = -10 + 11.1 = 1.1$. Thresholds are not reached. Conclusion: continue procedure.

Step 3:Variable 'Un'=1, $DC(Un = 1) = -10.3$, $Sum(DC) = 1.1 + (-10.3) = -9.2$. Thresholds are not reached. Conclusion: continue procedure.

Step 4:Variable 'C_CA125'=0, $DC(C_CA125 = 0) = -5.6$, $Sum(DC) = -9.2 + (-5.6) = -14.8$. $Sum(DC) < DC_{th}(A_2)$. Conclusion: Stop SNuP. Decision: benign form of tumour.

For this example four variables are enough to make a decision with 95% of confidence.

The next case contains missing values for some variables.

Case 2. This is a difficult case of ovarian cancer. It is for a woman aged 72, post-menopause, ascites, multilocular cyst, strong blood flow, smooth internal wall, no information on the level of CA125 (patient N 216). From Table I the thresholds for $Sum(DC)$ are 9.8 for malignant and -12.5 for benign ($\alpha = 0.05$, $\beta = 0.10$). DC values are taken from the Table IV.

Step 1:Variable 'Smooth'=1, $DC(Smooth = 1) = -10$, $Sum(DC) = -10$. Thresholds are not reached. Conclusion: continue procedure.

Step 2:Variable 'Col4'=1, $DC(Col4 = 1) = 11.1$, $Sum(DC) = -10 + 11.1 = 1.1$. Thresholds are not reached. Conclusion: continue procedure.

Step 3:Variable 'Un'=0, $DC(Un = 0) = 2.5$, $Sum(DC) = 1.1 + 2.5 = 3.6$. Thresholds are not reached. Conclusion: continue procedure.

Step 4:Variable 'C_CA125' value unknown, $Sum(DC) = 3.6$. Thresholds are not reached. Conclusion: continue procedure.

Step 5:Variable 'Ascites'=1, $DC(Ascites = 1) = 6.6$, $Sum(DC) = 3.6 + 6.6 = 10.2$. $Sum(DC) > DC_{th}(A_1)$. Conclusion: Stop SNuP. Decision: malignant form of tumour.

D. Performance evaluation

The performance of the model was assessed using ROC analysis and a 3-fold cross validation. A ratio of 2:1 between benign and malignant groups' sample sizes was taken from the initial data set. The results are presented in Table VI. The acceptable level of α and β errors was varied from 0.20 (20%) to 0.001 (0.1%).

As can be seen from the table, decreasing the *a priori* error from 0.15 to 0.001 augments the predictive accuracy of the model since the predictive positive and negative values are over 98%. However, the use of these strict boundaries decreases the overall accuracy, sensitivity and specificity levels from moderate 80% to less than 50%. There are two reasons for this situation. The first reason is linked to increased number of 'Group undefined' values. This happens when available information is not enough to make a decision with a specified level of confidence. The second reason of the poor classification when values of acceptable errors are small is in the nature of the variables in the model and their distribution inside the groups. When we raise the boundaries as high as ± 30 ($\alpha = 0.001$, $\beta = 0.001$) there are no definitive variables presence/absence of which we can make a decision and it is harder to reach the thresholds. From Table V the median number of variable required is 18 and in almost 50% of cases all 21 variables were involved in the procedure. Thus more variables with a low information measures and a low discriminative ability are involved into the SNuP which decrease the overall accuracy, sensitivity and specificity.

In Figure 3, the results of the classification based on SNuP are presented. 95% confidence boundaries of the ROC curve are calculated using a nonparametric approach described in [13]. In Figure 4, we compared the performance of the model based on SNuP with that of the Risk of Malignancy Index (RMI). The area under the curve (AUC) for SNuP was higher, but the difference is not statistically significant ($\mathrm{t} = 0.79$, $\mathrm{p} = 0.43$).

E. Comparison of the model performance with an expert assessment

It is almost impossible to reach 100% accuracy for the model except in situations where the number of cases is small and the variance of the input parameters is limited. The performance and usefulness of the model can be assessed by comparing the model to an expert opinion[1]. Previous findings [14] showed that no models can beat an expert sonologist.

[1]In our study the expert was Prof Dr Dirk Timmerman.

Assuming S is a gold standard value (0 - benign tumour, 1 - malignant tumour), M - model result (0 - benign tumour, 0.5 - undefined, 1 - malignant tumour), E - expert opinion (0 - benign tumour, 1 - malignant tumour). There are six possible situations in comparing the model and expert.

1) Model is correct. Expert is correct.
   $S = M \cup S = E$.
2) Model is correct. Expert is incorrect.
   $S = M \cup S \neq E$
3) Model is incorrect. Expert is correct.
   $S \neq M \cup M \neq 0.5 \cup S = E$
4) Model is incorrect. Expert is incorrect.
   $S \neq M \cup S \neq E$
5) Model's result is undefined. Expert is correct.
   $M = 0.5 \cup S = E$
6) Model's result is undefined. Expert is incorrect.
   $M = 0.5 \cup S \neq E$

In the above, conditions 1 and 4 represent a situation when the model and the expert agree. The rest of the conditions are more interesting. Conditions 2 and 6 are difficult cases for the diagnosis. Condition 2 is true when the expert misses something or reaches a conclusion based on wrong assumption. This may also be due to new knowledge obtained by the model. When the expert outperforms the model, condition 3 is true. Condition 5 is possible when there is enough information for the expert to come up with a correct conclusion but not enough for the model.

Table IX is a contingency table of all possible situations between the expert and the model. Modelling conditions included a 3-fold cross validation, 141 cases in the test set, α and β were equal 0.05, $p(A_i)$ was taken into account. The table shows that for the malignant form of tumour the agreement was reached in 93.2% cases (81.8% and 11.4%) whereas for the benign form it was only in two thirds - 67.1% (64.9% and 5.2%). The model was better than the expert in 7 (7.2%) cases of benign and 1 (2.3%) case of malignant tumours. The expert outperformed the model in 17 (17.6%) cases of benign neoplasm and 7 (15.9%) cases of ovarian cancer.

The results of the model-expert comparison show that an experienced sonologist is better than the model. However, the model is not supposed to be used on its own. It provides support for clinical decision making process as the clinician carries the responsibility for the diagnosis made.

F. Role CA125 in the model performance

Measurement of the tumour marker CA125 in the serum is very common in the diagnosis of ovarian cancer as well as during and after treatment. It has been shown that an abnormally raised level of CA125 is associated with malignancy [14]. However, many women with benign tumour or even healthy ones might have raised levels of CA125 which results in a high false alarm rate. On the other hand 10 to 20 percent of ovarian cancer patients have normal levels of CA125. Analysis of CA125 is quite expensive and the laboratory results need to be awaited. Therefore it is important to evaluate the role of CA125 in the preoperative differential diagnosis of adnexal masses and describe conditions in which the diagnosis will definitely benefit from CA125, or to identify conditions when CA125 can be omitted from diagnostic procedures.

To evaluate the role of CA125 in the model's performance we run SNuP with and without the variable CA125 during cross validation. The overall results are shown in the Table X. As can be seen from the table, CA125 does not significantly improve the classification performance, but brings more certainty to the decision making process by reducing the total number of undefined cases although the median number of variables stays the same.

Considering the last run of cross validation the more detailed results can be shown. For the test set of 141 cases the number of steps of SNuP was reduced in 21 cases, was increased in 11 cases and did not change for 109 patients. The informative measure of the variables is very important for classification results as it has an impact on the order of their use during SNuP. The list of variables sorted in descending order of information is: 'Smooth', 'Un', 'Asc', 'Cat CA125', 'Col4', 'Sol', 'Irreg', 'Pap', 'Bilat', 'Mul', 'MulSol', 'Meno', 'Col3', 'G.Glass', 'Sept', 'Lucent', 'UnSol', 'Shadows', 'Mixed', 'Haem', 'Low level'. CA125 is fourth in the list so it is not always taken into account during SNuP. CA125 was used for 80 patients out of 141. In 66 (82.5%) cases the absence of CA125 did not change the model's decision, and the rates per groups were 30 (78.9%) benign cases, 36 (85.7%) malignant cases. The exclusion of CA125 produced the worse results[2] in 9 (11.3%) cases and better results[3] in 5 (6.3%) cases.

IV. Conclusion and future work

In the paper we have presented the sequential non-uniform procedure which is an approach of modelling the preoperative diagnosis of adnexal masses. It can be considered as an extension of Wald's consecutive

[2]Worse results: from 'correct' to 'incorrect', from 'correct' to 'undefined', from 'undefined' to 'incorrect'

[3]Better results: from 'incorrect' to 'true', from 'incorrect' to 'undefined', from 'undefined' to 'correct'.

analysis where accumulation of diagnostic information makes it possible to use a minimal number of features in order to make a decision with any given level of confidence.

The advantages of the method include the use of a minimal number of variables, permissibility of cases with missed values, and interpretability of the model and results. The model can also incorporate prior knowledge of the distribution of the classes. From the end-user point of view (i.e. clinician) SNuP produces a model which is understandable and enables ranking of the input variables according to their discriminative abilities. One of the disadvantages of the SNuP is that only a binary classification can be made using the approach presented. However, a multiclass classification problem can be solved using a set of pair-wise models. Another limitation is the inability to consider the correlation between variables.

We applied the SNuP to the ovarian tumour data set. The task was to distinguish malignant and benign forms of this kind of neoplasm. The differential diagnosis of these conditions apart from clinical examination involves ultrasound methods, tumour markers, CT and MRI. It is important to find a trade-off between the cost and the number of the diagnostic procedures and the risk of missing a case when urgent surgical operation might be required.

The SNuP showed a high performance on a real data set during cross validation. The method is close to clinical thinking and can be used not only for research but also for educational purposes to demonstrate the inference process.

We plan to expand the diagnostic model with IOTA phase I data set [9]. The subject of the future work will be on multiclass generalisation [15], [16] and incorporation of continuous variables.

REFERENCES

[1] "Cancer incidence in five continents. Volume VIII," *IARC Sci Publ*, no. 155, pp. 1–781, 2002.

[2] "Survival of Cancer Patients in Europe: The EUROCARE-2 study," *IARC Sci Publ*, no. 151, pp. 1–572, 1999.

[3] I. Jacobs, D. Oram, J. Fairbanks, J. Turner, C. Frost, and J. Grudzinskas, "A risk of malignancy index incorporating ca 125, ultrasound and menopausal status for the accurate preoperative diagnosis of ovarian cancer," *Br J Obstet Gynaecol*, vol. 97, no. 10, pp. 922–929, Oct 1990.

[4] D. Timmerman, T. Bourne, A. Tailor, W. Collins, H. Verrelst, K. Vandenberghe, and I. Vergote, "A comparison of methods for preoperative discrimination between malignant and benign adnexal masses: the development of a new logistic regression model," *Am J Obstet Gynecol*, vol. 181, no. 1, pp. 57–65, Jul 1999.

[5] N. Aslam, S. Banerjee, J. V. Carr, M. Savvas, R. Hooper, and D. Jurkovic, "Prospective evaluation of logistic regression models for the diagnosis of ovarian cancer," *Obstet Gynecol*, vol. 96, no. 1, pp. 75–80, Jul 2000.

[6] D. Timmerman, H. Verrelst, T. Bourne, B. De Moor, W. Collins, I. Vergote, and J. Vandewalle, "Artificial neural network models for the preoperative discrimination between malignant and benign adnexal masses," *Ultrasound Obstet Gynecol*, vol. 13, no. 1, pp. 17–25, Jan 1999.

[7] C. Lu, T. Van Gestel, J. A. K. Suykens, S. Van Huffel, I. Vergote, and D. Timmerman, "Preoperative prediction of malignancy of ovarian tumors using least squares support vector machines," *Artif Intell Med*, vol. 28, no. 3, pp. 281–306, Jul 2003.

[8] E. Madu, V. Stalbovskaya, B. Hamadicharef, E. Ifeachor, S. Van Huffel, and D. Timmerman, "Preoperative ovarian cancer diagnosis using neuro-fuzzy approach," in *European Conference on Emergent Aspects on Clinical Data Analysis (EACDA 2005) September 28-30, 2005 - Pisa, Italy*, Sep 2005.

[9] D. Timmerman, A. C. Testa, T. Bourne, E. Ferrazzi, L. Ameye, M. L. Konstantinovic, B. Van Calster, W. P. Collins, I. Vergote, S. Van Huffel, and L. Valentin, "Logistic regression model to distinguish between the benign and malignant adnexal mass before surgery: a multicenter study by the international ovarian tumor analysis group," *J Clin Oncol*, vol. 23, no. 34, pp. 8794–8801, Dec 2005.

[10] E. Gubler, *Computational methods of analysis and recognition of pathological processes*. Medicina, Moscow, 1978.

[11] A. Wald, *Sequential Analysis*. Wiley, New York, 1947.

[12] T. Kailath and H. Poor, "Detection of stochastic processes," *IEEE Trans Inform Theory*, vol. 44, no. 6, pp. 2230–2259, Oct 1998.

[13] J. Tilbury, P. Van Eetvelt, J. Garibaldi, J. Curnow, and E. Ifeachor, "Reciever operator characteristic analysis for intelligent medical systems - a new approach for finding confidence intervals," *IEEE Trans Biomed Eng*, vol. 47, no. 7, pp. 952–963, 2000.

[14] D. Timmerman, "The use of mathematical models to evaluate pelvic masses; can they beat an expert operator?" *Best Pract Res Clin Obstet Gynaecol*, vol. 18, no. 1, pp. 91–104, Feb 2004.

[15] C. Baum and V. Veeravalli, "A sequential procedure for multihypothesis testing," *IEEE Trans Inform Theory*, vol. 40, no. 6, pp. 1994–2007, Nov 1994.

[16] V. P. Dragalin, A. G. Tartakovsky, and V. V. Veeravalli, "Multihypothesis sequential probability ratio tests - part i: Asymptotic optimality." *IEEE Trans Inform Theory*, vol. 45, no. 7, pp. 2448–2461, 1999.

LIST OF FIGURE CAPTIONS

Fig. 1. Examples of the SNuP inference process. Sum of DC_i on every step of procedure is indicated by arrow. Thresholds for A_1 and A_2 are denoted by bold solid lines. Hatched areas are indeterminate zones for A_1 and A_2 (malignant and benign groups). Case 1 and case 2 demonstrates SNuP with definitive variables (when one variables is enough to reach a threshold). Case 3 shows straight-forward classification of benign tumour in three steps. Case 4 is a difficult case of ovarian cancer.

Fig.2. Inference process of SNuP demonstrated on two cases of ovarian tumour. Sum of DC_i on every step of procedure is indicated by arrow. Thresholds for A_1 and A_2 are denoted by bold solid lines. Hatched areas are indeterminate zones for A_1 and A_2 (malignant and benign groups). For Case 2 the value of one variable is missed so the Sum(DC) for step 4 does not change.

Fig.3. ROC Curve of the SNuP model with 95% confidence boundaries).

Fig.4. ROC curve of the SNuP model and ROC curve of RMI applied to the test set.

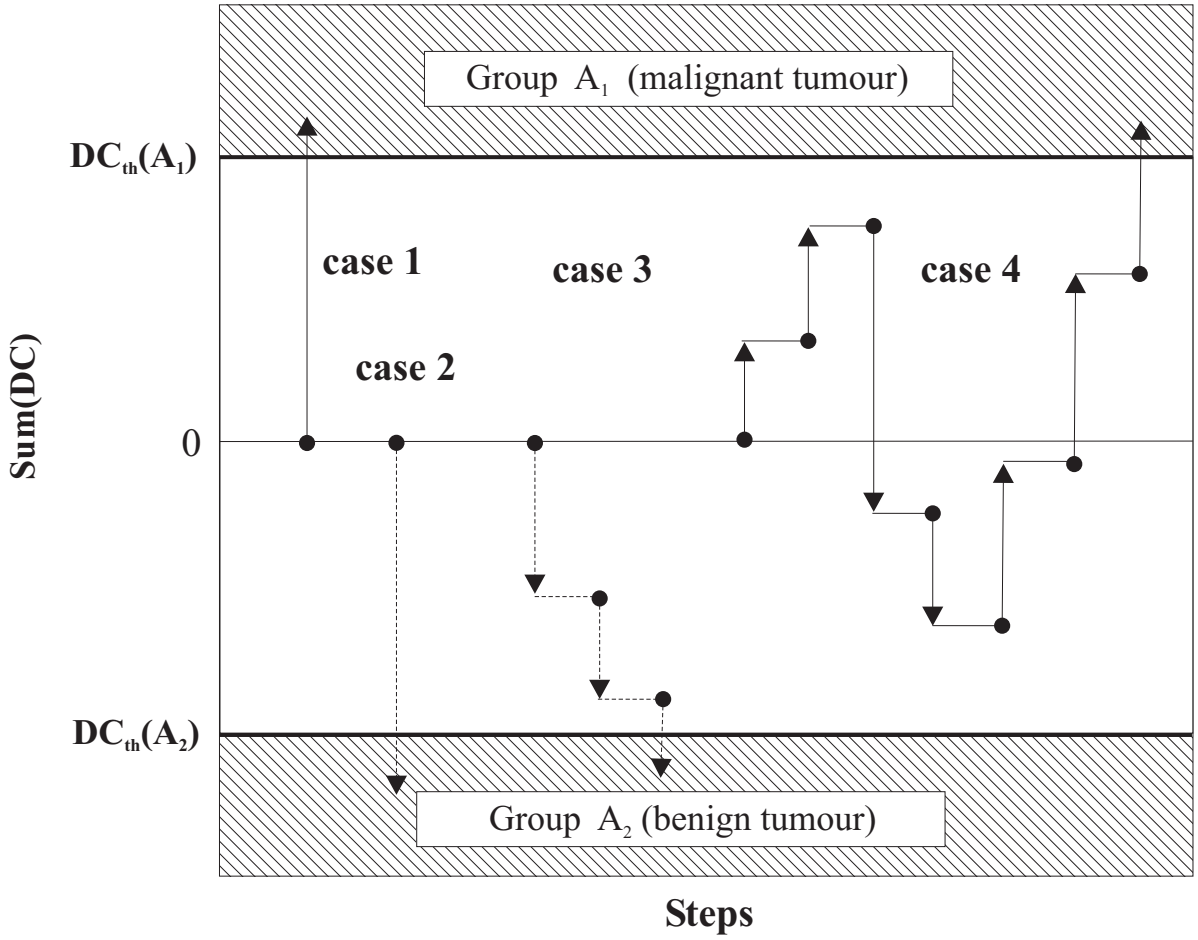

Group A1 (malignant tumour)
DCth(A1)
case 1
case 3
case 4
case 2
Sum(DC)
0
DCth(A2)
Group A2 (benign tumour)
Steps

Fig. 1.

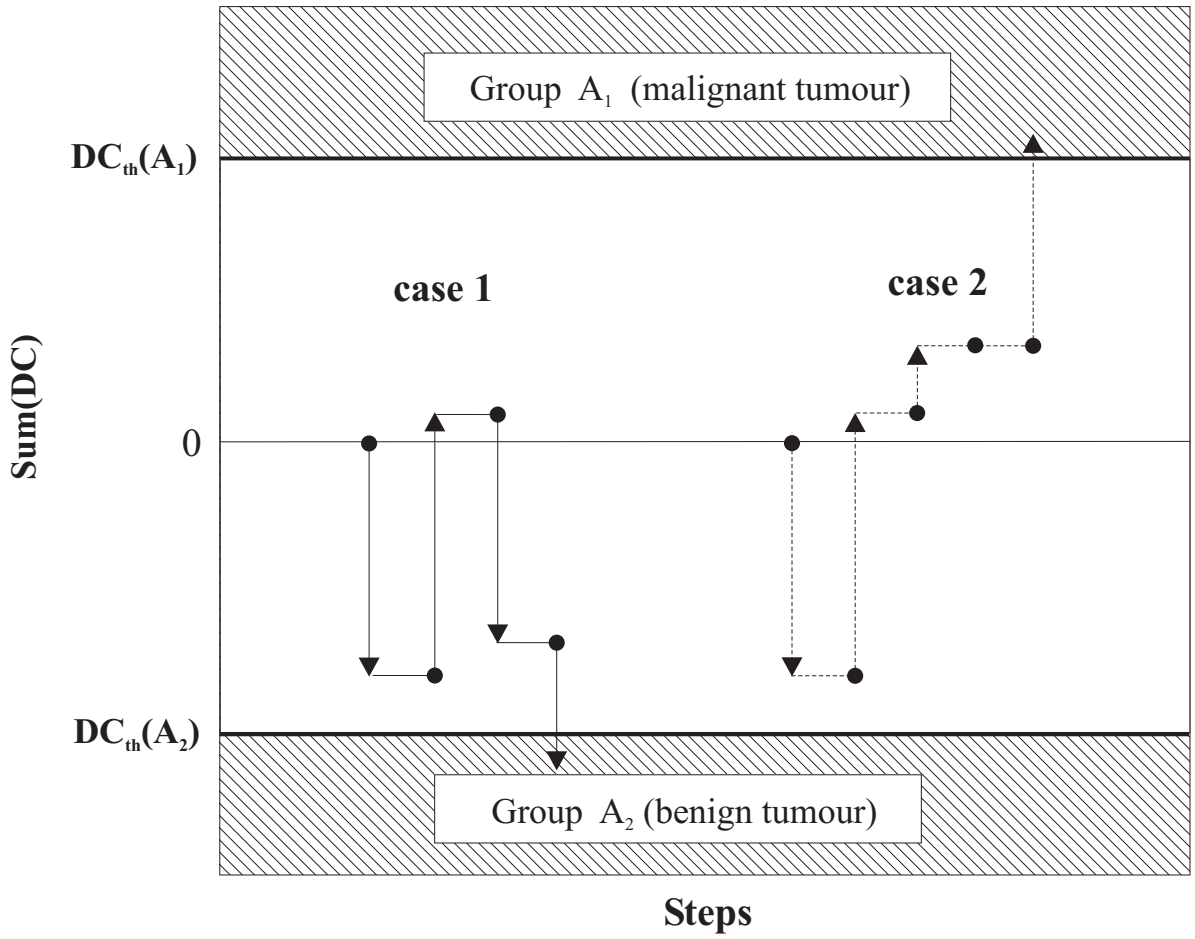

Group A1 (malignant tumour)
DCth(A1)
case 1
case 2
Sum(DC)
0
DCth(A2)
Group A2 (benign tumour)
Steps

Fig. 2.

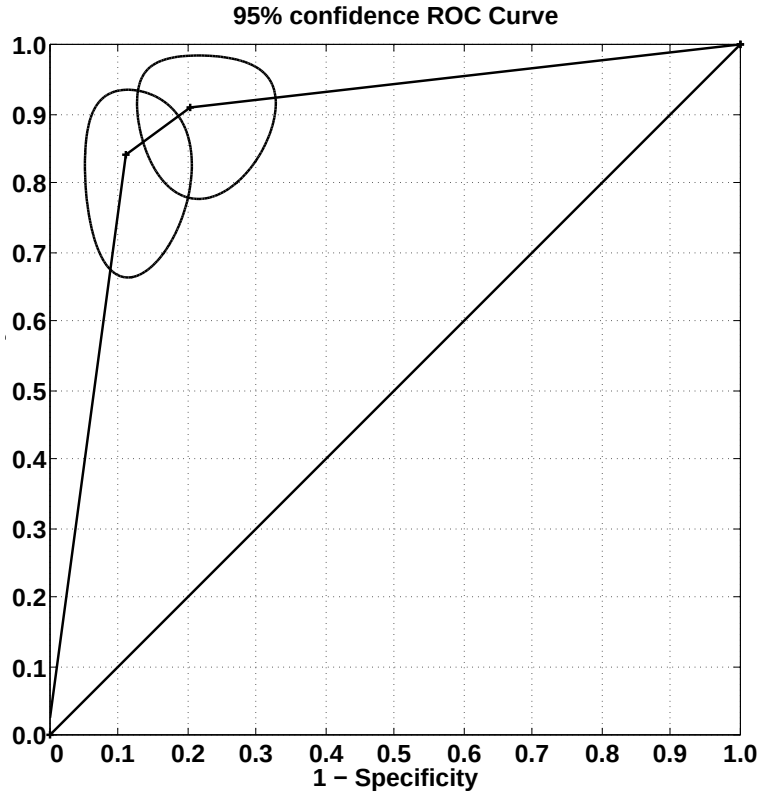

95% confidence ROC Curve
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
1 − Specificity

Fig. 3.

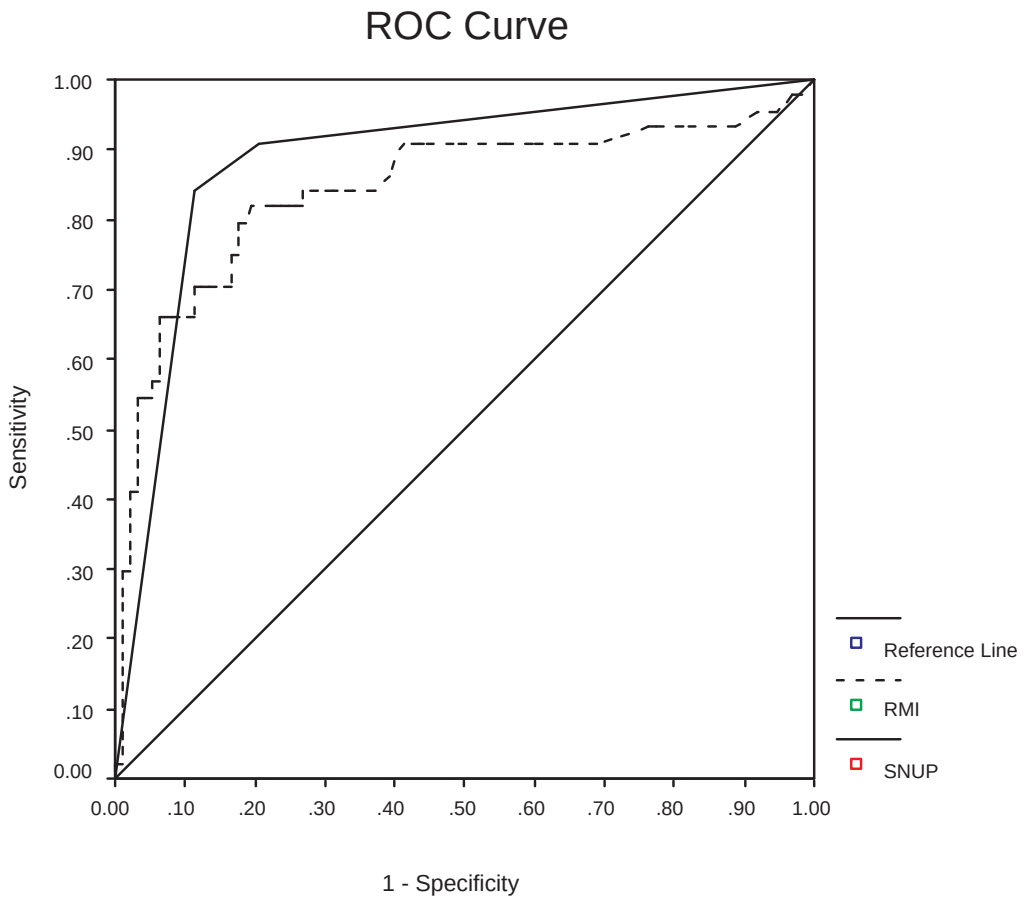

ROC Curve
Sensitivity
1.00
.90
.80
.70
.60
.50
.40
.30
.20
.10
0.00
0.00
.10
.20
.30
.40
.50
.60
.70
.80
.90
1.00
1 - Specificity
Reference Line
RMI
SNUP

Fig. 4.

TABLE I

THRESHOLDS FOR DIFFERENT LEVELS OF ACCEPTABLE ERRORS

α	β	$\frac{1-\alpha}{\beta}$	$\frac{\alpha}{1-\beta}$	$DC_{th}(A_1)$	$DC_{th}(A_2)$
0,10	0,10	9	0,111	9,5	-9,5
0,10	0,05	18	0,105	12,6	-9,8
0,05	0,10	9,5	0,056	9,8	-12,5
0,05	0,05	19	0,053	12,8	-12,8
0,01	0,10	9,9	0,011	10	-19,6
0,01	0,05	19,8	0,011	13	-19,6
0,001	0,10	10	0,001	10	-30

TABLE II

CONTINGENCY TABLE FOR ROC ANALYSIS

		Gold standard	
		Malignant	Benign
Model	Malignant	A	B
	Benign	C	D

TABLE III

DESCRIPTION OF THE DATA SET

Type of vars	Variable description (short name)	Type
Input vars	Number (Number)	nominal
	Age (Age)	cont.
	Menopause (Meno)	binary
	Amount of blood flow (Col score)	nominal
	Level of serum CA 125 (CA125)	cont.
	Pulsatility index (PI)	cont.
	Resistance index (RI)	cont.
	Peak systolic velocity (PSV)	cont.
	Time-averaged mean velocity (TAMX)	cont.
	Ascites (Asc)	binary
	Unilocular cyst (Un)	binary
	Unilocular solid (UnSol)	binary
	Multilocular cyst (Mul)	binary
	Multilocular solid (MulSol)	binary
	Solid tumour (Sol)	binary
	Bilateral mass (Bilat)	binary
	Smooth wall (Smooth)	binary
	Irregular wall (Irreg)	binary
	Papillations (Pap)	binary
	Septa $>$ 3 mm (Sept)	binary
	Acoustic shadows (Shadows)	binary
	Anechoic cystic content (Lucent)	binary
	Low level echogenicity (Low-level)	binary
	Mixed echogenicity (Mixed)	binary
	Ground glass cyst (G.Glass)	binary
	Hemorrhagic cyst (Haem)	binary
Output var	Pathology result (Path)	binary
Indices	Ultrasound score (Morph)	nominal
	Jacobs index (Jacobs)	nominal
	Risk of malignancy index (RMI)	cont.
Transformed	Rather strong blood flow (Col3)	binary
vars	Very strong blood flow (Col4)	binary
	CA125 $\leq$ 30 U/ml (C_CA125)	binary

TABLE IV

DIAGNOSTIC COEFFICIENTS AND INFORMATION MEASURES OF VARIABLES

No	Variable	Values	Malignant tumour			Benign tumour			$DC(x_{ij})$	$J(x_{ij})$	$J(x_i)$	J rank
			N	%	n	N	%	n				
1	Menopause	1	141	65,2	92	384	31,3	120	3,2	0,54	1,05	9
	(Meno)	0		34,8	49		68,7	264	-3,0	0,51		
2	Normal blood flow	1	141	34	48	384	15,4	59	3,4	0,32	0,42	14
	(Col3)	0		66	93		84,6	325	-1,1	0,10		
3	Strong blood flow	1	141	44	62	384	3,4	13	11,1	2,25	2,74	2
	(Col4)	0		56	79		96,6	371	-2,4	0,49		
4	Ascites	1	141	60,3	85	384	13,3	51	6,6	1,55	2,35	5
	(Asc)	0		39,7	56		86,7	333	-3,4	0,80		
5	Unilocular cyst	1	141	4,3	6	384	46,1	177	-10,3	2,15	2,67	3
	(Un)	0		95,7	135		53,9	207	2,5	0,52		
6	Unilocular solid	1	141	16,3	23	384	6,3	24	4,1	0,21	0,24	15
	(UnSol)	0		83,7	118		93,7	360	-0,5	0,03		
7	Multilocular cyst	1	141	5,7	8	384	28,6	110	-7,0	0,80	0,94	10
	(Mul)	0		94,3	133		71,4	274	1,2	0,14		
8	Multilocular solid	1	141	36,2	51	384	10,7	41	5,3	0,68	0,87	11
	(MulSol)	0		63,8	90		89,3	343	-1,5	0,19		
9	Solid tumour	1	141	37,6	53	384	8,3	32	6,6	0,97	1,22	8
	(Sol)	0		62,4	88		91,7	352	-1,7	0,25		
10	Bilateral mass	1	141	39	55	384	13,3	51	4,7	0,60	0,79	12
	(Bilat)	0		61	86		86,7	333	-1,5	0,19		
11	Smooth wall	1	141	5,7	8	384	56,8	218	-10,0	2,56	3,43	1
	(Smooth)	0		94,3	133		43,2	166	3,4	0,87		
12	Irregular wall	1	138	73,2	101	373	33,8	126	3,4	0,67	1,44	7
	(Irreg)	0		26,8	37		66,2	247	-3,9	0,77		
13	Papillations	1	141	53,9	76	384	12,2	47	6,5	1,36	1,94	6
	(Pap)	0		46,1	65		87,8	337	-2,8	0,58		
14	Septa $>$ 3 mm	1	141	31,2	44	384	13	50	3,8	0,35	0,44	13
	(Sept)	0		68,8	97		87	334	-1,0	0,09		
15	Acoustic shadows	1	141	5,7	8	384	12,2	47	-3,3	0,11	0,12	19
	(Shadows)	0		94,3	133		87,8	337	0,3	0,01		
16	Anechoic cystic content	1	141	28,4	40	384	43,5	167	-1,9	0,14	0,22	17
	(Lucent)	0		71,6	101		56,5	217	1,0	0,08		
17	Low level echogenicity	1	141	20,6	29	384	11,7	45	2,5	0,11	0,13	18
	(Low-level)	0		79,4	112		88,3	339	-0,5	0,02		
18	Mixed echogenicity	1	141	13,5	19	384	20,3	78	-1,8	0,06	0,07	21
	(Mixed)	0		86,5	122		79,7	306	0,4	0,01		
19	Ground glass cyst	1	141	8,5	12	384	19,8	76	-3,7	0,21	0,24	16
	(G.Glass)	0		91,5	129		80,2	308	0,6	0,03		
20	Hemorrhagic cyst	1	141	0,7	1	384	3,6	14	-7,1	0,10	0,10	20
	(Haem)	0		99,3	140		96,4	370	0,1	0,00		
21	CA125 $\leq$ 30	1	137	80,3	110	295	29,2	86	4,4	1,12	2,55	4
	(C_CA125)	0		19,7	27		70,8	209	-5,6	1,43		

TABLE V

NUMBER OF VARIABLES REQUIRED FOR CLASSIFICATION USING SNUP

Level of errors, α and β	Number of variables		
	Min	Median	Max
$\alpha = 0.20$, $\beta = 0.20$	1 (42.6%)	2 (22.0%)	21 (2.1%)
$\alpha = 0.15$, $\beta = 0.15$	1 (42.6%)	2 (22.0%)	21 (2.8%)
$\alpha = 0.10$, $\beta = 0.10$	2 (60.3%)	2 (60.3%)	21 (7.1%)
$\alpha = 0.05$, $\beta = 0.05$	2 (38.3%)	3 (19.9%)	21 (13.5%)
$\alpha = 0.01$, $\beta = 0.01$	4 (17.7%)	6 (25.5%)	21 (24.1%)
$\alpha = 0.001$, $\beta = 0.001$	5 (17.7%)	18 (0.7%)	21 (46.8%)
$\alpha = 0.20$, $\beta = 0.15$	1 (44.0%)	2 (18.4%)	21 (2.8%)
$\alpha = 0.15$, $\beta = 0.15$	1 (44.0%)	2 (18.4%)	21 (4.3%)
$\alpha = 0.10$, $\beta = 0.15$	1 (44.0%)	2 (18.4%)	21 (4.3%)
$\alpha = 0.05$, $\beta = 0.15$	1 (44.0%)	3 (7.8%)	21 (9.9%)
$\alpha = 0.01$, $\beta = 0.15$	1 (44.0%)	3 (7.8%)	21 (15.6%)
$\alpha = 0.001$, $\beta = 0.15$	1 (44.0%)	4 (6.4%)	21 (22.7%)

TABLE VI

ROC ANALYSIS OF DIAGNOSTIC MODEL INCLUDING ALL AVAILABLE VARIABLES

Receiver operating characteristic	Level of errors					
	0.20	0.15	0.10	0.05	0.01	0.001
Considering undefined values as incorrect classification						
Acc, %	82.3 ± 2.3	82.3 ± 2.3	81.3 ± 2.3	76.6 ± 2.5	68.1 ± 2.8	50.4 ± 3.0
Se, %	74.2 ± 2.6	77.3 ± 2.5	75.8 ± 2.6	73.5 ± 2.6	62.9 ± 2.9	35.6 ± 2.9
Sp, %	85.9 ± 2.1	84.5 ± 2.2	83.8 ± 2.2	78.0 ± 2.5	70.4 ± 2.7	57.0 ± 2.9
PPV, %	72.8 ± 2.7	72.4 ± 2.7	76.4 ± 2.5	79.4 ± 2.4	91.2 ± 1.7	97.0 ± 1.0
PNV, %	88.1 ± 1.9	89.5 ± 1.8	90.1 ± 1.8	93.5 ± 1.5	96.9 ± 1.0	98.3 ± 0.8
Values are calculated from the second point of the ROC curve						
Acc, %	82.5 ± 2.3	83.9 ± 2.2	86.3 ± 2.0	90.1 ± 1.8	96.5 ± 1.1	99.1 ± 0.6
Se, %	74.2 ± 2.6	78.0 ± 2.5	79.5 ± 2.4	87.9 ± 1.9	94.7 ± 1.3	97.7 ± 0.9
Sp, %	86.3 ± 2.1	86.6 ± 2.0	89.3 ± 1.8	91.1 ± 1.7	97.3 ± 1.0	99.7 ± 0.3
PPV, %	72.8 ± 2.7	72.4 ± 2.7	76.4 ± 2.5	79.4 ± 2.4	91.2 ± 1.7	97.0 ± 1.0
PNV, %	88.1 ± 1.9	89.5 ± 1.8	90.1 ± 1.8	93.5 ± 1.5	96.9 ± 1.0	98.3 ± 0.8
Undefined (M), rate (min-max)	0 - 0	0 - 1	1 - 2	4 - 10	12 - 18	20 - 32
Undefined (B), rate (min-max)	0 - 1	0 - 5	5 - 6	10 - 17	20 - 31	36 - 45
Undefined (Total), rate (min-max)	0 - 1	0 - 5	7 - 7	15 - 21	32 - 45	63 - 75

TABLE VII

CONFUSION MATRIX FOR THE TEST DATA SET

		Gold standard	
		Malignant	Benign
Model	Malignant	37	11
	Undefined	3	9
	Benign	4	77

TABLE VIII

AREA UNDER THE ROC CURVE

Model	AUC	Std. Error	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
SNuP	0.887	0.033	0.823	0.951
RMI	0.845	0.042	0.763	0.926

TABLE IX

PERFORMANCE OF THE MODEL (M) VS EXPERT (E) BY GROUPS

Conditions	Number of cases	
	Benign	Malignant
1. M correct. E correct.	63 (64.9%)	36 (81.8%)
2. M correct. E incorrect.	7 (7.2%)	1 (2.3%)
3. M incorrect. E correct.	5 (5.2%)	3 (6.8%)
4. M incorrect. E incorrect	5 (5.2%)	0 (0.0%)
5. M undefined. E correct.	12 (12.4%)	4 (9.1%)
6. M undefined. E incorrect.	5 (5.2%)	0 (0.0%)

TABLE X

PERFORMANCE OF THE MODEL WITH AND WITHOUT CA125 ($\alpha = 0.05$, $\beta = 0.05$)

	With CA125	Without CA125
Acc, %	88.6 $\pm$ 1.8	88.4 $\pm$ 0.6
Se, %	85.5 $\pm$ 6.6	83.8 $\pm$ 10.0
Sp, %	89.7 $\pm$ 3.9	90.5 $\pm$ 5.2
PPV, %	79.4 $\pm$ 4.7	81.2 $\pm$ 6.1
PNV, %	93.5 $\pm$ 2.1	92.7 $\pm$ 3.9
Undefined (M), rate (min-max)	4 - 10	6 - 8
Undefined (B), rate (min-max)	10 - 17	13 - 18
Undefined (Total), rate (min-max)	15 - 21	21 - 24
Median number of variables	4	4