Development of a clinical decision support system using genetic algorithms and Bayesian classification for improving the personalised management of women attending a colposcopy room

Panagiotis Bountris[1] ✉, *Elena Topaka*[1], *Abraham Pouliakis*[2], *Maria Haritou*[3], *Petros Karakitsos*[2], *Dimitrios Koutsouris*[1]

[1]*Biomedical Engineering Laboratory, School of Electrical and Computer Engineering, National Technical University of Athens, Athens 15773, Greece*
[2]*Department of Cytopathology, 'Attikon' University Hospital, National and Kapodistrian University of Athens, Athens 12462, Greece*
[3]*Institute of Communication and Computer Systems, Athens 10682, Greece*
✉ *E-mail: pbountris@biomed.ntua.gr*

Published in Healthcare Technology Letters; Received on 19th November 2015; Revised on 20th April 2016; Accepted on 29th April 2016

Cervical cancer (CxCa) is often the result of underestimated abnormalities in the test Papanicolaou (Pap test). The recent advances in the study of the human papillomavirus (HPV) infection (the necessary cause for CxCa development) have guided clinical practice to add HPV related tests alongside the Pap test. In this way, today, HPV DNA testing is well accepted as an ancillary test and it is used for the triage of women with abnormal findings in cytology. However, these tests are either highly sensitive or highly specific, and therefore none of them provides an optimal solution. In this Letter, a clinical decision support system based on a hybrid genetic algorithm – Bayesian classification framework is presented, which combines the results of the Pap test with those of the HPV DNA test in order to exploit the benefits of each method and produce more accurate outcomes. Compared with the medical tests and their combinations (co-testing), the proposed system produced the best receiver operating characteristic curve and the most balanced combination among sensitivity and specificity in detecting high-grade cervical intraepithelial neoplasia and CxCa (CIN2+). This system may support decision-making for the improved management of women who attend a colposcopy room following a positive test result.

1. Introduction: Worldwide, cervical cancer (CxCa) is ranked as the third most common cancer type and the fourth leading cause of cancer death in females. These cases and deaths occur mostly in developing countries (more than 85%) due to the lack of organised screening programmes that allow the detection of precancerous lesions and CxCa in early stage. In reality, organised CxCa screening is applied in rather limited countries, and in many developed countries it is a matter of opportunistic control. Even in well-organised programmes and despite the advances of screening, CxCa still remains a serious problem of public health in the developed world as well, due to the high percentage of detection failure [1]. Cervical intraepithelial neoplasias (CINs), a precancerous condition, is very common, almost one in ten women will have such abnormalities at primary screening.

The well-known Papanicolaou CxCa screening (Pap test) has reduced CxCa rates worldwide dramatically. Nowadays, prevention of CxCa is based on frequent and repetitive Pap tests, followed by colposcopical examinations and if required histological examination on biopsy material (i.e. when the test Pap or colposcopy is abnormal). Unfortunately, the evaluation of cervical cytology smears is a difficult task and may be accomplished only by very well trained medical personnel (cytopathologists); therefore, interpretation is influenced by subjective factors and prone to diagnostic error.

CxCa is almost caused by human papillomavirus (HPV), being the commonest sexually transmitted infection. There are more than 100 types of HPV infecting humans; however, only 14 are highly oncogenic and may cause CxCa. In addition, presence of HPV does not always lead to disease, as the infection may regress due to human immune system. The developments in understanding HPV infection and the cervical neoplasia natural history have resulted in the addition of the HPV DNA test along with the Pap test [2]. HPV DNA testing is nowadays used as ancillary test to test Pap and has been proposed its utilisation for primary screening. Owing to this, recently, there are many developed countries that included HPV DNA test in their official CxCa screening guidelines.

There are numerous studies attempting to analyse the role of HPV DNA test and compare it with Pap test [3–5]. A detailed analysis of these studies shows that the performance of the two tests differ significantly. They are either highly sensitive or highly specific, unfortunately not both at the same time; thus, today, there is no perfect test. Additionally, studies results are affected by the disease incidence and the prevalence and the HPV infection in the studied population; therefore, the application of a single test, despite it offers a level of protection, cannot determine reliably the real risk of individual women participating in CxCa screening programmes.

The meta-analysis of published studies [3–5] shows that the sensitivity (SN) of Pap test when combined with the HPV DNA test is higher than the SN of each individual method. Thus, the two tests complement effectively each other. In contrast, Pap test specificity (SP) when combined with HPV DNA test was lower than the specificities of each of the two methods. Regarding the positive predictive value (PPV), there are equivocal findings: some studies report PPV values similar for each method separately as well as when combined. Other studies report smaller values of PPV for the two methods combination. As expected, the negative predictive value (NPV) of the combination of the two tests is high, actually several studies report values of almost 100%.

Today, despite the advances in CxCa screening, there is no consensus for the optimal management of women when there are abnormal test results. A percentage of women that have atypical squamous cells of undetermined significance (ASCUS) or low-grade squamous intraepithelial lesion (LSIL) in cytology may in reality have high-grade CIN [i.e. CIN grade II (CIN2) or III

Table 1 Cases distribution

Histology	Pap test result					Total
	WNL	*ASCUS*	*LSIL*	*HSIL*	*CxCa*	
clinically negative	196	0	0	0	0	196 (26.5%)
negative	35	60	22	5	0	122 (16.5%)
CIN1	31	66	142	22	0	261 (35.3%)
CIN2/3	3	13	27	93	0	136 (18.4%)
CxCa	0	1	2	7	15	25 (3.4%)
total	265	140	193	127	15	740

(CIN3)]. These women are at very high risk to develop CxCa. Therefore, CIN2 is the decision threshold beyond which a case surgically treated. In contrast, women with CIN grade I (CIN1) are monitored strictly. Moreover, it is not infrequent that women with high-grade lesion (HSIL) in cytology; may in reality have CIN1 or even a normal histology. Therefore, the management options of ASCUS or LSIL Pap smears that are widely accepted are: either immediate colposcopy or cytological surveillance by frequent Pap tests.

The policy to immediate referral to colposcopy can easily result in colposcopy clinics overload, as well as, in over-intervention and/or over-treatment in the case of subtle colposcopic findings. Thus, women are exposed to the physical and psychological sequelae of unnecessary treatment, which in nulliparous women incorporates the danger of pre-term delivery. On the other hand, repeated Pap tests incorporate the risk of missing HSILs, increases non-conformance rates, increases the CxCa organised screening programmes (OSPs) cost, increases the social and psychological burden of women, and eventually the credibility of OSPs' becomes questionable.

Therefore, it is essential to correctly identify women who are at real risk of developing CxCa, and simultaneously to reduce unnecessary colposcopies and repeated smears. Clearly, there are required effort for a screening method presenting simultaneously high SN and high SP for the threshold of CIN2 or worst (CIN2+). On the basis of this requirement, in this Letter we present a clinical decision support system (CDSS) based on a hybrid genetic algorithms (GAs) – Bayesian classification framework, which presents balanced SN and SP in detecting CIN2+, by intelligently combining the results of Pap test and HPV DNA test.

2. Clinical data: For the purposes of the study, there were collected anonymised data from enrolled women. The study was headed by the Department of Cytopathology of the Medical School of Athens' University (Attikon University Hospital). The Institution's Ethical Review Board had approved the study and all procedures, and the participating women signed an informed patient consent form, permitting anonymous use of their data for the research.

The biological material was collected in ThinPrep® vials (liquid-based cytology). Participating women were referred to colposcopy because: (a) they had an abnormal Pap test or (b) they volunteered to participate in the study and accepted a colposcopical examination and the application of the test on their biological material, even if they had a normal Pap test (e.g. HPV positive women). In women that had a negative Pap test followed by negative colposcopy, there was no biopsy taken and those cases are considered as clinically negative.

The collected data had HPV DNA test and Pap test results, histological examination outcome (when available), patient demographic, and identification details. These data were stored in a database implemented for the study purpose. About, 740 cases with full tests' series were extracted from the database and used as anonymised data for further analysis (Table 1). Each data series (woman case) included: Pap test and the HPV DNA test results. The later was performed using the CLART® HPV 2 test (GENOMICA, Spain) that simultaneously detects 35 different high-risk (HR)- or low-risk (LR) HPV subtypes [2].

The cytological result for each woman was interpreted according to the Bethesda classification system formulated according to the Bethesda classification (TBS2001 system) [6] which assigns each case to the following categories (ranked): (a) within normal limits (WNLs), (b) ASCUS, (c) LSIL lesion, (d) HSIL lesion, and (e) squamous cell carcinoma (SCC) or adenocarcinoma (Adeno-Ca). If Pap test revealed ASCUS or higher cytological categories (ASCUS+) and there was a visible lesion during colposcopy, a cervical biopsy was performed. For those cases having a histological

Table 2 Description of the available features

Variable name, feature	Description	Value range
Pap test	the result of the cytological examination expressed according to Bethesda system	1: WNL 2: ASCUS 3: LSIL 4: HSIL 5: Cancer
HPV DNA arrays for the subtypes: 6, 11, 16, 18, 26, 31, 33, 35, 39, 40, 42, 43, 44, 45, 51, 52, 53, 54, 56, 58, 59, 61, 62, 66, 68, 70, 71, 72, 73, 81, 82, 83, 84, 85, 89	the existence of individual subtypes according to the HPV DNA examination	0 if the specific subtype is not found 1 if the specific subtype is found
HPV DNA (positive or negative)	positive if one or more subtypes found by the HPV DNA test	0 if none subtype found (negative test) 1 if one or more subtypes found (positive test)
HR-HPV DNA HR genotypes: 16, 18, 26, 31, 33, 35, 39, 45, 51, 52, 53, 56, 58, 59, 66, 68, 70, 73, 82, 85	positive if one or more HR subtypes found by the HPV DNA test	0 if none of the HR types found 1 if at least one of the HR types found
Very high-risk (VHR) HPV VHR genotypes: 16, 18, 31, 33, 45	positive if one or more of the very HR subtypes found	0 if none of the very HR types found 1 if at least one of the very HR types found
LR-HPV DNA LR genotypes: 6, 11, 40, 42, 43, 44, 54, 61, 62, 71, 72, 81, 83, 84, 89	positive if one or more LR subtypes found by the HPV DNA test	0 if none of the LR types found 1 if at least one of the LR types found

doi: 10.1049/htl.2015.0051

outcome, the histological diagnosis was used as the golden standard. No biopsies were obtained in clinically negative cases, which are defined as cases with negative cytology and negative colposcopy, as in these cases it is not ethical to obtain biological samples for histological examination.

For the formulation of the histological diagnosis (the golden standard), it was used the CIN grading system. The categories, ranked according to severity, are: (a) without evidence of malignancy (negative), (b) CIN1, (c) CIN2 or CIN3, and (d) SCC or Adeno-Ca.

Summarising, for each of the 740 women, a feature set consisting of 40 indicators, which derived from the applied tests, was created. The result of the cytological examination has been used according to the TBS2001. The results of the HPV DNA test examination were expressed as 35 individual variables (either positive or negative), one for each HPV DNA genotype. Additionally, other variables expressing the HPV DNA test results were added, for instance, the existence of HR or LR types was expressed as either positive or negative. The 40 used variables/features collected for each case are presented in Table 2.

3. Hybrid GA – Bayesian classification feature selection framework

3.1. Overview: Our purpose is to create a classification system that will effectively combine the results of the Pap test and the HPV DNA test. However, our focus is not on the classification's accuracy per se, but rather the creation of a system that yields the most balanced results in terms of SN and SP, as this is the requirement of the clinical context. To achieve this balanced outcome, a feature subset must be found, which, when fed to a classifier, would satisfy this goal. Therefore, the problem we face is more of a feature subset selection problem rather than one of classification.

In the present Letter, a hybrid GA – Bayesian classification [GA-Naïve-Bayes (NB)] framework is adopted in order to perform the task of feature subset selection. For this purpose, a GA and a NB classifier are intergraded in a wrapper manner.

3.2. Feature subset selection using GAs: GAs appear to be a robust technique in the field of feature selection due to their heuristic nature and their prominent ability in solving optimisation problems [7, 8]. GAs, as a search strategy 'wrapped' around a classifier, have been extensively researched and applied to many feature selection problems [9–12]. In wrapper type approaches, feature subset selection is performed by 'wrapping' around a learning method: the usefulness of a feature subset is directly judged by the estimated performance of a trained classifier. That is, for each feature combination, the performance of a classifier is estimated and the combination resulting to the best performance is selected as the best subset.

In GAs, a candidate solution (in this case a possible feature subset) is represented as a chromosome made up of genes (features) [7–12]. Each chromosome is defined as an individual of a population. There are three stages in a standard GA: initialisation, selection, and reproduction [7–12].

A GA starts by randomly creating an initial population (initialisation stage) of *Np* individuals (candidate feature subsets). Afterwards, each individual is evaluated using a fitness function and assigned a fitness value: typically, a classification metric is calculated to measure the performance of a candidate feature subset by using it as input to a classifier. In the selection stage, the fitter individuals, according to their fitness value, are selected to form a new population (the next generation) through reproduction. Reproduction includes operators such as cross-over, mutation, and elitism [7–12], which are designed in a way so that the 'offspring' will inherit properties of the parents. The new formed population is used in the next iteration of the algorithm and the whole process is repeated for a number of *Ng* generations. The GA terminates when a specific criterion or a combination of criteria is satisfied [7–12]. In the end, GA returns the solution (feature subset) with the highest fitness value.

3.3. Multivariate multinomial NB classification [13, 14]: NB is a classification algorithm that applies the Bayes' theorem with the 'naive' assumption of class-conditional independence between every pair of features. Using the training data, the method estimates the parameters of a probability distribution, assuming features are conditionally independent given the class. Then, NB classifies a new sample by estimating the posterior probabilities of that sample belonging to each class according to Bayes rule, and by assigning the sample to the class yielding the maximum posterior probability. To estimate the posterior probability for each class, the classes' prior probabilities and the features' distributions must be estimated. A class' prior may be calculated by estimating the class probability from the training set. Regarding the distributions of categorical features such as the ones encountered in this problem, the most appropriate distribution is the multivariate multinomial distribution.

Important reasons behind the choice of the NB as the classifier of the presented framework are as follows:

(a) NB is a non-parametric classifier, meaning that it is not required to perform parameter selection inside the GA. This fact is very important since a parameter selection process would increase the complexity of the search problem.
(b) NB is one of the fastest classifiers (both for training and classifying new data), which is very important for a heuristic search algorithm.
(c) Moreover, the fact that the NB calculates the posterior probabilities of each class is of great significance to the problem in hand, because they can be used to estimate the risk assessment odd of a woman having CIN2+.

3.4. Framework implementation: The detailed implementation (Fig. 1) is achieved as follows. At first, an initial population of 1200 chromosomes (candidate feature subsets) is randomly created. To represent the selected feature subsets, integer encoding is used [7]: a chromosome is set as a vector consisting of a set of integers that represent the index numbers of the selected features. That is, each gene takes an integer value (1–40, where 40 is the number of all available features), which represents an index to a specific feature.

Following, the fitness evaluation of each chromosome is taking place by employing an NB classifier and applying five-fold cross-validation (Fig. 1). The mean SN and SP over all folds are computed in order to create the fitness value of each chromosome.

As presented, in this problem the accuracy is not as important as the SP and the SN in terms of diagnostic value. The performance of a diagnostic test or a classifier can be quantified by the measures of diagnostic accuracy such as SN and SP, predictive values, likelihood ratios, the area under the receiver operating characteristic (ROC) curve, Youden's index (YI) and diagnostic odds ratio. Different diagnostic metrics relate to the different aspects of diagnostic procedure: some metrics are used to assess the discriminative property of the test, and others are used to assess its predictive ability. Since the objective of the Letter is a balanced combination of SN and SP, we chose to use as a fitness function the Youden's *J* statistic (also called YI) (1), which is a single statistic index that captures the performance of a diagnostic test by combining SN and SP. It is considered as a global measure of a diagnostic test's performance and it is used for comparison between tests. The GA, by trying to maximise the YI, is in fact attempting to find that combination of features which leads to the most balanced SP

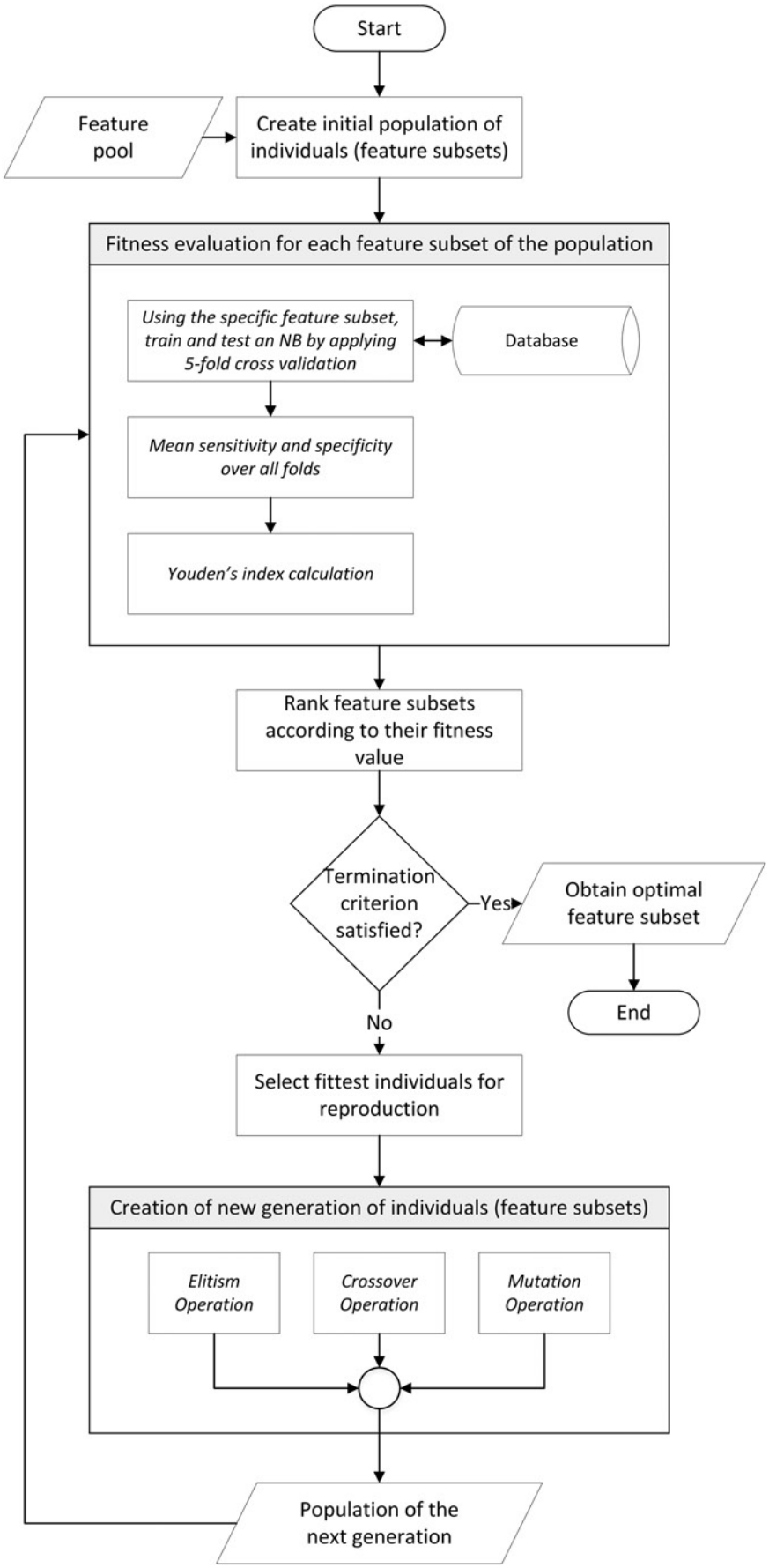

Fig. 1 *GA flowchart*

doi: 10.1049/htl.2015.0051

and SN

$$J = \text{sensitivity} + \text{specificity} - 1 \quad (1)$$

After a fitness value is assigned to each chromosome of the current population, the fitter chromosomes are chosen for reproduction using tournament selection (of size four).

To create the population of the next generation, three basic genetic operators are used (Fig. 1): elitism, cross-over, and mutation [7–12]. All the genetic operators are modified so as to be used with integer encoded chromosomes. Initially, an elitism operator is performed with an elite count of 120 (10% of the population): the 120 chromosomes of the current population with the highest fitness values are copied as they are, without any modifications, directly to the next population. For the cross-over operation, uniform cross-over is used with exchange probability of 0.5. The cross-over rate is set to 0.7 and it specifies the fraction of each population, other than elite children, that is created by the cross-over operator. With this value of the cross-over rate, in each generation there are 120 elite children, 756 cross-over children, and 324 mutation children. Subsequently, the mutation procedure is carried out using uniform mutation. Uniform mutation is applied at the genes of the chromosomes selected for mutation, with a probability rate of change equal to 0.2. Finally, the termination criterion that is used is a combination of two criteria. The GA terminates: (a) after 50 generations or (b) after 12 stall generations (i.e. the fitness value remains the same for 12 generations).

In integer encoding, the length of a chromosome is equal to the number of features that are selected to form the candidate feature subset. Thus, aiming to find the optimal feature subset, we executed the GA algorithm for every different chromosome length, i.e. for subsets with 2–39 features. Eventually, GA returns the best feature subset for each different chromosome length. Therefore, we can determine the optimal chromosome length which produces the feature subset with the highest fitness value for the specific problem.

4. Decision support system's architecture: An NB classifier in conjunction with the optimal feature subset have been integrated together to create a CDSS. The schematic diagram of the proposed CDSS is presented in Fig. 2. As depicted in Fig. 2, the examinations' results of each woman are used as inputs to the CDSS. First, the medical information is transformed to data appropriate for processing by the NB classifier, i.e. the complete feature set (40 features) is created. From this feature set, we

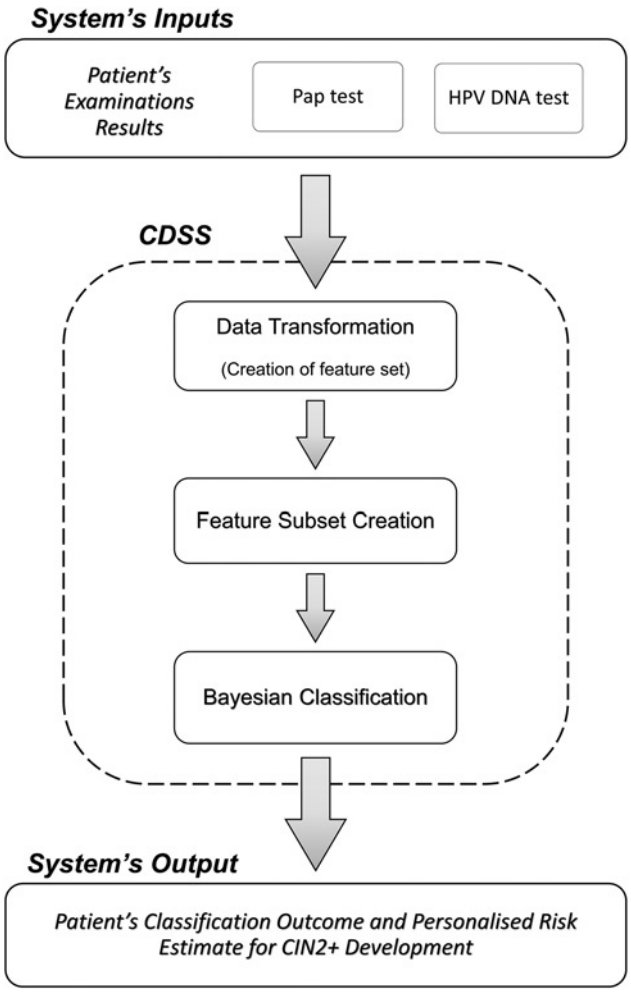

Fig. 2 *CDSS's architecture*

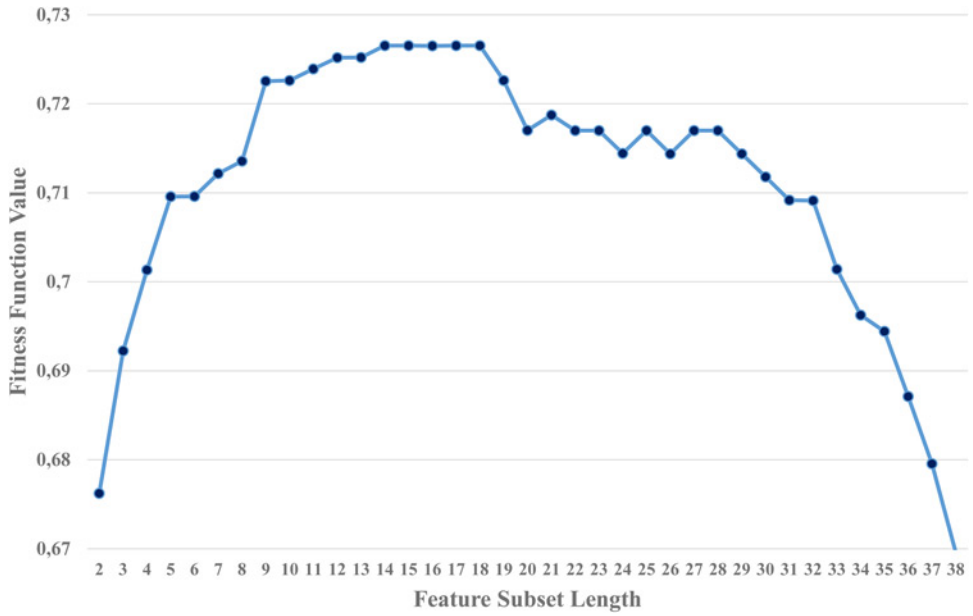

Fig. 3 *Fitness values of the best feature subsets for each different feature subset length*

select the features corresponding to the optimal feature subset, which emerged by the GA-NB framework, in order to create the input vector of the NB classifier. Following, the input vector is promoted to the trained NB classifier and the latter provides the classification outcome along with the posterior probabilities of each class, i.e. the probability of CIN1− and the probability of CIN2+. In this way, the CDSS provides risk estimates of CIN2+ development for each individual woman. The CDSS has been developed using the MATLAB® platform and was integrated into the previous constructed and presented web-based information system *CxCaDSS* [15].

5. Results: The available dataset (Table 1) was divided into two sets using stratified sampling: the two thirds (2/3) of the data (494 instances) was used for the feature subset selection process and the training of the CDSS, while the rest one third (1/3) of the data (246 instances) was used as a test set for the evaluation of the developed CDSS. Stratification was performed in order to ensure that each class is represented with approximately equal proportions in both subsets.

At first, utilising the first set (494 instances), we performed the GA-NB framework presented above in order to find the optimal feature subset. As described previously, we executed the GA algorithm for every different chromosome length, i.e. for subsets with 2–39 features, in order to determine the optimal chromosome length and eventually detect the optimal feature subset.

Fig. 3 depicts the fitness values of the best feature subsets for each different chromosome length. As it is shown in Fig. 3, the value of the fitness function is increasing while the number of features used is increased, and reaches its maximum value for a feature subset consisting of 14 features. As feature subsets, with more than 14 features, are created, the performance is getting worst, which means that the addition of more features does not provide useful information to the classifier. This fact means that only 14 from the 40 features contain important information and are complementing each other in order to obtain a classifier with balanced SN and SP, whereas the rest may be considered as irrelevant or redundant to the problem. These 14 features are as follows: Pap test, HPV DNA (positive or negative), VHR HPV, HPV 6, HPV 33, HPV 44, HPV 45, HPV 51, HPV61, HPV 62, HPV 66, HPV 68, HPV 82, and HPV 84.

Table 3 Performance of Pap test, HPV DNA test and the CDSS in detecting CIN2+

	SN, %	SP, %	PPV, %	NPV, %	YI	AUC
Pap test (cut-off ASCUS+)	98.1	47.2	34.2	98.9	0.45	0.73
Pap test (cut-off LSIL+)	90.7	72.0	47.6	96.5	0.63	0.81
Pap test (cut-off HSIL+)	70.4	95.9	82.6	92.0	0.66	0.83
HPV DNA test	94.4	66.8	44.3	97.7	0.61	0.80
HR HPV DNA	90.7	72.0	47.6	96.5	0.63	0.81
VHR HPV DNA	79.6	84.5	58.9	93.7	0.64	0.82
HPV 16/18	63.0	89.1	61.8	89.6	0.52	0.76
Pap test (ASCUS+) or HPV 16/18	98.1	46.1	33.8	98.9	0.44	0.72
Pap test (ASCUS+) and HPV 16/18	63.0	90.2	64.2	89.7	0.53	0.77
GA-NB CDSS	83.4	88.1	66.2	95.0	0.72	0.95

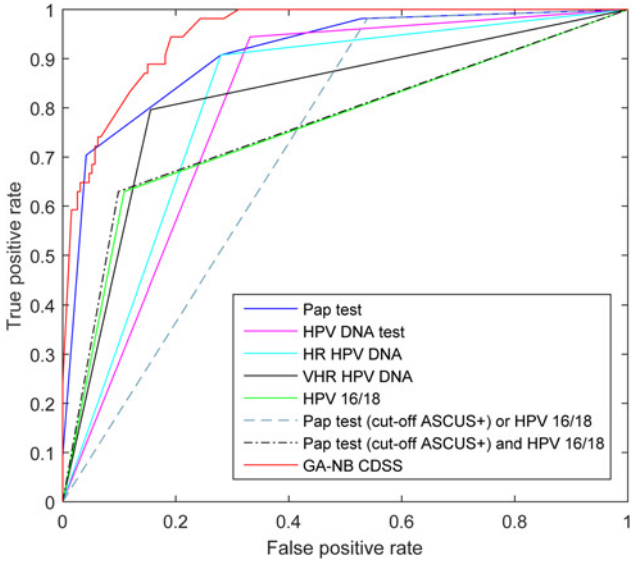

Fig. 4 *Comparison of ROC curves for Pap test, HPV DNA test and the GA-NB CDSS in detecting CIN2+*

The utilisation of irrelevant or redundant features increases the dimensionality of the feature space and as a result the complexity of the classification problem without any performance gain. Moreover, as stated in the so-called 'curse of dimensionality' (a major problem of pattern classification) [13]: for a fixed number of training samples, as the dimensionality increases, the classification rate of a classifier decreases after a peak, due to over-fitting. By observing Fig. 3, we can clearly state that the classifier suffers from over-fitting when a feature subset of more than 19 features is used. Moreover, it is worth noting that the fitness value achieved when only two features are used is much better compared with the values achieved when 38 or more features are used. These findings emerge the importance of the feature selection process in the construction of a classification system.

After the derivation of the optimal feature subset by the GA-NB framework, an NB classifier was trained with the 494 instances, using the specific feature subset. It should be noted that according to the work of Singhi and Liu [16], the practice of using the same dataset for feature selection and learning in classification problems is not inappropriate and the supposedly selection bias does not degrade the classification performance significantly. The trained NB classifier was eventually used as the classification module of the CDSS.

The developed CDSS was evaluated utilising the test set (246 cases). Table 3 presents the performance (on the test set) of the proposed system compared with the medical tests applied, in terms of SN, SP, PPV, NPV, YI, and area under the ROC curve (AUC), on the basis of detecting CIN2+. Fig. 4 presents the ROC curves for the CDSS and the medical tests in detecting CIN2+. The ROC curve of the CDSS was constructed using the provided posterior probabilities for CIN2+ for the cases of the test set. In comparison with the Pap test and the HPV DNA test (taking into consideration different positivity thresholds for these tests), the proposed architecture produced the most balanced results in terms of SN and SP, the best YI, the best AUC, and the best ROC curve. When ranking the tests by the maximal YI, which gives equal weight to SN and SP, the CDSS ranked highest, outperforming the medical tests.

6. Conclusions: Test Papanicolaou, despite being the most successful cancer prevention method, nowadays competes with HPV DNA typing. There are many reasons for this: test Pap requires experienced health professionals (cytopathologists or

doi: 10.1049/htl.2015.0051

cytotechnologists) to analyse glass slides through the microscope; this is a task requiring time and is prone to human errors. Additionally, if cell sampling is not performed correctly inadequate specimens are produced; therefore, women have to visit the health facilities one more time. On the other side, HPV DNA test can be performed in a batch manner, is less sensitive to human errors, requires less experienced personnel and eventually can cost less than test Pap. Today, developing countries choose to perform one of the two tests at frequent intervals. While in advanced economies performing both tests is usual and justified by the health economics, the process is facilitated because the two tests can be performed from biological material obtained in one visit.

Today, cases with HSIL cytology are prompted to immediate colposcopy, as the test Pap presents very high SP when the cut-off is HSIL+. However, the real question is the optimal management of the women with ASCUS or LSIL cytology or the HR-HPV positive women. According to the results, test Pap has very high SN when the cut-off is on ASCUS+ or LSIL+, while HPV DNA test has higher SP when the cut-off is on the existence of VHR or 16/18 subtypes. In both approaches this comes at the cost of SP for test Pap and SN for HPV DNA test. This is reflected to unnecessary referrals and missed positive cases, respectively. Obviously, a combinatorial approach is of great interest, as there is a promise of more balanced performance. However, simple algorithms for triage such as to refer for additional tests (such as colposcopy) when both tests are positive (logical 'AND') or when a single test is positive (logical 'OR') are not effective because the combination performance is still not optimally balanced (see Table 3 when we combined the cytologic approach which presents the best SN (Pap test with cut-off ASCUS+) with the HPV DNA approach with the best SP (HPV 16/18)). As one can observe from Table 3, a gain in SN is at the expense of SP and vice versa. Similar results have been presented in [17], where more combinations have been explored.

In comparison with the medical tests and their combinations (co-testing), the presented CDSS produced the most balanced results in terms of SN and SP, the best YI, the best AUC, and the best ROC curve.

These balanced outcomes make the proposed system a useful method for the everyday routine of the cytopathology laboratory or the gynaecological clinic. This system, by providing risk estimates for CIN2+ development, may be used as a 'third' opinion for women referred to colposcopy because of an ASCUS or LSIL cytology or an HR-HPV positive result. In this way, this CDSS may guide personalised management and therapeutic decisions, and reduce unnecessary colposcopies and treatments. By integrating the CDSS into the web-based information system *CxCaDSS* [15], it may serve as a web-based decision support system to physicians and medical researchers for the personalised management of women who attend a colposcopy room following a positive test result.

7. Funding and declaration of interests: This study was funded by the Greek Ministry of Development [General Secretariat for Research and Technology (GSRT)], Project 'HPVGuard', Cooperation 2011–2013 (code: 11ΣΥΝ_10_250) http://www.HPVGuard.org. Conflict of interest: none declared.

8 References

[1] Leyden W.A., Manos M.M., Geiger A.M., *ET AL.*: 'Cervical cancer in women with comprehensive health care access: attributable factors in the screening process', *J. Natl. Cancer Inst.*, 2005, **97**, (9), pp. 675–683

[2] Gomez-Roman J.J., Echevarria C., Salas S., *ET AL.*: 'A type-specific study of human papillomavirus prevalence in cervicovaginal samples in three different Spanish regions', *Acta Pathol. Microbiol. Immunol. Scand.*, 2009, **117**, (1), pp. 22–27

[3] Mayrand M.H., Duarte-Franco E., Rodrigues I., *ET AL.*: 'Human papillomavirus DNA versus Papanicolaou screening tests for cervical cancer', *N. Engl. J. Med.*, 2007, **357**, (16), pp. 1579–1588

[4] Cuzick J., Arbyn M., Sankaranarayanan R., *ET AL.*: 'Overview of human papillomavirus-based and other novel options for cervical cancer screening in developed and developing countries', *Vaccine*, 2008, **26**, (Suppl 10), pp. K29–41

[5] Naucler P., Ryd W., Tornberg S., *ET AL.*: 'Efficacy of HPV DNA testing with cytology triage and/or repeat HPV DNA testing in primary cervical cancer screening', *J. Natl. Cancer Inst.*, 2009, **101**, (2), pp. 88–99

[6] Henry M.R.: 'The Bethesda system 2001: an update of new terminology for gynecologic cytology', *Clin. Lab. Med.*, 2003, **23**, (3), pp. 585–603

[7] Mitchell M.: 'An introduction to genetic algorithms' (MIT, Cambridge, Massachusetts, USA, 1996)

[8] Goldberg D.E.: 'Genetic algorithms in search, optimization, and machine learning' (Addison-Wesley Publishing Co Reading, Massachusetts, USA, 1989)

[9] Yang J., Honavar V.: 'Feature subset selection using a genetic algorithm', *IEEE Intell. Syst. Appl.* , 1998, **13**, (2), pp. 44–49

[10] Huang C.-L., Wang C.-J.: 'A GA-based feature selection and parameters optimization for support vector machines', *Expert Syst. Appl.*, 2006, **31**, (2), pp. 231–240

[11] Tan K.C., Teoh E.J., Yu Q., *ET AL.*: 'A hybrid evolutionary algorithm for attribute selection in data mining', *Expert Syst. Appl.*, 2009, **36**, (4), pp. 8616–8630

[12] Xu L., Georgieva A., Redman C.W., *ET AL.*: 'Feature selection for computerized fetal heart rate analysis using genetic algorithms'. Conf. Proc. IEEE Engineering in Medicine and Biology Society, 2013, pp. 445–448

[13] Mitchell T.M.: 'Machine learning' (McGraw-Hill Boston, Massachusetts, USA, 1997)

[14] Sakka E., Prentza A., Koutsouris D.: 'Classification algorithms for microcalcifications in mammograms (Review)', *Oncol. Rep.*, 2006, **15**, pp. 1049–1055, Spec no

[15] Bountris P., Kotronoulas G., Tagaris T., *ET AL.*: 'CxCaDSS: a web-based clinical decision support system for cervical cancer'. Sixth European Conf. Int. Federation for Medical and Biological Engineering, 2015

[16] Singhi S.K., Liu H.: 'Feature subset selection bias for classification learning'. Proc. 23rd Int. Conf. on Machine Learning, Pittsburgh, Pennsylvania, 25–29 June 2006, pp. 849–856

[17] Bountris P., Haritou M., Pouliakis A., *ET AL.*: 'An intelligent clinical decision support system for patient-specific predictions to improve cervical intraepithelial neoplasia detection', *Biomed. Res. Int.*, 2014, p. 341483