HPVGuard: A Software Platform to Support Management and Prognosis of Cervical Cancer

Ioannis Tamposis, Evripidis Iordanidis, Leonidas Tzortzis
OraSys New Technologies S.A.
Larissa, Greece
{info, itamposis}@orasys.gr

Panagiotis Bountris, Maria Haritou, Dimitrios Koutsouris
Biomedical Engineering Laboratory
Institute of Communication and Computer Systems
National Technical University of Athens
Athens, Greece
{pbountris, mhari, dkoutsou}@biomed.ntua.gr

Abraham Pouliakis, Petros Karakitsos
Department of Cytopathology
"ATTIKON" University Hospital, University of Athens
Athens, Greece
{apouliak, pkaraki}@med.uoa.gr

Abstract—**Cervical cancer (CxCa) is one of the commonest reasons of women's mortality, although it can be prevented and treated if diagnosed early. Key to this is the regular examination with the test Papanikolaou but nowadays also with ancillary molecular biology tests. In this paper the authors present aspects of the architecture, design and implementation of the HPVGuard information system, a software platform capable to store and handle a multitude of medical examination data along with non-medical information. HPVGuard integrates artificial intelligence models that combine data from different medical examinations and produce an estimation of women's risk to develop CxCa. The application of HPVGuard proved that computerized systems supporting women control can be of extreme value. This is nowadays feasible via the use of inexpensive tools and can be made available to the end users as a web service on standard computers as well as on a variety of mobile terminals.**

Keywords - Cervical cancer; human papillomavirus (HPV); artificial intelligence models; cytopathology; distributed software applications; mobile health

I. INTRODUCTION

Cervical cancer (CxCa) is a serious hazard in the modern world, as it is the second most common cancer worldwide among women between 14 and 44 years old, and additionally, the third leading cause of cancer death [1]. Human papillomavirus (HPV) infection is proved to be associated with CxCa. This virus infects millions of women -actually HPV is the most common sexually transmitted infection. HPV is not a single virus, there are more than 100 HPV subtypes, among which 13 are high-risk (HR) oncogenic and may cause CxCa. The natural history of CxCa and its relation with HPV is today well studied [2]. Approximately 70% of women will be HPV infected during their lifetime, this is usually transient, however, the combination of high-risk HPV subtype and a compromised immune system along with persistence may lead to a progression towards cervical intraepithelial neoplasia (CIN), a precursor of CxCa. CIN should be detected and medically treated or monitored. It takes a lot of years to develop precancerous lesions (CIN) and CxCa [3], thus there is adequate time to identify these situations and act in a preventive way.

Test Papanikolaou (Pap test) is the most valuable tool in preventing cervical cancer [4]. The evaluation of the related smears is a task accomplished by well-trained cytopathologists. However, recently, HPV related tests have emerged. The typical test involves the detection of HPV DNA existence, and more specifically the detection of either specific individual subtypes, or the existence of HPV, or the existence of subtypes from the HR family or types 16 and 18. Additional tests involve the detection of HPV mRNA for the identification of the viral oncogenes E6/E7, which are linked to oncogenic activation. This mRNA typing can be performed via either nucleic acid sequence based amplification (NASBA) technique or Flow Cytometry technique, being used for the detection of cancer and CIN. Both techniques, provide increased positive predictive value and therefore may contribute to the reduction of unnecessary colposcopies [5]. Other approaches involve immunocytochemical detection of genetic effects, for example the overexpression of p16 or L1 proteins. In summary, nowadays CxCa control is linked not only to Pap test but to a wealth of other modern tests, the results of which require appropriate storage, processing and exploitation.

Despite the fact that the prevalence of cervical cancer decreased by 75% between 1955 and 1992 as a result of the CxCa screening programs using the methods introduced by George Papanikolaou, the mortality is still high. In Europe,

more than 50,000 women develop CxCa each year, about 25,000 finally die. The American Cancer Society reported about 12,200 new cervical cancer cases in USA in 2010 and 4,210 related deaths.

Obviously, due to the nature of the disease, it is required to have an organized control that has severe impact and can prevent the majority of CxCa. This benefit can be achieved only with the support of information technology both in the case of CxCa, but also for the management of other diseases as well [6, 7].

In this paper we present the design and implementation of HPVGuard, a software platform supporting CxCa control. The system stores the results of medical examinations such as Pap test and HPV tests, woman's medical history and non-medical data such as demographics. Moreover, it interacts with artificial intelligence (AI) components for CxCa risk estimation.

AI components are used for the combination of the results of various CxCa diagnostic techniques in order to provide clinicians with patient-specific predictions of diagnostic or prognostic outcomes and thus to identify women at true risk of developing CxCa. In this way, the HPVGuard platform serves as an integrated clinical decision support system for the management and prognosis of CxCa.

It is obvious that such IT systems are absolutely necessary in order to achieve efficient and continuous monitoring of women. Their design and implementation present significant challenges, like e.g. involving specialists from various health disciplines, i.e. gynecologists performing colposcopies, biologists conducting molecular tests, cytopathologists interpreting the Pap test along with the other data and pathologists diagnosing biopsies' material. Additionally, the security requirements are extremely strict because of the handling of sensitive personal data and medical records, and, finally, user training and acceptance, especially when such systems are deployed for the first time, is very crucial.

II. MATERIALS AND METHODS

A. Framework

The design of the integrated platform conforms to commonly accepted standards (common standards). It applies virtualization technologies (VMware ESXi [8]) to simplify the necessary infrastructure, it is installed and runs on virtual machines and therefore is cloud-ready. The architecture is 3-tier (database - application - client tier). It is web based and supports mobile functionality, which, though, complies with international certificates and is leading in its' kind.

In particular, at the database level the system uses Oracle 11g [9]. At the level of the application server the platform uses the GlassFish open source edition [10], a full featured enterprise ready and widely accepted application server. The application is developed with Oracle Application Express, an Oracle license free browser based rapid application development framework for developing rich interactive web and mobile applications. The web applications are very popular since they only need a web browser in order to operate, they are accessible from the internet and the maintenance and upgrading is easy as it is made in a single point.
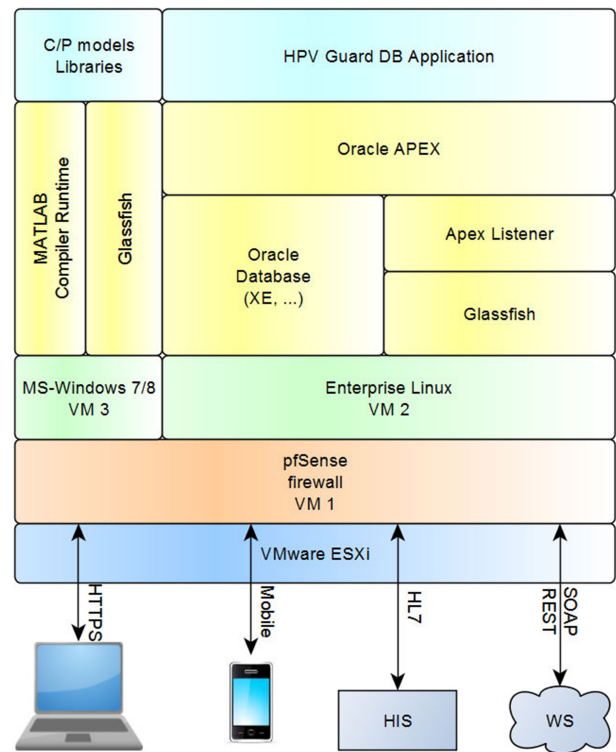

Figure 1. Diagram of HPVGuard Physical Architecture

As shown in Fig. 1, the HPVGuard platform runs on two independent VMs [8]. VM2 hosts the database and the application whereas VM3 hosts the AI components (MATLAB classification and prediction models). This architecture was preferred because: a) each subsystem has different technical requirements, b) each subsystem operates independently from each other and can be located separately.

B. Security

The platform is protected by an integrated firewall (pfSense [11]) hosted in VM1; the web server traffic is SSL encrypted, thus browsers accessing the server use https protocol. User access is by username and password that follow minimum requirements. All users are forced to change their passwords on initial login and every six months thereafter. Users are automatically logged out after 60 minutes of inactivity. Authenticated users are provided with different web interfaces, menu items and system operations based on their respective user role, such as: Doctor, Technologist or Typist which is assigned by a system administrator at the time of user registration. The system uses the Role Based Access Control (RBAC) approach to restrict system access to authorized users.

With regard to the data, there are three levels of access: "ALL" allows access to all data, "HPV" allows access to data characterized as HPV, and "LAB" allows access to data where the referring or performing institution is the same with the working institution. Anonymization removes sensitive personal information from the data, for users with limited rights either to view clinical data or to export data for research purposes.

C. Artificial Intelligence Components

The design and implementation of the artificial intelligence algorithms used in the HPVGuard platform, is presented in [12]. These algorithms were developed using the MATLAB platform [13], resulting to the classification and prediction models (C/P models). MATLAB algorithms were compiled in executable jar using the MATLAB Compiler and MATLAB Builder java toolbox [13]. C/P models were interpreted by the free-licensed Matlab Component Runtime (MCR) Library, which was installed in a server (VM2).

As presented in [12], the developed C/P models are based on Artificial Neural Networks (ANNs) which are used for the classification of cases into the following 4 classes corresponding to the cervical histology: Negative, CIN1, CIN2/3 and CxCa. The platform's architecture enable us to replace easily these models with improved ones or to utilize at the same time C/P models based on different AI techniques. For testing purposes, we have also integrated to the platform a C/P model based on Classification and Regression Trees [12, 14].

These C/P models take the results of the examinations as inputs and return to the system the classification outcome along with the posterior probabilities of class membership, providing in this way a prediction regarding the actual status of each woman. The clinical value of the C/P models for CxCa prevention is thoroughly described in [12].

D. Interoperability

Fig. 2 shows the interface between the database and the MATLAB's C/P models. GlassFish application server is hosted in VM2, as a provider of Internet services and, in this case, acts as servlet container. A servlet was developed which: 1) receives the examinations' results as HTTP request, 2) interfaces with MATLAB algorithms (jar) where it passes the data, 3) receives the results from the MATLAB algorithms (C/P models' output), 4) sends the C/P models' results to the database as HTTP response. The whole process starts from the database which executes an HTTP request to the servlet, as a result of the user's action.
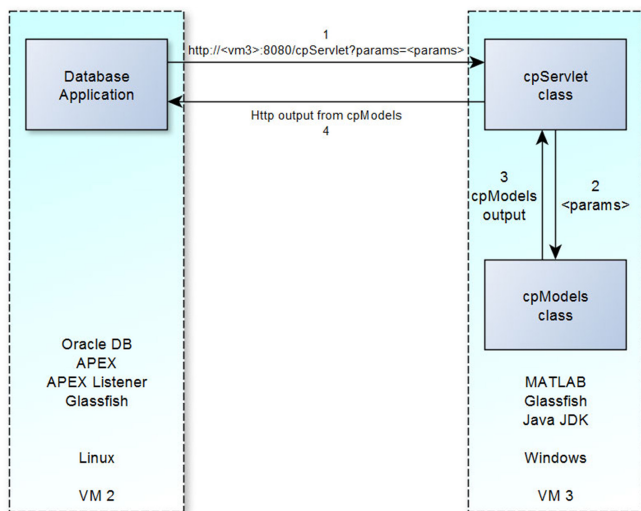

Figure 2. Diagram of interface between C/P models and DataBase Application
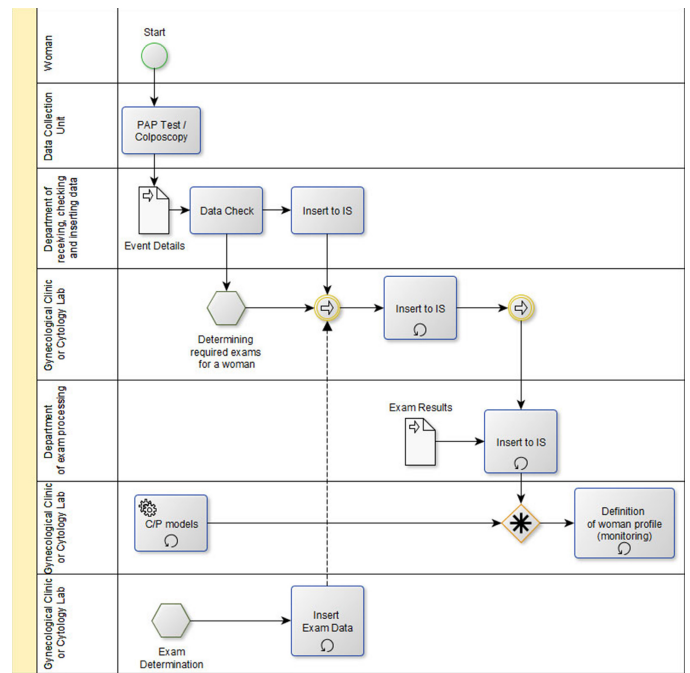

Figure 3.   Prediction Activity Diagram

The HL7 protocol [15] is used to retrieve and request clinical information from other systems. Web Services (RestFul, SOA) are used to provide/consume data to/from other systems (e.g. National Registries / Exam Referrals).

III.   RESULTS

Fig. 3 presents HPVGuard's main components and their interactions. Below, a usage scenario is presented.

Initially, a woman is registered in the patient archive, thus a Patient Record containing demographic and clinical information is created. Patient Record is the basic component for recording patients' health events like allergies, diseases, operations, medications and medical examinations. The primary use of such data is to provide clinical information to other systems like Clinical Information Systems (CIS). Another use of such data (commonly anonymized for privacy reasons) is for clinical research. However, this component encapsulates an essential feature; as shown in Fig. 4, Patient Timeline is a modern technology, providing physicians with the tools to make fast clinical decision as it presents the patient's medical history in a visual, graphic form. A major problem of this stage is identification of patients with unique identifiers such as SSN (Social Security Number), in order to eliminate or minimize multiple medical records for the same patient.

After registration, the physician creates a new visit for the woman and fills out the questionnaire with demographic and medical data, such as lifestyle habits, medical history and nutrition habits. The completion of a new visit ends with the request for exams to be performed for this woman. Exams may be selected via research protocols (predefined examination sets) or can be chosen separately from a list. These requests are assigned by the system an "On Hold" status and finally the physician sets the exam performing institution.
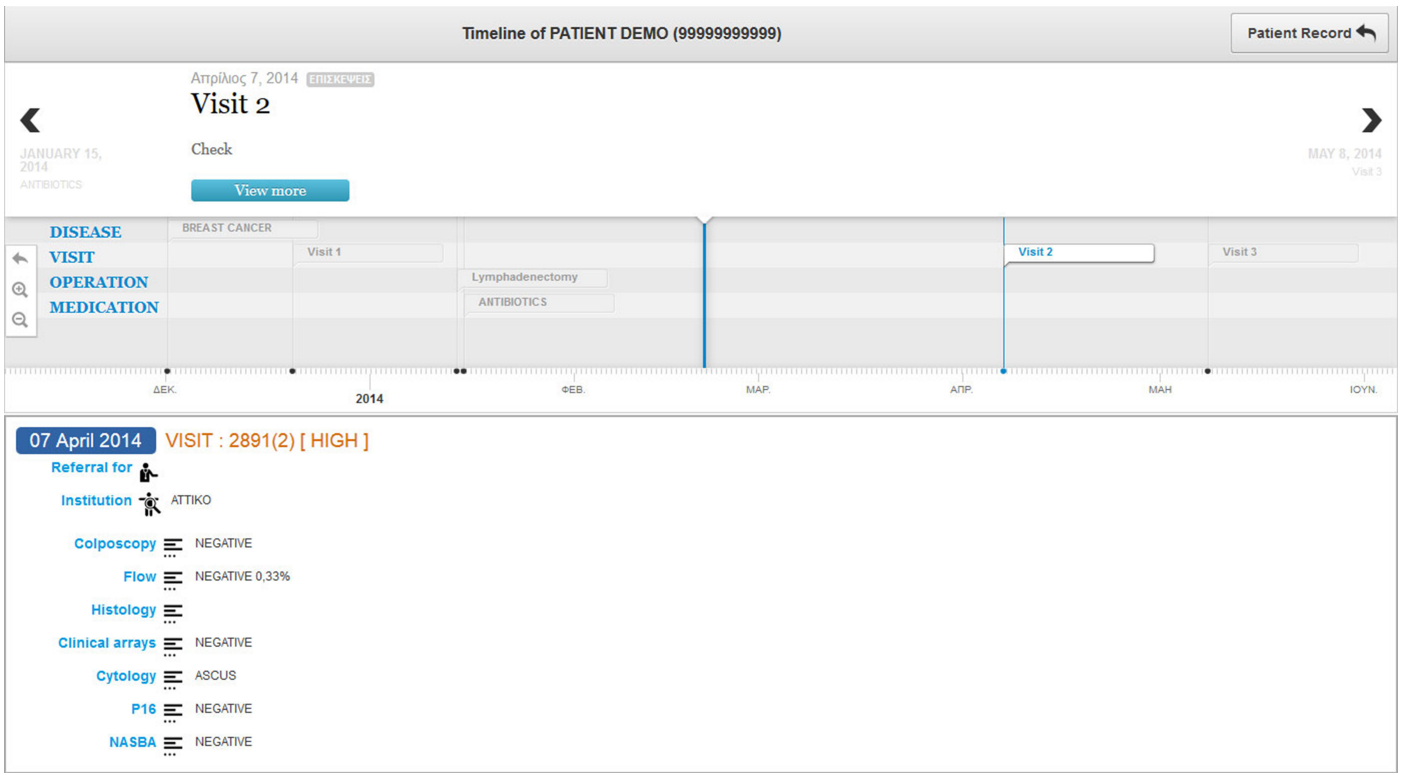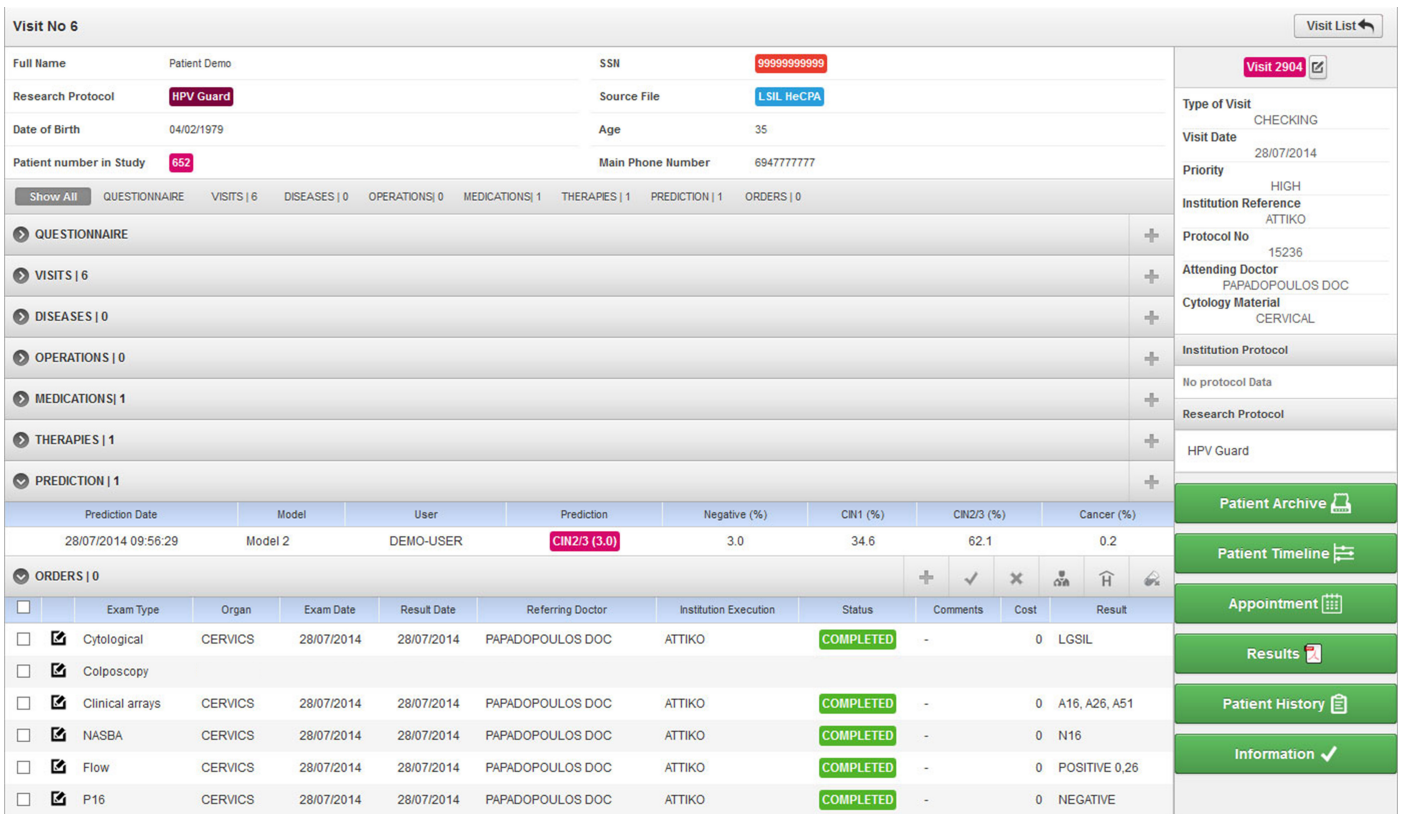
Figure 4.   HPVGuard's Patient Timeline

Figure 5.   HPVGuard's Visit Application Screen

The list of pending/requested examinations becomes available to each performing institution at their workstations in order to fill the exam results. When the institution inserts the exam results, exam status is changed to "Completed". If an exam is not planned to be performed, then the status is set to "Cancelled". When a result is filled, it is automatically included in the Patient Record. The system has special functions for notifying doctors and other users by e-mail or SMS (e.g. for exam results).

If specific exams have been completed, HPVGuard system provides physicians with the ability to execute the C/P models, aiming to identify the true risk of a woman to develop CxCa. For instance, in the snapshot of Fig. 5, for a case with low-grade squamous intraepithelial lesion (LGSIL) cytology, the C/P models provided a result of 62.3% probability of CIN2+, due to the existence of positive biomarkers (existence of high risk HPV, positive NASBA test for HPV type 16, and positive Flow Cytometry test), recommending in this way a possible referral to colposcopy of this woman due to high risk of developing high-grade CIN [12].

HPVGuard platform embeds features supporting mobile functionality and online access to specific tasks via mobile devices. Currently, HPVGuard supports access via smartphone/tablet to patient medical records, to visit details as well as functionality to exploit C/P models. Future considerations involve conditional and automated C/P model execution and dispatching of e-mails and SMSs to inform responsible physicians on their mobile devices.

IV. DISCUSSION AND CONCLUSIONS

Design and implementation of advanced information systems supporting CxCa prevention is feasible via freely available software. This is extremely important in countries such as Greece that face stringent economical situations. These modern technologies can make applications available from a single point located in the Cloud, thus facilitate maintenance, bandwidth availability and scalability in terms of processing power and storage.

These modern software development platforms and tools, along with the advancement of portable and hand held devices (multicore processors, screens with high resolution, high speed mobile telephony based data networks, modern security algorithms [16]) can make applications available for mobile users. This is a real advantage because CxCa prevention efforts are often based on not-fixed premises, i.e. often a group of specialists visit isolated villages to examine the population, cable based Internet is not always available in the examination premises and establishing land-line communications is not cost effective. In contrast, mobile data services are available in most places, thus the application of portable devices ranging from laptops to tablets and mobile phones via the HPVGuard platform can be a real benefit. In addition, the web based system access does not pose specific requirements for user terminals, being an application aspect that facilitates system use and adoption.

In relation to modern medical technology, the availability of a multitude of biomarkers along with the traditional well tested test Papanikolaou can be combined and provide a better estimation of the risk associated to each individual woman. In this field, statistical classifiers and more advanced techniques, such as ANNs, have been reported [12, 17, 18] and most of these have already produced excellent results in relatively small datasets. Information systems like the HPVGuard platform integrating modern C/P models provide tangible benefits not only for the individual women, but for the health professionals and the medical research community as well. To conclude, with the proposed architecture is now feasible to have a flexible web-based intelligent clinical decision support system available at anyplace/anytime.

REFERENCES

[1] A. Jemal, F. Bray, M. M. Center, *et al.*, "Global cancer statistics," *CA: A Cancer Journal for Clinicians,* vol. 61, pp. 69-90, Mar-Apr 2011.

[2] E. Paraskevaidis, A. Kaponis, V. Malamou-Mitsi, *et al.*, "The natural history of HPV infection of the uterine cervix. Long-term observational and histological data," *Anticancer Res,* vol. 22, pp. 1177-81, Mar 2002.

[3] N. F. Schlecht, S. Kulaga, J. Robitaille, *et al.*, "Persistent human papillomavirus infection as a predictor of cervical intraepithelial neoplasia," *JAMA,* vol. 286, pp. 3106-14, Dec 26 2001.

[4] Z. J. Hollenbeck, "The value of the Papanicolaou smear test," *Henry Ford Hosp Med Bull,* vol. 5, pp. 239-42, Dec 1957.

[5] C. Kottaridi, S. Tsiodras, A. Spathis, *et al.*, "Clinical performance of human papillomavirus E6, E7 mRNA flow cytometric assay compared to human papillomavirus DNA typing," *Anal Quant Cytol Histol,* vol. 33, pp. 305-10, Dec 2011.

[6] D. Iliopoulou, K. Giokas, S. Mougiakakou, *et al.*, "A Telematic System for Diabetes Management, Reporting and Patient Advice," *The Journal on Information Technology in Healthcare,* vol. 3, pp. 307-313, 2005.

[7] I. Kouris, S. Mougiakakou, L. Scarnato, *et al.*, "Mobile phone technologies and advanced data analysis towards the enhancement of diabetes self-management," *Int J Electron Healthc,* vol. 5, pp. 386-402, 2010.

[8] VMware. (2014). *VMware ESXi 5.5.0 Open Source*. Available: https://my.vmware.com/web/vmware/details?downloadGroup=ESXI550_OSS&productId=355

[9] Oracle Corporation. (2014). *Oracle Home Page*. Available: http://www.oracle.com

[10] Oracle Corporation. (2014). *GlassFish - World's first Java EE 7 Application Server*. Available: https://glassfish.java.net/

[11] pfSense. (2014). *pfSense Home Page*. Available: https://www.pfsense.org/

[12] P. Bountris, M. Haritou, A. Pouliakis, *et al.*, "An intelligent clinical decision support system for patient-specific predictions to improve cervical intraepithelial neoplasia detection," *Biomed Res Int,* vol. 2014, p. 341483, 2014.

[13] MathWorks. (2014). *MATLAB overview*. Available: http://www.mathworks.com/products/matlab/

[14] A. C. Stasis, E. N. Loukis, S. A. Pavlopoulos, and D. Koutsouris, "Using decision tree algorithms as a basis for a heart sound diagnosis decision support system," in *Information Technology Applications in Biomedicine, 2003. 4th International IEEE EMBS Special Topic Conference on*, 2003, pp. 354-357.

[15] Health Level Seven International. (2014). *Health Level Seven International*. Available: http://www.hl7.org/

[16] A. Giakoumaki, S. Pavlopoulos, and D. Koutsouris, "A multiple watermarking scheme applied to medical image management," *Conf Proc IEEE Eng Med Biol Soc,* vol. 5, pp. 3241-4, 2004.

[17] P. Karakitsos, A. Pouliakis, C. Meristoudis, *et al.*, "A preliminary study of the potential of tree classifiers in triage of high-grade squamous intraepithelial lesions," *Anal Quant Cytol Histol,* vol. 33, pp. 132-40, Jun 2011.

[18] P. Karakitsos, C. Chrelias, A. Pouliakis, *et al.*, "Identification of women for referral to colposcopy by neural networks: a preliminary study based on LBC and molecular biomarkers," *J Biomed Biotechnol,* vol. 2012, p. 303192, 2012.