Research & Reviews: Journal of Nursing & Health Sciences

Computational Tools for Human Papillomavirus (HPV) Risk Prediction by Fuzzy Logic

Ilka Kassandra Pereira Belfort[1]*, Sally Cristina Moutinho Monteiro[2], Allan Kardec Duailibe Barros[3]

[1]Student in Post-Graduation in Biotechnology, Northeast Biotechnology Network (RENORBIO), Brazil

[2]Professor of Post-Graduation in Biotechnology, Federal University of Maranhão, Brazil

[3]Professor of Post-Graduation Adult Health Program, Federal University of Maranhão, Brazil

Research Article

Received date: 03/11/2020

Accepted date: 19/12/2020

Published date: 26/12/2020

***For Correspondence**

Ilka Kassandra Pereira Belfort. Rua 04 Quadra 09 Casa 3 Residencial Primavera, São Luis, MA, Brazil. Tel: +55 (98) 99111-2694

E-mail: ilkabelfort@gmail.com

Keywords: Fuzzy, Medical informatics, Human papillomavirus, Pap test, Software validation.

ABSTRACT

Introduction: Human Papillomavirus (HPV) is one of the most common sexually transmitted infections (STIs) and responsible for approximately 99% of cervical cancers in the world. Thus, the objective of this work was to develop a computational tool for HPV risk prediction by fuzzy logic.

Material and Methods: This involves the development of a computational model using fuzzy logic tools to predict women with a greater predisposition to exposure and infection by HPV. The health, lifestyle, and sexual data of adult women were collected using a semi-structured questionnaire, oncotic cytology occurred from the analysis of the vaginal smear, and DNA-HPV research was carried out through the chain reaction of the Nested polymerase (PGMY09/11 (first-round PCR) and GP + 5/GP + 6 (second round PCR) using the Platinum ™ Taq DNA Polymerase system (Invitrogen ™, NY, USA). in the literature and later these were added to the results obtained from the analysis of the participants' data to construct the calculation with the determination of the risk.

Results: After the statistical analysis, the RISK set was concatenated into 3 sets (green, yellow, and red), the fuzzified data obtained as variables for availability in the risk calculator the following items: Green = [0-30%], low risk; Yellow = [31-50%], medium risk; RED = above 50%, high risk of HPV infection. os: 400 results of the epidemiological and cervical findings of the women participating in the research were used for training the software and 162 for system validation. After evidenced statistical data, the insertion of the results in the database started.

Conclusion: With the results obtained Fuzzy inference system can be as well adopted for the screening for HPV as this will in turn helps to reduce the mortality rate in cases with cancer. This expert system is user-friendly and carries out screening based on patients 'complain (clinical and laboratory data) to a medical expert.

INTRODUCTION

The World Health Organization (WHO) reports the occurrence of more than one million STI cases per day, worldwide. Approximately 357 million new infections are estimated each year, including HPV, chlamydia, gonorrhea, syphilis, and trichomoniasis. In Brazil, data from the Ministry of Health (MS) show that the population between 25 and 39 years of age are the most susceptible to contracting STIs [1]. The WHO also explains that there are inadequate screening programs, difficulties in accessing health services, absence of health education programs, early detection, and treatment, especially in developing countries [2].

HPV infection is one of the most frequent STIs in the world. It is estimated that 80% of the world population will come into contact with at least one type of HPV in their lifetime [3]. The virus is identified as an etiological state in almost 100% of cases of cervical cancer-CC [4]. Therefore, this condition is considered a priority public health problem, since the possibilities of cure are directly proportional to the early diagnosis and timely treatment of cancer [5].

Studies of the prevalence of HPV infection show that more than 630 million people, including men and women, are infected.

For Brazil, it is estimated that there are 9 to 10 million infected with this virus and that, every year, 700 thousand new cases occur [3]. Regarding the estimates of new cases of cervical cancer, the Cancer Institute estimated 16,590 in 2019 [6]. In this perspective, aiming at increasing the positive impacts in the analysis of the epidemiological situation of cervical cancer, there are necessary actions to increase the coverage of the cervical-vaginal lesions screening test, the improvement in the identification system of these lesions and cervical cancer and investment in health education to modify the population's exposure to risk factors for infection, as well as the dissemination of relevant information on this topic.

Primary health care is the main form of access for the population to strategic actions for the promotion, prevention, control, diagnosis, and treatment of STIs based on the availability of collective activities, distribution of male and female condoms, availability of the vaccine, and health guidelines, in addition to conducting the preventive cervical exam collection, thus initiating changes in the prevention, treatment and cure paradigms.

Thus, the performance of the primary level of health care can contribute to the improvement of indicators in the prevention and early diagnosis, and still use interventions in their risk factors, such as encouraging safe sex, reducing tobacco through the smoking program, and conducting the exam on time. It is worth mentioning the importance of having a preventive cervical exam regularly.

However, although the test is available throughout the public network, there are still taboos that revolve around the collection and the importance of prevention by women. On the other hand, health professionals still miss the opportunity to collect and inform at different times, especially for women who enter the health unit for other services. Therefore, there is a need to insert innovative methods in the active search and screening of these women faster.

Therefore, one of the possible improvements for this early search comes from technological developments. Regarding this evolution, bioinformatics appears as an instrument to aid screening, prediction, and early diagnosis of diseases, since it can be defined as research, development, and application of computational tools for the use of health data, including those to acquire, store, organize, archive, analyze this data [7]. Thus began the insertion of software, apps, games, among others, as learning methods for disease prediction, in addition to other functions. These methods emerged, mainly, to alert people to preventable diseases. And they are being successfully implanted/implemented in medicine [8].

Namely, the fuzzy approach has been used as an alternative for several areas, including Medicine. Its main advantage is the ease of dealing with linguistic terms and inaccurate and uncertain information, in addition to the low computational cost. Therefore, the use of these models in primary health care will be essential, inexpensive, and easy to handle by health professionals in the search for women quickly, especially those who are not looking for the exam, for possible diagnosis and brief referral for treatment. precursor lesions of cancer, if necessary, thus preventing disease progression and improving treatment efficiency.

Finally, the objective of the article was to create a computational tool using fuzzy logic to serve in the expansion of women's search strategies at an opportune time to perform the cervical preventive, thus contributing to the improvement of early diagnosis indicators that will reflect directly in the reduction of cervical cancer morbidity/mortality as a result of HPV.

MATERIALS AND METHODS

It is the development of software using fuzzy logic tools to screen women with a greater predisposition to risk exposure to HPV.

For the development of the computational model, data on human papillomavirus and its risk factors of women (≥18 years old) who sought the Primary Health Care Services of the Unified Health System of São Luis/MA were used, comparing them and validating them with the information on risk factors for HPV available in the scientific literature.

The study was approved by the Federal University of Maranhão Ethics Committee (number 2,383,604). All participants provided written informed consent.

Dataset collection and grouping of data sets

Participants included 562 women, aged 18-70 years, with active sexual life and users of Primary Care of the Unified Health System (SUS) of São Luís/MA. Exclusion criteria included menstruation on the day of the consult, hysterectomy, pregnancy, or at less than 45 days postpartum. All individuals answered a semi-structured questionnaire based on validated instruments that assessed sociodemographic characteristics, age, sexual behaviours, parity, smoking status, methods of contraception, and history of STIs.

Cervical epithelial tissue specimens were collected and tested. The presence of DNA/HPV was detected using nested polymerase chain reaction (Nested PCR) with the primer sets PGMY09/11 (first-round PCR) and GP+5/GP+6 (second round PCR) using the Platinum ™ Taq DNA Polymerase system (Invitrogen™, NY, USA).

According to the results, women who were infected with HPV were classified as DNA/HPV positive, and women no infected with HPV were classified as DNA/HPV negative. According to data in the literature, 6 (six) cofactors were selected as potential predictors and important for the development of the software, among them: Age, education, smoking, sexual behaviour, number of pregnancies, and use of oral contraceptives (OCA) [9].

Input variables

The input variables are the risk elements or factors which put a lady at a higher risk of getting HPV. The inputs are age, education, smoking, sexual behaviour, number of pregnancies, use of oral contraceptives. These factors were chosen as parameters for the starting point for evaluating the data collected and analysed in the exams collected from the women participating in the research. The analysis of epidemiological data was performed using the Statistical Package for the Social Sciences-SPSS version 22.

The whole data was divided into 400 data samples for training and 162 data samples for testing. The checking or validation set is used to check how generalized the trained set can be while the testing set is to evaluate how efficient the ANFIS can be in predicting HPV.

Fuzzy logic approach

The construction of the software followed supervised learning, where the algorithms learn from training data sets to predict results, where the output results are provided in the training process.

The development of the algorithm occurred in two phases:

a) The first was the PHP language and MySQL database for insertion of the collected data, separation, and standardization of the information. This stage aimed to search and compare faster through computer systems.

b) The second phase was the organization of the fuzzy sets to assemble the fuzzy logic in the system. In this work, the Trapeizodal membership function was used, where the algorithm performs the processing considering the limits of the interval in which the variable has full membership.

Following the construction of the models, it was necessary to divide the input data into degrees of risk, as well as the output set, which represented the final cloudy set. After defining the input and output sets, the base data was inserted in the software called HPV Risk Calculator. With the formation of fuzzified sets, the software base was built from the data sets, which was worked with the trapezoidal pertinence function in all variables. In the software development process, it is emphasized that it was created in such a way that a diffuse inference system is used that consists of a set of rules IF (antecedent) THEN (consequent), specifying a relationship between the diffuse sets of input and exit. Thus, a total of 38 rules for fuzzification were used to assemble the Mamdani Inference Method [10].

For the RISK output variable, the following percentages for evaluation are described: Very low-0 to 10%; Low-from 5 to 30%; Medium-from 20 to 50%; High-from 40% to 80%; Very high-from 70 to 100%.

Based on the indicators above, the software referred to the data reported in the interviews and exams of each research subject. After collecting the data, the calculation was performed to determine the risk, and then the analysis of these results, the RISK set was concatenated into 3 sets. Subsequently, the fuzzification of the data obtained the following items as variables for making the risk calculator available:

- Green = [0-30%], low risk;
- Yellow = [31-50%], medium risk;
- Red = over 50%, high risk.

RESULTS AND DISCUSSION

The calculator was built in two main phases: a collection of epidemiological data and biological samples from adult, sexually active women, users of SUS, and the development of predictive software to calculate the risk of infection with HPV.

It showed that the main risk factors for HPV in the studied population were smoking, age, education, sexual behavior, and use of oral contraceptives. These risk factors are consistent with the scientific literature that highlights the importance of guiding the modern lifestyle experienced by women, who, in general, acquire life habits that often constitute risks for certain diseases, which they do not even suspect subject.

These risk factors were used as input data for the software and consequently the construction of the calculator. First, software validation and training took place. Of all the women who were willing to participate in the epidemiological phase of the research, data from 400 participants were used to train the software and information from 162 to validate the system.

Therefore, the division into indicators (inbox named "indicator") was carried out, which were previously selected in the methodology phase (risk factors chosen from the literature compared to the risk factors observed in the epidemiological data collected). Then these "hypothetical" risk factors were tested with "real" information from the study participants. Then (inbox named "user") the relationship was made according to the selected indicator that served to prove whether or not the person had the positive variable or not, finally, the crossings of the selected indicator/variable and the positive result were performed or not. This determined whether the person was a candidate for exposure to the virus (inbox named "user-indicator").

Figure 1 shows this insertion model obtained from the results at the time of the learning steps (training) and after this, the moment of the assessment (validation) of the constructed model.

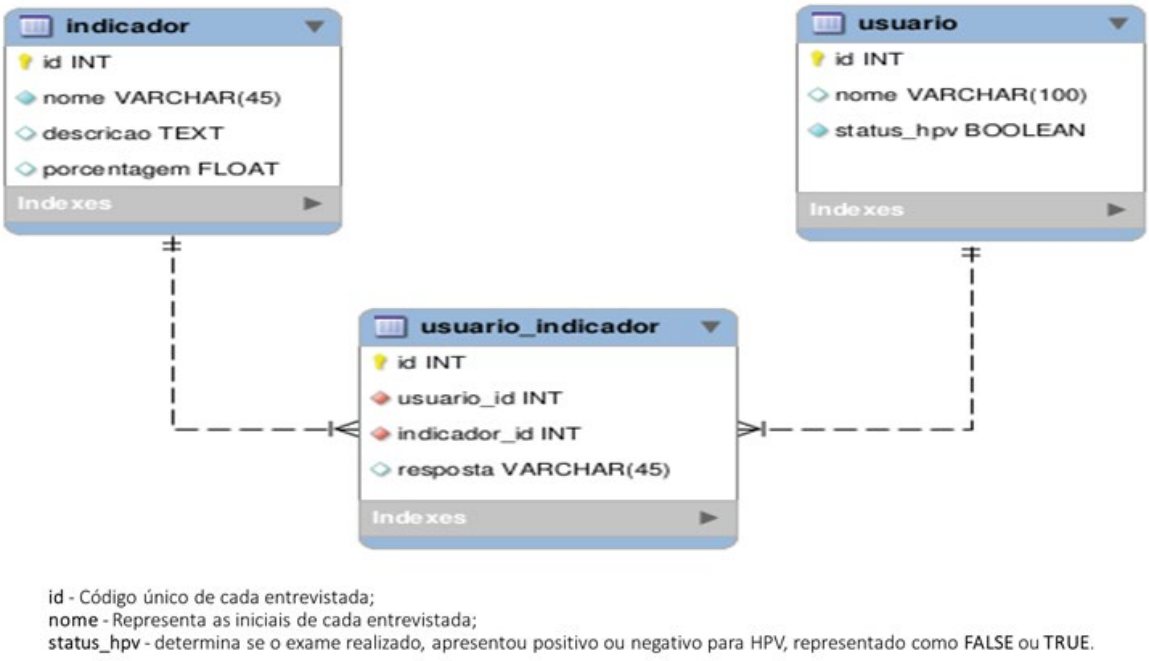

Figure 1. Results of the learning and validation step of the algorithm.

The software was hosted on Firebase (https://firebase.google.com/?hl=pt-br). The calculator's address was created by the author. Firebase is a Baas (Backend as a Service) for Google's Web and Mobile applications. Its launch and use started in 2004 and today it is considered a tool of the best option for some specific projects, due to the number of services offered, in addition to the ease of implementation (https://support.google.com/admob/answer/ 6360054? Hl = en-BR, 2019). The calculator was registered with the National Institute of Industrial Property (INPI) under the number BR512019000887-1.

After configuring the software, training, and validation, data were entered into the calculator, with an accuracy of 82% in the results.

When analyzing the model described here to the models, using fuzzy logic in the health area, available in the literature, it was possible to make some considerations. In the computational screening model to estimate the length of hospital stay, accuracy was also used to evaluate the data, however, it was not stated what percentage was reached by that, which reverberated in the inference about the lack of studies in this area using modeling. fuzzy, compromising the performance of comparisons for a more precise conclusion, nevertheless, there are articles in the national literature with the fuzzy application [11].

Evaluated in eighteen cases of oral squamous cell carcinoma (OSCC) the relationship between some cell cycle markers and HPV infection, conditionally to age, gender and certain habits of patients, and to assess the ability of fuzzy neurais networks (FNNs) in building up an adequate predictive model based on logic inference rules [12]. The study, although limited by small sample size.

We can also mention the model that other authors worked on fuzzy logic, the same aimed to evaluate the use of an intelligent computer system, using fuzzy logic as a method of reading by the specialist in predicting the risk of developing pre-neoplastic lesions. To build this software, the authors used 82 hypothetical cases (designed by a doctor) that encompassed different aspects of the daily practice of a woman's health specialist, including cases in which the specialist himself could face doubts in the assessment of the condition patient's clinic. The study had concluded that the fuzzy logic is an adequate reader of the specialist's thought that, if validated, can be used in the public health network, to carry out an organized schedule and consequent increase in the number of patients. However, it is emphasized that this software was built from hypothetical data, based on the experience of the specialist doctor, which emphasizes the importance of the work developed for the creation of the HPV risk calculator from real data collected from women with life active sexual [13].

HPV risk calculator

The software received the name of HPV Risk Calculator. The calculator in its first version has 3 data entry screens. Its handling lasts approximately 1 minute for the result **(Figure 2).**

This tool was designed to be initially used by health professionals in the Family Health Strategy (FHS) to assist in the active and early search of women who do not seek the Health Unit to collect cervical preventive. In this sense, to contribute to decreasing

the incidence of cervical cancer as cytological changes are discovered early, identifying this woman can be done to plan sequential preventive exams without the woman "missing" the procedure offered.

The software received the name of HPV Risk Calculator. The calculator in its first version has 3 data entry screens. Its handling lasts approximately 1 minute for the result **(Figure 2).**

Therefore, secondary screening or prevention will be represented by attracting women to perform the exam, through the opportunity to collect material for the exam, when performed assertively, it reduces the costs of surgeries and treatments for the state as well as an improvement the quality of life for women, family, and community.

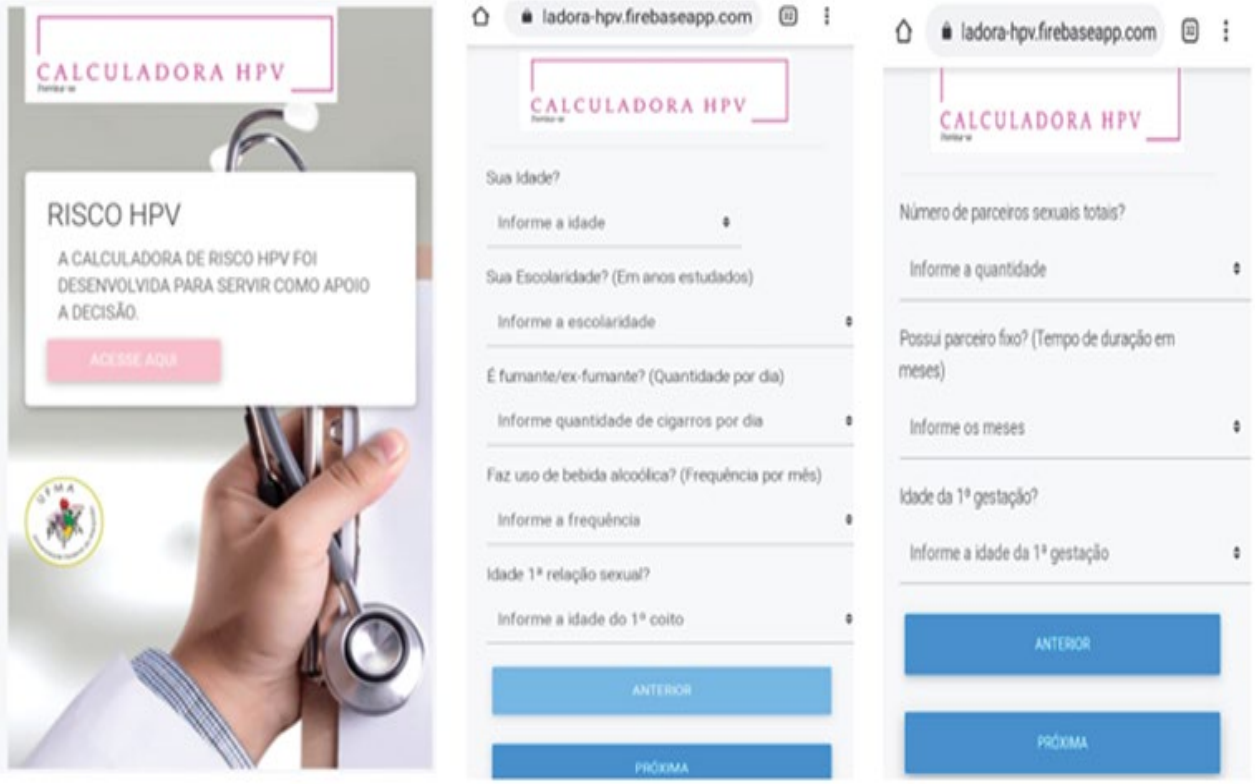

Figure 2. Own authorship.

The calculator is easy to use, very intuitive, and does not need manuals for access and/or a thorough evaluation of results. Due to responsive web design, it can be accessed from computers, tablets, and Smartphones. The user accesses the calculator through the link: https://calculadora-hpv.firebaseapp.com/, on the main screen of the system, there are the initial credits and there is only one button to initialize the questionnaire and a link to return to the main page, this screen gives the appropriate permissions to start the questionnaire. For the user, just click on "ACCESS HERE" and he will be redirected to the link https://calculadora-hpv.firebaseapp.com/calc in which he can insert the data.

CONCLUSION

With the results obtained Fuzzy inference system can be as well adopted for the screening for HPV as this will in turn helps to reduce the mortality rate in cases with cancer. This expert system is user-friendly and carries out screening based on patients 'complain (clinical and laboratory data) to the medical expert. This study presented a tool, of low financial cost, that can sort, in a satisfactory way, the duration of the average time of one minute, assuming a significant and important role so that health professionals can be prepared for a more fast and effective. It is software, so far, unique in the world for use in primary care, which is considered the place of entry of users in the Brazilian health system.

With the search for parameters for the development of the calculator, it was also observed that the demand for cervical prevention is low about the high power that the exam has for secondary prevention of cervical cancer. Finally, we consider the need to increase the number of collections for an accuracy range of 95% for optimal calculator validation.

REFERENCES

1. Ministry of Health. Sexually transmitted infections (STIs): what they are, what they are and how to prevent them. Ministry Cheers. 2019.
2. World Health Organization. Comprehensive control of cervical câncer: Essential practice guides. WHO. 2014.
3. Fedrizzi EN. Epidemiology of genital infection by HPV. Rev Bras Pat Trato Gen Inf. 2011;1:3-8.

4. Picconi MA. Human papilomavirus detection in cervical cancer prevention. Med. 2013;73:585-596.
5. World Health Organization. Fact sheet: HPV and cervical cancer. Brasilia. WHO. 2019.
6. National Cancer Institute. Estimate 2020: Incidence of cancer in Brazil. Rio de Janeiro: INCA. 2020.
7. Freitas R. Collective intelligence in bioinformatics: A systematic review of the literature [Dissertation]. Curitiba. 2019.
8. Drable RG, Moli ACA, Legey AP. Evaluation of the use of fuzzy logic to predict risk of Human Papilloma Virus. Reciis. 2014;8:344-358.
9. Ministry of Health. Support room for the strategic management of the Ministry of Health. SAGE. 2019.
10. Iancu, I. A mamdani type fuzzy logic controller: controls, concepts, theories and applications. Fuzzy Logic. 2012;16:325-350.
11. Coutinho KMV, et al. Fuzzy model estimating length of stay for cardiovascular diseases. Public health science. 2015;20:2585-2590.
12. Muzio LL, et al. Expression of cell cycle markers and human papillomavirus infection in oral squamous cell carcinoma: Use of fuzzy neural networks. Int J Cancer. 2005;115:717-723.
13. Dable RG, Mol ACA, Legey AP. Evaluation of the use of fuzzy logic to predict risk of human papilloma virus. Reciis. 2014;8:344-358.