Artificial Intelligence for Cervical Cancer Analysis

Ruth Reátegui[1], Priscila Valdiviezo-Diaz[2], Luis Barba-Guamán[3]

[1]Universidad Técnica Particular de Loja, Ecuador, rmreategui@utpl.edu.ec
[2]Universidad Técnica Particular de Loja, Ecuador, pmvaldiviezo@utpl.edu.ec
[3]Universidad Técnica Particular de Loja, Ecuador, lrbarba@utpl.edu.ec

ABSTRACT

Substantial works exist related to cervical cancer that is generally caused by human papillomavirus (HPV). These works have reported important results related to cervical cancer disease and HPV. Artificial intelligence applications are essential to diagnose and treat cervical cancer and so combat this disease in time. The purpose of this re- view is to summarize the published studies in recent years on the use of Artificial intelligence technologies in cases associated with cervical cancer. 26 articles related to HPV were analyzed based on three principal criteria: contributions of the paper, techniques, and algorithms of artificial intelligence, and datasets used. The analysis realized would help health staff to make decisions in the management of this disease.

Key words: artificial intelligence, cervical cancer, machine learning.

1. INTRODUCTION

Cervical cancer (CxCa) is the fourth most common cancer in women worldwide. In 2018, an estimated 570 000 women were diagnosed with CC and about 311 000 women died from this disease [1]. In the Region of the Americas cervical cancer is the second cause of death [2].

More than 100 types of human papillomavirus (HPV) exist, of which at least 14 cause cancer. Persistent infection with high-risk HPV (hrHPV), principally HPV16 and HPV18, is the cause of almost all CxCa worldwide [2].

CxCa is a preventable disease and curable with appropriate detection and treatment. Inequalities in access to primary prevention, screening, and treatment contribute to disparities in disease burden, mortality and treatment initiation, even in the same country [3] [4]. Nowadays, CC prevention is based on HPV vaccination, cytology test (Pap test), HPV test and colposcopy. The US Preventive Services Task Force (USP-STF)[5] recommends screening for CxCa considering the following condition:

- For women aged 21 to 29 years, every 3 years with cervical cytology alone
- For women aged 30 to 65 years, every 3 years with cervical cytology alone, every 5 years with hrHPV testing alone, or every 5 years with hrHPV testing in combination with cytology (cotesting).

Also, the American Society for Colposcopy and Cervical Pathology (ASCCP) recommends biopsies in non-pregnant women 25 years and older with very high risk of pre-cancer either immediate excisional treatment without biopsy confirmation, or colposcopy with multiple targeted biopsies [6].

Artificial intelligence (AI) had been applied in different fields and health is one of them. The amount of heterogeneous data produced by health institutions motivate the use of vary techniques or branch of AI such as machine learning, natural language processing, data/text mining, computer vision among others. With the emergence of big data, machine learning techniques are used in the medical research [7], the four major applications of healthcare that can benefit from AI techniques are prognosis, diagnosis, treatment, and clinical workflow [8].

Cancer is a disease analyzed by different AI techniques. Diagnosis and personalized cancer therapy have a great potential to be realized by AI systems [9]. Due to the advance of AI in cancer, the present work aims to review researches related with the application of some AI branches in CxCa analysis.

2. METHODOLOGY

The process to identify the relevant works that apply AI branches to analyze CxCa has the following steps:

- A preliminary search was made using SCOPUS Index with the keywords: cervical cancer, HPV, machine learning, datamining, artificial intelligence, data mining and text mining. We excluded the words medical image, computer vision, image processing, image classification, image detection and image

segmentation. Also, the search takes account articles from 2015 to 2021 in English language. This search had as result 43 articles.

- We eliminated articles that correspond with reviews, studies, and some articles that used image datasets. This selection has a result 26 articles.

- With the 26 articles, an exhaustive analysis was made in order to identify relevant characteristics related to AI and CxCa such as: contributions, algorithms and datasets.

3. RESULTS AND DISCUSSION

In this section relevant characteristics like contributions, algorithms and datasets are mentioned.

3.1 Articles Contributions

In recent years a great variety of works have been carried out around CxCa. Some authors used natural language processing (NLP), text mining and data mining techniques in order to extract information from different types of texts. For example, in [10] used NLP algorithms that incorporates machine learning and rule-based methods to extract diagnosis information from pathology reports. They focused on identify individuals with cancer or pre-cancer of cervix and anus. Furthermore, in [11] used NLP to identify the opinions on HPV vaccination reflected on Twitter over a 10-year period from 2008 to 2017. In addition, [12] applied text mining analysis to examine and compare Japanese newspaper articles related to HPV vaccination crisis. Some works applied data mining and statistical techniques to analyze HPV [13] and to identify virus in the typically ignored non-reference aligned sequence reads of MBD-seq experiments [14].

Other authors created platforms or decision support systems related with CxCa. In [15] the authors present a clinical decision-support system based on a hybrid genetic algorithm which combines the results of the Pap test and HPV DNA test to enhance accurate outcomes. They focused on feature selection problems to achieve balanced results in the sensitivity and specificity. Furthermore, [16] designed and implemented a platform to CxCa control. The system stores Pap test, HPV tests, woman's medical history and non-medical data. This platform used artificial neural networks for CxCa risk estimation. In [17] worked in a web-based decision support system using Probabilistic Neural Network and a Multilayer Perceptron Network that combine various CxCa diagnostic techniques to improve the accuracy of diagnosis and identify women at risk of developing CxCa. Also, [18] applied fuzzy logic inference to HPV detection. They focused on HPV types associated with cervical interstitial neoplasia (CIN) and CxCa.

Pre-processing and features selection are two important steps to improve machine learning tasks. In this sense, [19] used a Chicken Swarm Optimization algorithm for feature selection for the prediction of CxCa. Similarly, [20] proposes an integrative machine learning approach to analyze multiple gene expression profiles in CxCa to identify genetic markers to help in the diagnosis or prognosis of this type of cancer. Also, [15] integrated in a wrapper manner a hybrid genetic algorithm Bayesian classification (GA-Nave-Bayes (NB)) framework to feature selection. On the other hand, [21] presents a methodology for data pre-processing in cleaning data considering patient's health information related to CxCa.

Some works were focused on finding differences or common information around CxCa data. As an example, [22] used Apriori algorithm to identify whether any of the hrHPV genotypes, when co-infected, imply higher risk of developing CIN2/3 or above. Also, in [23] applied Decision Tree and Apriori algorithms on DNA sequences to analyzes both common and distinctive features of HPV Type 16, 18, and 58. Similarly, [24] applied Decision Tree Algorithm, Apriori Algorithm, and Support Vector Machine to analyze RNA sequences of highest risk and probably high-risk HPV to identify similarities and differences of those two categories. Furthermore, [25] used 138 cervical squamous cell carcinoma (CSCC) samples infected of HPV16; they used supervised clustering to identify and to characterize the differences between subgroups in terms of the HPVpca level, pathway activity, epigenetic dysregulation, somatic mutation frequencies, and likelihood of responding to chemo/immunotherapies. Also, in [26] a statistical and cluster analysis was developed in order to identify genes involved in both pregnancy and CxCa.

Every year increase the use of machine learning and artificial techniques in the genetic field. In [27] used tissue microarrays and machine learning enhanced profiling of STING signatures with immune subsets to analyze head and neck squamous cell carcinoma. In [28] presents some options for triage such as methylation, dual staining, machine learning, and artificial intelligence. Also, [29] combine genome comparison and machine learning techniques to identify a previously unnoticed insert near the 3-end of the E6 oncoprotein gene that is unique to hrHPV. In [30] the algorithm of random walk with restart (RWR) was applied to a protein-protein interaction (PPI) network to predict several novel HPV infection related genes.

Other works have been applied classification and prediction algorithms. In [31] the authors worked in the prediction of normal (NRML), high-grade intraepithelial lesion (HSIL) and cervical squamous cell carcinoma (CSCC) from exfoliated cell samples of cervix. They used PCA, LDA and

SVM as a classification technique. These authors present a spectroscopic modality by utilizing label-free ultrasensitive surface enhanced Raman scattering (SERS) technique to generate a differential spectral fingerprint for the prediction. Furthermore, [32] proposed a Gauss Newton representation-based algorithm (GNRBA) methods to classification of CxCa. They worked with four target variables: Hinselmann, Schiller, Biopsy, and Cytology. In [33] used capture sequencing and machine learning method to stratify CxCa lesions. They worked with Random Forest algorithm to risk stratification model for cervical precursor lesions based on CIN2 biomarkers. Also, they applied four clustering algorithms: k-prototypes, DBSCAN, HDBSCAN and PAM to explore association between biomarkers and sample categories (HPV and CIN1/CIN2). In [34] proposed a machine learning algorithm on imbalance data to classify Pap test results using the method of over-sampling.

Considering social media data, in [35] recollect, characterized and quantified 3 types of Twitter messages related to HPV vaccine. They used to supervise machine learning to classify tweets according to users, audience, sentiment, content, and vaccine characteristics.

Table 1 shows contribution and dataset used in the articles analyzed.

Table 1: Contributions and Datasets

Ref.	Contributions	Datasets
[10]	Pipeline-based on NLP algorithm that incorporates machine learning and rule-based methods to extract diagnosis elements from pathology report related to cervical and anus cancer.	Digital pathology reports from the HPV Vaccine Effectiveness Project.
[31]	Spectroscopic modality by utilizing label-free ultrasensitive SERS technique to generate a differential spectral fingerprint for the prediction of NRML, HSIL and CSCC from exfoliated cell samples of cervix.	Cell samples of cervix.
[27]	This work extends previous findings into the clinical domain using tissue microarrays and machine learning enhanced profiling of STING signatures with immune subsets. They focused on HNSCC.	Tumors from 246 patients.
[28]	An outlook to several other promising, yet not fully established, options for triage that have been proposed, including methylation, dual staining, machine learning, and artificial intelligence to analyzed CxCa.	Data from Swedish National Cervical Screening Registry, the Swedish Total Population Registry and the Swedish National Cancer Registry on individual level.
[19]	Chicken Swarm Optimization algorithm for feature selection, which can be used for the prediction of CxCa.	CxCa dataset from Venezuela.
[32]	Gauss Newton representation-based algorithm (GNRBA) methods in order to identify the classification of CxCa.	CxCa dataset from Venezuela.
[33]	Sequence and machine learning method to stratify the risk of cervical lesions and pro- vided valuable integrated triage strategies.	All 34 subjects from out-patient clinics of The First Affiliated Hospital of Sun Yat-sen University gave consent for retrieving their exfoliated cervical samples
[11]	An NLP framework to investigate the opinions on HPV vaccination reflected on Twitter over a 10-year period 20082017	Information collected from Twitter.
[12]	A text mining analysis to chronologically examine content distribution based on Japanese newspapers articles before and after the start of the HPV vaccination. Hierarchical cluster analysis to examine the appearance pattern of terms.	Information from newspapers.

[25]	Supervised clustering to discover subtypes and to characterize the differences between sub- groups in terms of the HPVpca level, pathway activity, epigenetic dysregulation, somatic mutation frequencies, and likelihood of responding to chemo/immunotherapies.	Data from GDC, Xena Public Data Hubs, cBio- Portal. Clinical and pathologic information from Firehose.
[13]	A framework to support clinical activity in HPV treatment. Data mining techniques have been proposed to analyze HPV.	Data coming from the microbiology unit of Magna Graecia University.
[29]	Genome comparison and machine learning techniques to identify a previously unnoticed insert near the 3-end of the E6 oncoprotein gene that is unique to hrHPV. Also, authors identify regions that differentiate be- tween high-risk and low-risk strains.	Nucleotide sequences of HPV and amino acid sequences of HPV E6 and E7 proteins from PaVE database.
[22]	A comparison of the performance of the BD Onclarity HPV assay with the Cobas HPV test in identifying cervical intraepithelial neoplasia 2/3 or above (CIN2/3+). Exploration of the CxCa detection specificity of Onclarity by machine learning.	Archive of the Cervical Cytology Laboratory, Department of Pathology, The University of Hong Kong.
[34]	Imbalance data on classification of cases Pap test results using the method of over-sampling.	Laboratory Test Data of cases Pap test results.
[23]	Identification of different types of high-risk HPV by analyzing the DNA sequences.	Dataset with Features of HPV Type 16, 18, and 58.
[30]	Identification of a group of potential novels HPV infection related genes in a PPI network. The algorithm RWR was applied on the PPI network using known HPV infection related genes as seed nodes.	HPV infection related genes from HPVbase.
[35]	Supervise machine learning to classified tweets on the basis of users, audience, sentiment, con- tent, and vaccine characteristics.	Information collected from Twitter.
[21]	Methodology of data pre-processing in cleaning data from patients health information in relation to CxCa.	CxCa dataset from Venezuela.
[26]	Data mining in placenta and CxCa gene expression to research genes involved in both pregnancy and cancer.	Placenta and CxCa genes samples. Information SAGE libraries. CxCa samples from the Hospital General at Clinica de Displasias in Mexico City.
[20]	Integrative machine learning approach to analyse multiple gene expression profiles in CxCa to identify genetic markers that associated with and may eventually aid in the diagnosis or prognosis of CxCa. They applied several machine learning methods to feature selection.	Information from NCBI.
[24]	Application of Decision Tree Algorithm, Apriori Algorithm, and Support Vector Machine to analyze RNA sequences of highest risk and probably high-risk human papillomaviruses in order to find out similarities and differences of those two categories.	Information from NCBI.
[18]	Occurrence of invasive CxCa based on the prevalence of HPV infection. They were focused on the HPV types associated with CIN and cancer.	Cervical epithelial tissue specimens.

[15]	Clinical decision support system based on a hybrid genetic algorithm Bayesian classification framework which combines the results of the Pap test and HPV DNA test to produce more accurate outcomes. They focus on feature selection.	Cases with HPV DNA and Pap test.
[16]	Software platform to support CxCa control. The system stores Pap test and HPV tests, woman's medical history and non-medical data such as demographics. The system uses AI components for CxCa risk estimation.	Pap test and HPV tests, woman's medical history and non-medical data.
[17]	A web-based clinical decision support system and its integration neural networks that com- bine various CxCa diagnostic techniques to in- crease the accuracy of diagnosis and identify women at true risk of developing CxCa.	Results of various CxCa diagnostic techniques.
[14]	Viruses related to cervix identification in the typically ignored non-reference aligned sequence reads of MBD-seq experiments. They applied statistical analysis, Pearsons Chi- squared test, Kruskal-Wallis test.	Two different collections of cervical samples and related cell lines.

3.2 Datasets

There are some Open Datasets for CxCa research. In this section, we describe datasets used by the majority of papers analyzed. Also, we present an overview of the use of these datasets and a description of features around them.

- **UCI Machine Learning Repository** integrates some datasets as a service to machine learning research. It was observed that the CxCa dataset collected from the Hospital Universitario de Caracas in Caracas, Venezuela, located in the UCI repository is widely used for research related to CxCa prediction. This dataset comprises statistic information, historic medical records and habits of 858 patients [19, 21, 32].
- **National Center for Biotechnology Information (NCBI)** provides access to biomedical and genomic information. Some studies [20, 24] that use information of this website analyzed HPV infection related genes information and RNA sequences of HPV viruses.
- **Twitter social network** used by researchers to analyzed Twitter messages related to HPV vaccine, cervical cancer, and Gardasil. Supervise machine learning is used to classify the tweets based on users, content, vaccine characteristics, etc. [35], and Natural Language Processing (NLP) to investigate the opinions on HPV vaccination reflected on Twitter [11].
- **The Swedish Cancer Register** covers the malignant cases of cancer registered every year in Sweden. This dataset is used to get information available about Cervical Screening Registry [28].
- **Hospital of Sun Yat-sen University.** The First Affiliated Hospital of Sun Yat-sen University is an important base for research, medical service, and medical education. Information on outpatient clinics of this hospital has been used to build a risk stratification model for cervical precursor lesions [33].

It was also observed that some of the works analyzed are based on databases with information from Laboratory Test Data of cases with HPV DNA and Pap test [23, 15], digital pathology [10], tumor samples [27, 25], and contents from newspapers related to HPV vaccination [12].

3.3 Machine Learning Task and Algorithms

Recently, several machine learning techniques have been applied in the health area, one of them is medical care, where radiological images, medical records of pathologies, diagnosis, follow-up and monitoring help in the prevention of diseases in patients. Machine learning techniques can also be applied to predict adverse reactions to medical treatments or the degree of compliance with treatment by patients.

From the literature review, we can observe the use of different machine learning techniques used in CxCa detection. We found that classification tasks use known techniques such as SVM, decision tree, artificial neural networks (ANN) and Bayesian networks; in clustering tasks, besides we can observe algorithms such as DBSCAN, HDBSCAN, hierarchical clustering, among others. Natural language processing has been used to identify words and phrases about cancer on the social network Twitter. In addition, there are different types of software that can be applied to different groups of diseases, these computer programs have been improving their learning through repeated and continuous use. Table 2 shows a list of algorithms applied in the articles analyzed.

Table 2: Algorithms for cervical cancer detection and analysis

Task	Algorithm	References
Classification and Prediction	Support Vector Machine (SVM)	[31], [29], [22], [24]
	Chicken Swarm Optimization	[19]
	GNRBA method for solving cancer classification	[32]
	Bayesian	[13], [15]
	KNN	[32]
	Decision tree	[22], [23], [24]
	Naive Bayes	[15]
	Random Forest	[33], [22]
	Artificial Neural Network (ANN)	[28], [16]
	Probabilistic Neural Network (PNN)	[17]
Dimensional Reduction	LDA, PCA	[31], [25]
Association rules	Apriori algorithm	[22], [23], [24]
Clustering	Supervised Hierarchical clustering	[25]
	Hierarchical clustering	[26]
	K-prototypes	[33]
	DBSCAN	[33]
	HDBSCAN	[33]
	PAM	[33]
Others	Kaplan-Meier	[27]
	NLP and Text mining	[11], [12]
	TIDE and Virusseq	[25]
	RNA	[29]
	ADASYN, ADASYN-KNN, SMOTE-N	[34]
	Random walk with restart (RWR) Jaccard coefficient	[30]
	Gene expression analysis	[20]
	Fuzzy Logic	[18]
	Statistical analysis, Pearson's Chi-squared test, and Kruskal-Wallis test	[26], [14]
	It is not specified.	[35]

4. CONCLUSION

Nowadays AI encompasses a considerable number of applications in all fields and the health is no exception. A vary of IA branches, techniques and algorithms are applied in researches related to CxCa. Most of the works are focused on classification or prediction of cervical cancer using different types of data. In this sense, researches use genes information, clinical test, medical records, demographic information and so on. In addition, an important number of works are interested in identifying differences of common information in HPV types that provoke CxCa. Those works use cluster analysis techniques. Despite the considerable number of works around CxCa, more research is needed in order to improve diagnosis, prevention and treatment of CxCa considering the different types of data and AI branches. The present work has some limitations, such as the fact that works that analyze images such as [36] were not considered. Also, a comparison of algorithms performance in similar task would be considered in future works.

REFERENCES

1. M. Arbyn, E. Weiderpass, L. Bruni, S de Sanjos, M. Saraiya, J. Ferlay, and F. Bray. **Estimates of incidence and mortality of cervical cancer in 2018: a worldwide analysis**. Lancet Glob Health. Feb;8(2):e191-e203, 2020. doi: 10.1016/S2214-109X(19)30482-6.
2. Pan American Health Organization. **Plan of Action for Cervical Cancer Prevention and Control 2018-2030**. Washington, D.C.: PAHO; 2018.
3. J.A. Hernández Vargas, P.X. Ramírez Barbosa, A.M. Valbuena-Garcia, L. Acuaña, and J. A. González-Diaz. **Factors associated with delays in time to treatment initiation in Colombian women with cervical cancer: A cross-sectional analysis**. Gynecologic Oncology Reports, vol. 35, 2021, pp. 100697.
4. P. Hung, W.E. Zahnd, H.M. Brandt, S.A. Adams, S. Wang, and J.M. Eberth. **Cervical cancer treatment initiation and survival: The role of residential proximity to cancer care.** Gynecologic Oncology, Volume 160, Issue 1, 2021, Pages 219-226, ISSN 0090-8258.
5. US Preventive Services Task Force. **Screening for Cervical Cancer: US Preventive Services Task Force Recommendation Statement.** JAMA. 2018, 320(7):674686. doi:10.1001/jama.2018.10897.
6. N. Wentzensen, L.S. Massad, E.J. Mayeaux, M.J. Khan, A.G. Waxman, M.H. Einstein, C. Conageski, M.H. Schiman, M.A. Gold, B.S. Apgar, D. Chelmow, K.K. Choma, T.M. Darragh, J.C. Gage, F.A.R. Garcia, R.S. Guido, J.A. Jeronimo, A. Liu, C.A. Mathews, M.M. Mitchell, A.B. Moscicki, A.P. Novetsky, T. Papasozomenos, R.B. Perkins, M.I. Silver, K.M. Smith, E.A. Stier, C.A. Tedeschi, C.L. Werner, and W.K. Huh. **Evidence-Based Consensus Recommendations for Colposcopy Practice for Cervical Cancer Prevention in the United States**. J Low Genit Tract Dis. 2017 Oct, 21(4):216-222.
7. A. Garg, and V. Mago. **Role of machine learning in medical research: A survey**. Computer Science Review, 40, (2021) doi.org/10.1016/j.cosrev.2021.100370
8. A. Qayyum, J. Qadir, M. Bilal, and A. Al-Fuqaha. **Secure and Robust Machine Learning for Healthcare: A Survey**. IEEE Reviews in Biomedical Engineering, Biomedical Engineering, IEEE Reviews in, IEEE Rev Biomed Eng, 14, 2021, pp. 156-180. doi:10.1109/RBME.2020.3013489
9. D.H. Suh, M. Kim, K. Kim, H.J. Kim, K. Lee, and J. Kim. **Major clinical research advances in gynecologic cancer in 2016: 10-year special edition**. Journal of Gynecologic Oncology, 28, 2017.

10. C.R. Oliveira. **Natural language processing for surveillance of cervical and anal cancer and precancer: Algorithm development and split-validation study**. JMIR Med. Informatics 8, 2020.

11. X. Luo, G. Zimet, and S. Shah, S. **A natural language processing framework to analyse the opinions on HPV vaccination reflected in twitter over 10 years (2008 - 2017)**. Hum. Vaccines Immunother. 15, 14961504, 2019.

12. T. Okuhara, H. Ishikawa, M. Okada, M. Kato, and T. Kiuchi. **Newspaper coverage before and after the HPV vaccination crisis began in Japan: A text mining analysis**. BMC Public Health 19, 2019.

13. D. Mirarchi, P. Vizza, G. Tradigo, G. Di Fatta, and P. Veltri. **On the use of mining techniques to analyse human papilloma virus dataset**. in Proceedings – 2018 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2018 (ed. Schmidt H. Griol D., W. H. B. J. Z. H. C. Z. H. X. D. J. Z. L.) 20412048, 2019. doi:10.1109/BIBM.2018.8621373

14. K. Mensaert. **Mining for viral fragments in methylation enriched sequencing data**. Front. Genet. 5, 2015.

15. P. Bountris. **Development of a clinical decision support system using genetic algorithms and Bayesian classification for improving the personalized management of women attending a colposcopy room.** Healthc. Technol. Lett. 3, 143149, 2016.

16. I. Tamposis. **HPVGuard: A software platform to support management and prognosis of cervical cancer.** in Proceedings of the 2014 4th International Conference on Wireless Mobile Communication and Healthcare - Transforming Healthcare Through Innovations in Mobile and Wireless Technologies, MOBIHEALTH2014401405, 2015.

17. P. Bountris, **CxCaDSS: A web-based clinical decision support system for cervical cancer**. in IFMBE Proceedings (ed. Lackovic I., V. D.) 45, 757760, 2015.

18. K. Abdelhalim, M. Abas, H. Mokhtar, and B. Khaoula. **Prevalence of risk factors for cervical cancer induced by papillomavirus infection: A fuzzy inference system modeling**. Der Pharm. Lett. 8, 106112, 2016.

19. A. K. Tripathi. **Prediction of Cervical Cancer Using Chicken Swarm Optimization**. Adv. Intell. Syst. Comput. 1087, 591604, 2020.

20. M. S. Tan, S.W. Chang, P. L. Cheah, and H.J. Yap. **Integrative machine learning analysis of multiple gene expression profiles in cervical cancer**. PeerJ 2018, 2018.

21. W. A. Nee, and M. Jayabalan. **Data preparation techniques for cervical cancer screening**. J. Adv. Res. Dyn. Control Syst. 10, 1925, 2018.

22. O. G. W. Wong. **Machine learning interpretation of extended human papillomavirus genotyping by onclarity in an Asian cervical cancer screening population**. J. Clin. Microbiol. 57, 2019.

23. D.E. Lee, and T. Yoon. **Analysis of high-risk human papillomavirus using decision tree and Apriori algorithm**. in ACM International Conference Proceeding Series 2226, 2018. doi:10.1145/3290818.3290830

24. Y. Cho, S. Burm, N. Choi, and T. Yoon. **Analysis of Human Papillomavirus Using Datamining-Apriori, Decision Tree, and Support Vector Machine (SVM) and its Application Field**. in MATEC Web of Conferences (ed. Kavakli M. Salami M.-J.E., B. M. A. B. M. L. S. C. H. P. M. A. A. M. A. B.) 56, 2016.

25. X. Lu. **Immune Signature-Based Subtypes of Cervical Squamous Cell Carcinoma Tightly Associated with Human Papillomavirus Type 16 Expression, Molecular Features, and Clinical Outcome**. Neoplasia (United States) 21, 591601, 2019.

26. K. Taniguchi-Ponciano. **The KISS1 gene overexpression as a potential molecular marker for cervical cancer cells**. Cancer Biomark. 22, 709719, 2018.

27. X. Luo. **HPV16 drives cancer immune escape via NLRX1-mediated degradation of STING**. J. Clin. Invest. 130, 16351652, 2020.

28. K. Sundstrm, E. Herweijer, and J. Wang. **Cervical screening in high-income countries: the need for quality assurance, adjunct biomarkers and rational adaptation to HPV vaccination**. Prev. Med. (Baltim), 2020. doi:10.1016/j.ypmed.2020.106382

29. N. Auslander, and E. V. Koonin, Y.I. Wolf, and S.A. Shabalina. **A unique insert in the genomes of high-risk human papillomaviruses with a predicted dual role in conferring oncogenic risk**. F1000Research 8, 2019.

30. L. Zhu, F. Su, Y. Xu, and Q. Zou. **Network-based method for mining novel HPV infection related genes using random walk with restart algorithm**. Biochim. Biophys. Acta - Mol. Basis Dis. 1864, 23762383, 2018.

31. V. Karunakaran. **Diagnostic spectro-cytology revealing differential recognition of cervical cancer lesions by label-free surface enhanced Raman fingerprints and chemometrics.** Nanomedicine Nanotechnology, Biol. Med. 29, 2020.

32. Z. Rustam, V.A.W Hapsari, and M.R. Solihin. **Optimal cervical cancer classification using Gauss-Newton representation based algorithm**. in AIP Conference Proceedings (ed. Mart T. Triyono D., A. I. T.) 2168, 2019.

33. R. Tian. **Risk stratification of cervical lesions using capture sequencing and machine learning method based on HPV and human integrated genomic profiles**. Carcinogenesis 40, 12201228, 2019.

34. Y.E. Kurniawati, A.E. Permanasari, and S. Fauziati. **Adaptive Synthetic-Nominal (ADASYN-N) and Adaptive Synthetic-KNN (ADASYN-KNN) for Multiclass Imbalance Learning on Laboratory Test Data**. in Proceedings – 2018 4th International Conference on Science and Technology, ICST 2018, 2018.

35. P.M. Massey. **What drives health professionals to tweet about HPV vaccine?** Identifying strategies for effective communication. Prev. Chronic Dis. 15, 2018.
36. T. Prabhakar, and S. Poonguzhali. **A comparison of classification methods to classify the breast lesions from ultrasound images using texture feature extraction**. International Journal of Advanced Trends in Computer Science and Engineering, 9 (5), 2020. doi: 10.30534/ijatcse/2020/01952020