
Elmer Diaz, Andres Ccopa, Lenis Wong

Abstract: Cervical Cancer is considered the fourth most common female malignancy worldwide and represents a major global health challenge. As a result, in recent years, various proposals and researches have been conducted. This study aims to analyze the data presented in current researches regarding cervical cancer and contribute to future research, all through the framework of literature review, based on 3 research questions: Q1: What are the risk factors that cause cervical cancer? Q2: What preventive measures are currently established for cervical cancer? and, Q3: What are the techniques to detect cervical cancer? Findings show that detection techniques are complementary since they are categorized under machine learning. Therefore, we recommend that further study be promoted in these techniques as they are helpful in the detection process. In addition, risk factors can be considered for a greater scope in detection, such as HPV infection, since it is the most relevant factor for the development of cervical cancer. Finally, we suggest to conduct further research on preventive measures for cervical cancer.

Keywords: Cervical cancer, Cervical cancer diagnosis, Machine learning.

I. INTRODUCTION

The Human Papilloma Virus (HPV) is considered the main risk factor for Cervical cancer, according to various supported analysis occurred from 1974 to 1976. Based on those researches, Meisels and Fortin suggested that the lesions could be differentiated between “benign, warty”, which do not progress to cervical cancer and "non-viral" precursor lesions that progress to cervical cancer. This concept was supported because of the slight dysplastic lesions, typical particles of HPV, were identified in the cervix [1]. Cervical cancer is a cellular alteration that originates in the epithelium of the cervix that initially manifests itself through the slow and progressive evolution of precursor lesions, which leads to cancer in situ (confined to the epithelial surface) or an invasive cancer where cells with malignant transformation cross the basement membrane [2].

Currently, cervical cancer is considered the fourth most common female malignancy worldwide and represents a major global health challenge. Approximately 90% of the 270,000 deaths from cervical cancer in 2015 occurred in underdeveloped countries, where mortality is 18 times higher than in developed countries [3].

For this reason, in recent years different proposals have emerged regarding risk factors, detection and prevention of cervical cancer. Therefore, in this study, a systematic literature review is carried out, focusing on risk factors that cause cervical cancer, existing prevention measures against this disease, and its detection techniques.

The present work is organized in the following order. Section 2 presents the research methodology for the systematic literature review on the appropriate analysis of cervical cancer. Section 3 presents the proposed taxonomy. Section 4 analyzes the findings where the proposed taxonomy is evaluated. Finally, section 5 concludes the study.

II. RESEARCH METHODOLOGY

For the systematic review of the literature, the methodology presented by Wong et al. [4], based on the guidelines of Kitchenham et al. [5], has been considered. It consists of following steps: (1) Review planning: In this first part, research questions are formulated and the review protocol is defined. (2) Review development: In this second part, primary studies are chosen according to the selection and exclusion guidelines set out above. (3) Review results: In this third part, analysis and statistics of previously selected studies are shown, and the respective details are presented.

A. Review Planning

The following research questions have been proposed to achieve the research purpose: Q1: What are the risk factors that cause cervical cancer? Q2: What preventive measures are there for cervical cancer? Q3: What are the techniques to detect cervical cancer?

The databases that were mainly used to define the search protocol are the following: SCIENCE DIRECT, IEEE Xplore Digital Library and Google Scholar. The research covers the period from January 2014 to September 2019.

The following search sequence was used: TITLE-ABS-KEY("cervical cancer") or TITLE-ABS-KEY("cervical uterine cancer") or TITLE-ABS-KEY("cervical intraepithelial neoplasia"), that have been applied in the title, abstract and keywords. After that, the selection and exclusion criteria, shown in Table I, were applied.

Table- I: Selection and exclusion criteria

<table>
<thead>
<tr>
<th>Selection criteria</th>
<th>Exclusion criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Studies related to the state of the art</td>
<td>Sources that are different from Journals and Proceeding</td>
</tr>
<tr>
<td>Risk factor proposals</td>
<td>Study language is different from English and Spanish</td>
</tr>
<tr>
<td>Different proposals for diagnosing cervical cancer</td>
<td>Diagnosis of cancer other than cervical cancer</td>
</tr>
<tr>
<td>Related to cervical cancer</td>
<td>Radiotherapy, chemotherapy and other treatments, therapies</td>
</tr>
</tbody>
</table>

B. Review Development

This phase explains the review process development considering the search statement, databases and selection and exclusion criteria found in Table 1. The flow chart of the review process is shown in Fig. 1.

Figure- 1: Systematic literature review process

C. Review Results

This phase shows the results of the systematic literature review. The result of the review process provided 36010 studies, which 25 were selected according to the selection and exclusion criteria. Those proposal studies were analyzed to answer the research questions. In the Table II, it is possible to see the number of studies selected by each type of source.

Table- II: Potentially eligible studies and selected studies

<table>
<thead>
<tr>
<th>Source</th>
<th>Potentially eligible studies</th>
<th>Selected studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Others</td>
<td>25012</td>
<td>4</td>
</tr>
<tr>
<td>Science Direct</td>
<td>6668</td>
<td>11</td>
</tr>
<tr>
<td>IEEE</td>
<td>4330</td>
<td>10</td>
</tr>
<tr>
<td>Total</td>
<td>36010</td>
<td>25</td>
</tr>
</tbody>
</table>

Fig. 2 shows the number of studies related to obtaining requirements between 2014 and 2019. These 25 studies correspond to different aspects of risk factors, diagnostic proposals and issues related to cervical cancer.

Figure- 2: Selected cervical cancer studies by year

III. PROPOSED TAXONOMY

According to the analysis of results obtained in the literature review, a taxonomy has been developed according to the research questions formulated: “Risk factors” (Q1), “Preventive measures” (Q2) y “Detection techniques” (Q3). Fig. 3 contains the taxonomy. The risk factors classification is related to studies on risk factors that influence the development of cervical cancer. Prevention measures will allow us to know the prevention measures that are taken to prevent the development of cervical cancer. Finally, diagnostic techniques will allow us to know what are the different methods or tests that are considered in different studies for cervical cancer diagnosis.

Figure- 3: Proposed framework for literature review

In summary, Table III shows the different studies found in the literature review according to the proposed taxonomy.

Table- III: Selected studies classification in the systematic review of the literature

<table>
<thead>
<tr>
<th>Classification</th>
<th>Source</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk Factors</td>
<td>[6,7,8,9,10]</td>
<td>5</td>
</tr>
<tr>
<td>Preventive Measures</td>
<td>[11,12,13,14,15]</td>
<td>5</td>
</tr>
<tr>
<td>Detection techniques</td>
<td>[16,17,18,19,20,21,22,23,24,25,26,27,28,29,30]</td>
<td>15</td>
</tr>
</tbody>
</table>

A. Risk Factors

Table IV shows the works related to risk factors that cause cervical cancer and their respective sources. The identified risk factors are HPV infection, sexual behavior, psychosocial, economic and cultural, health and reproduction, and quality of care.

Table- IV: Cervical cancer risk factors

<table>
<thead>
<tr>
<th>Risk Factors</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>HPV infection</td>
<td>[6,7]</td>
</tr>
<tr>
<td>Sexual conduct</td>
<td>[6,8]</td>
</tr>
<tr>
<td>Psychosocial, economic and cultural</td>
<td>[6,8]</td>
</tr>
<tr>
<td>Reproduction and health</td>
<td>[6,9]</td>
</tr>
<tr>
<td>Quality of care</td>
<td>[6,10]</td>
</tr>
</tbody>
</table>
• **HPV Infection:** Stewart et al. [6], argues that there are different risk factors related to cervical cancer in Sub-Saharan Africa (SSA) (Eastern, Central, Southern and Western). To analyze the risk factors and determine which are the most relevant, they used the Mortality Incidence Ratio (IMR) and the database provided by GLOBOCAN 2012 from IARC. They were able to find that one of the most significant factors is the high prevalence of HIV in populations, and this affects a higher risk of contracting HPV, which is one of the most relevant factors for developing cervical cancer. According to Fani et al. [7], HPV is one of the causes of cervical neoplasia to cervical cancer progression. A review of articles was carried out, 19 articles of those were selected for analysis, indicating that HPV16 was the main factor causing cervical cancer cases in Iranian women.

• **Sexual behavior:** Stewart et al. [6], argues that another significant factor is related to the lack of condom use with a non-regular partner, this high incidence of unsafe sex is due to a lack of health education and insufficient government awareness and intervention programs. Sharma y Pattanshetty et al. [8], argues that the number of cancer cases varies throughout the world and, of these, undeveloped countries bear the greatest burden. Based on the results, they conclude that marital status (married), age at menarche (13-14 years), history of vaginal itching, age of first intercourse (<18 years), at least one abortion, parity (> 3) are important risk factors for cervical cancer.

• **Psychosocial, economic and cultural:** Stewart et al. [6], argues that another influential factor is rural populations, since they have little access to medical care, poorer health literacy and lack of cancer awareness, therefore, patients only present in a medical center when symptoms are considerable. Sharma y Pattanshetty et al. [8], adds that education is an important factor in relation to cervical cancer. In their statistical studies, an inversely proportional relationship between the level of study (literacy) and cervical cancer is presented.

• **Health and reproduction:** Stewart et al. [6], argues that the high parity factor is related to a higher frequency of unprotected sexual intercourse that results in a higher exposure to HPV and a higher risk of developing cervical cancer. Also, it is related to the lack of knowledge about health or access to contraceptives, due to an inefficient health system. High parity may also reflect gender inequality in access to education. According to Xu et al. [9], smoking and combined oral contraceptives were classified as carcinogenic to humans, according to IARC, and its evaluation of the evidence has shown a causal association between these agents and cervical cancer. His research asserts that, among Australian women aged 30 to 44 years, current users of hormonal contraceptives and current smokers had a higher risk of developing CIN 2/3, and a longer duration of use and a higher intensity of exposure lead to a greater increase in risk.

• **Quality of care:** Stewart et al. [6], argues that the lack of quality of care is related to the fact that governments do not provide their populations with a good quality in education and health, which affects the lack of screening tests and prevention programs. Zahras y Rustman [10], highlight the importance of neural networks to help doctors easily classify some risk factors for cervical cancer, leading to better quality of care.

### B. Preventive Measures

Table V shows the works related to preventive measures against the diagnosis of cervical cancer and their respective sources. The identified prevention measures are health, vaccination and early detection.

<table>
<thead>
<tr>
<th>Preventive measures</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health</td>
<td>[11]</td>
</tr>
<tr>
<td>Vaccine</td>
<td>[12,13,14]</td>
</tr>
<tr>
<td>Early detection</td>
<td>[15]</td>
</tr>
</tbody>
</table>

• **Health:** Okunade et al. [11], argues how trace elements like zinc, selenium, and copper can help prevent cervical cancer. The study was carried out at the Teaching Hospital of the University of Lagos, where the results obtained reflected that zinc and selenium levels were low in patients with cervical cancer, compared with control patients, but that the copper levels were not relevant. The authors suggest that zinc and selenium supplements can prevent the occurrence of cervical cancer.

• **Vaccine:** Smith et al. [12], conducts research on the implementation of the New Zealand National Cervical Detection Program (NCSP) to determine whether the changes being proposed for detection and prevention benefit Maori women and other women to the same extent. The authors adopted the comprehensive model, Policy1-Cervix, in which they were able to analyze that the disparity continues to occur, but these HPV vaccination measures greatly favor the prevention of cervical cancer. Almazrou et al. [13] argues that after analyzing after evaluating the knowledge of the doctors of the Rey Abdul-Aziz Medical City regarding cervical cancer and the vaccine, they obtained that 98% of the doctors knew of cervical cancer, but almost half did not consider it fatal. But overall, 61% of doctors showed a good understanding of cervical cancer, and that most doctors recommended the HPV vaccine as a preventive measure for cervical cancer. Sankaranarayanan et al. [14], argues that two recombinant HPV vaccines are available that contain virus-like particles (VLPs). Both vaccines have remarkable immunogenicity and protection capable of preventing 70% of cervical cancers. Among the conclusions, the author highlights that the WHO currently recommends a two-dose HPV vaccination program for girls with a minimum interval of six months between doses.

• **Early detection:** Firmino-Machado et al. [15], analyzes a low-cost strategy for the prevention of cervical cancer, which involves making invitations through text messages and phone calls, which can increase adherence by 15%, and therefore they propose to evaluate the effectiveness of this strategy. The study was carried out in Portugal. The authors with this study demonstrated that a gradual invitation to perform early detection is more effective than standard care.

C. Detection Techniques

Table VI shows the works related to cervical cancer detection techniques and their respective sources. Detection techniques identified are Pap Smear, Colposcopy, and Biopsy.

<table>
<thead>
<tr>
<th>Detection methods</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canny edge detector</td>
<td>[16]</td>
</tr>
<tr>
<td>Clustering algorithms</td>
<td>[17, 18, 19]</td>
</tr>
<tr>
<td>Classification algorithms</td>
<td>[18, 19, 20, 21, 22, 23]</td>
</tr>
<tr>
<td>Deep learning</td>
<td>[22, 23, 24, 25, 26]</td>
</tr>
<tr>
<td>Others</td>
<td>[27, 28, 29, 30]</td>
</tr>
</tbody>
</table>

- **Canny edge detector**: Mustafa et al. [16], proposes to improve the detection of cancerous tissues from normal tissues by applying the Canny edge detector and using standard low-resolution images. After pre-processing the images, they used an algorithm in Java, where they were able to successfully identify normal and abnormal cervixes, with an accuracy of 90%.

- **Clustering algorithms**: Kaaviya et al. [17], uses the Fuzzy C-means (FCM) clustering algorithm for the segmentation of individual cells. They used this algorithm to segment Pap test images, where they adequately identified each component of the cell (nucleus, cytoplasm, and non-cellular component). Kuko and Pourhomayoun [18] use techniques to extract, segment, and classify abnormalities in cervical cells using images obtained from Pap tests. To achieve this, they applied the K-means algorithm and the Watershed algorithm. Bhuvaneshwari and Poornima et al. [19], they use the Fuzzy C-means clustering algorithm for the segmentation of the Pap test images. They managed to separate the nucleus and cytoplasm of the cell.

- **Classification algorithms**: Kuko and Pourhomayoun [18], applied the random forest model to classify cells as abnormal and normal, where they obtained an accuracy of 90.37% with a sensitivity and specificity of 96.33% and 83.59% respectively. Bhuvaneshwari and Poornima et al. [19], they used the KNN algorithm and used MATLAB 2017 to classify cells. In their training and testing they obtained an accuracy of 95%, and they conclude that this detection is useful for the pathologist to effectively give a diagnosis and treatment. Nehra et al. [20], proposes a method of detecting and classifying images as cancerous or non-cancerous. After applying the scale on the images obtained by colposcopy, the Gray Level Coexistence Matrix (GLCM) was constructed. The extracted features were used to classify images using Support Vector Machines (SVM). Analysis showed that the linear core function provided the highest accuracy of 96.61%. D. Kashyap et al. [21], propose a method to classify Papanicolaou test images, where they perform the extraction of geometric and texture characteristics. They use PCA for feature selection. Then they proceed to the classification where they applied the polynomial SVM technique, obtaining an accuracy of 95%. Hyeon et al. [22] designs and proposes a model that automatically classifies normal and abnormal states of cervical cells from microscopic images using a convolutional neural network. As a result, the support vector machine showed the best performance with 78% performance. Rohmatillah et al. [23] provides a method of classifying cervical cells. The method consists of three stages: feature extraction using CNN, feature reduction using PCA and LDA, and classification using SVM and DNN with softmax as output activation function. In the case of SVM, the results obtained are 95.8%, 99.3% and 99.1% in precision, sensitivity and specificity respectively.

- **Deep learning**: Hyeon et al. [22], as mentioned above, automatically classifies normal and abnormal states of cervical cells using VGGNet-16, which is a convolutional neural network (CNN) model previously trained in the ImageNet Large Scale Visual Recognition Challenge, and various classifiers machine learning. Rohmatillah et al. [23], previously mentioned, analyzes a second classifier based on a DNN, for this case it obtains a precision, sensitivity and specificity of 94.8%, 97.53% and 99.38% respectively. Sharma et al. [24], proposes an algorithm of segmentation and automatic classification. Preprocessing is accomplished by improving the contrast of the image. Then, the core of the images is segmented and 22 image texture characteristics are calculated using the gray level matched matrix (GLCM). Then, the characteristics obtained are passed through a neural network to classify the images in their respective category. The neural network is made using the backpropagation algorithm with 2 hidden layers. The precision achieved is 92%. Ghoneim et al. [25], apply convolutional neural networks for the classification of images of the Pap test, and the diagnosis of cervical cancer. They inserted an ELM-based classifier, and finally they introduced an autoencoder (AE)-based classifier. For the research they applied three CNN models, (1) shallow architecture, (2) VGG16 Net, (3) CaffeNet. The best results were obtained with the ELM classifier since it offered an accuracy of 99.7% for the two-class classification and an accuracy of 97.2% for the seven-class classification. Devi et al. [26], argues that artificial neural networks are specifically used in many medical applications with precision in performance results.

- **Others**: Makkonen et al. [17], argues that any Pap test reduces the risk of cervical cancer in a period of 5 years before diagnosis among women aged 35 to 39, but at younger ages the effect is small. Wentzsen et al. [28], argues that taking multiple targeted biopsies during colposcopy improves detection of prevalent pre-cancers. In their research, they demonstrated that better identification of cervical pre-cancers can be achieved in colposcopy through risk assessment based on baseline cytology and HPV tests, as well as colposcopic visual examination. Jagtap et. al [29], uses biopsy images and a statistical method for the detection of early-stage tumors, and their quantitative classification in cancerous grades. For the analysis of the images they used correlation methods and statistical moments. Also, they used a classification algorithm to compare the values of the statistical moments, with which they achieved a sensitivity of 100% and a specificity of 100% for the classification between the pre-cancer stages (CIN) and the section of normal tissue.
Song et al. [30], evaluates the efficacy of random biopsy in the diagnosis of high-grade squamous intraepithelial lesions or carcinomas (HSIL +) omitted by colposcopy-directed biopsy. In conclusion, he argues that random biopsy is not effective in the negative quadrant in women with positive colposcopy, but should be performed in women with cytological HSIL + but negative colposcopy, or in those with cytological LSIL or HGSI + and positive HPV but negative colposcopy.

IV. ANALYSIS OF RESULTS

The analysis and the results rescued from the review are detailed below.

A. Risk Factors (Q1)

According to the results obtained in the systematic analysis of the literature, 5 works refer to the “risk factors” that cause cervical cancer, which represent 20% (See Table 1) of the total number of works reviewed. Furthermore, the work carried out by Stewart covers all the risk factors that have been analyzed in the systematic review of the literature, since it covers different risk factors that cause cervical cancer [6]. HPV infection, behavior sexual, psychosocial, economic and cultural, health and reproduction; and quality of care. Also, we can note that artificial neural networks help to identify risk factors, for instance, study [10].

B. Preventive Measures (Q2)

In the analysis of the literature, 5 related works were obtained regarding “prevention measures” for cervical cancer, which represent 20% of the total number of works reviewed (See Table 1), where the prevention measure The most used is the "vaccine" of the human papilloma virus, in the works [12, 13, 14], since it is related to the factor that is most influential which is the Human Papilloma Virus (HPV), for example in [13], they found that most doctors recommend the vaccine as a preventive measure.

C. Detection Techniques (Q3)

According to the results obtained in the systematic literature analysis about cervical cancer detection techniques, they represent 60% of the total number of works reviewed (See Table 1). Also, we can see that most of the works that apply artificial intelligence combine grouping, classification and deep learning techniques or algorithms, which cover the preprocessing steps, characteristic extraction, characteristics selection and classification, helping cervical cancer diagnosis in the works [17 - 26], such is the case of the work [25], which applies the deep learning technique, obtaining an accuracy of 99.7%.

V. CONCLUSION

In this article, we presented a systematic literature review of 36010 articles related to cervical cancer, in which the summary of 725 articles were reviewed and 25 of those reviewed were obtained. The articles were examined based on the framework proposed in Figure 3, where the analysis of the research findings was related to 3 research questions indicated in section 2. Table 3 presents the analyzed information and how they were divided by the means of the research questions. It is observed that most of the studies are related to "Detection techniques" and few others related to "Risk factors" or "Prevention measures". Additionally, the study shows a correlation between the presented detection techniques, such as the grouping, classification and deep learning algorithm techniques, which are all categorized under machine learning. As a result, it is recommended that further investigation in these techniques be promoted as they provide positive outcomes by assisting experts in cervical cancer detection. Likewise, risk factors could be added to these techniques for a greater scope in detection, for instance HPV infection, as it is one of the most relevant factors for developing cervical cancer [6]. Moreover, it is recommended to invest in researching preventive measures for cervical cancer through effective and efficient means, since it is one of the most preventable cancers.

REFERENCES


