

# Cervical Cancer Screening

## Summary

This dataset is composed of responses from 858 patients and 36 variables focusing on the prediction of indicators or diagnosis of cervical cancer. The dataset provides demographic information, habits, and historic medical records of the 858 patients from Hospital Universitario de Caracas in Caracas, Venezuela. A number of the patients did not answer some of the questions due to privacy concerns.

## Key Facts

<b>Date Created</b>	2017-03-03
<b>Date Modified</b>	2017-03-03
<b>Version</b>	2017-03-03
<b>Update Frequency</b>	Never
<b>Complexity</b>	Simple
<b>Temporal Coverage</b>	2017
<b>Spatial Coverage</b>	Caracas, Venezuela
<b>Source</b>	UCI Machine Learning Repository; Universidad Central de Venezuela, Caracas, Venezuela; INESC TEC & FEUP, Porto, Portugal
<b>Source License URL</b>	N/A
<b>Source License Requirements</b>	N/A
<b>Source Citation</b>	Kelwin Fernandes, Jaime S. Cardoso, and Jessica Fernandes. 'Transfer Learning with Partial Observability Applied to Cervical Cancer Screening.' Iberian Conference on Pattern Recognition and Image Analysis. Springer International Publishing, 2017.
<b>Keywords</b>	Cervical Cancer, Cervical Cancer Symptoms, Signs of Cervical Cancer, Causes of Cervical Cancer, Cervical Cancer Risk Factors, Sexually Transmitted Diseases, Acquired Immune Deficiency Syndrome, Human Immunodeficiency Virus, HIV/AIDS, Human Papillomavirus

## Other Titles and Uses

- Risk Factors for Cervical Cancer
- Predication of Indicators of Cervical Cancer
- Diagnosis of Cervical Cancer

## Description

Despite the possibility of prevention with regular cytological screening, cervical cancer is one of the significant causes of mortality in low-income countries killing more than a quarter of a million cases per year. This is because resources are very limited and patients have poor adherence to routine screening due to lack of awareness. In addition, prediction of individual patient's risk and best screening strategy during diagnosis has become a challenge with the existence of several diagnostic methods and physician's subjective preferences, usually based on expertise and comfort. Hence, prediction of cervical cancer using automated methods or computed aided diagnosis (CAD) system would require data from each source - modality and expertise.

This study was conducted to create a predictive model of transfer learning (TL) from one source to another, such as modality to an expert, in order to accurately predict risk for cervical cancer and consequently diagnose cervical cancer among patients.

## Schema

Field Name	Type	Description	Properties
<b>Age_of_Respondents</b>	Integer	A featured risk factor for cervical cancer, this represents the age of patients from Hospital Universitario de Caracas who responded to the questions on demographic information, habits, and historic medical records; some did not answer some questions due to privacy	Level: Ratio
<b>Number_of_Sexual_Partners</b>	Integer	Number of sexual partners as a featured risk factor for cervical cancer	Level: Ratio
<b>First_Sexual_Intercourse</b>	Integer	Age of first sexual intercourse as a featured risk factor for cervical cancer	Level: Ratio
<b>Number_of_Pregnancies</b>	Integer	Number of pregnancies as a featured risk factor for cervical cancer	Level: Ratio
<b>Is_Smoking</b>	Boolean	Smoking as a featured risk factor for cervical cancer; answers whether patient is smoking or not	
<b>Smoking_in_Years</b>	Number	Length of smoking in years as featured risk factor for cervical cancer	Level: Ratio



Field Name	Type	Description	Properties
<b>Smoking_in_Packs_per_Year</b>	Number	Number of cigarette packs consumed per year of smoking as a featured risk factor for cervical cancer	Level: Ratio
<b>Is_On_Hormonal_Contraceptives</b>	Boolean	Use of hormonal contraceptive as a featured risk factor for cervical cancer; answers whether patient is on contraceptives or not	
<b>Hormonal_Contraceptives_in_Years</b>	Number	Number of years on hormonal contraceptives as a featured risk factor for cervical cancer	Level: Ratio
<b>Is_On_IUD</b>	Boolean	Use of intrauterine device (IUD) as a featured risk factor for cervical cancer; answers whether patient is on IUD	
<b>IUD_in_Years</b>	Number	Number of years on IUD as a featured risk factor for cervical cancer	Level: Ratio
<b>Is_Diagnosed_with_STDs</b>	Boolean	Patient diagnosis of STDs as a featured risk factor for cervical cancer; answers whether patient has been diagnosed with STD	
<b>Number_of_Years_with_STDs</b>	Integer	Number of years with STDs acquired as featured risk factor for cervical cancer	Level: Ratio

Field Name	Type	Description	Properties
<b>Is_STD_Condylomatosis</b>	Boolean	If STD is categorized as condylomatosis	
<b>Is_STD_Cervical_Condylomatosis</b>	Boolean	If STD is categorized as cervical condylomatosis	
<b>Is_STD_Vaginal_Condylomatosis</b>	Boolean	If STD is categorized as vaginal condylomatosis	
<b>Is_STD_Vulvoperineal_Condylomatosis</b>	Boolean	If STD is categorized as vulvoperineal condylomatosis	
<b>Is_STD_Syphilis</b>	Boolean	If STD is categorized as syphilis	
<b>Is_STD_Pelvic_Inflammatory_Disease</b>	Boolean	If STD is categorized as inflammatory disease	
<b>Is_STD_Genital_Herpes</b>	Boolean	If STD is categorized as genital herpes	
<b>Is_STD_Molluscum_Contagiosum</b>	Boolean	If STD is categorized as molluscum contagiosum	
<b>Is_STD_AIDS</b>	Boolean	If STD is categorized as Acquired Immune Deficiency Syndrome (AIDS)	
<b>Is_STD_HIV</b>	Boolean	If STD is categorized as Human Immunodeficiency Virus (HIV)	
<b>Is_STD_Hepatitis_B</b>	Boolean	If STD is categorized as hepatitis B	
<b>Is_STD_HPVP</b>	Boolean	If STD is categorized as Human Papillomavirus (HPV)	
<b>Number_of_STD_Diagnosis</b>	Integer	Number of STDs diagnosed as a featured risk factor for cervical cancer	Level: Ratio

Field Name	Type	Description	Properties
Time_Since_First_STD_Diagnosis	Integer	Time since first STD diagnosis	Level: Ratio
Time_Since_Last_STD_Diagnosis	Integer	Time since last STD diagnosis	Level: Ratio
Is_Diagnosis_Cancer	Boolean	If patient is diagnosed with cancer or no	
Is_Diagnosis_CIN	Boolean	If patient is diagnosed with cervical intraepithelial neoplasia (CIN) or no	
Is_Diagnosis_HPVP	Boolean	If patient is diagnosed with human papillomavirus (HPV) or no	
Is_Diagnosed	Boolean	If patient is diagnosed with	
Is_Screening_Hinselmann	Boolean	If screening strategy used to predict the patient's risk of cervical cancer is colposcopy using acetic acid done	
Is_Screening_Schiller	Boolean	If screening strategy used to predict the patient's risk of cervical cancer is colposcopy using Lugol iodine	
Is_Screening_Cytology	Boolean	If screening used to predict the patient's risk of cervical cancer is Cytology	
Is_Screening_Biopsy	Boolean	If screening used to predict the patient's risk of cervical cancer is Biopsy	

## Sample Records

Field Name	Sample 1	Sample 2	Sample 3
Age_of_Respondents	30	22	18
Number_of_Sexual_Partners	3	2	3
First_Sexual_Intercourse	16	17	16
Number_of_Pregnancies	3	1	1
Is_Smoking			
Smoking_in_Years			
Smoking_in_Packs_per_Year			
Is_On_Hormonal_Contraceptives			
Hormonal_Contraceptives_in_Years			
Is_On_IUD			
IUD_in_Years			
Is_Diagnosed_with_STDs			
Number_of_Years_with_STDs			
Is_STD_Condylomatosis			
Is_STD_Cervical_Condylomatosis			
Is_STD_Vaginal_Condylomatosis			
Is_STD_Vulvoperineal_Condylomatosis			
Is_STD_Syphilis			
Is_STD_Pelvic_Inflammatory_Disease			
Is_STD_Genital_Herpes			
Is_STD_Molluscum_Contagiosum			
Is_STD_AIDS			
Is_STD_HIV			
Is_STD_Hepatitis_B			
Is_STD_HPVP			
Number_of_STD_Diagnosis			
Time_Since_First_STD_Diagnosis			
Time_Since_Last_STD_Diagnosis			
Is_Diagnosis_Cancer	false	false	false
Is_Diagnosis_CIN	false	false	false

Field Name	Sample 1	Sample 2	Sample 3
Is_Diagnosis_HPVP	false	false	false
Is_Diagnosed	false	false	false
Is_Screening_Hinselmann	false	false	false
Is_Screening_Schiller	false	false	false
Is_Screening_Cytology	false	false	false
Is_Screening_Biopsy	false	false	false