



## Cervical cancer (Risk Factors)

Donated on 2017-03-03

319 views

0 citations

## General Information

### Abstract

This dataset focuses on the prediction of indicators/diagnosis of cervical cancer. The features cover demographic information, habits, and historic medical records.

## Quick Facts

### Dataset Characteristics

Multivariate

### Associated Tasks

Classification

### Subject Area

Life

### DOI

None

**# of Instances**

858

**# of Views**

319 views

**Creators**

Kelwin Fernandes

Jaime Cardoso

Jessica Fernandes

**Descriptive Questions****For what purpose was the dataset created?**

N/A

**Who funded the creation of the dataset?**

N/A

**What do the instances that comprise the dataset represent?**

(int) Age (int) Number of sexual partners (int) First sexual intercourse (age) (int) Num of pregnancies (bool) Smokes (bool) Smokes (years) (bool) Smokes (packs/year)  
(bool) Hormonal Contraceptives (int) Hormonal Contraceptives (years) (bool) IUD  
(int) IUD (years) (bool) STDs (int) STDs (number) (bool) STDs:condylomatisis (bool)  
STDs:cervical condylomatisis (bool) STDs:vaginal condylomatisis (bool)  
STDs:vulvo-perineal condylomatisis (bool) STDs:syphilis (bool) STDs:pelvic  
inflammatory disease (bool) STDs:genital herpes (bool) STDs:molluscum  
contagiosum (bool) STDs:AIDS (bool) STDs:HIV (bool) STDs:Hepatitis B (bool)  
STDs:HPV (int) STDs: Number of diagnosis (int) STDs: Time since first diagnosis  
(int) STDs: Time since last diagnosis (bool) Dx:Cancer (bool) Dx:CIN (bool) Dx:HPV  
(bool) Dx (bool) Hinselmann: target variable (bool) Schiller: target variable (bool)  
Cytology: target variable (bool) Biopsy: target variable

**Are there recommended data splits?**

N/A

**Does the dataset contain data that might be considered sensitive in any way?**

N/A

**Was there any data preprocessing performed?**

N/A

**Has the dataset been used for any tasks already?**

N/A

## Additional Information

The dataset was collected at 'Hospital Universitario de Caracas' in Caracas, Venezuela. The dataset comprises demographic information, habits, and historic medical records of 858 patients. Several patients decided not to answer some of the questions because of privacy concerns (missing values).

## Citation Requests/Acknowledgements

N/A

## Tabular Data Properties



**Does this dataset contain missing values?**

True

**What symbol is used to indicate missing data?**

1

**Number of Attributes**

36

## Features



Attribute Name	Role	Type	Description	Units	Missing Values
----------------	------	------	-------------	-------	----------------

Attribute Name	Role	Type	Description	Units	Missing Values
	N/A	N/A	N/A	N/A	False
	N/A	N/A	N/A	N/A	False
	N/A	N/A	N/A	N/A	False
	N/A	N/A	N/A	N/A	False
	N/A	N/A	N/A	N/A	False

## Papers Citing This Dataset



N/A

## Keywords

N/A

## License

This dataset is licensed under a [Creative Commons Attribution 4.0 International \(CC BY 4.0\)](#) license.

This allows for the sharing and adaptation of the datasets for any purpose, provided that the appropriate credit is given.

### UC Irvine Machine Learning Repository

Supported by National Science Foundation

Contact: [ml-repository@ics.uci.edu](mailto:ml-repository@ics.uci.edu)

[Make a Feature Request or Bug Report](#)

[Home](#)   [Datasets](#)   [Donate a Dataset](#)   [About Us](#)   [Contact Us](#)   [CML](#)