

## EARLY PREDICTION OF CERVICAL CANCER USING DATA MINING TECHNIQUES

*K. Gomathi<sup>1</sup>, Dr. D. Shanmuga Priyaa<sup>2</sup>*

### ABSTRACT

Cancer is the one of the major problem in the World and also in Medical Industry. Cancer can start in any part of the body and it can spread to other parts also. It is uncontrollable and it has many types. This study applied the data mining techniques in cervical cancer data for prediction earlier stage. The pre-processed data set consists of 850 records, which have all the available 36 fields from the dataset. This paper mainly focused on comparison various data mining algorithms based on Accuracy, Recall, F-measure and Precision.

**Keywords :** Data Mining, Cervical cancer, Naive Bayes, MLP, Decision tree.

### I. INTRODUCTION

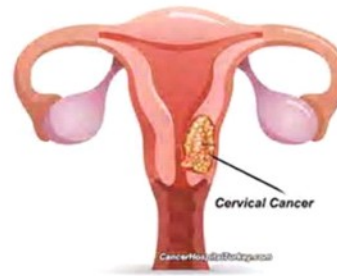
Cervical cancer is the one of the common cancer among women in the world wide. The number of peoples is made a diagnosis with cervical cancer every year. In recent survey about cervical cancer that 12,820 cases of invasive cervical cancer will be identified and an estimated 4,210 deaths from cervical cancer will occur in 2017[1].

---

<sup>1</sup>Research Scholar, Department of Computer Science, Karpagam Academy of Higher Education, Karpagam University, Coimbatore

<sup>2</sup>Professor, Department of Information Technology, Karpagam Academy of Higher Education, Karpagam University, Coimbatore

Cervical cancer is a one type of cancer that occurs in the cells of the cervix — the lower part of the uterus that connects to the vagina. Various strains of the human papillomavirus (HPV), a sexually transmitted infection, play a function in causing most cervical cancer [2].



### II. CAUSES OF CERVICAL CANCER

- Infect by Human papillomavirus (HPV), it can spread though illegal sexual contacts.
- Women are who taking birth control pills, and early sexual contact.
- HPV virus can cause abnormal growth of cervical cells.
- Another reason for causing cervical cancer is Smoking- it can develop the growth of cancer cells.
- Regular pelvic examination and Pap smear test can detect precancerous in the cervix

- Precancerous in the cervix can be treated with cryosurgery, cauterization and laser surgery.

#### Symptoms of Cervical Cancer

Cervical cancer symptoms are not caused by the Precancerous cervical cell and early cancer of cervix [3]. So, the Pap Screening and HPV test can only help to identify the cancerous cell and the development of cervical cancer.

Common Symptoms of cervical cancer disease are abnormal and irregular vaginal bleeding, ache during sexual contact and pain in vaginal discharge.

- Abnormal bleeding in time of
  - Menstrual time
  - During Sexual Contact
  - After Washing of Vaginal
  - After a pelvic examination
  - After menopause period
- Pelvic pain is not related to menstrual time.
- Heavy or abnormal discharge to may be dilute or bulky and it have a foul smell.
- Frequent urination and also pain during pass of urine

#### Risk Factors of Cervical Cancer [4]:

**HPV (Human papillomavirus)** – During sexual contact no. of virus can be transmit; few no's of virus can cause cervical cancer.

**Many sexual partners** - Irregular sexual contact.

**Early sexual activity**- Sexually arrested in childhood.

**Smoking** – Smoking can increase the growth of disease cells and it can develop cancer.

**A weakened immune system** -transplant recipients taking immunosuppressive medicines.

**Long-term mental stress** - women who experience high levels of stress over a sustained period may be less able to fight off HPV.

**Giving birth at a very young age** - women who gave birth before the age of 17 are much more to develop cervical cancer.

**Several pregnancies** - women who have more than three children in separate pregnancies are more to grow cervical cancer.

**Contraceptive pill** - long-term use of some common contraceptive pills slightly increases a woman's risk.

**Other sexually transmitted infections (STIs)**. Having other STIs — such as chlamydia, gonorrhea, syphilis and HIV/AIDS — increases your risk of HPV.

**Socio-economic status** - studies in several countries have revealed that women in deprived areas have significantly higher rates of cervical cancer.

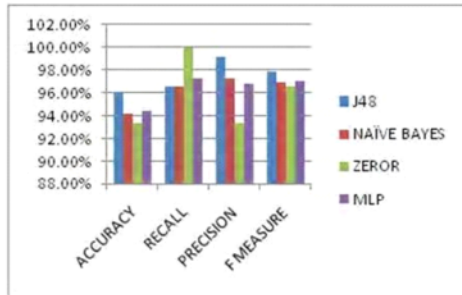
Signs of advanced cervical cancer include [5]:

- Weight loss
- Fatigue

- Back pain
- Leg pain or swelling
- Leakage of urine or feces from the vagina
- Bone fractures

**III. EXPERIMENTATION**

In this work, prediction of Cancer cervix is analyzed by Data Mining [6]. It plays an important role of predictions in medical field. Cervical Cancer predicted by using of Naïve Bayes, J48, ZeroR, MultiLayerPerceptron algorithms. Collection of data from University Hospital of Caracas in our work, the data set contains 800 records and 36 attributes based on age, habits, and historic medical records and etc. Based on biopsy attribute, Comparison of the performance of various algorithms was used under the techniques in terms of Precision, Recall, F-measure and Accuracy [7] to determine the best predictor for the cervical cancer.



**Accuracy**

Accuracy can calculate by the number of correct predictions and it can divided by the total number

of dataset. The best accuracy value is 1.0, worst value is 0.0.

$$ACC = \frac{TP+TN}{TP+TN+FN+FP} = \frac{TP+TN}{P+N}$$

**Recall (Or) Sensitivity**

Recall or Sensitivity is calculated as the no. of correct positive predictions divide by the total number of positives Prediction. It is called as recall (REC) or true positive rate (TPR). The best sensitivity value is 1.0, the worst is 0.0.

$$SN = \frac{TP}{TP+FN} = \frac{TP}{P}$$

**Precision**

Precision is calculated with the number of correct positive predictions and it can divide by the total number of positive predictions. It is known as positive predictive value (PPV). The best precision value is 1.0, but the worst case is 0.0.

$$PREC = \frac{TP}{TP+FN}$$

**F measure**

The F measures can measure the accuracy values and is defined as the weighted harmonic mean of the precision and recall of the test. It has another name is F1 score or F scores [8]

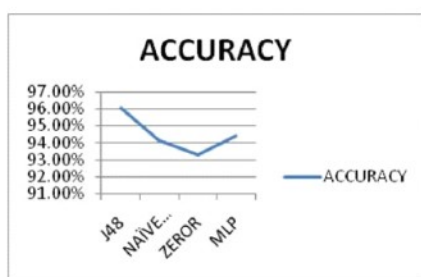
$$F_1 = \frac{2 \cdot PREC \cdot REC}{PREC + REC}$$

ALGORITHM	ACCURACY	RECALL	PRECISION	F MEASURE
J48	96.03%	0.965	0.992	0.978
NAÏVE BAYES	94.17%	0.965	0.972	0.969
ZEROR	93.30%	1	0.933	0.965
MLP	94.42%	0.972	0.968	0.97

Table 1

**RESULTS**

In this study, the accuracy of various data mining techniques is compared. The above result shows that J48 has highest accuracy compared than other techniques.



**IV. CONCLUSION**

In this study, we have taken various classification methods and it's compared the results of various algorithms on weka datamining tool, basis of accuracy, sensitivity, recall and precision. According to the table 1, we can determine that J48 has the highest accuracy percentage of 96.03. So J48 algorithm is better than other algorithms.

**REFERENCES**

1. <https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2017/cancer-facts-and-figures-2017.pdf>
2. <http://www.mayoclinic.org/diseases-conditions/cervical-cancer/home/ovc-20210887>
3. <http://www.nccc-online.org/hpvcervical-cancer/cervical-cancer-overview/>
4. <https://www.medicalnewstoday.com/articles/159821.php>
5. <http://www.cancercenter.com/cervical-cancer/symptoms>
6. R. Vidya , G. M. Nasira, "Prediction of Cervical Cancer using Hybrid Induction Technique: A Solution for Human Hereditary Disease Patterns", Indian Journal of Science and Technology, Volume 9, Issue 30, August 2016
7. <https://classeval.wordpress.com/introduction/basic-evaluation-measures/>
8. [https://metacademy.org/graphs/concepts/f\\_measure](https://metacademy.org/graphs/concepts/f_measure)