

Prediction of Cervical Cancer using Machine Learning Framework

*Barnali Dey, Shreya Bhaduri, Anjali Sharma, Megha Baitha, Swikriti Debnat, Nandita Sarkar
Annwasha Banerjee Majumder*

JIS College of Engineering, Kalyani, West Bengal, India

annwasha.banerjee@jiscollege.ac.in

Available online at: <http://jacsai.org/>

Abstract— Cervical cancer is the one of common cancer that leads to death of many women across the world. It caused by papillomaviruses (HPV) virus. Some of the cases no symptoms are being observed. Early detection of any disease can save life. In case of cervical cancer early detection is very much required. Machine learning can helps in these aspects. In this paper we have proposed a method to detect the risk of cervical cancer using K- Nearest neighbor and Decision Tree techniques. The dataset used are collected from Kaggle . Accuracy achieved are 88% and 92%.

Keywords— Cervical Cancer, Decision Tree, Machine Learning, K Nearest neighbor

I. INTRODUCTION

Cervical cancer one of the most common cancer in women. According to report of World Health Organization, 311 000 women died from the disease out of these cancer.[1]. All women have the risk of cervical cancer and women at age 30s are high risk. The main cause of cervical cancer is the long last infection due to human papillomavirus (HPV). Early detection cervical cancer can save lives Here the machine learning is playing the role. There are different machine learning methods using which early disease detection is possible. In this paper a model has been proposed to detect the risk of cervical cancer at an early stage using K Nearest Neighbor and Decision Tree.

- A. **Cervical Cancer** –It is a type of cancer occurs at cell of cervix. Cervix is the connecting unit of vagina and uterus.
- B. **K Nearest Neighbor- K Nearest Neighbor** is the simplest supervised learning mechanism. It can be used both classification and regression. Here any object is classified based on the class of K number of nearest neighbors.
- C. **Decision Tree- Decision Tree** is also a very efficient supervised learning mechanism used for classification and prediction. It is based on flowchart like structure.

II. RELATED WORK

In [3] authors have proposed a Bayes Net algorithm which has achieved accuracy of 96.38%.In [4] Decision Tree, Logistic Regression and Random Forest model has proposed. A multi-layer perceptron based model to predict cervical cancer has been proposed [5]. In[6] authors proposed a model to predict cervical cancer considering factors sexually transmitted disease (STDs), intra-uterine device (IUD), hormonal contraceptives and the age at which first sexual intercourse happens.

In [7] a deep neural network based model has been proposed to predict cervical cancer. In [8] a model has proposed to identify abnormalities of cervix. In [9] another model has been proposed using ensemble mechanism.. In [10] authors have proposed a model to predict cervical cancer using decision tree classifier. In[11] a machine learning based model to predict cervical cancer has been proposed using Long Short Term Memory and Random forest.

III. METHODOLOGY

In this paper a machine learning based cervical cancer prediction model has been proposed using K Nearest Neighbor and Decision Tree approaches. By applying the KNN and Decision tree accuracy achieves are :88% and 92%.

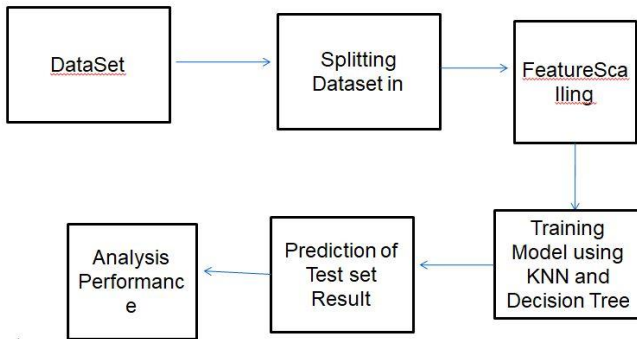


Fig1: Model Flow Diagram

Date set for the training has been collected from Kaggle. The factors are considered are as : Age, Number of sexual partners, First sexual intercourse, Num of pregnancies, Smokes, Smokes (years), Smokes (packs/year), Hormonal Contraceptives, Hormonal Contraceptives (years), IUD, IUD (years), STDs, STDs (number), STDs: condylomatosis, STDs: cervical condylomatosis, STDs: vaginal condylomatosis, STDs: vulvo-perineal condylomatosis, STDs: syphilis, STDs: pelvic inflammatory disease, STDs: genital herpes, STDs: molluscum contagiosum, STDs: AIDS, STDs: HIV, STDs: Hepatitis B, STDs: HPV, STDs: Number of diagnosis, Dx: Cancer, Dx: CIND, Dx: HPV, Dx: Hinselmann, Schiller, Citology. Following figure describes the dataset.

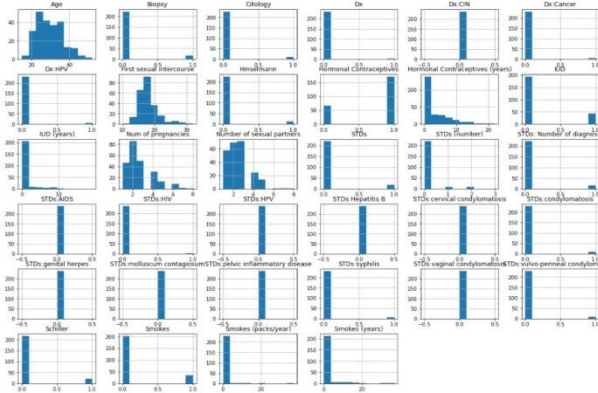


Fig 2: Dataset Description.

IV. RESULTS AND DISCUSSION

In this proposed model good level of accuracy achieved in KNN and Decision Tree also, Decision Tree provides better performances than KNN. Following figures show the performances of both the technique.

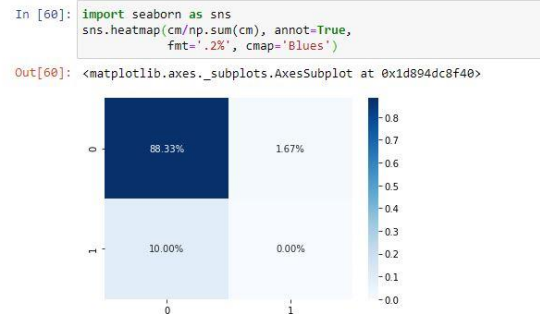


Fig 3: Confusion Matrix of KNN

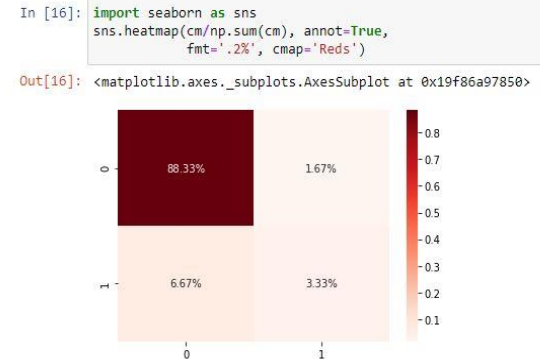


Fig 4: Confusion Matrix Decision Tree

V. CONCLUSION AND FUTURE SCOPE

Cervical cancer prediction model has been proposed in this paper using K Nearest neighbor and Decision tree. Decision tree has achieved better performance than the KNN. This model would help to predict cervical cancer at an early stage.

References

- [1] https://www.who.int/health-topics/cervical-cancer#tab=tab_1
- [2] https://www.cdc.gov/cancer/cervical/basic_info/index.htm
- [3] Ashay Singh, Ankur Singh Bist. (2020) OSX malware detection : Challenges and solutions. Journal of Information and Optimization Sciences 41:2, pages 379-385.
- [4] Riham Alsmariy1 , Graham Healy2 , Hoda Abdelhafez3 Predicting Cervical Cancer using Machine Learning Methods(IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 11, No. 7, 2020
- [5] Wenyng Yang, Xin Gou, Tongqing Xu, Xiping Yi, and Maohong Jiang. 2019. Cervical Cancer Risk Prediction Model and Analysis of Risk Factors based on Machine Learning. In Proceedings of the 2019 11th International Conference on Bioinformatics and Biomedical Technology (ICBBT'19). Association for Computing Machinery, New York, NY, USA, 50–54. DOI: <https://doi.org/10.1145/3340074.3340078>
- [6] S. F. Abdoh, M. Abo Rizka, and F. A. Maghraby, "Cervical cancer diagnosis using random forest classifier with SMOTE and feature reduction techniques," IEEE Access, vol. 6, pp. 59475–59485, 2018.
- [7] K. Rayavarapu and K. K. V. Krishna, "Prediction of Cervical Cancer using Voting and DNN Classifiers," Proceedings of the 2018

- International Conference on Current Trends towards Converging Technologies, ICCTCT 2018. 2018.
- [8] Daniëlle van der Waal, Ruud L.M. Bekkers, Stéfanie Dick, Charlotte H. Lenselink, Leon F.A.G. Massuger, Willem J.G. Melchers, Channa E. Schmeink, Albert G. Siebers, Mireille J.M. Broeders, Risk prediction of cervical abnormalities: The value of sociodemographic and lifestyle factors in addition to HPV status, Preventive Medicine, Volume 130,2020,
- [9] Alam, TalhaMahboob and Khan, Muhammad MilhanAfzal and Iqbal, Muhammad Atif and Abdul, Wahab and Mushtaq, Mubbashar, Cervical Cancer Prediction through Different Screening Methods Using Data Mining (October 23, 2019). (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 10, No. 2, 2019 , Available at SSRN: <https://ssrn.com/abstract=3474371>
- [10] JaswinderSingh1 ,Sandeep Sharma2Prediction of Cervical Cancer Using Machine Learning TechniquesInternational Journal of Applied Engineering Research ISSN 0973-4562 Volume 14, Number 11 (2019) pp. 2570-2577 © Research India Publications. <http://www.ripublication.com>

Authors Profile

B.Dey- She is B.Tech(IT) graduate from JIS College of Engineering. She is interested in Machine learning and Artificial Intekkgience.

ShreyaBhaduri

- She is B.Tech(IT) graduate from JIS College of Engineering. She is interested in Machine learning and Artificial Intekkgience.

A.Sharma- She is B.Tech(IT) graduate from JIS College of Engineering. She is interested in Machine learning and Artificial Intekkgience.

B.Dey- She is B.Tech(IT) graduate from JIS College of Engineering. She is interested in Machine learning and Artificial Intekkgience.

M.Baitha- She is B.Tech(IT) graduate from JIS College of Engineering. She is interested in Machine learning and Artificial Intekkgience.

S Debnath

She is B.Tech(IT) graduate from JIS College of Engineering. She is interested in Machine learning and Artificial Intekkgience.

N Sarkar

She is B.Tech(IT) graduate from JIS College of Engineering. She is interested in Machine learning and Artificial Intekkgience.

A.Banerjee

She is an Asst. Professor of department of Infromation Technology , JIS College of