IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 4, April 2016

Cervical Cancer stage prediction using Decision Tree approach of Machine Learning

Sunny Sharma

Research Scholar, Department of Computer Science, Guru Nanak Dev University, Amritsar, India

Abstract: Cervical cancer is the largest common cause of cancer deaths in women around the world. It affects the cervix in the female reproductive system which leads to death. To identify the stages of cervical cancer the decision trees are very helpful. They classify the stages of the cervical cancer and help the oncologist to detect the cancer. The proposed way use the data set obtained from (http://www.igcs.org) & predict the stages of cervical cancer using See5 tool.

Keywords: Cervical Cancer prediction; Machine Learning; See5, Decision Tree; C5.

I. INTRODUCTION

The body is made up of lots of living cells. Normal body cells grow, partition into new cells, and pass away in an orderly manner. Cancer begins when cells in a part of the body start to grow out of control. Cancer cell growth is diverse from normal cell growth. Instead of dying or pass away, cancer cells continue to grow and form new, unusual cells. Cells become cancer cells because of damage to DNA structure. We can say cancer is one of the syndrome in which the cells are partitioned & replicated in uncontrolled way. Cervical cancer is one of the most affecting cancers in women worldwide now these days. Its rate of occurrence is around 80% in low & middle income countries or in low socio-economic groups of countries & around 20% in higher income countries or in developed nations. The main problem with cervical cancer is that it cannot be detected as it doesn't show any symptoms until the final stages normally [5] [7].

The machine learning is the technique in which decision boundaries are explored. The Rule base mining technique and Decision trees plays vital role for decision making as well as helpful in machine learning. Another major advantage of using decision tree approach is the white box nature of this approach. It becomes easier to analyse and comprehend the course of decision making process.

1.1 Decision Trees

Decision Tree refers to hierarchical structure of the problem. it has root node& few other nodes. Rather than having incoming edges it has outgoing edges. Middle level nodes are called testing nodes & leaf nodes are called terminal nodes as well as decision nodes. The ability of Decision tree is that it can illustrate or depict the decision among different attributes. The objective is to predict the stage of cervical cancer stage target attribute based on several input variables [9][11].

1.2 C5 Algorithm

In 1987 Quinlan proposed C5.0 algorithm. It deals with the formation of decision tree using the selection of best optimized attribute from the given information with the help of the Max Gain method. Basically in C5 the maximum information gain is calculated using the formula.

Gain= I(X)-Remainder(X)

Where I(X) is the Information content & Remainder(X) is remainder information these are evaluated as:

$$I(X) = \sum_{m} P(a)^{*}-\log_2 P(a)$$

$$\text{Remainder}(X) = \sum_{i=1} q(i) I(\%Pi, \%Ni)$$

Where X is the set of different data samples, P(a) is the probability of occurrence of a value, a is the variable range, %Pi & %Ni are the fraction of positive & Negative examples on branch [13][14].

1.3 Stages of Cervical cancer

1.3.1 Stage 0-Carcinoma in Situ:

Stage 0 is carcinoma in situ i.e origin of abnormal cells in the inner-most lining of the cervix. They become cancer & affect nearby customary tissue.

1.3.2 Stage I

At this stage presence of cancer is in cervix only.

- Stage I-A: with the help of microscope cancer can be seen in cervix tissues. Further detail is expressed as.
  - Stage I-A (1): cancer depth is not more than 3 millimeters and it is not greater than 7 millimeters wide in tissues.
  - Stage I-A (2): cancer depth is greater than 3 but not greater than 5 millimeters, but not more than 7 millimeters wide.
- Stage I-B detail is expressed as.
  - Stage I-B (1): the depth is greater than 5 millimeters & greater than 7 millimeters wide.
  - Stage I-B (2): the not more then 4 centimeter wider cancer can be seen without the help of microscope.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 4, April 2016

1.3.3 Stage II

At this stage cancer has spread beyond the cervix but not towards the pelvic wall or towards the inferior third of the vagina. With deep penetration of cancer further detail can be expressed as.

- Stage II-A: Cancer increased from cervix towards the superior two third of the vagina, but not towards tissues of the uterus.
  - Stage II-A (1): without the help of microscope tumor of not more than 4 centimeters can be seen.
  - Stage II-A (2): without the help of microscope tumor of more than 4 centimeters can be seen.
- Stage IIB: Cancer stretched from cervix towards the tissues of the uterus.

1.3.4 Stage III

At this stage cancer stretched towards the inferior third of the vagina, Pelvic wall can cause kidney troubles. With deep penetration of cancer further detail can be expressed as.

- Stage III-A: Cancer has stretched towards the inferior third of the vagina but not towards pelvic wall.
- Stage III-B: Cancer has stretched towards the pelvic wall, tumor has become huge enough to chunk the ureters which increase the size of kidneys or can stop kidneys working.

1.3.5 Stage IV

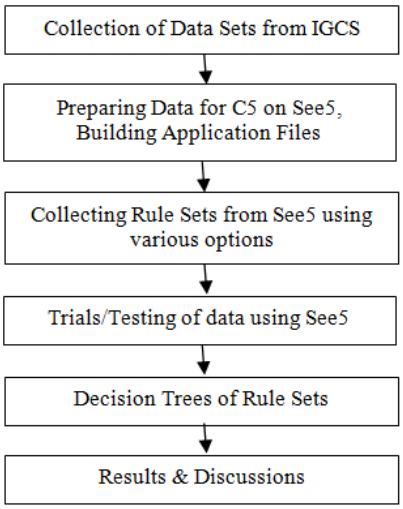
At this stage cancer stretched towards the bladder/rectum, or other parts of the body.

- Stage IV-A: Cancer stretched towards nearly connected organs like rectum/bladder.
- Stage IV-B: Cancer stretched towards other parts of the body like liver/lungs/bones/distant lymph nodes [3]
- Common types of treatments for cervical cancer are as follow: Surgery/Radiation therapy/ Chemotherapy & combination of them.

II. LITERATURE SURVEY

In 2009 **A. Satija et al.** revealed the statistics of cervical cancer in India. It is primarily caused by human papilloma virus (HPV) infection with the vaccination much progress has been made in the prevention and control of cervical cancer [2]. In 2009 G. Jayalalitha et al. discussed technique of grading cervical cancer images according to the cell formation of tissues. They expressed the use of Box Counting Method (DB) and Harfa Programme software to detect the fractional dimension and calculate the variation of intensity and texture complexity of cancer cell images [4]. In 2009 the **C. Todd et al.** proposed computer assisted algorithm for the classification of cervical cancer using digitized histology images of biopsies. Texture analysis of the nuclei structure is very important for the classification of cervical cancer histology[6]. In 2010 **M. Ross et al.** explore the focus towards the use of histology images for the classification of cervical cancer [8]. In 2010 **S.Allwin et al.** proposes approach which use textural properties to classify the various malignancies in cervical cyto images [10]. In 2011 **C. Balleyguier et al.** proposed the guidelines for staging & follow up of patients which suffered from uterine cervical cancer & provides the radiologists with a framework & expressed the importance towards adequate patient preparation, protocol optimization and MRI reporting expertise are essential to achieve high diagnosis accuracy [11]. In 2012 **M. Singh et al.** use the decision trees to predict the protein classes through see5 tool and express the importance of decision tress in bioinformatics. In 2014 **S. Sharma et al.** describe the importance of evolutionary multi objective optimization algorithms in bioinformatics [15].

III. METHODOLOGY USED



IV. RESULTS AND DISCUSSIONS

The Pearson correlation between various features like Node PET, Clin Diameter, MRI Vol, Uterine Body, Status, Rel Primary, Rel pelvic, Rel Abdo, Rel Supraclav, Rel Distant is obtained and shown in the table. Which describe that Rel Abdo, Rel Distant, Status, Histology & Rel Primary plays vital role in decision making. Correlation graph for various features is shown in Fig.1

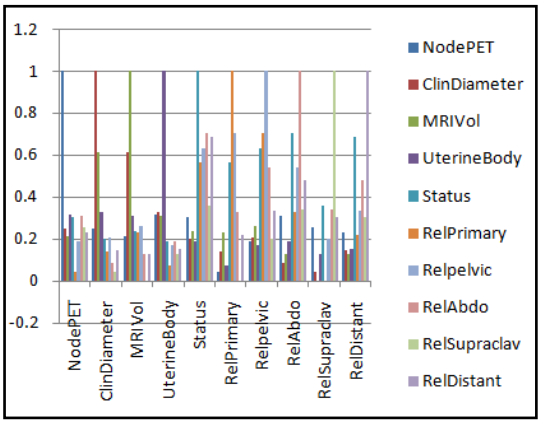


Fig. 1 Correlation between various features

DOI 10.17148/IJARCCE.2016.5488

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

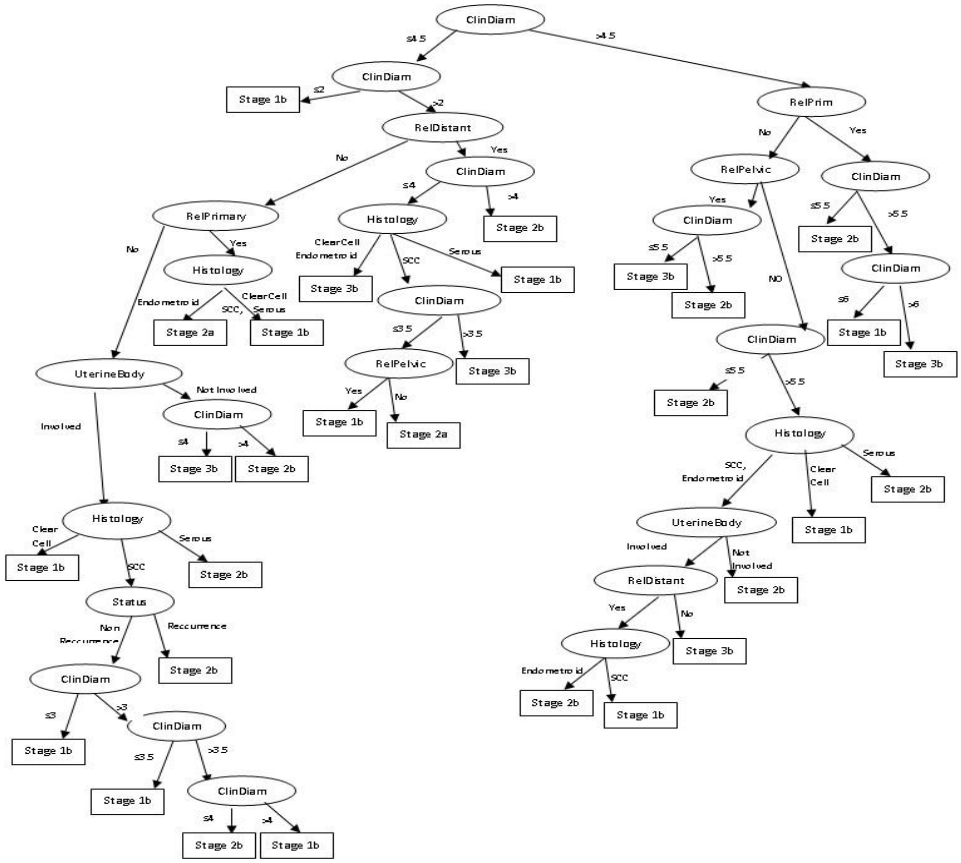
IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 4, April 2016

TABLE 1: Features Correlation

Correlations										
	Node PET	Clin Diameter	MRI Vol	Uterine Body	Status	Rel Primary	Rel pelvic	Rel Abdo	Rel Supraclav	Rel Distant
Node PET	1	0.25	0.214	0.313	0.303	0.04	0.19	0.311	0.252	0.229
ClinDiameter	0.25	1	**0.615**	0.329	0.201	0.137	0.204	0.087	0.04	0.143
MRIVol	0.214	0.615	1	0.311	0.237	0.228	0.258	0.13	0.003	0.128
Uterine Body	0.313	0.329	0.311	1	0.187	0.073	0.167	0.186	0.127	0.152
Status	0.303	0.201	0.237	0.187	1	0.563	0.629	**0.702**	0.357	**0.684**
Rel Primary	0.04	0.137	0.228	0.073	0.563	1	0.704	0.326	-0.008	0.219
Rel pelvic	0.19	0.204	0.258	0.167	**0.629**	**0.704**	1	0.538	0.202	0.332
Rel Abdo	0.311	0.087	0.13	0.186	0.702	0.326	0.538	1	0.338	0.478
Rel Supraclav	0.252	0.04	0.003	0.127	0.357	-0.008	0.202	0.338	1	0.305
Rel Distant	0.229	0.143	0.128	0.152	0.684	0.219	0.332	0.478	0.305	1

The correlations of various features are shown in Table1 and decision tree based on various features have been obtained using different options like Rule sets, Sort by Utility, Boosting, Winnowing, Advance Pruning, etc. present in the See5 tool



For the data set of 237 patients with 10 features, the accuracy of the different techniques was calculated shown in Fig 2. The C5 algorithm gives 67.5% accuracy using advance pruning option.

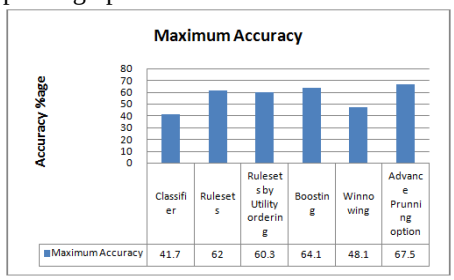


Fig. 2 Shows the Maximum accuracies obtained

V. CONCLUSION

In the field of oncology huge amount of information is processed to diagnose the cancer as well as for treatment of patient. A decision tree based computerized program classified the stages of cancer & helpful for oncologist to identify the stage of cancer.

The proposed way classifies the stages of the cervical cancer and recognizes similar diagnostic cases in the data using see5 Tool. The data of the cervical cancer patients is provided as input and based on some classification rules are extracted the stage of the cancer is detected from decision trees.

DOI 10.17148/IJARCCE.2016.5488

REFERENCES

[1]. www.igcs.org/professional/Education/treatmentResources/CervicalCaDB.html.
[2]. A. Satija, "Cervical cancer in India", South Asia Centre for Chronic Disease, 2009.
[3]. http://www.cancer.gov/cancertopics/pdq/treatment/cervical/Patient/page2.
[4]. G. Jayalalitha and R. Uthayukumar, "Recognition of Cervical cancer based on Fractal Dimension", International Conference on Advances in Recent Technologies in Communication and Computing, 2009.
[5]. K.Jayant, R.S Rao, "Improved stage at diagnosis of Cervical cancer with increased cancer awareness in rural Indian population", International Journal Cancer,1995, Vol.63, pp 161-163
[6]. C.Todd , Rahmdwati, G.Naghdy, "Cervical Cancer Classification Using Gabor Filters", First IEEE International Conference on Healthcare and Informatics, Imaging and Systems Biology, 2011
[7]. Cevical Cancer Overview", American Cancer Society, http://www.cancer.org/acs/groups/cid/documents/webcontent/003042-pdf.pdf.
[8]. Montse Ross and Rahmadwati, "Classification Cervical Cancer Using Histology Images", Second international Conference on Computer Engineering and Application, 2010.
[9]. B. Bergeron, "Bioinformatics Computing", pp 257-270, 2002.
[10]. S.Allwin and S.Pradeep Kumar, "Classification of stages of Maligancies using Textron signature of Cervical Cyto Image", Computational Intelligence and Computing Research (ICCIC), 2010.
[11]. J. Han and M. Kamber, "Data Mining Concepts and Techniques", Morgan Kaufmann Publishers, USA pp 279-322, 2003.
[12]. C. Balleyguier et al., "Staging of uterine cervical cancer with MRI", Journal on European Radiology, 2011
[13]. I. Friedberg, "Automated Protein Function Prediction- the Genomic Challenge", Briefings in Bioinformatics, vol 7, no.3, pp 225-242.
[14]. L.J. Jensen, R. Gupta, N. Blom, D. Devos, J. Tamames C. Kesmir, H. Nielsen, H.H. Stærfeldt, K. Rapacki, C. Workman C.A.F. Andersen, S. Knudsen, A. Krogh, A.Valencia and S. Brunak , "Prediction of Human Protein Function from Post-Translational Modifications and Localization Features", Journal of Molecular Biology, vol. 319, issue 5,pp 1257-1265, 2002.
[15]. MS Sharma," A Review towards Evolutionary Multi objective optimization Algorithms", An International Journal of Engineering Sciences, Vol13/37-Vol13, Vol. 3, Issue December 2014.
[16]. M. Singh, G. Singh, S. Sharma," Human Protein Function Prediction from Sequence Derived Features using See5 ", International Journal of Scientific & Engineering Research Volume 3, Issue 7, July-2012

BIOGRAPHY

Sunny Sharma is a Research Scholar at Department of computer Science of Guru Nanak Dev University, Amritsar India. He received his MCA degree in Computer Science from Guru Nanak Dev University, Amritsar, Punjab and Cleared UGC NET, GATE & now pursuing Ph.D. in Computer Science from Guru Nanak Dev University, Amritsar Pb. (India). He has published 04 International and 08 National research papers. His main field of research interest is Bio-Informatics, Machine Learning and Data mining. He works on the Prediction of Protein function & Structure, Rule Mining, Machine Learning.