A Novel Medical Support System for the Social Ecology of Cervical Cancer: A Research to Resolve the Challenges in Pap Smear Screening and Prediction at Firm Proportion

Mrs. R. Vidya and Dr. G.M. Nasira

1Research Scholar, Department of Computer science, Manonmaniam Sundaranar University, Tirunelveli.
2Assistant Professor, Department of Computer Science, Chikkanur government college, Tirupur.

INTRODUCTION

Cancer is a large group of diseases (over 200) characterized by uncontrolled growth and spread of abnormal cells. When, gynecological cancer is taken into account the mostly affected cancers are Breast cancer, ovarian cancer, cervical cancer & Uterus Cancer. The high rate of it which cannot be predicted too is Cervical Cancer as shown in Fig 1. There are two classifications i.e. Benign (Non-cancerous) and Malignant (Cancerous) in this DSS.

Table 1: Global - Common Cancer.

<table>
<thead>
<tr>
<th></th>
<th>Female</th>
<th>Developing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Developed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breast</td>
<td></td>
<td>Breast</td>
</tr>
<tr>
<td>Ovarian</td>
<td></td>
<td>Cervix</td>
</tr>
<tr>
<td>Lung</td>
<td></td>
<td>Stomach</td>
</tr>
</tbody>
</table>

Anybody, he or she can be young or old can get cancer. What causes Cancer? Lifestyle, Family History and Environment are three major risk factors for cancer. 67% of all cancers are caused by the way we live as shown in Figure 2.

Factors to Contribute – Causes Of Cancer:

Worldwide cervical cancer is the second leading cause of death among women. India ranks as the first in common cancer among women. In urban areas cervix cancer account for over 40% of cancer while in rural places the percentage of cancer prevails for 65% when overall compared with 100% of cancer affected persons as per the information from the cancer CNCI (Chittranjan National Cancer Institute Registry.) as shown in figure 3.
The lower part of the uterus or womb persistent is where cervical cancer occurs and it is can be said as cervix as shown in figure 4, which is HPV (Human Papilloma Virus). It is sexually transmitted. Infection usually with certain high risk types of HPV can cause abnormal cells to develop on the cervix. Regular Pap tests can detect these abnormal cells at an early stage, when they can usually be treated quickly and easily. The most symptoms of cervical cancer are abnormal bleeding then Pelvic pain, pain during intercourse and vaginal bleeding. Without regular Pap test the abnormal cells may remain undetected and could develop into cervical cancer, usually over many years. Treatments for cervical cancer can include surgery, chemotherapy and radiotherapy. The best way for women to protect themselves against cervical cancer is to have the HPV vaccine when aged 35 - 40 years and then to have regular Pap test.

Treatment depends on the stage of the cancer if it has been advanced then it can include surgery, radiation or chemotherapy. Prevention of cervical cancer can be improved by i) Avoiding smoking ii) Delaying first intercourse iii) Using a condom to reduce the risk of HPV iv) Regular Pap test and getting vaccinated against HPV as shown in figure 5.

Methodology:

Early detection is important for a better prognosis, so using Regression Tree we can predict the cancer. Linear, non-linear and generalized linear model of regression can be used for prediction. Classification and prediction are two forms of data analysis that can be used to extract models describing important data classes or to predict future data trends.

CART system was proposed in 1984 by Breiman, he elucidates that every leaf will store a valid prediction continuously and this is considered as the average value for the attributes which is predicted. This predicted attributes is taken as the training tuples that reaches leafs. When compared with linear regression, Regression model trees are more accurate even when the data are not as easy as in simple linear model.

Cart Specimen:

First on starting with prediction, the new sample is injected from the top to down of the tree model. The sample data ends up in one node and it is assigned with a label as terminal node. This particular procedure is recapitulated for all data over all tree nodes in a collaborative way and the average sum of all the nodes are reported as “random forest prediction”.

Algorithm:

The classification and regression trees (CART) methodology is known prediction in cancer. The basic CART building algorithm is a greedy algorithm. In the implementation of CART the dataset is split into the two sub groups that are most different with the respect to the outcome.

Method 1: Recursive Partitioning:
1. Take all of our data
2. Consider all possible values of all variables
3. Select the variable / value (X= t1) that produces the greatest “separation” ion the target. (X=t1) is called a “Split”.
4. If X< t, then send the data to the “left” otherwise send data point to the “right”.
5. Now repeat the same process on these two “nodes” we can get a tree. CART only uses binary splits.

Method 2: CART Algorithm:

Each and every tree along with it nodes from start to terminal is erected using this algorithm.
1. Let the training cases inputted be considered as N.
2. Let the variables in the classifier can be allocated as M
3. The input variables number is taken as m and it is used to determine the node of the tree. The decision is predicted using m<M
4. From the training cases the training set has be to be chosen and it can be named as n. this n would be a replacement for all other available training cases.
5. The rest training cases can be exploited for finding the error in the tree nodes by forecasting their classes
6. Choose m variables on two way base for decision of the node in the trees.
7. The best split is calculated based on these variables of m from the training set.
8. So each tree grows fully and not even a single node is pruned from the process.

First on starting with prediction, the new sample is injected from the top to down of the tree model. The sample data ends up in one node and it is assigned with a label as terminal node.

CART advantages:
1) Non parametric
2) Automatically performs variable selection
3) Use any combination of continuous/discrete variables
4) Ability to automatically bin massively categorical variables into a few categories.
5) Good for prediction of cancer in medical field
6) Discovers interactions among variables.

Model Architecture:

![Model Architecture Diagram]

**Fig. 1:** Other Gynecological Cancer when compared gives that cervical cancer is more dreadful than all other cancers.

![Fig. 1: Other Gynecological Cancer]

**Fig. 2:** Causes of cancer.

![Fig. 2: Causes of Cancer]

**Figure 3:** The cancer affecting ration in rural and urban places.
**Fig. 4:** Explaining what is cervix and where it is formed (Image systematized by us).

**Fig. 5:** Risk Factor of cervical cancer (Image systematized by us).

**Fig. 6:** The model architecture how the methodology works with an example. Here Y stands for Yes & N stands for no the variables are most on the form of either categorical or numerical.
Experiments and Results:

In the present study the dataset for training cases are collected by conducting a survey from doctors and by collecting the case sheets of the patients. The case sheet gives the detailed description from their history to current treatment taken based on those references the data has been collected. The training data is acquired from the training cases, which comprises of 100 records, 75 records for training and 25 for testing. In this paper the above data has been analyzed using MATLAB and received result with an accuracy of 83.871% as shown in figure 8 & 9.

Fig. 7: Model architecture with an example. Here H stands for High risk and L stands for Low risk.

Fig 8: Regression tree->prediction, Identifiers for X10327 is GAS7.

Fig. 9: Regression tree, Using Gene 3233 – Biospy features, we can draw the tree.
Discussion:
The cancer is very worse now a days and it is very hard to predict. Many people just ignore their problem and when they feel it too worse to bear their pain they approach to hospitals. The test taken for the patients carry many days to find the real cause whether they have cancer or not, the process of screening should be quick and it is ought to be effective also. The data collected at the first instance should be taken into account and with that dataset the prediction should be made. To make it possible we have implemented a methodology for the medical support system.

Conclusion:
This paper implements a new CART algorithm which provides a new way in the medical system to identify cervix cancer at a fast pace. The algorithm is tested and it provides accuracy to a beneficiary level of 83% which can really bring a new improvement in the challenges to predict cervical cancer screening. On our further work we are trying to increase the efficiency of the percentage by applying RFT laterally with it.

REFERENCES


David M. Eddy, “Screening for cervical cancer”. American college of physicians”, 113(3).


Jan, M.M. Walboomers, “Human papilloma virus is a necessary cause of invasive cervical cancer worldwide”. Wiley Online Library, 189(1).

Julie, M. David, Kannan Balakrishnan, “Significance of classification techniques in prediction of learning disabilities”. IJAIA.


