Early Detection of ovarian cancer (BARCA 1 & BARCA 2 MUTATION) risk prediction for low income country using Data mining technology: Bangladesh

ABSTRACT

Background: Ovarian cancer is the most lethal gynecological cancer which incidence increasing day by day in developing countries. More than 80% ovarian cancers occur in women over the age of 50. Therefore, identification of genetic factors including mutations in the BRCA1 and BRCA2 gene (breast cancer genes) as well as others factors is very crucial in developing novel strategy of ovarian cancer prevention.

Methodology: This study was carried out in 521 cancer and non-cancer patients’ data is collected from different diagnostic centre and data pre-processed. Then a structured questionnaire was used containing details of ovarian cancer risk factors including age, menopause end age, problem during pregnancy, first sex age, any infection in genital area, affected by ovarian cancer, abortion, pregnancy, BMI, menopause age after 50, food habit, obesity, excessive alcohol, late Menopause, early Menopause, hormone therapy, exercise, previous exposure to other sexually transmitted infections (STIs), marital status, genetic risk, outdoor activities and affected by other cancer basis on the previous studies. Results: After pre-processing data is clustered using K-means clustering algorithm for identifying relevant and non-relevant data to ovarian Cancer. Next significant frequent patterns are discovered using AprioriTid shown in Table 1 and Decision Tree algorithm shown in Table 2. This ovarian cancer risk prediction system will be helpful in detection of a patient’s predisposition to ovarian cancer. Specifically there were no work of ovarian cancer risk prediction system using data mining or Statistical approaches.

REFERENCES

6. Australian Lung Cancer Collaborative update on cancer scenario of Bangladesh.
7. South Asia Cancer 2013 Oct 2;24(7-8).
8. Li Ka Shing Faculty of Medicine, University of Hong Kong.
10. The authors are grateful to all of the subjects who participated in this research.

RESULTS & Analysis

The frequency table wascontrived by comparing the results of SPSS and WEKA. Both statistical and data mining approaches shows same frequency. Age range between 35 and 65 where the mean age was 50 approximately 521 Bangladesh women data were analyzed. Here 121 women were not affected (control group) and 421 women were affected by ovarian cancer (case group). We performed the data visualization analysis and clustering using data mining technology. The mean approach of this study to finalize the data analysis with decision tree algorithm based tree by which we can predict if a person is affected by ovarian cancer. If we have “problem during pregnancy” then if that “abortion” then yes she has the possibility to have ovarian cancer here the tree is based on the highly significant 10 factors not all the factors.

CONCLUSIONS

In conclusion, as Bangladesh is a low incoming and population country, most of the women are not aware of distressful ovarian cancer disease because of lacking education. Moreover a vast number of health care system society mostly women conceive ignorance of discuss on ovarian cancer with others. This scheme has worked on ovarian cancer and different data mining techniques has been updated day-by-day. In the paper the association relations of factors has been detected with ovarian cancer and the possibility of prevention and the obtained high 521 Bangladesh women data were used to increase awareness among women about different factors. It will be highly needed to maintain and implement the current work and further research would depict to future researchers to find out some new error in the meantime to save the women from this atrocious cure. Otherwise one day the whole system would by us or any relatives or anyone of the society.