

DATA MINING

EARLY DETECTION AND PREDICTION OF CANCER

Srinath - CED15B033
Pruthvi - COE15B019
Sasidhar - COE15B029
Nikhil - CED15I010

Abstract

Introduction

Cancer is one of the most common diseases in the world that results in majority of death. It is caused by uncontrolled growth of cells in any of the tissues or parts of the body. Cancer may occur in any part of the body and may spread to several other parts. Only early detection of cancer at the beginning stage and prevention from spreading to other parts in malignant stage could save a person's life which is very important. A number of studies in developed countries have shown that cancer incidence varies between people with different levels of education. A high incidence of breast cancer has been found among those with high levels of education whereas the other classes has been found for the incidence of cancers of the stomach, lung and uterine cervix. Such differences in cancer risks associated reflect in the differences in lifestyle factors and exposure.

Importance

So, early detection and prevention of cancer plays a very important role in reducing deaths caused by cancer. Identification of genetic and environmental

factors is very important in developing novel methods to detect and prevent cancer. This study describes the association between cancer incidence pattern and risk levels of various factors by developing a risk prediction system for different types of cancer which helps in diagnosis.

Fuzzy idea(s) on techniques

Therefore, techniques like multi layered methods with clustering and decision tree(tentative methods) to build a cancer risk prediction system can be proposed here which predicts lung, breast, oral, cervix, stomach and blood cancers and they also may be user friendly, time and cost saving. Data mining technologies such as classification, clustering and prediction are used to identify potential cancer patients from the data set. The gathered data is preprocessed, fed into the database and can be classified using some clustering algorithm and cancer and non cancer patient data can be predicted.

Finally a prediction system is developed based upon the user input to analyze risk levels which help in estimating the rise level of that particular person, so that he/she can get the treatment of that particular cancer in the early stage of it. This detection also helps by saving person's time/money before going for clinical and lab tests that's usually expensive and time consuming now a days.

Conclusion

So the bottom line of the story is to develop an application that runs some data mining algorithms applied on some data set, using which we will be predicting the person cancer's risk level based upon his input/data given by him/her.