Analysing Critical Factors Associated with Perceived Risk for Major Types of Cancers in Sri Lanka — using Data Mining Techniques

M.L. Parakrama[1], K.A.K.D. Kodikara[1*] and S.H.D. Senanayake[1]

[1]*Department of Computer Science and Technology, Uva Wellassa University, Badulla, Sri Lanka*

Cancer is one of the world recognized critical cause of morbidity and mortality. Identification of demographic, health and life style factors are significant to detect and prevent from the cancers. Since the Sri Lankan public has lack of knowledge about influence of those factors on cancer risk, the research has been conducted using data mining techniques to develop significant patterns in order to identify the critical factors towards the most recorded cancer types in Sri Lanka (Breast, Oral and Oesophagus cancers). Hence, the ultimate objective of the research is to increase the awareness among Sri Lankan public which will be helpful to take preventive approaches. The research was based on the patients' data within a five-year period of time, which was collected from the sources of *Apeksha* hospital, *Maharagama.* By following the Knowledge Discovery in Databases process, data collection and preprocessing were completed manually. Data was analysed through the data mining tool called *Weka*. The research was used Random Forest classification technique, *k-mean* clustering technique, and *Apriori* association rule mining. From the research it has been identified that the gender, number of children, breast feeding, menopause, weight, age at menarche as the critical factors towards the breast cancers. The gender, marital status, weight, smoking, consumption of alcohol, betel chewing were identified as critical factors towards the oral cancers. The gender, consumption of alcohol, occupation, weight and age were identified as critical factors towards the Oesophagus cancers. The patterns were obtained from the above identified critical factors with their approximate values.

Keywords: Cancer, Data mining, Weka, KDD Process