NCON-PGR_2022_032

Classification of Cervical Cancer using Random Forest

***B.Mohd Bashah*[1]*, *K.M.N. Ku Khalif*[2] *and* *N.A. Ramli*[3]**

[1]Centre for Foundation Studies, International Islamic University Malaysia, 26300 Gambang, Pahang, Malaysia.

[2,3]Centre for Mathematical Sciences, College of Computing and Applied Sciences, Universiti Malaysia Pahang, 26300 Gambang, Pahang, Malaysia.

[2,3]Centre for Artificial Intelligence & Data Science, Universiti Malaysia Pahang, 26300 Gambang, Pahang, Malaysia.

**Corresponding author*: bahirahmohdbashah@gmail.com

Abstract

Cervical cancer is the second most common cancer among Malaysian women between 15 to 44 although the morbidity and the mortality of cervical cancer have been decreasing in recent years. Developing supervised models for cervical cancer is a challenging task. By gleaning deeper insights from the data, data mining knowledge has capability to learn from data, identify the patterns with meaningful in that they lead to some advantages in many real-world applications. In this research, the cervical cancer risk classification model was used by using data mining approach which consider Decision Tree and Random Forest algorithm. These two models have been implemented by using JupyterLab on the UCI datasets. Model evaluation has been conducted to identify the robust data mining algorithm in the prediction of cervical cancer risk. The model gives 67% for the precision and 95% of accuracy. However, decision tree is the best method compared to Random Forest since Random Forest has the lowest AUC which indicated that it is the worse model. To improve this study, other method such as Artificial Neural Network, Support Vector Machine or ensemble classifiers can be applied to the dataset to see if there is a better model to predict cervical cancer.

Keywords: Cervical cancer; Data mining; Classification; Random forest.