Enhancing the Risk Prediction of Cervix Cancer Using RFE, SMOTE, Deep Learning and Ensemble Learning Methods

Melisa Krisnawati, ByungGook Lee, Hyotaek Lim

Abstract

Cervical cancer is one of the significant global health concerns among the world. This study analyses a dataset of cervical cancer that contains demographic, behavioral, and medical history variables. The aim of this research is to predict the factors of cervix cancer using ensemble learning (many machine learning methods that combined together for producing better prediction) like Random Forest (RF), Decision Tree (DT), and Logistic Regression (LR) that showed promising results on the prediction. The dataset also contains missing values that can have an impact on the performance of the machine learning models. That's why in this research we use RFE (Recursive Feature Elimination), also SMOTE (Synthetic Minority Oversampling Technique) to optimize the results. Ways to detect cervical cancer based on this dataset is by looking at the positive biopsy results, so we can reduce the cervical cancer cases by early detection and make prevention strategies effectively. Additionally, we are also adding deep learning methods as the enhancement methods that can provide insights to the important feature for cervical cancer identification. The outcome of this research is the suggested method can efficiently identify and have the potential at improving the diagnosis and management of cervix cancer.

Funding

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2020R1A2C1008589).