

Title : Aprecise machine learning model: Detecting cervical cancer using feature selection and explainable AI

Author Names : Rashiduzzaman Shakil , Sadia Islam, Bonna Akter

Abstract: Cervical cancer is a cancer that remains a significant global health challenge all over the world. Due to improper screening in the early stages, and healthcare disparities, a large number of women are suffering from this disease, and the mortality rate increases day by day. Hence, in these studies, we presented a precise approach utilizing six different machine learning models (decision tree, logistic regression, naïve bayes, random forest, k nearest neighbors, support vector machine), which can predict the early stage of cervical cancer by analyzing 36 risk factor attributes of 858 individuals. In addition, two data balancing techniques—Synthetic Minority Oversampling Technique and Adaptive Synthetic Sampling—were used to mitigate the data imbalance issues. Furthermore, Chi-square and Least Absolute Shrinkage and Selection Operator are two distinct feature selection processes that have been applied to evaluate the feature rank, which are mostly correlated to identify the particular disease, and also integrate an explainable artificial intelligence technique, namely Shapley Additive Explanations, for clarifying the

Keywords: Cervical cancer SMOTE ADASYN Chi-square LASSO Machine learning Decision tree Explainable AI SHAP

<https://doi.org/10.1016/j.jpi.2024.100398>

<https://www.sciencedirect.com/science/article/pii/S2153353924000373>