

Comparative Evaluation of Accuracy of Selected Machine Learning Classification Techniques for Diagnosis of Cancer: A Data Mining Approach

Authors : Rajvir Kaur, Jeewani Anupama Ginige

Abstract : With recent trends in Big Data and advancements in Information and Communication Technologies, the healthcare industry is at the stage of its transition from clinician oriented to technology oriented. Many people around the world die of cancer because the diagnosis of disease was not done at an early stage. Nowadays, the computational methods in the form of Machine Learning (ML) are used to develop automated decision support systems that can diagnose cancer with high confidence in a timely manner. This paper aims to carry out the comparative evaluation of a selected set of ML classifiers on two existing datasets: breast cancer and cervical cancer. The ML classifiers compared in this study are Decision Tree (DT), Support Vector Machine (SVM), k-Nearest Neighbor (k-NN), Logistic Regression, Ensemble (Bagged Tree) and Artificial Neural Networks (ANN). The evaluation is carried out based on standard evaluation metrics Precision (P), Recall (R), F1-score and Accuracy. The experimental results based on the evaluation metrics show that ANN showed the highest-level accuracy (99.4%) when tested with breast cancer dataset. On the other hand, when these ML classifiers are tested with the cervical cancer dataset, Ensemble (Bagged Tree) technique gave better accuracy (93.1%) in comparison to other classifiers.

Keywords : artificial neural networks, breast cancer, classifiers, cervical cancer, f-score, machine learning, precision, recall

Conference Title : ICHIHIT 2018 : International Conference on Health Informatics and Health Information Technology

Conference Location : London, United Kingdom

Conference Dates : April 24-25, 2018