

Early Gastric Cancer Prediction from Diet and Epidemiological Data Using Machine Learning in Mizoram Population

Authors : Brindha Senthil Kumar, Payel Chakraborty, Senthil Kumar Nachimuthu, Arindam Maitra, Prem Nath

Abstract : Gastric cancer is predominantly caused by demographic and diet factors as compared to other cancer types. The aim of the study is to predict Early Gastric Cancer (ECG) from diet and lifestyle factors using supervised machine learning algorithms. For this study, 160 healthy individual and 80 cases were selected who had been followed for 3 years (2016-2019), at Civil Hospital, Aizawl, Mizoram. A dataset containing 11 features that are core risk factors for the gastric cancer were extracted. Supervised machine algorithms: Logistic Regression, Naive Bayes, Support Vector Machine (SVM), Multilayer perceptron, and Random Forest were used to analyze the dataset using Python Jupyter Notebook Version 3. The obtained classified results had been evaluated using metrics parameters: minimum_false_positives, brier_score, accuracy, precision, recall, F1_score, and Receiver Operating Characteristics (ROC) curve. Data analysis results showed Naive Bayes - 88, 0.11; Random Forest - 83, 0.16; SVM - 77, 0.22; Logistic Regression - 75, 0.25 and Multilayer perceptron - 72, 0.27 with respect to accuracy and brier_score in percent. Naive Bayes algorithm out performs with very low false positive rates as well as brier_score and good accuracy. Naive Bayes algorithm classification results in predicting ECG showed very satisfactory results using only diet cum lifestyle factors which will be very helpful for the physicians to educate the patients and public, thereby mortality of gastric cancer can be reduced/avoided with this knowledge mining work.

Keywords : Early Gastric cancer, Machine Learning, Diet, Lifestyle Characteristics

Conference Title : ICMLT 2020 : International Conference on Machine Learning and Technology

Conference Location : Singapore, Singapore

Conference Dates : March 30-31, 2020