BEST: International Journal of Management Information Technology and Engineering (BEST: IJMITE)
ISSN (P): 2348–0513, ISSN (E): 2454–471X
Vol. 9, Issue 2, Dec 2021, 47–56

IMPORTANCE OF ACTIVE LEARNING FOR OPTIMAL DETECTION OF DISEASES

CHANDNI CHATTERJEE, PRANAB HAZRA, SUVADEEP GUHA*, AVIJIT SAHA & RAJU DAS

Department of Electronics and Communication, Narula Institute of Technology, Kolkata, India

ABSTRACT

Availability of data is very easy and most of the data are unlabeled. If we want to take some decision from the available data, it is required to be processed and data should be labeled. This unlabeled data is a big problem in the field of machine learning. Data scientists are proficient for analyzing with more data than they have and it's the point where active learning comes into picture. In machine learning, active learning acts as a subset where learning algorithm interact with user by putting query to label data with the desired outputs. In active learning a query unit is of the same type as the target concept to be learned. Alternative query is introduced in the context of multiple-instance active learning (MIL) where instances are grouped into bags, and it is the bags, rather than instances, that are labeled for training. A bag is labeled negative if and only if all of its instances are negative and positive, even if at least one of its instances is positive. In this paper we can see at higher specificity the Area Under Receiver operating characteristics (ROC) Curve (AUC) increases. Combined use of MIL and AL will reduce the labeling effort and it will do accurate detection of disease through classification.

KEYWORDS: Active Learning, Multiple Instance Learning, Machine Learning, Bag, ROC, Labeled Data, Unlabeled Data.