

Example abstract 4modIA

Here is an example of an AI abstract with the different sections colour coded

Background context; Existing gap(s) in knowledge; **Aims;** **Methods;** **Results;** **Conclusion;**
Keywords

(modified from Sayin et al., 2021, A review and experimental analysis of active learning over crowdsourced data. Artificial Intelligence Review, Vol 54 (5283-5305)

Creating training data is a major bottleneck for developing machine learning, especially for deep learning systems. Active learning provides a cost-effective means for creating training data by selecting the most informative instances for labeling. Labels in real applications are often collected from crowdsourcing, which engages online crowds for data labeling at scale. Despite the importance of using crowdsourced data in the active learning process, an analysis of how the existing active learning approaches behave over crowdsourced data is currently missing. We aim to review the existing active learning approaches and test a set of benchmarking ones on crowdsourced datasets. We conducted a comprehensive and systematic survey of the recent research on active learning in the hybrid human-machine classification setting, where crowd workers contribute labels (often noisy) to either directly classify data instances or to train machine learning models. We identified three categories of state of the art active learning methods according to whether and how predefined queries employed for data sampling, namely fixed-strategy approaches, dynamic-strategy approaches, and strategy-free approaches. We conducted an empirical study on their cost-effectiveness. We show that the performance of the existing active learning approaches is affected by many factors in hybrid classification contexts, such as the noise level of data, label fusion technique used, and the specific characteristics of the task. Various challenges remain in designing active learning strategies for hybrid classification problems. One potential direction is to balance the AL and human vs machine contribution in hybrid crowd-ML classification problems.

Keywords: Active learning · Crowdsourcing · Human in the loop · Classification