

## Active Features Determination: A Unified Framework

**Authors :** Meenal Badki

**Abstract :** We address the issue of active feature determination, where the objective is to determine the set of examples on which additional data (such as lab tests) needs to be gathered, given a large number of examples with some features (such as demographics) and some examples with all the features (such as the complete Electronic Health Record). We note that certain features may be more costly, unique, or laborious to gather. Our proposal is a general active learning approach that is independent of classifiers and similarity metrics. It allows us to identify examples that differ from the full data set and obtain all the features for the examples that match. Our comprehensive evaluation shows the efficacy of this approach, which is driven by four authentic clinical tasks.

**Keywords :** feature determination, classification, active learning, sample-efficiency

**Conference Title :** ICDEMLA 2024 : International Conference on Data Engineering and Machine Learning Applications

**Conference Location :** Havana, Cuba

**Conference Dates :** February 01-02, 2024