4.1.1 Datasets

# Original

30 datasets from the OpenML repository [33] were used for comparison. These were the 20 most run binary datasets, and 10 most run multiclass datasets. The datasets were restricted to less than 15000 instances, less than 5 classes, and no missing values. These datasets are from a variety of domains, and have a varying number of features (with both categorical and numeric), classes, and instances. The datasets offer a broad range to ensure generalisability of the proposed method. These datasets are summarised in Table 1.

# Condensed

30 datasets from the OpenML repository [33] were used for comparison. These were the 20 most run binary datasets, and 10 most run multiclass datasets. The datasets were restricted to less than 15000 instances, less than 5 classes, and no missing values. These datasets are from a variety of domains, and have a varying number of features (with both categorical and numeric), classes, and instances. The datasets offer a broad range to ensure generalisability of the proposed method.