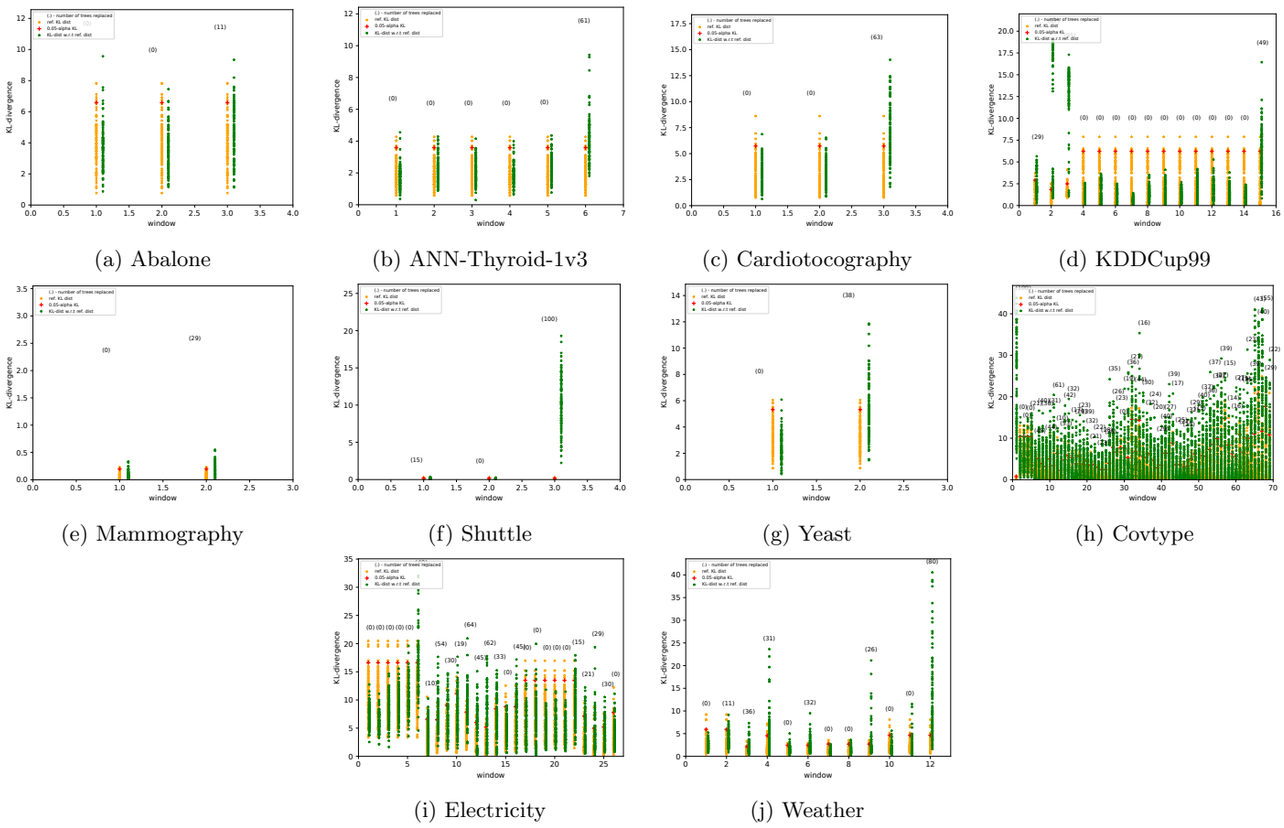
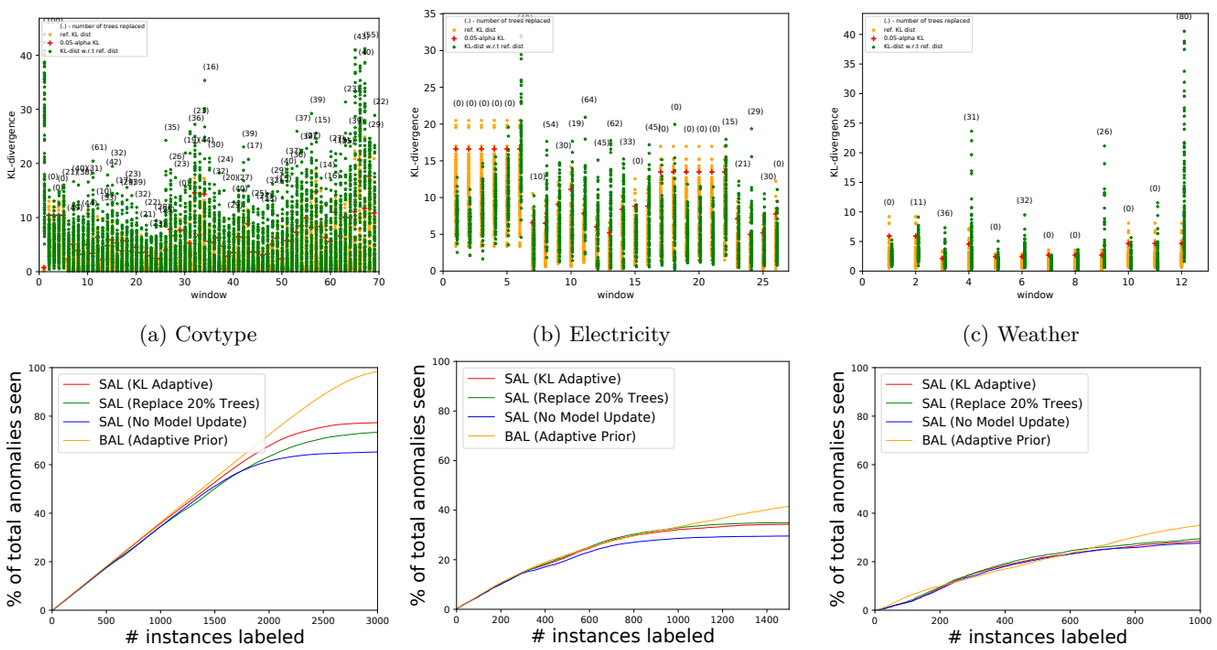


(a) Abalone (b) ANN-Thyroid-1v3 (c) Cardiotocography (d) KDDCup99

(e) Mammography (f) Shuttle (g) Yeast (h) Covtype

(i) Electricity (j) Weather

Concept drift detection in datasets using tree-structures. The total number of trees in the forest model is held constant at **100** for all datasets. Each dataset has a streaming window size commensurate with its total size: *Abalone*(512), *ANN-Thyroid-1v3*(512), *Cardiotocography*(512), *Covtype*(4096), *Electricity*(1024), *KDDCup99*(4096), *Mammography*(4096), *Shuttle*(4096), *Weather*(1024), and *Yeast*(512). The last data window in each dataset usually has much fewer instances and therefore its distribution is very different from the previous window despite there being no data drift. **Therefore, ignore the drift in the last window.** We did not expect *Abalone*, *ANN-Thyroid-1v3*, *Cardiotocography*, *KDDCup99*, *Mammography*, *Shuttle*, and *Yeast* to have much drift in data, and this can also be seen in the plots where most of the windows in the middle of streaming did not result in too many trees being replaced (the numbers in the parenthesis are mostly zero). We also did not expect *Covtype* to have much drift in its default ordering (from its public repository); however, after seeing the plots we realize that there is a definite drift due to which a significant number of trees need to be replaced with each new window of data. *Electricity* and *Weather* are standard streaming datasets with expected concept drift, which shows up in the plots.



(a) Covtype (b) Electricity (c) Weather

(d) Covtype (e) Electricity (f) Weather

Integrated drift detection and label feedback with Stream Active Learner (SAL). The query budget and the stream window size for each dataset was set as: *Covtype*$(3000, 4096)$, *Electricity*$(1500, 1024)$, and *Weather*$(1000, 1024)$. The max memory limit was set equal to the window size. We query 20 labels every time a window of data arrives. The all remaining query budget is used with the final set of instances in memory after the last window has been processed. All experiments employed *Retention Type 1* where only the most anomalous instances fitting the memory limit are retained in memory. When a new window of data arrives: `SAL (KL Adaptive)` dynamically determines which trees to replace, `SAL (Replace 20% Trees)` replaces 20% oldest trees, and `SAL (No Model Update)` keeps the trees fixed after initially training with the first window of data. `BAL (Adaptive Prior)` is the Batch Active Learner without any memory limitation and is the most optimistic setup because all data is available in memory right from the start. `SAL (KL Adaptive)` **is the best performing streaming strategy** on these datasets and **also competitive** with `BAL (Adaptive Prior)`.