Table A: Performance on various workload scenarios in GMQ (lower is better), where the dominant operation (insert/delete/update) occurs twice as frequently as each of the other two operations.

| Data | Method | Mild Drift | | | Severe Drift | | |
|---|---|---|---|---|---|---|---|
| | | Insert-heavy | Delete-heavy | Update-heavy | Insert-heavy | Delete-heavy | Update-heavy |
| STATS | Fine-tune | 3.18 | 3.36 | 4.97 | **3.16** | 3.81 | 5.12 |
| | PostgreSQL | 120.24 | 166.94 | 172.68 | 216.78 | 227.56 | 258.96 |
| | ALECE | 18.58 | 17.25 | 20.18 | 30.67 | 26.92 | 32.15 |
| | DDUp | 4.37 | 4.45 | 5.62 | 6.89 | 6.72 | 6.96 |
| | FLAIR | **2.85** | **2.96** | **3.76** | 3.35 | **3.26** | **3.96** |
| Job-light | Fine-tune | **1.65** | 1.94 | 2.28 | 6.42 | 7.21 | 7.72 |
| | PostgreSQL | 7.96 | 7.68 | 8.45 | 25.15 | 26.89 | 29.93 |
| | DeepDB | 29.16 | 28.79 | 29.88 | 39.53 | 36.79 | 42.98 |
| | ALECE | 11.31 | 10.89 | 11.78 | 25.72 | 26.84 | 27.35 |
| | DDUp | 3.53 | 3.87 | 3.96 | 9.08 | 9.76 | 10.18 |
| | FLAIR | 1.68 | **1.59** | **1.92** | **6.26** | **6.58** | **7.21** |

Table B: Performance comparison on abrupt concept drift indicated by $D_{KL} > 3$ in GMQ (lower is better).

| Data | PostgreSQL | ALECE | DDUp | Fine-tune | FLAIR |
|---|---|---|---|---|---|
| STATS | 176.38 | 12.63 | 6.91 | 5.75 | **4.15** |
| Job-light | 19.41 | 16.24 | 6.65 | 6.25 | **3.26** |

Table C: Performance comparison with common concept drift learning strategies in GMQ (lower is better), including training from scratch (Retraining, RT), Fine-tuning (FT), and Knowledge Distillation (KD).

| Data | Scenario | RT | FT | KD | FLAIR |
|---|---|---|---|---|---|
| STATS | Mild Drift | 4.97 | 5.35 | 5.79 | **4.49** |
| | Severe Drift | 5.59 | **5.02** | 10.95 | 5.47 |
| Job-light | Mild Drift | 3.25 | 2.45 | 4.16 | **2.36** |
| | Severe Drift | 8.21 | 8.09 | 10.96 | **7.95** |

Table D: Performance comparison on TPC-H benchmark dataset under concept drift, where DeepDB is omitted due to PK-FK limitation.

| Data | Method | Mild Drift | | | | | | | | | Severe Drift | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | GMQ | Q-error | | | | P-error | | | | GMQ | Q-error | | | | P-error | | | |
| | | | 50% | 75% | 90% | 95% | 50% | 75% | 90% | 95% | | 50% | 75% | 90% | 95% | 50% | 75% | 90% | 95% |
| TPC-H | Fine-tune† | **3.75** | 1.76 | 5.32 | 13.09 | 64.81 | 1.21 | 2.65 | 8.54 | 28.16 | **6.11** | 2.85 | 7.76 | 18.53 | 97.89 | 1.92 | 3.26 | 18.21 | 36.24 |
| | PostgreSQL | 36.15 | 18.63 | 42.27 | 76.18 | 1182.54 | 2.97 | 8.14 | 27.25 | 136.37 | 88.75 | 18.21 | 82.66 | 120.12 | 8725.98 | 3.26 | 5.79 | 71.54 | 119.65 |
| | ALECE | 8.97 | 2.26 | 6.88 | 25.74 | 166.92 | 2.47 | 3.15 | 19.62 | 97.74 | 38.05 | 14.60 | 58.79 | 84.85 | 182.79 | 2.68 | 3.78 | 32.46 | 106.96 |
| | DDUp | 6.58 | 2.12 | 5.56 | 18.32 | 68.08 | 1.89 | 2.96 | 9.37 | 47.67 | 9.65 | 2.54 | 10.12 | 24.82 | 99.28 | 2.11 | 3.60 | 16.45 | 32.97 |
| | FLAIR | **3.62** | 1.92 | 5.68 | 12.96 | 60.82 | 1.92 | 2.15 | 8.33 | 22.87 | **5.67** | 2.46 | 6.72 | 21.66 | 89.32 | 1.97 | 3.16 | 15.21 | 26.60 |

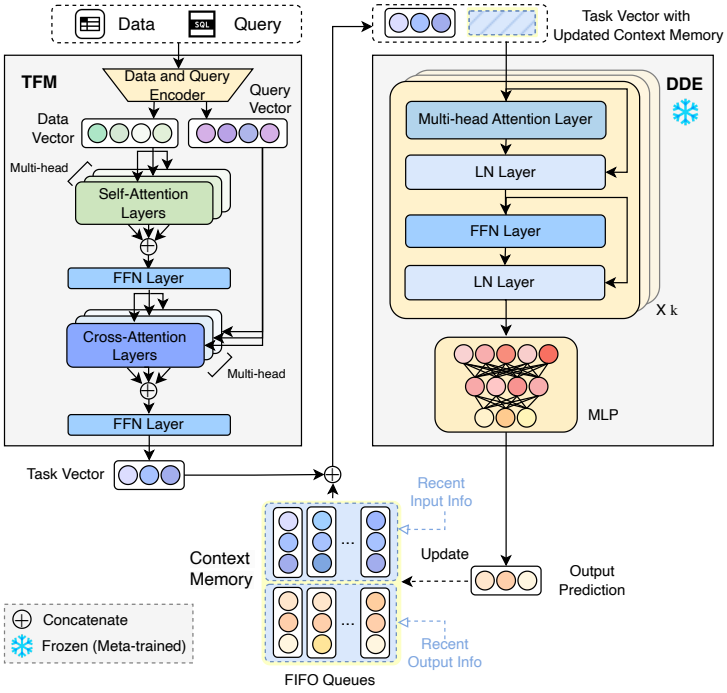


Figure A: The architecture of FLAIR with in-context adaptation via dynamic context memory constructed from recent observed input-output pairs, enabling efficient adaptation to emerging concepts in a single feedforward pass without parameter updates.

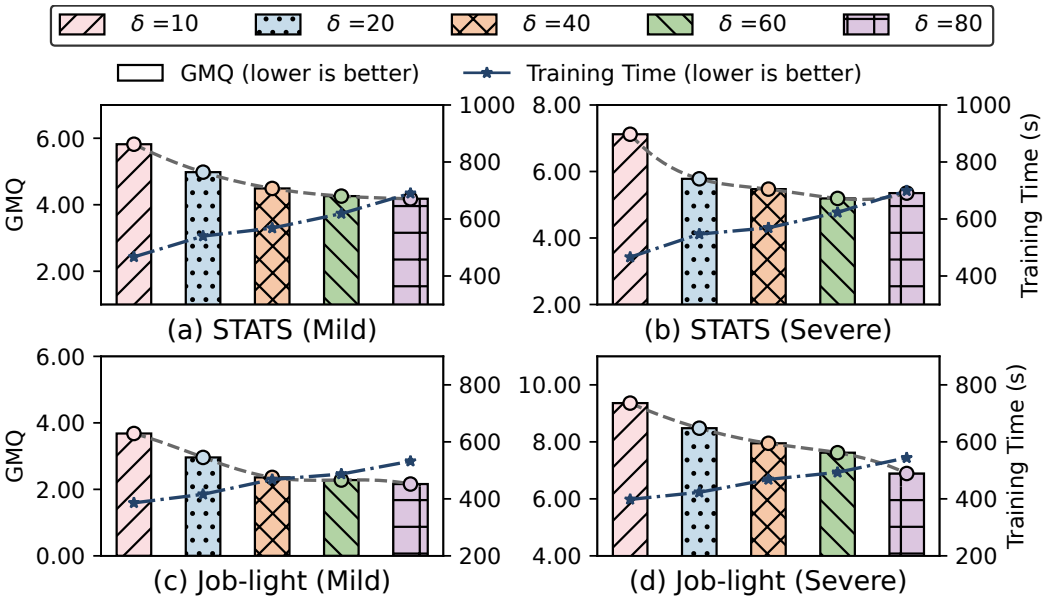


Figure B: Sensitivity analysis of the bin number $\delta$ on model effectiveness and efficiency.

Table E: Summary and comparison of prevailing concept drift learning methods.

| Category | Core Mechanism | Effectiveness | Efficiency | Representative Work |
|---|---|---|---|---|
| Lazy Methods | Retraining | Moderate (depends on retraining quality) | Low (due to full model retraining) | AiRStream (Machine Learning 2022)<br>CDTMSW (Information Sciences 2022) |
| Incremental Methods | Incremental Model Update | Moderate (depends on update strategy) | Low (due to frequent updates) | Deltagrad (ICML 2020)<br>ADF (Pattern Recognition 2022)<br>AIMED (Information Systems 2023)<br>ICICLE (ICCV 2023) |
| Ensemble Methods | Model Pool | Moderate (limited by model pool size) | Low (due to maintaining multiple models) | ARCUS (KDD 2022)<br>Targeted EL (TKDE 2024)<br>OBAL (AAAI 2024) |
| FLAIR | In-context Adaptation | High (context-aware) | High (no parameter updates) | - |

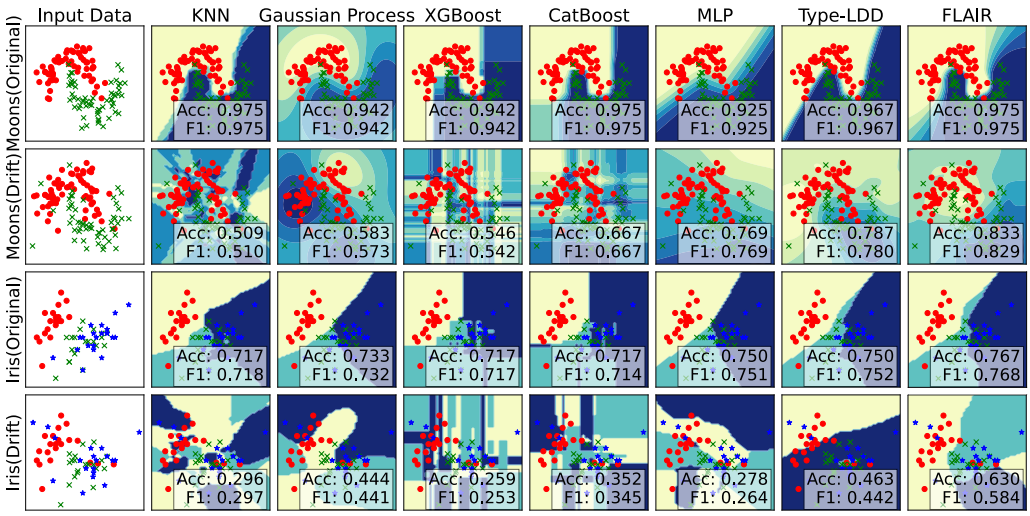


Figure C: Decision boundaries and model performance on data classification task under concept drift.

Table G: Ablation study on latency and partial feedback in GMQ metric, where each setting is evaluated over 5 independent runs and we report the average performance.

| Data | Scenario | FLAIR (Delay 5%) | FLAIR (Delay 10%) | FLAIR (Partial 5%) | FLAIR (Partial 10%) | FLAIR |
|---|---|---|---|---|---|---|
| STATS | Mild | 4.79 | 5.32 | 4.52 | 4.88 | 4.49 |
| | Severe | 5.62 | 5.92 | 5.53 | 5.61 | 5.47 |
| Job-light | Mild | 2.45 | 2.56 | 2.40 | 2.49 | 2.36 |
| | Severe | 8.10 | 8.26 | 8.06 | 8.12 | 7.95 |

Table H: Model efficiency on various benchmark datasets.

| Data | Inference Throughput (query/s) | Memory Footprint (MB) | Inference Latency (ms/query) | Storage Overhead (MB) |
|---|---|---|---|---|
| STATS | 109.17 | 4.97 | 9.16 | 47.68 |
| Job-light | 121.21 | 5.02 | 8.25 | 47.23 |
| TPC-H | 93.28 | 5.11 | 10.82 | 47.72 |