

One-Pass Online Learning under Evolving Feature Data Streams: A Non-Parametric Model (The Supplementary Materials)

Han Zhou, Hongpeng Yin and Bin Wang

This is the supplementary file for the paper entitled *One-Pass Online Learning under Evolving Feature Data Streams: A Non-Parametric Model*. We provide more detailed theoretical analysis, experimental settings and results here. The code is available at <https://github.com/Kan9594/One-Pass-Online-Learning-under-Evolving-Feature-Data-Streams-A-Non-Parametric-Model>

We will ensure that the supplementary material is submitted at a later stage, as the journal's guidelines.

CONTENTS

S1	Symbols and Description	2
S2	Theoretical Proofs	2
S3	Experimental Settings	5
S3-A	Public Datasets	5
S3-B	Settings under General Evolution	5
S3-C	Settings under Synchronized Evolution	6
S3-D	Settings under Overlap Evolution	6
S3-E	Settings of OLEF-KL	7
S3-F	Implementation Details	7
S4	Experiments under General Evolution	8
S4-A	Comparison Results	8
S4-B	Error Curves	8
S4-C	Nemenyi Test	8
S4-D	Parameter Study	9
S4-E	Kernel Comparison	12
S5	Comparison under Synchronized Evolution	15
S5-A	Comparison Results	15
S5-B	Error Curves	16
S5-C	Nemenyi Test	18
S5-D	Parameter Study	18
S6	Experiments under Overlap Evolution	21
S6-A	Comparison Results	21
S6-B	Nemenyi Test	21
S7	Time Analysis	22
S7-A	Theoretical Time complexity Analysis	22
S7-B	Experimental Time comparison	22
References		22

S1. SYMBOLS AND DESCRIPTION

Symbols	Description	Symbols	Description
t	Number of current iteration.	T	Number of total iteration.
T_N	Iteration that the last feature evolution occurs.	$\mathbb{T} = \{1, \dots, T\}$	Set of iteration numbers.
\mathbf{d}_t	Feature set in t -th iteration.	d_t	The dimension of the feature space at the t -th iteration.
N	Times of feature evolving occurs.	\mathcal{D}	Data stream.
$\mathbf{x}_t \in \mathbb{R}^{d_t}$	Sample at the t -th iteration.	$y_t, \hat{y}_t \in \{-1, +1\}$	The true label, predicted label for \mathbf{x}_t .
\mathcal{H}	Hilbert space.	f_t, g_t	Hypothesis in the Hilbert space.
H	Square norm of any hypothesis.	\mathcal{L}	Convex loss function.
$\mathcal{K}(\cdot, \cdot)$	Kernel function.	$J_{i,j}$	Jaccard metric between \mathbf{d}_i and \mathbf{d}_j .
α_i	Coefficient for i -th support vector.	ρ	Decay factor.
η	Learning rate	σ	Bandwidth of Gaussian kernel.
ϵ	Threshold for truncation.		

S2. THEORETICAL PROOFS

This section derives the performance guarantee of the OLEF-KL algorithm. Specifically, two propositions are established. The first one discusses the case when using updating rule in Eq.(8) ($\rho = 1$), and the second one discusses the case when using updating rule in Eq.(9) ($\rho < 1$) based on the analysis of the former one.

Let the g represent a arbitrary classifier in the Hilbert space \mathcal{H}_K . The squared norm of g in \mathcal{H} , denoted as $H^2 = \|g\|_{\mathcal{H}_K}^2$, measures the length or magnitude of the hypothesis g in this space. Denote $\mathcal{L}(f_t; \mathbf{x}_t, y_t)$ as $\mathcal{L}(f_t(\mathbf{x}_t))$ for convenience. The size of the data stream is T . We denote the number of occurrences of feature evolution as N , denote the time that the last feature evolution occurs as T_N and $T_N << T$. Based on the above notations, we can establish the regret bound for the proposed method during the overall online learning procedure.

Proposition 1: Let $f_t, t \in \mathbb{T}$ be a sequence of classifiers generated by the updating rule in Eq.(8). If we assume that $K(\mathbf{x}_i, \mathbf{x}_j) \leq 1$, and the loss function $\mathcal{L}(f_t(\mathbf{x}_t))$ is L -Lipschitz, then for any $g_t \in \mathcal{H}$, the regret of the learned classifier can be bounded as follows:

$$\sum_{t \in \mathbb{T}} \mathcal{L}(f_t(\mathbf{x}_t)) - \sum_{t \in \mathbb{T}} \mathcal{L}(g_t(\mathbf{x}_t)) \leq \frac{H^2}{2\eta} + \eta L^2 T + NLHT_N + \eta NL^2 T_N^2. \quad (\text{S1})$$

If we set the learning rate as $\eta_t = 1/\sqrt{t}$, the algorithm achieves a sublinear regret $\mathcal{O}(\sqrt{T})$.

Proof 1 (Proof of Proposition 1):

The updating rule in Eq.(8) for f_t has two forms based on whether the feature evolution occurs or not at the iteration t . This proof considers two cases separately. Let the set $\{\mathbf{d}_{t-1} = \mathbf{d}_t | t \in \mathbb{T}\}$ as \mathbb{T}_1 and the set $\{\mathbf{d}_{t-1} \neq \mathbf{d}_t | t \in \mathbb{T}\}$ as \mathbb{T}_2 . Clearly, we have $\mathbb{T} = \{\mathbb{T}_1 \cup \mathbb{T}_2\}$.

Case 1: $t \in \mathbb{T}_1$. In these trials, the updating rule in Eq.(8) for f_t can be rewrite as:

$$f_{t+1} = f_t - \eta \nabla \mathcal{L}(f_t(\mathbf{x}_t)) \mathcal{K}(\mathbf{x}_t, \cdot). \quad (\text{S2})$$

Thus, we have the following equation:

$$\begin{aligned} \|f_t - g_t\|_{\mathcal{H}_K}^2 - \|f_{t+1} - g_t\|_{\mathcal{H}_K}^2 &= \|f_t - g_t\|_{\mathcal{H}_K}^2 - \|f_t - g_t - \eta \nabla \mathcal{L}(f_t(\mathbf{x}_t)) \mathcal{K}(\mathbf{x}_t, \cdot)\|_{\mathcal{H}_K}^2 \\ &= 2\eta \nabla \mathcal{L}(f_t(\mathbf{x}_t)) \langle \mathcal{K}(\mathbf{x}_t, \cdot), f_t - g_t \rangle - \|\eta \nabla \mathcal{L}(f_t(\mathbf{x}_t)) \mathcal{K}(\mathbf{x}_t, \cdot)\|_{\mathcal{H}_K}^2. \end{aligned} \quad (\text{S3})$$

Using the reproducing property of \mathcal{H}_K , we have $\langle \mathcal{K}(\mathbf{x}_t, \cdot), f \rangle = f(\mathbf{x}_t)$. Recalling that we also have $\nabla \mathcal{L} \leq L$ and $\mathcal{K}(\mathbf{x}_i, \mathbf{x}_j) \leq 1$. Then, the above formulation can be written as:

$$\|f_t - g_t\|_{\mathcal{H}_K}^2 - \|f_{t+1} - g_t\|_{\mathcal{H}_K}^2 \geq 2\eta \nabla \mathcal{L}(f_t(\mathbf{x}_t)) [f_t(\mathbf{x}_t) - g_t(\mathbf{x}_t)] - \eta^2 L^2. \quad (\text{S4})$$

Rearrange it as:

$$\nabla \mathcal{L}(f_t(\mathbf{x}_t)) [f_t(\mathbf{x}_t) - g_t(\mathbf{x}_t)] \leq \frac{\|f_t - g_t\|_{\mathcal{H}_K}^2 - \|f_{t+1} - g_t\|_{\mathcal{H}_K}^2}{2\eta} + \frac{\eta L^2}{2}. \quad (\text{S5})$$

Due to the convexity of the loss function, we have:

$$\mathcal{L}(f_t(\mathbf{x}_t)) - \mathcal{L}(g_t(\mathbf{x}_t)) \leq \nabla \mathcal{L}(f_t(\mathbf{x}_t)) [f_t(\mathbf{x}_t) - g_t(\mathbf{x}_t)] \quad (\text{S6})$$

Summing over $|\mathbb{T}_1|$, we have the regret in the trials $t \in \mathbb{T}_1$:

$$\begin{aligned} \sum_{t \in \mathbb{T}_1} \{\mathcal{L}(f_t(\mathbf{x}_t)) - \mathcal{L}(g_t(\mathbf{x}_t))\} &\leq \sum_{t \in \mathbb{T}_1} \{\nabla \mathcal{L}(f_t(\mathbf{x}_t)) [f_t(\mathbf{x}_t) - g_t(\mathbf{x}_t)]\} \\ &\leq \frac{\sum_{t \in \mathbb{T}_1} \{\|f_t - g_t\|_{\mathcal{H}_K}^2 - \|f_{t+1} - g_t\|_{\mathcal{H}_K}^2\}}{2\eta} + \frac{\eta L^2}{2} |\mathbb{T}_1| \end{aligned} \quad (\text{S7})$$

Case 2: $t \in \mathbb{T}_2$. The updating rule for f_t includes three steps: 1) dimension adjustment of support vectors; 2) inclusion of new support vector; 3) Jaccard coefficient adaptation. In this situation, the updating rule in trials $|\mathbb{T}_2|$ can be expressed as:

$$\begin{aligned} f_t &= \sum_{\tau=1}^{t-1} \alpha_\tau J_{\tau,t} \mathcal{K}(\Pi_{\mathfrak{d}_t} \mathbf{x}_\tau, \cdot), \\ f'_t &= f_t - \underbrace{\eta \nabla \mathcal{L}(f_t) \mathcal{K}(\mathbf{x}_t, \cdot)}_{\Phi_{A_t}}, \\ f_{t+1} &= f'_t + \underbrace{\sum_{\tau \in \mathbb{T} \setminus \{t\}} \eta \Delta J_{\tau,t} \nabla \mathcal{L}(f_\tau) \mathcal{K}(\Pi_{\mathfrak{d}_t} \mathbf{x}_\tau, \cdot)}_{\Phi_{B_t}}, \end{aligned} \quad (\text{S8})$$

where $\Delta J_{\tau,t} = \Delta J_{\tau,t-1} - \Delta J_{\tau,t}$ is the Jaccard coefficient difference. Note that, it follows $0 \leq \Delta J_{\tau,t} < 1$ and $\sum_{t \in \mathbb{T}} \Delta J_{\tau,t} \leq 1$ for any $\tau \in \mathbb{T}$.

Based on the above updating rule and notations, the Eq.(S3) has the form:

$$\begin{aligned} \|f_t - g_t\|_{\mathcal{H}_K}^2 - \|f_{t+1} - g_t\|_{\mathcal{H}_K}^2 &= \|f_t - g_t\|_{\mathcal{H}_K}^2 - \|f_t - g_t + \Phi_{B_t} - \Phi_{A_t}\|_{\mathcal{H}_K}^2 \\ &= 2 \langle f_t - g_t, \Phi_{A_t} - \Phi_{B_t} \rangle - \|\Phi_{B_t} - \Phi_{A_t}\|_{\mathcal{H}_K}^2. \end{aligned} \quad (\text{S9})$$

Using the reproducing property $\langle \mathcal{K}(\mathbf{x}_t, \cdot), f \rangle = f(\mathbf{x}_t)$, the first term of Eq.(S9) is:

$$\langle f_t - g_t, \Phi_{A_t} - \Phi_{B_t} \rangle = \eta \nabla \mathcal{L}(f_t(\mathbf{x}_t)) [f_t(\mathbf{x}_t) - g_t(\mathbf{x}_t)] + \langle f_t - g_t, -\Phi_{B_t} \rangle. \quad (\text{S10})$$

Thus, rearranging Eq.(S9) yields:

$$\nabla \mathcal{L}(f_t(\mathbf{x}_t)) [f_t(\mathbf{x}_t) - g_t(\mathbf{x}_t)] = \frac{\|f_t - g_t\|_{\mathcal{H}_K}^2 - \|f_{t+1} - g_t\|_{\mathcal{H}_K}^2}{2\eta} + \frac{2 \langle f_t - g_t, \Phi_{B_t} \rangle + \|\Phi_{B_t} - \Phi_{A_t}\|_{\mathcal{H}_K}^2}{2\eta} \quad (\text{S11})$$

Using the convexity of loss function and summing over $|\mathbb{T}_2|$ achieve the regret in trials \mathbb{T}_2 :

$$\begin{aligned} \sum_{t \in \mathbb{T}_2} \{\mathcal{L}(f_t(\mathbf{x}_t)) - \mathcal{L}(g_t(\mathbf{x}_t))\} &\leq \sum_{t \in \mathbb{T}_2} \{\nabla \mathcal{L}(f_t(\mathbf{x}_t)) [f_t(\mathbf{x}_t) - g_t(\mathbf{x}_t)]\} \\ &\leq \frac{\sum_{t \in \mathbb{T}_2} \{\|f_t - g_t\|_{\mathcal{H}_K}^2 - \|f_{t+1} - g_t\|_{\mathcal{H}_K}^2\}}{2\eta} + \frac{\sum_{t \in \mathbb{T}_2} \{2 \langle f_t - g_t, \Phi_{B_t} \rangle + \|\Phi_{B_t} - \Phi_{A_t}\|_{\mathcal{H}_K}^2\}}{2\eta} \end{aligned} \quad (\text{S12})$$

Note that, since we assume that feature evolution occurs N times, we then have $|\mathbb{T}_2| = N$. Thus, we have:

$$\begin{aligned} \sum_{t \in \mathbb{T}_2} \langle f_t - g_t, \Phi_{B_t} \rangle &= \sum_{t \in \mathbb{T}_2} \sum_{\tau \in \mathbb{T} \setminus \{t\}} \eta \Delta J_{\tau,t} \nabla \mathcal{L}(f_\tau(\mathbf{x}_\tau)) [f_\tau(\Pi_{\mathfrak{d}_t} \mathbf{x}_\tau) - g_\tau(\Pi_{\mathfrak{d}_t} \mathbf{x}_\tau)] \\ &\leq \eta N L H \cdot \text{MAX}\{\mathbb{T}_2\} \\ &\leq \eta N L H T_N, \end{aligned} \quad (\text{S13})$$

where $T_N = \text{MAX}\{\mathbb{T}_2\}$ is the maximum number in the set \mathbb{T}_2 , denoting the time the last feature evolution occurs. Meanwhile, we have:

$$\begin{aligned} \sum_{t \in \mathbb{T}_2} \|\Phi_{B_t} - \Phi_{A_t}\|_{\mathcal{H}_K}^2 &= \sum_{t \in \mathbb{T}_2} \left\| \sum_{\tau \in \mathbb{T} \setminus \{t\}} \eta \Delta J_{\tau,t} \nabla \mathcal{L}(f_\tau) \mathcal{K}(\Pi_{\mathfrak{d}_t} \mathbf{x}_\tau, \cdot) - \eta \nabla \mathcal{L}(f_t) \mathcal{K}(\Pi_{\mathfrak{d}_t} \mathbf{x}_t, \cdot) \right\|_{\mathcal{H}_K}^2 \\ &\leq \eta^2 N L^2 (\text{MAX}\{\mathbb{T}_2\})^2 \\ &\leq \eta^2 N L^2 T_N^2 \end{aligned} \quad (\text{S14})$$

Combining Eq.(S12), Eq.(S13) and Eq.(S14) yields:

$$\sum_{t \in \mathbb{T}_2} \{\mathcal{L}(f_t(\mathbf{x}_t)) - \mathcal{L}(g_t(\mathbf{x}_t))\} \leq \frac{\sum_{t \in \mathbb{T}_2} \{\|f_t - g_t\|_{\mathcal{H}_K}^2 - \|f_{t+1} - g_t\|_{\mathcal{H}_K}^2\}}{2\eta} + N L H T_N + \eta N L^2 T_N^2. \quad (\text{S15})$$

Combing Eq.(S7) and Eq.(S15) achieves the overall regret of the proposed algorithm:

$$\sum_{t \in \mathbb{T}} \mathcal{L}(f_t(\mathbf{x}_t)) - \sum_{t \in \mathbb{T}} \mathcal{L}(g_t(\mathbf{x}_t)) \leq \frac{H^2}{2\eta} + \eta L^2 T + N L H T_N + \eta N L^2 T_N^2. \quad (\text{S16})$$

This completes the proof. ■

Remark 1: In Proposition 1, setting the learning rate as $\eta_t = 1/\sqrt{t}$ achieves a regret in the order of $\mathcal{O}(\sqrt{T})$. This means that the cumulative loss disagreement across f_t and g_t is sublinear with respect to the time horizon. The sublinearity implies that the time average regret $\frac{\text{Regret}(T)}{T}$ is asymptotically non-positive as the number of data number goes to infinity. This suggests that, asymptotically, the learning model sequence incrementally generated by Algorithm 1 has performance no worse than any model g_t in \mathcal{H}_K .

Remark 2: From the Proposition 1, we can see that the regret under evolving features is unsurprisingly related to N and T_N . Note that, we have $T_N << T$ as N is finite and T approaches to infinity. Therefore, the regret of the proposed algorithm still grows at a sublinear rate.

Next, let us consider the updating rules with exponentially weighted decay strategy in Eq.(9) ($\rho < 1$). Based on the same notations and assumptions as previously introduced, we then establish the following regret bound.

Proposition 2: Using the same notations and assumptions in Proposition 1, when $\rho < 1$, the regret of the learned classifier generated by Eq.(9) is sublinear ($\mathcal{O}(\sqrt{T})$) by setting ρ_t as $1 - \frac{1}{\sqrt{t}}$ and η_t as $\frac{1}{\sqrt{t}}$.

$$\sum_{t \in \mathbb{T}} \mathcal{L}(f_t(\mathbf{x}_t)) - \sum_{t \in \mathbb{T}} \mathcal{L}(g_t(\mathbf{x}_t)) \leq \frac{H^2}{\eta} + \eta L^2 T + (1 - \rho)LHT + NLHT_N + \eta NL^2 T_N^2 + NH^2. \quad (\text{S17})$$

Proof 2 (Proof of Proposition 2):

Similarly to the previous proof, we consider two cases separately: 1) without feature evolution and 2) with feature evolution.

Case 1: $t \in \mathbb{T}_1$. In these trials, the updating rule in Eq.(9) for f_t can be rewrite as:

$$f_{t+1} = f_t - \eta \nabla \mathcal{L}(f_t(\mathbf{x}_t)) \mathcal{K}(\mathbf{x}_t, \cdot) - (1 - \rho)f_t. \quad (\text{S18})$$

Thus, we have the following equation:

$$\begin{aligned} \|f_{t+1} - g_t\|_{\mathcal{H}_K}^2 &= \|f_t - g_t - \eta \nabla \mathcal{L}(f_t(\mathbf{x}_t)) \mathcal{K}(\mathbf{x}_t, \cdot) - (1 - \rho)f_t\|_{\mathcal{H}_K}^2 \\ &= \|f_t - g_t\|_{\mathcal{H}_K}^2 - 2\eta \nabla \mathcal{L}(f_t(\mathbf{x}_t)) \langle \mathcal{K}(\mathbf{x}_t, \cdot), f_t - g_t \rangle - (2 - 2\rho) \langle f_t, f_t - g_t \rangle \\ &\quad + \|\eta \nabla \mathcal{L}(f_t(\mathbf{x}_t)) \mathcal{K}(\mathbf{x}_t, \cdot) + (1 - \rho)f_t\|_{\mathcal{H}_K}^2 \\ &\leq \|f_t - g_t\|_{\mathcal{H}_K}^2 - 2\eta \nabla \mathcal{L}(f_t(\mathbf{x}_t)) [f_t(\mathbf{X}_t) - g_t(\mathbf{X}_t)] + \eta^2 L^2 + (1 - \rho)(2\eta LH + H^2). \end{aligned} \quad (\text{S19})$$

Rearranging it yields:

$$\nabla \mathcal{L}(f_t(\mathbf{x}_t)) [f_t(\mathbf{x}_t) - g_t(\mathbf{x}_t)] \leq \frac{\|f_t - g_t\|_{\mathcal{H}_K}^2 - \|f_{t+1} - g_t\|_{\mathcal{H}_K}^2}{2\eta} + \frac{\eta L^2}{2} + (1 - \rho)(LH + \frac{H^2}{2\eta}) \quad (\text{S20})$$

Using the convexity of the loss function and summing the above inequality over $|\mathbb{T}_1|$ yields:

$$\sum_{t \in \mathbb{T}_1} \{\mathcal{L}(f_t(\mathbf{x}_t)) - \mathcal{L}(g_t(\mathbf{x}_t))\} \leq \frac{\sum_{t \in \mathbb{T}_1} \left\{ \|f_t - g_t\|_{\mathcal{H}_K}^2 - \|f_{t+1} - g_t\|_{\mathcal{H}_K}^2 \right\}}{2\eta} + \frac{\eta L^2}{2} |\mathbb{T}_1| + (1 - \rho)(LH + \frac{H^2}{2\eta}) |\mathbb{T}_1|. \quad (\text{S21})$$

Case 2: $t \in \mathbb{T}_2$. Using the same notations in Eq.(S8), the updating rule can be rewritten as:

$$f_{t+1} = f_t - \Phi_{A_t} + \Phi_{B_t} - (1 - \rho)f_t. \quad (\text{S22})$$

and then we have:

$$\begin{aligned} \|f_{t+1} - g_t\|_{\mathcal{H}_K}^2 &= \|f_t - g_t - \Phi_{A_t} + \Phi_{B_t} - (1 - \rho)f_t\|_{\mathcal{H}_K}^2 \\ &= \|f_{t+1} - g_t\|_{\mathcal{H}_K}^2 - 2 \langle f_t - g_t, \Phi_{A_t} - \Phi_{B_t} \rangle - 2(1 - \rho) \langle f_t, f_t - g_t \rangle + \|\Phi_{A_t} - \Phi_{B_t} + (1 - \rho)f_t\|_{\mathcal{H}_K}^2 \\ &\leq \|f_{t+1} - g_t\|_{\mathcal{H}_K}^2 - 2\eta \nabla \mathcal{L}(f_t(\mathbf{x}_t)) [f_t(\mathbf{X}_t) - g_t(\mathbf{X}_t)] + 2 \langle f_t - g_t, \Phi_{B_t} \rangle - 2(1 - \rho) \langle f_t, f_t - g_t \rangle \\ &\quad + \|\Phi_{A_t} - \Phi_{B_t}\|_{\mathcal{H}_K}^2 + \|(1 - \rho)f_t\|_{\mathcal{H}_K}^2 \\ &\leq \|f_{t+1} - g_t\|_{\mathcal{H}_K}^2 - 2\eta \nabla \mathcal{L}(f_t(\mathbf{x}_t)) [f_t(\mathbf{X}_t) - g_t(\mathbf{X}_t)] + \eta LHT_N + \eta^2 L^2 T_N^2 + H^2. \end{aligned} \quad (\text{S23})$$

Rearranging it and summing over $|\mathbb{T}_2|$ yields:

$$\sum_{t \in \mathbb{T}_2} \{\mathcal{L}(f_t(\mathbf{x}_t)) - \mathcal{L}(g_t(\mathbf{x}_t))\} \leq \frac{\sum_{t \in \mathbb{T}_2} \left\{ \|f_t - g_t\|_{\mathcal{H}_K}^2 - \|f_{t+1} - g_t\|_{\mathcal{H}_K}^2 \right\}}{2\eta} + LHT_N |\mathbb{T}_2| + \eta L^2 T_N^2 |\mathbb{T}_2| + H^2 |\mathbb{T}_2| \quad (\text{S24})$$

Combining Eq.(S21) and Eq.(S24) and using $|\mathbb{T}_2| = N$ lead to the regret under the exponentially weighted decay:

$$\sum_{t \in \mathbb{T}} \mathcal{L}(f_t(\mathbf{x}_t)) - \sum_{t \in \mathbb{T}} \mathcal{L}(g_t(\mathbf{x}_t)) \leq \frac{H^2}{\eta} + \eta L^2 T + (1 - \rho)LHT + NLHT_N + \eta NL^2 T_N^2 + NH^2. \quad (\text{S25})$$

This completes the proof. ■

Remark 3: Unsurprisingly, compared with the regret in Proposition 1, the regret when setting $\rho < 1$ incurs extra regret. If one sets ρ_t as $1 - \frac{1}{\sqrt{t}}$ and η_t as $\frac{1}{\sqrt{t}}$, these two terms can be in $\mathcal{O}(\sqrt{T})$. Consequently, the overall regret when using the exponentially weighted decay strategy ($\rho < 1$) can also be sublinear with respect to the time horizon. The sublinearity suggests that, asymptotically, the learning model sequence incrementally generated by Algorithm 1 when $\rho < 1$ has performance no worse than any model g_t in $\mathcal{H}_{\mathcal{K}}$.

S3. EXPERIMENTAL SETTINGS

A. Public Datasets

We selected a set of 15 publicly available datasets from the UCI Repository. The selection process was executed with a certain level of randomness, ensuring that our dataset collection is representative of diverse sizes and domains. A comprehensive description of these 15 binary-class datasets can be found in Table S1. To simulate feature evolution, we have drawn upon prior research [1] to guide our settings of feature evolution. Specifically, we generated a random feature set, denoted as \mathbb{D}_{rand} , which is viewed as disappeared features. Simultaneously, we mapped original data into an alternative feature space \mathbb{D}_{new} through the application of random Gaussian matrices. Consequently, we obtained appearing features. For details of the settings of feature evolution, we direct the reader to the forthcoming subsections.

TABLE S1: The details of datasets. \mathbb{D}_{orig} , \mathbb{D}_{rand} and \mathbb{D}_{new} are the feature set in the original, random and new feature space, respectively.

	#Samples	$ \mathbb{D}_{orig} $	$ \mathbb{D}_{rand} $	$ \mathbb{D}_{new} $
HTRU2	17898	8	4	12
TUANDROMD	4464	240	120	30
credit-a	653	15	7	22
dna	949	180	90	30
ijcmn1	49990	22	11	30
magic04	19020	15	5	15
mushroom	8124	21	10	30
musk2	6598	166	83	30
phishing	11055	68	34	30
rna	59535	8	4	12
sensors	10638	48	24	30
spambase	4601	57	28	30
splice	3175	60	30	30
svmguide1	7089	4	2	6
wdbc	569	30	15	30

B. Settings under General Evolution

The simulation of general evolving feature data streams was conducted as shown in Figure S1. Each dataset comprises two sets of features: the random feature set and the original feature set. Particularly, to simulate feature disappearance, random features were sequentially removed, and to simulate feature appearance, original features were introduced across S periods, where S takes on values from the set $\{2, 3, 4, 5\}$. For instance, when $S = 2$, half of the features from \mathbb{D}_{orig} were provided in the first period, and the remaining features were introduced in the subsequent period. Concurrently, all features from \mathbb{D}_{rand} were initially available, but they were systematically removed throughout the entire process until none of them remained.

We compared the proposed OLEF-KL with six algorithms in this scenario.

- **NPA** [2] and **NOGD** [3]: NOGD and NPA represent traditional one-pass online learning algorithms. In our setup, when the feature space undergoes changes, both NOGD and NPA are initiated from the beginning.
- **OLSF** [4] and **OLSF_{OGD}**: OLSF is a model primarily designed for trapezoidal evolution scenarios, employing vector augmentation to incorporate newly appeared features while maintaining model sparsity to forget disappeared features. It employs a PA-like optimization method, and we also make slight extensions (OLSF_{OGD}) by incorporating OGD-like optimization methods.
- **OLVF** [5]: The OLVF algorithm deals with the evolving features by dynamically projecting the classifier and the training instances onto their shared feature subspace. Particularly, the projection weights were dynamically adjusted by learning projection confidence.
- **OLI2DS** [6]: The OLI2DS algorithm is an extension of OLVF. It replaces the projection weights by feature uncertainty and introduces further modifications to align itself with the challenges posed by class imbalance and concept drift.

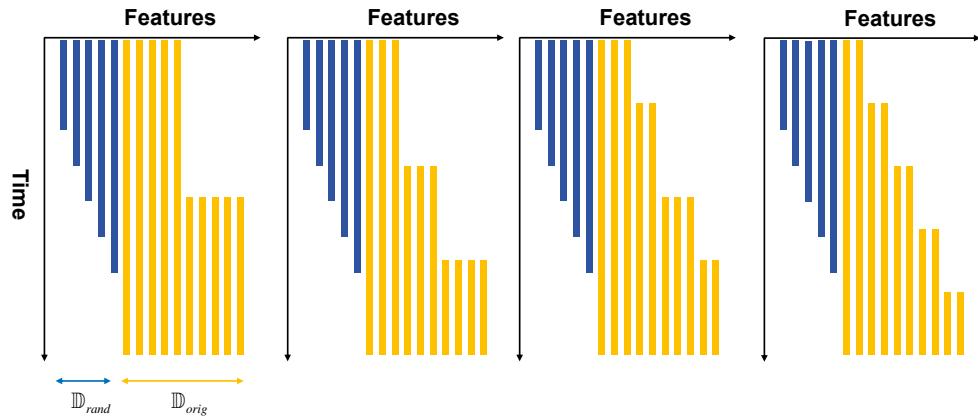


Fig. S1: Visualization of generated general evolving data streams: scenarios ranging from two to five periods.

C. Settings under Synchronized Evolution

The simulation of evolving feature data streams was conducted as shown in Figure S2. For each dataset, it was systematically partitioned into a predetermined number of periods denoted as S , where S takes on values from the set $\{2, 3, 4, 5\}$. In the initial period, we introduced a combination of features sourced from 100% of \mathbb{D}_{rand} , and features originating from $\frac{100}{S}$ % of \mathbb{D}_{orig} . Subsequently, in the second period, a dynamic adjustment was made. Specifically, $\frac{100}{S-1}$ % of the features from \mathbb{D}_{rand} were removed. Meanwhile, an analogous proportion of $\frac{100}{S-1}$ % of features from \mathbb{D}_{orig} , distinct from those present in the first period, were introduced. These deliberate adjustments were made to stimulate the phenomenon of feature disappearance and appearance within the evolving feature data streams.

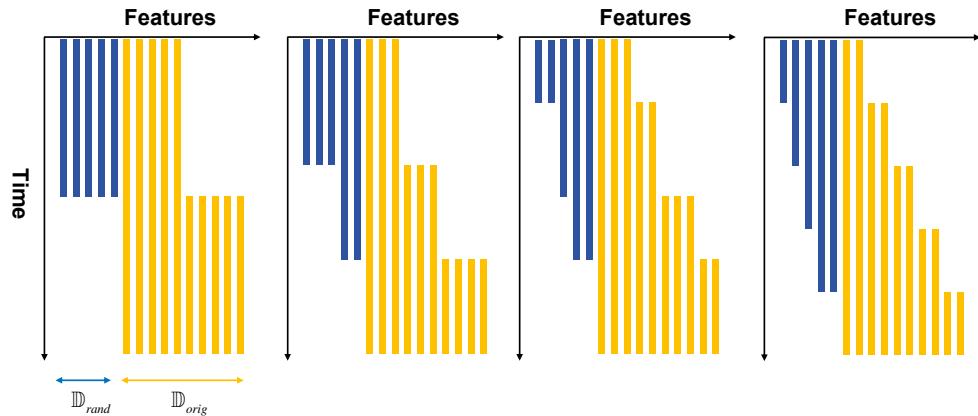


Fig. S2: Visualization of generated synchronized evolving data streams: scenarios ranging from two to five periods.

We compared the proposed OLEF-KL with **NOGD**, **NPA**, **OLSF_{OGD}**, **OLSF** and two additional methods:

- **FF**: First Features (FF) learns a linear classifier based on features received in the first period over all the learning procedure.
- **OPID** [7]: The OPID algorithm is able to learn from evolving feature streams by compressing information of disappeared features, and expanding to include the appeared features.

D. Settings under Overlap Evolution

This scenario simulates evolving feature data streams as shown with a overlap, as shown in Figure S3. Specifically, each data stream was divided into three periods. The first period exclusively consisted of synthetic features \mathbb{D}_{orig} , while the second period included both the original and new features, stimulating the overlap period. In the third period, the original features were removed, leaving only the new features \mathbb{D}_{new} .

In this scenario, we compare OLEF-KL with five methods, including **OLSF_{OGD}** and **OLSF** as described in the previous section and the following algorithms:

- **FESL** [1]: FESL involves the recovery of features that have previously disappeared, followed by the construction of two distinct models for both the recovered features and the current features. The ensemble of these two models is accomplished through the utilization of two distinctive ensemble methods, denoted as **FESL-c** and **FESL-s**.
- **PUFE** [8]: PUFE is an extension of the FESL framework, with an additional focus on addressing missing values during the overlapping period of feature evolution.

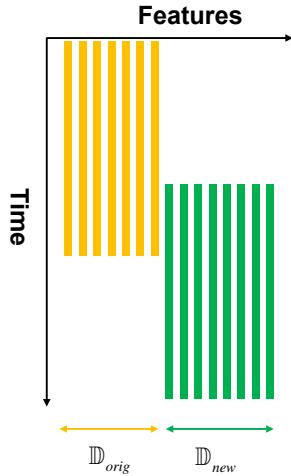


Fig. S3: Visualization of generated overlap evolving data streams.

E. Settings of OLEF-KL

There are two crucial parameters η , ρ needing to be determined in OLEF-KL. Empirically, we have set η to a value of 1 and ρ to 0.999 as our initial configuration. Unless otherwise specified, we use Gaussian kernel function in the proposed method, denoted as $\mathcal{K}(\mathbf{x}_i, \mathbf{x}_j) = e^{\|\mathbf{x}_i - \mathbf{x}_j\|^2/\sigma}$, with σ assigned a value of 0.2. The comprehensive exploration of parameter sensitivity will be addressed in forthcoming subsections.

F. Implementation Details

Our experimental methodology adhered to the Test-Then-Train strategy. Namely, each algorithm was presented with an unlabeled sample and made a prediction. Subsequently, the true label was received for the update of models. We performed a total of 20 independent runs for each dataset and each algorithm. The outcomes were presented as the average performance metrics, accompanied by their respective standard deviations. The assessment of performance was conducted through the measurement of accuracy (ACC) and the Area Under the Curve (AUC).

S4. EXPERIMENTS UNDER GENERAL EVOLUTION

This section evaluates the performance of OLEF-KL under general evolution scenarios.

A. Comparison Results

The classification results, measured by ACC and AUC, for 15 generated evolving feature data streams are presented in Table S2, Table S3, Figure S4 and Figure S5. Notably, our proposed OLEF-KL algorithm exhibits superior performance relative to other methods in most instances, particularly with regard to ACC. This enhanced performance can be attributed to the inherent flexibility of our non-parametric model. Our non-parametric algorithm refrains from imposing strong assumptions regarding the functional form of the classifier and the underlying distribution of data streams. Consequently, it excels at modeling intricate, nonlinear relationships within data streams. Therefore, our algorithm is particularly advantageous in scenarios where the true distribution of data or features remains unknown. In terms of AUC performance, as depicted in Table S3, our algorithm falls slightly short of OLI2DS, particularly in datasets such as *ijcnn1*, *musk2*, *rna*, and *spambase*. This performance differential can be attributed to the design of a cost-sensitive loss function in OLI2DS to address imbalanced data streams. It is worth noting that our research primarily focuses on the development of algorithms for learning from evolving feature data streams, and addressing data imbalance falls outside the scope of this study. This area, however, presents an interesting topic for future research.

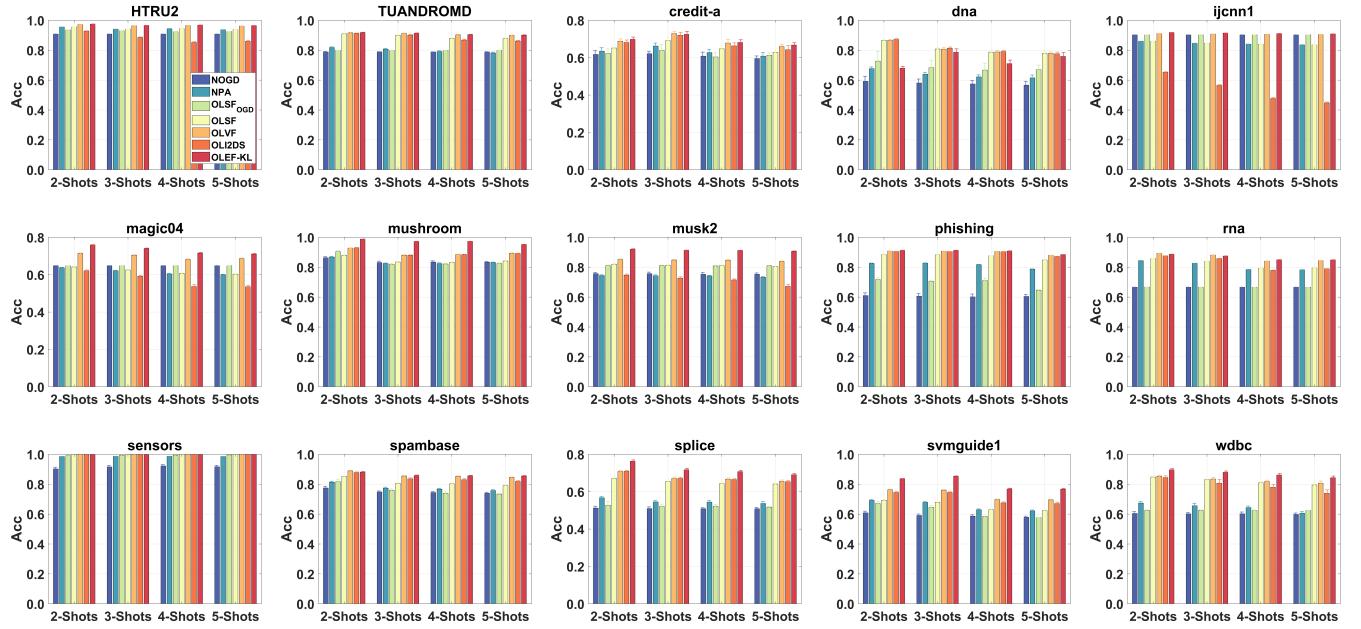


Fig. S4: The performance comparison under general evolution scenarios in terms of ACC.

B. Error Curves

Figure S6 illustrates the error rate trends of various algorithms under 2-period general evolving scenarios. Due to the page limitation, only four datasets are selected to be reported here. As we can see, all the curves generally decreases over time, indicating the fact that the performance of all methods becomes better with more data coming in this period. Particularly, the red curves, which correspond to the OLEF-KL algorithm, exhibit a more obvious reduction in error rates. This trend indicates the better performance of OLEF-KL when learning from feature evolution.

C. Nemenyi Test

Furthermore, to conduct a more rigorous assessment of the relative performance across different methods with respect to ACC and AUC, we employ the Nemenyi test. Within this analytical framework, our approach is considered as the benchmark or dominating approach. The difference between the average ranks of our approach and each of the other comparative approaches is judged by reference to the critical difference statistic, which is defined as follows: $CD = q_\alpha \sqrt{K(K+1)/6N}$.

In the context of the Nemenyi test, the critical difference is computed as 1.1627, with parameters set at $q_{0.05} = 2.984$, $K = 7$, and $N = 60$. Subsequent analysis, as depicted in Figure S7, reveals that OLEF-KL holds a significant advantage over competing methods, including NOGD, NPA, OLSF_{OGD}, and OLSF, across both the ACC and AUC metrics. Furthermore, our proposed method exhibits a marginal superiority compared to OLVF. Whereas, OLI2DS displays a slight advantage in AUC, leveraging its specialized design for imbalanced data. Despite this, OLEF-KL remains competitively effective in comparison.

TABLE S2: The Performance comparison in general evolution scenarios, in term of ACC (mean \pm std). •/○ indicate that OLEF-KL is significantly better/worse than competitors (hypothesis supported by paired t-tests at 0.05 significant level.). The best results are indicated in **bold**.

Datasets	Periods	NOGD	NPA	OLSF _{OGD}	OLSF	OLVF	OLI2DS	OLEF-KL
HTRU2	2	0.908 \pm 0.000•	0.955 \pm 0.001•	0.936 \pm 0.001•	0.956 \pm 0.001•	0.973 \pm 0.001•	0.928 \pm 0.003•	0.975\pm0.001
	3	0.908 \pm 0.000•	0.940 \pm 0.001•	0.930 \pm 0.001•	0.942 \pm 0.001•	0.963 \pm 0.001•	0.885 \pm 0.004•	0.966\pm0.001
	4	0.908 \pm 0.000•	0.943 \pm 0.001•	0.924 \pm 0.001•	0.944 \pm 0.001•	0.965 \pm 0.001•	0.852 \pm 0.004•	0.968\pm0.001
	5	0.908 \pm 0.000•	0.936 \pm 0.002•	0.924 \pm 0.001•	0.938 \pm 0.001•	0.962 \pm 0.001•	0.861 \pm 0.007•	0.964\pm0.001
TUANDROMD	2	0.789 \pm 0.002•	0.819 \pm 0.005•	0.799 \pm 0.000•	0.909 \pm 0.004•	0.918\pm0.003	0.912 \pm 0.004•	0.917\pm0.003
	3	0.789 \pm 0.002•	0.806 \pm 0.005•	0.799 \pm 0.000•	0.902 \pm 0.004•	0.914\pm0.004	0.901 \pm 0.004•	0.913\pm0.003
	4	0.788 \pm 0.002•	0.789 \pm 0.005•	0.798 \pm 0.000•	0.877 \pm 0.004•	0.900 \pm 0.002•	0.866 \pm 0.005•	0.903\pm0.002
	5	0.788 \pm 0.002•	0.784 \pm 0.004•	0.798 \pm 0.000•	0.879 \pm 0.005•	0.900\pm0.003	0.860 \pm 0.006•	0.900\pm0.003
credit-a	2	0.608 \pm 0.017•	0.633 \pm 0.015•	0.621 \pm 0.025•	0.651 \pm 0.015•	0.680 \pm 0.015•	0.672 \pm 0.018•	0.689\pm0.011
	3	0.622 \pm 0.021•	0.662 \pm 0.017•	0.646 \pm 0.027•	0.689 \pm 0.015•	0.727\pm0.017	0.718 \pm 0.013•	0.730\pm0.014
	4	0.611 \pm 0.019•	0.630 \pm 0.021•	0.598 \pm 0.020•	0.651 \pm 0.021•	0.683\pm0.015	0.659 \pm 0.016•	0.687\pm0.011
	5	0.604 \pm 0.019•	0.610 \pm 0.020•	0.614 \pm 0.020•	0.629 \pm 0.017•	0.666\pm0.013	0.640 \pm 0.014•	0.668\pm0.012
dna	2	0.591 \pm 0.028•	0.680 \pm 0.011•	0.716 \pm 0.081•	0.866 \pm 0.008○	0.869 \pm 0.007○	0.872\pm0.008○	0.683 \pm 0.017
	3	0.581 \pm 0.021•	0.636 \pm 0.015•	0.678 \pm 0.034•	0.805 \pm 0.007○	0.811\pm0.007○	0.811\pm0.008○	0.777 \pm 0.019
	4	0.576 \pm 0.029•	0.624 \pm 0.013•	0.669 \pm 0.032•	0.787 \pm 0.010○	0.786 \pm 0.011○	0.792\pm0.009○	0.708 \pm 0.019
	5	0.569 \pm 0.013•	0.611 \pm 0.015•	0.685 \pm 0.031•	0.782\pm0.007○	0.781\pm0.011○	0.783\pm0.010○	0.752 \pm 0.022
ijcnn1	2	0.902 \pm 0.000•	0.860 \pm 0.001•	0.903 \pm 0.000•	0.861 \pm 0.001•	0.911 \pm 0.001•	0.654 \pm 0.006•	0.919\pm0.001
	3	0.902 \pm 0.000•	0.846 \pm 0.001•	0.903 \pm 0.000•	0.847 \pm 0.001•	0.908 \pm 0.001•	0.568 \pm 0.005•	0.914\pm0.000
	4	0.902 \pm 0.000•	0.840 \pm 0.001•	0.903 \pm 0.000•	0.841 \pm 0.001•	0.906 \pm 0.001•	0.478 \pm 0.004•	0.911\pm0.001
	5	0.902 \pm 0.000•	0.837 \pm 0.001•	0.903 \pm 0.000•	0.837 \pm 0.001•	0.905 \pm 0.001•	0.445 \pm 0.004•	0.909\pm0.000
magic04	2	0.648 \pm 0.000•	0.637 \pm 0.003•	0.648 \pm 0.000•	0.641 \pm 0.003•	0.713 \pm 0.002•	0.618 \pm 0.004•	0.759\pm0.002
	3	0.648 \pm 0.000•	0.620 \pm 0.004•	0.648 \pm 0.000•	0.625 \pm 0.004•	0.704 \pm 0.002•	0.593 \pm 0.006•	0.741\pm0.002
	4	0.648 \pm 0.000•	0.605 \pm 0.003•	0.648 \pm 0.000•	0.608 \pm 0.004•	0.683 \pm 0.003•	0.532 \pm 0.007•	0.717\pm0.003
	5	0.648 \pm 0.000•	0.599 \pm 0.003•	0.648 \pm 0.000•	0.603 \pm 0.003•	0.686 \pm 0.003•	0.531 \pm 0.007•	0.710\pm0.003
mushroom	2	0.864 \pm 0.006•	0.869 \pm 0.003•	0.903 \pm 0.007•	0.882 \pm 0.004•	0.929 \pm 0.001•	0.930 \pm 0.001•	0.988\pm0.001
	3	0.827 \pm 0.009•	0.826 \pm 0.004•	0.816 \pm 0.009•	0.835 \pm 0.004•	0.880 \pm 0.003•	0.879 \pm 0.003•	0.971\pm0.001
	4	0.832 \pm 0.008•	0.824 \pm 0.005•	0.817 \pm 0.008•	0.832 \pm 0.005•	0.884 \pm 0.004•	0.883 \pm 0.004•	0.973\pm0.001
	5	0.830 \pm 0.007•	0.831 \pm 0.003•	0.825 \pm 0.006•	0.842 \pm 0.003•	0.892 \pm 0.002•	0.891 \pm 0.002•	0.952\pm0.002
musk2	2	0.755 \pm 0.007•	0.746 \pm 0.005•	0.809 \pm 0.004•	0.821 \pm 0.004•	0.852 \pm 0.003•	0.752 \pm 0.007•	0.921\pm0.002
	3	0.757 \pm 0.012•	0.742 \pm 0.006•	0.810 \pm 0.004•	0.814 \pm 0.004•	0.849 \pm 0.004•	0.732 \pm 0.008•	0.914\pm0.002
	4	0.755 \pm 0.009•	0.742 \pm 0.004•	0.808 \pm 0.004•	0.813 \pm 0.004•	0.847 \pm 0.003•	0.713 \pm 0.008•	0.912\pm0.002
	5	0.753 \pm 0.010•	0.733 \pm 0.004•	0.810 \pm 0.004•	0.805 \pm 0.004•	0.841 \pm 0.005•	0.669 \pm 0.009•	0.906\pm0.002
phishing	2	0.601 \pm 0.013•	0.825 \pm 0.003•	0.718 \pm 0.014•	0.883 \pm 0.003•	0.908 \pm 0.002•	0.906 \pm 0.002•	0.911\pm0.002
	3	0.607 \pm 0.012•	0.827 \pm 0.002•	0.710 \pm 0.008•	0.883 \pm 0.003•	0.907 \pm 0.002•	0.905 \pm 0.002•	0.913\pm0.002
	4	0.601 \pm 0.016•	0.818 \pm 0.003•	0.707 \pm 0.021•	0.877 \pm 0.002•	0.904 \pm 0.001•	0.903 \pm 0.002•	0.909\pm0.002
	5	0.607 \pm 0.014•	0.789 \pm 0.004•	0.649 \pm 0.006•	0.849 \pm 0.003•	0.878 \pm 0.002•	0.871 \pm 0.003•	0.884\pm0.003
rna	2	0.666 \pm 0.000•	0.843 \pm 0.001•	0.667 \pm 0.000•	0.858 \pm 0.001•	0.893\pm0.001○	0.875 \pm 0.002•	0.887 \pm 0.001
	3	0.666 \pm 0.000•	0.826 \pm 0.001•	0.667 \pm 0.000•	0.840 \pm 0.001•	0.882\pm0.001○	0.858 \pm 0.001•	0.875 \pm 0.001
	4	0.666 \pm 0.000•	0.784 \pm 0.002•	0.667 \pm 0.000•	0.793 \pm 0.001•	0.840 \pm 0.001•	0.778 \pm 0.002•	0.849\pm0.001
	5	0.666 \pm 0.000•	0.782 \pm 0.001•	0.667 \pm 0.000•	0.795 \pm 0.001•	0.844 \pm 0.001•	0.789 \pm 0.002•	0.849\pm0.001
sensors	2	0.901 \pm 0.009•	0.985 \pm 0.001•	0.988 \pm 0.006•	0.999\pm0.000○	0.999\pm0.000○	0.999\pm0.000○	0.998 \pm 0.000
	3	0.911 \pm 0.011•	0.986 \pm 0.001•	0.993 \pm 0.005•	0.999\pm0.000○	0.999\pm0.000○	0.999\pm0.000○	0.998 \pm 0.000
	4	0.918 \pm 0.010•	0.987 \pm 0.001•	0.994 \pm 0.006•	0.999\pm0.000○	0.999\pm0.000○	0.999\pm0.000○	0.998 \pm 0.000
	5	0.917 \pm 0.011•	0.986 \pm 0.001•	0.996 \pm 0.004•	0.999\pm0.000○	0.999\pm0.000○	0.999\pm0.000○	0.998 \pm 0.000
spambase	2	0.775 \pm 0.014•	0.815 \pm 0.005•	0.815 \pm 0.016•	0.851 \pm 0.004•	0.890\pm0.003○	0.879 \pm 0.004•	0.881 \pm 0.003
	3	0.750 \pm 0.009•	0.778 \pm 0.007•	0.759 \pm 0.011•	0.807 \pm 0.005•	0.857\pm0.003	0.835 \pm 0.007•	0.858\pm0.002
	4	0.746 \pm 0.006•	0.769 \pm 0.005•	0.738 \pm 0.005•	0.804 \pm 0.006•	0.853 \pm 0.005•	0.830 \pm 0.007•	0.857\pm0.003
	5	0.742 \pm 0.006•	0.759 \pm 0.005•	0.737 \pm 0.009•	0.788 \pm 0.004•	0.846 \pm 0.004•	0.820 \pm 0.007•	0.855\pm0.003
splice	2	0.516 \pm 0.008•	0.568 \pm 0.009•	0.536 \pm 0.027•	0.672 \pm 0.008•	0.706 \pm 0.007•	0.707 \pm 0.007•	0.766\pm0.005
	3	0.510 \pm 0.006•	0.548 \pm 0.010•	0.519 \pm 0.011•	0.659 \pm 0.009•	0.674 \pm 0.007•	0.674 \pm 0.008•	0.716\pm0.006
	4	0.512 \pm 0.009•	0.546 \pm 0.010•	0.518 \pm 0.006•	0.641 \pm 0.006•	0.665 \pm 0.007•	0.663 \pm 0.007•	0.706\pm0.005
	5	0.511 \pm 0.006•	0.537 \pm 0.008•	0.518 \pm 0.009•	0.642 \pm 0.008•	0.658 \pm 0.007•	0.658 \pm 0.007•	0.690\pm0.005
svmguide1	2	0.603 \pm 0.013•	0.692 \pm 0.005•	0.666 \pm 0.010•	0.693 \pm 0.005•	0.764 \pm 0.005•	0.743 \pm 0.005•	0.836\pm0.004
	3	0.594 \pm 0.009•	0.680 \pm 0.005•	0.647 \pm 0.013•	0.681 \pm 0.005•	0.761 \pm 0.005•	0.746 \pm 0.005•	0.854\pm0.003
	4	0.584 \pm 0.007•	0.629 \pm 0.007•	0.583 \pm 0.003•	0.629 \pm 0.007•	0.700 \pm 0.005•	0.679 \pm 0.007•	0.769\pm0.004
	5	0.578 \pm 0.006•	0.623 \pm 0.007•	0.577 \pm 0.004•	0.623 \pm 0.007•	0.694 \pm 0.005•	0.669 \pm 0.008•	0.766\pm0.004
wdbc	2	0.599 \pm 0.011•	0.665 \pm 0.017•	0.625 \pm 0.003•	0.843 \pm 0.010•	0.852 \pm 0.011•	0.831 \pm 0.015•	0.900\pm0.005
	3	0.601 \pm 0.008•	0.659 \pm 0.016•	0.624 \pm 0.003•	0.833 \pm 0.011•	0.834 \pm 0.007•	0.820 \pm 0.017•	0.880\pm0.009
	4	0.599 \pm 0.014•	0.646 \pm 0.018•	0.				

TABLE S3: The Performance comparison in general evolution scenarios, in term of AUC (mean \pm std). \bullet/\odot indicate that OLEF-KL is significantly better/worse than competitors (hypothesis supported by paired t-tests at 0.05 significant level.). The best results are indicated in **bold**.

Datasets	Periods	NOGD	NPA	OLSF _{OGD}	OLSF	OLVF	OLI2DS	OLEF-KL
HTRU2	2	0.498 \pm 0.005 \bullet	0.857 \pm 0.004 \bullet	0.573 \pm 0.006 \bullet	0.862 \pm 0.004 \bullet	0.883 \pm 0.003 \bullet	0.890\pm0.004\odot	0.886 \pm 0.004
	3	0.500 \pm 0.009 \bullet	0.800 \pm 0.006 \bullet	0.550 \pm 0.011 \bullet	0.809 \pm 0.006 \bullet	0.821 \pm 0.007 \bullet	0.861\pm0.006\odot	0.838 \pm 0.006
	4	0.502 \pm 0.008 \bullet	0.811 \pm 0.005 \bullet	0.529 \pm 0.011 \bullet	0.816 \pm 0.005 \bullet	0.830 \pm 0.005 \bullet	0.848\pm0.006	0.849\pm0.006
	5	0.496 \pm 0.009 \bullet	0.784 \pm 0.007 \bullet	0.526 \pm 0.010 \bullet	0.795 \pm 0.006 \bullet	0.814 \pm 0.006 \bullet	0.840\pm0.005\odot	0.829 \pm 0.006
TUANDROMD	2	0.505 \pm 0.010 \bullet	0.708 \pm 0.009 \bullet	0.501 \pm 0.008 \bullet	0.885 \pm 0.006 \bullet	0.897 \pm 0.005 \odot	0.904\pm0.005\odot	0.891 \pm 0.004
	3	0.504 \pm 0.016 \bullet	0.697 \pm 0.009 \bullet	0.501 \pm 0.015 \bullet	0.877 \pm 0.005 \bullet	0.895\pm0.006\odot	0.892 \pm 0.005	0.892 \pm 0.006
	4	0.499 \pm 0.010 \bullet	0.681 \pm 0.008 \bullet	0.498 \pm 0.010 \bullet	0.847 \pm 0.006 \bullet	0.869\pm0.004	0.862 \pm 0.005 \bullet	0.871\pm0.006
	5	0.499 \pm 0.009 \bullet	0.669 \pm 0.010 \bullet	0.498 \pm 0.008 \bullet	0.849 \pm 0.006 \bullet	0.864\pm0.005	0.854 \pm 0.006 \bullet	0.865\pm0.005
credit-a	2	0.608 \pm 0.023 \bullet	0.636 \pm 0.018 \bullet	0.621 \pm 0.038 \bullet	0.653 \pm 0.017 \bullet	0.688\pm0.017	0.664 \pm 0.028 \bullet	0.688\pm0.015
	3	0.618 \pm 0.024 \bullet	0.663 \pm 0.013 \bullet	0.635 \pm 0.042 \bullet	0.690 \pm 0.015 \bullet	0.739\pm0.015	0.714 \pm 0.035 \bullet	0.737\pm0.017
	4	0.606 \pm 0.023 \bullet	0.629 \pm 0.024 \bullet	0.580 \pm 0.042 \bullet	0.651 \pm 0.026 \bullet	0.694\pm0.017	0.641 \pm 0.041 \bullet	0.693\pm0.014
	5	0.600 \pm 0.022 \bullet	0.608 \pm 0.021 \bullet	0.611 \pm 0.035 \bullet	0.627 \pm 0.018 \bullet	0.672\pm0.011	0.630 \pm 0.031 \bullet	0.672\pm0.014
dna	2	0.591 \pm 0.035 \bullet	0.681 \pm 0.017 \bullet	0.726 \pm 0.113 \bullet	0.871 \pm 0.008 \odot	0.878\pm0.013\odot	0.875\pm0.009\odot	0.710 \pm 0.025
	3	0.585 \pm 0.027 \bullet	0.641 \pm 0.015 \bullet	0.702 \pm 0.075 \bullet	0.807 \pm 0.008 \bullet	0.821\pm0.015	0.807 \pm 0.015 \bullet	0.824\pm0.015
	4	0.581 \pm 0.034 \bullet	0.633 \pm 0.014 \bullet	0.689 \pm 0.073 \bullet	0.790 \pm 0.011 \odot	0.790 \pm 0.022 \odot	0.792\pm0.017\odot	0.720 \pm 0.019
	5	0.571 \pm 0.020 \bullet	0.617 \pm 0.018 \bullet	0.691 \pm 0.085 \bullet	0.783\pm0.007\odot	0.786\pm0.016\odot	0.785\pm0.015\odot	0.769 \pm 0.029
ijcnn1	2	0.501 \pm 0.005 \bullet	0.561 \pm 0.006 \odot	0.501 \pm 0.005 \bullet	0.565 \pm 0.005 \bullet	0.554 \pm 0.006 \odot	0.681\pm0.005\odot	0.539 \pm 0.006
	3	0.499 \pm 0.004 \bullet	0.530 \pm 0.004 \odot	0.499 \pm 0.004 \bullet	0.532 \pm 0.004 \odot	0.522 \pm 0.004 \odot	0.627\pm0.004\odot	0.517 \pm 0.004
	4	0.501 \pm 0.003 \bullet	0.522 \pm 0.004 \odot	0.501 \pm 0.003 \bullet	0.522 \pm 0.004 \odot	0.513 \pm 0.003 \odot	0.601\pm0.004\odot	0.512 \pm 0.003
	5	0.501 \pm 0.004 \bullet	0.517 \pm 0.005 \odot	0.501 \pm 0.004 \bullet	0.517 \pm 0.005 \odot	0.508 \pm 0.005 \odot	0.584\pm0.004\odot	0.508 \pm 0.005
magic04	2	0.500 \pm 0.005 \bullet	0.607 \pm 0.004 \bullet	0.499 \pm 0.005 \bullet	0.613 \pm 0.004 \bullet	0.677 \pm 0.003 \bullet	0.635 \pm 0.005 \bullet	0.723\pm0.003
	3	0.502 \pm 0.003 \bullet	0.590 \pm 0.005 \bullet	0.502 \pm 0.003 \bullet	0.596 \pm 0.004 \bullet	0.661 \pm 0.003 \bullet	0.615 \pm 0.006 \bullet	0.703\pm0.003
	4	0.499 \pm 0.004 \bullet	0.574 \pm 0.005 \bullet	0.498 \pm 0.004 \bullet	0.579 \pm 0.005 \bullet	0.639 \pm 0.004 \bullet	0.566 \pm 0.007 \bullet	0.683\pm0.003
	5	0.501 \pm 0.005 \bullet	0.568 \pm 0.004 \bullet	0.501 \pm 0.005 \bullet	0.573 \pm 0.003 \bullet	0.637 \pm 0.003 \bullet	0.566 \pm 0.006 \bullet	0.676\pm0.003
mushroom	2	0.860 \pm 0.014 \bullet	0.867 \pm 0.003 \bullet	0.894 \pm 0.016 \bullet	0.879 \pm 0.004 \bullet	0.922 \pm 0.002 \bullet	0.924 \pm 0.002 \bullet	0.984\pm0.001
	3	0.816 \pm 0.027 \bullet	0.822 \pm 0.004 \bullet	0.806 \pm 0.026 \bullet	0.832 \pm 0.004 \bullet	0.882 \pm 0.004 \bullet	0.885 \pm 0.003 \bullet	0.960\pm0.003
	4	0.829 \pm 0.021 \bullet	0.820 \pm 0.006 \bullet	0.819 \pm 0.018 \bullet	0.828 \pm 0.006 \bullet	0.886 \pm 0.004 \bullet	0.888 \pm 0.004 \bullet	0.963\pm0.003
	5	0.818 \pm 0.026 \bullet	0.827 \pm 0.003 \bullet	0.819 \pm 0.023 \bullet	0.838 \pm 0.004 \bullet	0.894 \pm 0.003 \bullet	0.897 \pm 0.003 \bullet	0.948\pm0.003
musk2	2	0.531 \pm 0.010 \bullet	0.628 \pm 0.008 \bullet	0.521 \pm 0.007 \bullet	0.722 \pm 0.006 \odot	0.706 \pm 0.007 \bullet	0.797\pm0.004\odot	0.711 \pm 0.010
	3	0.524 \pm 0.014 \bullet	0.616 \pm 0.008 \bullet	0.516 \pm 0.013 \bullet	0.709 \pm 0.006 \odot	0.691 \pm 0.007 \odot	0.795\pm0.006\odot	0.677 \pm 0.009
	4	0.525 \pm 0.011 \bullet	0.614 \pm 0.009 \bullet	0.515 \pm 0.009 \bullet	0.708 \pm 0.008 \odot	0.685 \pm 0.006 \odot	0.786\pm0.006\odot	0.675 \pm 0.008
	5	0.525 \pm 0.009 \bullet	0.591 \pm 0.008 \bullet	0.518 \pm 0.008 \bullet	0.687 \pm 0.007 \odot	0.662 \pm 0.008 \odot	0.778\pm0.005\odot	0.650 \pm 0.008
phishing	2	0.571 \pm 0.021 \bullet	0.825 \pm 0.003 \bullet	0.761 \pm 0.009 \bullet	0.884 \pm 0.003 \bullet	0.912 \pm 0.002 \bullet	0.909 \pm 0.003 \bullet	0.919\pm0.002
	3	0.579 \pm 0.021 \bullet	0.826 \pm 0.002 \bullet	0.751 \pm 0.007 \bullet	0.884 \pm 0.003 \bullet	0.910 \pm 0.002 \bullet	0.909 \pm 0.003 \bullet	0.917\pm0.002
	4	0.567 \pm 0.019 \bullet	0.817 \pm 0.003 \bullet	0.743 \pm 0.010 \bullet	0.877 \pm 0.003 \bullet	0.908 \pm 0.002 \bullet	0.906 \pm 0.003 \bullet	0.913\pm0.003
	5	0.579 \pm 0.018 \bullet	0.790 \pm 0.004 \bullet	0.663 \pm 0.006 \bullet	0.852 \pm 0.003 \bullet	0.890 \pm 0.003 \bullet	0.875 \pm 0.008 \bullet	0.899\pm0.003
rna	2	0.501 \pm 0.002 \bullet	0.815 \pm 0.002 \bullet	0.501 \pm 0.002 \bullet	0.832 \pm 0.001 \bullet	0.866 \pm 0.001 \odot	0.889\pm0.001\odot	0.857 \pm 0.002
	3	0.499 \pm 0.002 \bullet	0.795 \pm 0.002 \bullet	0.499 \pm 0.002 \bullet	0.811 \pm 0.002 \bullet	0.851 \pm 0.001 \odot	0.873\pm0.001\odot	0.843 \pm 0.001
	4	0.500 \pm 0.002 \bullet	0.735 \pm 0.002 \bullet	0.500 \pm 0.002 \bullet	0.746 \pm 0.002 \bullet	0.782 \pm 0.003 \bullet	0.833\pm0.002\odot	0.794 \pm 0.002
	5	0.499 \pm 0.003 \bullet	0.735 \pm 0.002 \bullet	0.500 \pm 0.003 \bullet	0.750 \pm 0.002 \bullet	0.790 \pm 0.002 \bullet	0.836\pm0.002\odot	0.797 \pm 0.002
sensors	2	0.902 \pm 0.013 \bullet	0.985 \pm 0.001 \bullet	0.991 \pm 0.012 \bullet	0.999\pm0.000\odot	0.999\pm0.001\odot	0.999\pm0.000\odot	0.998 \pm 0.000
	3	0.910 \pm 0.013 \bullet	0.986 \pm 0.001 \bullet	0.993 \pm 0.008 \bullet	0.999\pm0.000\odot	0.999\pm0.001\odot	0.999\pm0.000\odot	0.998 \pm 0.000
	4	0.921 \pm 0.010 \bullet	0.986 \pm 0.001 \bullet	0.995 \pm 0.007 \bullet	0.999\pm0.000\odot	0.999\pm0.001\odot	0.999\pm0.000\odot	0.998 \pm 0.000
	5	0.917 \pm 0.013 \bullet	0.986 \pm 0.001 \bullet	0.995 \pm 0.008 \bullet	0.999\pm0.000\odot	0.999\pm0.001\odot	0.999\pm0.000\odot	0.998 \pm 0.000
spambase	2	0.675 \pm 0.031 \bullet	0.793 \pm 0.005 \bullet	0.713 \pm 0.038 \bullet	0.834 \pm 0.005 \bullet	0.865 \pm 0.005 \bullet	0.873\pm0.004\odot	0.836 \pm 0.006
	3	0.630 \pm 0.017 \bullet	0.752 \pm 0.008 \bullet	0.623 \pm 0.020 \bullet	0.785 \pm 0.005 \bullet	0.820 \pm 0.006 \odot	0.837\pm0.005\odot	0.802 \pm 0.004
	4	0.629 \pm 0.015 \bullet	0.742 \pm 0.006 \bullet	0.595 \pm 0.012 \bullet	0.783 \pm 0.006 \bullet	0.815 \pm 0.007 \odot	0.836\pm0.006\odot	0.808 \pm 0.005
	5	0.630 \pm 0.016 \bullet	0.731 \pm 0.005 \bullet	0.599 \pm 0.024 \bullet	0.765 \pm 0.005 \bullet	0.805 \pm 0.004 \bullet	0.825\pm0.008\odot	0.805 \pm 0.006
splice	2	0.508 \pm 0.011 \bullet	0.566 \pm 0.010 \bullet	0.511 \pm 0.032 \bullet	0.671 \pm 0.008 \bullet	0.709 \pm 0.009 \bullet	0.703 \pm 0.013 \bullet	0.773\pm0.004
	3	0.505 \pm 0.008 \bullet	0.548 \pm 0.009 \bullet	0.506 \pm 0.015 \bullet	0.660 \pm 0.009 \bullet	0.680 \pm 0.007 \bullet	0.669 \pm 0.014 \bullet	0.729\pm0.011

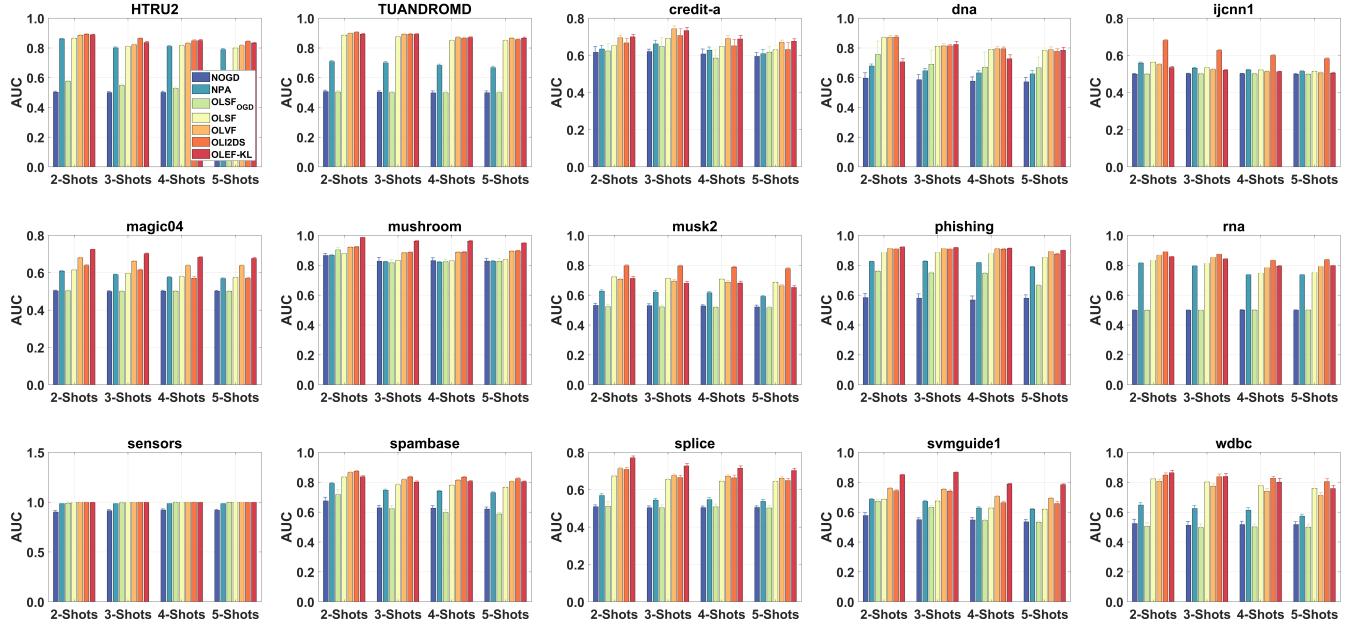


Fig. S5: The performance comparison under general evolution scenarios in terms of AUC.

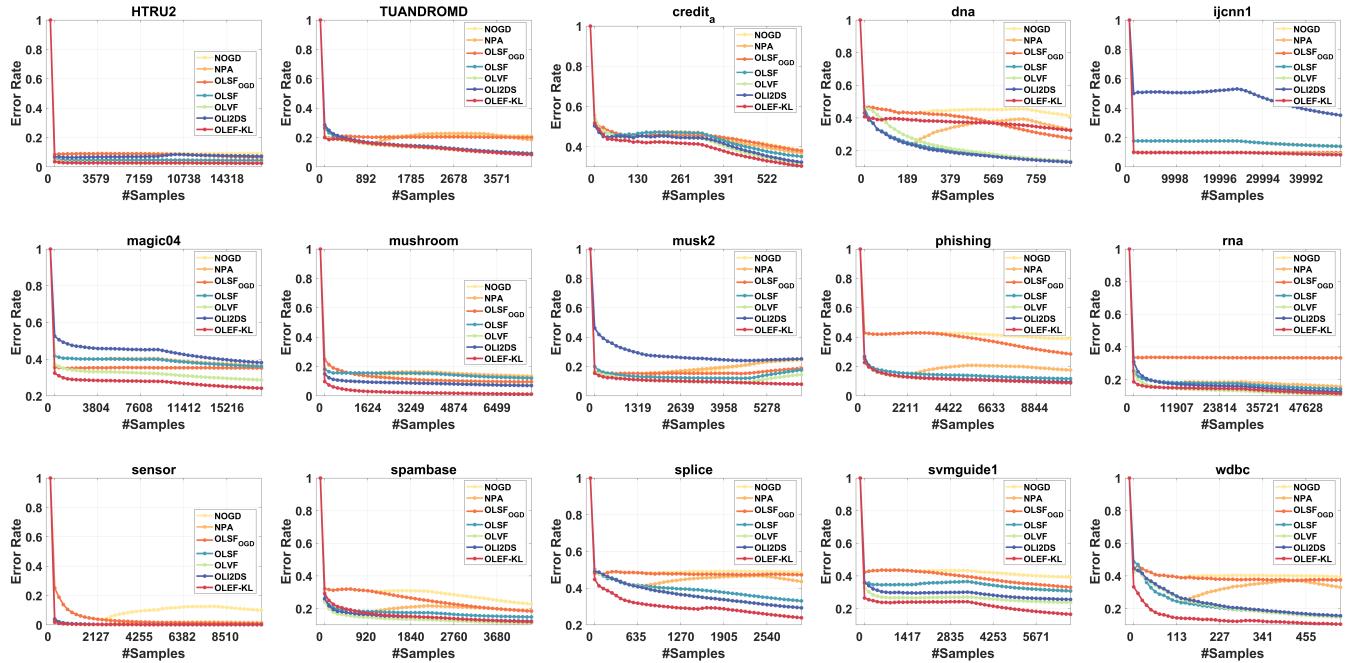


Fig. S6: The error rate trends of various methods across four datasets under 2-period general evolution scenarios.

capturing intricate patterns inherent in the data. However, it is crucial to note that when selecting the value of $\rho = 1$ may come at the cost of reduced computational efficiency. Hence, a selection of the decay parameter is important to keep a balance between optimizing performance and maintaining computational efficiency.

- Parameter η : The insights derived from Figure S9 underscore the dataset-specific nature of the learning rate η . It is evident that neither excessively large nor exceedingly small values of η can ensure a satisfactory model performance. Empirical evidence suggests that setting η to a value of 0.1 serves as a practical guideline to attain a satisfying performance level across datasets.
- Parameter σ : The findings presented in Figure S10 emphasize that the utilization of excessively large values for the parameter σ within the Gaussian kernel results in unacceptable performance. This outcome can be principally attributed to the enlargement of σ , which leads to a wide and flat Gaussian shape. Within this context, the kernel allocates nearly

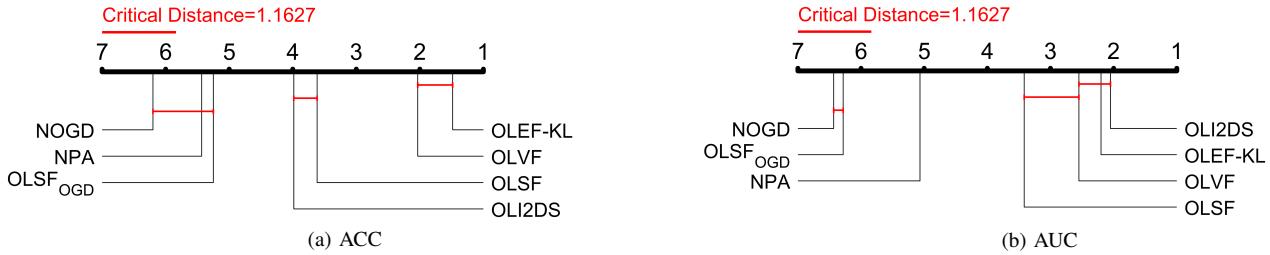


Fig. S7: OLEF-KL against competitors under general evolving scenarios with the Nemenyi test at significance level $\alpha = 0.05$.

uniform weights to all data points across the feature space. Consequently, the kernel smooths out variations in the data, making it challenging for the model to capture fine-grained patterns in the data.

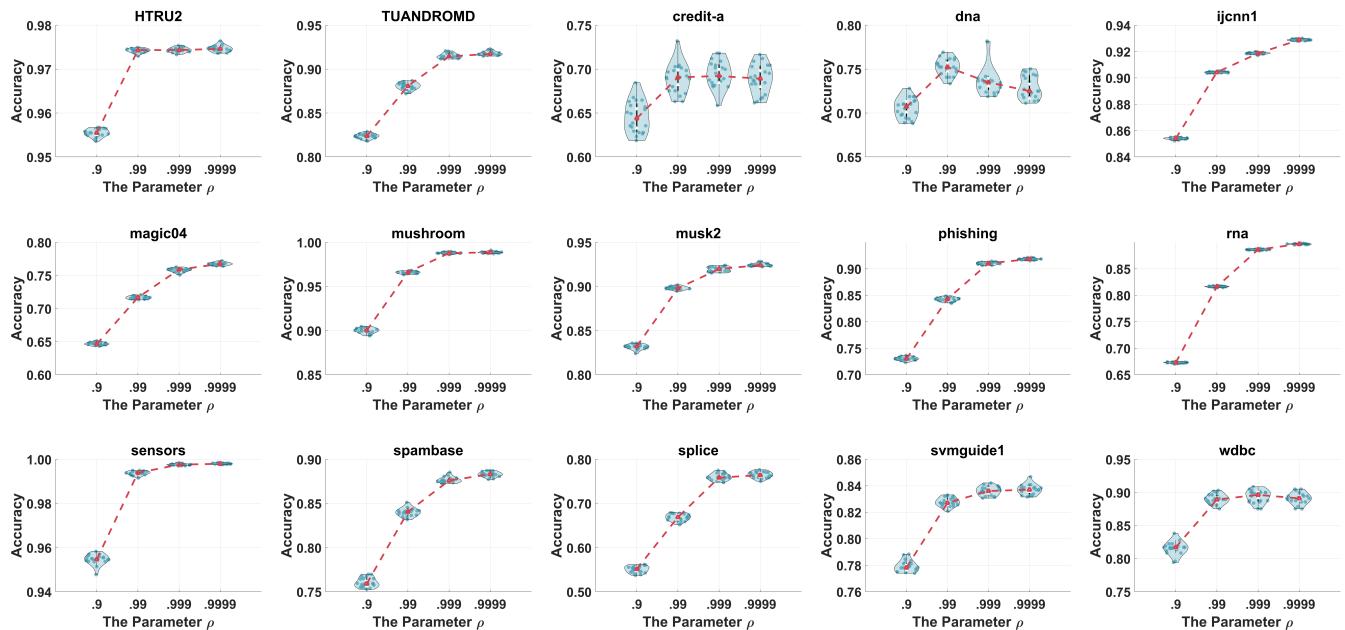


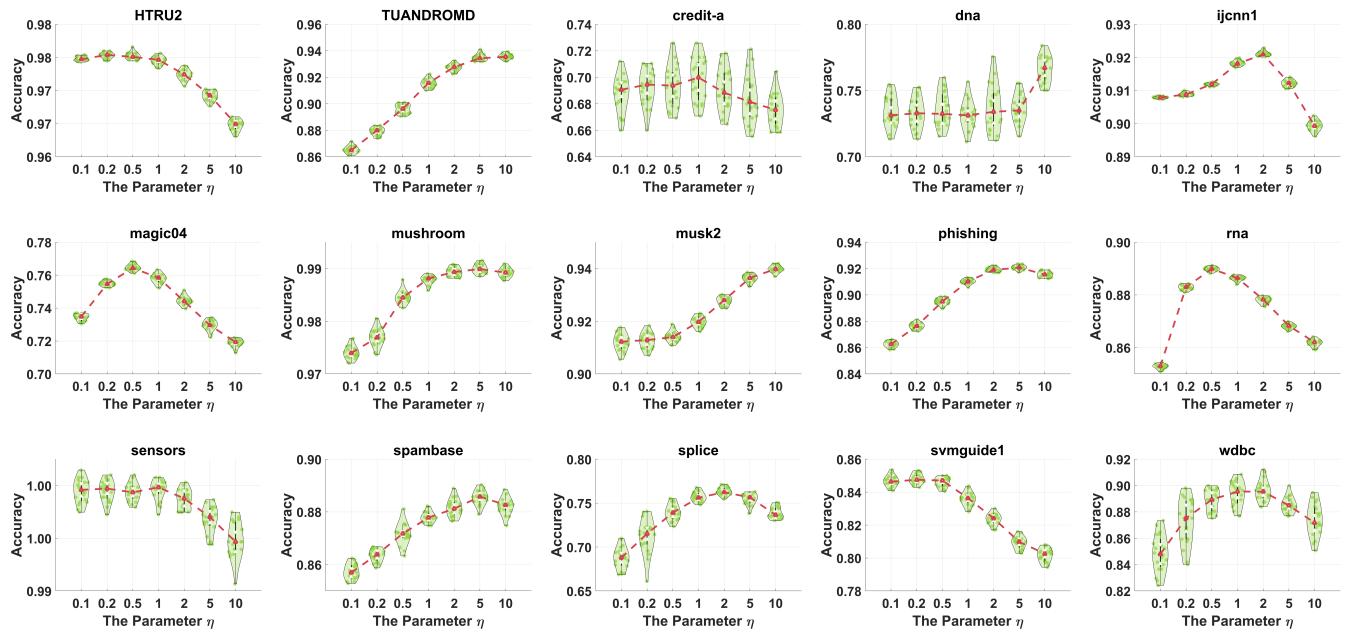
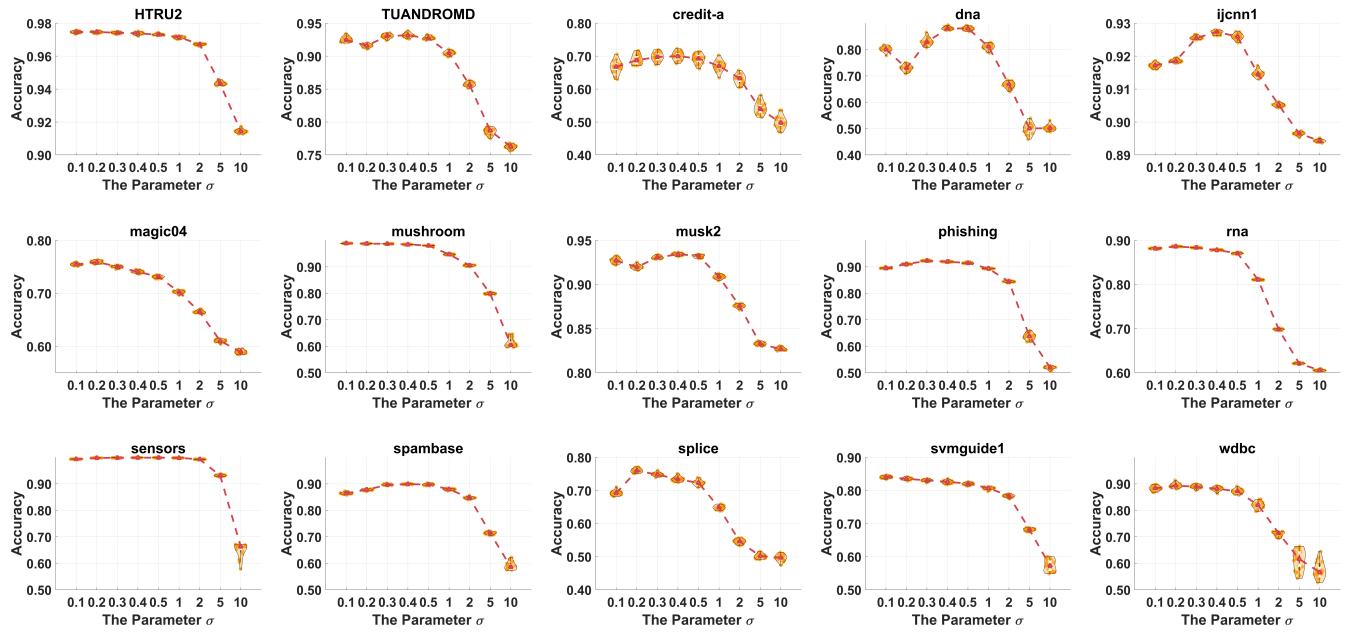
Fig. S8: The parameter ρ study under 2-period general evolution scenarios.

E. Kernel Comparison

This section conducts a comparative analysis of the performance of OLEF-KL by employing two distinct kernel functions: the linear kernel function and the Gaussian kernel function. The assessment of model performance is based on the metric of ACC, and the outcomes are presented in Figure S11.

As we can see, OLEF-KL with the Gaussian kernel function, outperforms its counterpart employing the linear kernel in most cases. The superiority of the Gaussian kernel can be attributed to its nonlinearity within the model, a feat that is beyond the purview of the linear kernel. Consequently, the Gaussian kernel excels in scenarios where the data exhibits complex patterns, thereby offering superior learning performance.

It is worth noting that in situations where the dataset is inherently linearly separable, the performance differential between the two kernel functions tends to diminish. For example, in the case of the *HTRU2* and *sensors* datasets, the linear kernel achieves results similar to those of the Gaussian kernel. This observation reinforces the idea that the choice of kernel function should be guided by the nature of the data.

Fig. S9: The parameter η study under 2-period general evolution scenarios.Fig. S10: The parameter σ study under 2-period general evolution scenarios.

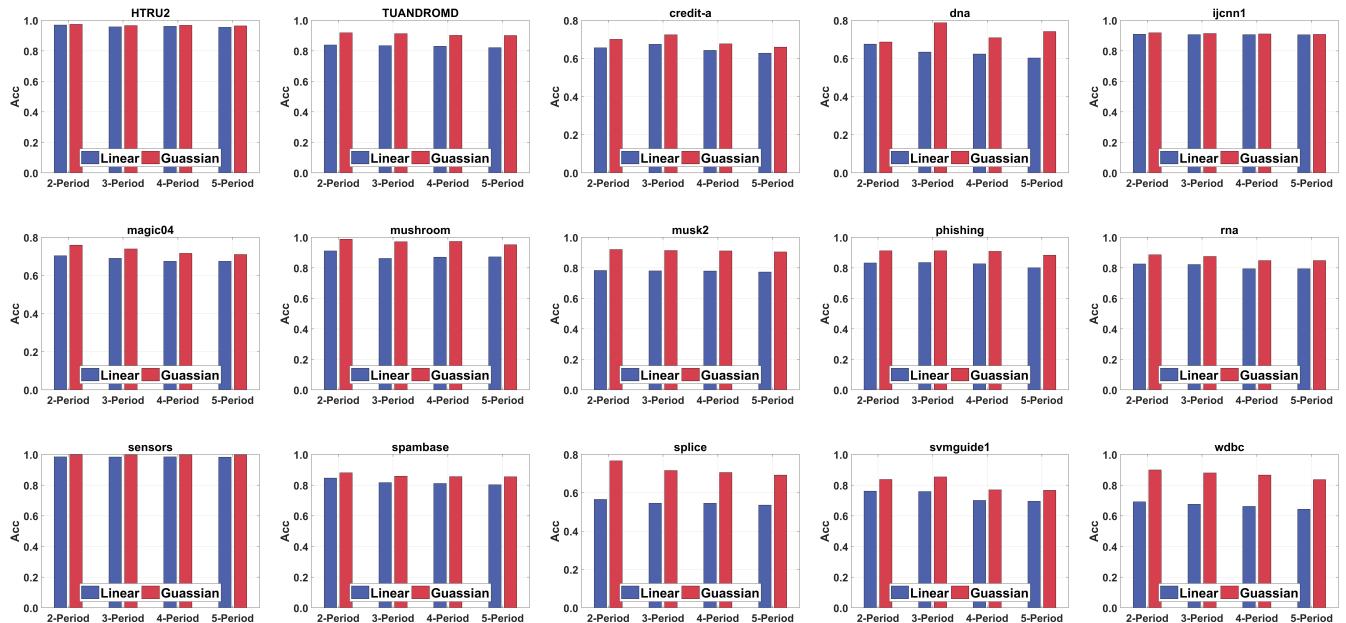


Fig. S11: The performance when OLEF-KL using different kernel functions.

S5. COMPARISON UNDER SYNCHRONIZED EVOLUTION

Within this subsection, our objective is to assess the performance of OLEF-KL under evolving scenarios, which is visually represented in Figure S2.

A. Comparison Results

The classification results for 15 synchronized evolving feature data streams are presented in Table S4, Table S5, in terms of ACC. Meanwhile, Table S5 and Figure S13 show the results based on AUC. Notably, our proposed OLEF-KL algorithm exhibits superior performance relative to other methods in most instances, particularly with regard to ACC. This enhanced performance can be attributed to the inherent flexibility of our non-parametric model.

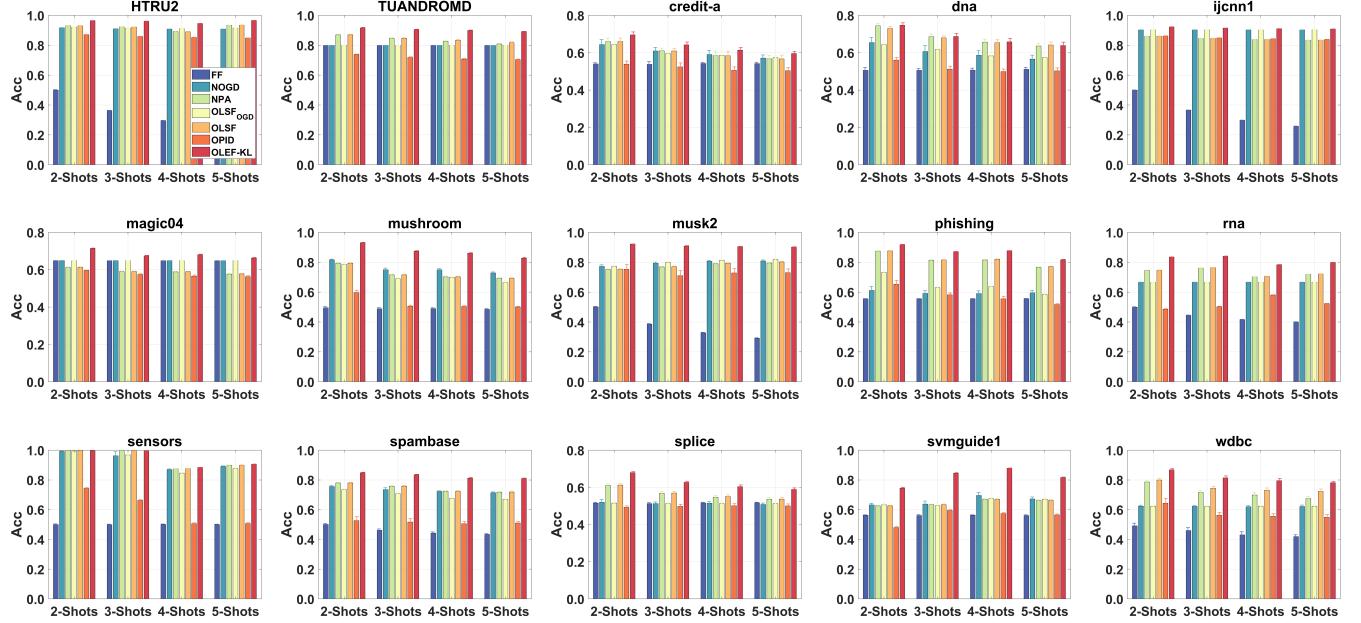


Fig. S12: The performance comparison under synchronized evolution scenarios in terms of ACC.

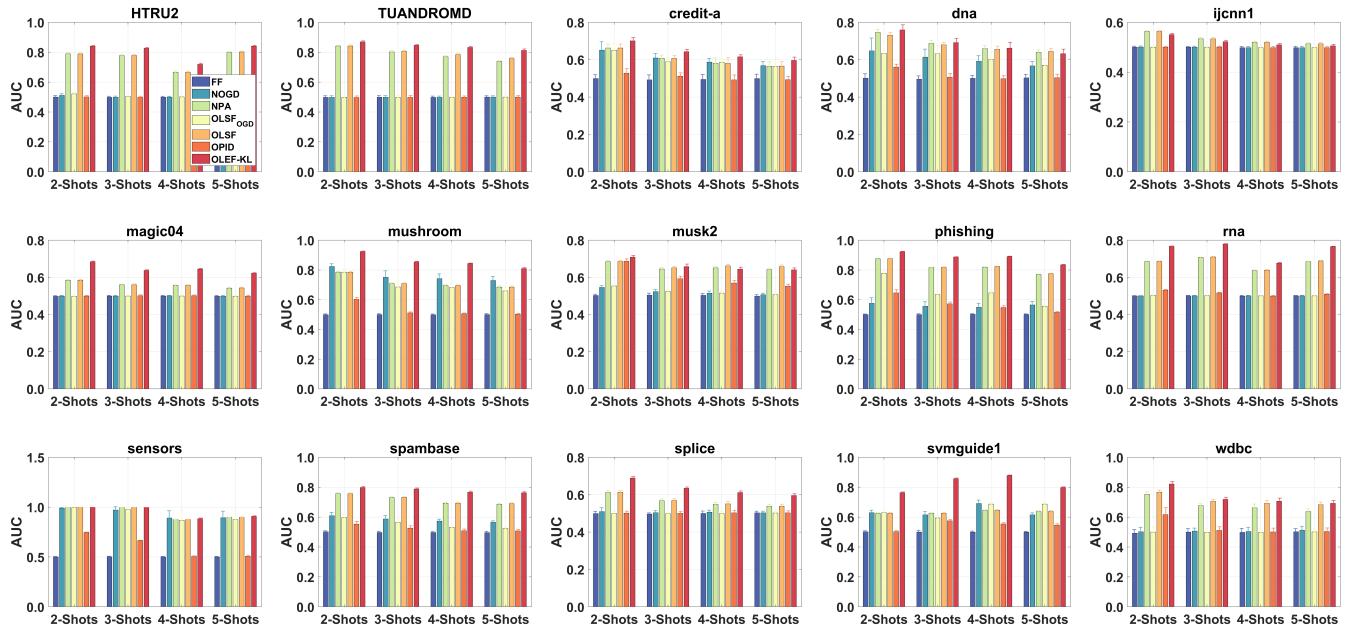


Fig. S13: The performance comparison under synchronized evolution scenarios in terms of AUC.

TABLE S4: The Performance comparison in synchronized evolution scenarios, in term of ACC (mean \pm std). •/○ indicate that OLEF-KL is significantly better/worse than competitors (hypothesis supported by paired t-tests at 0.05 significant level.). The best results are indicated in **bold**.

Datasets	Periods	FF	NOGD	NPA	OLSF _{OGD}	OLSF	OPID	OLEF-KL
HTRU2	2	0.500 \pm 0.002•	0.917 \pm 0.001•	0.931 \pm 0.002•	0.921 \pm 0.001•	0.930 \pm 0.002•	0.871 \pm 0.001•	0.964\pm0.001
	3	0.363 \pm 0.002•	0.910 \pm 0.001•	0.923 \pm 0.001•	0.914 \pm 0.001•	0.921 \pm 0.002•	0.858 \pm 0.001•	0.961\pm0.001
	4	0.296 \pm 0.002•	0.908 \pm 0.000•	0.891 \pm 0.003•	0.911 \pm 0.001•	0.889 \pm 0.002•	0.852 \pm 0.001•	0.945\pm0.001
	5	0.255 \pm 0.002•	0.908 \pm 0.000•	0.934 \pm 0.001•	0.915 \pm 0.002•	0.934 \pm 0.001•	0.848 \pm 0.001•	0.965\pm0.001
TUANDROMD	2	0.798 \pm 0.000•	0.798 \pm 0.000•	0.869 \pm 0.005•	0.799 \pm 0.001•	0.869 \pm 0.005•	0.738 \pm 0.003•	0.916\pm0.003
	3	0.798 \pm 0.000•	0.798 \pm 0.000•	0.845 \pm 0.006•	0.799 \pm 0.000•	0.846 \pm 0.006•	0.717 \pm 0.005•	0.905\pm0.002
	4	0.799 \pm 0.000•	0.798 \pm 0.000•	0.826 \pm 0.005•	0.799 \pm 0.000•	0.834 \pm 0.004•	0.708 \pm 0.004•	0.898\pm0.003
	5	0.799 \pm 0.000•	0.798 \pm 0.000•	0.808 \pm 0.005•	0.799 \pm 0.000•	0.817 \pm 0.005•	0.703 \pm 0.004•	0.891\pm0.002
credit-a	2	0.540 \pm 0.006•	0.644 \pm 0.026•	0.660 \pm 0.019•	0.644 \pm 0.018•	0.660 \pm 0.019•	0.538 \pm 0.017•	0.695\pm0.016
	3	0.538 \pm 0.015•	0.609 \pm 0.019•	0.610 \pm 0.014•	0.594 \pm 0.014•	0.609 \pm 0.013•	0.524 \pm 0.022•	0.642\pm0.015
	4	0.542 \pm 0.006•	0.590 \pm 0.022•	0.584 \pm 0.025•	0.587 \pm 0.023•	0.583 \pm 0.025•	0.506 \pm 0.019•	0.614\pm0.014
	5	0.542 \pm 0.008•	0.571 \pm 0.017•	0.567 \pm 0.018•	0.573 \pm 0.012•	0.567 \pm 0.020•	0.503 \pm 0.017•	0.596\pm0.012
dna	2	0.506 \pm 0.015•	0.654 \pm 0.028•	0.744\pm0.012	0.643 \pm 0.022•	0.730 \pm 0.009•	0.561 \pm 0.013•	0.747\pm0.013
	3	0.506 \pm 0.009•	0.606 \pm 0.032•	0.686\pm0.015	0.617 \pm 0.020•	0.679\pm0.011	0.512 \pm 0.016•	0.687\pm0.017
	4	0.507 \pm 0.010•	0.587 \pm 0.026•	0.656\pm0.016	0.583 \pm 0.020•	0.654\pm0.016	0.499 \pm 0.013•	0.658\pm0.019
	5	0.511 \pm 0.011•	0.566 \pm 0.021•	0.636\pm0.013	0.574 \pm 0.017•	0.641\pm0.016	0.503 \pm 0.016•	0.638\pm0.018
ijcnn1	2	0.500 \pm 0.001•	0.903 \pm 0.000•	0.860 \pm 0.001•	0.903 \pm 0.000•	0.860 \pm 0.001•	0.862 \pm 0.002•	0.922\pm0.001
	3	0.366 \pm 0.001•	0.903 \pm 0.000•	0.846 \pm 0.001•	0.903 \pm 0.000•	0.846 \pm 0.001•	0.849 \pm 0.002•	0.915\pm0.001
	4	0.298 \pm 0.001•	0.903 \pm 0.000•	0.838 \pm 0.001•	0.903 \pm 0.000•	0.838 \pm 0.001•	0.843 \pm 0.002•	0.911\pm0.001
	5	0.258 \pm 0.001•	0.903 \pm 0.000•	0.834 \pm 0.001•	0.903 \pm 0.000•	0.834 \pm 0.001•	0.838 \pm 0.002•	0.908\pm0.001
magic04	2	0.648 \pm 0.000•	0.649 \pm 0.000•	0.612 \pm 0.003•	0.648 \pm 0.000•	0.612 \pm 0.003•	0.595 \pm 0.003•	0.713\pm0.003
	3	0.648 \pm 0.000•	0.648 \pm 0.000•	0.590 \pm 0.003•	0.648 \pm 0.000•	0.590 \pm 0.003•	0.574 \pm 0.004•	0.674\pm0.003
	4	0.648 \pm 0.000•	0.648 \pm 0.000•	0.588 \pm 0.003•	0.648 \pm 0.000•	0.588 \pm 0.003•	0.566 \pm 0.005•	0.680\pm0.003
	5	0.648 \pm 0.000•	0.648 \pm 0.000•	0.577 \pm 0.003•	0.648 \pm 0.000•	0.577 \pm 0.003•	0.563 \pm 0.005•	0.663\pm0.003
mushroom	2	0.496 \pm 0.007•	0.816 \pm 0.005•	0.793 \pm 0.004•	0.783 \pm 0.004•	0.793 \pm 0.004•	0.597 \pm 0.015•	0.930\pm0.002
	3	0.489 \pm 0.008•	0.750 \pm 0.011•	0.715 \pm 0.006•	0.689 \pm 0.009•	0.715 \pm 0.006•	0.506 \pm 0.006•	0.874\pm0.003
	4	0.490 \pm 0.006•	0.750 \pm 0.007•	0.702 \pm 0.005•	0.700 \pm 0.007•	0.702 \pm 0.005•	0.504 \pm 0.005•	0.861\pm0.003
	5	0.485 \pm 0.005•	0.729 \pm 0.007•	0.692 \pm 0.006•	0.663 \pm 0.013•	0.692 \pm 0.006•	0.501 \pm 0.004•	0.828\pm0.004
musk2	2	0.501 \pm 0.004•	0.772 \pm 0.010•	0.751 \pm 0.004•	0.773 \pm 0.008•	0.752 \pm 0.004•	0.754 \pm 0.030•	0.921\pm0.001
	3	0.386 \pm 0.004•	0.793 \pm 0.008•	0.768 \pm 0.005•	0.799 \pm 0.005•	0.771 \pm 0.003•	0.710 \pm 0.035•	0.908\pm0.002
	4	0.328 \pm 0.003•	0.806 \pm 0.007•	0.789 \pm 0.003•	0.813 \pm 0.004•	0.793 \pm 0.004•	0.728 \pm 0.029•	0.905\pm0.001
	5	0.292 \pm 0.003•	0.808 \pm 0.007•	0.795 \pm 0.004•	0.818 \pm 0.003•	0.802 \pm 0.005•	0.730 \pm 0.025•	0.902\pm0.002
phishing	2	0.555 \pm 0.002•	0.612 \pm 0.026•	0.874 \pm 0.003•	0.731 \pm 0.011•	0.876 \pm 0.003•	0.653 \pm 0.025•	0.918\pm0.002
	3	0.556 \pm 0.001•	0.591 \pm 0.018•	0.813 \pm 0.003•	0.629 \pm 0.006•	0.815 \pm 0.003•	0.581 \pm 0.013•	0.871\pm0.002
	4	0.556 \pm 0.002•	0.591 \pm 0.018•	0.815 \pm 0.003•	0.637 \pm 0.011•	0.819 \pm 0.004•	0.555 \pm 0.017•	0.876\pm0.002
	5	0.556 \pm 0.002•	0.596 \pm 0.015•	0.765 \pm 0.003•	0.584 \pm 0.005•	0.770 \pm 0.003•	0.517 \pm 0.006•	0.816\pm0.003
rna	2	0.500 \pm 0.002•	0.667 \pm 0.000•	0.745 \pm 0.002•	0.665 \pm 0.000•	0.746 \pm 0.002•	0.486 \pm 0.004•	0.835\pm0.001
	3	0.444 \pm 0.002•	0.667 \pm 0.000•	0.761 \pm 0.002•	0.667 \pm 0.000•	0.764 \pm 0.002•	0.501 \pm 0.003•	0.840\pm0.001
	4	0.416 \pm 0.002•	0.667 \pm 0.000•	0.702 \pm 0.002•	0.667 \pm 0.000•	0.704 \pm 0.002•	0.580 \pm 0.003•	0.782\pm0.002
	5	0.400 \pm 0.002•	0.666 \pm 0.000•	0.719 \pm 0.002•	0.667 \pm 0.000•	0.722 \pm 0.002•	0.523 \pm 0.003•	0.798\pm0.002
sensors	2	0.500 \pm 0.004•	0.991 \pm 0.005•	0.998 \pm 0.001•	0.995 \pm 0.004•	0.998\pm0.001 ○	0.744 \pm 0.005•	0.997 \pm 0.000
	3	0.499 \pm 0.004•	0.962 \pm 0.027•	0.996 \pm 0.001○	0.968 \pm 0.029•	0.997\pm0.001 ○	0.663 \pm 0.004•	0.995 \pm 0.001
	4	0.501 \pm 0.003•	0.869 \pm 0.005•	0.875 \pm 0.002•	0.846 \pm 0.022•	0.876 \pm 0.002•	0.505 \pm 0.006•	0.883\pm0.002
	5	0.500 \pm 0.003•	0.891 \pm 0.004•	0.898 \pm 0.002•	0.876 \pm 0.017•	0.899 \pm 0.002•	0.506 \pm 0.006•	0.906\pm0.002
spambase	2	0.501 \pm 0.007•	0.755 \pm 0.007•	0.780 \pm 0.005•	0.735 \pm 0.008•	0.780 \pm 0.005•	0.527 \pm 0.026•	0.849\pm0.003
	3	0.462 \pm 0.008•	0.735 \pm 0.012•	0.758 \pm 0.005•	0.707 \pm 0.006•	0.759 \pm 0.005•	0.517 \pm 0.023•	0.835\pm0.003
	4	0.443 \pm 0.008•	0.723 \pm 0.004•	0.724 \pm 0.006•	0.675 \pm 0.005•	0.724 \pm 0.006•	0.506 \pm 0.015•	0.812\pm0.004
	5	0.434 \pm 0.006•	0.714 \pm 0.006•	0.717 \pm 0.007•	0.669 \pm 0.003•	0.719 \pm 0.007•	0.510 \pm 0.011•	0.809\pm0.004
splice	2	0.515 \pm 0.005•	0.520 \pm 0.014•	0.611 \pm 0.008•	0.516 \pm 0.006•	0.612 \pm 0.007•	0.493 \pm 0.008•	0.679\pm0.007
	3	0.513 \pm 0.006•	0.512 \pm 0.010•	0.568 \pm 0.009•	0.513 \pm 0.008•	0.569 \pm 0.009•	0.498 \pm 0.011•	0.627\pm0.006
	4	0.517 \pm 0.004•	0.515 \pm 0.008•	0.548 \pm 0.008•	0.513 \pm 0.010•	0.551 \pm 0.009•	0.501 \pm 0.011•	0.604\pm0.009
	5	0.517 \pm 0.003•	0.508 \pm 0.008•	0.536 \pm 0.010•	0.513 \pm 0.008•	0.537 \pm 0.009•	0.501 \pm 0.010•	0.589\pm0.008
svmguide1	2	0.563 \pm 0.003•	0.632 \pm 0.010•	0.625 \pm 0.005•	0.632 \pm 0.010•	0.625 \pm 0.005•	0.480 \pm 0.008•	0.746\pm0.005
	3	0.561 \pm 0.006•	0.638 \pm 0.020•	0.633 \pm 0.005•	0.626 \pm 0.020•	0.633 \pm 0.005•	0.596 \pm 0.005•	0.846\pm0.003
	4	0.562 \pm 0.003•	0.696 \pm 0.020•	0.669 \pm 0.004•	0.674 \pm 0.010•	0.670 \pm 0.004•	0.574 \pm 0.006•	0.879\pm0.002
	5	0.561 \pm 0.005•	0.672 \pm 0.013•	0.661 \pm 0.005•	0.671 \pm 0.007•	0.662 \pm 0.005•	0.564 \pm 0.007•	0.816\pm0.004
wdbc	2	0.492 \pm 0.018•	0.624 \pm 0.006•	0.786 \pm 0.013•	0.624 \pm 0.004•	0.800 \pm 0.012•	0.643 \pm 0.032•	0.868\pm0.008
	3	0.460 \pm 0.020•	0.623 \pm 0.005•	0.716 \pm 0.012•	0.623 \pm 0.004•	0.745 \pm 0.012•	0.562 \pm 0.020•	0.814\pm0.013
	4	0.431 \pm 0.021•	0.620 \pm 0.007•	0.699 \pm 0.017•	0.624 \pm 0.004•	0.7		

TABLE S5: The Performance comparison in synchronized evolution scenarios, in term of AUC (mean \pm std). \bullet/\odot indicate that OLEF-KL is significantly better/worse than competitors (hypothesis supported by paired t-tests at 0.05 significant level.). The best results are indicated in **bold**.

Datasets	Periods	FF	NOGD	NPA	$OLSF_{OGD}$	OLSF	OPID	OLEF-KL
HTRU2	2	0.500 \pm 0.009 \bullet	0.512 \pm 0.010 \bullet	0.789 \pm 0.007 \bullet	0.521 \pm 0.011 \bullet	0.788 \pm 0.007 \bullet	0.501 \pm 0.009 \bullet	0.840\pm0.005
	3	0.498 \pm 0.008 \bullet	0.499 \pm 0.008 \bullet	0.777 \pm 0.006 \bullet	0.503 \pm 0.009 \bullet	0.778 \pm 0.006 \bullet	0.498 \pm 0.008 \bullet	0.827\pm0.005
	4	0.499 \pm 0.006 \bullet	0.500 \pm 0.006 \bullet	0.666 \pm 0.008 \bullet	0.501 \pm 0.006 \bullet	0.666 \pm 0.008 \bullet	0.500 \pm 0.007 \bullet	0.719\pm0.008
	5	0.500 \pm 0.006 \bullet	0.500 \pm 0.006 \bullet	0.798 \pm 0.005 \bullet	0.516 \pm 0.008 \bullet	0.802 \pm 0.005 \bullet	0.499 \pm 0.007 \bullet	0.841\pm0.005
TUANDROMD	2	0.498 \pm 0.011 \bullet	0.498 \pm 0.011 \bullet	0.841 \pm 0.006 \bullet	0.499 \pm 0.011 \bullet	0.841 \pm 0.007 \bullet	0.497 \pm 0.011 \bullet	0.870\pm0.006
	3	0.499 \pm 0.011 \bullet	0.499 \pm 0.011 \bullet	0.802 \pm 0.007 \bullet	0.499 \pm 0.011 \bullet	0.805 \pm 0.007 \bullet	0.498 \pm 0.013 \bullet	0.846\pm0.006
	4	0.498 \pm 0.009 \bullet	0.498 \pm 0.009 \bullet	0.769 \pm 0.007 \bullet	0.498 \pm 0.009 \bullet	0.784 \pm 0.006 \bullet	0.499 \pm 0.009 \bullet	0.832\pm0.007
	5	0.500 \pm 0.009 \bullet	0.500 \pm 0.009 \bullet	0.739 \pm 0.006 \bullet	0.500 \pm 0.009 \bullet	0.758 \pm 0.007 \bullet	0.500 \pm 0.010 \bullet	0.812\pm0.010
credit-a	2	0.499 \pm 0.020 \bullet	0.651 \pm 0.045 \bullet	0.663 \pm 0.021 \bullet	0.650 \pm 0.036 \bullet	0.662 \pm 0.022 \bullet	0.528 \pm 0.024 \bullet	0.701\pm0.018
	3	0.492 \pm 0.027 \bullet	0.609 \pm 0.025 \bullet	0.608 \pm 0.017 \bullet	0.589 \pm 0.025 \bullet	0.607 \pm 0.015 \bullet	0.512 \pm 0.018 \bullet	0.643\pm0.013
	4	0.495 \pm 0.027 \bullet	0.587 \pm 0.018 \bullet	0.581 \pm 0.029 \bullet	0.587 \pm 0.022 \bullet	0.580 \pm 0.030 \bullet	0.492 \pm 0.026 \bullet	0.615\pm0.011
	5	0.499 \pm 0.024 \bullet	0.568 \pm 0.023 \bullet	0.566 \pm 0.023 \bullet	0.565 \pm 0.020 \bullet	0.566 \pm 0.024 \bullet	0.493 \pm 0.017 \bullet	0.596\pm0.018
dna	2	0.501 \pm 0.024 \bullet	0.648 \pm 0.068 \bullet	0.746\pm0.017	0.634 \pm 0.061 \bullet	0.731 \pm 0.015 \bullet	0.561 \pm 0.015 \bullet	0.759\pm0.028
	3	0.496 \pm 0.017 \bullet	0.614 \pm 0.043 \bullet	0.687\pm0.017	0.631 \pm 0.045 \bullet	0.679 \pm 0.013 \bullet	0.507 \pm 0.020 \bullet	0.692\pm0.022
	4	0.501 \pm 0.016 \bullet	0.592 \pm 0.029 \bullet	0.659\pm0.017	0.602 \pm 0.040 \bullet	0.655\pm0.017	0.498 \pm 0.016 \bullet	0.662\pm0.031
	5	0.503 \pm 0.019 \bullet	0.567 \pm 0.024 \bullet	0.640 \pm 0.013 \bullet	0.571 \pm 0.037 \bullet	0.644\pm0.016\odot	0.504 \pm 0.018 \bullet	0.632 \pm 0.024
ijenn1	2	0.500 \pm 0.004 \bullet	0.501 \pm 0.004 \bullet	0.564\pm0.004\odot	0.501 \pm 0.004 \bullet	0.564\pm0.004\odot	0.500 \pm 0.003 \bullet	0.551 \pm 0.004
	3	0.500 \pm 0.004 \bullet	0.501 \pm 0.004 \bullet	0.534\pm0.005\odot	0.500 \pm 0.004 \bullet	0.534\pm0.005\odot	0.501 \pm 0.004 \bullet	0.522 \pm 0.004
	4	0.498 \pm 0.005 \bullet	0.498 \pm 0.005 \bullet	0.520\pm0.005\odot	0.498 \pm 0.005 \bullet	0.520\pm0.005\odot	0.498 \pm 0.005 \bullet	0.510 \pm 0.005
	5	0.498 \pm 0.006 \bullet	0.498 \pm 0.006 \bullet	0.513\pm0.005\odot	0.498 \pm 0.006 \bullet	0.513\pm0.005\odot	0.498 \pm 0.005 \bullet	0.505 \pm 0.006
magic04	2	0.498 \pm 0.004 \bullet	0.499 \pm 0.004 \bullet	0.585 \pm 0.004 \bullet	0.498 \pm 0.004 \bullet	0.585 \pm 0.004 \bullet	0.500 \pm 0.003 \bullet	0.684\pm0.003
	3	0.499 \pm 0.004 \bullet	0.500 \pm 0.004 \bullet	0.560 \pm 0.004 \bullet	0.499 \pm 0.004 \bullet	0.560 \pm 0.004 \bullet	0.501 \pm 0.005 \bullet	0.637\pm0.004
	4	0.500 \pm 0.003 \bullet	0.500 \pm 0.003 \bullet	0.556 \pm 0.003 \bullet	0.500 \pm 0.003 \bullet	0.557 \pm 0.003 \bullet	0.501 \pm 0.005 \bullet	0.645\pm0.003
	5	0.499 \pm 0.004 \bullet	0.499 \pm 0.004 \bullet	0.543 \pm 0.004 \bullet	0.499 \pm 0.004 \bullet	0.543 \pm 0.004 \bullet	0.499 \pm 0.005 \bullet	0.622\pm0.004
mushroom	2	0.499 \pm 0.006 \bullet	0.823 \pm 0.019 \bullet	0.785 \pm 0.004 \bullet	0.786 \pm 0.016 \bullet	0.785 \pm 0.004 \bullet	0.602 \pm 0.013 \bullet	0.924\pm0.003
	3	0.500 \pm 0.006 \bullet	0.751 \pm 0.043 \bullet	0.708 \pm 0.006 \bullet	0.686 \pm 0.040 \bullet	0.707 \pm 0.006 \bullet	0.511 \pm 0.007 \bullet	0.854\pm0.005
	4	0.498 \pm 0.006 \bullet	0.742 \pm 0.031 \bullet	0.695 \pm 0.005 \bullet	0.683 \pm 0.032 \bullet	0.695 \pm 0.005 \bullet	0.506 \pm 0.005 \bullet	0.843\pm0.004
	5	0.500 \pm 0.007 \bullet	0.729 \pm 0.027 \bullet	0.684 \pm 0.007 \bullet	0.659 \pm 0.026 \bullet	0.685 \pm 0.007 \bullet	0.503 \pm 0.006 \bullet	0.809\pm0.006
musk2	2	0.503 \pm 0.007 \bullet	0.546 \pm 0.009 \bullet	0.685 \pm 0.007 \bullet	0.554 \pm 0.009 \bullet	0.686 \pm 0.006 \bullet	0.687 \pm 0.012 \bullet	0.709\pm0.009
	3	0.504 \pm 0.011 \bullet	0.523 \pm 0.011 \bullet	0.645 \pm 0.008 \bullet	0.524 \pm 0.011 \bullet	0.651 \pm 0.008 \bullet	0.593 \pm 0.013 \bullet	0.658\pm0.014
	4	0.503 \pm 0.010 \bullet	0.515 \pm 0.011 \bullet	0.651 \pm 0.006 \odot	0.515 \pm 0.011 \bullet	0.662\pm0.007\odot	0.570 \pm 0.013 \bullet	0.645 \pm 0.010
	5	0.499 \pm 0.008 \bullet	0.506 \pm 0.008 \bullet	0.642 \pm 0.008 \bullet	0.507 \pm 0.008 \bullet	0.658\pm0.008\odot	0.553 \pm 0.011 \bullet	0.641 \pm 0.010
phishing	2	0.500 \pm 0.004 \bullet	0.576 \pm 0.037 \bullet	0.875 \pm 0.003 \bullet	0.777 \pm 0.007 \bullet	0.876 \pm 0.002 \bullet	0.645 \pm 0.022 \bullet	0.923\pm0.002
	3	0.500 \pm 0.006 \bullet	0.556 \pm 0.030 \bullet	0.816 \pm 0.003 \bullet	0.636 \pm 0.010 \bullet	0.818 \pm 0.002 \bullet	0.572 \pm 0.012 \bullet	0.887\pm0.002
	4	0.502 \pm 0.005 \bullet	0.549 \pm 0.027 \bullet	0.818 \pm 0.004 \bullet	0.646 \pm 0.013 \bullet	0.822 \pm 0.004 \bullet	0.547 \pm 0.012 \bullet	0.890\pm0.003
	5	0.501 \pm 0.004 \bullet	0.564 \pm 0.024 \bullet	0.770 \pm 0.003 \bullet	0.557 \pm 0.006 \bullet	0.774 \pm 0.003 \bullet	0.515 \pm 0.005 \bullet	0.834\pm0.004
rna	2	0.500 \pm 0.002 \bullet	0.500 \pm 0.002 \bullet	0.685 \pm 0.003 \bullet	0.503 \pm 0.002 \bullet	0.686 \pm 0.003 \bullet	0.532 \pm 0.003 \bullet	0.768\pm0.002
	3	0.501 \pm 0.003 \bullet	0.501 \pm 0.003 \bullet	0.707 \pm 0.002 \bullet	0.501 \pm 0.003 \bullet	0.710 \pm 0.002 \bullet	0.517 \pm 0.003 \bullet	0.780\pm0.002
	4	0.500 \pm 0.004 \bullet	0.500 \pm 0.004 \bullet	0.637 \pm 0.003 \bullet	0.500 \pm 0.004 \bullet	0.639 \pm 0.004 \bullet	0.500 \pm 0.004 \bullet	0.677\pm0.003
	5	0.500 \pm 0.003 \bullet	0.500 \pm 0.003 \bullet	0.685 \pm 0.002 \bullet	0.500 \pm 0.003 \bullet	0.688 \pm 0.002 \bullet	0.510 \pm 0.003 \bullet	0.765\pm0.003
sensors	2	0.501 \pm 0.005 \bullet	0.990 \pm 0.007 \bullet	0.997 \pm 0.001 \bullet	0.994 \pm 0.007 \bullet	0.998\pm0.001\odot	0.745 \pm 0.005 \bullet	0.997 \pm 0.001
	3	0.502 \pm 0.006 \bullet	0.971 \pm 0.033 \bullet	0.996 \pm 0.001 \bullet	0.977 \pm 0.034 \bullet	0.997\pm0.001\odot	0.664 \pm 0.005 \bullet	0.996 \pm 0.001
	4	0.500 \pm 0.004 \bullet	0.891\pm0.073	0.875 \pm 0.002 \bullet	0.869\pm0.096	0.876 \pm 0.002 \bullet	0.505 \pm 0.006 \bullet	0.884\pm0.007
	5	0.501 \pm 0.006 \bullet	0.894\pm0.066	0.898 \pm 0.003 \bullet	0.876\pm0.088	0.899 \pm 0.003 \bullet	0.507 \pm 0.007 \bullet	0.907\pm0.006
spambase	2	0.502 \pm 0.007 \bullet	0.610 \pm 0.021 \bullet	0.757 \pm 0.006 \bullet	0.596 \pm 0.021 \bullet	0.756 \pm 0.006 \bullet	0.553 \pm 0.018 \bullet	0.798\pm0.006
	3	0.497 \pm 0.008 \bullet	0.588 \pm 0.022 \bullet	0.731 \pm 0.005 \bullet	0.565 \pm 0.015 \bullet	0.732 \pm 0.005 \bullet	0.527 \pm 0.015 \bullet	0.788\pm0.007
	4	0.497 \pm 0.006 \bullet	0.573 \pm 0.013 \bullet	0.694 \pm 0.007 \bullet	0.533 \pm 0.011 \bullet	0.694 \pm 0.006 \bullet	0.509 \pm 0.011 \bullet	0.766\pm0.008
	5	0.496 \pm 0.009 \bullet	0.565 \pm 0.010 \bullet	0.687 \pm 0.006 \bullet	0.525 \pm 0.008 \bullet	0.689 \pm 0.006 \bullet	0.508 \pm 0.009 \bullet	0.763\pm0.006
splice	2	0.499 \pm 0.012 \bullet	0.509 \pm 0.021 \bullet	0.612 \pm 0.010 \bullet	0.500 \pm 0.012 \bullet	0.613 \pm 0.009 \bullet	0.501 \pm 0.011 \bullet	0.689\pm0.008
	3	0.497 \pm 0.006 \bullet	0.503<math					

Specifically, the error curve of FF demonstrates an unfortunate increase, suggesting its inability to effectively learn from evolving feature conditions. This observation underscores the inadequacy of employing the original fixed feature space in the presence of feature evolution. Conversely, the red curves, which correspond to the OLEF-KL algorithm, exhibit a more obvious reduction in error rates. This trend indicates the capacity of OLEF-KL to promptly adapt to feature evolution.

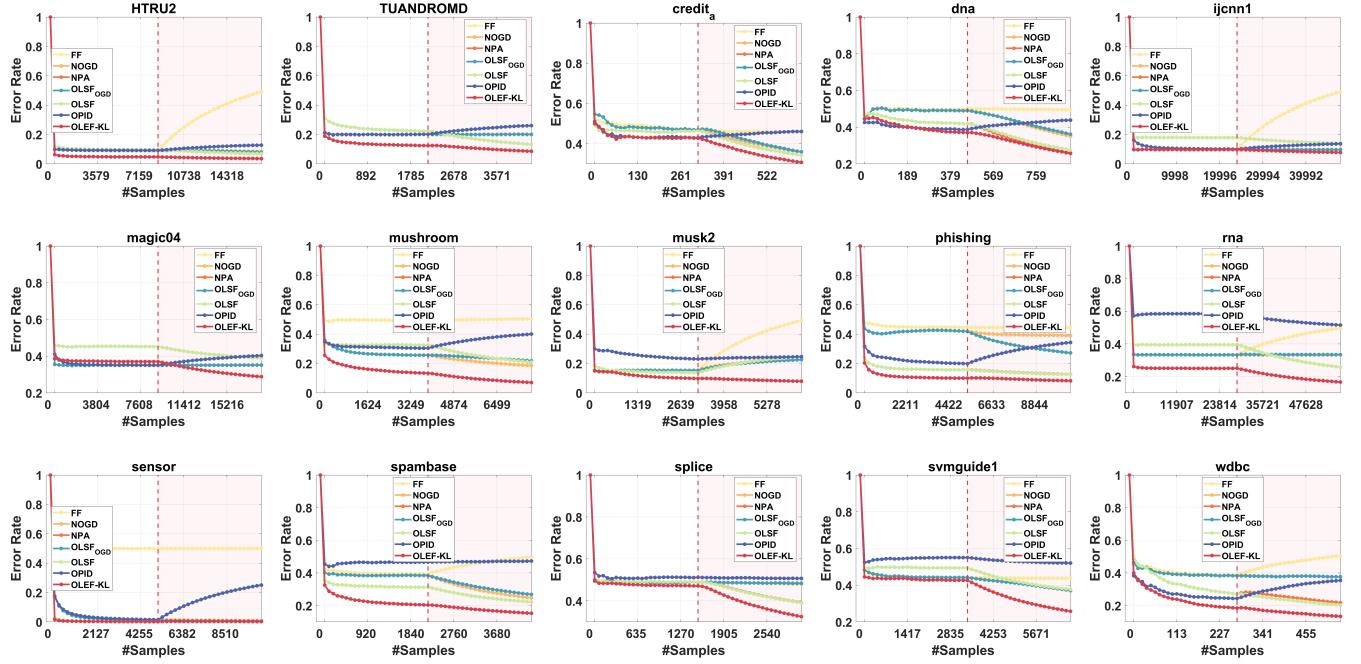


Fig. S14: The error rate trends of various methods under 2-period synchronized evolving scenarios, with a vertical red dashed line indicating the timing of feature evolution.

C. Nemenyi Test

The Nemenyi tests under synchronized evolving scenarios are provided in Figure S15. The critical distance is calculated as 1.1627 in this situation ($q_{0.05} = 2.984, K = 7, N = 60$). It is evident from Figure S15 that OLEF-KL exhibits a substantial advantage over all competing methods, across both the ACC and AUC metrics. These observations underscore the effectiveness of OLEF-KL on learning synchronized evolving feature data streams.

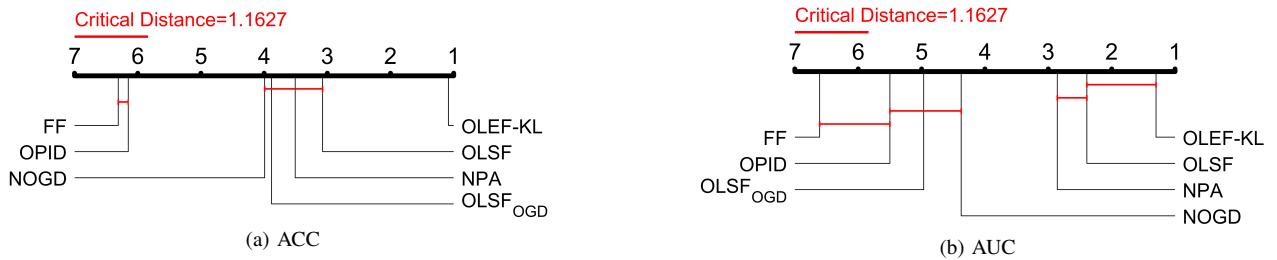


Fig. S15: OLEF-KL against competitors under synchronized evolving scenario with the Nemenyi test at significance level $\alpha = 0.05$.

D. Parameter Study

We proceed to examine the influence of key parameters within the OLEF-KL algorithm under synchronized evolution scenarios. Similar to the previous study, We chose values for the parameters as follows: ρ from the set $\{0.9, 0.99, 0.999, 0.9999\}$, η from the set $\{0.1, 0.2, 0.5, 1, 2, 5, 10\}$, and σ from the set $\{0.1, 0.2, 0.3, 0.4, 0.5, 1, 2, 5, 10\}$. During the analysis of each parameter, we maintain the other parameters at constant values, specifically, $\rho = 0.999$, $\eta = 1$, and $\sigma = 0.2$. We employ accuracy to measure the performance and the outcomes are depicted from Figure S16 and Figure S18. From these outcomes, we can see the key observations are similar to those from the parameter study under general evolution scenarios.

- Parameter ρ : The observations from Figure S16 reveal a positive correlation between the performance of OLEF-KL and the magnitude of the decay parameter ρ , but the value of $\rho = 1$ may come at the cost of reduced computational efficiency.
- Parameter η : Empirical evidence from Figure S17 suggests that setting η to a value of 0.1 serves as a practical guideline to attain a satisfying performance level across datasets.
- Parameter σ : The findings presented in Figure S18 emphasize that the utilization of excessively large values for the parameter σ within the Gaussian kernel results in unacceptable performance.

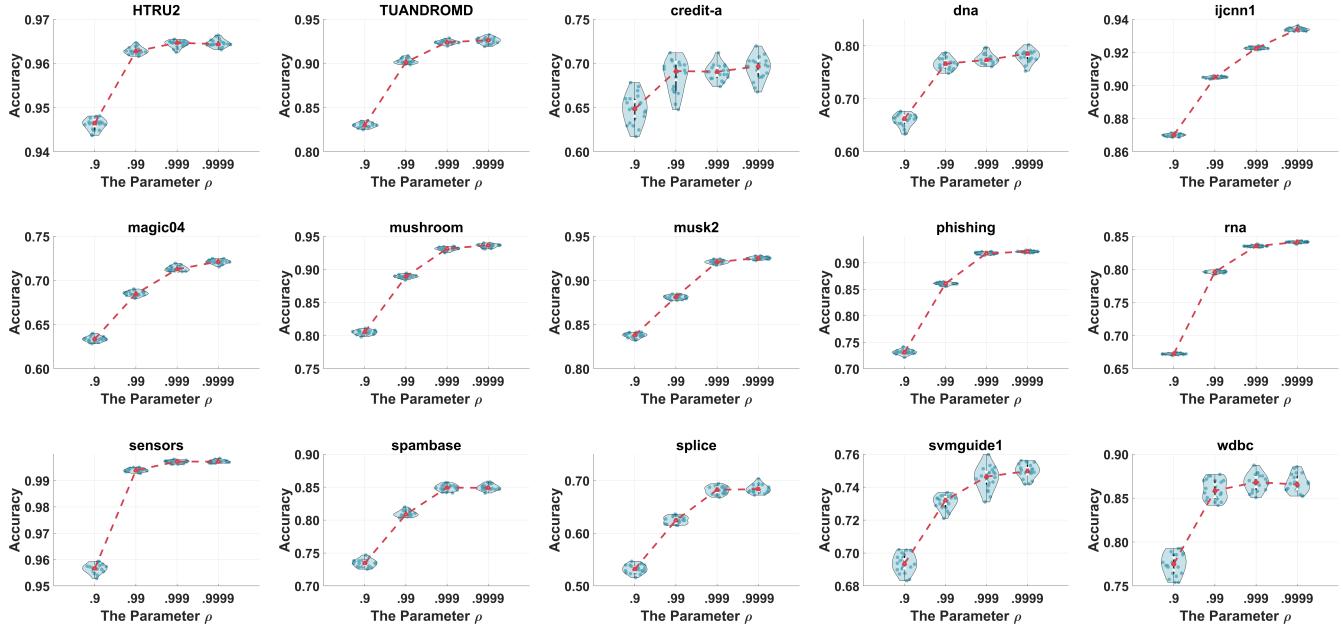


Fig. S16: The parameter ρ study under 2-period synchronized evolution scenarios.

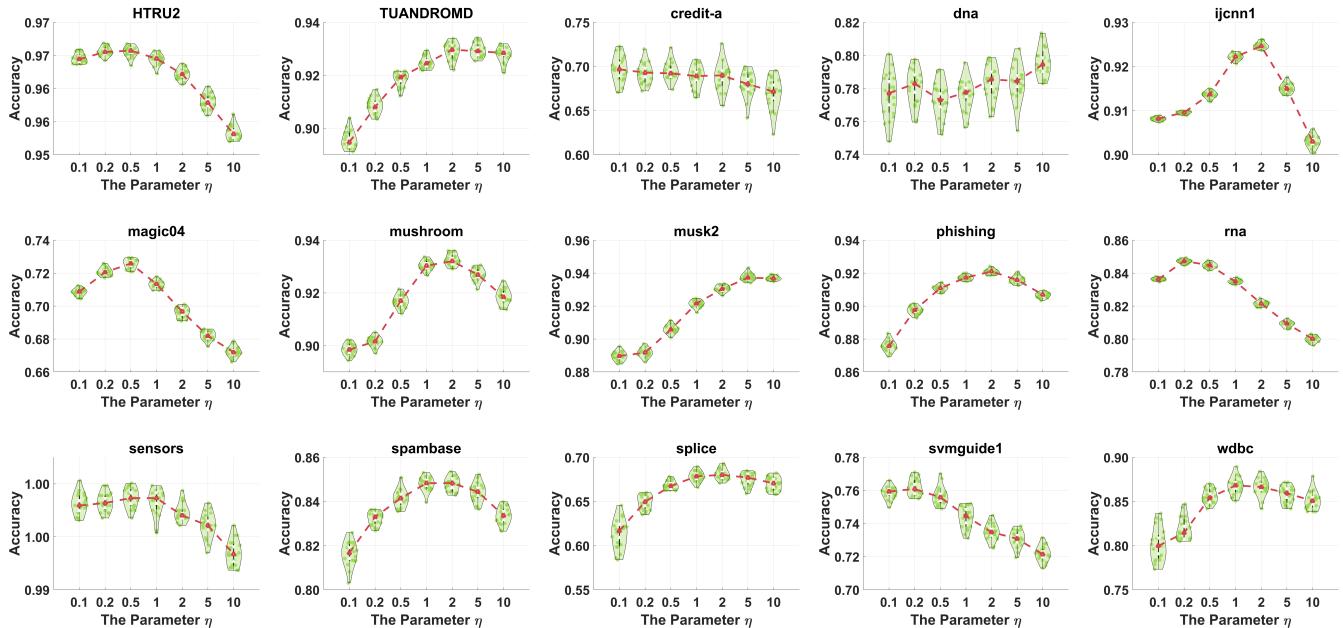


Fig. S17: The parameter η study under 2-period synchronized evolution scenarios.

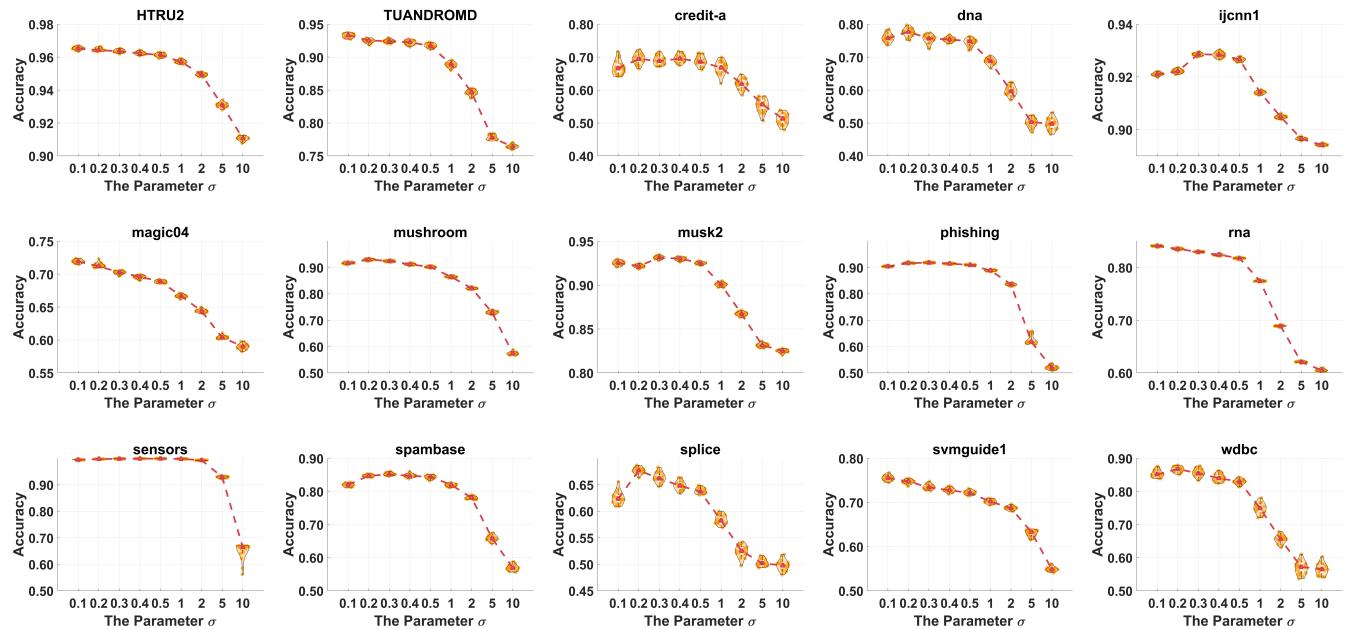


Fig. S18: The parameter σ study under 2-period synchronized evolution scenarios.

S6. EXPERIMENTS UNDER OVERLAP EVOLUTION

A. Comparison Results

Tables S6 provides the classification outcomes measured through the metrics of ACC and AUC. Despite OLEF-KL not being explicitly designed for scenarios involving feature evolution with overlap, it demonstrates commendable performance and yields satisfactory results. This notable achievement can be attributed to the inherent flexibility within our non-parametric model. In addition, the results obtained by FESL and PUFE show evidence that the act of recovering disappeared features and leveraging them in subsequent periods exhibits the promise of significantly enhancing model performance. This avenue of exploration presents an interesting area for the extension of our non-parametric method.

TABLE S6: The Performance comparison in the evolving with overlap scenarios (mean \pm std). \bullet/\odot indicate that OLEF-KL is significantly better/worse than competitors (hypothesis supported by paired t-tests at 0.05 significant level.). The best results are indicated in **bold**.

Metric	Datasets	NOGD	NPA	FESL-c	FESL-c	PUFE	OLEF-KL
ACC	HTRU2	0.929 \pm 0.001•	0.946 \pm 0.005•	0.924 \pm 0.001•	0.923 \pm 0.001•	0.924 \pm 0.001•	0.974 \pm 0.001
	TUANDROMD	0.828 \pm 0.004•	0.864 \pm 0.013•	0.816 \pm 0.010•	0.816 \pm 0.010•	0.817 \pm 0.010•	0.923\pm0.009
	credit-a	0.628 \pm 0.021•	0.707 \pm 0.020•	0.674 \pm 0.025•	0.670 \pm 0.020•	0.673 \pm 0.024•	0.826\pm0.009
	dna	0.612 \pm 0.026•	0.665 \pm 0.013•	0.689\pm0.034	0.683\pm0.035	0.690\pm0.037	0.691\pm0.015
	ijcnn1	0.902 \pm 0.001•	0.889 \pm 0.002•	0.903 \pm 0.001•	0.902 \pm 0.002•	0.902 \pm 0.002•	0.928 \pm 0.003
	magic04	0.675 \pm 0.004•	0.631 \pm 0.005•	0.681 \pm 0.006•	0.682 \pm 0.006•	0.682 \pm 0.006•	0.755 \pm 0.008
	mushroom	0.837 \pm 0.022•	0.879 \pm 0.006•	0.855 \pm 0.015•	0.854 \pm 0.015•	0.855 \pm 0.015•	0.992\pm0.001
	musk2	0.619 \pm 0.009•	0.598 \pm 0.008•	0.625 \pm 0.011•	0.619 \pm 0.009•	0.620 \pm 0.007•	0.896\pm0.006
	phishing	0.675 \pm 0.003•	0.738 \pm 0.019•	0.718 \pm 0.032•	0.720 \pm 0.028•	0.717 \pm 0.032•	0.819\pm0.022
	rna	0.758 \pm 0.048•	0.809 \pm 0.012•	0.780 \pm 0.042•	0.781 \pm 0.042•	0.782 \pm 0.042•	0.889\pm0.008
	sensors	0.666 \pm 0.004•	0.672 \pm 0.003•	0.732\pm0.050 ○	0.730\pm0.043 ○	0.709\pm0.057 ○	0.677 \pm 0.003
	spambase	0.748 \pm 0.006•	0.792 \pm 0.017•	0.755 \pm 0.009•	0.751 \pm 0.006•	0.753 \pm 0.007•	0.868\pm0.008
	splice	0.517 \pm 0.007•	0.560 \pm 0.007•	0.560 \pm 0.011•	0.561 \pm 0.012•	0.565 \pm 0.014•	0.632\pm0.009
	svmguide1	0.646 \pm 0.017•	0.720 \pm 0.006•	0.707 \pm 0.040•	0.709 \pm 0.041•	0.710 \pm 0.042•	0.895\pm0.002
	wdbc	0.623 \pm 0.009•	0.658 \pm 0.020•	0.623 \pm 0.009•	0.623 \pm 0.009•	0.623 \pm 0.008•	0.826\pm0.014
AUC	HTRU2	0.687 \pm 0.009•	0.846 \pm 0.011•	0.928\pm0.003 ○	0.926\pm0.003 ○	0.928\pm0.003 ○	0.888 \pm 0.005
	TUANDROMD	0.533 \pm 0.011•	0.801 \pm 0.026•	0.784 \pm 0.036•	0.780 \pm 0.035•	0.781 \pm 0.036•	0.846\pm0.030
	credit-a	0.535 \pm 0.036•	0.701 \pm 0.019•	0.762 \pm 0.029•	0.747 \pm 0.025•	0.760 \pm 0.027•	0.837\pm0.013
	dna	0.575 \pm 0.043•	0.660 \pm 0.015•	0.770\pm0.027 ○	0.756\pm0.027 ○	0.764\pm0.028 ○	0.685 \pm 0.015
	ijcnn1	0.502 \pm 0.005•	0.681\pm0.006 ○	0.691\pm0.032 ○	0.683 \pm 0.034	0.686 \pm 0.036	0.673 \pm 0.011
	magic04	0.523 \pm 0.008•	0.591 \pm 0.006•	0.714\pm0.014 •	0.709\pm0.015 ○	0.712\pm0.015 ○	0.703 \pm 0.012
	mushroom	0.830 \pm 0.023•	0.882 \pm 0.005•	0.913 \pm 0.008•	0.912 \pm 0.007•	0.913 \pm 0.007•	0.992\pm0.001
	musk2	0.687 \pm 0.009•	0.657 \pm 0.012•	0.787\pm0.009 ○	0.774\pm0.007 ○	0.779\pm0.005 ○	0.753 \pm 0.013
	phishing	0.547 \pm 0.006•	0.730 \pm 0.021•	0.805\pm0.028	0.801\pm0.027	0.799\pm0.029	0.809 \pm 0.025
	rna	0.585 \pm 0.054•	0.794 \pm 0.012•	0.866\pm0.058	0.866\pm0.059	0.867\pm0.056	0.880 \pm 0.008
	sensors	0.672 \pm 0.073	0.674 \pm 0.004•	0.834\pm0.040 ○	0.836\pm0.033 ○	0.814\pm0.045 ○	0.676 \pm 0.004
	spambase	0.749 \pm 0.007•	0.792 \pm 0.018•	0.804 \pm 0.012•	0.798 \pm 0.007•	0.799 \pm 0.007•	0.848\pm0.009
	splice	0.503 \pm 0.012•	0.559 \pm 0.011•	0.587 \pm 0.016•	0.587 \pm 0.017•	0.593 \pm 0.021•	0.627\pm0.011
	svmguide1	0.551 \pm 0.031•	0.723 \pm 0.007•	0.843 \pm 0.009•	0.843 \pm 0.009•	0.846 \pm 0.009•	0.891\pm0.003
	wdbc	0.504 \pm 0.024•	0.647 \pm 0.017•	0.669 \pm 0.029•	0.664 \pm 0.019•	0.675 \pm 0.032•	0.816\pm0.016
win-tie-lose		29-1-0	29-0-1	21-2-7	22-3-6	22-3-6	-

B. Nemenyi Test

Figure S19 presents the results of Nemenyi tests under overlapping evolving scenarios. In this context, the critical distance is computed as 1.9469 ($q_{0.05} = 2.984, K = 6, N = 60$). The analysis reveals that OLEF-KL demonstrates a significant advantage over all competing methods in terms of ACC. Moreover, concerning AUC, OLEF-KL performs competitively compared to PUFE and FESL-c.

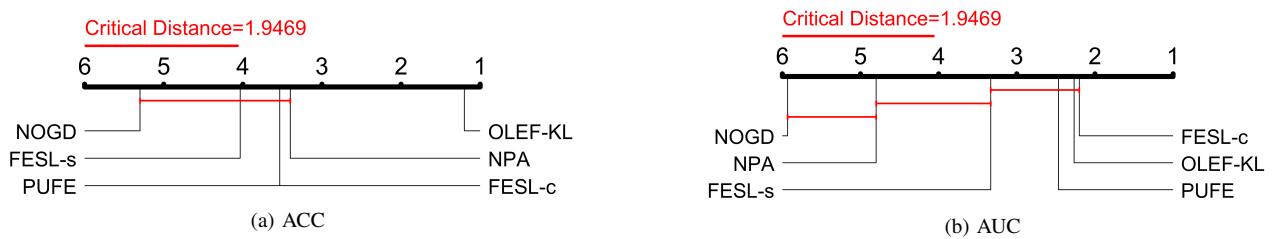


Fig. S19: OLEF-KL against competitors under overlap evolving scenario with the Nemenyi test at significance level $\alpha = 0.05$.

S7. TIME ANALYSIS

A. Theoretical Time complexity Analysis

Table S7 provides a theoretical time analysis of different methods within the one-pass online learning setting. Traditional algorithms such as NOGD and NPA, not designed to handle feature evolution, are labeled as "not applicable" in terms of time complexity during the feature adaptation procedure. As observed in the table, our proposed method generally demands more computational time, attributed to the utilization of a non-parametric model. Notably, when $\rho = 1$, the time complexity of OLEF-KL₁ increases with time. To address this, we advocate the use of a time-decay parameter, ρ , to ensure computational costs within a specified budget.

TABLE S7: The time complexity analysis of different methods under the one-pass online learning setting.

	Sample Learning	Feature Adaptation
NOGD	$\mathcal{O}(d_1)$	Not Applicable
NPA	$\mathcal{O}(d_1)$	Not Applicable
OLSF_{OGD}	$\mathcal{O}(d_t)$	$\mathcal{O}(1)$
OLSF	$\mathcal{O}(d_t)$	$\mathcal{O}(1)$
OLI2DS	$\mathcal{O}(d_t)$	$\mathcal{O}(1)$
OLVF	$\mathcal{O}(d_t)$	$\mathcal{O}(1)$
OPID	$\mathcal{O}(d_t ^3)$	$\mathcal{O}(d_t ^2)$
FESL	$\mathcal{O}(d_t)$	$\mathcal{O}(B^3)$
PUFE	$\mathcal{O}(d_t)$	$\mathcal{O}(B^3)$
OLEF-KL_{ρ}	$\mathcal{O}\left(\frac{\ln \epsilon}{\ln \rho} d_t \right)$	$\mathcal{O}\left(\frac{\ln \epsilon}{\ln \rho} d_t \right)$
OLEF-KL₁	$\mathcal{O}(t d_t)$	$\mathcal{O}(t d_t)$

- B is the number of samples in the overlap evolving scenarios.
- OLEF-KL _{ρ} is the proposed method when $\rho < 1$.

B. Experimental Time comparison

Table S8 presents the computation time (in seconds) for various algorithms in the generated evolving feature data streams. A notable observation is that OLEF-KL ($\rho = 1$) generally exhibits longer execution times compared to many of the other algorithms across a wide range of datasets. This suggests that OLEF-KL may be computationally more demanding. The rationale behind this computational discrepancy lies in the expanding nature of OLEF-KL models. As the number of data points increases, OLEF-KL models incrementally includes more support vectors to adapt to the evolving data landscape. A crucial technique for reducing the computational demands of OLEF-KL is the incorporation of a decay parameter $\rho < 1$. This parameter allows for the removal of older support vectors within the model. As a consequence, the computational complexity can be reduced.

TABLE S8: The time comparison of different methods under 2-period synchronized evolution scenarios.

Time(s)	NOGD	OPID	NPA	OLVF	OLI2DS	OLEF _{0.9}	OLEF ₁
HTRU2	0.056	8.468	0.068	0.236	0.415	0.484	0.649
TUANDROMD	0.033	6.415	0.033	0.144	0.186	0.410	1.407
credit-a	0.003	0.315	0.003	0.011	0.017	0.025	0.023
dna	0.028	1.374	0.024	0.046	0.068	0.272	0.465
ijcnn1	0.171	24.541	0.184	0.736	1.304	1.620	13.281
magic04	0.078	9.209	0.084	0.312	0.495	0.587	2.953
mushroom	0.032	4.015	0.034	0.127	0.209	0.266	0.665
musk2	0.046	7.379	0.047	0.153	0.268	0.406	4.764
phishing	0.058	6.431	0.052	0.192	0.337	0.445	3.63
rna	0.231	28.391	0.229	0.843	1.407	1.679	11.844
sensors	0.041	5.550	0.039	0.165	0.272	0.398	1.463
spambase	0.025	2.536	0.024	0.083	0.138	0.181	0.848
splice	0.020	1.853	0.019	0.065	0.101	0.126	0.48
svmguide1	0.031	3.495	0.031	0.115	0.179	0.203	0.368
wdbc	0.003	0.284	0.003	0.010	0.016	0.018	0.018

REFERENCES

- [1] B.-J. Hou, L. Zhang, and Z.-H. Zhou, "Learning with feature evolvable streams," in *Advances in Neural Information Processing Systems*, 2017, p. 1416–1426.
- [2] K. Crammer, O. Dekel, J. Keshet, S. Shalev-Shwartz, and Y. Singer, "Online passive-aggressive algorithms," *Journal of Machine Learning Research*, vol. 7, no. 19, pp. 551–585, 2006.
- [3] M. Zinkevich, "Online convex programming and generalized infinitesimal gradient ascent," *International Conference on Machine Learning*, pp. 928–936, 2003.
- [4] Q. Zhang, P. Zhang, G. Long, W. Ding, C. Zhang, and X. Wu, "Online learning from trapezoidal data streams," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 10, pp. 2709–2723, 2016.
- [5] E. Beyazit, J. Alagurajah, and X. Wu, "Online learning from data streams with varying feature spaces," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, 2019, pp. 3232–3239.

- [6] D. You, J. Xiao, Y. Wang, H. Yan, D. Wu, Z. Chen, L. Shen, and X. Wu, "Online learning from incomplete and imbalanced data streams," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 10, pp. 10650–10665, 2023.
- [7] C. Hou and Z.-H. Zhou, "One-pass learning with incremental and decremental features," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 11, pp. 2776–2792, 2017.
- [8] B.-J. Hou, L. Zhang, and Z.-H. Zhou, "Prediction with unpredictable feature evolution," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 10, pp. 5706–5715, 2021.