# KAN-MAMMOTE: Kolmogorov-Arnold Networks with Mamba and Mixture for Time Embedding

[Your Name]
[Your Affiliation/Department]

Submitted: [Date]

## Abstract

Modeling continuous-time systems with asynchronous events is a core challenge in modern ML. We propose **KAN-MAMMOTE**, a framework composed of two synergistic modules: **(i) K-MOTE**, an adaptive time encoder with a **Mixture-of-Experts** selecting among four expert basis types (Fourier, Spline, RKHS Kernel, Wavelet); and **(ii) Continuous-Mamba**, a sequential model using learned relative and absolute time embeddings. This approach dynamically adapts to the time structure, overcoming limitations in fixed encodings like LeTE and ensuring strong performance on complex temporal patterns.

## 1 Introduction

Continuous-time dynamic systems are ubiquitous (finance, climate, biology). Standard encodings like LeTE impose rigid basis families (e.g., B-splines, Fourier), limiting adaptivity. Further, sequence models like RNNs or Mamba assume discrete time, creating mismatch for irregular data.

K-MAMMOTE addresses these by:

- Selecting basis type *per timestamp* via MoE
- Using learned **absolute** and **relative** encodings
- Embedding sequences with **continuous-time-aware** Mamba

## 2 KAN-MAMMOTE Modules

### 2.1 K-MOTE: Adaptive Time Encoding

K-MOTE uses an MoE router to compute:

$$\alpha_e = \text{Softmax}(\text{MLP}_{\text{router}}([t_k, \Delta t_k, ...]))$$
$$\text{K-MOTE}(t_k) = \sum_{e \in \text{Top-}K_{\text{top}}} \alpha_e \cdot \phi_e(t_k)$$

Each **expert** $\phi_e$ is a learnable basis:

- **Fourier-KAN**: Learn non-integer frequencies $\lambda_j$.
- **Spline-KAN**: Use B-splines and MatrixKAN optimization.
- **RKHS-KAN**: Gaussian mixture kernel with learned anchors.
- **Wavelet-KAN**: Localized basis with learnable scales $s_j$.

### 2.2 Continuous-Mamba: Sequence Integration

- Input timestamps $t_k$ and $t_{k-1}$ are encoded by K-MOTE, each timestamps is encoded independently.
- Each of them is transformed by FasterKAN, then differenced to yield $\Delta t$ embedding.
- This $\Delta t$ embedding will will concatenate with mamba parameter in Mamba's internal SSM and

pass a mlp layer, becoming a new mamba parameter for internal SSM updates:

$$h_k = \overline{A}_k h_{k-1} + \overline{B}_k u_k$$

# 3 Regularization

- **Sobolev** $L^2$: Encourages smooth derivatives.
- **Total Variation**: Preserves sharp changes.
- **Load Balance Loss**: Ensures expert usage.