二、研究方案：

1.研究目标、研究内容和拟解决的关键问题：

1.1 研究目标

本研究拟解决不规则采样时间序列分类中存在的不规则时空建模困难、少数类识别率低等问题。通过构建神经随机微分方程与连续时间 Transformer 协同框架，设计类别感知的分类模型，以实现对不规则时序的精确连续表示建模，提高天文观测和医疗检测等领域不规则不均衡时序数据分析的准确性和可靠性。

具体而言，针对不规则采样和长间隔预测问题，采用线性神经 SDE 构建连续路径表示并精确计算状态转移；针对不规则时序的时序依赖建模问题，引入连续时间 Transformer 架构，通过连续时间的注意力机制有效捕捉长时间、不均匀跨度的复杂模式；针对类别不平衡问题，结合混合重采样和类别感知分组注意力机制，从数据层和模型层协同优化。最终构建一个对复杂时序数据具有强大适应能力的分类框架，为相关领域提供高效可靠的数据分析工具。

1.2 研究内容

(1) 基于 Neural SDE 和连续时间 Transformer 的不规则时序连续时间建模

针对单独使用 Neural SDE 缺乏全局依赖建模能力、单独使用连续时间 Transformer 无法刻画系统内在动力学的根本缺陷，提出两者结合的架构。Neural SDE 虽能通过微分方程描述连续演化过程，但其马尔可夫性假设限制了对历史信息的利用，难以捕获远距离时间点之间的复杂依赖；连续时间 Transformer 虽能通过注意力机制建模任意时间点间的关联，但缺乏对系统动力学规律的显式建模，在长时间外推时容易失真。通过将微分方程的动力学建模能力与 Transformer 的全局依赖捕获能力相结合，既保证了物理规律的连续性约束，又实现了跨时间尺度的灵活建模，解决了单一方法在不规则时序建模中的根本局限。

(2) 基于混合重采样和 CGA 的类别不平衡时序分类

针对只使用重采样导致模型过度依赖合成样本、只使用 CGA 在极度不平衡下难以学到有效表示的困境，提出数据-模型协同优化策略。仅依靠混合重采样虽能平衡训练集分布，但合成样本与真实样本的分布差异可能误导模型学习，且无法解决测试集仍然不平衡的问题；仅依靠 CGA 虽能增强少数类特征学习，但在训练样本极度稀缺时，注意力机制缺乏足够的样本多样性来学习鲁棒的类别特定模式。通过重采样为 CGA 提供更丰富的训练信号，同时 CGA 的类别感知机制补偿合成样本的分布偏差，实现了从数据增强到特征学习的全流程优化，有效解决了极度不平衡

场景下的分类问题。

1.3 拟解决的关键问题

(1) 如何实现不规则采样时序的连续建模

不规则采样时间序列的核心挑战在于观测点时间分布的随机性和采样间隔的巨大差异，传统深度学习模型依赖固定时间步长假设，难以自然处理这种不规则性。现有方法往往采用简单的插值或时间编码技术，不仅造成信息损失，更无法准确捕获时序数据中的动态演化规律。在医疗监测数据根据临床需要采集、天文观测受天气和设备限制形成高度不规则的观测模式等实际场景中，数据呈现出复杂的时空耦合特性，如何构建一个统一的深度学习架构，既能自然处理不规则采样特性，又能准确建模连续时间动态，同时在长时间间隔情况下保持预测精度并量化预测不确定性，是本研究需要解决的首要关键问题。

(2) 如何解决不规则采样时序的类别不均匀的问题

在时间序列分类任务中的类别不平衡是一个普遍且严重的问题，在天文光变数据、医疗诊断等关键应用中，目标类别的样本往往只占总体的极小比例。这种不平衡不仅体现在样本数量上，更体现在时序模式的复杂度差异上——少数类往往具有更加多样和复杂的时序演化模式，但由于样本稀缺而难以被充分学习。现有解决方案通常将时序建模和类别平衡作为两个独立问题分别处理，忽视了它们之间的内在联系，简单的重采样技术难以保持时序数据的连贯性和物理意义，通用的类别平衡损失函数未能考虑时序特有的上下文依赖性，如何在设计专门的架构，使少数类在特征空间中获得充分表达的同时保持时序建模的准确性，实现不同类别间的有效信息交互，从数据层和模型层协同优化构建端到端的解决方案，是本研究面临的另一个关键挑战。

2. 拟采取的研究方法（或技术路线、实验方案）及可行性分析

2.1 研究框架概述

本研究提出一种创新的两阶段串联架构来处理不规则采样时间序列的不均衡分类问题。该架构的核心思想是将连续时间建模与类别感知机制有机结合，通过先建模、后平衡的策略系统性地解决不规则采样和类别不均衡的双重挑战。

具体而言，第一阶段采用基于神经随机微分方程的连续时间建模方法，将不规则采样的离散观测嵌入到连续时间随机过程中。这一阶段包含两个关键组件：(1) 线性神经随机微分方程（LNSDE）作为基础框架，通过解析解形式实现高效的状态演化建模；(2) 连续时间注意力机制，通过重参数化技术在保持 Transformer 并行计算

优势的同时，实现对任意时间间隔依赖关系的精确捕捉。这两个组件的协同作用确保了对不规则时间序列动态特性的准确建模。

第二阶段在连续时间表示的基础上，通过混合重采样策略和改进的类别感知分组注意力 (CGA) 机制处理类别不均衡问题。数据层面采用时序感知的智能重采样，在平衡类别分布的同时保持时间连贯性；模型层面设计双路径注意力架构，为不同类别，特别是少数类提供专属的特征学习通道。

这种串联结构的合理性体现在三个方面：首先，连续时间建模为后续的类别平衡处理提供了高质量的特征表示，避免了在原始不规则数据上直接进行重采样可能破坏时间依赖性的问题；其次，在连续时间域中进行类别平衡操作能够更好地保持时序的连贯性和动态特性；最后，两阶段的解耦设计使得每个模块可以独立优化，提高了方法的灵活性和可扩展性。通过 LNSDE 的解析求解能力和连续注意力机制的全局建模能力，第一阶段为第二阶段的类别平衡处理奠定了坚实基础，确保了整个框架在处理复杂时序分类任务时的有效性和鲁棒性。

2.2 不规则采样时间序列的连续时间建模

（1）线性神经随机微分方程(LNSDE)框架

不规则采样时间序列在实际应用中普遍存在，传统的离散时间模型通过插值或时间编码处理不规则采样，不可避免导致信息损失。本研究采用线性神经随机微分方程 (LNSDE) 作为连续时间建模的基础框架，将离散观测嵌入到连续随机过程中。

LNSDE 将系统状态演化建模为：

$$dz_t = f_\theta(z_t, t)dt + g_\phi(z_t, t)dW_t \tag{1}$$

其中z_t为隐状态，f_θ为漂移函数，g_ϕ为扩散函数，dW_t为维纳过程增量。为实现高效推理，采用线性化形式：$dz_t = A_t z_t dt + b_t dt + \sigma_t dW_t$

这种线性化设计的优势在于存在解析解：

$$z_t = \Phi(t, s)z_s + \int_s^t \Phi(t, r)b_r dr + \int_s^t \Phi(t, r)\,\sigma_r dW_r \tag{2}$$

其中$\Phi(t, s)$为状态转移矩阵。解析解避免了数值积分的误差累积，在长时间跨度预测中尤为重要。

（2）连续时间注意力机制

为克服传统 Transformer 在处理不规则时间序列时的局限，本研究设计了连续

时间注意力机制。该机制通过常微分方程建模注意力分数在连续时间域的演化，能够精确捕捉任意时间间隔内的依赖关系。

对于查询时刻t_j和历史时刻t_i，连续注意力权重定义为：

$$\alpha_i(t_j) = \frac{\int_{t_i}^{t_j} q(\tau) \cdot k_i(\tau)^T d\tau}{t_j - t_i} \tag{3}$$

为保持计算效率，采用重参数化技术将积分映射到标准区间：

$$\alpha_i(t_j) \approx \frac{1}{2} \sum_{p=1}^{P} \gamma_p \tilde{q}_{i,j}(\xi_p) \cdot \tilde{k}_{i,j}(\xi_p)^T \tag{4}$$

这种设计保持了 Transformer 的并行计算优势，同时实现了连续时间建模能力。通过多头注意力扩展，不同注意力头可以专注于不同时间尺度的依赖模式，进一步提升了模型的表达能力。

2.3 基于混合重采样和改进 CGA 的类别不平衡时序分类

针对不规则采样时间序列分类中的类别分布严重失衡问题，本研究提出基于混合重采样策略和改进类别感知分组注意力（CGA）机制的综合解决方案。该方案从数据层和模型层协同出发，系统性地解决少数类识别率低的核心挑战。

在数据层面，我们设计了智能混合重采样策略来平衡类别分布 2。对于少数类样本，采用改进的 SVM-SMOTE 算法进行智能合成。该算法首先识别少数类中靠近决策边界的关键样本，然后通过线性插值生成新的合成样本。具体地，对于少数类样本x_i及其 k 近邻中的样本x_j，合成样本生成公式为：

$$x_{new} = x_i + \lambda \cdot (x_j - x_i) \tag{5}$$

其中$\lambda \in [0,1]$为随机插值系数。特别地，在时间序列场景下，我们不仅在特征空间进行插值，还需要保持时序的连贯性。因此，对于时间戳t_i和t_j，相应的时间插值为：

$$t_{new} = t_i + \lambda \cdot (t_j - t_i) \tag{6}$$

这确保生成的样本既符合原始数据的特征分布，又保持合理的时间间隔模式。

对于多数类样本，我们采用改进的 Repeated ENN（Edited Nearest Neighbor）方法进行选择性欠采样[2]。该方法通过迭代清理多数类中的噪声和边界样本，判定规

则为：

$$R(x_i) = \sum_{j \in kNN(x_i)} \mathbb{I}(y_j \neq y_i) > \theta \cdot k \tag{7}$$

其中 $kNN(x_i)$表示样本x_i的 k 个近邻，$\mathbb{I}(\cdot)$为指示函数，$\theta \in (0,1)$为噪声判定阈值。当 $R(x_i)$为真时，样本x_i被识别为噪声或边界样本并被移除。这种方法保留最具代表性的核心样本，在减少样本数量的同时保持数据质量。

在模型层面，我们提出了基于类别感知分组的注意力机制。该机制的核心创新在于为每个类别 c 构建独立的特征表示路径，同时设计了类间语义交互模块。具体而言，输入特征 $X \in \mathbb{R}^{N \times d}$首先通过类别感知编码器生成类别特定的表示：

$$Z_c = Attention_c(XW_c^Q, XW_c^K, XW_c^V) \tag{8}$$

其中 $W_c^Q, W_c^K, W_c^V \in \mathbb{R}^{d \times d_c}$ 为类别 c 的专属投影矩阵，d_c为类别特定的隐藏维度。这种设计确保了少数类能够获得充分的表示能力，不会被多数类的特征所掩盖。

为了避免类别间信息的完全隔离，我们设计了基于语义相似度的类间交互机制。不同类别特征表示之间的语义相似度计算为：

$$S_{ij} = \frac{Z_i^T Z_j}{\| Z_i \| \| Z_j \|} \tag{9}$$

基于相似度矩阵 S，类间信息交换通过门控机制进行：

$$\tilde{Z}_i = Z_i + \sum_{j \neq i} \sigma(S_{ij} - \tau) \cdot Gate_{ij}(Z_j) \tag{10}$$

其中 $\sigma(\cdot)$为 sigmoid 函数，τ 为相似度阈值，$Gate_{ij}$为可学习的门控函数。这种设计既保护了各类别（特别是少数类）的独特特征，又能够利用类别间的互补信息提升整体分类性能。

此外，我们还引入了自适应的加权交叉熵损失函数：

$$\mathcal{L} = -\sum_{i=1}^{N} \sum_{c=1}^{C} w_c \cdot y_{ic} \log(\hat{y}_{ic}) \tag{11}$$

其中权重$w_c = \frac{N}{N_c \cdot C}$ N 为总样本数，N_c为类别 c 的样本数，C 为类别总数。这种机

制确保模型在整个训练过程中持续关注少数类的学习，避免在训练后期被多数类主导。通过上述数据层和模型层的协同优化，本方法能够有效解决不规则时序分类中的类别不平衡问题。

2.4 可行性分析

（1）理论基础成熟可靠

本研究涉及的核心理论方法均有坚实的数学基础和广泛的研究支撑。神经随机微分方程（Neural SDE）理论已在金融建模、物理系统模拟等领域得到验证，其解析解特性和不确定性量化能力为不规则时序建模提供了可靠保障。线性 SDE 的解析解形式已被证明能够避免数值积分的误差累积问题，特别适用于长时间跨度预测。连续时间注意力机制虽然相对较新，但其基于 ODE 的理论框架已在多项研究中被验证有效，重参数化技术确保了计算效率。类别不平衡处理方面，SMOTE、Repeated ENN 等重采样方法已经过大量实证检验，在不同领域被证实能够有效改善少数类识别性能。这些成熟理论的有机结合为本研究提供了坚实的理论支撑。

（2）技术方案切实可行

本研究提出的两阶段串联架构具有良好的工程可行性。首先，LNSDE 的线性化设计使得状态转移计算可以通过矩阵运算高效实现，避免了复杂的数值积分过程。连续时间注意力机制通过重参数化保持了 Transformer 的并行计算优势，无需修改底层计算框架即可在现有深度学习平台上实现。其次，混合重采样策略和 CGA 机制均可作为独立模块集成到整体框架中，模块化设计提高了系统的可维护性和可扩展性。各模块之间通过标准的张量接口进行数据交互，确保了系统集成的便利性。此外，本研究的关键算法均有开源实现可供参考，降低了技术实现的难度和风险。

（3）数据资源充分可得

时间序列分类领域已积累了大量公开数据集，为本研究提供了充分的实验资源。这些数据集不仅具有不规则采样特性，还大多数存在类别不平衡问题，完全符合本研究的应用场景。数据集的多样性也保证了方法的泛化性验证。同时，这些数据的收集和使用均遵守相关法律和伦理规范，确保研究的合规性。

（4）研究条件完备充足

本研究团队在时间序列分析、深度学习和不平衡分类等相关领域具有丰富的研究经验和技术积累。团队成员曾参与多项国家级科研项目，在顶级会议和期刊发表相关论文，具备扎实的理论功底和工程实践能力。硬件方面，实验室配备了高性能