**1. Title & Author Information**

* 标题示例：
  + *"TCA-Enhanced Foundation Model for Cross-Hospital Medical Classification on Small and Imbalanced Data"*
  + *"* Domain-Adaptive Tabular Foundation Models with TCA: A Novel Approach for Imbalanced Medical Data *"*
* 作者及机构
* 通讯作者邮箱
* 贡献说明（CRediT）

**2. Abstract（摘要）**

* **Background**：小样本、类别不平衡、域偏移在医学AI中的挑战
* **Methods**：提出预训练Transformer+TCA跨域适应框架
* **Results**：在两个真实医院数据集（A: 295条，B: 190条）上验证AUC/SEN/SPE
* **Interpretation**：模型具备可推广性和临床部署潜力

**3. Introduction（引言）**

**3.1 医学场景问题**

* 医疗AI部署的主要瓶颈：数据异质性、小样本、高偏移
* 多中心部署难以泛化

**3.2 方法学背景**

* Transformer 在表格数据的成功应用（预训练）
* TCA 作为经典、稳健的无监督域对齐方法

**3.3 技术挑战与动机**

* 表格基础模型在域外泛化不足
* 小样本条件下UDA方法稳定性差
* 医疗数据分布偏移 + 不平衡加剧模型崩溃

**3.4 研究目标与贡献**

* 集成预训练表格Transformer与TCA
* 提出TCA-enhanced Foundation Model
* 实证其在跨医院、小样本任务中的有效性
* 对比多个UDA方法，验证其稳定性与优势

**4. Methods（方法）**

**4.1 数据集与样本**

* 数据集A（某医院，n=295）与数据集B（另一医院，n=190）
* 包括病人特征、临床标签（异常/正常）
* 病种分布与样本分布（附表格）
* 伦理声明与匿名化处理

**4.2 数据预处理**

* 缺失值处理、标准化方法
* 特征筛选：递归特征消除（RFE）
* 类别变量编码方式

**4.3 模型架构**

**4.3.1 表格预训练模型（Tabular Transformer）**

* 输入：结构化样本（列视为token）
* 编码器：多头注意力编码结构
* 表示学习：输出 pooled representation 表征样本

**4.3.2 Transfer Component Analysis（TCA）**

* 核函数（RBF）映射
* MMD距离度量
* 映射矩阵学习 → 降维+对齐
* 对齐后送入分类器

**4.3.3 集成架构**

* Pipeline：输入 → Transformer → TCA映射 → 分类器
* 支持端到端或部分冻结训练
* 架构图（可附Mermaid或LaTeX图）

**4.4 训练策略**

* 冻结Transformer，仅训练TCA+Classifier
* 类别加权、采样策略应对不平衡
* 超参数搜索范围（如TCA维度/γ参数）
* 实现框架（PyTorch）

**4.5 验证设计**

* 内部验证（数据A的5折交叉验证）
* 外部验证（train on A，test on B）
* 方法对比：不使用TCA / 使用CORAL / 使用DANN等
* 多种随机种子重复实验

**4.6 评估指标**

* AUC、ACC、F1、Sensitivity、Specificity
* 95% CI
* ROC曲线 + Calibration Curve + Decision Curve
* PCA / t-SNE前后可视化
* 域距离对比（MMD/Wasserstein）

**5. Results（结果）**

**5.1 数据描述**

* 表格展示A/B医院特征分布
* 患者统计学特征差异（年龄、性别、疾病类型）

**5.2 模型性能**

* 表格：内外部验证性能（TCA vs Non-TCA）
* 图：ROC、t-SNE图等

**5.3 方法对比**

* 与ML方法（XGBoost、RF）、DL模型、UDA方法对比
* 表格+统计显著性检验（如DeLong检验）

**5.4 子组分析**

* 性别、年龄、分中心性能对比
* 高/低特征重要性子集泛化能力

**5.5 可解释性分析**

* 特征重要性排序（SHAP、线性权重）
* 医生友好解释举例

**5.6 鲁棒性分析**

* 随机子集扰动、特征缺失模拟
* 多次实验方差分析

**6. Discussion（讨论）**

**6.1 主要发现**

* 强化了TCA在小样本场景的域泛化能力
* 表格基础模型的潜力首次在跨域医学数据中系统验证

**6.2 临床价值**

* 可部署性强：泛化+稳定+轻量级
* 辅助决策系统的可行性

**6.3 技术创新**

* 预训练表格Transformer + TCA首次集成
* 提供了一种新型高效的无监督跨域学习路径

**6.4 局限性**

* 数据量少，标签质量可能存在主观偏差
* 未测试大规模中心（>5医院）

**6.5 未来方向**

* 联邦学习 + TCA集成
* 多源域适应（MSDA）扩展
* 联动LMM/LLM进行多模态预测

**7. Conclusion（结论）**

* 提出一种TCA增强的表格预训练模型，适用于小样本跨域医学分类任务
* 实证其在真实医疗数据上的可推广性
* 为未来多中心AI部署奠定了方法基础

**8. Contributors**

**9. Declaration of Interests**

**10. Code & Data Availability**

**11. References**

**12. Supplementary Materials**