

黄力恒

liheng.huang@outlook.com | 广州 | 个人主页



教育背景

香港城市大学 生物统计学硕士

2025.09 – 2026.10

GPA: 3.83/4.3

华东交通大学 数据科学与大数据技术学士

2020.09 – 2024.06

平均分: 84.3/100; 优秀毕业论文奖 (学院前 1%)

工作经历

康景堂有限公司 实习生 (业务系统开发与报表分析)

2026.05 – 2026.08

- 基于开源系统定制诊所收支管理工具, 关联患者、服务项目、收支明细与次卡账户, 构建按日、月、年及服务项目汇总的经营报表, 支持账务核对。
- 设计次卡发放、核销和人工调整的流水规则, 实现记录修改、删除及 Excel 导入撤回后的余额同步, 处理重复导入和同名客户识别问题。
- 实现导入前备份、人工调整原因记录和历史流水查询, 支持数据恢复及账务变更追溯。

深圳前海新心数字科技有限公司 数据分析实习生

2024.01 – 2024.06

- 使用逻辑回归、XGBoost 及 WOE / IV 方法参与客户特征分析与优先级排序, 支持短信营销客群筛选。
- 使用 PostgreSQL 编写客户、交易及信贷数据的多表查询, 整理金融业务数据并生成统计报表。

项目经历

基于 Transformer 的癌症生存分析 [GitHub]

2025.04 – 2025.08

- 基于 MSK-CHORD 纵向临床记录, 使用 SQLite 整合多源事件并构建 Landmark 序列样本, 实现训练侧类别编码、数值标准化及变长序列批处理, 控制患者间与未来信息泄漏。
- 使用 PyTorch 实现类别嵌入、Transformer 编码器、掩码池化与自定义 Cox 损失, 处理序列填充、并列事件时间及极端风险分数下的数值稳定性。

双区域纵向 CT 建模与跨中心预测 [项目说明]

2026.06 – 至今

- 面向治疗前后肿瘤与淋巴结 CT, 结合 SLIC 局部划分与预训练 ResNet18 多层特征, 通过训练折内 PCA 降维和 GMM 聚类构建影像表型, 形成包含局部异质性及淋巴结集合信息的 102 维患者表征。
- 融合 50 维空间与纵向特征, 联合建模肿瘤与淋巴结的影像表型及治疗前后变化; 在 827 名患者上采用患者级十折与三中心留一验证, 评估术后病理淋巴结状态预测的判别、校准及跨中心泛化表现。

多元函数型生存数据的 Cox Filtration 方法 [GitHub]

2026.05 – 2026.07

- 将多尺度共享表征学习扩展至右删失生存结局, 基于 Cox 偏似然的得分与信息几何学习导联间分组, 逐层提取共享与特有风险成分, 并通过残差化学习后续风险方向。
- 实现 Fusion 与 Directional 两类分组机制, 分别依据导联风险贡献的一致性和标准化风险方向构造候选; 利用递归变换将预测重构为“导联 × 时间”载荷, 解释不同导联及时间位置的贡献。
- 在 154,744 条十二导联心电记录上开展多划分验证; 全训练规模下, Directional 平均成分数从 72 降至 36.5, 平均 C-index 为 0.7460。

更多项目请见我的个人主页

技能

编程与数据: Python、SQL、R、PostgreSQL、Excel

机器学习与工具: PyTorch、scikit-learn、XGBoost、Transformer、特征工程、SHAP、Git、Linux

统计与实验: 回归分析、假设检验、抽样与区间估计、模型评价

语言能力: 普通话、粤语、英语 (雅思 7.0、英语六级)