

# 0.目录

---

## 0.目录

### 1. 整理迁移学习的结果

1.1. 迁移学习的结果简述:

1.2. 迁移学习总结

1.3. 去除Morgan分子指纹图片展示

1.4. 使用Morgan分子指纹图片展示

### 2.根据热力图检查去掉分子指纹后性能变化

2.1. 情况1: SGF-LR

2.2. 情况2: SGF-RF

2.3. 情况3: SGF-XGB

2.4. 情况4: SIF-LR

2.5. 情况5: SIF-RF

2.6. 情况6: SIF-RF

2.7. 性能指标总结

### 3. 最终总结

3.0 口语化总结

3.1 迁移效果

3.2 性能指标 (去掉分子指纹后的影响)

3.3 最佳组合

3.4 数据集内在特性总结

# 1. 整理迁移学习的结果

---

## 1.1. 迁移学习的结果简述:

---

4268数据集和WO2017数据集，两个重要数据集的迁移效果：

去除Morgan后：

- 在SIF任务上，三个模型的迁移F1均为0.6上下浮动（比使用Morgan情况稍好）。
- 在SGF任务上，则在0.7上下浮动。

不去除Morgan：

- 在SIF任务上，三个模型的迁移F1均为0.6左右。
- 在SGF任务上，则在0.7左右浮动

两个数据集在其他数据集上的迁移效果：

去除Morgan：

- 在SIF任务上，在其他模型上迁移效果一般。
- 在SGF任务上，在US98数据集上迁移效果很好（但是有异常值）

不去除Morgan：

- SIF上在其他数据集上的迁移效果一般（除了US2014，但是这个数据集太少了不考虑）
- SGF上，仍然在US98数据集上迁移效果比较好

**Morgan对迁移的影响：**

- 去掉Morgan会让迁移的F1分数更高

**特殊情况**

- 在SGF-RF的任务中，可以看出US98在4268和WO2017数据集上的迁移效果不错，并且去掉Morgan会让迁移F1分数提升。

**异常情况**

- 数据集导致很多情况无法查看迁移效果。
- SGB-XGB任务上，不去除分子指纹，WO2017向US98数据集的迁移效果突然下降 ( $F1=0.333$ )
- 4268和WO2017数据集向US98迁移时会出现 $F1=1$ 的情况，感觉不太正常。

## 1.2. 迁移学习总结

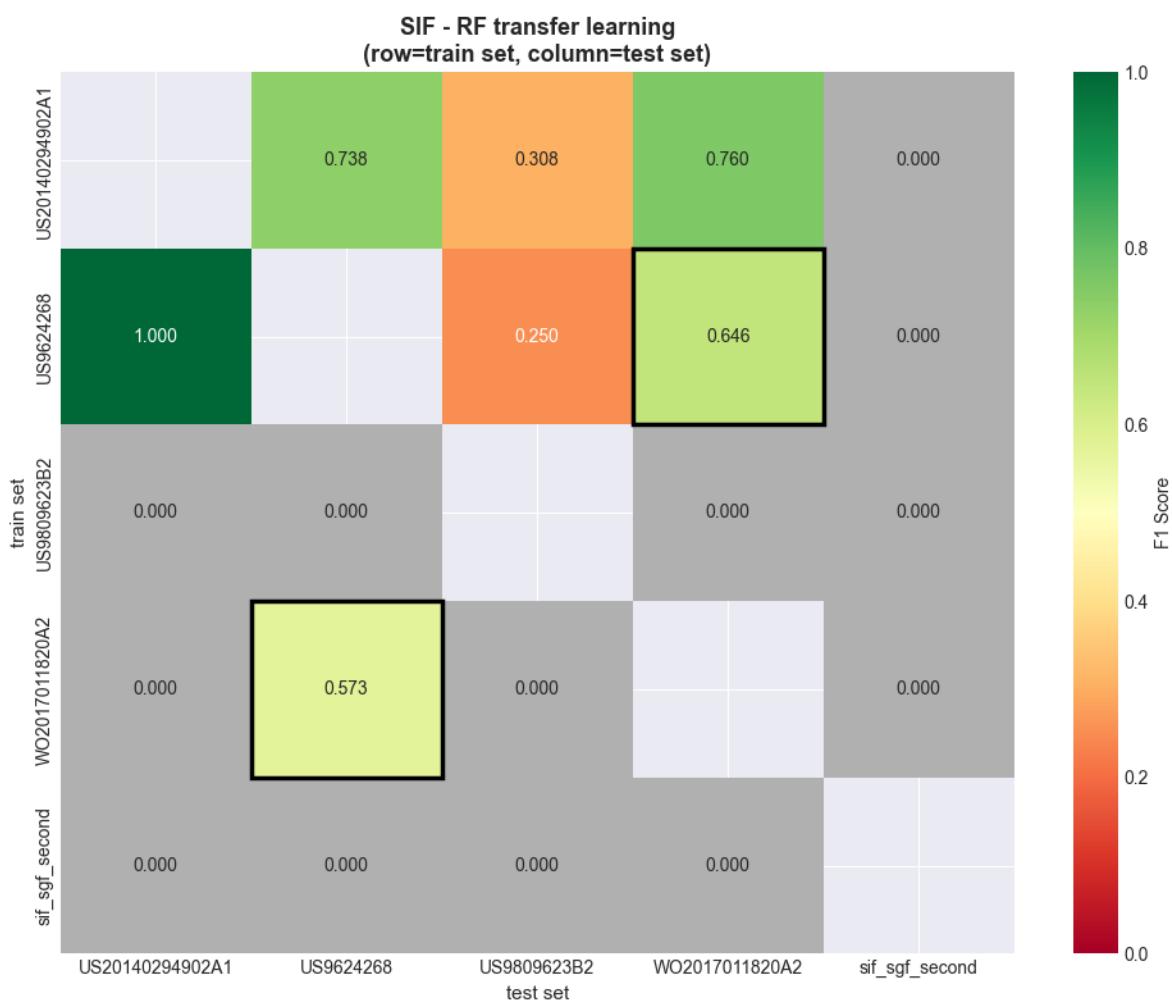
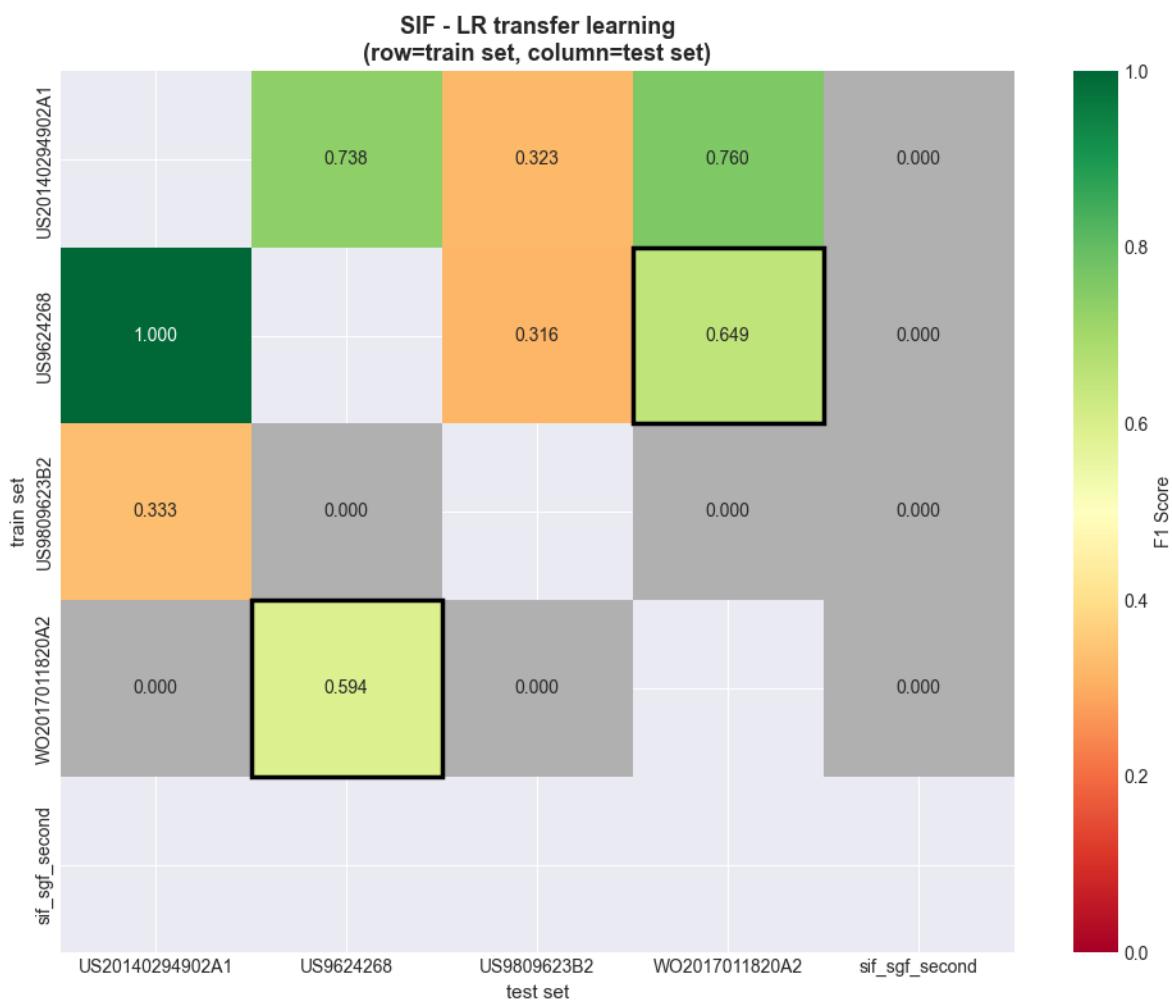
---

- Morgan去掉后带来的整体趋势是迁移效果更好
- US98数据集可能也有比较好的迁移效果

## 1.3. 去除Morgan分子指纹图片展示

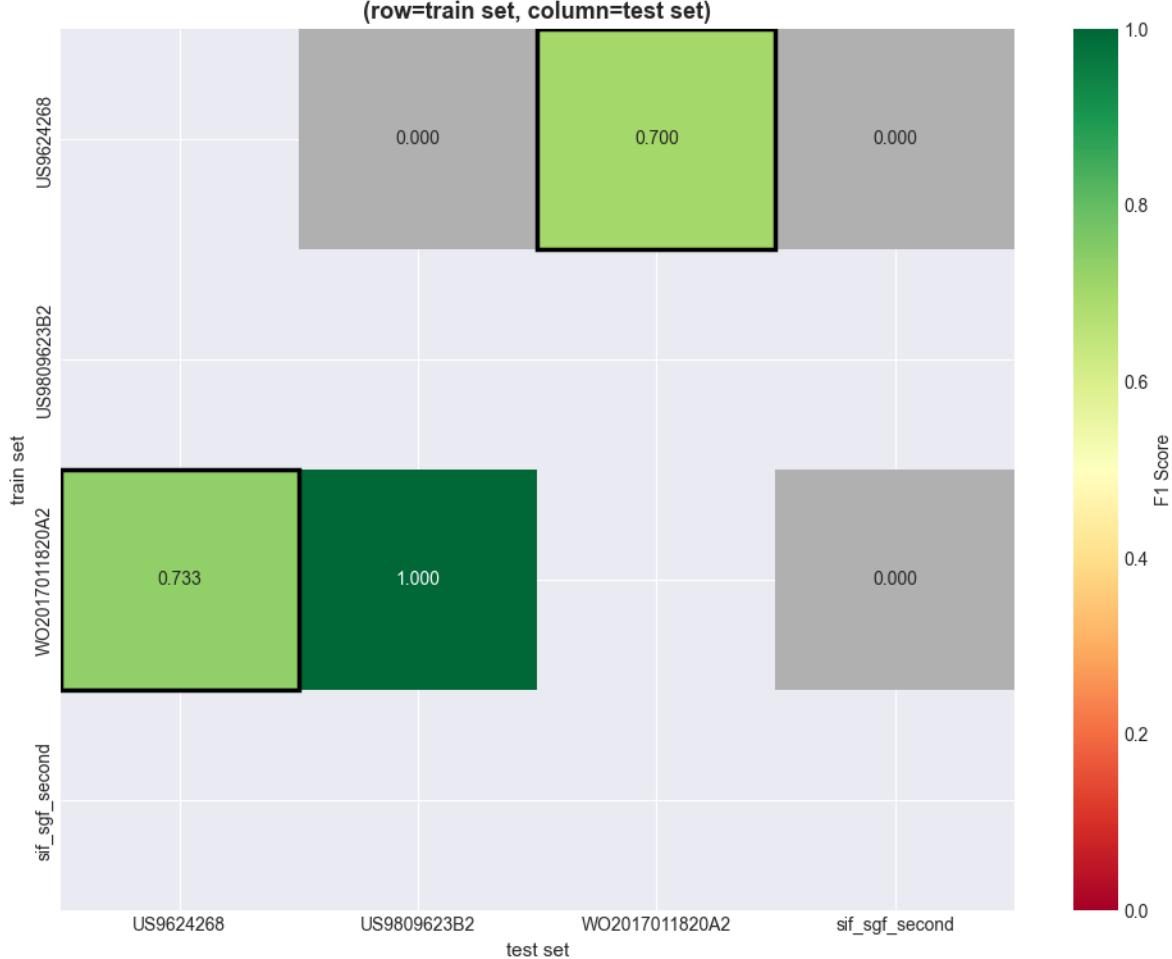
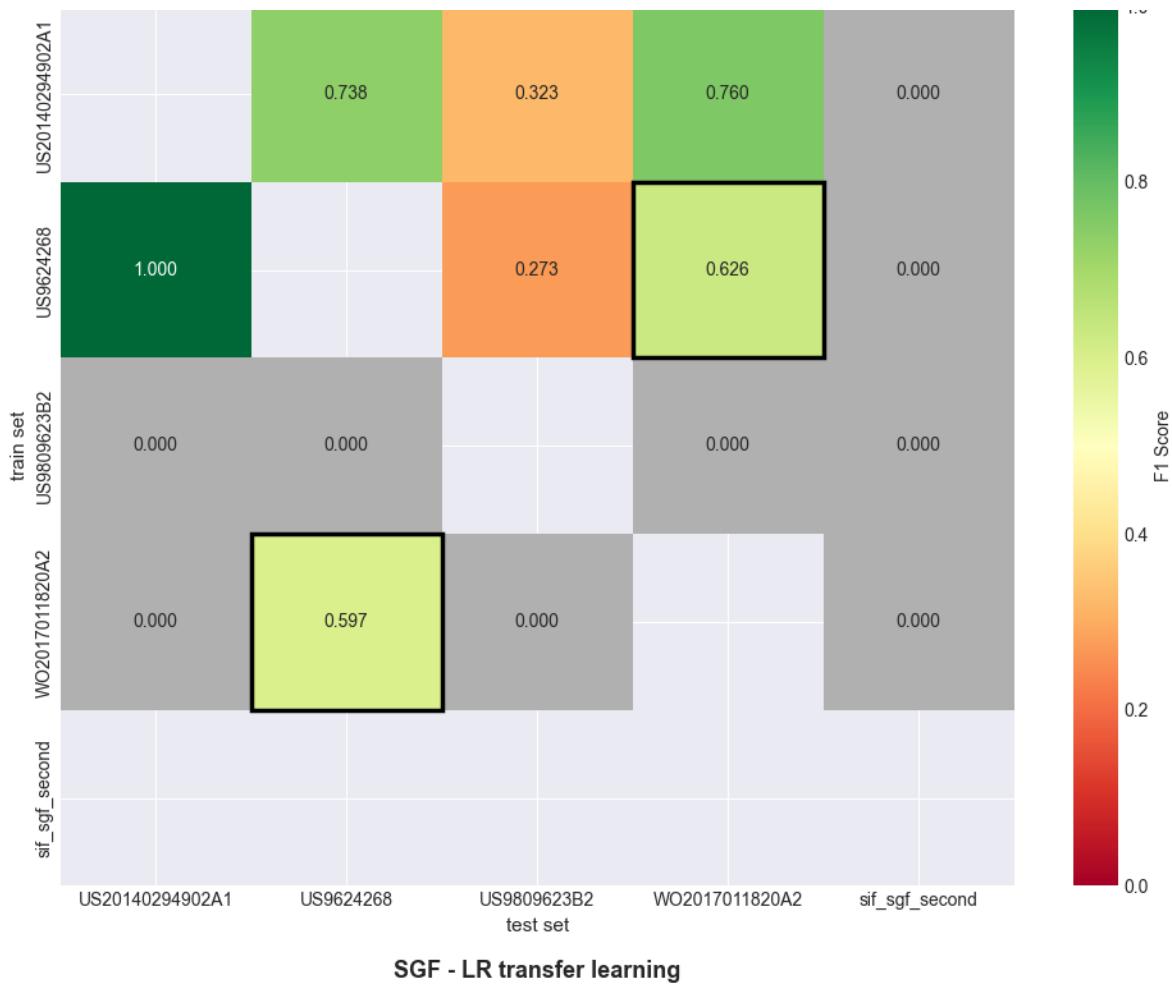
---

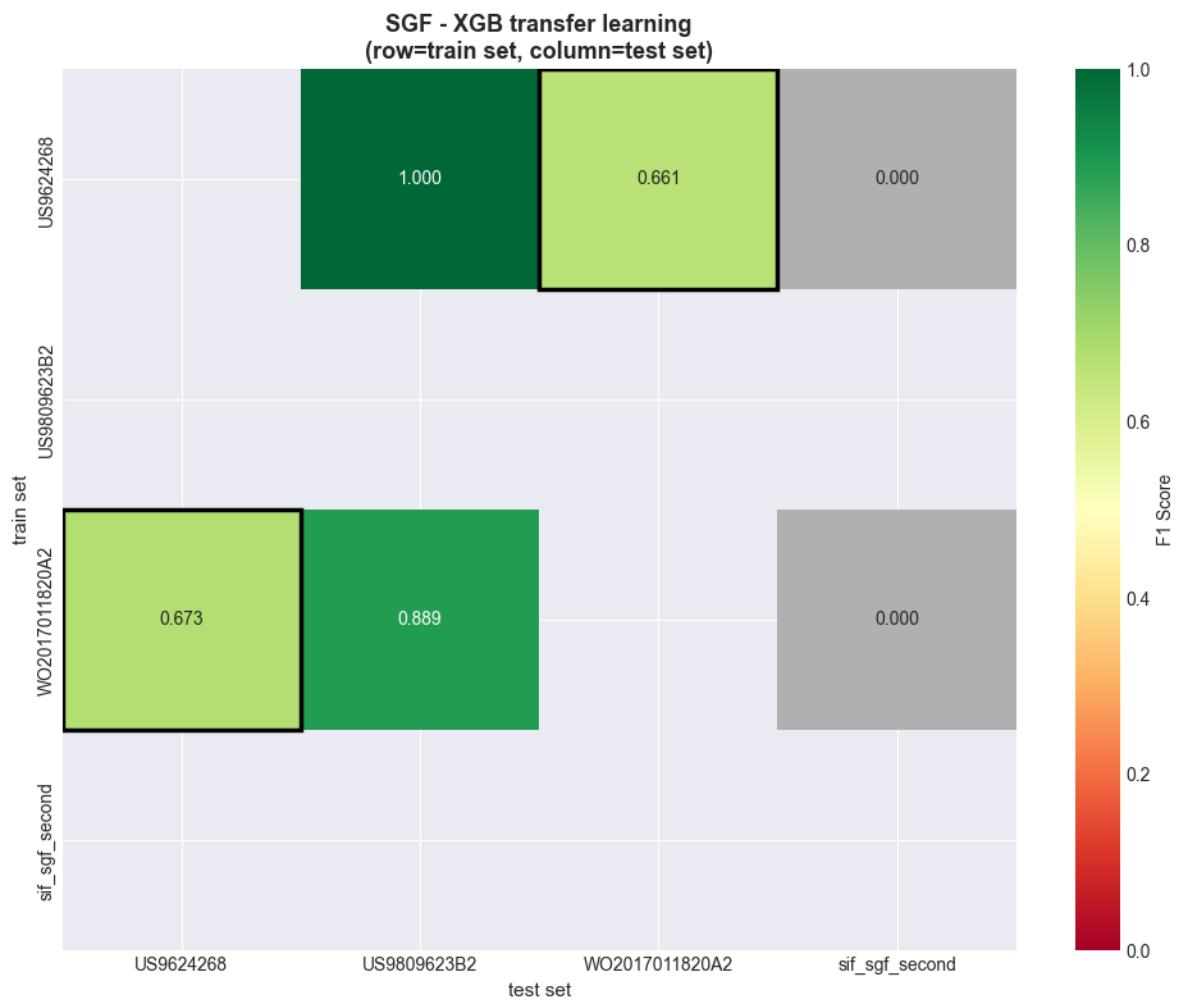
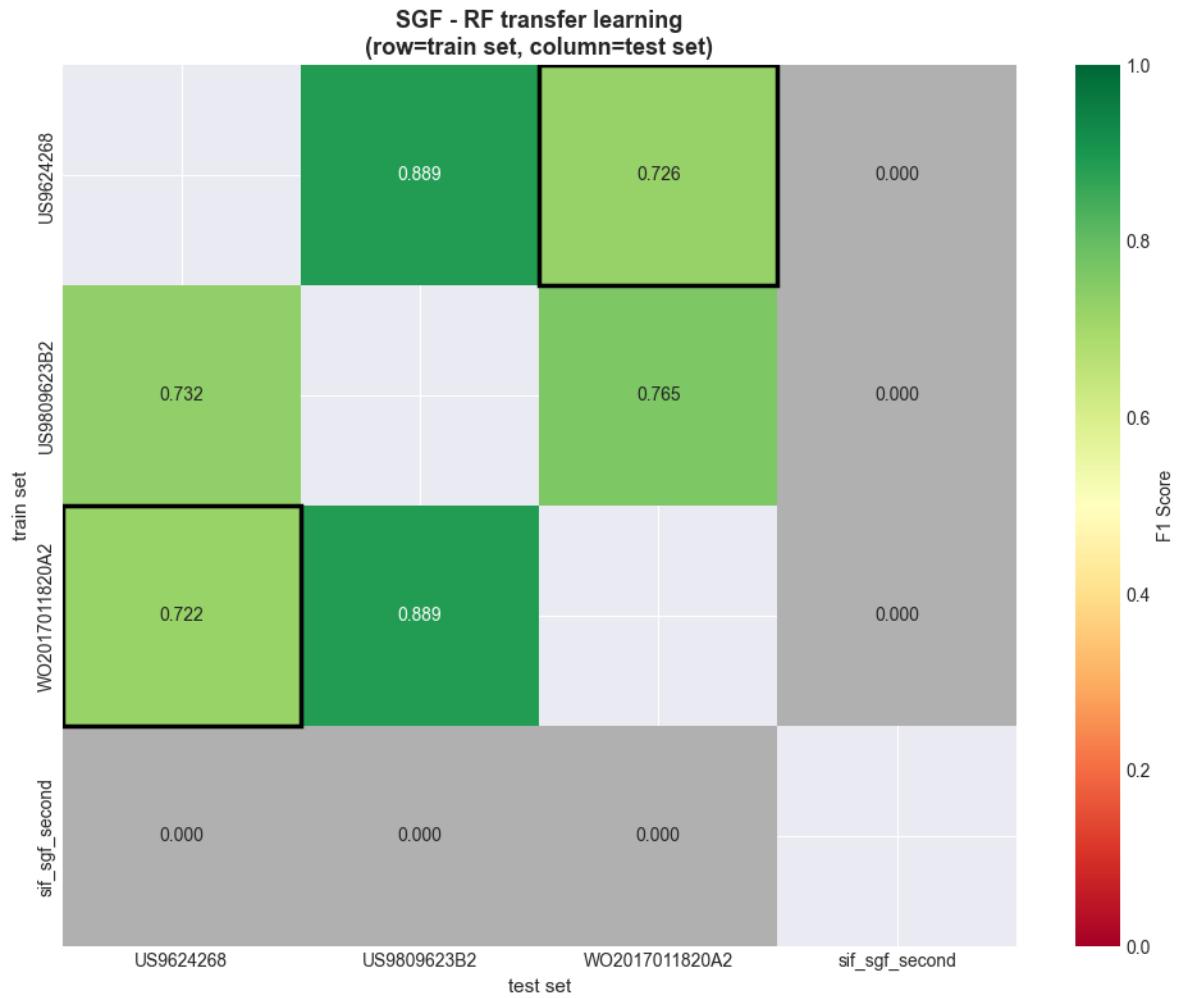




**SIF - XGB transfer learning**  
(row=train set, column=test set)

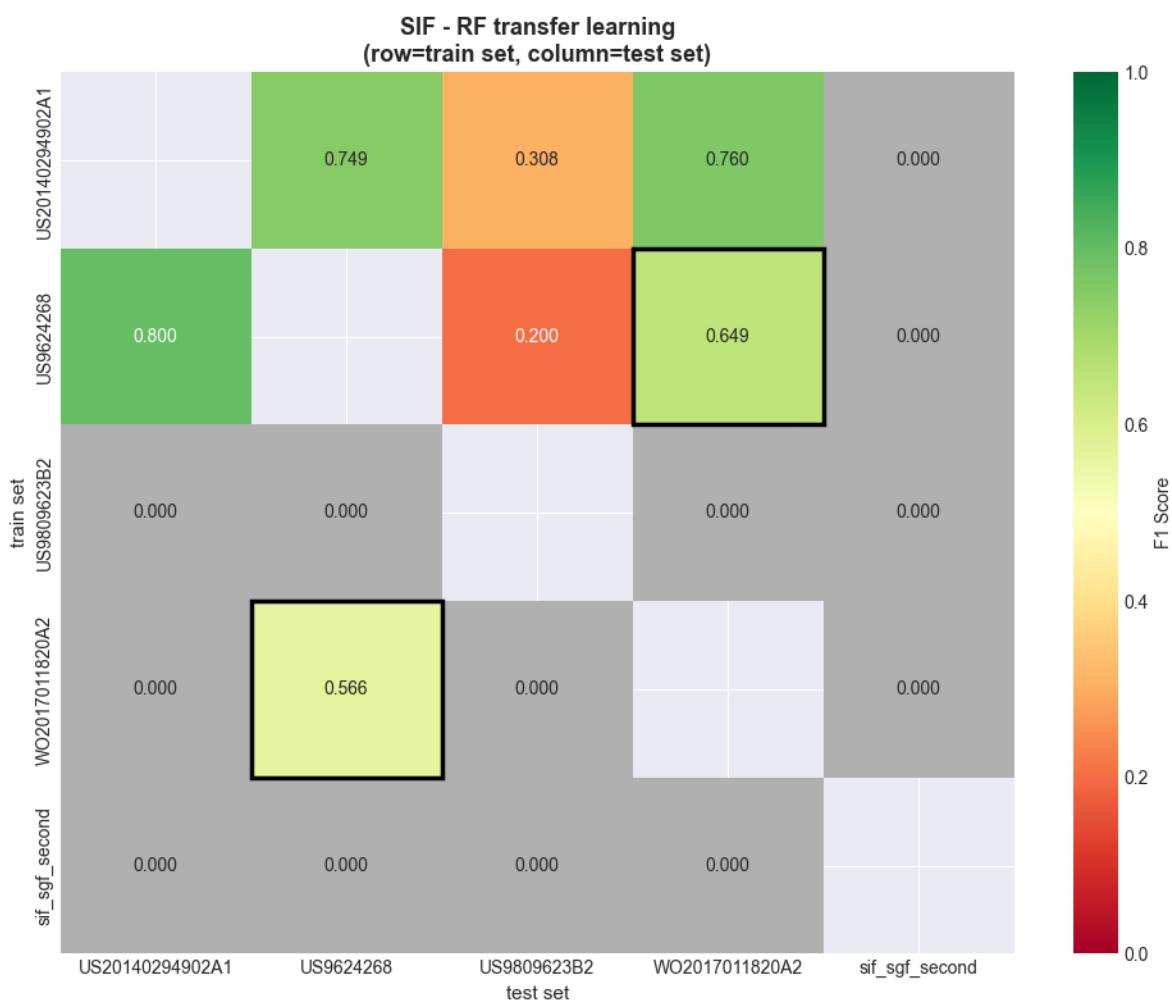
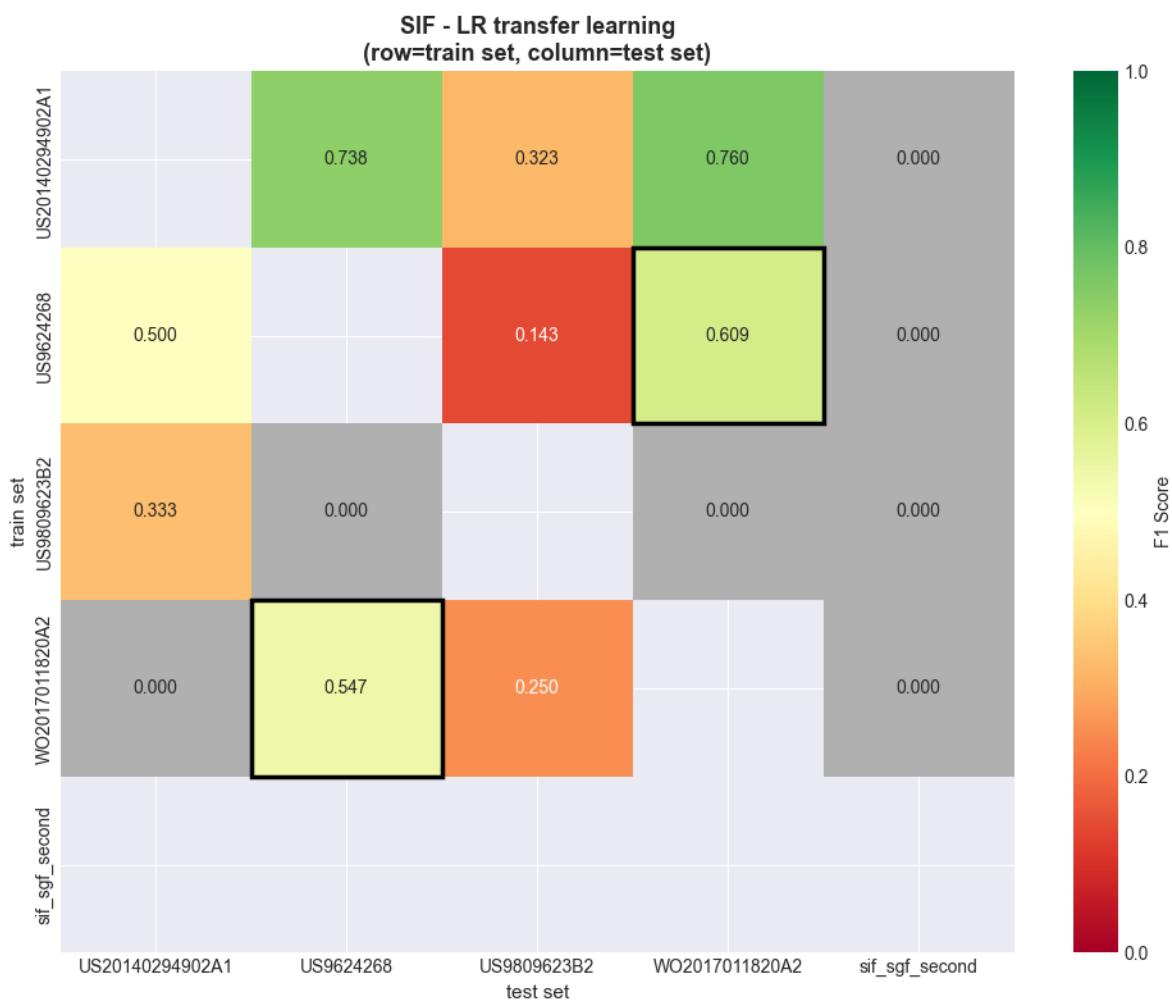
1.0





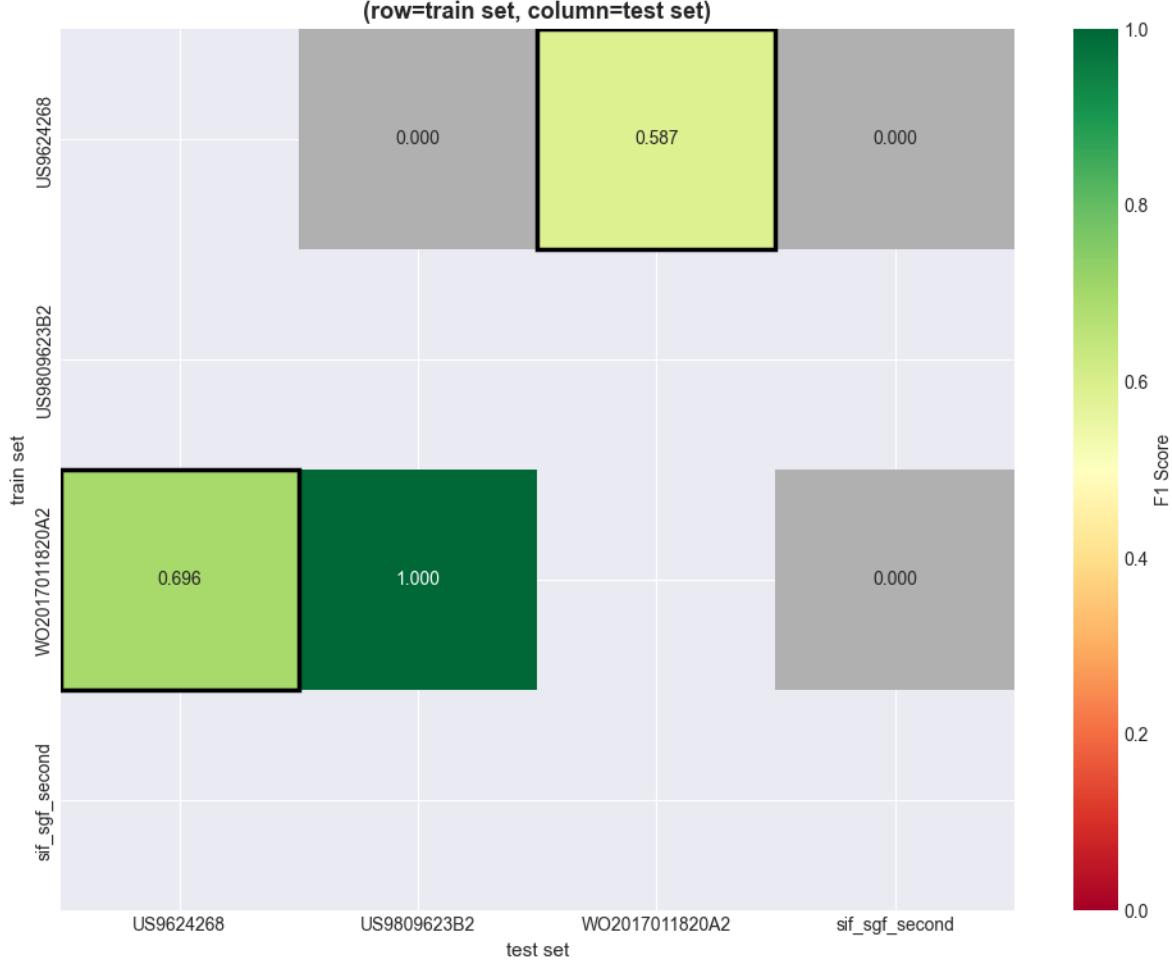
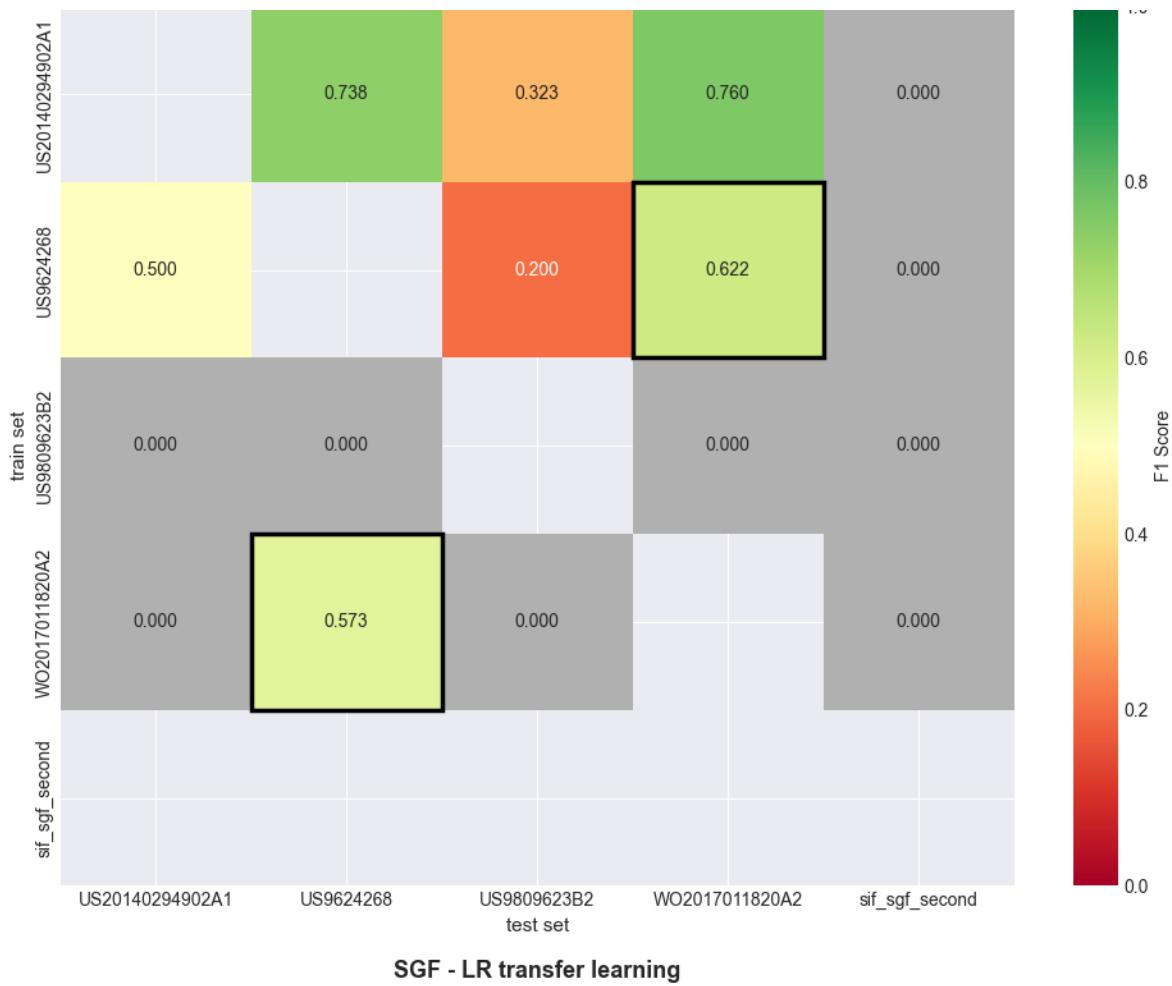
## 1.4. 使用Morgan分子指纹图片展示

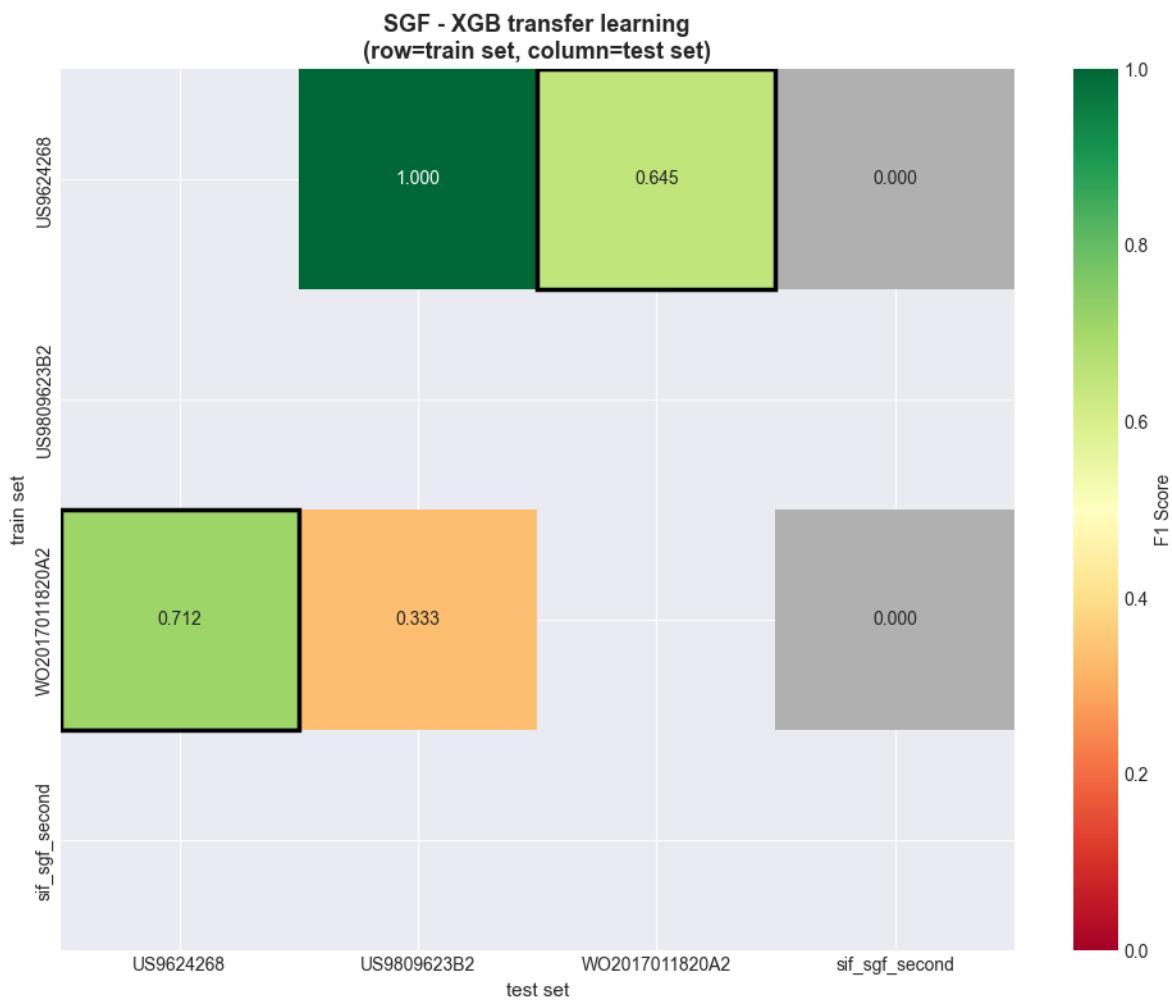
---



**SIF - XGB transfer learning**  
(row=train set, column=test set)

1 0





## 2. 根据热力图检查去掉分子指纹后性能变化

(热力图绘图的title有错误) 计算公式应为 (means: no-with, std: with-no)

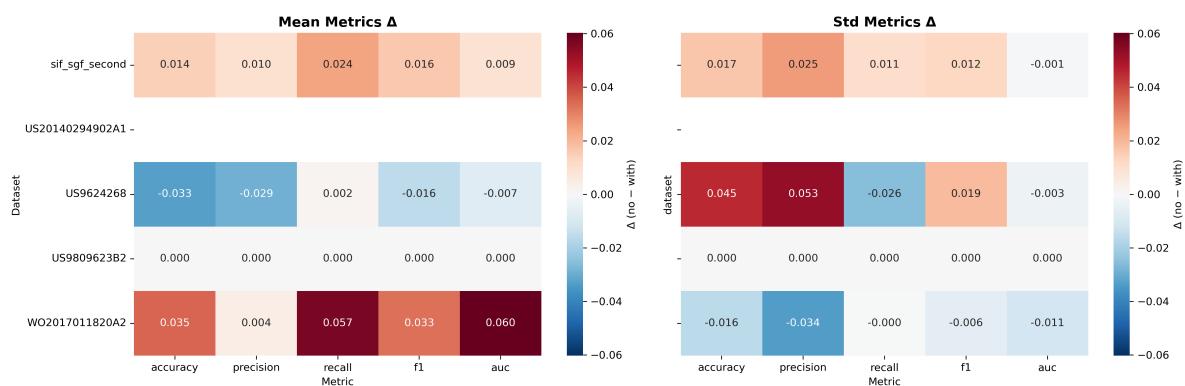
这样两边都是越红越好, 越蓝越差

### 2.1. 情况1: SGF-LR

US9624268: 正确率有所下降, 稳定性提升, 但不多 (其中AUC还下降了)

WO2017011820A2: 正确率大幅上升, 稳定性下降程度和US9624268上很接近

SGF - LR | Effect of With vs No

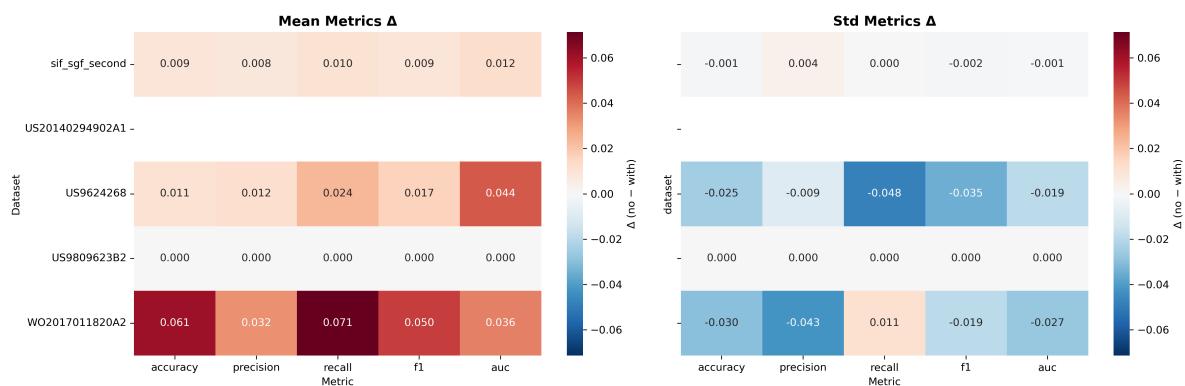


### 2.2. 情况2: SGF-RF

US9624268: 正确率提升, 稳定性下降

WO2017011820A2: 正确率大幅上升, 稳定性下降程度和US9624268上很接近

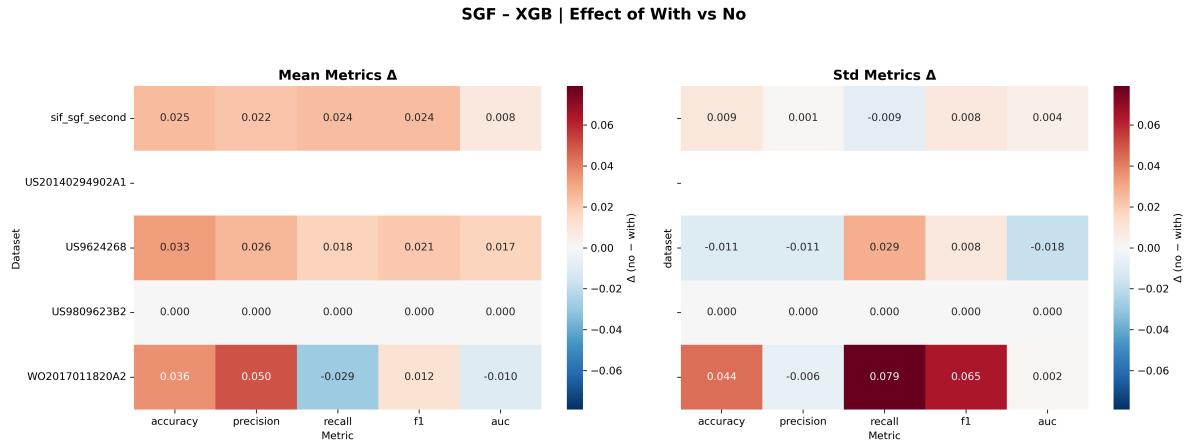
SGF - RF | Effect of With vs No



### 2.3. 情况3: SGF-XGB

US9624268: 正确率提升, 稳定性略微波动 (程度不大, std的指标在0.01左右波动)

WO2017011820A2: 正确率略微波动, 稳定性提升 (F1提升明显, 但是AUC几乎没变化)

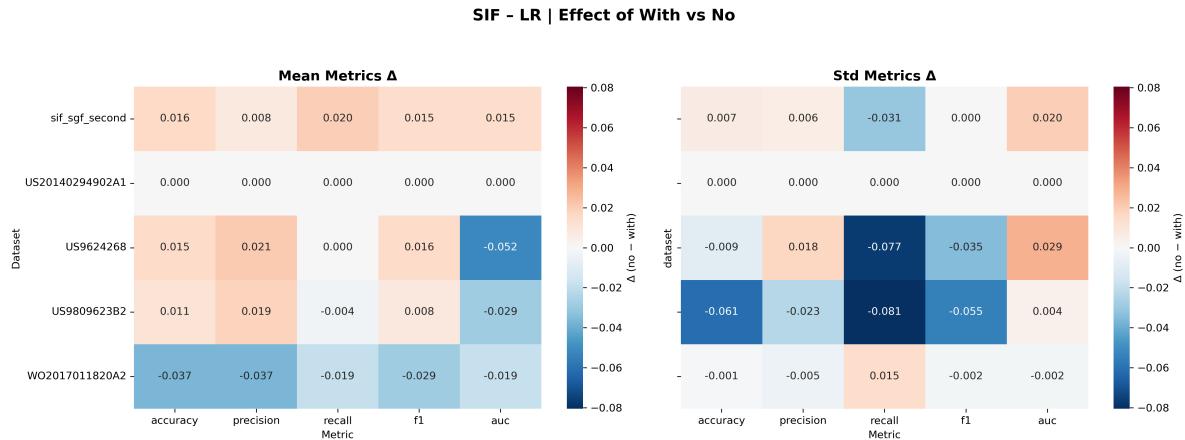


## 2.4. 情况4: SIF-LR

US9624268: 正确率略微提升, 稳定性下降

US9809623B2: 正确率略有上升, 稳定性下降明显 (std指标f1下降大约0.05, 但AUC提高了大约0.03)

WO2017011820A2: 正确率下降, 稳定性略微波动几乎不变

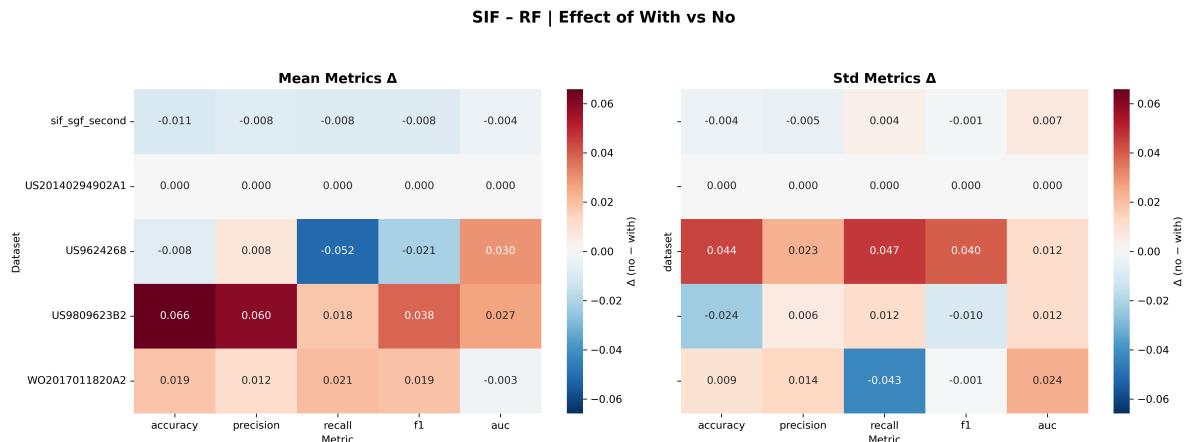


## 2.5. 情况5: SIF-RF

US9624268: 正确率下降, 稳定性明显提升

US9809623B2: 正确率明显提升, 稳定性波动但总体变化不大

WO2017011820A2: 正确率提升, 稳定性小幅波动 (std波动值大约在0.02)



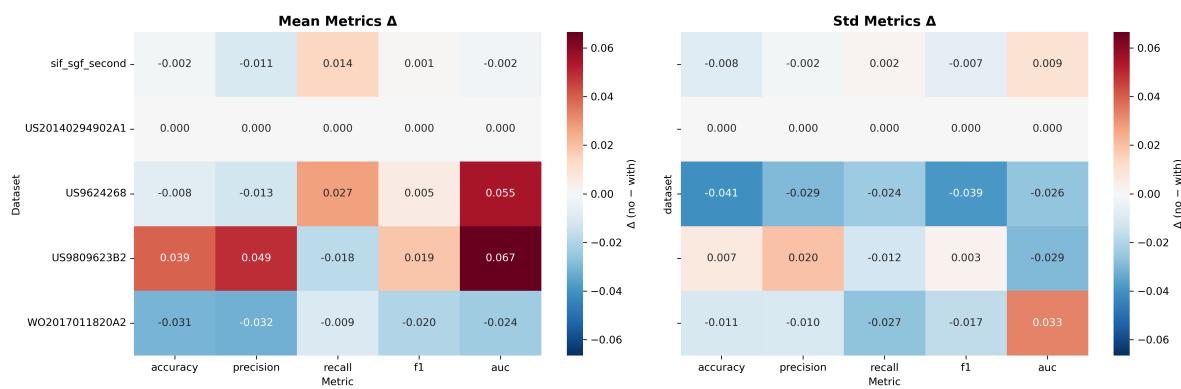
## 2.6. 情况6: SIF-RF

US9624268: 正确率小幅提升 (mean\_auc提升明显), 稳定性小幅度下降

US9809623B2: 正确率提升, 稳定性下降

WO2017011820A2: 正确率下降, 稳定性小幅下降

SIF - XGB | Effect of With vs No



## 2.7. 性能指标总结

- 整体趋势上，去掉Morgan带来的效果是正确率倾向于提升，模型稳定性趋向于下降
- 若是选择一种组合进行预测，SGF-RF最能体现出上述特点
- US9624268 & WO2017011820A2 数据集表现比较稳定，US9809623B2在不同任务中变化幅度比较大

## 3. 最终总结

### 3.0 口语化总结

- Morgan去掉后带来的整体趋势是迁移效果更好
- US98数据集可能也有比较好的迁移效果
- 整体趋势上，去掉Morgan带来的效果是正确率倾向于提升，模型稳定性趋向于下降
- 若是选择一种组合进行预测，SGF-RF最能体现出上述特点
- US9624268 & WO2017011820A2 数据集表现比较稳定，US9809623B2在不同任务中变化幅度比较大

### 3.1 迁移效果

关键发现	详细分析
去 Morgan 提升整体迁移效果	宏观趋势显示，移除 Morgan 分子指纹后，模型在多数迁移任务中的 F1 Score 普遍有所提升。
US98 数据集迁移潜力高	US9809623B2 上训练后迁移到另外两个可靠数据集，有比较好的迁移效果（但是出现了0.333和1.000的异常情况）。F1 Score 达到 1.000 表明 US98 数据集要么区分度极高，要么存在数据泄露风险。需要进一步验证，排除测试集的样本独立性问题

## 3.2 性能指标（去掉分子指纹后的影响）

指标	整体趋势
正确率（主要考虑F1/AUC）	倾向于提升或保持稳定
模型稳定性 (std)	普遍趋向于下降（稳定性降低）

## 3.3 最佳组合

- SGF-RF 组合：在去除 Morgan 后实现了正确率的最大幅度提升，但也带来了最明显的稳定性下降。

## 3.4 数据集内在特性总结

根据不同数据集在迁移任务中的表现，可总结其固有特性：

数据集	迁移表现特性	备注
US9624268 & WO2017011820A2	相互迁移性能稳定	无
US9809623B2	迁移性能好，但是在不同任务中性能变化幅度大，还有一场数值F1=0.333和F1=1.000。	迁移性好，但高度敏感