CTU
CZECH TECHNICAL UNIVERSITY IN PRAGUE

# Graph Neural Networks for Identification of Robust Biomarkers from Multi-Omics Data

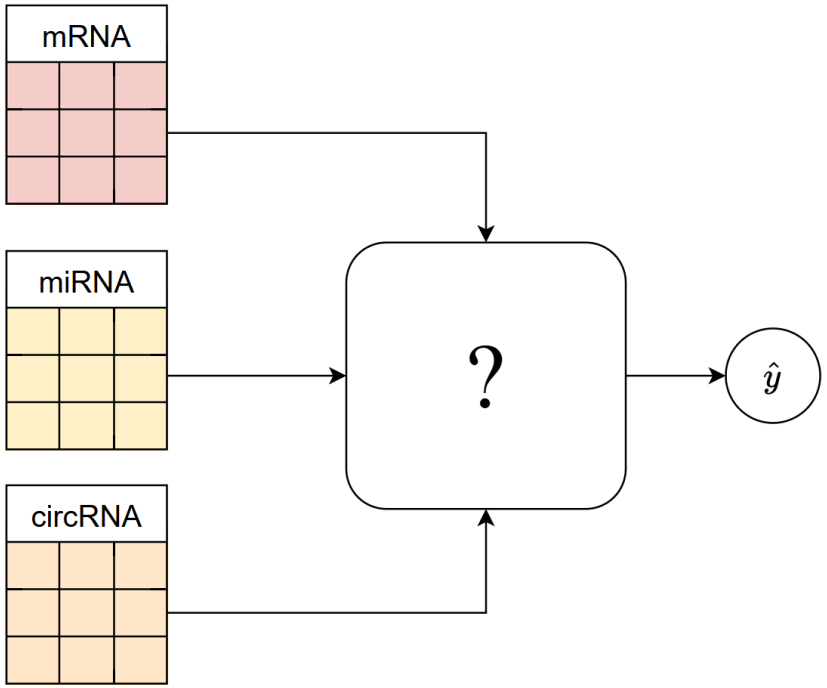
Bc. Jan Lubojacký

Supervisor: Ing. Jiří Kléma, Ph.D.
Project reviewer: Ing. Jan Drchal, Ph.D.

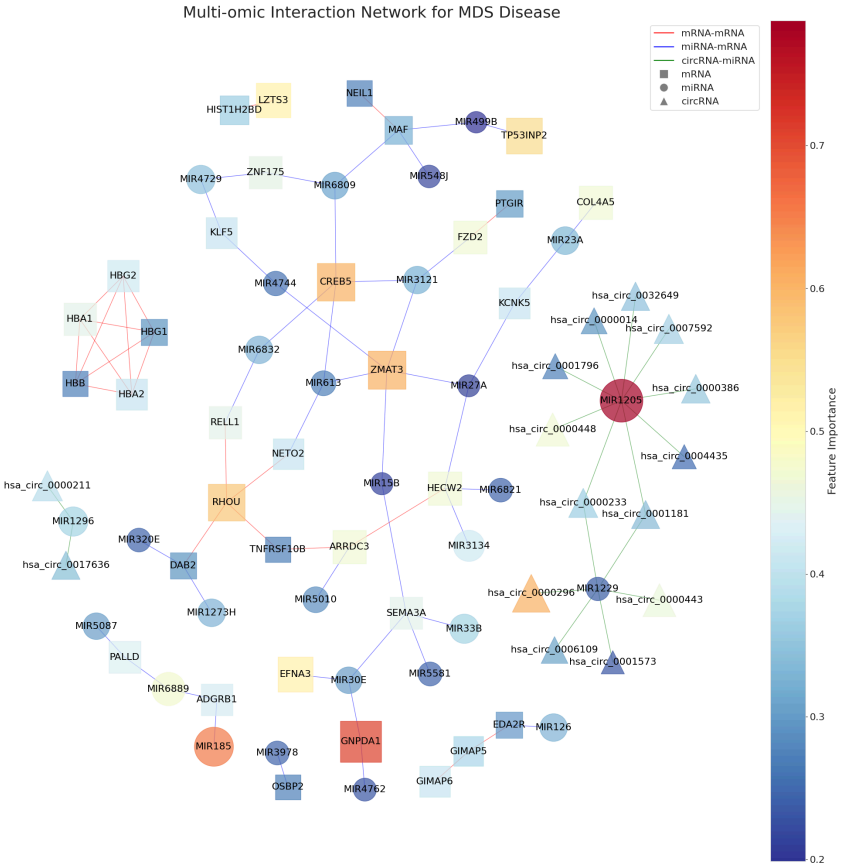
Medical Electronics and Bioinformatics
Faculty of Electrical Engineering
Department of Cybernetics

CTU
CZECH TECHNICAL UNIVERSITY IN PRAGUE

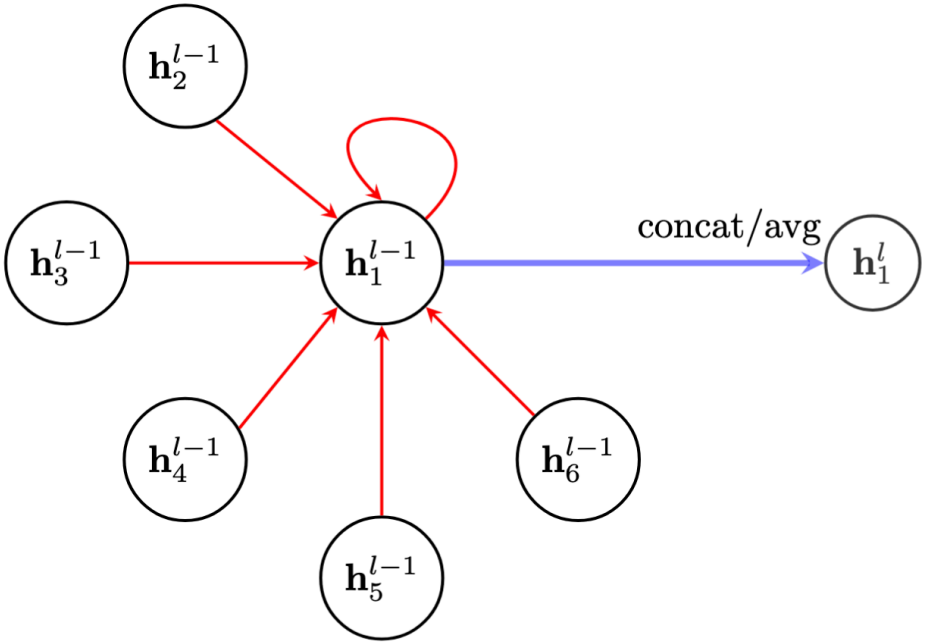
# Making predictions




# Extracting biomarkers



# Graph Neural Networks




$$\boldsymbol{h}_u^l = \mathrm{UPDATE}\left(\boldsymbol{h}_u^{l-1}, \mathrm{AGGREGATE}\left(\boldsymbol{h}_v^{l-1}, \forall v \in \mathcal{N}(u)\right)\right)$$

CTU
CZECH TECHNICAL UNIVERSITY IN PRAGUE

$$\boldsymbol{h}_u^l = \text{UPDATE}\left(\boldsymbol{h}_u^{l-1}, \text{AGGREGATE}\left(\boldsymbol{h}_v^{l-1}, \forall v \in \mathcal{N}(u)\right)\right)$$

- **Graph Convolutional Networks**

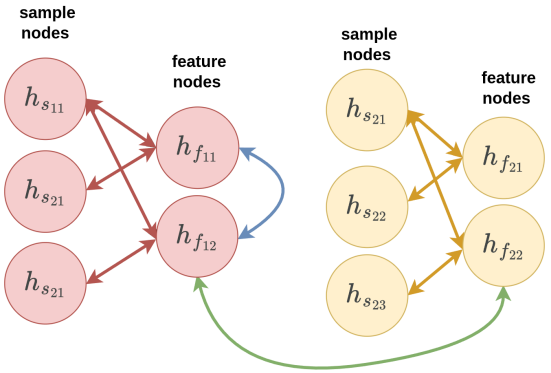
$$\boldsymbol{h}_u^l = \sigma\left(\boldsymbol{W}_{\text{self}}^l \boldsymbol{h}_u^{l-1} + \boldsymbol{W}_{\text{neigh}}^l \sum_{v \in \mathcal{N}(u)} \boldsymbol{h}_v^{l-1} + \boldsymbol{b}^l\right)$$

- **Graph Attention Networks**

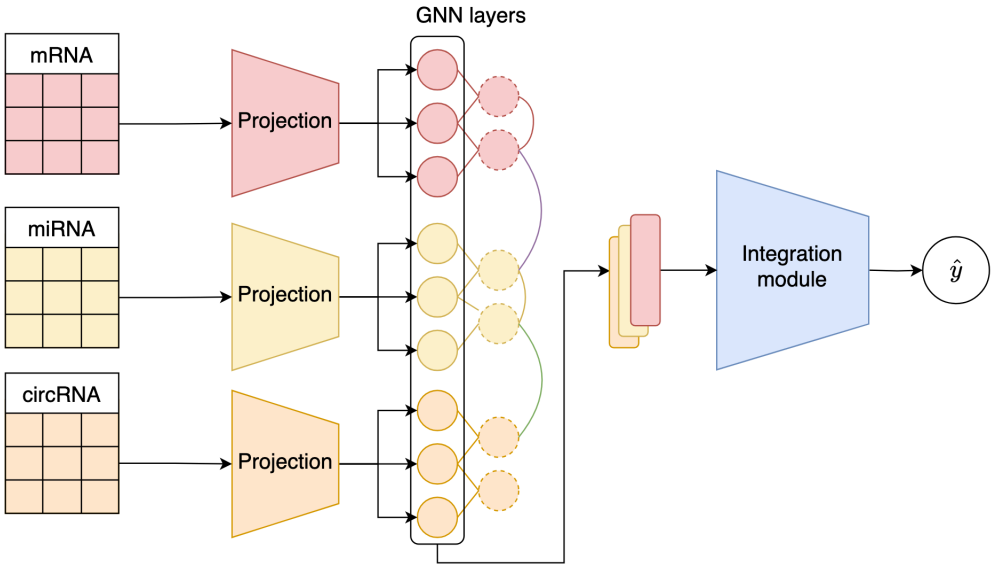
$$\boldsymbol{h}_u^l = \sigma\left(\sum_{v \in \mathcal{N}(u)} \alpha(v,u) \boldsymbol{W}^l \boldsymbol{h}_v^{l-1}\right)$$

CTU
CZECH TECHNICAL UNIVERSITY IN PRAGUE

# Bipartite GNN



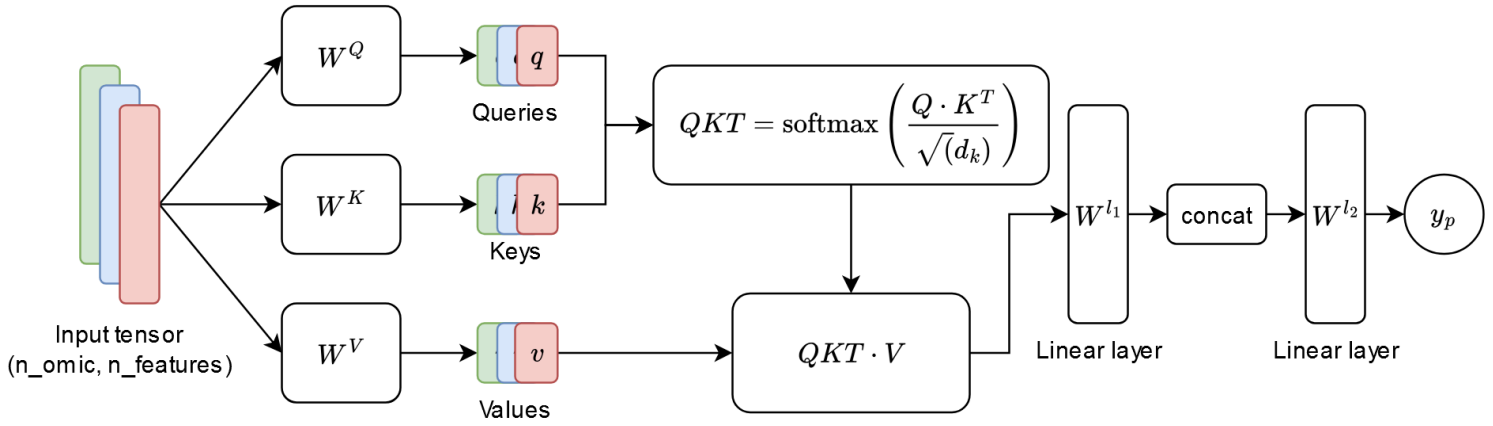

Input graph structure




Model architecture

## Integration modules

- **linear integrator**
  - $\hat{y} = \sigma\left(\boldsymbol{W} \cdot ||_{i=1}^{m}\ \boldsymbol{x}^{(i)} + \boldsymbol{b}\right)$
- **view correlation discovery network**
  - $C_{a_1,\ldots,a_m} = \prod_{i=1}^{m} x_{a_i}^{(i)};\ a = 1, \ldots, m;\ \boldsymbol{C} \in \mathbb{R}^{|x_1| \times \ldots \times |x_m|}$
  - $\hat{y} = \mathrm{VCDN}(\boldsymbol{C})$
- **attention integrator**



CTU
CZECH TECHNICAL UNIVERSITY IN PRAGUE

# Models

- KNN
- Linear SVMs
- XGBoost
- Linear NN
- MOGONET
- BipartiteGNN

CTU
CZECH TECHNICAL UNIVERSITY IN PRAGUE

## Models

- KNN
- Linear SVMs
- XGBoost
- Linear NN
- MOGONET
- BipartiteGNN

## Datasets

- TCGA-BRCA
  - 483 samples
  - 4 classes
  - mRNA, miRNA, CNV, DNA methylation
- MDS
  - 3 tasks
    - Disease (74 samples)
    - Risk (53 samples)
    - Mutation (26 samples)
  - 2 classes

CTU
CZECH TECHNICAL UNIVERSITY IN PRAGUE

# Models

- KNN
- Linear SVMs
- XGBoost
- Linear NN
- MOGONET
- BipartiteGNN

# Datasets

- TCGA-BRCA
  - 483 samples
  - 4 classes
  - mRNA, miRNA, CNV, DNA methylation
- MDS
  - 3 tasks
    - Disease (74 samples)
    - Risk (53 samples)
    - Mutation (26 samples)
  - 2 classes

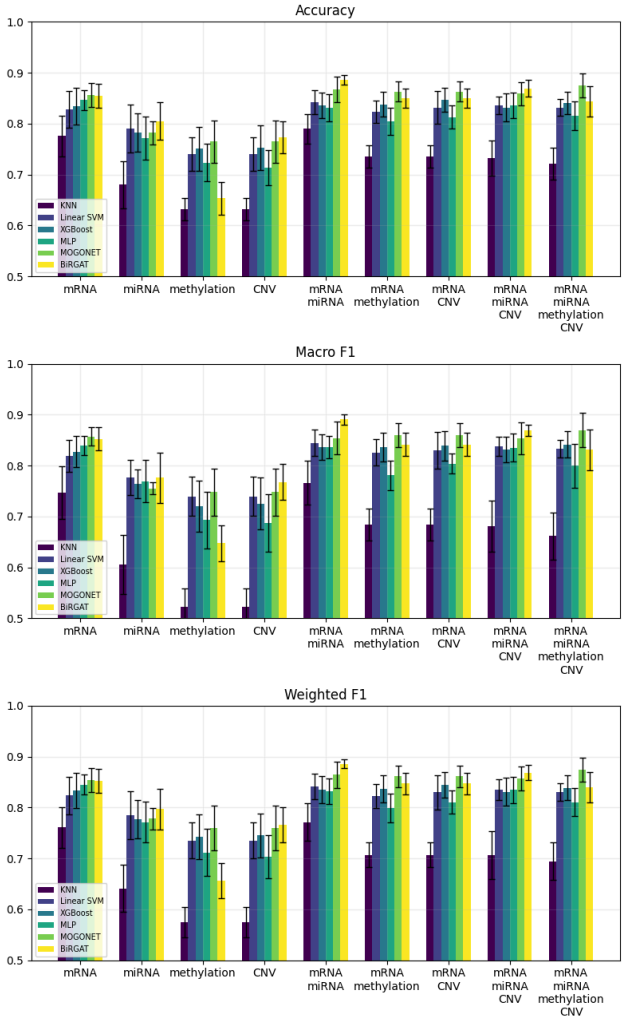
# Evaluation

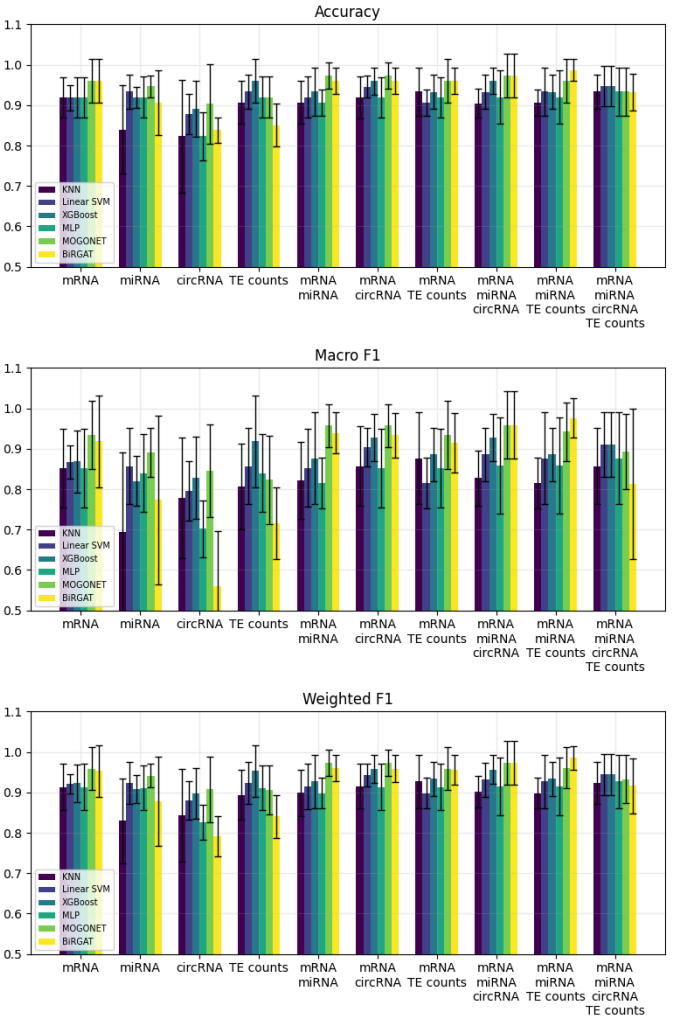
```
for each trial in 1 to N:

    1. sample hyperparameters

    for each CV split:
        2. split data
        3. feature selection
        4. normalize features
        5. fit model
        6. test model
        7. record performance P

    8. Average P across folds
    9. If P > P_best: P_best=P

return P_best
```

CTU
CZECH TECHNICAL UNIVERSITY IN PRAGUE

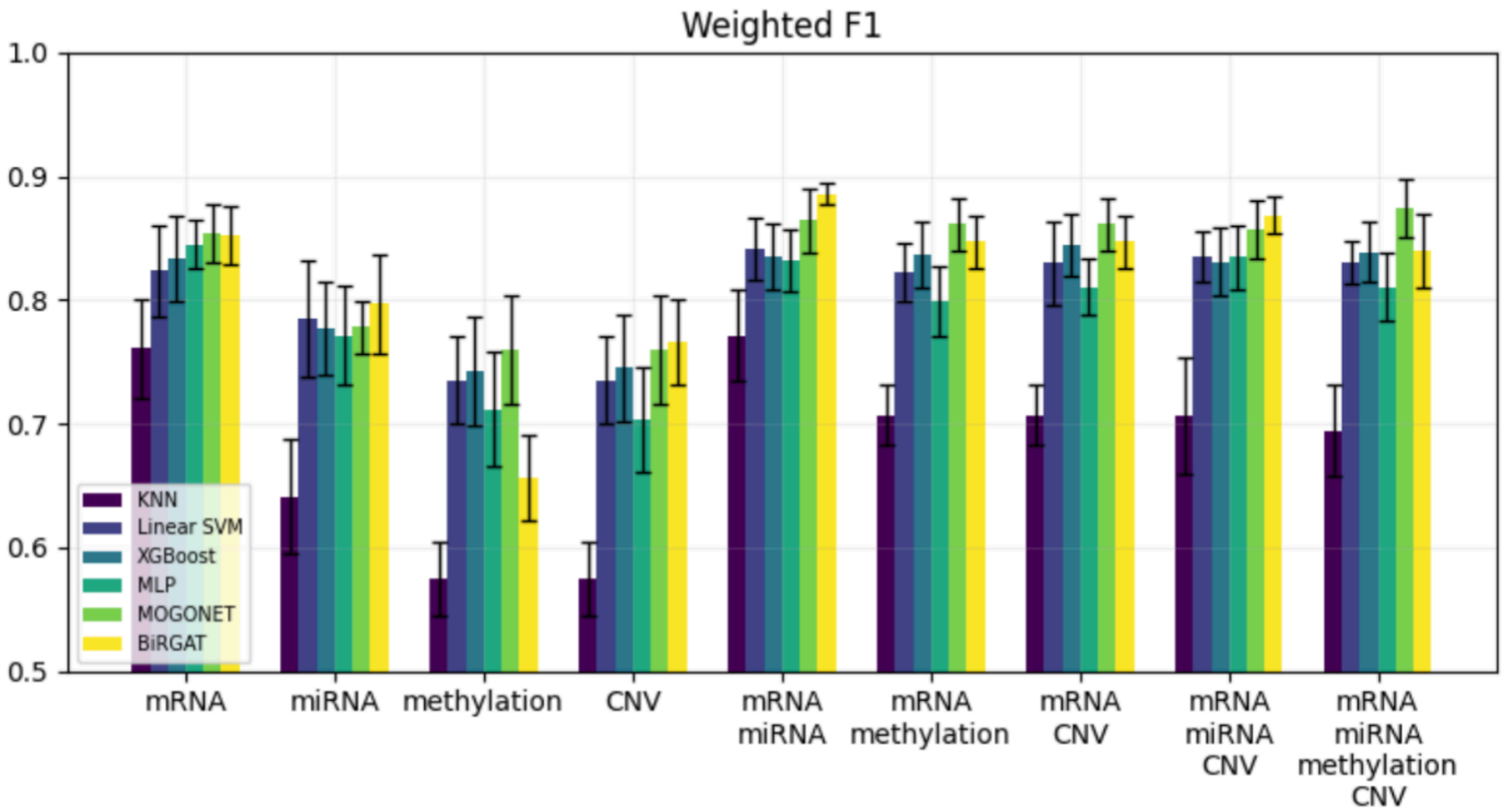
# Results

CTU
CZECH TECHNICAL UNIVERSITY IN PRAGUE




- **TCGA-BRCA results**
- **MDS disease results**

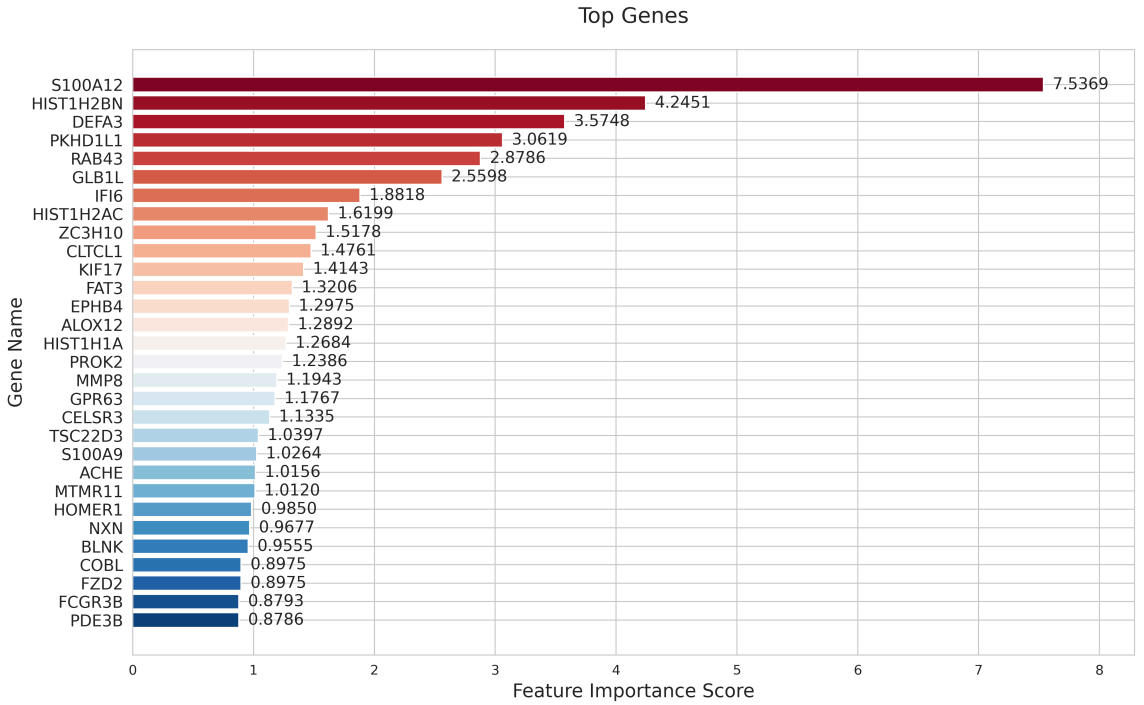



- **Model performances in terms of Weighted F1 score on the TCGA-BRCA dataset**

CTU
CZECH TECHNICAL UNIVERSITY IN PRAGUE

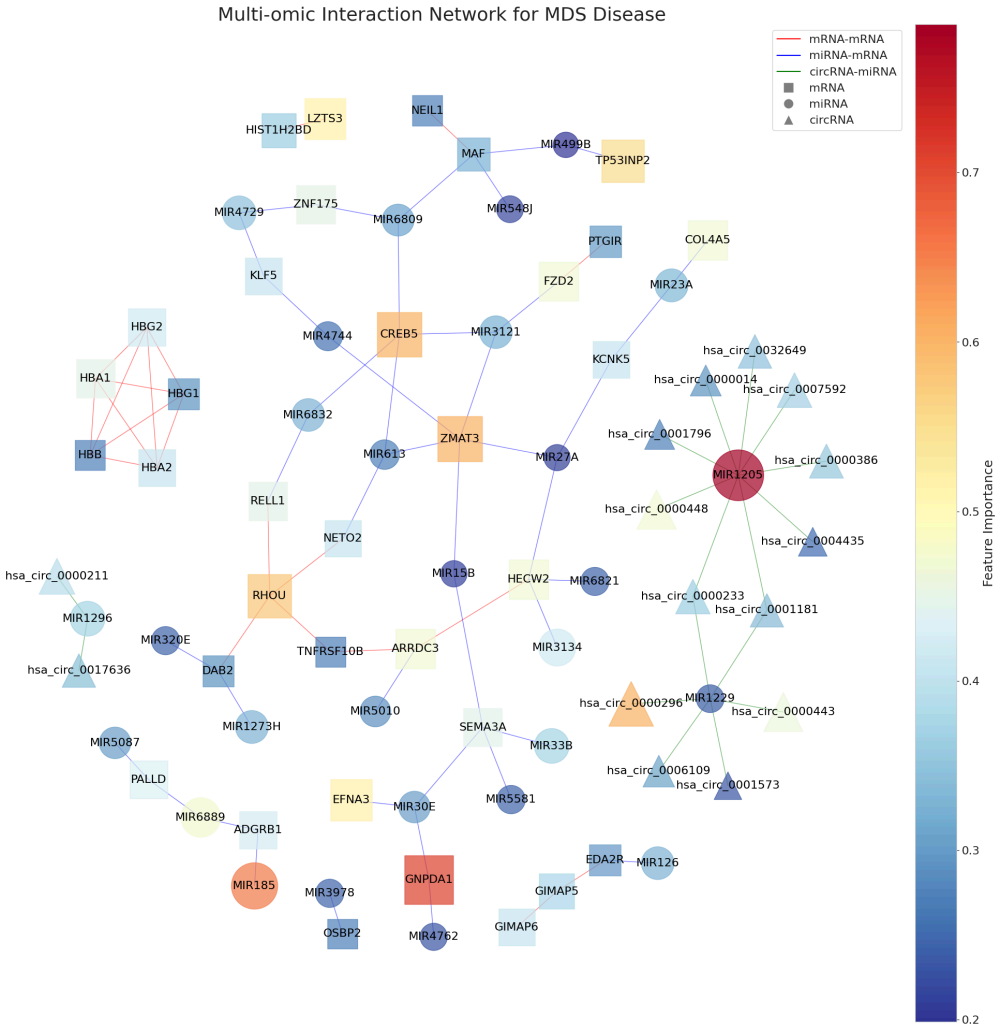
# Extracting Biomarkers

- XGBoost linear booster weights
- Feature permutation for GNNs
- Comparison against traditional methods




- feature importances (MDS risk, XGBoost)

CTU
CZECH TECHNICAL UNIVERSITY IN PRAGUE




- **structural biomarkers (MDS disease, Mogonet)**

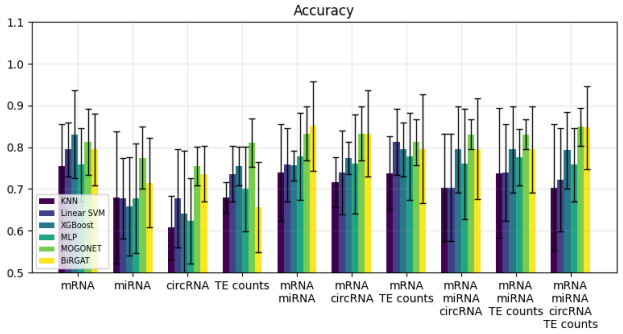

# Thank You
## for your attention

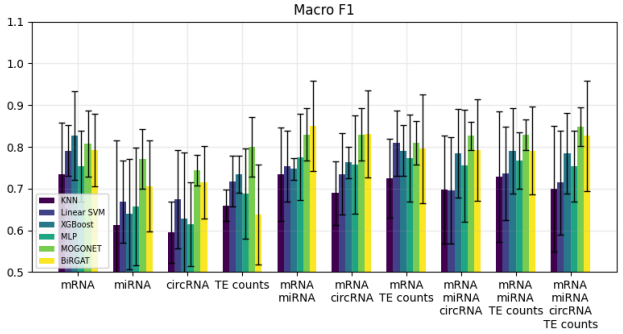
CTU
CZECH TECHNICAL UNIVERSITY IN PRAGUE

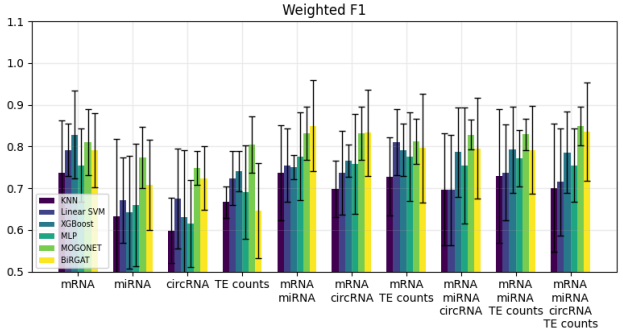
# Reviewer question

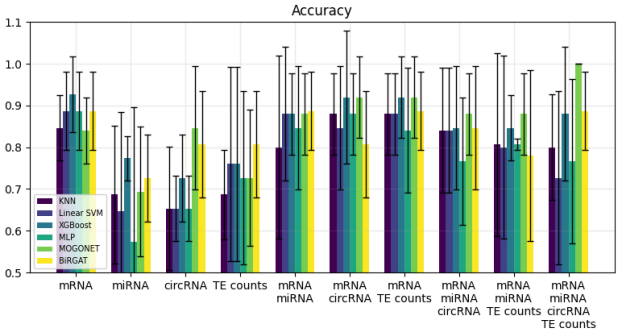
- U sloupcových grafů v obrázcích 8.1, 8.2 a 8.3 mi není zřejmý přesný význam černých čar posazených na vrchní část každého sloupce. Předpokládám, že se jedná o konfidenční intervaly. Zejména u obr. 8.3 jsou tyto intervaly značně široké a pro jednotlivé metody se výrazně překrývají. Můžete na základě grafů říci něco o statistické významnosti výsledků?

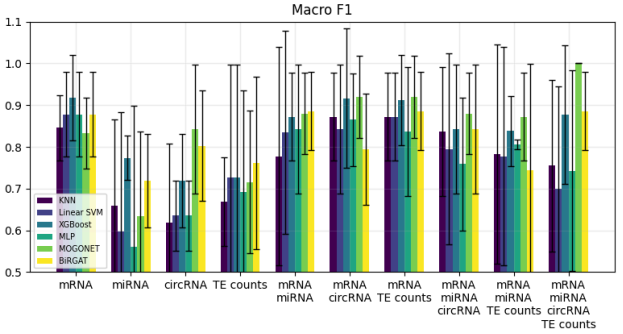
CTU
CZECH TECHNICAL UNIVERSITY IN PRAGUE

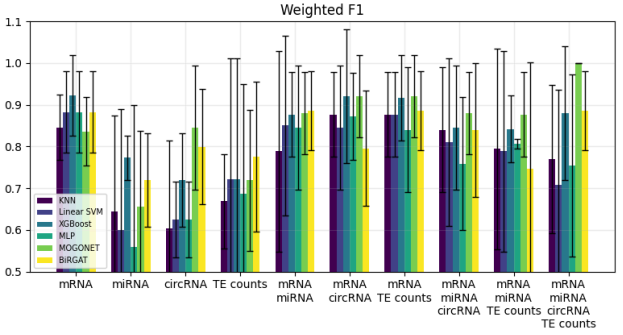



8.2

8.3