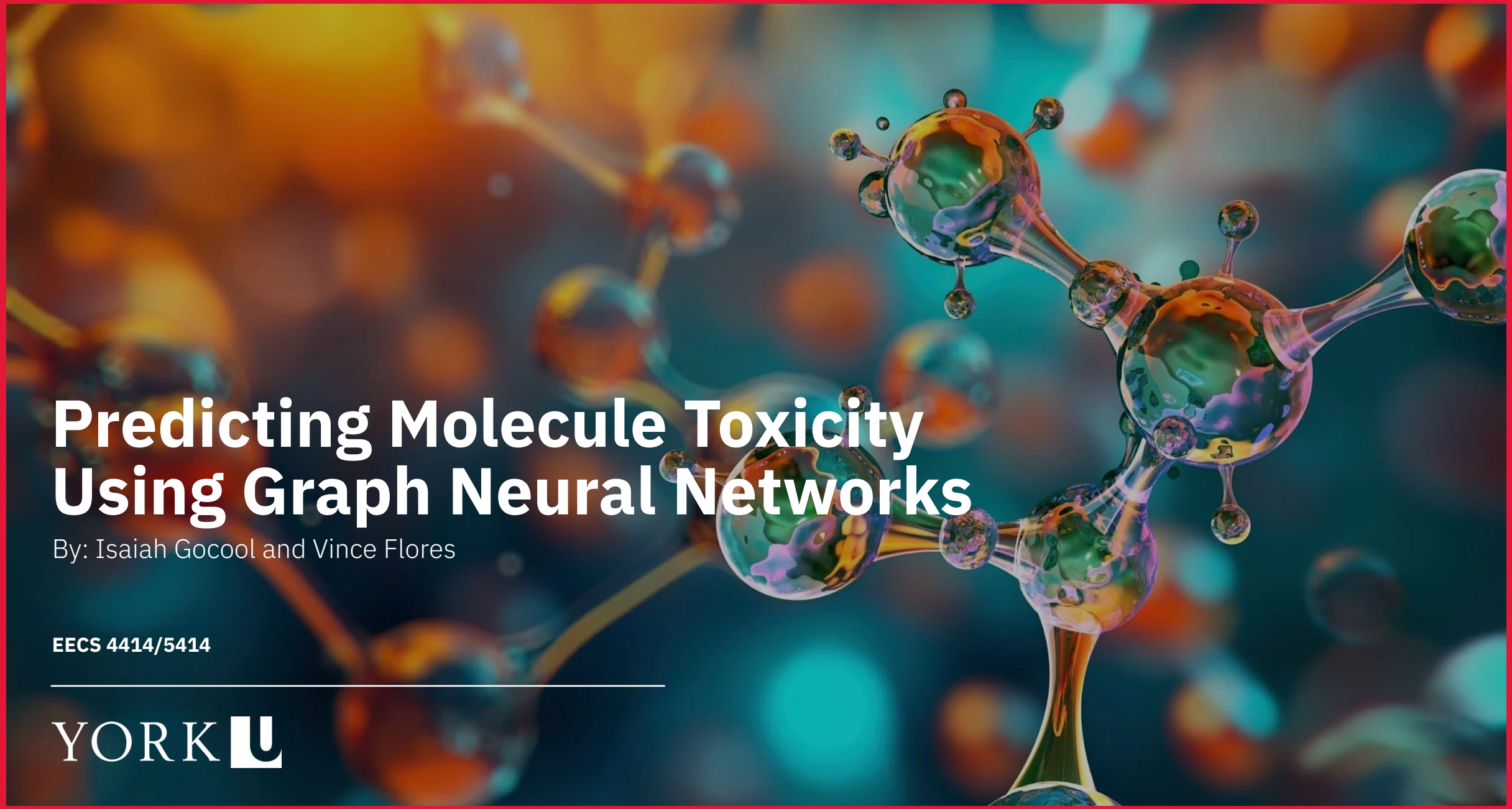


Predicting Molecule Toxicity Using Graph Neural Networks

By: Isaiah Gocool and Vince Flores

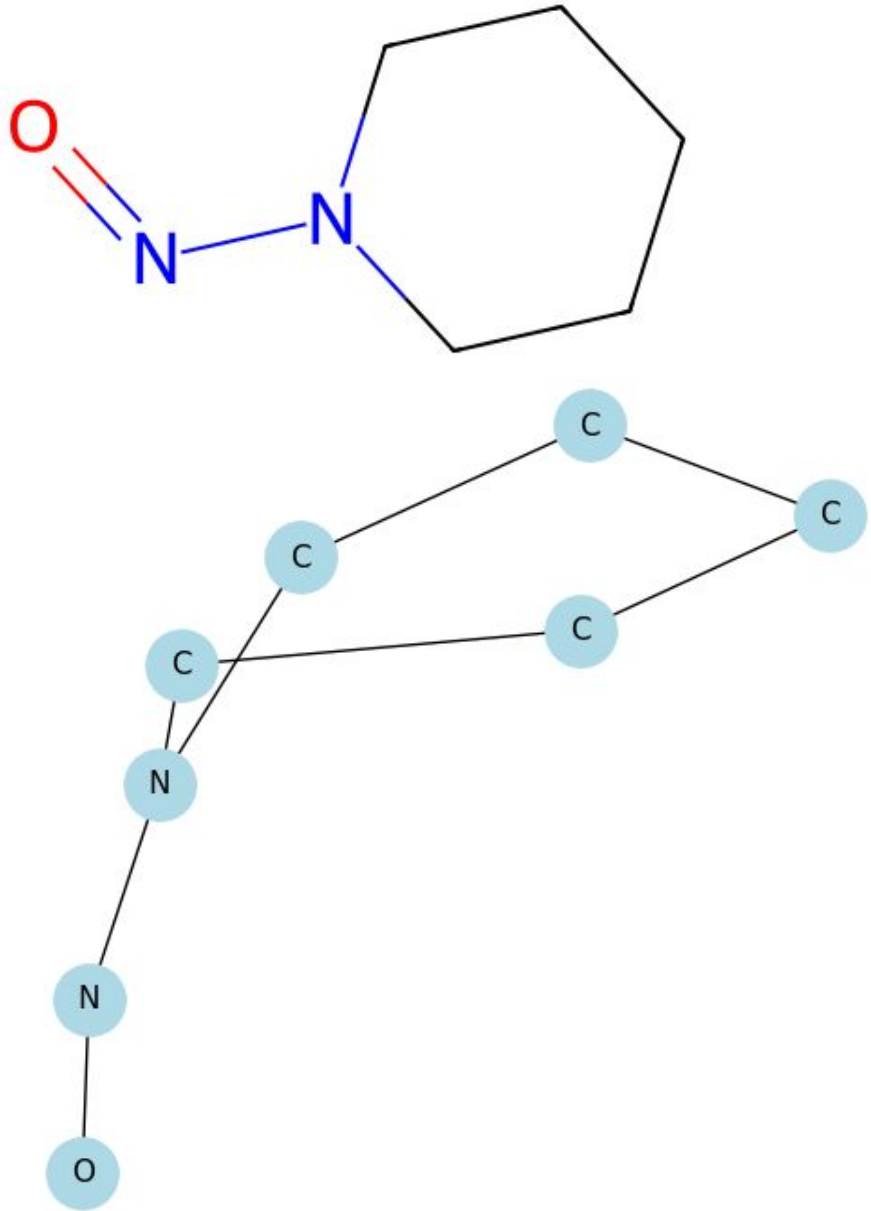
EECS 4414/5414

YORK U

Introduction

- Over $1 billion is spent annually on developing new pharmaceutical drugs
- Only 12% of drugs entering clinical trials are approved by the FDA
- Molecule toxicity is one of the leading problems affecting pharmaceutical drug approval
- Highlights the need for strong predictive models that detect molecule toxicity early in drug development

YORK U



Why Graph Neural Networks?

- Molecular structures can easily be modelled through the use of graphs
  - Nodes = Atoms
  - Edges = Bonds/bond types
- Neighbourhood aggregation used for graph classification
- GNNs can be used for multi-label classification tasks such as label prediction

YORK U

Problem Definition

Problem 1

Are GNNs better at predicting molecular properties than standard machine learning methods?

Problem 2

Which GNN is better for this task, Graph Convolutional Networks (GCNs), or Graph Attention Networks (GATs)?

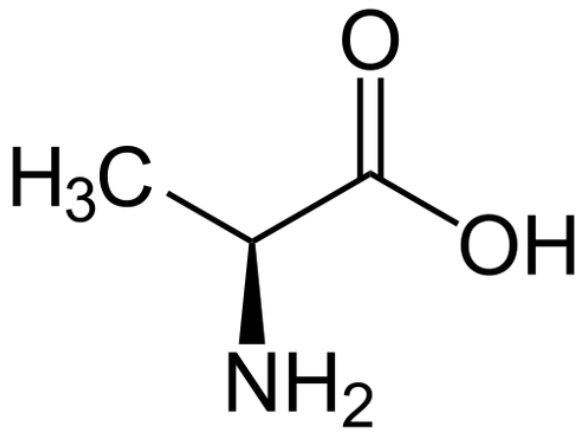
Problem 3

What are the limitations of GNNs when predicting molecular properties?

Dataset

- Tox21 Dataset:
- 7823 Molecules
- Imbalanced
- 12 Labels
  - Each label represents a toxicity target
  - Toxicity targets are biological effects that chemicals are tested for
  - Binary label represents if target has been activated

	NR-AR	NR-AR-LBD	NR-AhR	NR-Aromatase	NR-ER	NR-ER-LBD	NR-PPAR-gamma	SR-ARE	SR-ATAD5	SR-HSE	SR-MMP	SR-p53	mol_id	smiles
0	0.0	0.0	1.0	NaN	NaN	0.0	0.0	1.0	0.0	0.0	0.0	0.0	TOX3021	CCOc1ccc2nc(S(N)(=O)=O)sc2c1
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	NaN	0.0	NaN	0.0	0.0	TOX3020	CCN1C(=O)NC(c2ccccc2)C1=O
2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.0	NaN	0.0	NaN	NaN	TOX3024	CC[C@]1(O)CC[C@H]2[C@@H]3CCC4=CCCC[C@@H]4[C@H]...
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	NaN	0.0	NaN	0.0	0.0	TOX3027	CCCN(CC)C(CC)C(=O)Nc1c(C)cccc1C
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	TOX20800	CC(O)(P(=O)(O)O)P(=O)(O)O



N[C@H](C)C(=O)O

SMILES Notation

Class Labels

NR-AR, NR-AR-LBD, NR-AhR, NR-Aromatase, NR-ER, NR-ER-LBD, NR-PPAR-gamma

SR-ARE, SR-ATAD5, SR-HSE, SR-MMP, SR-p53

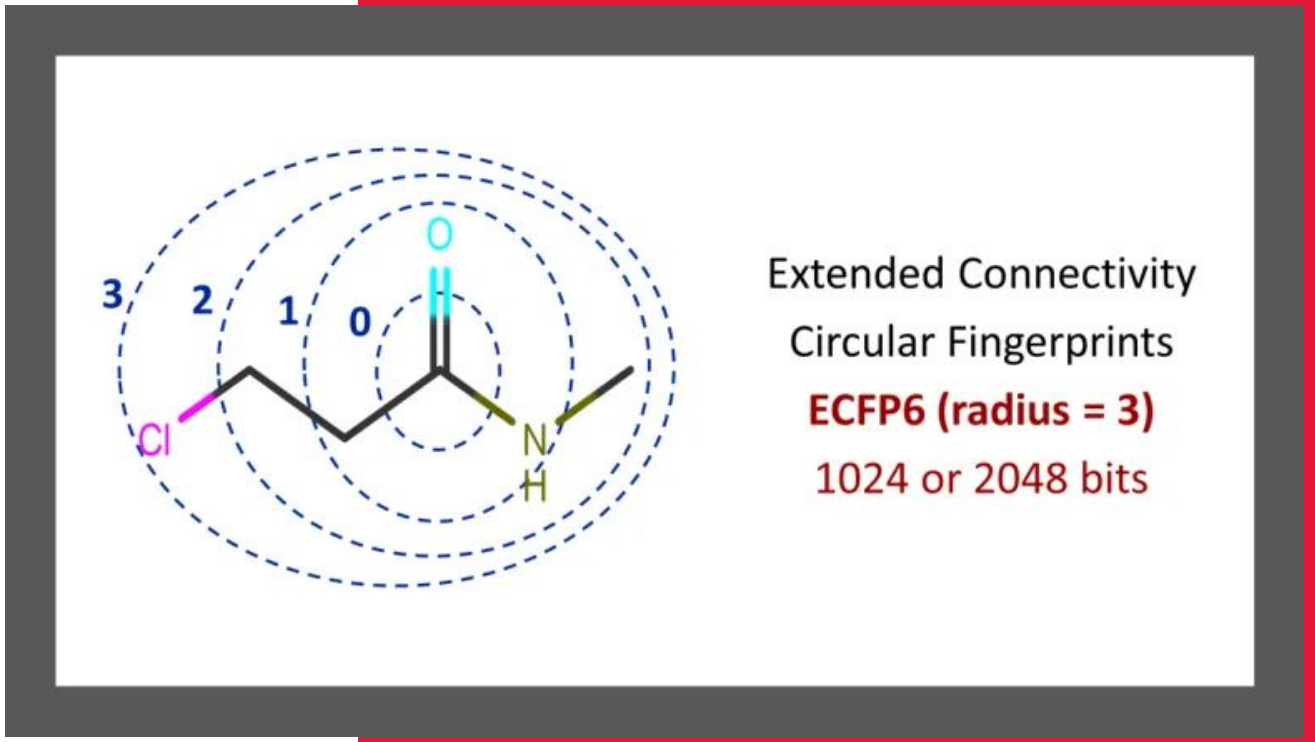
YORK U

Methodology

Extended Connectivity Fingerprint (ECFP)

Basic SMILES to ECFP FLow

1. For each atom collect info (atom type, bond, etc)
2. For each atom collect info from its neighbors
3. For each atom collect info from its neighbors of neighbors
4. Combine all and encode into 2048 bits



YORK U

Methodology

Baseline Machine Learning Model: Random Forest Classification Model

- 1000 trees
- Bootstrap sampling (each tree learns from a random sample)
- weight class set to “balanced”
- Input: ECFP
- Output 12 labels

Graph Convolutional Networks (GCNs)

$$h'_v = \mathrm{ReLU}\left(W \cdot \frac{1}{|\mathcal{N}(v)|} \sum_{u \in \mathcal{N}(v)} h_u \right)$$

- A deep learning model that can be used on graphs
- Nodes are updated using the average properties of its neighbours
- Does not take into account which neighbours are more important

h'_v = The updated/predicted node v

h_u = The feature vector of neighbour node u

$\mathcal{N}(v)$ = All neighbours of node v

W = Learnable weight matrix

ReLU = Rectified Linear Unit

Graph Attention Networks (GATs)

$$h'_v = \mathrm{ReLU}\left(\sum_{u \in \mathcal{N}(v)} \alpha_{vu} W h_u\right)$$

- A deep learning model that can be used on graphs
- Nodes are updated using the average properties of its neighbours
- Focuses on the most important neighbours through the use of an attention coefficient
- Allows us to identify important substructures

h'_v = The updated/predicted node v

h_u = The feature vector of neighbour node u

α_{vu} = Attention coefficient of node u to node v

W = Learnable weight matrix

ReLU = Rectified Linear Unit

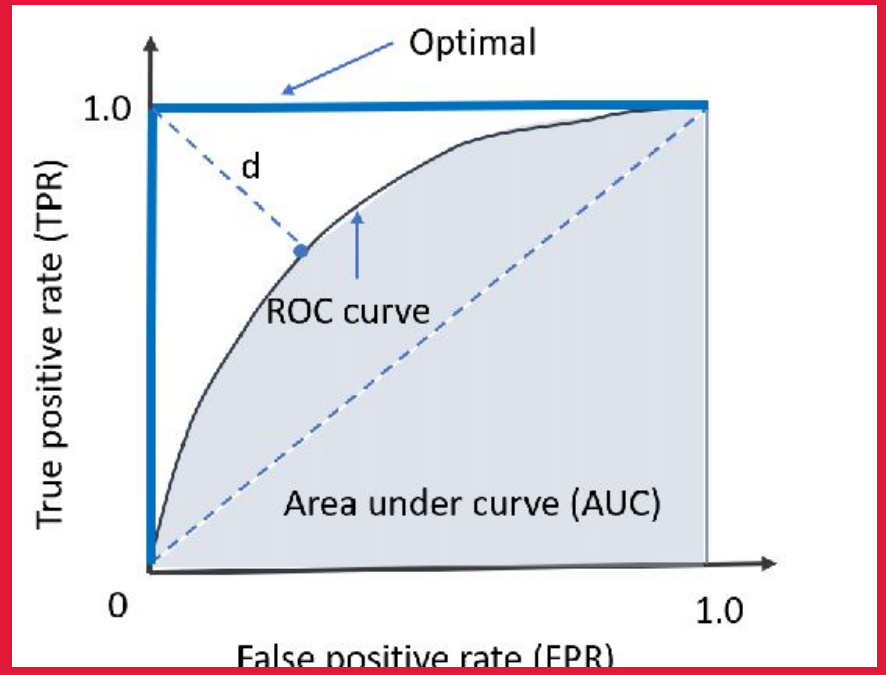
Optimization Metrics

AUC-ROC:

- AUC-ROC summarizes an ROC curve by computing the area under it
- AUC-ROC is between 0 and 1, closer to 1 means the model was more accurate
- Used to calculate the accuracy of each toxicity target

Accuracy:

- Will be used to check for overall toxicity
- Measured by calculating the percentage of correct predictions on the testing dataset



YORK U

Classification Metrics

Metric	Description
AUC ROC	Area under the receiver operating characteristic curve
Accuracy	number of molecules classified as toxic correctly over the total number of molecules in the testing dataset
Precision	percentage of the molecules classified as toxic that are actually toxic
Recall	percentage of molecules that are toxic that are correctly labeled as toxic
F1-score	harmonic mean of precision and recall

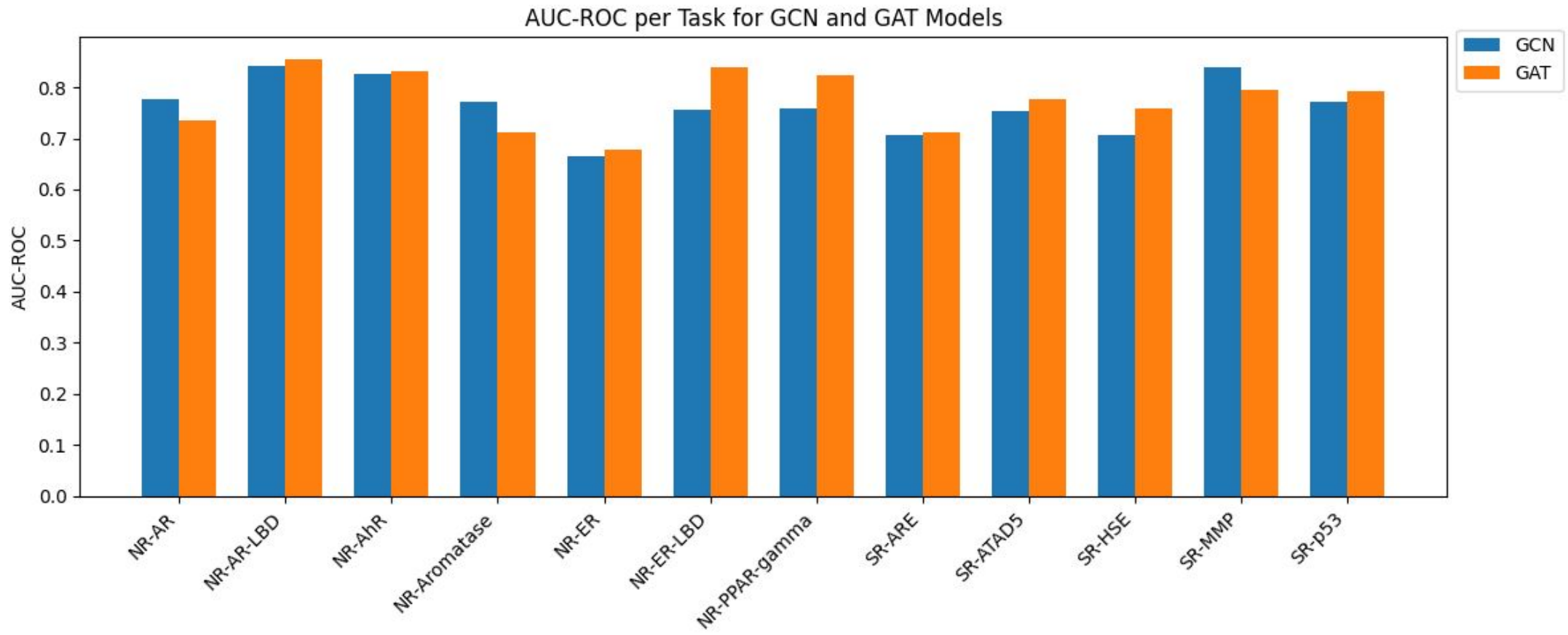
Precision = True Toxic / (True Toxic + False Toxic)

Recall= True Toxic / (True Toxic + False Non-Toxic)

Experiments & Results: GCN vs. GAT

Model	Validation Loss	Test Loss	Accuracy
GCN	0.2109	0.2300	0.7565
GAT	0.2097	0.2252	0.7622

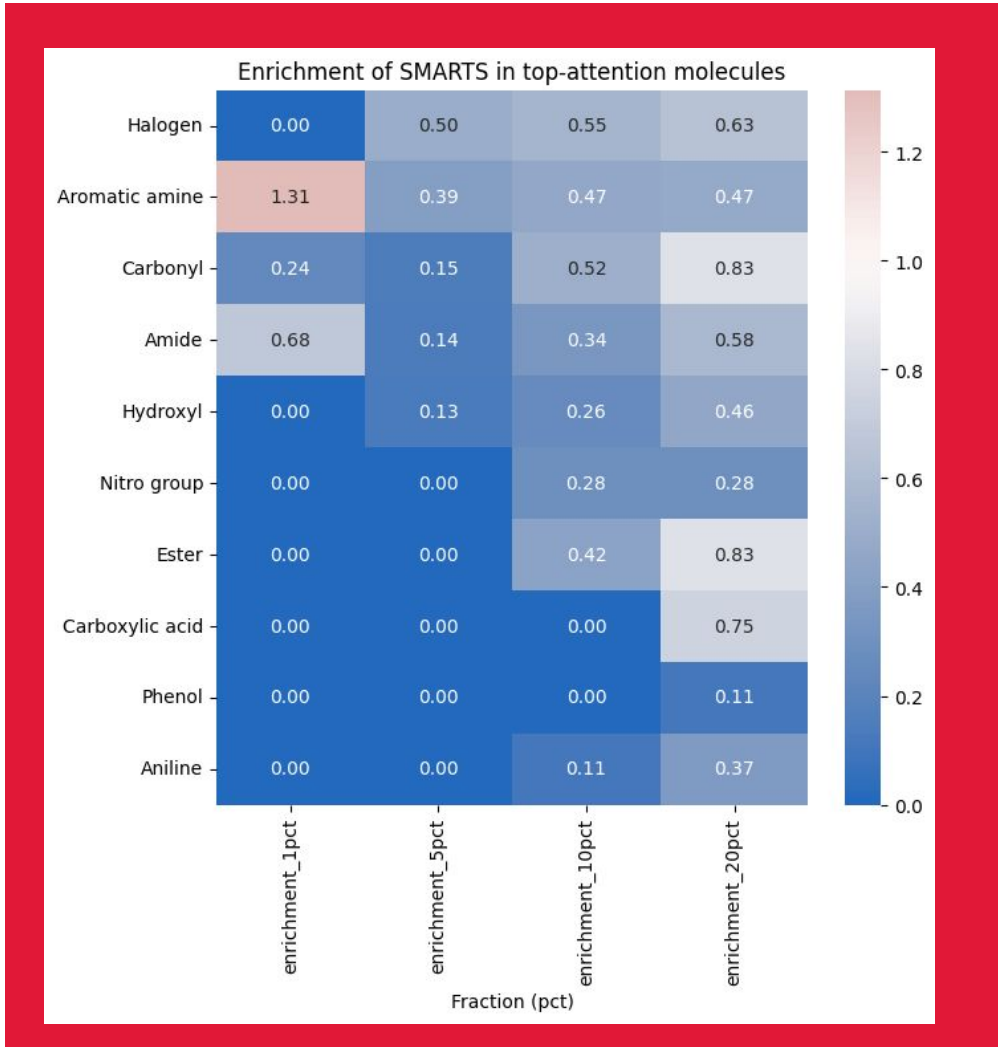
Experiments & Results: GCN vs. GAT



YORK U

Enrichment Heatmap

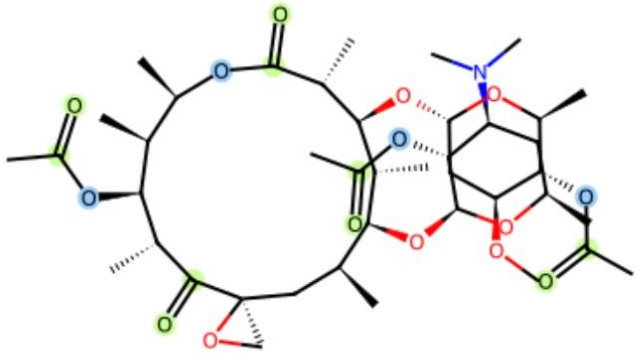
- A list of molecules known as toxicophores were used as SMARTS for this heatmap
- Toxicophores are chemical substructures that are known to contribute to toxicity
- Fraction represents top x% of top-attention molecules
- Enrichment scores > 1.0 means the model pays extra attention to molecules containing the substructure



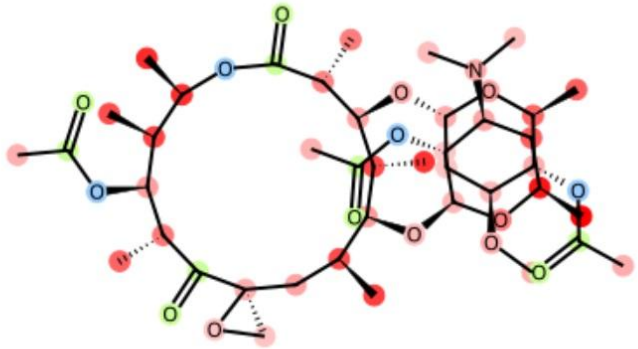
Molecules With Disagreeing Predictions

Index: 621
SMILES: CO[C@H]1C[C@H](O[C@@H]2[C@@H](C)C(=O)O[C@H](C)[
[C@H](C)C[C@H](N(C)C)[C@H]3OC(C)=O)[C@H]2C)O[C@@H](C)[
GCN_pred: 0.214
GAT_pred: 0.033
abs_diff: 0.181
Correct model: GAT
Tasks disagreeing: ['NR-AR', 'NR-AR-LBD']

GAT Attention:

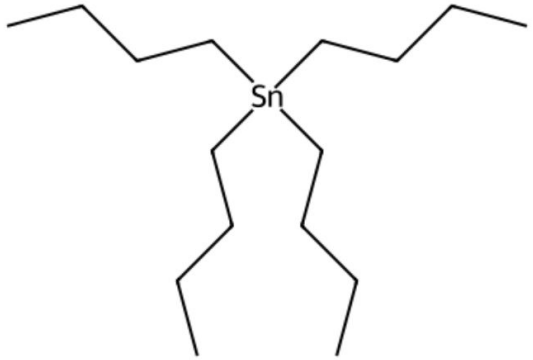


GCN Saliency:

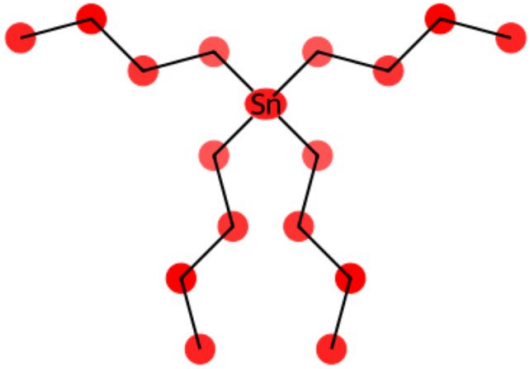


Index: 1
SMILES: CCCC[Sn](CCCC)(CCCC)CCCC
GCN_pred: 0.123
GAT_pred: 0.440
abs_diff: 0.346
Correct model: GCN
Tasks disagreeing: ['NR-AR-LBD', 'NR-ER-LBD', 'NR-PPAR-gamma', 'SR-ARE', 'SR-ATAD5', 'SR-HSE', 'SR-MMP', 'SR-p53']

GAT Attention:



GCN Saliency:



YORK U

Experiments & Results: GNN VS Random Forest

Average From All 12 Labels

Model	AUC ROC	Accuracy	Precision	Recall	F1-score
GCN	0.732384	0.756492	0.195589	0.516800	0.257348
GAT	0.751418	0.762239	0.189231	0.533661	0.260866
RFC	0.791187	0.620113	0.657729	0.250197	0.334193

AUC ROC: GAT > GCN, GNN ≈ RFC

Accuracy : GNN > RFC

Precision: RFC > GCN > GAT

Recall: GAT > GCN > RFC

F1-score: RFC > GAT > GCN

Precision = True Toxic / (True Toxic + False Toxic)

Recall= True Toxic / (True Toxic + False Non-Toxic)

Conclusions and Next Steps

- GNNs are better at predicting toxicity of a molecule than RandomForest
- GATs achieve a better overall accuracy than GCNs, and achieve a higher AUC-ROC in general
- There are some toxicity targets where GCNs achieve a higher AUC-ROC score than GATs
  - NR-AR
  - NR-Aromatase
  - SR-MMP
- The ideal GNN model would use GCNs to predict tasks where it scored a higher AUC-ROC, and then GATs for the rest of the tasks
- Further fine-tuning on hyperparameters should be performed to achieve a better performance
- Testing should also be performed on datasets like ClinTox or ToxCast
- Better handle Imbalance in the dataset

Image Sources

https://static.vecteezy.com/system/resources/previews/023/041/928/non_2x/hydrogen-molecule-or-atom-abstract-structure-for-science-or-medical-background-clear-blue-water-concept-of-chemical-model-connections-atoms-3d-rendering-generat-ai-free-photo.jpg

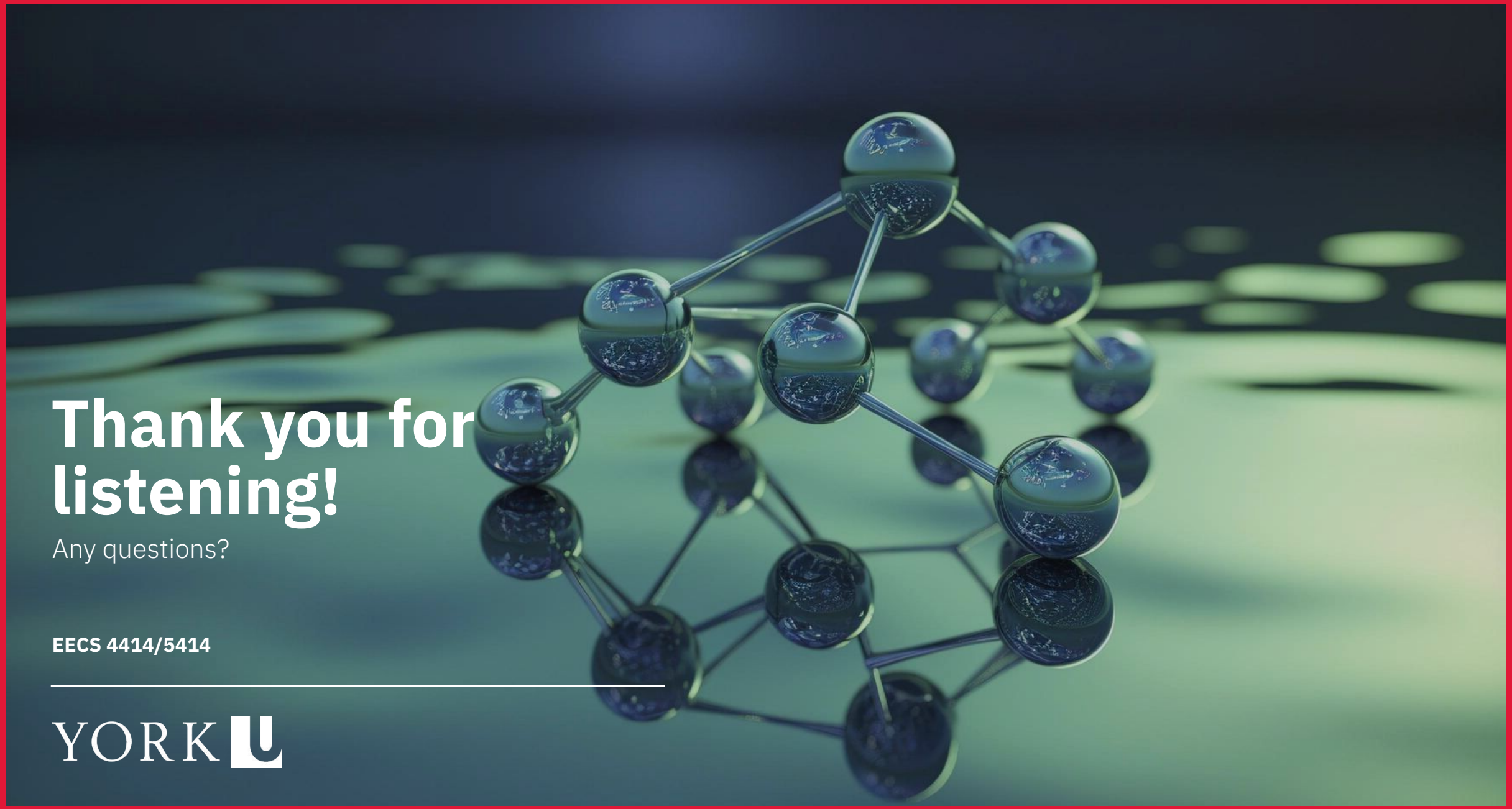
https://search.brave.com/images?q=Alanine&context=W3sic3JjIjoiaHR0cHM6Ly91cGxvYWQud2lraW1lZGlhLm9yZy93aWtpcGVkaWEvY29tbW9ucy90aHVtYi85LzkwL0wtQWxhbmluXy1fTC1BbGFuaW5lLnN2Zy81MTJweC1MLUFsYW5pbl8tX0wtQWxhbmluZS5zdmcucG5nIiwidGV4dCI6IkFsYW5pbmUgaW4gbm9uLWlvbmljIGZvcm0iLCJwYWdlX3VybCI6Imh0dHBzOi8vZW4ud2lraXBlZGlhLm9yZy93aWtpL0FsYW5pbmUifV0%3D&sig=52962ada3b8c65eca52010a14984e2de1bb20c0b5a9363d0d7840024bfb0a9ca&nonce=90161c11278f26a339926c58982c6ad4&source=infoboxImg

https://drzinph.com/ecfp6-fingerprints-in-python-part-3/

https://www.freepik.com/photos/medical-store

https://github.com/nikhilgawai/ROC_AUC_Curve

https://rakemgroup.co.uk/wp-content/compressx-nextgen/uploads/2025/01/shutterstock_2499664001-scaled.jpg.webp


Thank you for listening!
Any questions?
EECS 4414/5414
YORK U