EXPLAINABILITY OF GNN DECISIONS APPLICATION TO CHEMOINFORMATICS

Mariana Brito Azevedo[1], Jean-Luc Lamotte[1], Luc Brun[1], Pierre Héroux[2]

[1]Université Caen Normandie, ENSICAEN, CNRS, Normandie Univ, GREYC UMR 6072, 14000 Caen, France

[2]Université Caen Normandie, UNIROUEN, UNIHAVRE, INSA Rouen, LITIS, 76000 Rouen, France

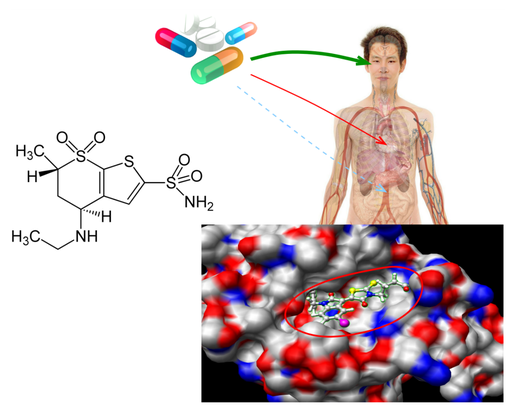
mariana.brito-azevedo@unicaen.fr, jean-luc.lamotte@unicaen.fr, luc.brun@unicaen.fr, pierre.heroux@univ-rouen.fr

In collaboration with *Centre d'Études et de Recherche sur le Médicament de Normandie (CERMN)*

GREYC
Electronics and Computer Science Laboratory

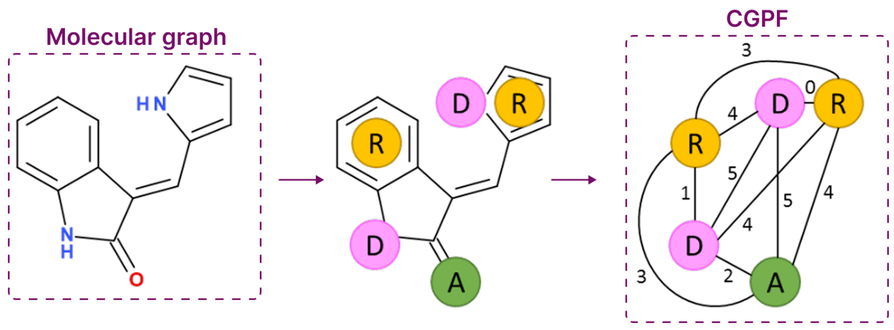
1. Motivation

- The drug development process is **costly**, **long** and needs to go through **several phases** for final approval by health authorities;
- The use Graph Neural Networks (**GNN**) and **explainability techniques** to facilitate this process.



2. Molecules as graphs

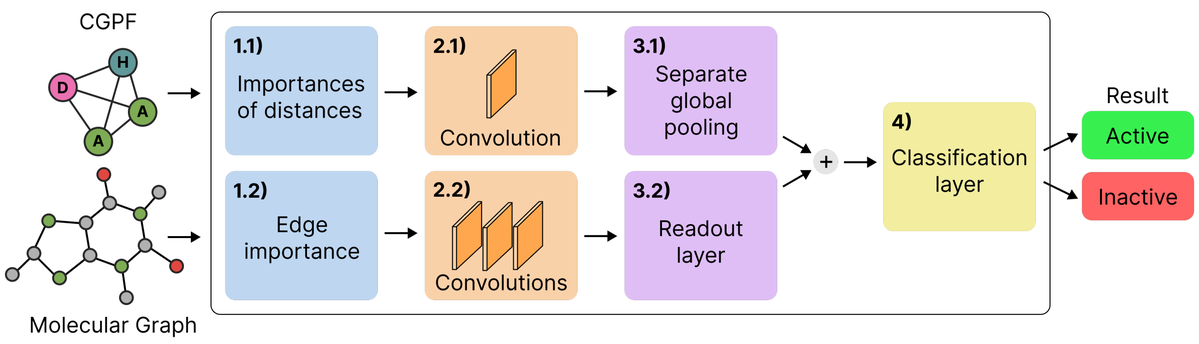
- **Molecular Graph :** atoms and chemical bonds;
- **Complete Graph of Pharmacophoric Features (CGPF)** [1] **:** pharmacophoric features and topological distances.



3. GNNs for activity classification

MCP-GNN [2]

- Molecular and Complete Pharmacophoric Features Graph Neural Network;
- GNN model for classifying **active** and **inactive** molecules;

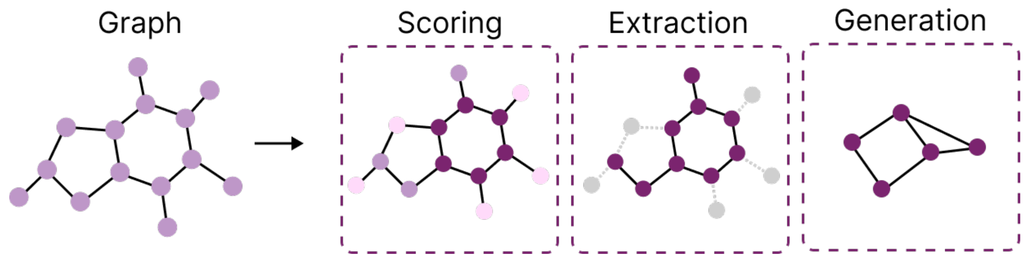


4. Explainability

Main goal

To understand **why** a molecule is active on a biological target.

Types of explanation [3]



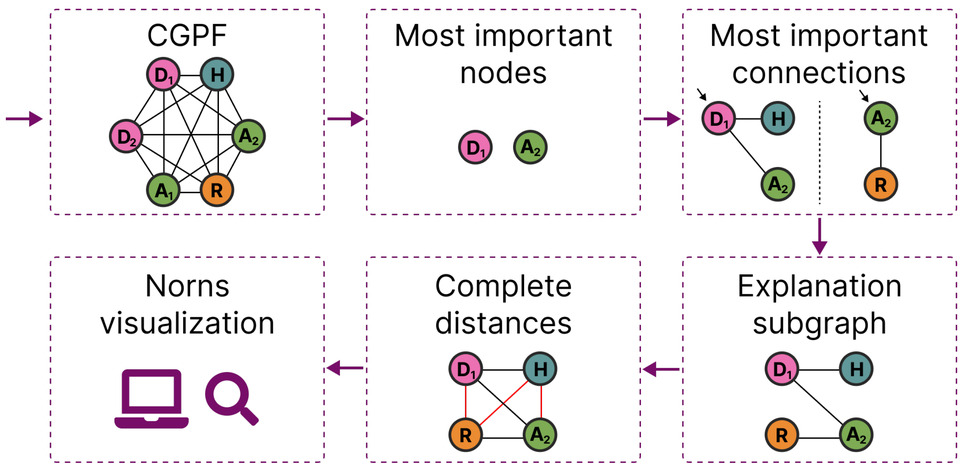
5. Model's performance

Table 1 : Results with BCR-ABL [4] and kinase datasets

Dataset	Model	ACC	AUC	MCC	MEAN F1
AKT1	RG-MPNN	0.875±0.016	0.935±0.012	0.748±0.032	0.873±0.016
3967 mol.	AttentiveFP	0.877±0.010	0.941±0.006	0.753±0.021	0.872±0.018
2175/1792	MCP-GNN	**0.892±0.012**	**0.949±0.009**	**0.783±0.023**	**0.890±0.012**
AURKA	RG-MPNN	0.829±0.030	0.889±0.034	0.644±0.063	0.821±0.034
3886 mol.	AttentiveFP	0.846±0.016	0.916±0.013	0.681±0.033	0.840±0.017
2290/1596	MCP-GNN	**0.860±0.016**	**0.928±0.014**	**0.710±0.034**	**0.854±0.017**
BCR-ABL	RG-MPNN	0.868±0.025	0.936±0.015	0.739±0.049	0.868±0.028
1479 mol.	AttentiveFP	0.858±0.052	0.926±0.028	0.721± 0.104	0.877±0.027
773/706	MCP-GNN	**0.883±0.024**	**0.949±0.017**	**0.761±0.050**	**0.878±0.025**
BRAF	RG-MPNN	0.895±0.020	0.934±0.033	0.709±0.060	0.877±0.023
4824 mol.	AttentiveFP	0.922±0.013	0.955±0.014	0.783±0.038	0.894±0.024
3629/1132	MCP-GNN	**0.932±0.011**	**0.960±0.013**	**0.808±0.033**	**0.901±0.016**
CK1	RG-MPNN	**0.832±0.049**	0.780±0.064	0.397±0.190	0.626±0.120
807 mol.	AttentiveFP	0.827±0.039	0.833±0.055	0.439±0.111	0.699±0.055
155/632	MCP-GNN	0.825±0.036	**0.852±0.029**	**0.517±0.062**	**0.718±0.044**
EGFR	RG-MPNN	0.854±0.014	0.923±0.013	0.709±0.028	0.854±0.015
8788 mol.	AttentiveFP	0.860±0.014	0.938±0.009	0.721±0.028	0.860±0.014
4513/4275	MCP-GNN	**0.874±0.011**	**0.946±0.006**	**0.749±0.023**	**0.874±0.011**
PIM1	RG-MPNN	0.881±0.018	0.942± 0.012	0.708±0.042	0.869±0.017
4507 mol.	AttentiveFP	0.908±0.014	0.956±0.011	0.763±0.039	0.883±0.019
3321/1186	MCP-GNN	**0.919±0.012**	**0.970±0.009**	**0.786±0.031**	**0.892±0.016**

6. Model's explanations

- First method developed to find an **explanation subgraph** for CGPF data to explain **active molecules**.

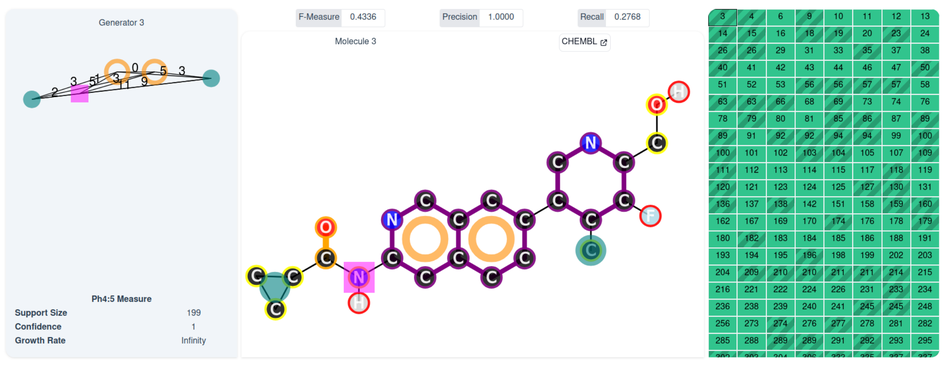


Finding most important nodes :

$$Importance_i =< v_i, V^{active}_{T(i)} - V^{inactive}_{T(i)} >$$

Finding most important connections :

$$Contribution_{i,j} = e_{i,j} < W_2 X_{T(j)}, V^{active}_{T(i)} - V^{inactive}_{T(i)} >$$



1. The pharmacophore net- work : A computational method for exploring structure–activity relationships from a large chemical data set (2018)
2. Graph Neural Network Based on Molecular and Pharmacophoric Features for Drug Design Applications (2025)
3. Adapted from : Graph-Based Explainable AI : A Comprehensive Survey (2024)
4. Data obtnained from *Centre d'Études et de Recherche sur le Médicament de Normandie (CERMN)*

Symposium MaDICS, May 2025, Toulouse, France.