



Deep GONet: Self-explainable deep neural network based on Gene Ontology for phenotype prediction from gene expression data

Victoria Bourgeais¹, Farida Zehraoui¹, Mohamed Ben Hamdoun¹, et Blaise Hanczar¹

CAp'2021 – Session Interprétabilité (16 Juin 2021)

¹ Université Paris-Saclay, Univ Evry, IBISC

Médecine de précision

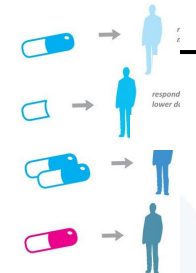


APPRENTISSAGE
AUTOMATIQUE



Prédiction de phénotypes

- Diagnostic
- Pronostic
- Réponse à un traitement
- ...



Médecine de précision



Données
d'expression de gènes

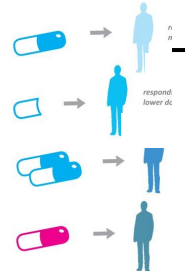


APPRENTISSAGE
AUTOMATIQUE

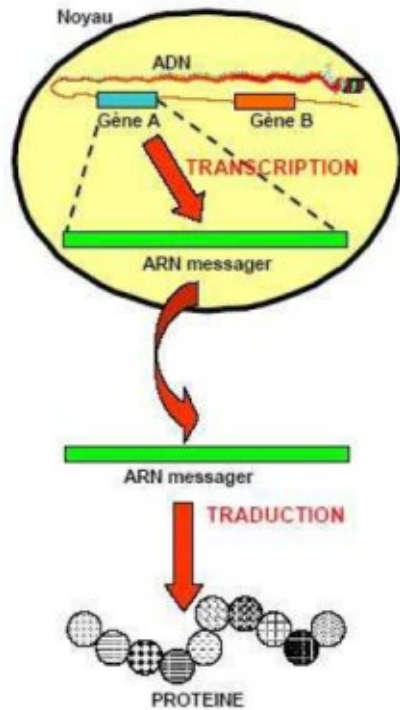


Prédiction de phénotypes

- Diagnostic
- Pronostic
- Réponse à un traitement
- ...

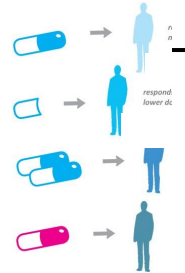


Médecine de précision



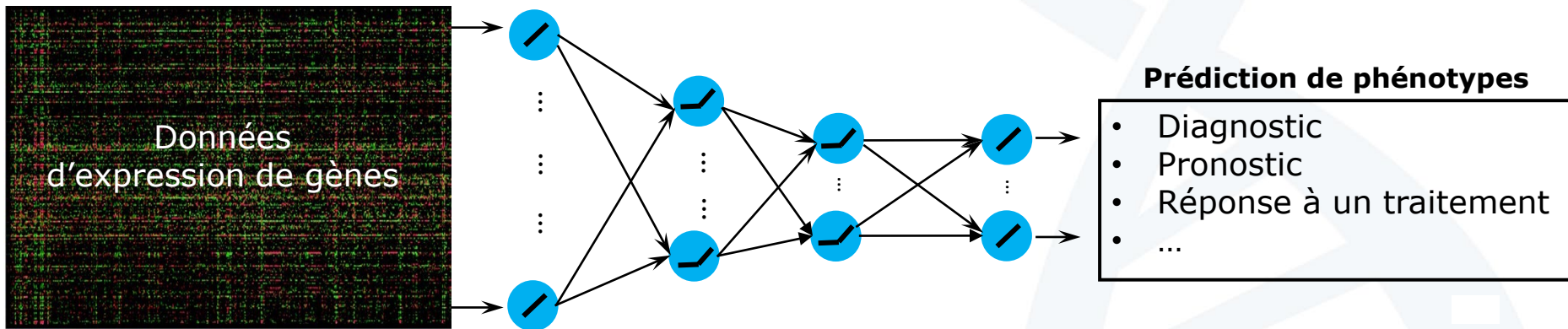
Prédiction de phénotypes

- Diagnostic
- Pronostic
- Réponse à un traitement
- ...



Médecine de précision

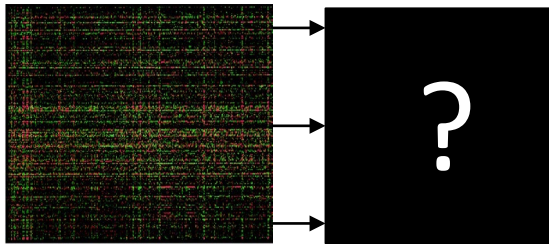
Apprentissage profond appliqué aux données d'expression de gènes



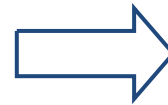
Zou, J., Huss, M., Abid, A., Mohammadi, P., Torkamani, A., & Telenti, A. (2019). A primer on deep learning in genomics. *Nature Genetics* 51, 12-18.
Hanczar, B., Henriette, M., Ratovomanana, T., & Zehraoui, F. (2018). Phenotypes Prediction from Gene Expression Data with Deep Multilayer Perceptron and Unsupervised Pre-training. *International Journal of Bioscience, Biochemistry and Bioinformatics*, 8(2), 125-131.

Inconvénient majeur : non-interprétabilité

Effet « boîte noire »



Prédiction de phénotypes



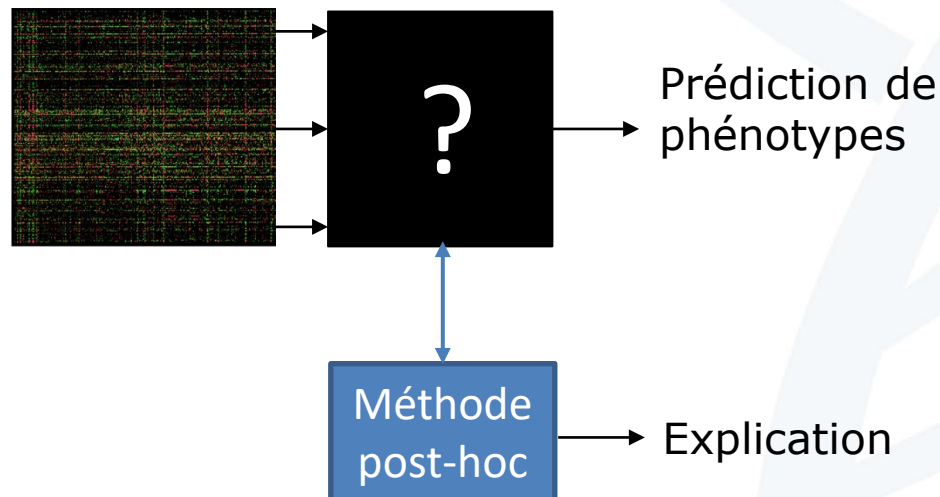
- **Explications nécessaires** pour les médecins et patients
- **Fiabilité et validation** du modèle
- Aide à la **découverte** de nouveaux biomarqueurs

Comment interpréter ?

- **Méthodes post-hoc**
- **Modèles self-explainable**

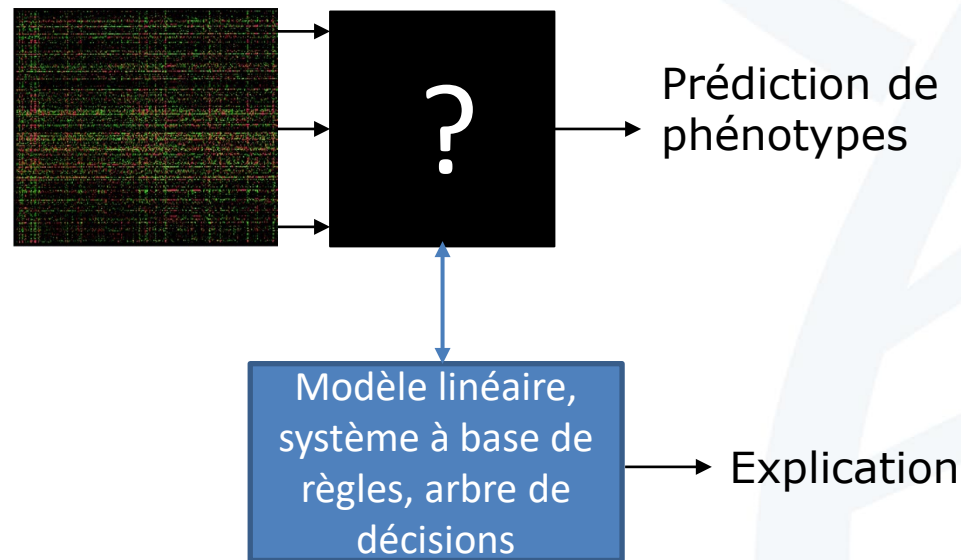
Comment interpréter ?

- **Méthodes post-hoc** : méthodes d'interprétation pour expliquer les prédictions d'un modèle "boîte noire".



Comment interpréter ?

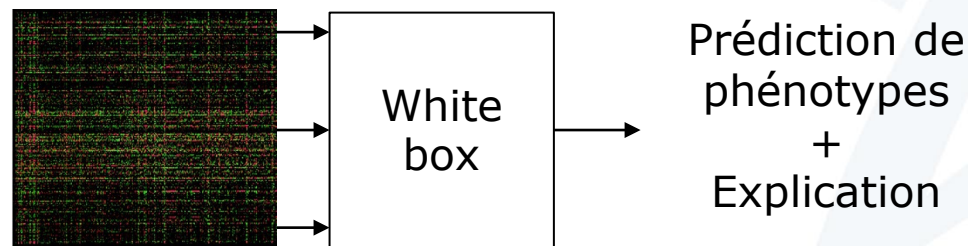
- **Méthodes post-hoc** : méthodes d'approximation



Exemples : LIME, DeepRED

Comment interpréter ?

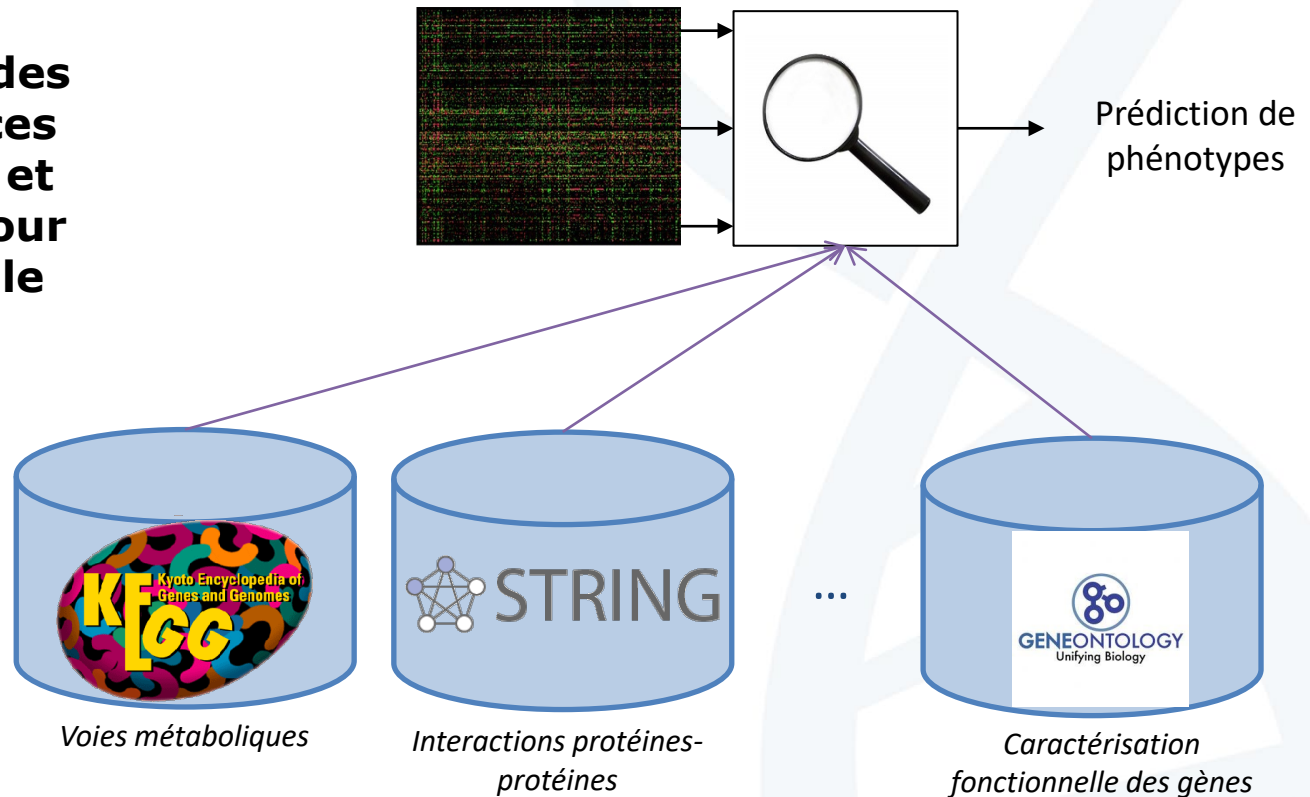
- **Modèles self-explainable** : ces modèles sont par construction interprétables.



Exemple : SENN

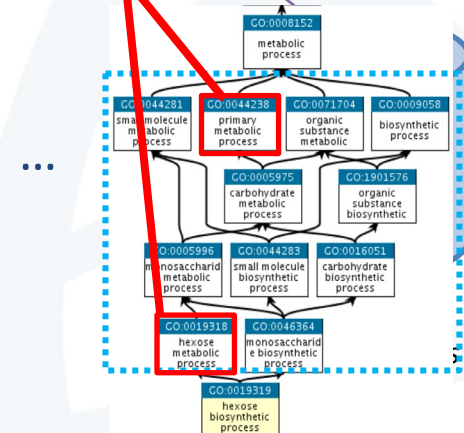
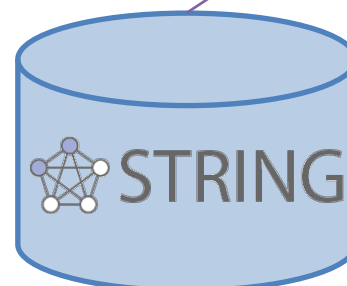
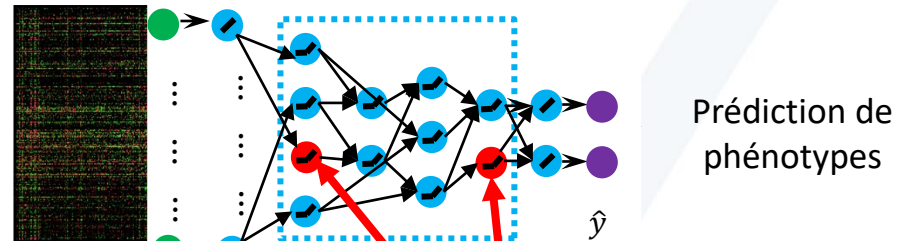
Interprétation biologique du modèle

Intégration des connaissances biologiques et médicales pour interpréter le modèle



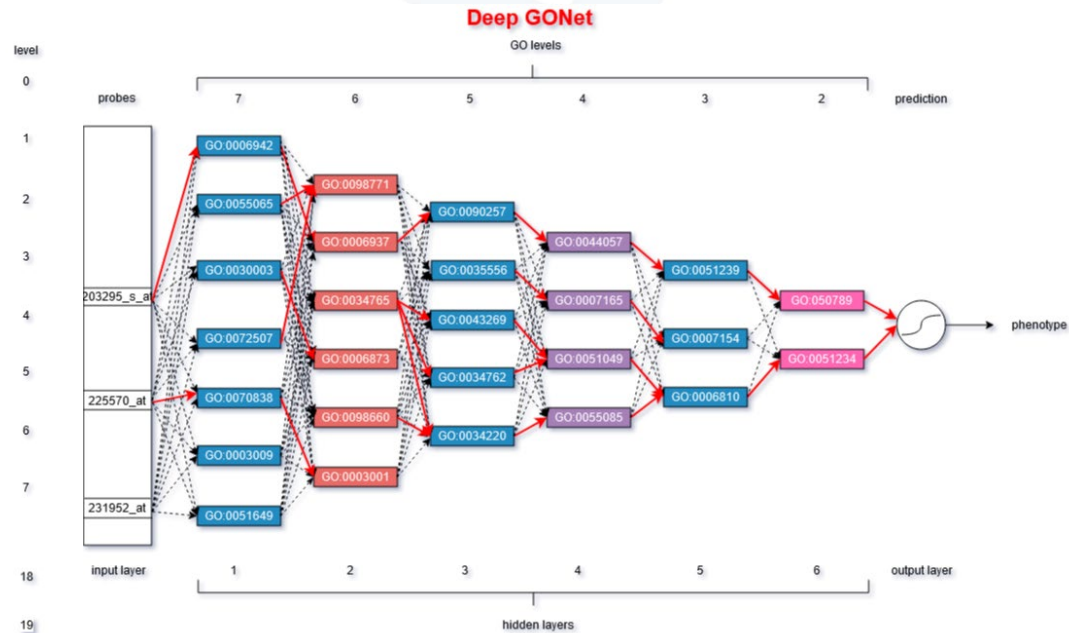
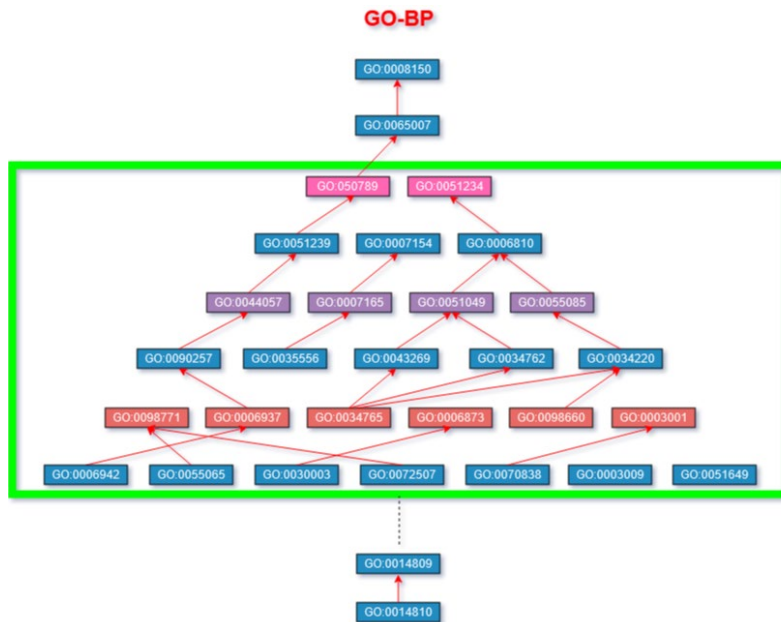
Interprétation biologique du modèle

Intégration des connaissances biologiques et médicales pour interpréter le modèle

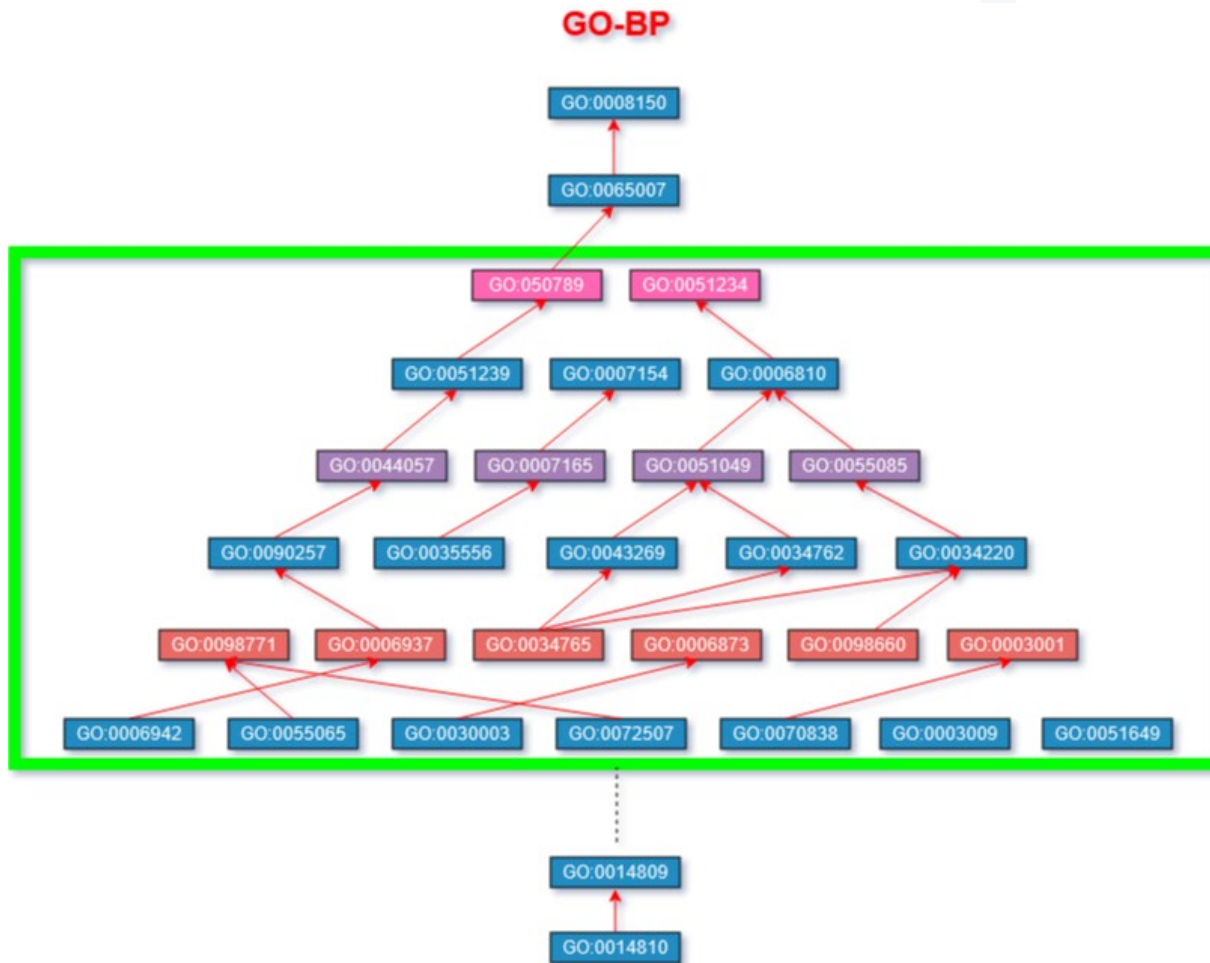


Outil proposé : Deep GONet

- Nouveau réseau de neurones *self-explainable* totalement connecté, contraint par les annotations de la Gene Ontology (GO)

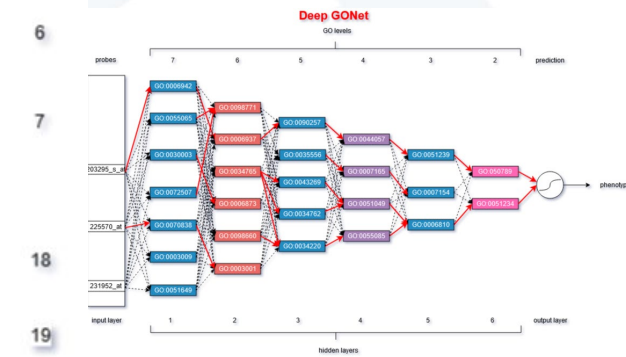


Outil proposé : Deep GONet



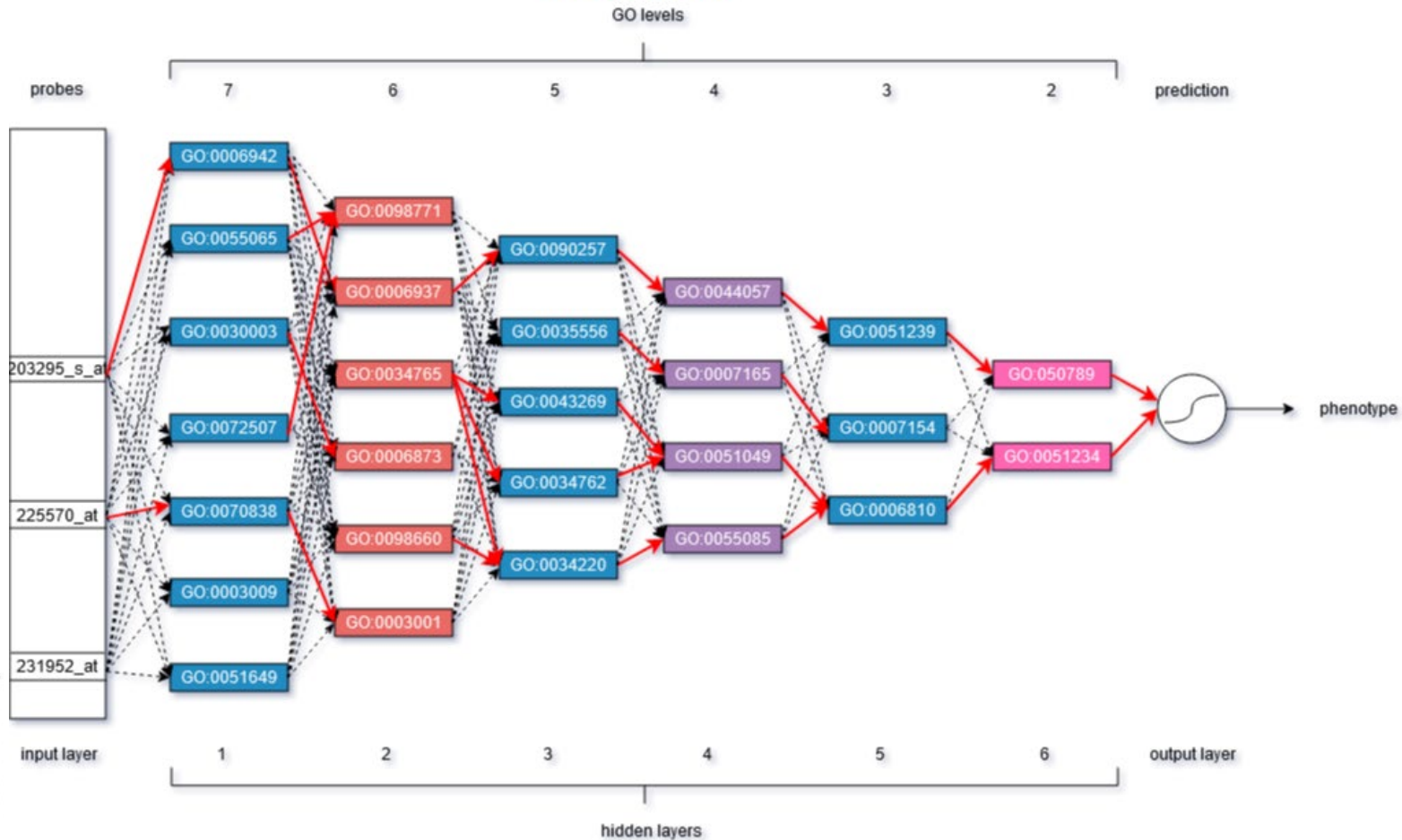
level

- 0
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 18
- 19

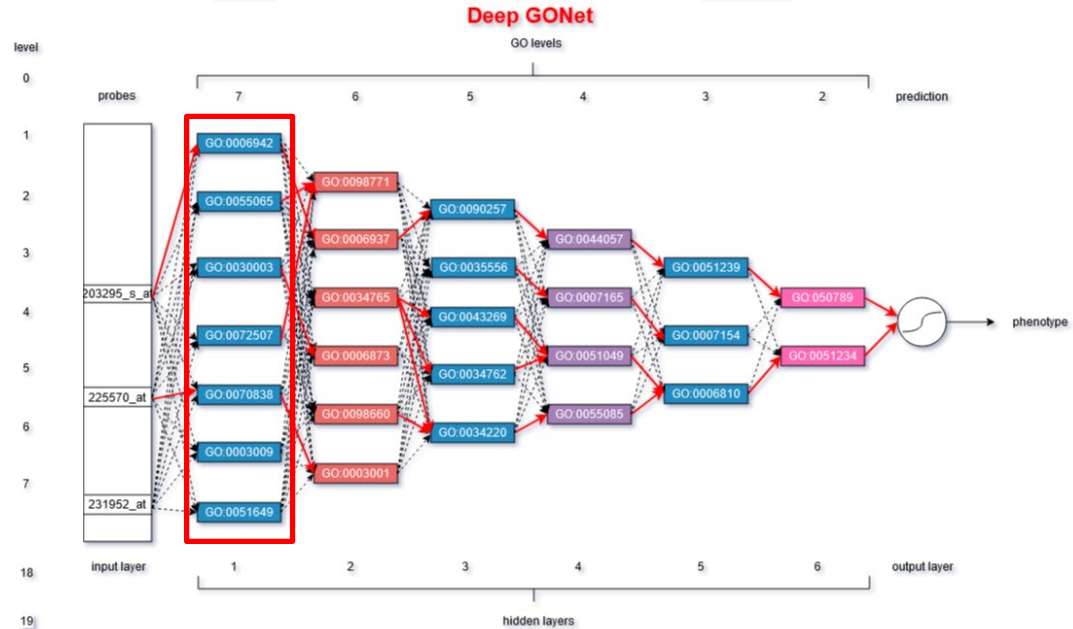
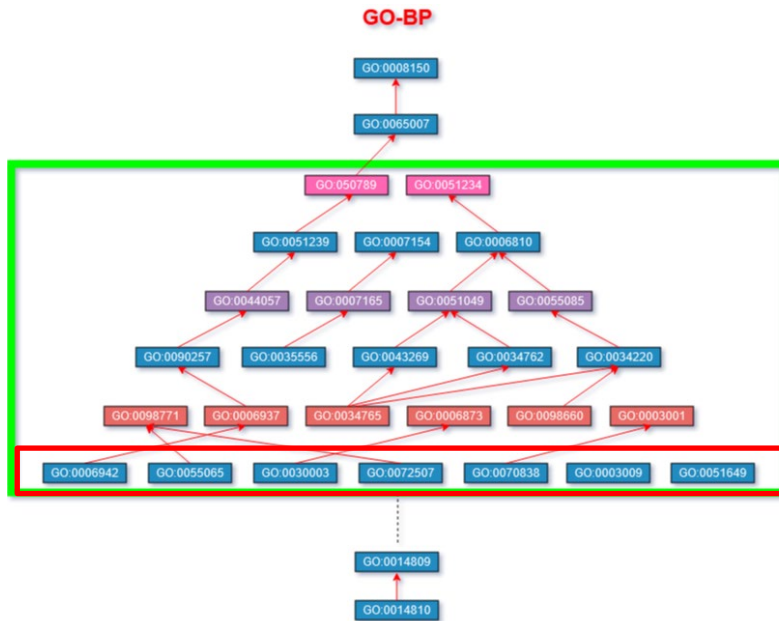


Outil proposé : Deep GONet

Deep GONet

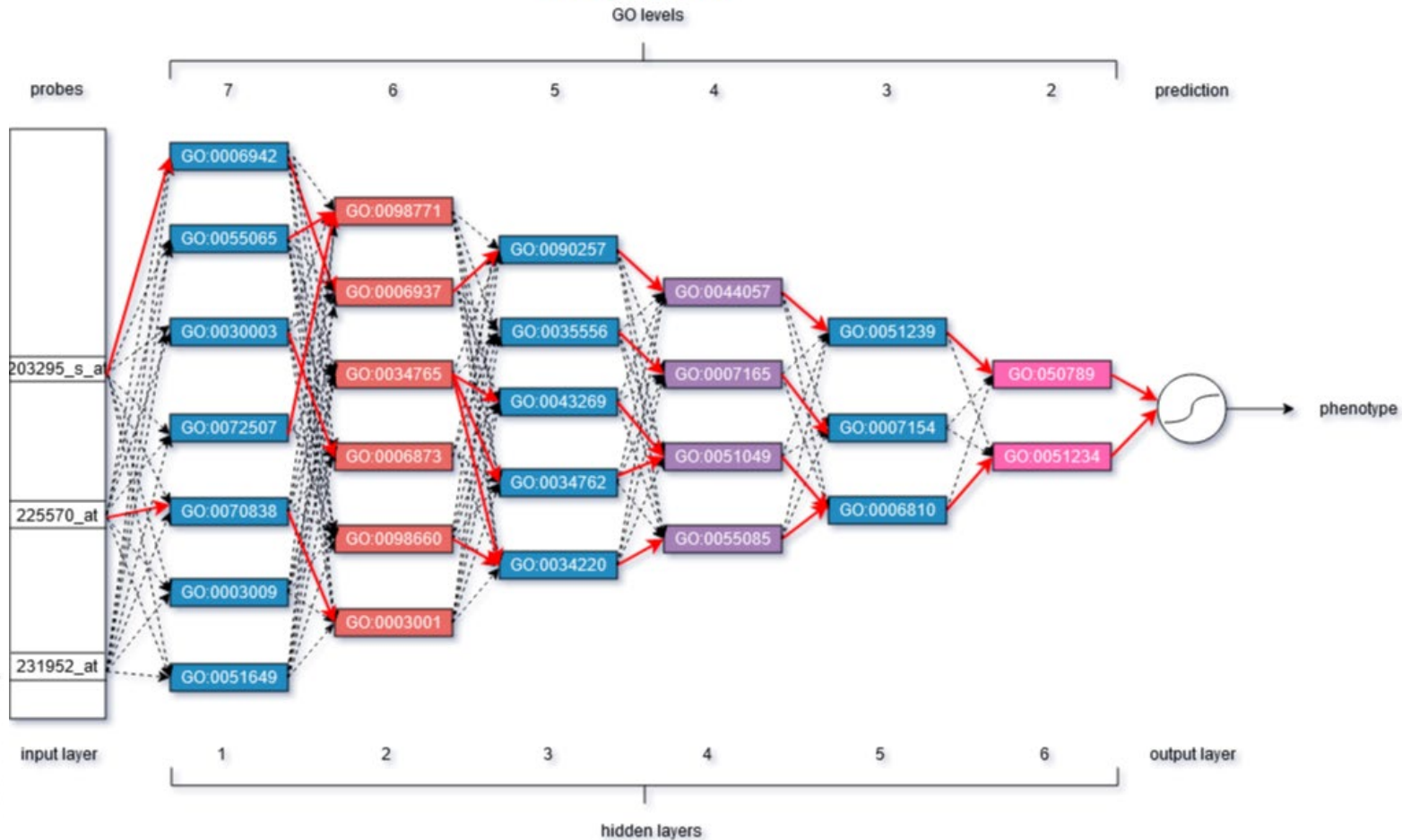


Outil proposé : Deep GONet



Outil proposé : Deep GONet

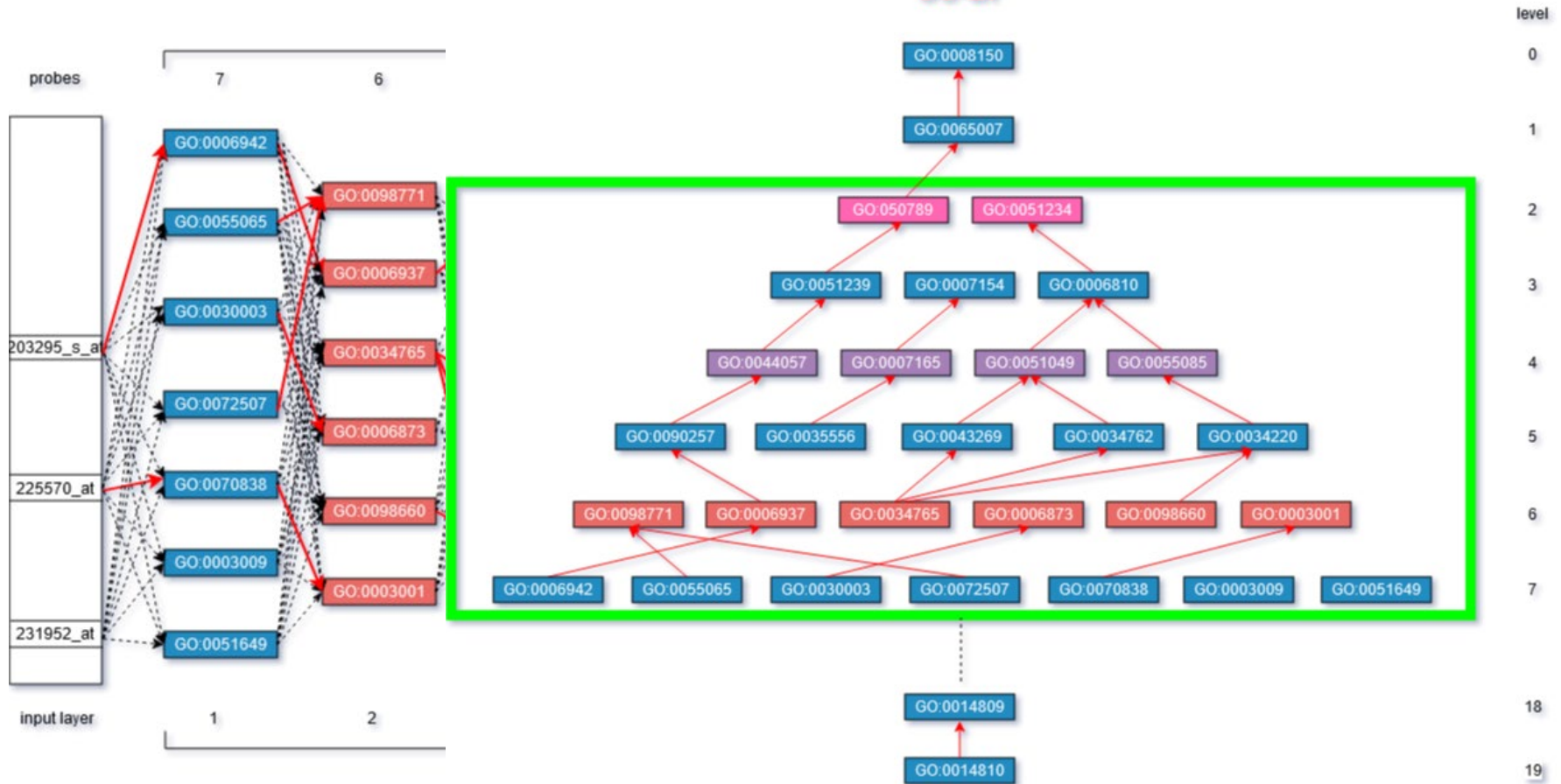
Deep GONet



Outil proposé : Deep GONet

Deep GONet

GO-BP



Régularisation du réseau

- Terme de régularisation personnalisé :

$L_{GO} = \sum_{l=1}^L \|W^{(l)} \otimes (1 - C^{(l)})\|_2^2$ où $W^{(l)}$ est la matrice de poids de la couche (l) et $C^{(l)}$ est la matrice d'adjacence associée telle que

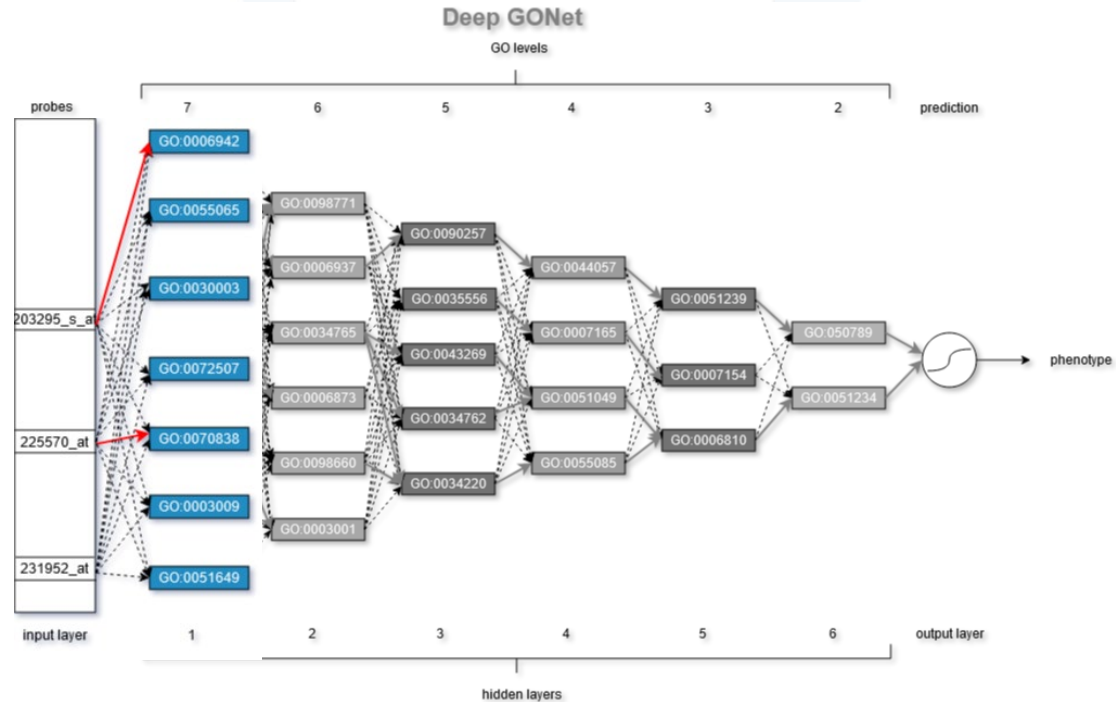
$$c_{j,i} = \begin{cases} 1 & \text{si le terme GO } j \text{ est un enfant du terme GO } i \text{ de niveau adjacent} \\ 0 & \text{sinon} \end{cases}$$

- Fonction de coût :

$$L = \sum_{i=1}^N \sum_{k=1}^K (-y_{i,k} \log \hat{y}_{i,k}) + \alpha L_{GO}$$

Régularisation du réseau

$$C_{1,2} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$



Régularisation du réseau

- Terme de régularisation personnalisé :

$L_{GO} = \sum_{l=1}^L \|W^{(l)} \otimes (1 - C^{(l)})\|_2^2$ où $W^{(l)}$ est la matrice de poids de la couche (l) et $C^{(l)}$ est la matrice d'adjacence associée telle que

$$c_{j,i} = \begin{cases} 1 & \text{si le terme GO } j \text{ est un enfant du terme GO } i \text{ de niveau adjacent} \\ 0 & \text{sinon} \end{cases}$$

- Fonction de coût :

$$L = \sum_{i=1}^N \sum_{k=1}^K (-y_{i,k} \log \hat{y}_{i,k}) + \alpha L_{GO}$$

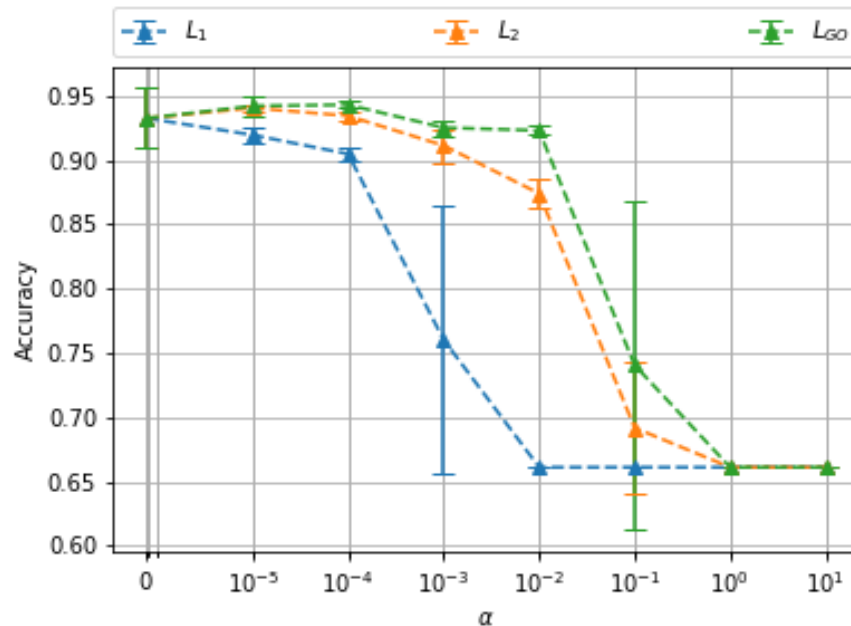
Résultats

- Jeu de données (ID : E-MTAB-3732) :
 - Données Microarray provenant d'une étude croisée (Chip array: Affymetrix HG-U133Plus2).
 - Patients : 22309 (différents types de tissu)
 - Gènes : 54675
 - Prédiction évaluée : détection de cancer (classification binaire)

Torrente, A., Lukk, M., Xue, V., Parkinson, H., Rung, J., & Brazma, A. (2016). Identification of Cancer Related Genes Using a Comprehensive Map of Human Gene Expression. *PLOS ONE*, 11(6), e0157484.

Résultats

- Comparaison avec l'état de l'art

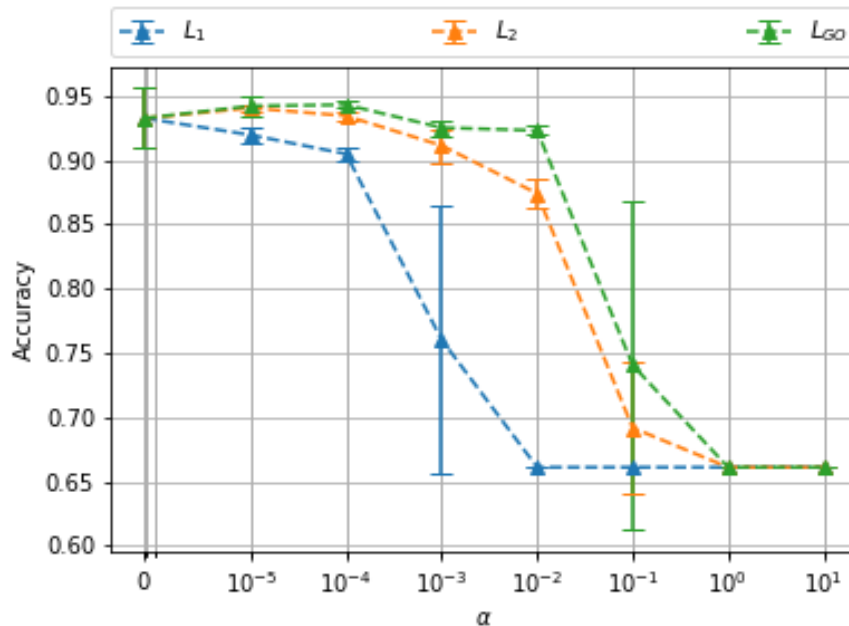


- $L_1 = \sum_{l=1}^L |W^{(l)}|$

- $L_2 = \sum_{l=1}^L \|W^{(l)}\|_2^2$

Résultats

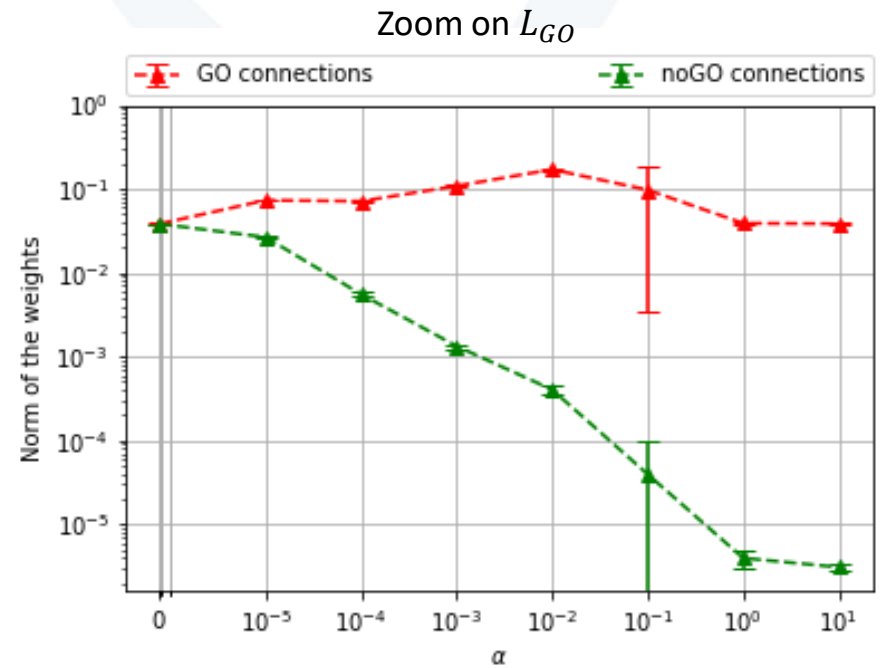
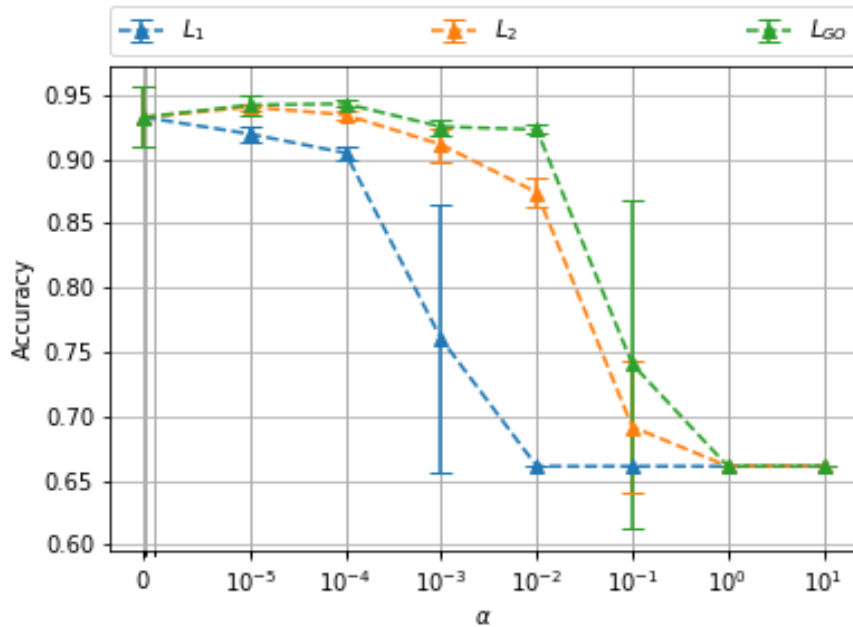
- Comparaison avec l'état de l'art



Modèle	Accuracy
RF	0.904
SVM	0.948
XGBoost	0.936
Classical DL	0.951
Deep GONet	0,950

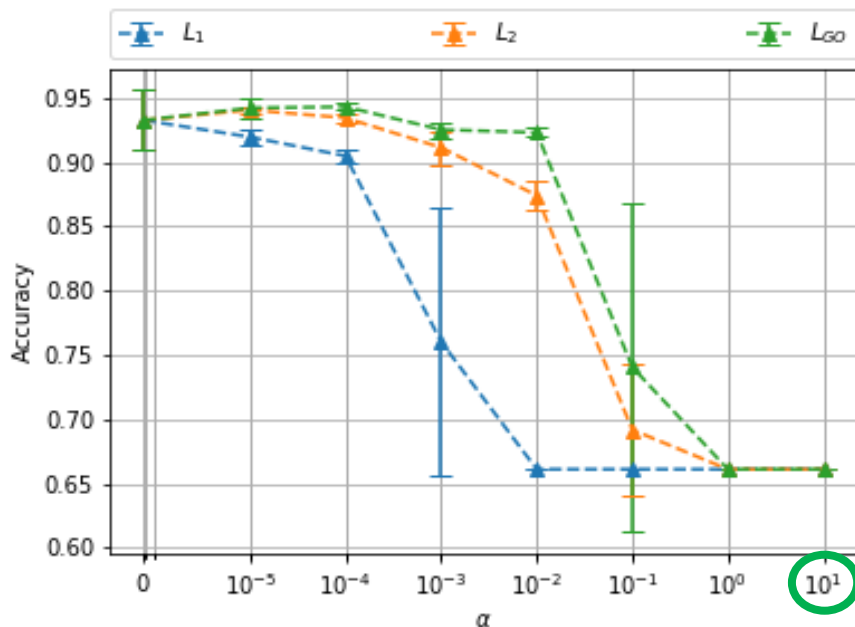
Résultats

- Analyse des connexions du modèle

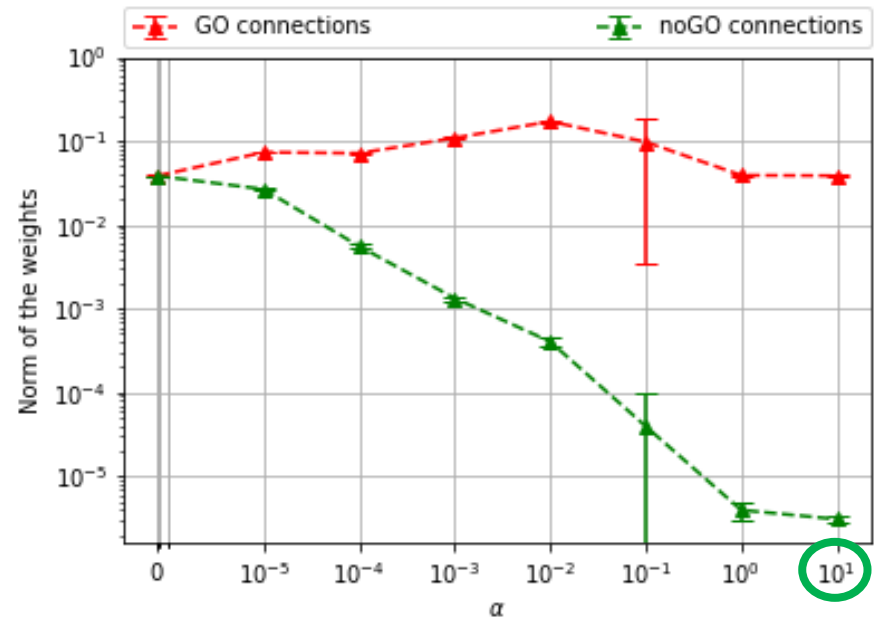


Résultats

- Analyse des connexions du modèle



Zoom on L_{GO}

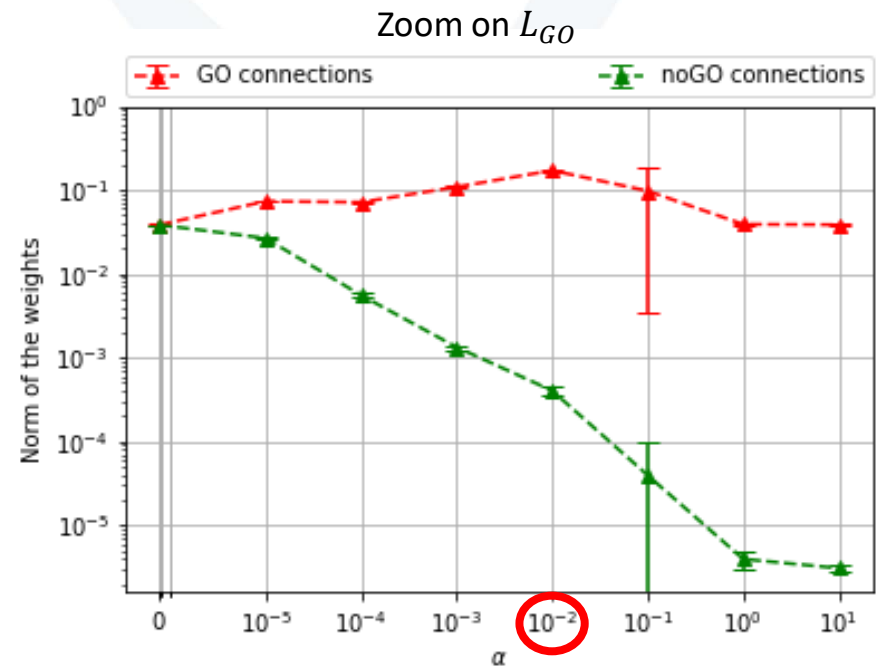
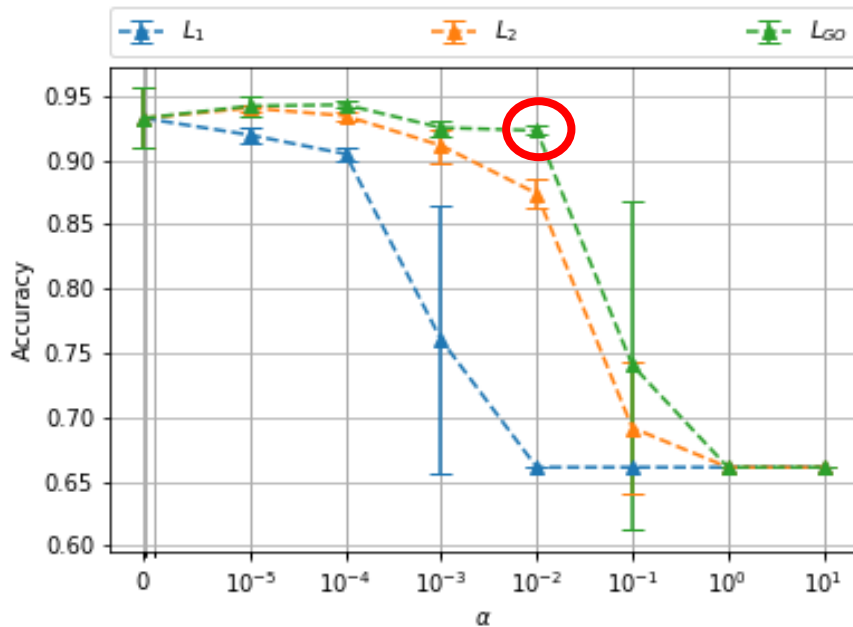


Capacité de prédiction faible

Équivalent à un modèle avec que des connexions GO

Résultats

- Analyse des connexions du modèle



valeur choisie
acc=0.925

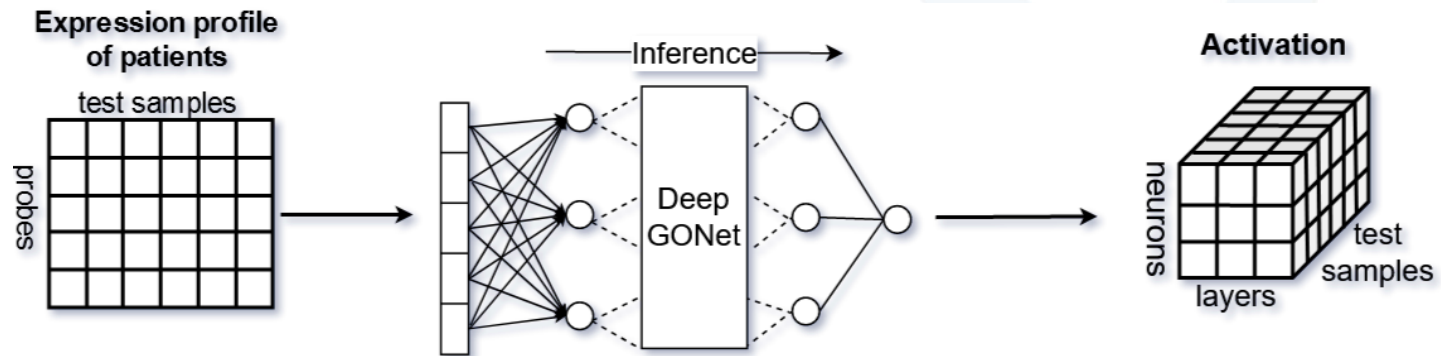
Etudes de cas

- Notre outil permet de fournir une interprétation biologique des résultats à 3 niveaux :
 - Tissus (global)
 - Maladie
 - Patient (local)

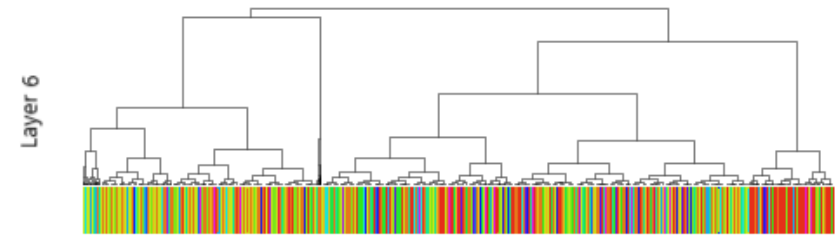
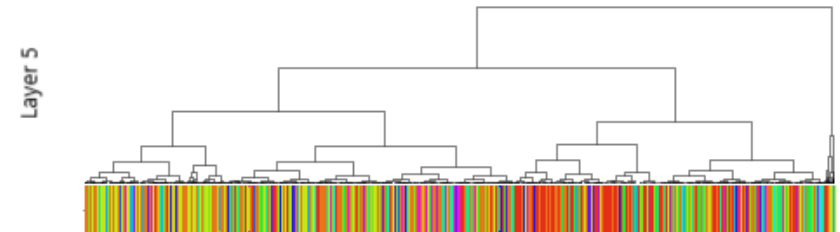
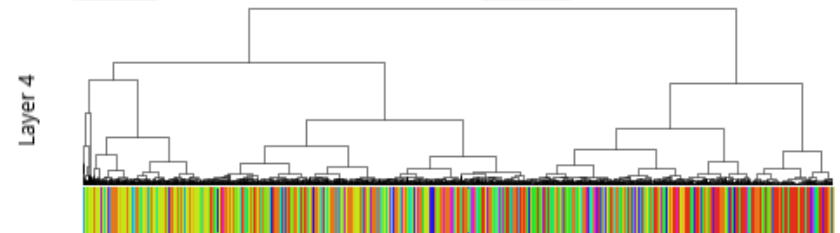
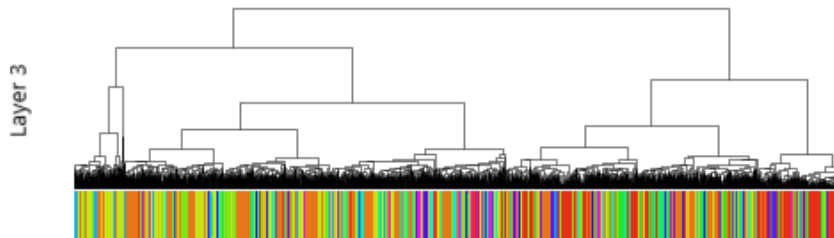
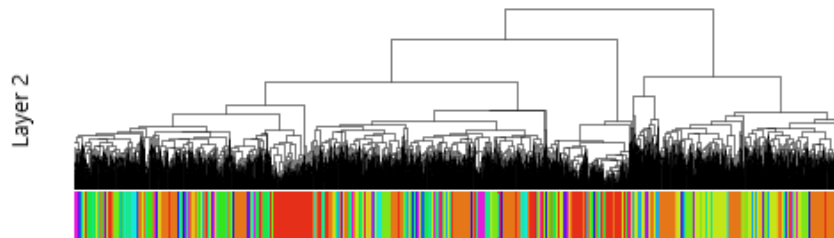
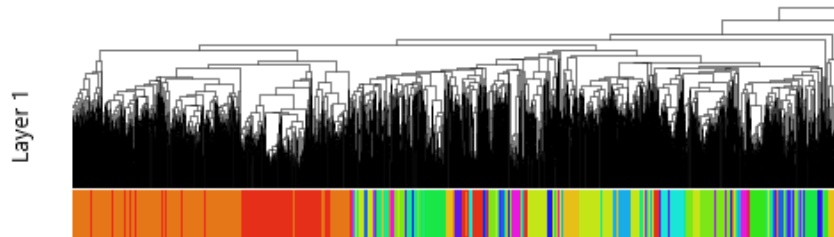
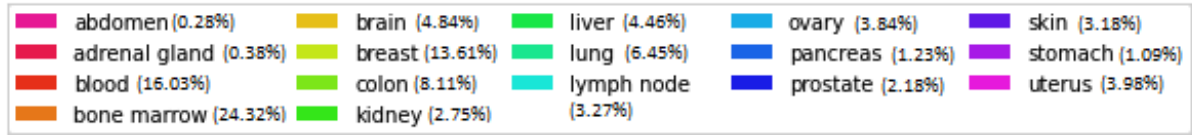
Etudes de cas

- Notre outil permet de fournir une interprétation biologique des résultats à 3 niveaux :
 - **Tissus** (global)
 - Maladie
 - Patient (local)

Etudes de cas – tissus



Etudes de cas – tissus



Etudes de cas

- Notre outil permet de fournir une interprétation biologique des résultats à 3 niveaux :
 - Tissus
 - Maladie
 - **Patient** (local)

Etudes de cas – patient

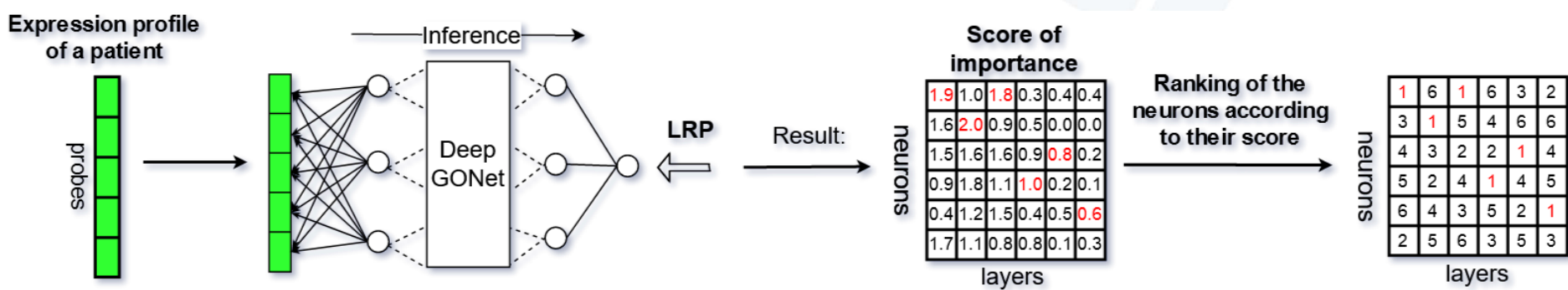
- Mesure additionnelle d'interprétation : score de pertinence (Layer-Wise Relevance Propagation (LRP))

$$R_i^{(l)} = \sum_{j=0}^{N_{l+1}} \frac{a_i^{(l)} w_{i,j}}{\sum_k a_k^{(l)} w_{k,j} + \epsilon} R_j^{(l+1)}$$

Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K.-R., & Samek, W. (2015). On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation. *PLOS ONE*, 10(7), e0130140.

Etudes de cas – patient

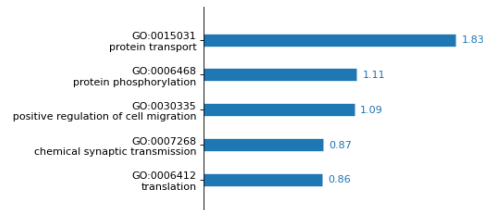
- Mesure additionnelle d'interprétation : score de pertinence (Layer-Wise Relevance Propagation (LRP))



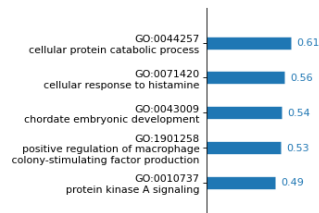
Etudes de cas – patient

- Exemple sur un patient prédit correctement cancer avec une probabilité de 0.99

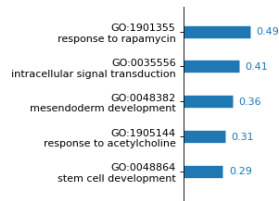
First Hidden Layer



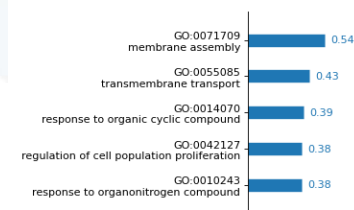
Second Hidden Layer



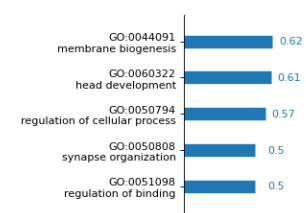
Third Hidden Layer



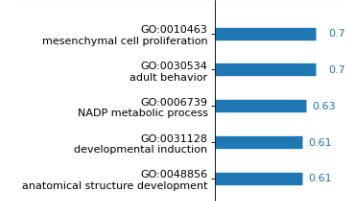
Fourth Hidden Layer



Fifth Hidden Layer



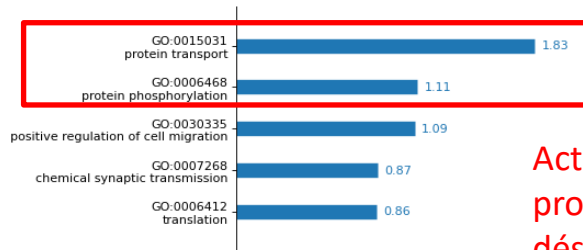
Sixth Hidden Layer



Etudes de cas – patient

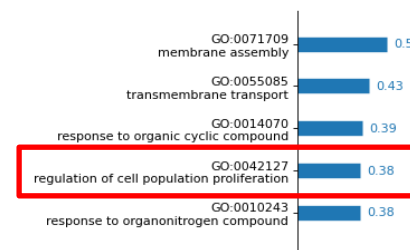
- Exemple sur un patient prédit correctement cancer avec une probabilité de 0.99

First Hidden Layer



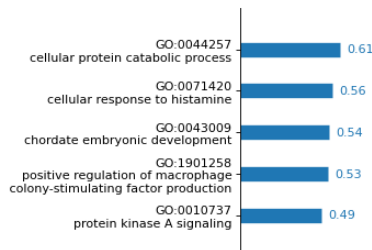
Activité protéique désordonnée

Fourth Hidden Layer

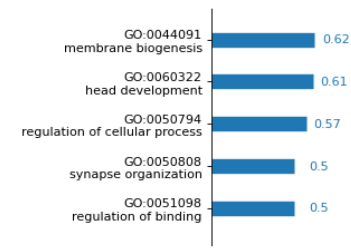


Prolifération incontrôlée des cellules cancéreuses

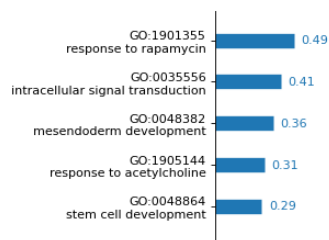
Second Hidden Layer



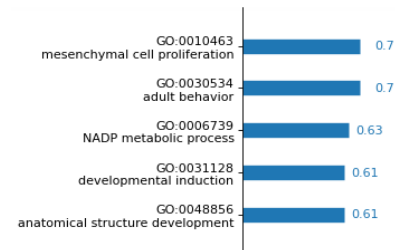
Fifth Hidden Layer



Third Hidden Layer



Sixth Hidden Layer



Conclusions et perspectives

Nous proposons un outil pour :

- **l'interprétation biologique** du modèle et de ses prédictions, compréhensibles par les biologistes et médecins
- **la découverte biologique** des gènes pas encore définis biologiquement

Conclusions et perspectives

Limites et travaux futurs :

- Intégration que de certains niveaux et relations **de l'ontologie**
⇒ Explorer de nouvelles architectures de réseaux de neurones (Graph Neural Network)
- Une **seule connaissance introduite**
⇒ Intégrer d'autres connaissances biologiques (pathways...)

Merci de votre attention !

Git : <https://forge.ibisc.univ-evry.fr/vbourgeois/DeepGONet>

Correspondance : victoria.bourgeois@univ-evry.fr,
blaise.hanczar@univ-evry.fr

Références

- Bourgeais, V., Zehraoui, F., Hamdoune, M. B., & Hanczar, B. (2021). Deep GONet: Self-explainable deep neural network based on Gene Ontology for phenotype prediction from gene expression data. 19th Asia Pacific Bioinformatics Conference (APBC 2021), Tainan, Taiwan.
- Hanczar, B., Zehraoui, F., Issa, T., & Arles, M. (2020). Biological interpretation of deep neural network for phenotype prediction based on gene expression. *BMC Bioinformatics*, 21(1).
- Hanczar, B., Henriette, M., Ratovomanana, T., & Zehraoui, F. (2018). Phenotypes Prediction from Gene Expression Data with Deep Multilayer Perceptron and Unsupervised Pre-training. *International Journal of Bioscience, Biochemistry and Bioinformatics*, 8(2), 125-131.

Références (2)

- Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82–115.
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why Should I Trust You?”: Explaining the Predictions of Any Classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144.
- Zilke, J. R., Loza Mencía, E., & Janssen, F. (2016). DeepRED – Rule Extraction from Deep Neural Networks. *Discovery Science* (pp. 457–473). Springer International Publishing.
- Melis, D. A., & Jaakkola, T. (2018). *Towards Robust Interpretability with Self-Explaining Neural Networks*. 7775–7784.
- Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K.-R., & Samek, W. (2015). On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation. *PLOS ONE*, 10(7), e0130140.
- Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), 206–215.