

# Stage en deep-learning pour la médecine de précision

**Mots-clés :** Deep learning, Transformers, Data augmentation, Données génomiques, Diagnostique médicale.

## **Contexte :**

L'apprentissage profond (Deep Learning) est une avancée majeure de l'intelligence artificielle de ces dernières années. Cette approche de l'apprentissage automatique consiste à apprendre à un réseau de neurones de grande taille à réaliser une tâche de prédiction à l'aide d'un ensemble de données d'apprentissage. L'apprentissage profond s'est rapidement imposé comme un standard dans plusieurs domaines en pulvérisant les records des précédentes méthodes de l'état de l'art. Ses domaines de prédilection sont principalement l'analyse d'images et le traitement du langage naturel. Un des futurs enjeux majeurs de cette approche est son application à la médecine personnalisée.

Nos thèmes de recherche se concentrent plus spécifiquement sur la prédiction de phénotypes (diagnostiques, pronostiques, réponse aux traitements...) à partir de données d'expression de gènes. L'expression génique représente l'activité des gènes d'un individu quantifié par le nombre de brins d'ARN messager identifié dans un prélèvement. Les modèles de réseaux de neurones profonds permettent de prendre en compte la grande complexité et les nombreuses interactions entre gènes pour calculer des prédictions fiables.

## **Sujet :**

Le modèle du Transformer est devenu une référence dans de nombreux domaines, comme le traitement du langage ou la vision artificielle. Les résultats prometteurs de cette architecture reposent sur des mécanismes d'attention [1]. Les mécanismes d'attention peuvent être utilisés pour améliorer la représentation d'un vecteur en incorporant des interactions calculées dynamiquement entre ses différents éléments.

Dans le cadre de la médecine personnalisée, le mécanisme d'attention permettrait de calculer des interactions spécifiques à chaque patient. Cependant l'application du mécanisme d'attention tel que proposé par Vaswani et al est coûteux en temps et en mémoire pour des vecteurs de grandes dimensions. Les données d'expression de gènes sont de très grandes dimensions, par exemple dans un génome humain on considère qu'il y a plus de 30 000 gènes, on dépasse alors les limites matérielles.

Nous avons proposé une première approche pour appliquer le mécanisme d'attention sur les données d'expression de gènes : AttOmics [2]. Dans cette approche les gènes associés à des groupes basés sur la connaissance ou basé sur les niveaux d'expression similaire.

Une limite de cette approche est que l'on ne considère pas directement les interactions entre les différents gènes, or de nombreux gènes sont connus pour interagir entre-eux. Afin de pouvoir considérer ces interactions, des méthodes permettant des améliorations de l'efficacité mémoire et du temps de calcul du mécanisme d'attention sont nécessaires. L'utilisation d'implémentations optimisées tel que FlashAttention [3] ou la réduction de la précision des calculs permettrait d'améliorer l'efficacité mémoire [4,5]. Des approches d'approximations du mécanisme d'attention tel que Linformer ou Nyströmformer sont aussi des pistes pour appliquer le mécanisme d'attention sur de grand vecteur [6].

Le travail consistera à faire un état de l'art des différentes méthodes de calcul de l'attention et identifier les méthodes les plus adaptées à notre problématique. Il faudra ensuite adapter ces méthodes, les implémenter en python/pytorch et les tester sur des jeux de données publiques.

### **Profil recherché :**

- Étudiant Master 2 Recherche ou en dernière année d'école d'ingénieur de formation informatique ou mathématiques appliquées.
- Connaissances des bibliothèques de machine learning (Scikit-learn) et de deep learning (Pytorch)
- De solides compétences en apprentissage profond sur une diversité d'architectures
- Autonomie et curiosité pour la recherche scientifique

### **Votre candidature :**

- Type de contrat : Stage 5-6 mois
- Lieu : Évry
- Début du contrat : 2024
- Envoyez votre CV et les relevés de notes de vos deux dernières années à [blaise.hanczar\[at\]univ-evry.fr](mailto:blaise.hanczar[at]univ-evry.fr)

### **Références :**

[1] Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; Polosukhin, I. Attention Is All You Need. In *Advances in Neural Information Processing Systems*; Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R., Eds.; Curran Associates, Inc., 2017; Vol. 30.

[2] Beaudé, A., Rafiee Vahid, M., Augé, F., Zehraoui, F. & Hanczar, B. AttOmics: attention-based architecture for diagnosis and prognosis from omics data. *Bioinformatics* vol. 39 i94-i102 (2023).

[3] Dao, T.; Fu, D. Y.; Ermon, S.; Rudra, A.; Re, C. FlashAttention: Fast and Memory-Efficient Exact Attention with IO-Awareness. In *Advances in Neural Information Processing Systems*; Oh, A. H., Agarwal, A., Belgrave, D., Cho, K., Eds.; 2022.

[4] Portes, J.; Trott, A. R.; Havens, S.; KING, D.; Venigalla, A.; Nadeem, M.; Sardana, N.; Khudia, D.; Frankle, J. MosaicBERT: How to Train BERT with a Lunch Money Budget. In *Workshop on Efficient Systems for Foundation Models @ ICML2023*; 2023.

[5] Izsak, P., Berchansky, M. & Levy, O. How to Train BERT with an Academic Budget. *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing* (2021).

[6] Tay, Y., Dehghani, M., Bahri, D. & Metzler, D. Efficient Transformers: A Survey. *ACM Computing Surveys* vol. 55 1-28 (2022).