

Quantification de l'incertitude dans le cas des signaux médicaux avec bruit d'annotation

Mathilde Dupouy¹, Yamil Vindas², Thibaut Dambry³, Blaise K. Guépié⁴, Philippe Delachartre¹

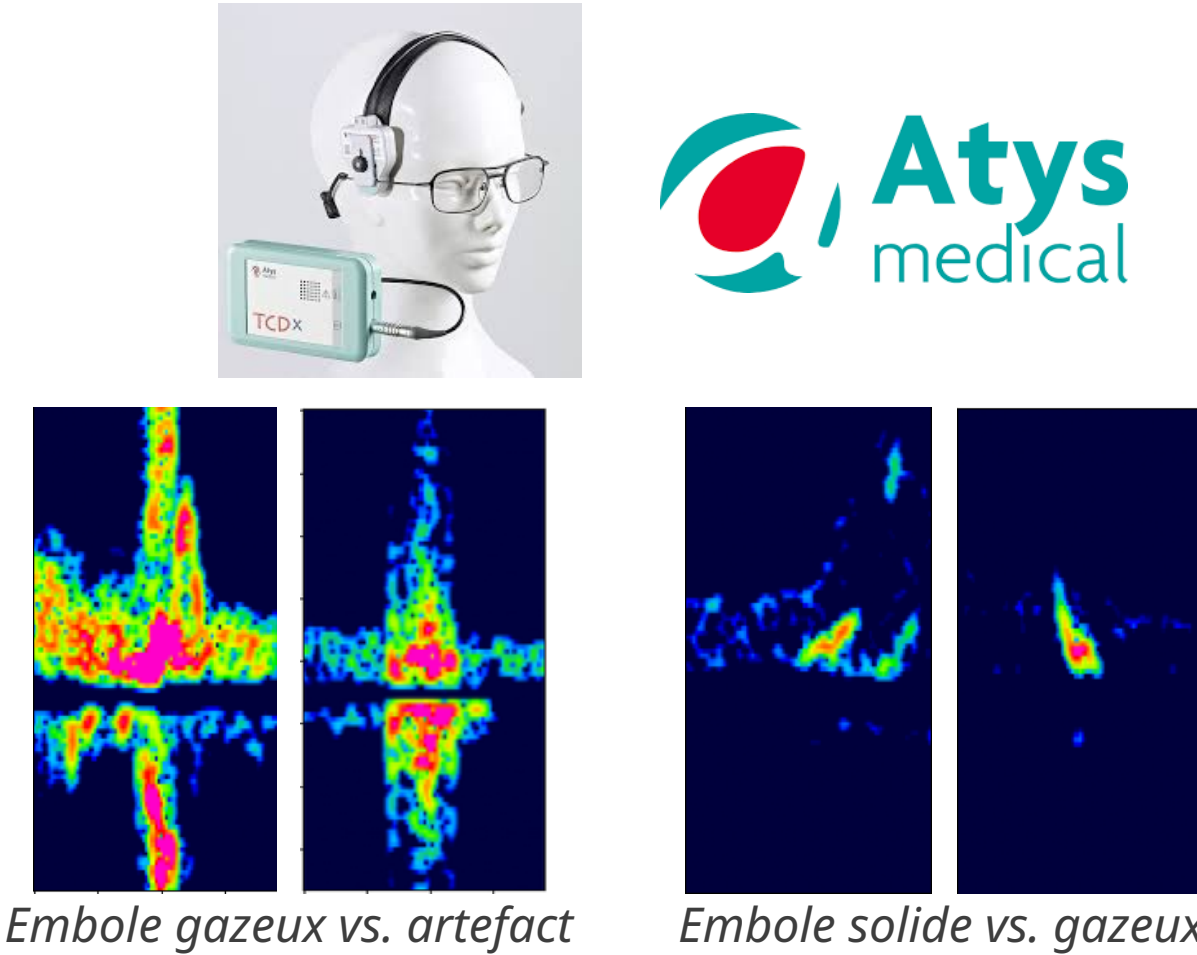
¹CREATIS; CNRS (UMR 5220); INSERM (U1206); INSA Lyon; Université de Lyon, France; ²Data Science for Digital Health, Geneva University, Genève, Suisse; ³Atys Medical, Soucieu-en-Jarrest, France; ⁴Université de Technologie de Troyes, LIST3N, Troyes, France

Introduction

Les annotations bruitées sont courantes.

- **Complexité** de la tâche (observation partielle de la vérité)
- **Biais** de l'annotateur (subjectivité et expérience de l'expert)
- Processus **répétitif / automatique** [1]
- Amplifié dans le **contexte médical** [2]

Holter de Doppler transcrânien TCD-X (haut) et ambiguïté entre les signaux d'embolies sur les spectrogrammes obtenus ▶



Indiquer l'incertitude d'une prédiction aide à la décision [3, 4]

S'il y a du **conflit** entre classe (**incertitude aléatoire AU**) :

- Clinicien : **observation insuffisante** pour conclure
- Chercheur : **entrées additionnelles** nécessaires

Si le modèle est **ignorant** sur la donnée (**incertitude épistémique EU**) :

- Clinicien : classification **non fiable**
- Chercheur : **généralisation** du modèle (plus de données, augmentations...)

Si l'impact du bruit d'annotation sur les performances est étudié, ses effets en terme de quantification d'incertitude restent sous-explorés.

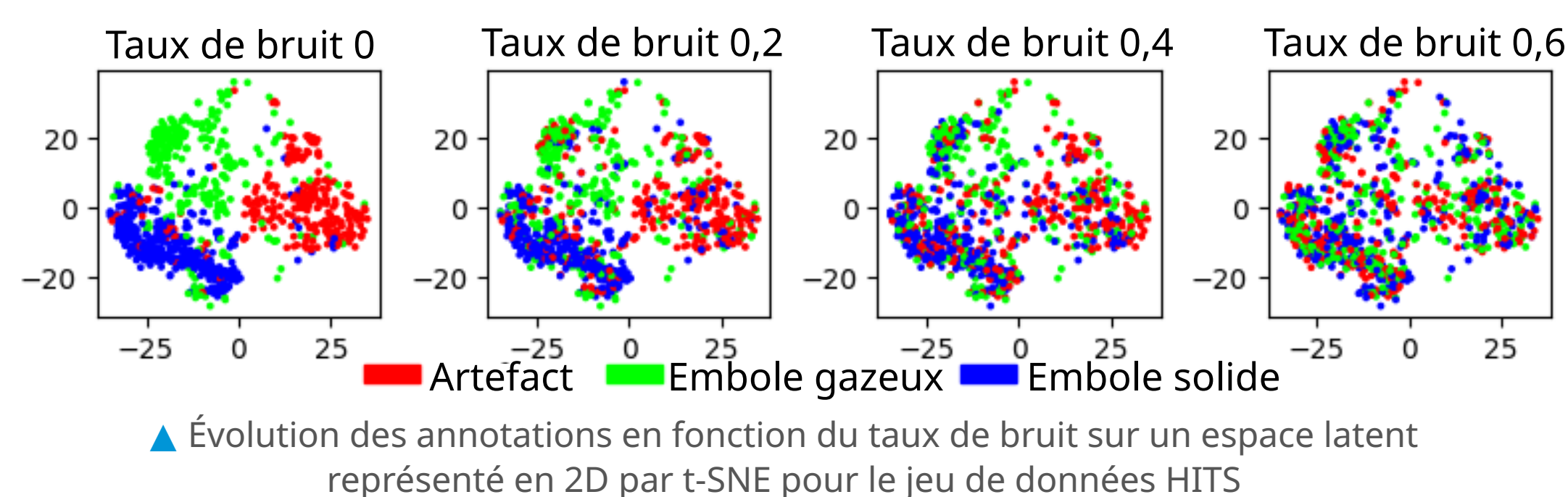
Contribution

Impact d'un bruit d'annotation symétrique sur la quantification d'incertitude par Monte Carlo Dropout

Méthode

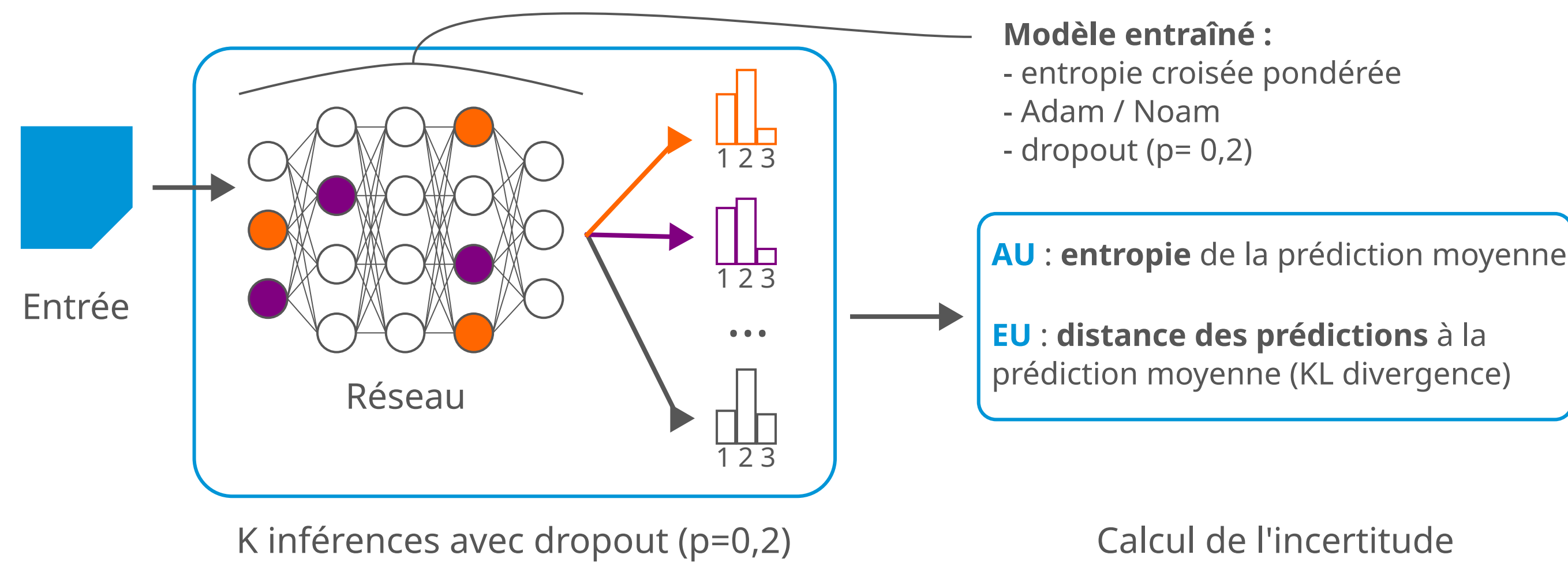
Bruit d'annotation symétrique (*random flip*) [5]

Le bruit d'annotation symétrique simule un changement de classe avec une probabilité fixée (ou **taux de bruit**), uniformément sur les autres classes.



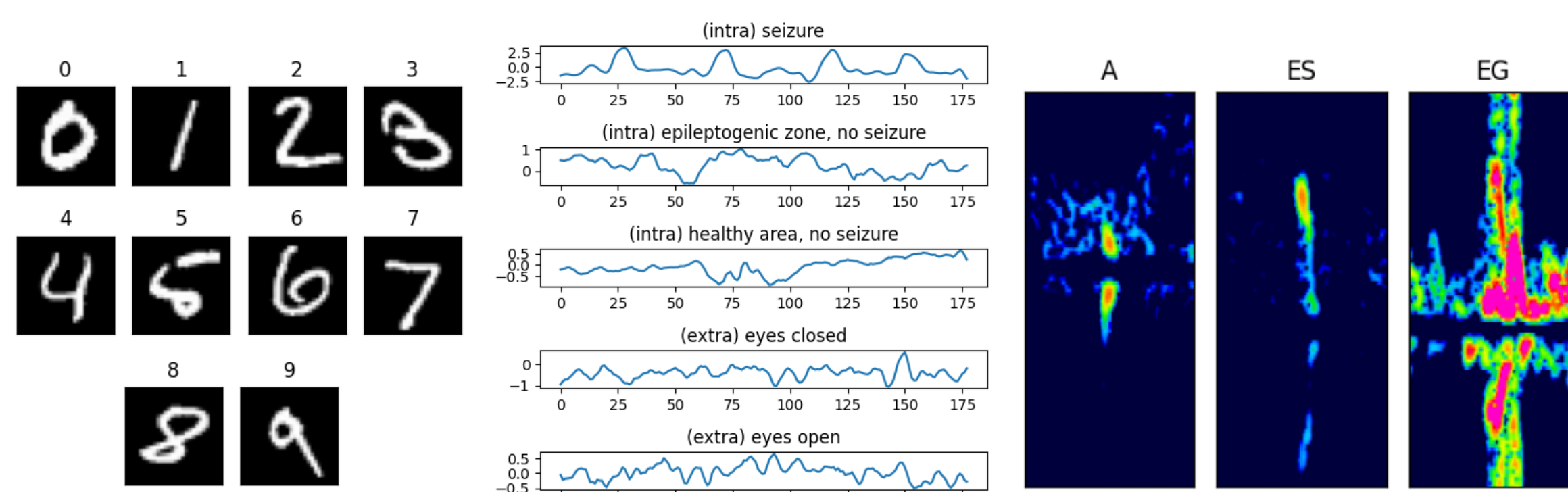
Quantification d'incertitude par Monte Carlo dropout [6]

Cette méthode est la plus utilisée en imagerie médicale pour la quantification d'incertitude [7].



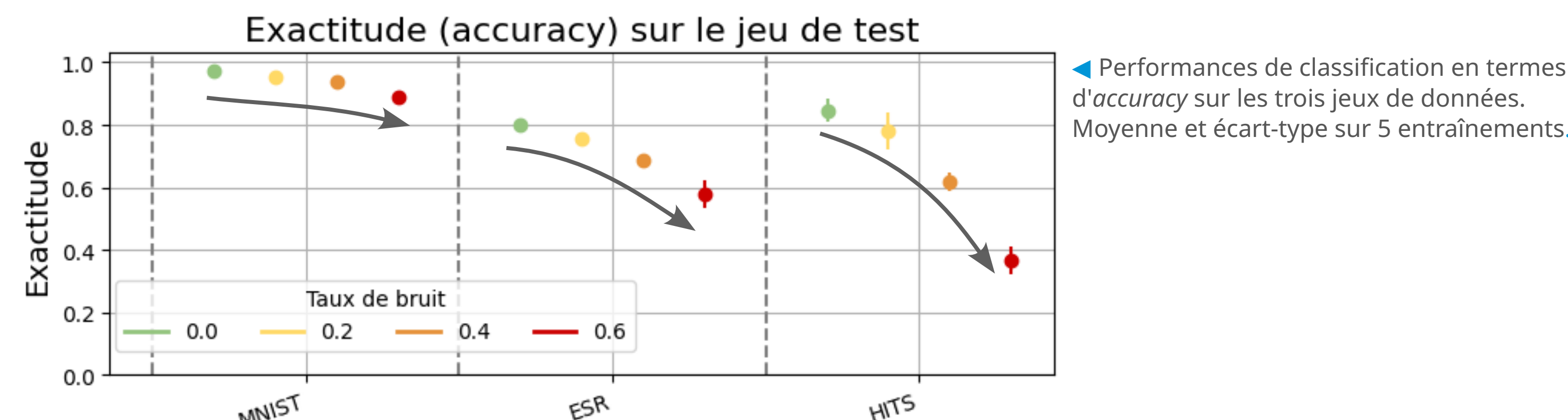
Jeux de données et architectures

	MNIST [8]	ESR [9]	HITS [10]
Description	Images de chiffres manuscrits	Signaux EEG	Spectrogrammes de Doppler transcrânien
Nombre de classes	10	5	3
Taille	70 000	11 500	1545
Balance	86% train, 14% test	80% train, 20% test	64% train, 36% test
Entrée	Image 1x15x15	Signal 1x178	Image 3x224x96
Architecture	2D CNN [10]	1D CNN-Transformer [10]	2D CNN [10]
Nombre de paramètres	56k	110k	1.682M

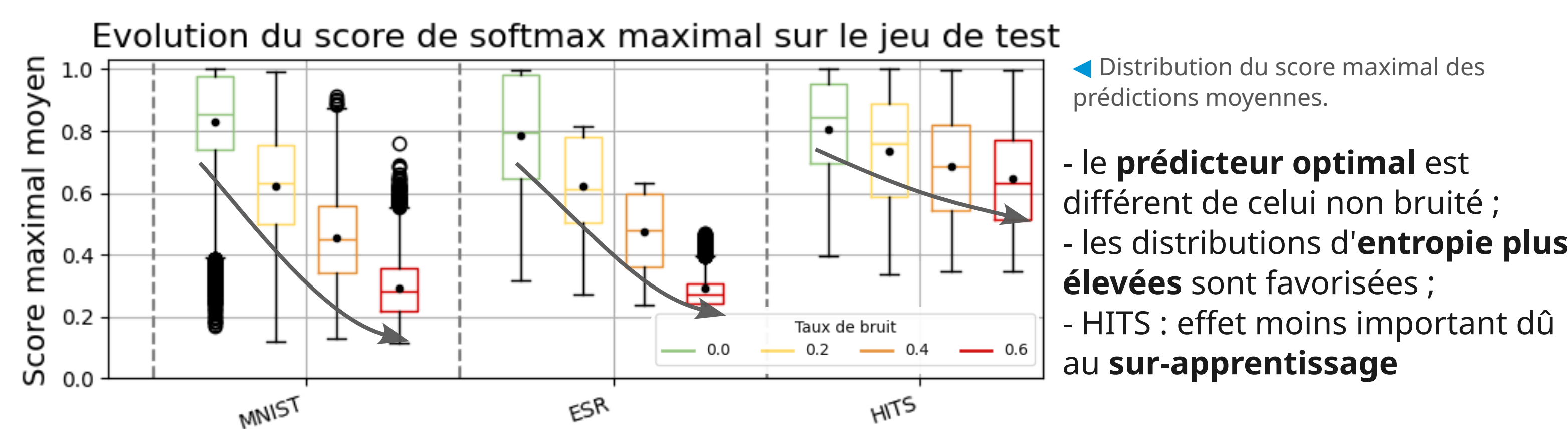


Résultats

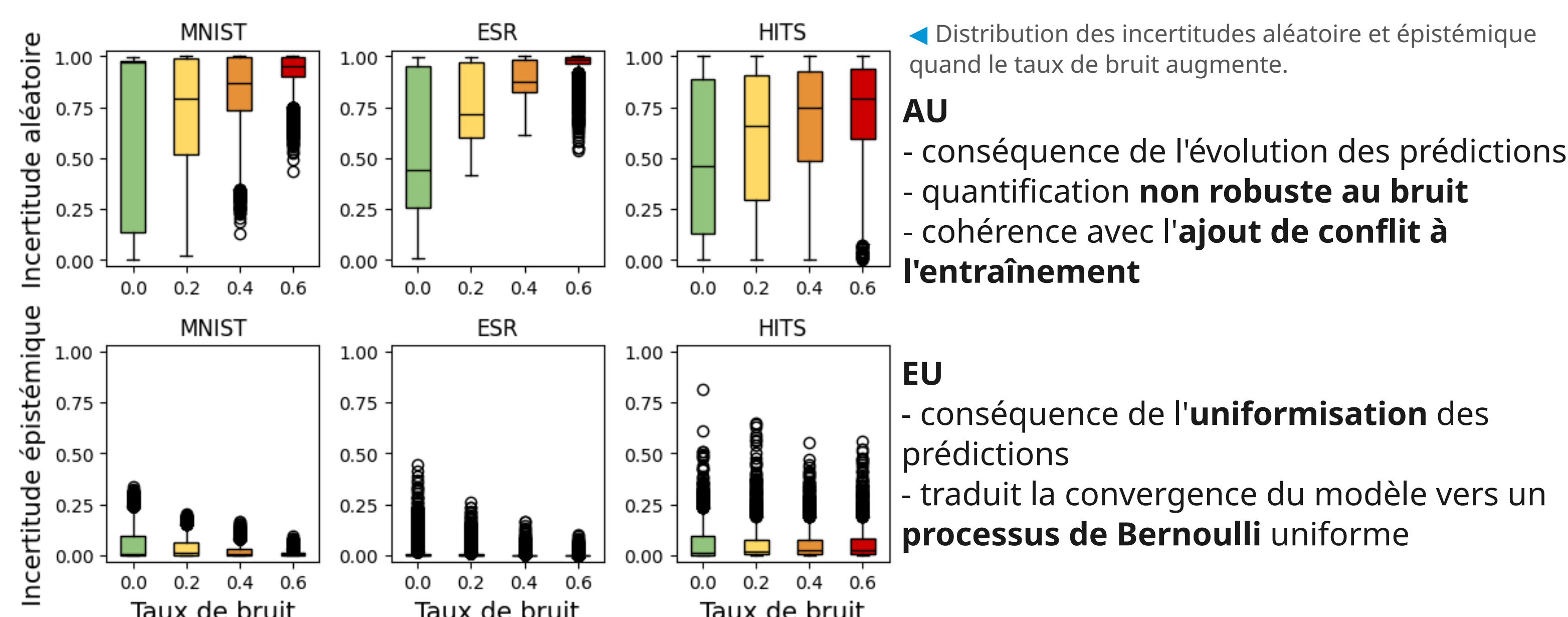
Le bruit d'annotation diminue les performances de classification.



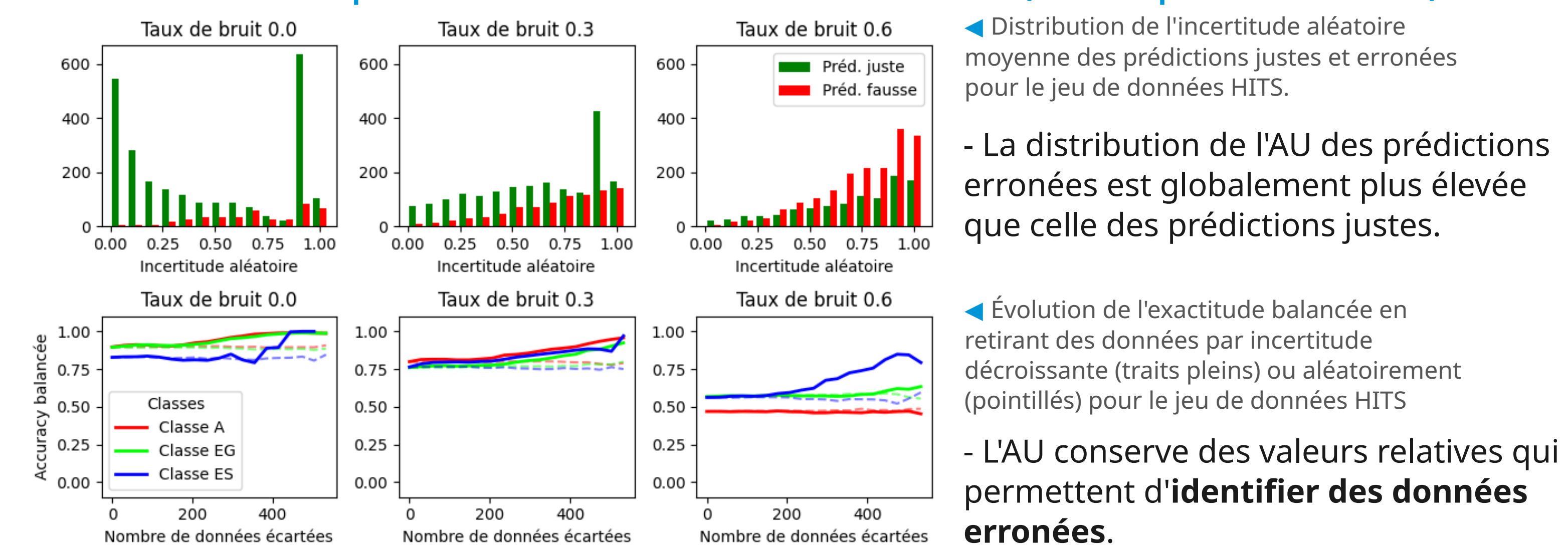
Les prédictions s'uniformisent, deviennent "moins certaines".



En conséquence, AU augmente et EU diminue.



AU reste fiable pour la détection des erreurs (exemple sur HITS).



Conclusion

- Incertitude aléatoire (AU) traduit l'augmentation de conflit à l'entraînement
- Incertitude épistémique (EU) traduit l'effondrement du modèle
- AU permet d'identifier des données erronées (dans un scénario de bruit simple)

Références

- [1] C. G. Northcutt, A. Athalye, and J. Mueller, "Pervasive Label Errors in Test Sets Destabilize Machine Learning Benchmarks," in Neural Information Processing Systems (NeurIPS 2021) Track on Datasets and Benchmarks, Virtual, Dec. 2021.
- [2] J. Shi, K. Zhang, C. Guo, Y. Yang, Y. Xu, and J. Wu, "A survey of label-noise deep learning for medical image analysis," Medical Image Analysis, vol. 95, p. 103166, Jul. 2024.
- [3] T. Löh, M. Ingrisch, and E. Hüllermeier, "Towards Aleatoric and Epistemic Uncertainty in Medical Image Classification," in Artificial Intelligence in Medicine.
- [4] E. Hüllermeier and W. Waegeman, "Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods" Mach Learn, vol. 110, no. 3, pp. 457–506, Mar. 2021.
- [5] B. Frenay and M. Verleysen, "Classification in the Presence of Label Noise: A Survey," IEEE Trans. Neural Netw. Learning Syst., vol. 25, no. 5, pp. 845–869, May 2014.
- [6] Y. Gal and Z. Ghahramani, "Dropout as a bayesian approximation: Representing model uncertainty in deep learning," in international conference on machine learning, PMLR, 2016, pp. 1050–1059.
- [7] B. Lambert, F. Forbes, S. Doyle, H. Dehaene, and M. Dojat, "Trustworthy clinical AI solutions: A unified review of uncertainty quantification in Deep Learning models for medical image analysis," Artificial Intelligence in Medicine, vol. 150, p. 102830, Apr. 2024.
- [8] L. Deng, "The MNIST Database of Handwritten Digit Images for Machine Learning Research [Best of the Web]," IEEE Signal Processing Magazine, vol. 29, no. 6, pp. 141–142, Nov. 2012.
- [9] R. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. Elger, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," Physical review. E, Statistical, nonlinear, and soft matter physics, vol. 64, p. 061907, Jan. 2002.
- [10] Y. Vindas, B. K. Guépié, M. Almar, E. Roux, and P. Delachartre, "An hybrid cnn-transformer model based on multi-feature extraction and attention fusion mechanism for cerebral emboli classification," in Machine Learning for Healthcare, Durham, United States, Aug. 2022.