

Utilisation du Deep Learning pour l'analyse de surface par spectrométrie de masse par temps de vol sur la plateforme MOSAIC-Andromède

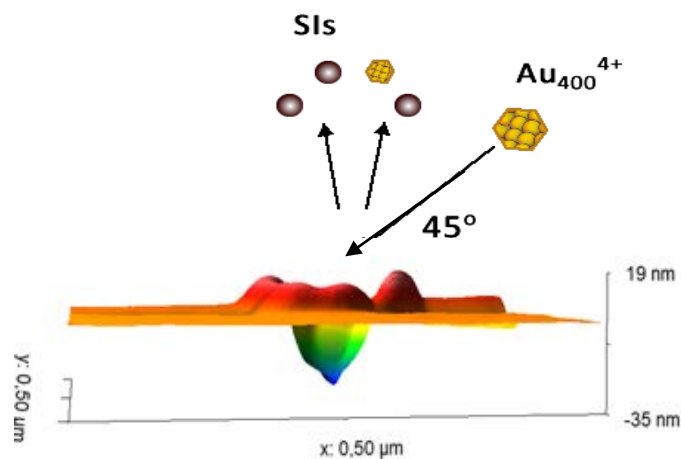
Françoise Bouvet, Service Développement, Département Informatique

Isabelle Ribaud, Plateforme MOSAIC, Pôle Nucléaire

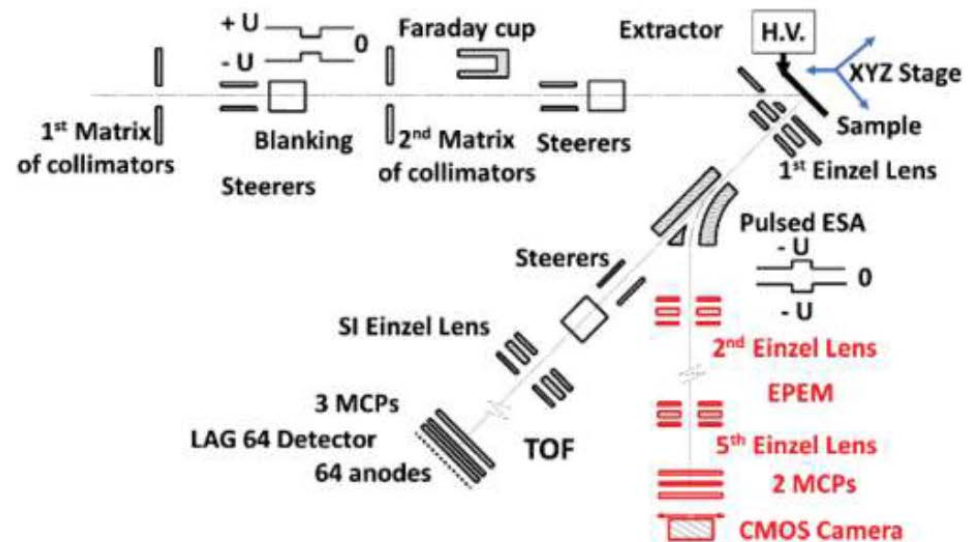


Spectrométrie de masse d'ions secondaires à temps de vol (ToF-SIMS)

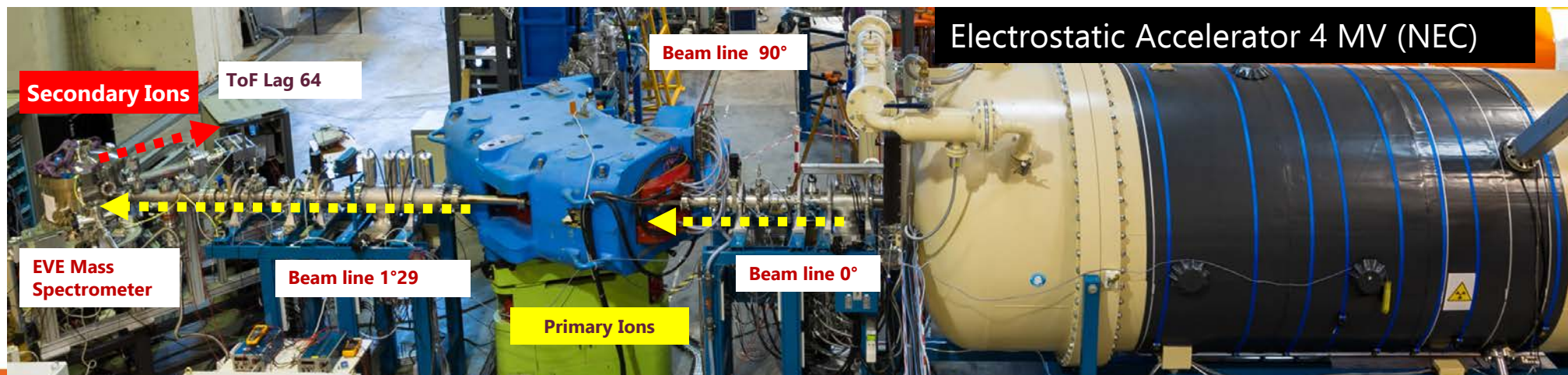
Analyse de surface d'échantillons complexes (biologie, micrométéorites, microélectronique,...) par spectrométrie de masse des ions secondaires à l'aide de faisceaux de nanoparticules d'or énergétique.



Experimental set-up

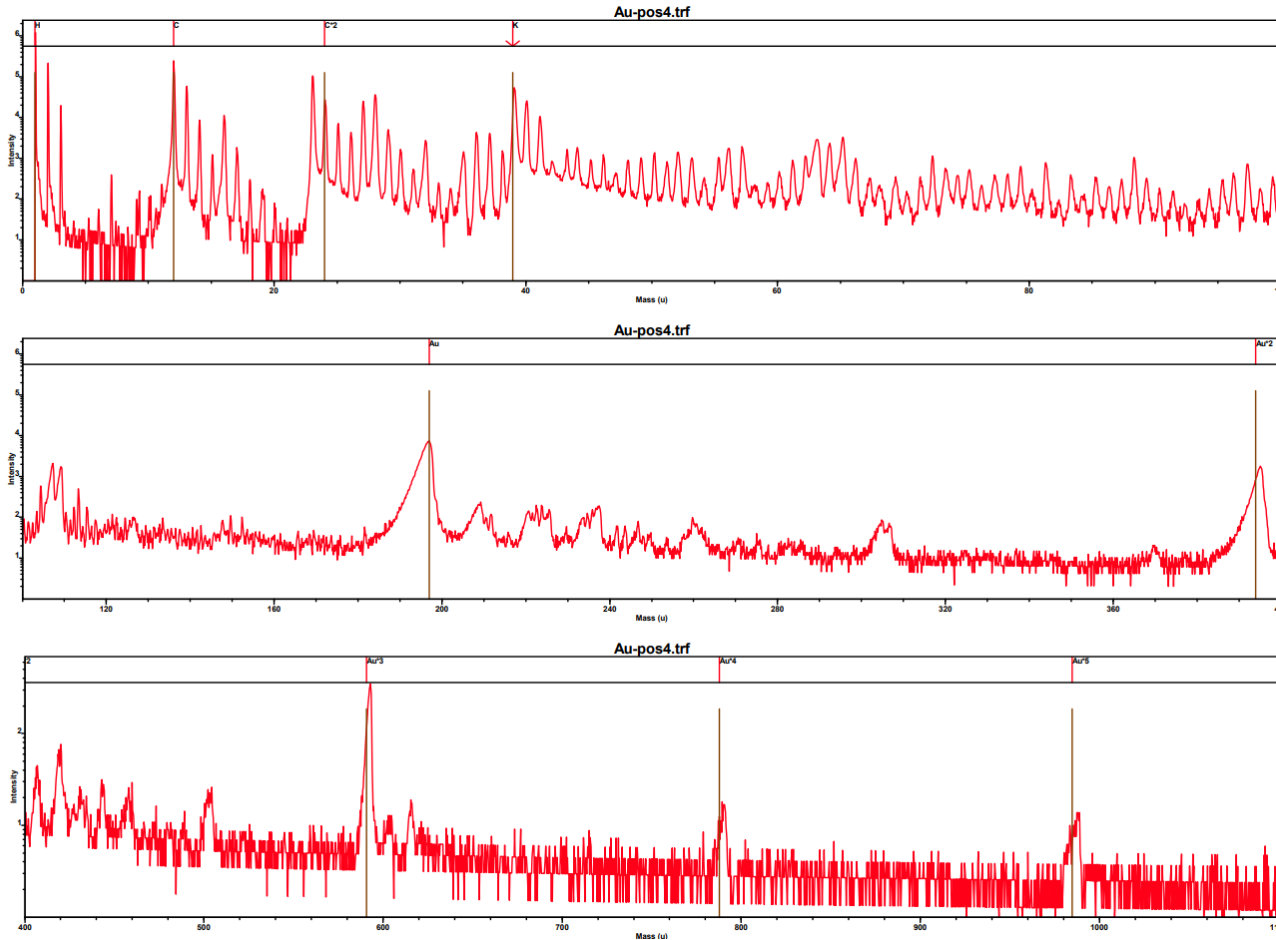


Plateforme Mosaic Andromède





Spectre de masse



Spectre positif d'une cible d'or

Objectifs :

- Identifier les pics de manière **rapide** et **automatique**, aide à l'interprétation
- Développer des modèles prédictifs de classification : fragmentation, clusters, fingerprint, substrat ...

Difficultés :

- Analyse des spectres ToF-SIMS longue et complexe (plusieurs centaines de pics de masse !)
- Mécanismes d'émission secondaire spécifiques aux ions primaires (Au_{400}^{4+} 12MeV), absence de base de données de référence
- Volume de données réelles faibles



Prétraitement des données

1. Conversion du temps de vol en masse/charge

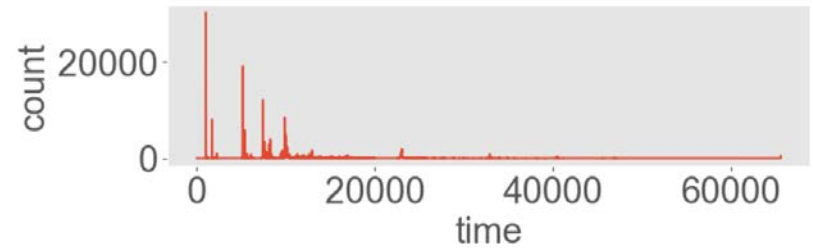
$$\frac{1}{2}mv^2 = qV \quad v = \frac{L}{t}$$

$$\frac{m}{q} = \frac{2Vt^2}{L^2}$$

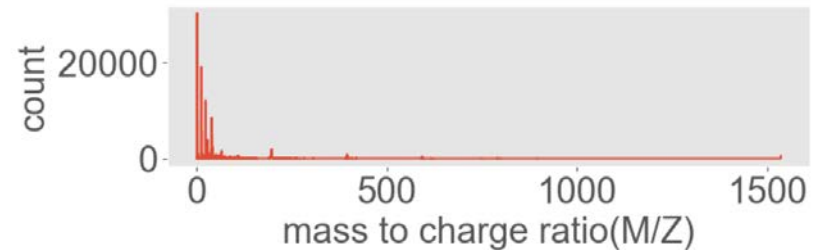
$$\frac{\mathbf{m}}{\mathbf{z}} = (\mathbf{A} \times \mathbf{t} + \mathbf{B})^2$$

où :

- m est la masse
- v : vitesse
- q : charge
- V tension d'accélération
- L : longueur du temps de vol



<Figure size 640x480 with 0 Axes>



2. Interpolation -> points régulièrement espacés

- Interpolation linéaire (T)
- Interpolation de Hermite (T * 3)
- Interpolation par Spline cubique (T * 7)

3. Normalisation par le nombre d'ions primaires

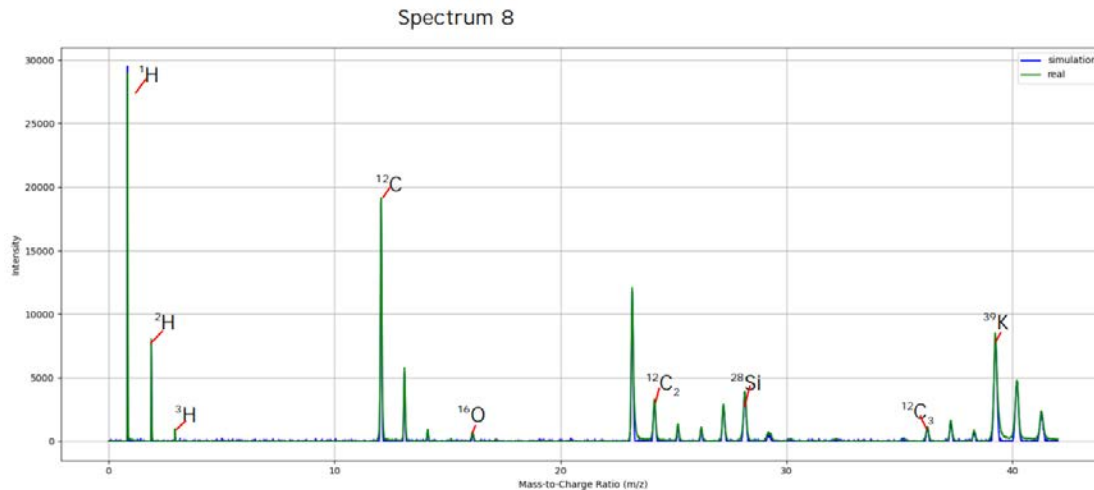


Simulation de spectres

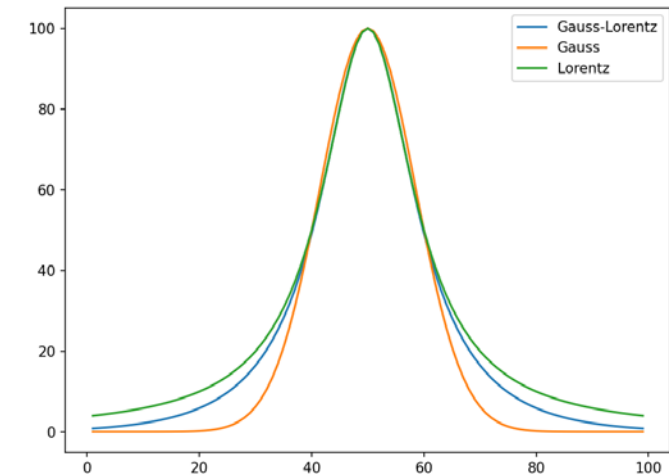
- Extraction des caractéristiques de l'élément à simuler (m/z, FWHM, rapport isotopique)
- Simulation du pic par un produit de fonction de Gauss-Lorentz
- Ajout de bruit (bruit gaussien + décalage aléatoire en m/z)

$$f(x) = H_1 \cdot \frac{\exp\left(- (1 - m_1) \cdot \ln(2) \cdot \frac{(x-x_{01})^2}{\beta_1^2}\right)}{1 + m_1 \cdot \frac{(x-x_{01})^2}{\beta_1^2}}$$

Fonction de Gauss-Lorentz

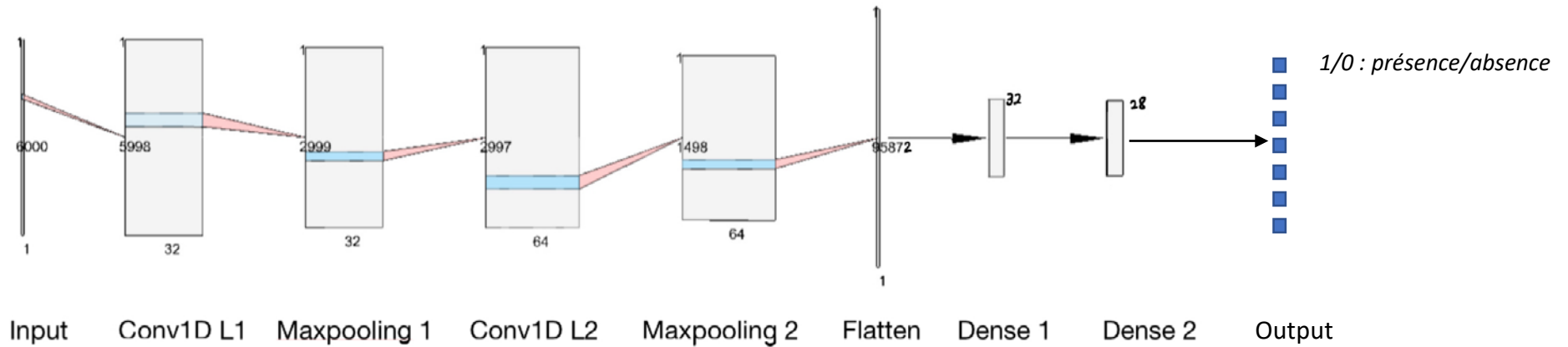


Comparaison de pics simulés et réels





Réseau de convolution 1D



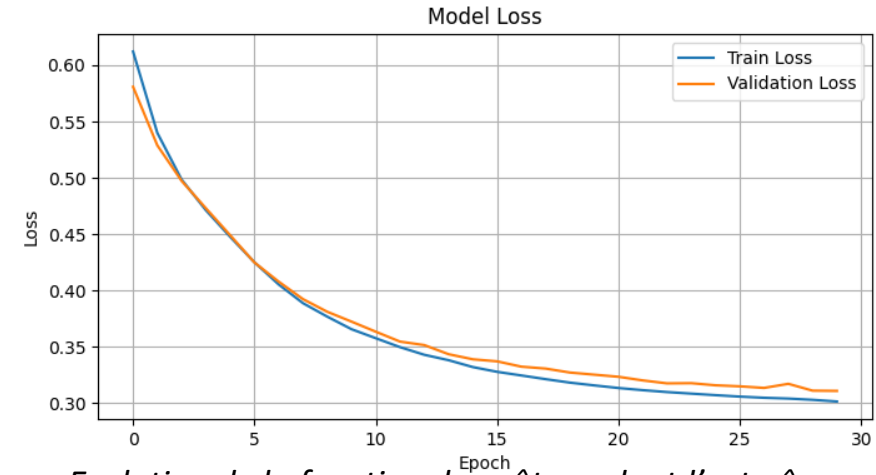
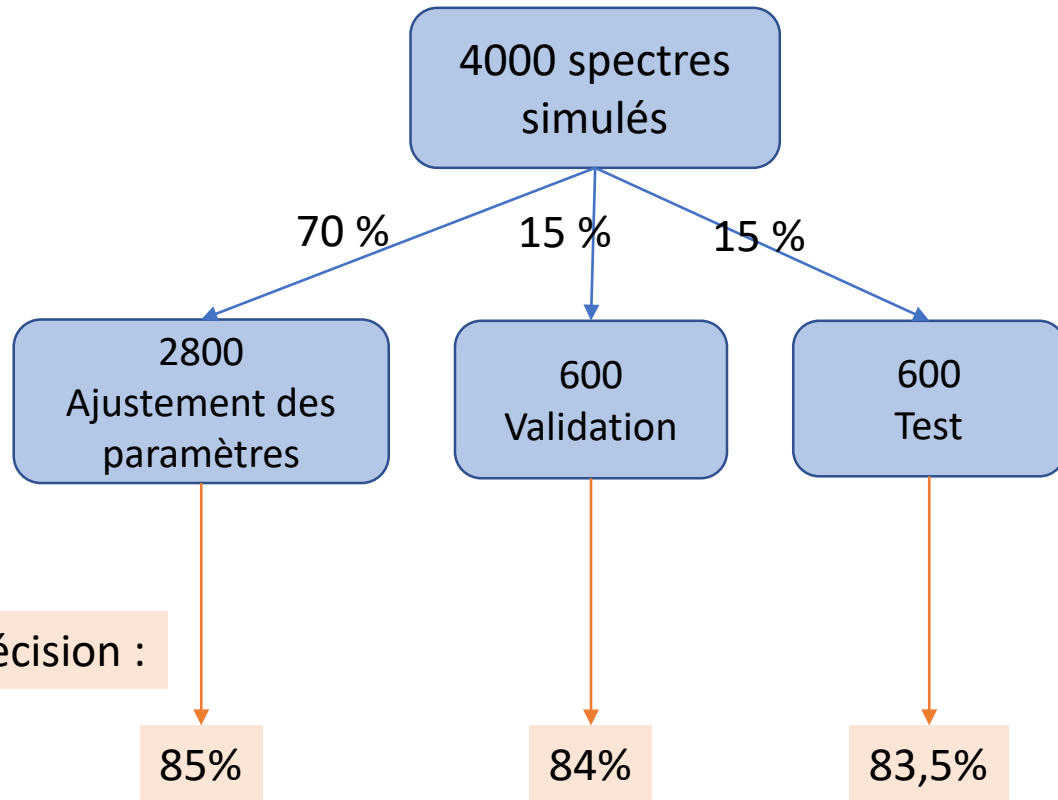
- Entrée : spectre de 6000 points
- Sortie : 7 neurones, un pour chaque élément

Fonction de coût : binary cross-entropy

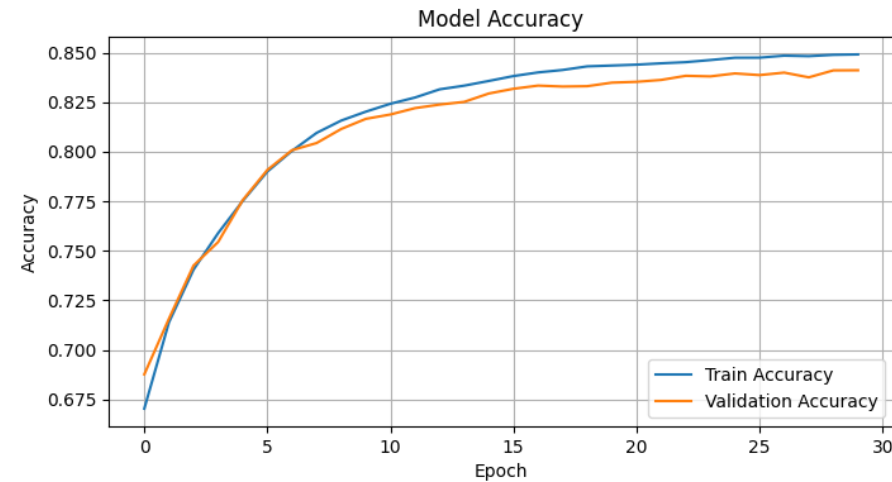
$$Accuracy = \frac{vrais\ positifs + vrais\ négatifs}{nombre\ total}$$



Entraînement



Evolution de la fonction de coût pendant l'entraînement



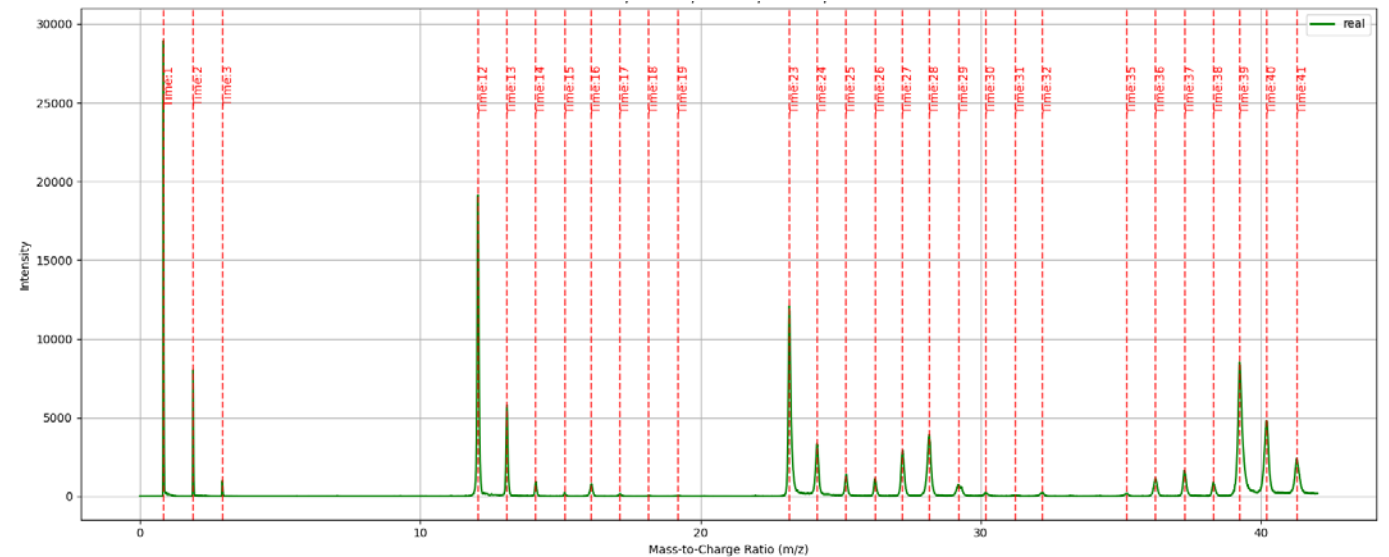
Evolution de la précision (accuracy) pendant l'entraînement



Performance du modèle sur données réelles

Application à des spectres réels :

- 8 spectres traités manuellement
- Précision globale : 80,3%





- Première étude sur des spectres de masse obtenus au moyen d'une sonde de nanoparticules d'or à très haute énergie (12 MeV)
- Résultats satisfaisants pour une cible métallique d'or avec 7 éléments sur la gamme en masse m/z [0-40]

Suite de l'étude :

- Améliorer l'algorithme de simulation
- Augmenter le nombre de spectres réels et simulés
- Augmenter le nombre d'éléments identifiés pour élargir l'analyse à toute la gamme en masse