

# Apprentissage profond multitâche pour la prédiction de la récidence du cancer utilisant l'entropie d'Havrda-Charvat

Thibaud BROCHET<sup>1</sup>, Jérôme LAPUYADE-LAHORGUE<sup>1</sup>, Alexandre HUAT<sup>1,2,3</sup>, Sébastien THUREAU<sup>1,2</sup>, David PASQUIER<sup>4</sup>, Isabelle GARDIN<sup>1,2</sup>, Romain MODZELEWSKI<sup>2</sup>, David GIBON<sup>3</sup>, Juliette THARIAT<sup>5</sup>, Vincent GRÉGOIRE<sup>6</sup>, Pierre VERA<sup>2</sup>, Su RUAN<sup>1</sup>

<sup>1</sup>LITIS, Eq. Quantif, University of Rouen  
22 Nd Gambetta, 76183 Rouen, France

<sup>2</sup>Centre Henri Becquerel  
1 rue d'Amiens, 76038 Rouen, France

<sup>3</sup>Société Aquilab  
Lille, France

<sup>4</sup>Centre Oscar Lambret  
Dpt de radiothérapie, Lille, France

<sup>5</sup>CLCC François Baclesse  
Dpt de radiothérapie, Caen, France

<sup>6</sup>Centre Léon Berard  
Dpt de radiothérapie, Lyon, France

jerome.lapuyade-lahorgue@univ-rouen.fr

**Résumé** – Dans cet article, nous utilisons l'entropie croisée d'Havrda-Charvat, généralisation de l'entropie croisée de Shannon comme fonction de perte dans un réseau de neurones de type U-Net. Nous proposons une architecture multitâche avec plusieurs bases de données pour prédire la récidence après traitement de patients atteints de cancers ORL ainsi que du poumon à partir d'images CT et d'informations sur les patients. L'entropie croisée de Havrda-Charvat est paramétrée par le paramètre  $\alpha$  et coïncide avec celle de Shannon pour  $\alpha = 1$ . Dans cet article, nous étudions la qualité de la prédiction de la récidence en fonction du paramètre  $\alpha$ . L'optimum pour notre jeu de données a lieu pour  $\alpha = 0.2$  avec 69% de prédictions correctes contre 64% avec celle de Shannon.

**Abstract** – In this article, we use Havrda-Charvat cross-entropy, a generalization of Shannon cross-entropy as a loss function in a U-Net neural network. We propose a multitasking architecture with multiple databases to predict post-treatment recurrence of patients with head-neck and lung cancers from CT images and patient information. The Havrda-Charvat cross-entropy is parametrized by the parameter  $\alpha$  and it becomes Shannon cross-entropy when the parameter  $\alpha$  is equal to 1. In this article, we study the quality of recidivism prediction according to the parameter  $\alpha$ . The optimum for our dataset takes place for  $\alpha = 0.2$  with 69% of correct predictions against 64% with Shannon.

## 1 Introduction

Cet article est consacré à l'étude des fonctions de perte dans les réseaux de neurones profonds [1] pour la prédiction des résultats dans les cancers du poumon et ORL. Lorsqu'elle est utilisée pour la prédiction catégorique, la fonction de perte est généralement une entropie croisée dérivée d'une fonction d'entropie. Les entropies sont une mesure de l'information des signaux aléatoires. De telles mesures d'information sont liées au désordre produit par l'échantillonnage de variables aléatoires. Dans [2] sont présentées plusieurs entropies différentes. Dans la classification et la prédiction, la fonction de perte basée sur l'entropie croisée mesure les différences entre les probabilités

prédites et les probabilités réelles de Dirac. Dans la plupart des réseaux de neurones utilisés pour la prédiction, l'entropie croisée basée sur Shannon est la plus courante et est largement utilisée pour la segmentation, la classification ou la détection et de nombreuses autres applications [3]. Dans [2], différentes manières de définir une entropie et sa divergence associée sont détaillées. Parmi elles, l'entropie de Shannon peut être étendue en remplaçant le logarithme par une autre fonction ayant des propriétés de convexité similaires. Les entropies croisées, quant à elles, peuvent être définies en remplaçant la mesure de comptage (resp. mesure de Lebesgue pour le cas continu) par une dérivée de Radon-Nykodim entre deux mesures de probabilités. Parmi les généralisations de l'entropie de Shannon,

on peut citer Renyi [4], Tsallis [5] et Havrda-Charvat [6]. Les entropies de Tsallis et Havrda-Charvat étant en fait la même entropie mais ayant été étudiées à l'origine dans des domaines différents; à savoir la physique statistique pour l'entropie de Tsallis et la théorie de l'information pour l'entropie d'Havrda-Charvat. Dans cet article, nous nous focalisons sur l'entropie croisée basée sur Havrda-Charvat. Cette entropie croisée a la particularité d'être paramétrée par un paramètre  $\alpha$  et de retrouver l'entropie de Shannon lorsque  $\alpha = 1$ .

La nouveauté de cet article réside dans la comparaison des performances des entropies de Shannon et Havrda-Charvat dans le cadre de la prédiction de la récurrence du cancer à partir d'images TDM et de données cliniques pour des patients atteints de cancers ORL ou du poumon. De plus, nous étudions particulièrement la valeur du paramètre  $\alpha$  pour voir quel est son impact sur la prédiction de ses récurrences dans les deux types de cancer. Comme les données médicales sont toujours petites, nous combinons donc deux ensembles de données afin d'augmenter les performances de l'apprentissage. Notre architecture est basée sur un apprentissage multi-tâches comprenant les tâches suivantes : reconstruction d'images, classification des cancers et prédiction des récurrences. L'article est organisé de la manière suivante. Dans une première section, nous rappelons comment l'entropie croisée catégorique de Shannon est définie et comment elle peut être généralisée à Havrda-Charvat et la deuxième section est consacrée aux expériences et à la comparaison entre les deux entropies.

L'article est organisé de la manière suivante. Dans une première section, nous rappelons comment l'entropie croisée catégorique de Shannon est définie et comment elle peut être généralisée à Havrda-Charvat et la deuxième section est consacrée aux expériences et à la comparaison entre les deux entropies.

## 2 Entropie

Comme nos travaux concernent uniquement la classification et prédiction binaires, nous nous détaillons alors le cas de variables aléatoires à valeurs dans un espace fini.

### 2.1 Entropie croisée de Havrda-Charvat

Comme détaillé dans [2], l'entropie de Shannon peut être généralisée de différentes manières. Une des façons de généraliser l'entropie de Shannon est de partir de son expression :

$$H(q) = - \sum_{i=1}^k h(q_i) \quad (1)$$

où  $h_1(u) = u \log(u)$  en remplaçant  $h_1$  par une fonction  $h$ , qui comme  $h_1$ , est convexe et telle que  $h(1) = 0$ . La fonction choisie pour l'entropie de Havrda-Charvat est :

$$h_\alpha(u) = \frac{u^\alpha - u}{\alpha - 1}, \quad (2)$$

où  $\alpha > 0$ . Celle-ci est alors donnée par :

$$H_\alpha(q) = \frac{1}{\alpha - 1} \times \left[ 1 - \sum_{i=1}^k q_i^\alpha \right] \quad (3)$$

L'entropie croisée associée est alors donnée par :

$$H_\alpha(q : p) = \frac{1}{\alpha - 1} \times \left[ 1 - \sum_{i=1}^k q_i^{\alpha-1} p_i^\alpha \right] \quad (4)$$

Etant donnée que  $p_i$  vaut soit 0 ou 1, alors cette formule se résume en :

$$H_\alpha(q : p) = \frac{1}{\alpha - 1} \times \left[ 1 - \sum_{i=1}^k q_i^{\alpha-1} p_i \right] \quad (5)$$

Comme pour l'entropie croisée classique, l'entropie croisée de Havrda-Charvat impose que la distribution prédite  $q$  soit la plus proche possible de  $p$  lorsque  $p$  est une distribution de Dirac.

En effet, si  $p = \delta_{i_0}$ , alors  $H_\alpha(q : p) = \frac{1 - q_{i_0}^{\alpha-1}}{\alpha-1}$  et la fonction  $u \rightarrow \frac{1-u^{\alpha-1}}{\alpha-1}$  est décroissante et son minimum est atteint à 1.

## 3 Architecture du réseau de neurones

L'architecture proposée pour la prédiction de la récurrence est un réseau neuronal multitâche de type U-Net comprenant deux branches capables d'effectuer des tâches de classification et de prédiction respectivement. L'architecture est représentée sur la figure 1.

Trois tâches principales sont menées conjointement par le ré-

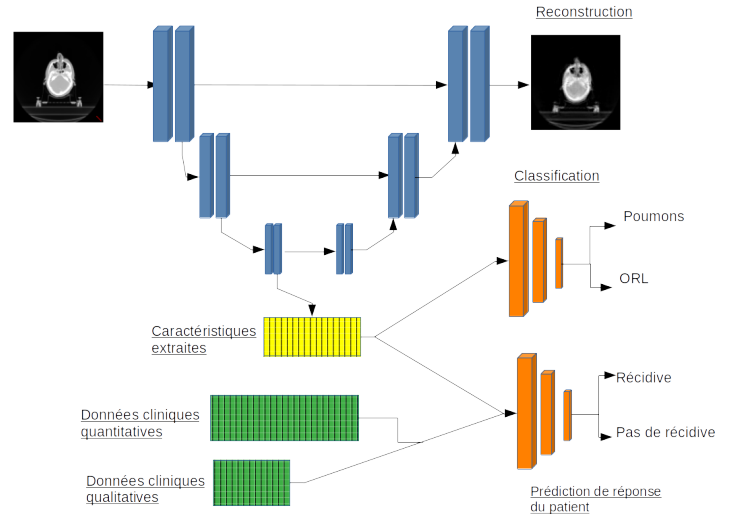


FIGURE 1 – Architecture du réseau neuronal multitâche pour la prédiction de la récurrence (T3) à l'aide de deux autres tâches (T1 : reconstruction d'image, T2 : classification du cancer)

seau. T1 est la tâche de reconstruction, spécifique à la partie U-Net du réseau. Il permet de déterminer que les caractéristiques extraites par la partie descendante du réseau sont à la

fois pertinentes pour la prédiction et la classification et représentatives de l'ensemble du scanner. La fonction de perte utilisée dans cette tâche est l'erreur quadratique moyenne. Il est défini comme suit.

$$L_{\text{rec}} = \frac{1}{N} \sum_{n=1}^N \|\vec{y}_n - \hat{y}_n\|^2, \quad (6)$$

où  $N$  est le nombre d'images TDM (nombre de patients),  $\vec{y}_n$  la vraie image et  $\hat{y}_n$  l'image estimée;  $\|\cdot\|$  étant la norme euclidienne.

T2 est la tâche de classification. Elle est utilisée pour déterminer, à partir des caractéristiques décrites précédemment et des données cliniques le type de cancer. Il fonctionne avec des couches entièrement connectées et aboutit à une classification binaire. La fonction de perte utilisée dans cette branche est l'entropie croisée de Shannon, telle que décrite dans l'équation 7.

$$L_{\text{classif}} = -\frac{1}{N} \sum_{n=1}^N [p_n \times \log(q_n) + (1 - p_n) \times \log(1 - q_n)], \quad (7)$$

où  $p_n = 1$  si le n-ième patient a un cancer ORL et  $p_n = 0$  s'il a un cancer du poumon,  $q_n \in [0, 1]$  est la probabilité prédite à la sortie du réseau qu'il s'agisse d'un cancer ORL.

T3 est la tâche de prédiction. Elle est utilisée pour déterminer, à partir de la même entrée que la classification, si le patient actuel risque de rencontrer une récurrence de son cancer. La fonction de perte de la tâche de prédiction, définie par (8) fera l'objet des tests. Nous déterminerons celui qui donne les résultats les plus précis.

$$L_{\text{pred},\alpha} = \frac{1}{\alpha - 1} \times \left[ 1 - \frac{1}{N} \sum_{i=1}^N (q_n^{\alpha-1} p_n' + (1 - q_n)^{\alpha-1} (1 - p_n')) \right], \quad (8)$$

où  $p_n' = 1$  si il y a récurrence pour le n-ième patient, 0 sinon et  $q_n'$  correspond à la probabilité de récurrence estimée par le réseau.

La fonction de perte totale du réseau est la somme des trois contributions :

$$L_{\text{total}} = L_{\text{rec}} + L_{\text{classif}} + L_{\text{pred},\alpha} \quad (9)$$

## 4 Expérimentations

Nous utilisons deux petites bases de données, l'une est composée de 434 patients souffrant d'un cancer ORL et l'autre de 146 patients souffrant d'un cancer du poumon. Les images TDM utilisées dans le réseau sont redimensionnées aux dimensions suivantes : (128 x 128 x 64 voxels). Quant aux informations sur les patients utilisées comme données d'entrée dans le réseau de neurones, elles sont de deux types : les informations quantitatives et les informations qualitatives. Parmi les données cliniques quantitatives figurent le taux d'hémoglobine (en g/dL), le taux de lymphocytes (en Giga/L), le taux de leucocytes (en Giga/L), le taux de thrombocytes (en Giga/L), le taux d'albumine (en g/L), la durée du traitement (en jours), la dose d'irradiation totale (en Gy), le nombre de fractions, la dose

moyenne par fractions (en Gy), le poids du patient au début et à la fin du traitement. Parmi les données qualitatives figurent le sexe du patient, sa consommation de tabac (fumeur, non fumeur ou ancien fumeur), si utilisation de chimiothérapie inductive (oui ou non), si utilisation de chimiothérapie concomitante (oui ou non) et TGM (tumeur, ganglions ou métastases).

### 4.1 Méthode d'évaluation

Étant donné que l'étude est menée sur de petites bases de données, une stratégie de validation des résultats est nécessaire. Nous proposons d'utiliser la validation croisée k-fold. La validation croisée est une procédure de rééchantillonnage utilisée pour évaluer des modèles sur des petites bases de données. Dans notre travail, nous utilisons une validation croisée quintuple. De plus, comme mesure d'évaluation, nous utilisons le taux de bonne prédiction. Nous calculons également la sensibilité et la spécificité de la prédiction. Voici la formule utilisée pour les trois calculs (10), (11), (12).

$$\text{Sensitivité} = \frac{VP}{VP + FN} \quad (10)$$

$$\text{Spécificité} = \frac{VN}{VN + FP} \quad (11)$$

$$\text{Taux de prédictions correctes (TPC)} = \frac{TP}{N} \quad (12)$$

où  $VP$  est le nombre de vrais positifs,  $FN$  celui des faux positifs,  $VN$  celui des vrais négatifs et  $FP$  celui des faux positifs.

### 4.2 Résultats

Les résultats obtenus lors des tests sont affichés dans cette section. Les images reconstruites sont principalement utilisées afin de vérifier la pertinence des caractéristiques extraites pour les autres tâches. Par conséquent, la performance de la reconstruction n'est pas très importante ici, car l'objectif est la prédiction de la récurrence. Les images originales et celles reconstruites sont présentées dans la figure 2.

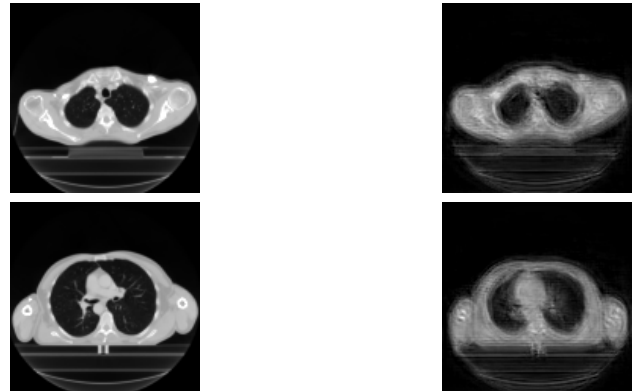


FIGURE 2 – Images d'entrées : images originales (gauche) vs images reconstruites (droite)

## 4.2.1 Entropie croisée de Havrda-Charvat

Concernant l'entropie croisée de Havrda-Charvat, il a été proposé d'étudier les variations de précision lorsque son hyperparamètre  $\alpha$  varie de 0.1 à 1.3. La  $p$ -value affichée mesure si les résultats acquis par Havrda-Charvat sont statistiquement différents de ceux obtenus en utilisant celle de Shannon avec une validation croisée quintuple. Deux conditions doivent être satisfaites pour que l'on accepte que l'entropie de Havrda-Charvat donne de meilleurs résultats que l'entropie de Shannon. Un seuil de 0.05 est pris pour la  $p$ -value, et la moyenne des résultats doit être cinq fois supérieure à celle obtenue avec Shannon. Les résultats sont décrits dans le tableau 1. Quand

TABLE 1 – Résultats obtenus par la fonction de perte issue de l'entropie croisée de Havrda-Charvat en fonction de  $\alpha$  ( $p$ -values inférieures à 0.05 sont soulignées en bleu).

$\alpha$	TPC	Ecarts-type	p-value
0.1	0.62	0.08	0.36
0.2	<b>0.69</b>	<b>0.03</b>	<b>0.01</b>
0.3	0.62	0.03	0.24
0.4	0.63	0.09	0.41
0.5	<b>0.66</b>	<b>0.01</b>	<b>0.02</b>
0.6	0.65	0.06	0.28
0.7	<b>0.67</b>	<b>0.04</b>	<b>0.02</b>
0.8	0.63	0.04	0.31
0.9	0.60	0.04	0.04
<b>1.0</b>	<b>0.64</b>	<b>0.02</b>	<b>N/A (Shannon's entropy)</b>
1.1	0.68	0.05	0.11
1.2	0.63	0.03	0.46
1.3	0.66	0.06	0.23

on regarde les résultats obtenus via Havrda-Charvat, on peut confirmer que pour la plupart des valeurs de  $\alpha$ , le taux d'erreur de prédiction n'est pas supérieur à celui obtenu par la fonction de perte de Shannon. On peut également observer que la fonction de perte issue de Havrda-Charvat peut donner de meilleurs résultats que celle de Shannon dans plusieurs cas. Cependant, il est difficile de savoir a priori quelle valeur de  $\alpha$  est bonne pour une application donnée. Son choix reste un défi, mais ici nous pouvons démontrer que l'entropie de Shannon n'est pas toujours le meilleur choix.

Pour démontrer l'avantage de notre réseau multi-tâches, nous proposons d'évaluer l'importance de chaque tâche du réseau avec Havrda-Charvat en utilisant le  $\alpha$  le plus favorable. Les résultats sont affichés dans le tableau 2.

TABLE 2 – Résultats obtenus pour différentes combinaisons de tâches avec Havrda-Charvat en prenant  $\alpha = 0.2$ .

Tâches	TPC	Sensitivity	Specificity
T2,T3	0.64	0.37	0.79
T1,T3	0.65	0.18	0.90
T1,T2,T3	0.69	0.52	0.77

## 5 Conclusion

Dans cet article, nous avons établi que, pour nos données et dans plusieurs cas, l'entropie croisée de Havrda-Charvat est plus performante que la fonction de perte de Shannon. La meilleure performance moyenne de Havrda-Charvat est de 69% de prédiction de récurrence correcte tandis que celle de Shannon est de 64%. Dans les applications médicales, même une amélioration de 1 ou 2% est intéressante. Cela donne la formule de Havrda-Charvat comme le meilleur candidat pour des recherches plus approfondies sur l'ensemble de données. Dans ces recherches, nous pourrions adapter l'entropie croisée binaire de Havrda-Charvat pour qu'elle devienne catégorique. Elle permettrait de réaliser des prédictions multi-classes, notamment en estimant la durée entre la fin du traitement du cancer et la récurrence. Un autre axe de développement pourrait être de trouver un moyen d'automatiser la détermination de la valeur appropriée de  $\alpha$  pour une application donnée.

## Références

- [1] Q. Wang, Y. Ma, K. Zhao, and et al., "A Comprehensive Survey of Loss Functions in Machine Learning," *Annals of Data Science*, 2020.
- [2] J. M. Amigó, S. G. Balogh, and S. Hernández, "A Brief Review of Generalized Entropies," *Entropy*, vol. 20, 2018.
- [3] U. Ruby and V. Yendapalli, "Binary cross entropy with deep learning technique for Image classification," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 9, 10 2020.
- [4] Valeria Andreieva and Nadiia Shvai, "Generalization of cross-entropy loss function for image classification," *Mohyla Mathematical Journal*, vol. 3, pp. 3–10, 01 2021.
- [5] R Roselin et al., "Mammogram image classification : Non-shannon entropy based ant-miner," *International Journal of Computational Intelligence and Informatics*, vol. 4, 2014.
- [6] T. Brochet, J. Lapuyade-Lahorgue, S. Bougleux, M. Salaün, and S. Ruan, "Deep learning using havrda-charvat entropy for classification of pulmonary optical endomicroscopy," *IRBM*, vol. 42, no. 6, pp. 400–406, 2021.
- [7] S. Kumar and G. Ram, "A Generalization of the Havrda-Charvat and Tsallis Entropy and Its Axiomatic Characterization," *Abstract and Applied Analysis*, vol. 2014, pp. 1–8, 2014.
- [8] A. Amyar, S. Ruan, I. Gardin, and et al., "3D RPET-NET : Development of a 3D PET Imaging Convolutional Neural Network for Radiomics Analysis and Outcome Prediction," *IEEE Trans. on Radiations and Plasma Medical Sciences*, vol. 3, no. 2, pp. 225–231, 2019.