




Article

A Quantitative Comparison between Shannon and Tsallis–Havrda–Charvat Entropies Applied to Cancer Outcome Prediction

Thibaud Brochet¹, Jérôme Lapuyade-Lahorgue¹, Alexandre Huat^{1,2,3}, Sébastien Thureau^{1,2} , David Pasquier⁴ , Isabelle Gardin^{1,2}, Romain Modzelewski^{1,2}, David Gibon³, Juliette Thariat⁵, Vincent Grégoire⁶, Pierre Vera^{1,2} and Su Ruan^{1,*} 

¹ LITIS, Quantif, University of Rouen, 76000 Rouen, France; thibaud.brochet@univ-rouen.fr (T.B.); jerome.lapuyade-lahorgue@univ-rouen.fr (J.L.-L.); alexandre.huat@univ-rouen.fr (A.H.); sebastien.thureau@chb.unicancer.fr (S.T.); isabelle.gardin@chb.unicancer.fr (I.G.); romain.modzelewski@chb.unicancer.fr (R.M.); pierre.vera@chb.unicancer.fr (P.V.)

² Centre Henri Becquerel, 76038 Rouen, France

³ Société Aquilab, 59120 Lille, France; david.gibon@aquilab.com

⁴ Département de Radiothérapie, Centre Oscar Lambret, 59000 Lille, France; d-pasquier@o-lambret.fr

⁵ Département de Radiothérapie, CLCC Francois Baclesse, 14000 Caen, France; juliette.thariat@cfb.unicancer.fr

⁶ Département de Radiothérapie, Centre Léon Berard, 69008 Lyon, France; vincent.gregoire@crcl.fr

* Correspondence: su.ruan@univ-rouen.fr



Citation: Brochet, T.;

Lapuyade-Lahorgue, J.; Huat, A.; Thureau, S.; Pasquier, D.; Gardin, I.; Modzelewski, R.; Gibon, D.; Thariat, J.; Grégoire, V.; et al. A Quantitative Comparison between Shannon and Tsallis–Havrda–Charvat Entropies Applied to Cancer Outcome Prediction. *Entropy* **2022**, *24*, 436. <https://doi.org/10.3390/e24040436>

Academic Editor: Anne Humeau-Heurtier

Received: 7 February 2022

Accepted: 18 March 2022

Published: 22 March 2022

Corrected: 13 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: In this paper, we propose to quantitatively compare loss functions based on parameterized Tsallis–Havrda–Charvat entropy and classical Shannon entropy for the training of a deep network in the case of small datasets which are usually encountered in medical applications. Shannon cross-entropy is widely used as a loss function for most neural networks applied to the segmentation, classification and detection of images. Shannon entropy is a particular case of Tsallis–Havrda–Charvat entropy. In this work, we compare these two entropies through a medical application for predicting recurrence in patients with head–neck and lung cancers after treatment. Based on both CT images and patient information, a multitask deep neural network is proposed to perform a recurrence prediction task using cross-entropy as a loss function and an image reconstruction task. Tsallis–Havrda–Charvat cross-entropy is a parameterized cross-entropy with the parameter α . Shannon entropy is a particular case of Tsallis–Havrda–Charvat entropy for $\alpha = 1$. The influence of this parameter on the final prediction results is studied. In this paper, the experiments are conducted on two datasets including in total 580 patients, of whom 434 suffered from head–neck cancers and 146 from lung cancers. The results show that Tsallis–Havrda–Charvat entropy can achieve better performance in terms of prediction accuracy with some values of α .

Keywords: deep neural networks; Shannon entropy; Tsallis–Havrda–Charvat entropy; generalized entropies; recurrence prediction; head–neck cancer; lung cancer

1. Introduction

This paper is devoted to studying the loss function based on Tsallis–Havrda–Charvat entropy in deep neural networks [1] for the prediction of outcomes in lung and head–neck cancers. When used for categorical prediction, the loss function is generally a cross-entropy related to a given entropy function. Indeed, cross-entropy-based loss functions are appropriate for evaluating how a probability distribution is close to the Dirac distribution. In deep neural networks, the output is a probability for each class obtained from a softmax activation function, and the Dirac distribution concentrated on one class represents the ground truth. There are several ways to compare these distributions [2–4]. Entropy-based metrics, such as divergences and cross-entropies, are the most common because they are the most appropriate way to sum up the informative content of a distribution, as explained

in [5]. In Ref. [6], different entropies are presented. In classification and prediction, the cross-entropy is derived from an entropy measure and used as a loss function measuring the difference between the predicted probability and the real Dirac probability. To sum up, in most neural networks used for prediction, Shannon-related cross-entropy is the most common and widely used for segmentation [7], classification [8], or detection [9] and many other applications [10–13]. The reason why Shannon is the most used is twofold: first, because Shannon was the first entropy in the domain of information theory, and secondly, because it is extensive in the sense that the entropy of a multivariate distribution whose margins are independent is the sum of the marginal entropies. This last property makes the calculation of Shannon entropy easy. In Ref. [14], different ways of choosing an entropy and an associated divergence are detailed. Among them, Shannon entropy can be extended by replacing the logarithm by another function. Cross-entropies can be defined by replacing the counting measure (resp. Lebesgue measure for continuous case) by a Radon–Nykodim derivative between probability measures. Shannon’s entropy can be generalized on other entropies such as Renyi [15] and Tsallis–Havrda–Charvat [16,17]. In this paper, we are interested in a particular generalization of Shannon cross-entropy: Tsallis–Havrda–Charvat cross-entropy [18]. This class of entropies has the particularity of being parametrized with one parameter α and we recover Shannon entropy when the value of the parameter equal to 1. The relevance and possibilities of Tsallis–Havrda–Charvat in the medical field have been discussed before and this paper expands on the Tsallis–Havrda–Charvat formula.

Tsallis–Havrda–Charvat entropy was introduced independently by Tsallis [19] in the context of statistical physics and by Havrda and Charvat [20] in the context of information theory. Tsallis–Havrda–Charvat entropy has been used in publications in several fields, including medical imaging [21,22]. Tsallis–Havrda–Charvat entropy has rarely been used in deep learning, especially because of the difficulties with interpreting the hyperparameters α . However, there exist some scientific articles on this issue. In Ref. [23], Tsallis–Havrda–Charvat entropy is used to reduce the loss while classifying an image. In Ref. [18], the authors define Tsallis–Havrda–Charvat entropy in terms of axiomatization and propose a generalization based on this. In Ref. [24], the maximization of the entropy measure is studied for different classes of entropies such as Tsallis–Havrda–Charvat; the maximization of Tsallis–Havrda–Charvat entropy under constraints appears to be a way to generalize Gaussian distributions. In our previous work [17], Tsallis–Havrda–Charvat cross-entropy is used for the detection of noisy images in pulmonary microendoscopy. To capitalize and improve on this previous work, we propose to use deep learning in this paper. It allows for taking advantage of the previously used architecture by tuning and improving it to achieve better results. Deep learning is also very relevant for our field of study.

Deep learning has been widely developed in the medical field for classification or segmentation tasks [25–27]. Classification can be used to identify automatically the kind of cancer from which the patient is suffering [28,29] or the relevant outcomes after treatment, such as survival expectation [30] or relation to the treatment [31]. Recurrence in cancer after treatment is one of the main concerns for physicians [32], as it can dramatically impact the outcome for patients and their life expectancy. It would be beneficial for treatment selection if one could predict whether a recurrence will occur. Some studies have been carried out using CT scan images and clinical data. To our knowledge, there is no article using Tsallis–Havrda–Charvat for recurrence prediction [17]. The novelty of this article lies in the performance comparison of Shannon and Tsallis–Havrda–Charvat entropies in the context of cancer recurrence prediction with CT scan data combined with clinical information for patients affected by head and neck (H&N) or lung cancers (examples of the analyzed CT images are displayed in Figure 1). Moreover, we decided to study the parameter value in particular to examine its impact in order to predict these recurrences in both kinds of cancer. As medical data are generally scarce, the choice of a good entropy is important even if it can improve the performance only by 1 or 2 percent.

The paper is organized as follows. In the first section, we recall how categorical Shannon cross-entropy is defined and how it can be generalized for Tsallis–Havrda–Charvat. The second section is devoted to experiments and a comparison between the two entropies.

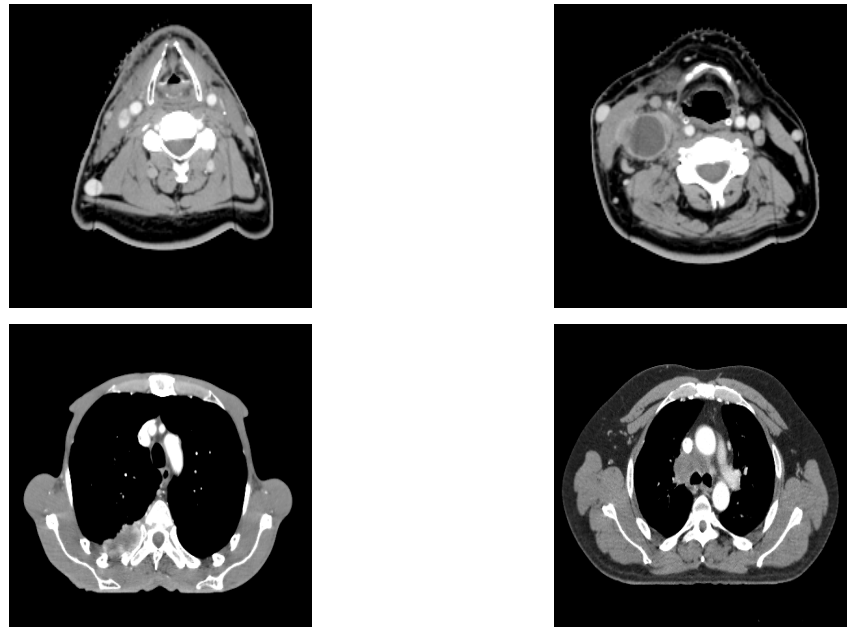


Figure 1. Input images: head–neck CT (**above**) and lung CT (**below**).

2. Entropy

As our problem concerns binary prediction, we focused only on finite-state random variables whose state-space was provided with the counting measure. Obviously, these results can be generalized for the finite-dimensional vectorial space \mathbb{R}^n provided with the Lebesgue measure.

2.1. Shannon Entropy and Related Cross-Entropy

For a discrete random variable Y , taking its values in $\Omega = \{1, \dots, k\}$ with respective probabilities p_1, \dots, p_k , Shannon entropy is defined by:

$$H(p) = - \sum_{i=1}^k \log(p_i) \times p_i \quad (1)$$

Shannon is minimal and equal to 0 if $p_i = 1$ for one i and 0 otherwise; it is maximal if Y is uniformly distributed. The corresponding cross-entropy is given by:

$$H(q : p) = - \sum_{i=1}^k \log(q_i) p_i \quad (2)$$

Generally, in a classification problem, the true distribution p is a Dirac distribution δ_{i_0} , where i_0 is the class of the data. In this case, the cross-entropy is $H(p; q) = -\log(q_{i_0}) \geq 0$. As a consequence, q_{i_0} is closer to 0 and $H(p; q)$ is higher. Furthermore, minimizing the cross-entropy forces q to be as close as possible to the distribution p , and as this last one is the Dirac distribution δ_{i_0} , the minimum of the cross-entropy is 0.

2.2. Tsallis–Havrda–Charvat Cross-Entropy

There are several ways to generalize Shannon entropy, as explained in [14]. Shannon entropy can be expressed by:

$$H(q) = - \sum_{i=1}^k h(q_i) \tag{3}$$

where $h(u) = u \log(u)$. h is a convex function such that $h(1) = 0$. The idea is to choose another function satisfying the same properties. Tsallis–Havrda–Charvat entropy is defined by choosing:

$$h_\alpha(u) = \frac{u^\alpha - u}{\alpha - 1}, \tag{4}$$

where $\alpha > 0$ and is given by:

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

The associated cross-entropy is given by:

$$H_\alpha(q : p) = \frac{1}{\alpha - 1} \times \left[1 - \sum_{i=1}^k q_i^{\alpha-1} p_i \right] \tag{6}$$

As for classical cross-entropy, Tsallis–Havrda–Charvat cross-entropy forces the predicted distribution q to be as close as possible to p when p is a Dirac distribution.

3. Neural Network Architecture for Recurrence Prediction

The proposed architecture used for recurrence prediction is a multitask neural network with a U-Net backbone, with one branch able to perform prediction tasks and another to reconstruct the input image for extracting features to help prediction. The architecture is presented in Figure 2.

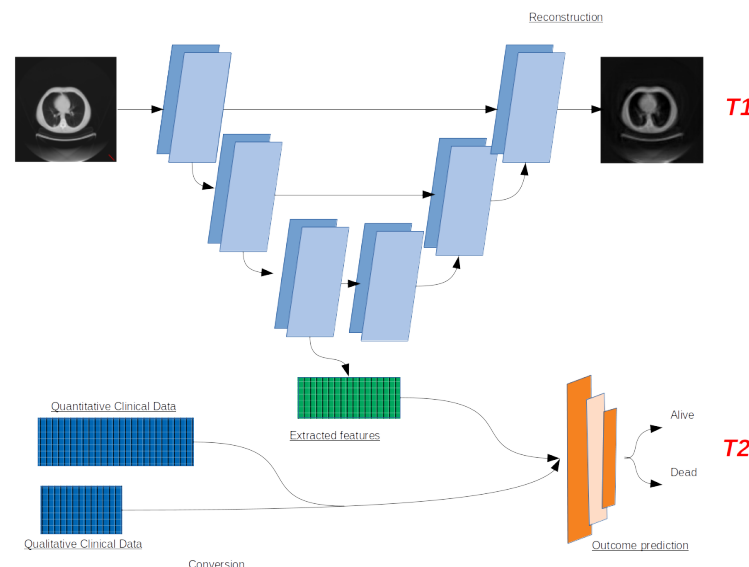


Figure 2. Architecture of the multitask neural network for recurrence prediction (T2) with the help of another task (T1: image reconstruction).

The U-Net backbone is composed of three convolutional layers skipped by concatenations to add the features extracted from descending convolutions to the ascending ones. Within the convolution layers, we use ReLU activation functions.

At the bottom of the network, the extracted features are sent as inputs to one branch of fully connected layers in charge of making a decision path to determine whether the patient is at risk of recurrence.

Two main tasks are jointly carried out by the network. T1 is the reconstruction task, specific to the U-Net part of the network. It allows for determining whether the features extracted by the descending part of the U are relevant for prediction and classification and representative of the whole CT scan at the same time. The loss function used in this task is the mean squared error. It is defined as follows:

$$L_{\text{rec}} = \frac{1}{N} \sum_{n=1}^N \|\vec{y}_n - \hat{y}_n\|^2, \quad (7)$$

where \vec{y}_n represents the true data values for the n -th patient and \hat{y}_n is the estimated output from the network, with $\|\cdot\|$ being the Euclidean norm and N the number of patients.

The mean squared error computes the squared distance between the predicted output and the input image volumes. This function allows for comparing the predicted images and the true ones voxel-by-voxel and to train the network to recompose the images from the extracted features.

T2 is the prediction task. It is used to determine, from the same input, whether the current patient is at risk of encountering a recurrence of their cancer. It is constructed with fully connected layers and ends up in a binary classification. The prediction task's loss function was the subject of our tests. We compared two entropies through this task.

The first, and most commonly used, was Shannon's binary cross-entropy.

$$L_{\text{pred},1} = -\frac{1}{N} \sum_{n=1}^N [p_n \times \log(q_n) + (1 - p_n) \times \log(1 - q_n)], \quad (8)$$

where p_n is the true class, $p_n = 1$ if recurrence, $p_n = 0$ otherwise, and q_n is the estimated probability of recurrence, with N being the number of patients.

The second entropy was the generalized formula, Tsallis–Havrda–Charvat binary cross-entropy.

$$L_{\text{pred},\alpha} = \frac{1}{\alpha - 1} \times \left[1 - \frac{1}{N} \sum_{n=1}^N (q_n^{\alpha-1} p_n + (1 - q_n)^{\alpha-1} (1 - p_n)) \right] \quad (9)$$

Binary cross-entropies are loss functions that are able to compare binary predictions with ground truths, which makes them relevant for our binary labels.

The total loss function of the network is the sum of the two losses. The choice of this total loss function was motivated by different experiments in which we used different weights for the individual loss functions. It appears that the sum with equal weights provided the best results.

$$L_{\text{total}} = L_{\text{rec}} + L_{\text{pred},\alpha} \quad (10)$$

The prediction branch is the subject of interest for this article, with the reconstruction task being used to help the prediction in the feature extraction step.

Regarding the execution time and its change owing to the increasing or decreasing complexity of the network, we proposed to lock our network at a certain level of complexity and find an acceptable execution time of 5–7 h for a number of epochs of 100. This was considered to be a good compromise between obtaining a sufficient number of epochs to achieve meaningful results and not obtaining so many as to cause overfitting. We decided to use convolution layers in our U-Net backbone because of the limited computational power of available machines and thus the need to limit the number of parameters in the network. Allowing for a few hours of computations enabled the experiments to be run at night so that the results could be available and ready for analysis the next day.

4. Experiments

4.1. Datasets

The datasets were composed of 580 patients, among which 434 suffered from head–neck cancer and 146 from lung cancer. Both datasets were small in size. We chose to conduct experiments on the two subsets (head–neck and lung) separately. Indeed, the optimal value of α depends on the kind of data used. Moreover, we had already tested the total dataset of 580 and the results were poor. We therefore chose to show only the results for separate datasets.

CT images used in the neural network were resized with an image resolution of $128 \times 128 \times 64$ voxels. The patient information used as input data in the neural network was of two kinds, namely quantitative and qualitative, as shown in Tables 1 and 2.

Table 1. Quantitative clinical data processed through the network.

Clinical Data	Modality
Hemoglobin	g/dL
Lymphocytes	Giga/L
Leucocytes	Giga/L
Thrombocytes	Giga/L
Albumin	g/L
Treatment duration	Days
Total irradiation dose	Gy
Number of fractions	/
Average dose per fraction	Gy
Weight at the start and end of treatment	kg

Table 2. Qualitative clinical data processed through the network.

Clinical Data	Modality
Gender	M/F
Tabacology	Smoker, non-smoker, former smoker
Use of induction chemotherapy	Yes/No
Use of concomitant chemotherapy	Yes/No
TNM	Tumor, Node, Metastasis

Our experiments consisted of comparing the accuracy of Tsallis–Havrda–Charvat and Shannon for both datasets.

4.2. Evaluation Method

Since the study was conducted on small datasets (434 and 146 patients), a result validation strategy was required. We proposed to use the k-fold cross-validation relevant for small data validation. In our work, we used a five-fold cross-validation.

The procedure unfolds as follows:

- Shuffle the dataset;
- Split it randomly into 5 subsets;
- For each subset:
 - Take the subset as a test dataset;
 - Take the other sets as training data;
 - Fit the model to the training data and evaluate it on the test dataset;
 - Retain the evaluation score;
- Summarize the skill of the model from the samples of model evaluation scores.

Furthermore, accuracy was proposed as a metric for evaluation. It consisted, in this case, in comparing the values of the ground truth and the prediction and summing all these occurrences over the size of the dataset.

4.3. Results

The results achieved during the tests are displayed in this section. Reconstructed images are mainly used in order to show the relevance of extracted features for the prediction. Therefore, their performance is not very important here, because the objective is the prediction of recurrence. The original images and reconstructed ones are featured in Figure 3. We can see that the reconstructed images are similar to the input images, meaning that our network is able to recover input images.

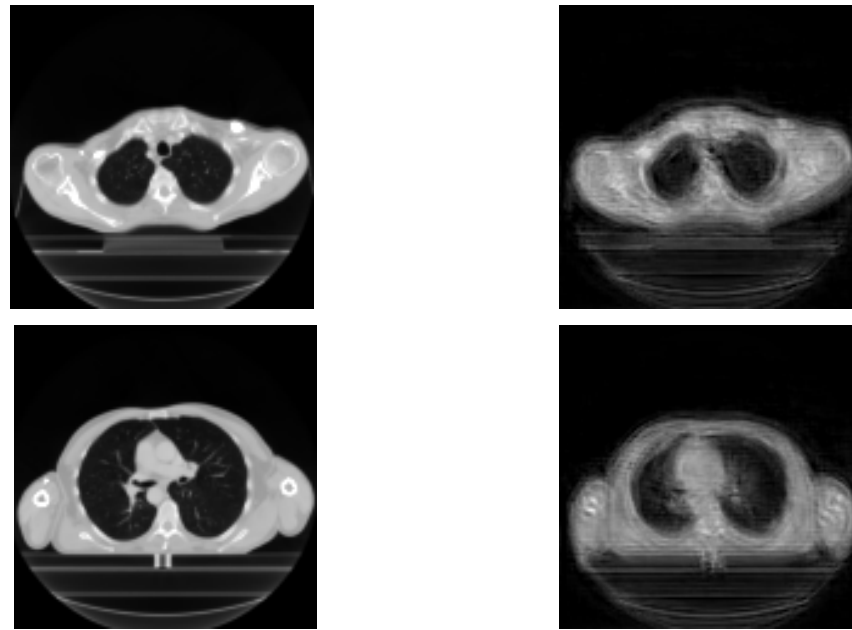


Figure 3. Input images: original images (left) vs. reconstructed images (right).

The loss of information generates uncertainty in each image. A possible improvement would be to use a fuzzy image processor to improve the quality of the obtained images, as described in [33].

Comparison Results

Regarding Tsallis–Havrda–Charvat, we studied its hyperparameter α as varying from 0.1 to 2.0. When $\alpha = 1$, this entropy corresponds to Shannon entropy. The displayed p -value measures whether the results acquired by Tsallis–Havrda–Charvat five-fold cross-validation are statistically different from Shannon’s. Two conditions must be satisfied in order to accept the Tsallis–Havrda–Charvat entropy as providing better results than Shannon entropy: the average of the five-fold results must be superior to Shannon and the p -value must be smaller than 0.05. The results are described in the following tables.

The results achieved for the dataset of head–neck cancers are described in Table 3.

Regarding the dataset containing lung cancers, the results achieved are described in Table 4.

After fine-tuning, we obtained a set of optimal hyperparameters.

The results achieved via the Tsallis–Havrda–Charvat formula confirm that, for most values of the hyperparameter α , the final accuracy is not superior to the accuracy obtained by Shannon’s loss function. It can also be observed that the loss function derived from Havrda–Charvat equation can provide better results than Shannon’s in some cases. However, it is difficult to know a priori what value of α is good for an application. Its choice is still a challenge.

Table 3. Accuracy obtained by loss function derived from Tsallis–Havrda–Charvat entropy in a function of α for the head–neck cancer dataset (p -values lower than 0.05 and accuracies higher than Shannon’s are highlighted in blue).

α	5-Fold					Average	SD	p -Value
	1	2	3	4	5			
0.1	0.68	0.53	0.6	0.58	0.63	0.60	0.06	0.01
0.3	0.60	0.70	0.70	0.70	0.43	0.63	0.12	0.28
0.5	0.58	0.58	0.60	0.70	0.73	0.64	0.07	0.27
0.7	0.85	0.70	0.60	0.70	0.65	0.70	0.09	0.25
0.9	0.58	0.60	0.60	0.60	0.68	0.61	0.04	0.07
1.0	0.75	0.65	0.68	0.58	0.70	0.67	0.06	N/A (Shannon entropy)
1.1	0.68	0.75	0.75	0.73	0.75	0.73	0.03	0.09
1.3	0.63	0.73	0.68	0.75	0.70	0.70	0.05	0.32
1.5	0.70	0.78	0.85	0.80	0.88	0.80	0.07	0.03
1.7	0.80	0.73	0.63	0.75	0.78	0.74	0.07	0.07
1.9	0.75	0.73	0.73	0.75	0.83	0.76	0.05	0.02
2.1	0.68	0.63	0.60	0.62	0.575	0.63	0.04	0.12
2.3	0.73	0.73	0.73	0.73	0.7	0.72	0.01	0.09
2.5	0.75	0.58	0.68	0.68	0.6	0.66	0.07	0.34
2.7	0.68	0.53	0.45	0.63	0.6	0.58	0.09	0.04
2.9	0.73	0.73	0.73	0.75	0.73	0.73	0.01	0.07
3.1	0.7	0.55	0.65	0.57	0.63	0.62	0.06	0.02
3.3	0.65	0.65	0.65	0.58	0.55	0.62	0.05	0.07
3.5	0.73	0.75	0.73	0.75	0.70	0.73	0.02	0.08
3.7	0.73	0.70	0.60	0.60	0.58	0.64	0.07	0.20
3.9	0.68	0.70	0.55	0.58	0.53	0.61	0.08	0.09

Table 4. Accuracy obtained by loss function derived from Tsallis–Havrda–Charvat entropy in a function of α for the lung cancer dataset (p -values lower than 0.05 and accuracies higher than Shannon’s are highlighted in blue).

α	5-Fold					Average	SD	p -Value
	1	2	3	4	5			
0.1	0.58	0.58	0.47	0.58	0.63	0.57	0.06	0.23
0.3	0.58	0.58	0.58	0.58	0.68	0.60	0.04	0.13
0.5	0.63	0.58	0.52	0.58	0.52	0.56	0.03	0.26
0.7	0.58	0.58	0.63	0.63	0.58	0.60	0.03	0.13
0.9	0.63	0.53	0.68	0.47	0.53	0.57	0.08	0.21
1.0	0.73	0.47	0.47	0.47	0.47	0.52	0.12	N/A (Shannon entropy)
1.1	0.63	0.63	0.52	0.52	0.52	0.56	0.06	0.18
1.3	0.68	0.53	0.53	0.47	0.53	0.55	0.08	0.15
1.5	0.58	0.53	0.53	0.47	0.53	0.53	0.04	0.44
1.7	0.73	0.47	0.63	0.57	0.42	0.56	0.12	0.37
1.9	0.69	0.63	0.58	0.53	0.68	0.62	0.07	0.04
2.1	0.79	0.84	0.73	0.79	0.79	0.79	0.04	0.004
2.3	0.84	0.78	0.84	0.73	0.84	0.81	0.05	0.002
2.5	0.79	0.84	0.74	0.68	0.63	0.74	0.08	0.007
2.7	0.79	0.74	0.69	0.74	0.74	0.74	0.04	0.003
2.9	0.79	0.79	0.79	0.79	0.73	0.78	0.03	0.004
3.1	0.74	0.74	0.78	0.78	0.68	0.74	0.04	0.008
3.3	0.79	0.79	0.74	0.74	0.74	0.76	0.03	0.003
3.5	0.74	0.73	0.68	0.33	0.58	0.61	0.17	0.12
3.7	0.68	0.63	0.47	0.63	0.73	0.63	0.09	0.07
3.9	0.78	0.53	0.47	0.47	0.63	0.58	0.13	0.07

We highlighted the results providing both better accuracy and a significant p -value in blue. The most promising values were achieved for α equal to 1.5 with head–neck cancers and between 1.9 and 3.5 with lung cancer. In this regard, we can state that the results obtained by Tsallis–Havrda–Charvat can be significantly improved compared to Shannon.

When analyzing the lung cancer results, a plateau can be easily noticed where Tsallis–Havrda–Charvat achieves better results than Shannon. This involves a specific set of values of α , from 1.9 to 3.5, in which the Tsallis–Havrda–Charvat loss function is significantly more efficient than Shannon’s.

5. Discussion

It has been determined that the Tsallis–Havrda–Charvat loss function performs equally well or better than Shannon cross-entropy, depending on the value of its hyperparameter α . It can be said that the Tsallis–Havrda–Charvat loss function, depending on the value of its hyperparameter α , can fit a wider array of input data and can potentially yield better results.

Conversely, we can state that, based on the calculated p-values and standard deviations, Tsallis–Havrda–Charvat entropy seems more unstable than Shannon entropy, as its standard deviation may reach 0.12, where Shannon’s is only 0.06. Furthermore, when looking at the p-values for several values of α , the results of Havrda–Charvat are not statistically different from Shannon’s. The instability of the results obtained using Tsallis–Havrda–Charvat could be explained by the fact that, despite the five-fold method, the data are still too scarce to reach a stable answer. In addition, data are tridimensional, making it more difficult than with 2D images to extract relevant features and thereby complicating the network’s tasks even further. In the analysis of 3D images, multiple slices must be taken into account in order to make a decision, which drastically increases the number of variables to be learned by the neural network. Moreover, in order to be usable for analysis by the neural network, the data are all supposed to have the same size. This is why, as for reasons of computational power, it was decided to have the data resized to $128 \times 128 \times 64$ voxels. This size, despite still being full of information for the network, implies a loss of information for wider, larger and deeper images.

Nevertheless, the value of α plays a large part in the behavior of the loss function, and it is the key element that can be set to fit the input data, but the questions remains: what α fits which data? The choice of the value of the hyperparameter α remains a challenge as it depends strongly on the kind of data. In perspective, it would be interesting to develop an algorithm for selecting automatically the value of this hyperparameter in order to fit it as accurately as possible to the data. The aim is to reach the plateau, or area, of α where the Tsallis–Havrda–Charvat loss function provides consistently better results and features a smaller SD and p-value. A further analysis needs to be conducted on the link between input data and the location of the best α area in order to determine the kind of extracted feature and the kind of neuronal path that cause one area to be more efficient than another. For instance, in our case, the question is about which key feature of the input lung cancer images leads to better values between 1.9 and 3.5 and which key feature of the input H&N cancer images lead to better results for α equal to 1.5.

6. Conclusions

In this article, we established that, for our data and in some cases, Tsallis–Havrda–Charvat cross-entropy performs better than a Shannon-based loss function. Tsallis–Havrda–Charvat performed best on the head–neck dataset and lung dataset, at 80% and 81%, respectively, of correct recurrence prediction, while Shannon’s results for these two datasets were 67% and 52%, respectively. This makes the Tsallis–Havrda–Charvat formula the best candidate for further research on these datasets. For further research we might adapt Tsallis–Havrda–Charvat binary cross-entropy to a categorical cross-entropy. This would allow for making multi-class predictions, including estimating the time between the end of cancer treatment and recurrence. Another axis of evolution could be finding a way to automatically determine the proper value of α for a given application.

Author Contributions: Conceptualization, S.R. and J.L.-L.; methodology, S.R., T.B. and J.L.-L.; software, T.B. and J.L.-L.; validation, S.R.; formal analysis, T.B.; writing—original draft preparation, T.B.; writing—review and editing, J.L.-L. and S.R.; supervision, S.R.; project administration, S.R.; data curation, A.H., S.T., D.P., I.G., R.M., D.G., J.T., V.G. and P.V.; resources, S.T., D.P., I.G., R.M., J.T., V.G. and P.V.; medical expertise, P.V.; investigation, T.B., J.L.-L., A.H., S.T., D.P., I.G., R.M., D.G., J.T., V.G., P.V. and S.R. All authors have read and agreed to the published version of the manuscript.

Funding: This project was co-financed by the European Union with the European regional development fund (ERDF, 18P03390/18E01750/18P02733), by the Haute-Normandie Regional Council via the M2SINUM project and by the PRPO project from the Cancéropôle Nord-Ouest, France.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data are not publicly available due to being the property of Centre Henri Becquerel.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

MDPI	Multidisciplinary Digital Publishing Institute
DOAJ	Directory of open access journals
TLA	Three-letter acronym
LD	Linear dichroism
H&N	Head–neck
H-Cd	Havrda–Charvat
Gy	Gray

References

1. Wang, Q.; Ma, Y.; Zhao, K.; Tian, Y. A Comprehensive Survey of Loss Functions in Machine Learning. *Ann. Data Sci.* **2020**, *9*, 187–212. [[CrossRef](#)]
2. Chung, J.K.; Kannappan, C.T.; Sahoo, P.K. Measures of distance between probability distributions. *J. Math. Anal. Appl.* **1989**, *138*, 280–292. [[CrossRef](#)]
3. Budzynski, R.J.; Kontracki, W.; Krolak, A. Applications of distance between distributions to gravitational wave data analysis. *Class. Quantum Gravity* **2007**, *25*, 015005. [[CrossRef](#)]
4. Serrurier, M. An informational distance for estimating the faithfulness of a possibility distribution, viewed as a family of probability distributions, with respect to data. *Int. J. Approx. Reason.* **2013**, *54*, 919–933. [[CrossRef](#)]
5. Kullback, S.; Leibler, R.A. On information and sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86. [[CrossRef](#)]
6. Amigó, J.M.; Balogh, S.G.; Hernández, S. A Brief Review of Generalized Entropies. *Entropy* **2018**, *20*, 813. [[CrossRef](#)] [[PubMed](#)]
7. Ma, Y.; Liu, Q.; Qian, Z.-B. Automated Image Segmentation Using Improved PCNN Model Based on Cross-entropy. In Proceedings of the 2004 International Symposium on Intelligent Multimedia, Video and Speech Processing, Hong Kong, China, 20–22 October 2004.
8. Mannor, S.; Peleg, D.; Rubinstein, R. The cross entropy method for classification. In Proceedings of the 22nd International Conference on Machine Learning, Bonn, Germany, 7–11 August 2005.
9. Qu, Z.; Mei, J.; Liu, L.; Zhou, D.Y. Crack Detection of Concrete Pavement With Cross-Entropy Loss Function and Improved VGG16 Network Model. In Proceedings of the 30th International Telecommunication Networks and Applications Conference (ITNAC), Melbourne, Australia, 25–27 November 2020.
10. Silva, L.; Marques de Sá, J.; Alexandre, L.A. Neural network classification using Shannon’s entropy. In Proceedings of the ESANN 2005 Proceedings—13th European Symposium on Artificial Neural Networks, Bruges, Belgium, 27–29 April 2005; pp. 217–222.
11. Rajinikanth, V.; Thanaraj, K.P.; Satapathy, S.C.; Fernandes, S.L.; Dey, N. Shannon’s Entropy and Watershed Algorithm Based Technique to Inspect Ischemic Stroke Wound. In *Smart Intelligent Computing and Applications*; Springer: Singapore, 2018; Volume 2, pp. 23–31.
12. Ruby, U.; Yendapalli, V. Binary cross entropy with deep learning technique for Image classification. *Int. J. Adv. Trends Comput. Sci. Eng.* **2020**, *9*, 5393–5397. [[CrossRef](#)]
13. Ramos, D.; Franco-Pedroso, J.; Lozano-Diez, A.; Gonzalez-Rodriguez, J. Deconstructing Cross-Entropy for Probabilistic Binary Classifiers. *Entropy* **2018**, *20*, 208. [[CrossRef](#)] [[PubMed](#)]
14. Basseville, M. *Information: Entropies, Divergences et Moyenne*; Technical Report; INRIA: Rennes, France, 2010.

15. Andreieva, V.; Shvai, N. Generalization of Cross-Entropy Loss Function for Image Classification. *Mohyla Math. J.* **2021**, *3*, 3–10. [[CrossRef](#)]
16. Roselin, R. Mammogram Image Classification: Non-Shannon Entropy based Ant-Miner. *Int. J. Comput. Intell. Inf.* **2014**, *4*, 33–43.
17. Brochet, T.; Lapuyade-Lahorgue, J.; Bougleux, S.; Salaün, M.; Ruan, S. Deep Learning Using Havrda-Charvat Entropy for Classification of Pulmonary Optical Endomicroscopy. *IRBM* **2021**, *42*, 400–406. [[CrossRef](#)]
18. Kumar, S.; Ram, G. A Generalization of the Havrda-Charvat and Tsallis Entropy and Its Axiomatic Characterization. *Abstr. Appl. Anal.* **2014**, *2014*, 505184. [[CrossRef](#)]
19. Tsallis, C. Possible generalization of Boltzmann-Gibbs statistics. *J. Stat. Phys.* **1988**, *52*, 479–487. [[CrossRef](#)]
20. Havrda, J.; Charvat, F. Quantification method of classification processes. Concept of structural alpha-entropy. *Kybernetika* **1967**, *3*, 30–35.
21. Khader, M.; Ben Hamza, A. An Entropy-Based Technique for Nonrigid Medical Image Alignment. In Proceedings of the 14th International Conference on Combinatorial Image Analysis (IWCIA'11), Madrid, Spain, 23–25 May 2011; Springer: Berlin/Heidelberg, Germany, 2011; pp. 444–455.
22. Waleed, M.; Ben Hamza, A. Medical image registration using stochastic optimization. *Opt. Lasers Eng.* **2010**, *48*, 1213–1223. [[CrossRef](#)]
23. Ramezani, Z.; Pourdarvish, A. Transfer learning using Tsallis entropy: An application to Gravity Spy. *Phys. A Stat. Mech. Appl.* **2021**, *561*, 125273. [[CrossRef](#)]
24. Karmeshu, J. (Ed.) *Entropy Measures, Maximum Entropy Principle and Emerging Applications*; Springer: Berlin/Heidelberg, Germany, 2003.
25. Zhou, T.; Canu, S.; Vera, P.; Ruan, S. Latent Correlation Representation Learning for Brain Tumor Segmentation with Missing MRI Modalities. *IEEE Trans. Image Process.* **2021**, *30*, 4263–4274. [[CrossRef](#)]
26. Amyar, A.; Modzelewski, R.; Li, H.; Ruan, S. Multi-task Deep Learning Based CT Imaging Analysis For COVID-19 Pneumonia: Classification and Segmentation. *Comput. Biol. Med.* **2020**, *126*, 104037. [[CrossRef](#)]
27. Jin, L.; Shi, F.; Chun, Q.; Chen, H.; Ma, Y.; Wu, S.; Hameed, N.U.F.; Mei, C.; Lu, J.; Zhang, J.; et al. Artificial intelligence neuropathologist for glioma classification using deep learning on hematoxylin and eosin stained slide images and molecular markers. *Neuro-Oncology* **2021**, *23*, 44–52. [[CrossRef](#)]
28. Sirinukunwattana, K.; Domingo, E.; Richman, S.D.; Redmond, K.L.; Blake, A.; Verrill, C.; Leedham, S.J.; Chatzipli, A.; Hardy, C.; Whalley, C.M.; et al. Image-based consensus molecular subtype (imCMS) classification of colorectal cancer using deep learning. *Gut* **2021**, *70*, 544–554. [[CrossRef](#)]
29. Noorbakhsh, J.; Farahmand, S.; Foroughi, P.; Namburi, S.; Caruana, D.; Rimm, D.; Soltanieh-Ha, M.; Zarringhalam, K.; Chuang, J.H. Deep learning-based cross-classifications reveal conserved spatial behaviors within tumor histological images. *Nat. Commun.* **2020**, *11*, 6367. [[CrossRef](#)]
30. Doppalapudi, S.; Qiu, R.G.; Badr, Y. Lung cancer survival period prediction and understanding: Deep learning approaches. *Int. J. Med. Inform.* **2021**, *148*, 104371. [[CrossRef](#)] [[PubMed](#)]
31. Amyar, A.; Ruan, S.; Gardin, I.; Chatelain, C.; Decazes, P.; Modzelewski, R. 3D RPET-NET: Development of a 3D PET Imaging Convolutional Neural Network for Radiomics Analysis and Outcome Prediction. *IEEE Trans. Radiat. Plasma Med. Sci.* **2019**, *3*, 225–231. [[CrossRef](#)]
32. Jiao, X.; Wang, Y.; Wang, F.; Wang, X. Recurrence pattern and its predictors for advanced gastric cancer after total gastrectomy. *Medicine* **2020**, *99*, e23795. [[CrossRef](#)] [[PubMed](#)]
33. Versaci, M.; Calcagno, S.; Morabito, F.C. Image Contrast Enhancement by Distances Among Points in Fuzzy Hyper-Cubes. In Proceedings of the Part II 16th International Conference on Computer Analysis of Images and Patterns, CAIP 2015, Valletta, Malta, 2–4 September 2015; Springer: Berlin/Heidelberg, Germany, 2015; Volume 9257, pp. 494–505. [[CrossRef](#)]