Warren F
Cardiff University
School of Medicine

Spezi E
Cardiff University
School of Engineering

Paisey S
Cardiff University
School of Medicine

Smith R
Cardiff University
School of Medicine

Lai Y
Cardiff University
School of Computer Science and Informatics

AI AND DEEP LEARNING

The Use of Mutual Information and Entropy as an Acquisition Function to be Used in Active Learning

This work has explored the theoretical principles of Mutual Information and Entropy and their use as combined acquisition score. The work looked at mutual information and entropy combined score values produced from a deep learning model trained to segment preclinical CT scans, which supported the investigation into its use in an active learning pipeline as an information metric, to guide the selection of data from the unlabelled data pool to incorporate into the training set. The work has shown that the combined acquisition function shows promise. Further refinement and validation are necessary to fully establish its utility in active learning tasks.

Keywords:
Mutual information, active learning, image segmentation, acquisition function, entropy.

Corresponding author:
WarrenF2@cardiff.ac.uk

F. Warren, S. Paisey, Y. Lai, E. Spezi, and R. Smith, 'The Use of Mutual Information and Entropy as an Acquisition Function to be Used in Active Learning', *Proceedings of the Cardiff University School of Engineering Research Conference 2024*, Cardiff, UK, 2024, pp. 19-22.

doi.org/10.18573/conf3.f

INTRODUCTION

Medical image segmentation is a fundamental and necessary task performed when analysing CT images. Deep learning has been shown to be useful in medical image segmentation. However, it faces issues including the requirement of large amounts of labelled data to act as the ground truth reference during training. Obtaining annotated data by experts is an expensive and time-consuming task [1]. An approach used to reduce the need for large amounts of annotated data is Active Learning (AL) [2].

The goal of AL is to select the most valuable and informative data from an unlabelled data pool to be labelled and added to the training set, to minimize the overall amount of data needed to be labelled. In AL, instead of labelling a large dataset, it iteratively acquires labels from an expert only for the most informative or uncertain data points from a unlabelled data pool. Following each acquisition step, the recently labelled data points are incorporated into the training set, and the model undergoes retraining. This iterative procedure continues until a satisfactory level of accuracy is attained [3].

In AL, the informativeness or uncertainty of new data points is assessed by an acquisition function. Acquisition functions can include entropy, least confident, margin and density-based sampling [4]. Effective metrics are essential for guiding the selection of informative samples. It is the most crucial component and the primary distinguishing factor among AL methods in the literature.

The main aim of this work is to present a base model trained on a minimal dataset that can be used for preclinical segmentation and utilise the model to show how an information metric, along with an uncertainty metric, can be used in AL as an acquisition function to guide the selection of data to incorporate into the training set.

MATERIALS AND METHODS

Deep Learning Model

The dataset used to train the base model was obtained from the Positron Emission Tomography Imaging Centre (PETIC). The data consisted of 76 preclinical CT scans of mice, all with corresponding ground truths with 6 labels corresponding to background, skeleton, liver, kidney, left tumour and right tumour. The ground truths were manually outlined by an expert user from PETIC. 20 randomly selected scans including their ground truths were used in a 80:20 split for training, where 16 samples were used for the initial training and 4 samples were used for testing. The remaining 56 scans created the unlabelled data pool. Additional preprocessing of the dataset was performed including resampling and resizing the scans to make the dimensions coincide together.

The base model was developed to predict the segmentation masks of the data utilizing NiftyNet open-source platform [5]. For the network architecture, a dense VNet was used with an initial learning rate of 0.001 along with the Adaptive Moment estimation (Adam) optimiser. The model was trained on the training set for 3000 iterations on whole images. Following training inference was performed on the test scans within the testing set to produce predicted segmentations and on the unlabelled data pool to produce the predicted segmentation and the probability masks.

The metric employed to evaluate the performance of the model and thus segmentation was the Dice Score Coefficient (DSC). The DSC shows the similarity between the predicted segmentation and the ground truth by measuring the overlap and is calculated using Eq. 1.

$$DSC = \frac{2 \times |GT \cap SM|}{|GT| + |SM|} \quad (1)$$

Where GT is the ground truth mask and SM is the segmented predicted mask. | GT ∩ SM | is the number of pixels common to both the segmented predicted mask and the ground truth mask. |GT| is the total number of pixels in the ground truth mask and |SM| is the total number of pixels in the segmented predicted mask [6]. A DSC of 0.7 and above is considered to have good overlap [7].

Active Learning Acquisition Function

Two acquisition functions were combined to create one final score, which is applied following inference of the unannotated pool. The final score consists of the uncertainty function which calculates the uncertainty associated with the prediction of the model for image x. The entropy is used as the measure of predictive uncertainty. The entropy (H) of an image x for a n-class task is defined in Eq. 2.

$$H(x) = -\sum_{i-1}^{n} P(y_i|x) \log\big(P(y_i|x)\big) \quad (2)$$

$P(y_i \backslash x)$ is the probability that the current sample x is predicated to be of class y_i. A higher entropy corresponds to a higher uncertainty, thus suggesting lower overall confidence [8].

The second part of the acquisition function is the Mutual Information (MI). In general, MI evaluates the degrees of relatedness between two variables [9]. In terms of images calculates the similarity between two images. MI of two random variables A and B is defined by Eq. 3.

$$I(A,B) = \sum_{a,b} P_{A,B}(a,b) \log \frac{P_{A,B}(a,b)}{P_A(a)P_B(b)} \quad (3)$$

Where $P_A(a)$ and $P_B(b)$ are the marginal probability functions and $P_{A,B}(a,b)$ is the joint probability function. MI measures the degree of dependence of A and B by measuring the distance between the joint probability and the probability associated with the case of complete independence [10]. The MI is 0 if the two variables are independent, meaning there is no relationship between them and if the two variables are dependent MI is greater than 0 [11].

The MI is normalised to produce the normalised mutual information (NMI) using Eq. 4 proposed by [12].

$$\tilde{I}(A,B) = \frac{H(A) + H(B)}{H(A,B)} \quad (4)$$

Where $H(A)$ is the entropy of A and $H(B)$ is the entropy of B and $H(A,B)$ is the joint entropy of A and B.

The final scoring function is calculated by combining MI and entropy and can be expressed as Eq. 5.

$$Score(x) = \alpha \cdot Uncertainty(x) - \beta \cdot MI(x, X_{train}) \quad (5)$$

$MI(x, X_{train})$ represents the NMI between the image from the unannotated data pool and the training dataset and the $Uncertainty(x)$ represents the uncertainty with the model's prediction of the segmentation for image x from the unannotated data pool.

The terms are weighted by coefficients and that control the contribution of uncertainty and NMI to the overall score. By adjusting these coefficients, one aspect can be prioritised over the other based on the specific requirements of the learning task.

The images with the highest combined score would be selected. These images represent those with a high entropy which indicates high uncertainty meaning the model has lower confidence in the correct label for that image and low MI meaning they contain information that the current training set doesn't contain.

Implementation
The DSC was calculated for each label in the predicted segmentation of the test images and the average taken across them.

The NMI was calculated on a label basis, by making comparisons of histograms of CT intensity values in the training set versus the unannotated data pool. The background label was ignored, to focus on the relevant regions of interest as background can introduce noise and unnecessary complexity. Each image within the unannotated pool has 16 values for each label and the average is taken per label. The overall average NMI value is the taken per image and these were normalised for the set.

To calculate the entropy of each image in the unannotated data pool, the predicted segmentations are used and the entropy is calculated for each pixel and the average is taken across the image. These were then normalised for the set.

The two calculations were combined as shown in Eq. 45 to provide a final score for each image in the unannotated data pool. To demonstrate the effectiveness of the combined acquisition score the DSC is used, as a low DSC indicates poor segmentation performance, suggesting that the sample is informative and should be added to the training pool. If the combined acquisition function consistently identifies samples with low DSC, it indicates that the combined function is effectively selecting the most informative samples.

RESULTS

The predicted segmentations were analysed using the DSC for each label in the test images and the average taken across them as shown in Table.1 Test image 4 had no left tumour present.

	Labels						
Test Image	**0**	**1**	**2**	**3**	**4**	**5**	**Average**
1	1.0	0.9	0.7	0.6	0.4	0.5	0.7
2	1.0	0.8	0.6	0.6	0.6	0.5	0.7
3	1.0	0.9	0.5	0.7	0.7	0.6	0.7
4	1.0	0.9	0.5	0.8	N/A	0.4	0.7

Table 1. Dice coefficient score for test images.

The images within the unannotated data pool are labelled 0-56 for display purposes and ease, where each number refers to a different image.

The average NMI and the average entropy per image were combined and the values are demonstrated in the bar chart in Fig. 1. The bar chart is ordered from lowest to highest where each bar represents a different image and the height of the bar represents the combined score consisting of entropy and NMI for that image. Within the graph, a line has been drawn, where the line acts as a cut-off point. All data to the right of the bar represent images that would be sent to the oracle for annotation as they represent images with both high entropy (indicating high uncertainty) and low mutual information (indicating information not present in the current training set) and all data to the right of the bar would not be annotated.

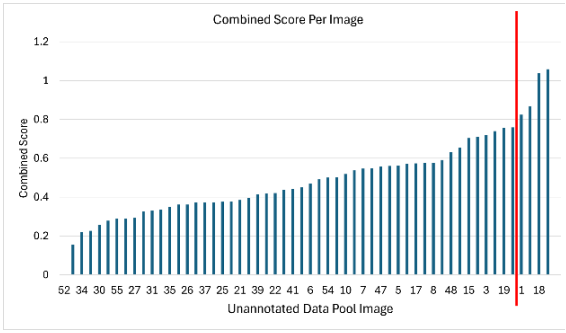

Fig. 1. Combined score per image of the image in the unannotated data pool.

The DSC were calculated for each image, ignoring the background, in the unannotated data pool, where the average DSC has a value of 0.6 and ranges from 0.2 to 0.8.

Based on the analysis, the combined acquisition score exhibits a Pearson correlation coefficient of -0.25 with the DSC, reflecting a slight negative correlation as depicted in Fig. 2. The p-value obtained was 0.059.

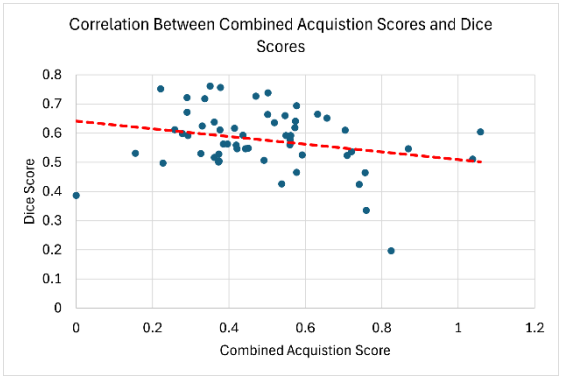

Fig. 2. Correlation plot between the combined acquisition score and the dice score.

DISCUSSION

The dice coefficient values showed the model performed well as all test scan segmentations had an average dice score of 0.7 showing good overlap when compared to the ground truth.

The Pearson correlation coefficient of -0.25 reflects a slight negative correlation. This suggests that higher acquisition scores tend to be associated with lower DSC, indicating that the function may have a tendency, albeit modest, to select more informative samples for model enhancement.

While the p-value of 0.059 slightly exceeds the conventional significance threshold of $P<0.05$, it remains close to this boundary, hinting at a potential trend where the acquisition function targets challenging samples for labelling. This near-significant result suggests that our acquisition function may be on the cusp of demonstrable effectiveness in enhancing the active learning process. Additionally, it's important to note that the effectiveness of the acquisition function does not necessarily depend on a linear relationship. The nuanced nature of data distribution and the complexity of the learning model might mean that non-linear relationships are also of substantial value in improving model accuracy. This insight provides an encouraging direction for further refinement and potential adjustments to the acquisition strategy to better capture and utilize informative samples for improving the model performance.

Acknowledgements
I would like to acknowledge the continual support and guidance received from Dr Rhodri Smith and also Dr Stephen Paisey who provided the data and advice. I would also like to acknowledge IPOCH who have provided background support.

Conflicts of Interest
The authors declare no conflict of interest.

REFERENCES

[1] S. Boehringer, A. Sanaat, H. Arabi, and H. Zaidi, 'An active learning approach to train a deep learning algorithm for tumor segmentation from brain MR images', *Insights Imaging*, vol. 14, no. 1, p. 141, Aug. 2023. doi.org/10.1186/s13244-023-01487-6

[2] J. Sourati, M. Akcakaya, J. Dy, T. Leen, and D. Erdogmus, 'Classification Active Learning Based on Mutual Information', *Entropy*, vol. 18, no. 2, p. 51, Feb. 2016. doi.org/10.3390/e18020051

[3] A. Kirsch, J. van Amersfoort, and Y. Gal, 'BatchBALD: Efficient and Diverse Batch Acquisition for Deep Bayesian Active Learning', *arXiv*, 28 Oct. 2019. doi.org/10.48550/arXiv.1906.08158.

[4] S. Budd, E. C. Robinson, and B. Kainz, 'A survey on active learning and human-in-the-loop deep learning for medical image analysis', *Medical Image Analysis*, vol. 71, p. 102062, Jul. 2021. doi.org/10.1016/j.media.2021.102062

[5] E. Gibson et al., 'NiftyNet: a deep-learning platform for medical imaging', *Computer Methods and Programs in Biomedicine*, vol. 158, pp. 113–122, May 2018. doi.org/10.1016/j.cmpb.2018.01.025

[6] S. M. Sundara and R. Aarthi, 'Segmentation and Evaluation of White Blood Cells using Segmentation Algorithms', in *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, Tirunelveli, India: IEEE, Apr. 2019, pp. 1143–1146. do.org/10.1109/ICOEI.2019.8862724

[7] A. P. Zijdenbos, B. M. Dawant, R. A. Margolin, and A. C. Palmer, 'Morphometric analysis of white matter lesions in MR images: method and validation', *IEEE Trans. Med. Imaging*, vol. 13, no. 4, pp. 716–724, Dec. 1994. doi.org/10.1109/42.363096

[8] P. Ren *et al.*, 'A Survey of Deep Active Learning', *ACM Comput. Surv.*, vol. 54, no. 9, pp. 1–40, Dec. 2022. doi.org/10.1145/3472291

[9] M. Noshad, J. Choi, Y. Sun, A. Hero, and I. D. Dinov, 'A data value metric for quantifying information content and utility', *J. Big Data*, vol. 8, no. 1, p. 82, Dec. 2021. doi.org/10.1186/s40537-021-00446-6

[10] H. Chen, 'Mutual Information: A Similarity Measure for Intensity Based Image Registration', in *Advanced Image Processing Techniques for Remotely Sensed Hyperspectral Data*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 89–108. doi.org/10.1007/978-3-662-05605-9_4

[11] L. Zheng, 'Using mutual information as a cocitation similarity measure', *Scientometrics*, vol. 119, no. 3, pp. 1695–1713, Jun. 2019. doi.org/10.1007/s11192-019-03098-9

[12] D. Hill, 'Across-Modality Registration Using Intensity-Based Cost Functions', in *Handbook of Medical Image Processing and Analysis*, Elsevier, 2009, pp. 613–628. doi.org/10.1016/B978-012373904-9.50047-7

Cynhadledd Ymchwil Peirianneg Caerdydd
Cardiff Engineering Research Conference

E. Spezi and M. Bray (eds.) 2024. *Proceedings of the Cardiff University School of Engineering Research Conference 2024.* Cardiff: Cardiff University Press.
doi.org/10.18573/conf3

Cardiff University School of Engineering Research Conference 2024 was held from 12 to 14 June 2024 at Cardiff University.

The work presented in these proceedings has been peer reviewed and approved by the conference organisers and associated scientific committee to ensure high academic standards have been met.

Cardiff University Press | Gwasg Prifysgol Caerdydd

First published 2024

Cardiff University Press
Cardiff University, Trevithick Library
First Floor, Trevithick Building, Newport Road
Cardiff CF24 3AA

cardiffuniversitypress.org

Editorial design and layout by
Academic Visual Communication

ISBN: 978-1-9116-5351-6 (PDF)