Adaptive Curriculum Query Strategy for Active Learning in Medical Image Classification

Siteng Ma[1], Honghui Du[1](✉), Kathleen M. Curran[2], Aonghus Lawlor[1], and Ruihai Dong[1](✉)

[1] Insight Centre for Data Analytics, School of Computer Science, University College Dublin, Ireland
siteng.ma@ucdconnect.ie, {honghui.du,aonghus.lawlor,ruihai.dong}@ucd.ie
[2] School of Medicine, University College Dublin, Ireland
kathleen.curren@ucd.ie

Abstract. Deep active learning (AL) is commonly used to reduce labeling costs in medical image analysis. Deep learning (DL) models typically exhibit a preference for learning from easy data and simple patterns before they learn from complex ones. However, existing AL methods often employ a fixed query strategy for sample selection, which may cause the model to focus too closely on challenging-to-classify data. The result is a deceleration of the convergence of DL models and an increase in the amount of labeled data required to train them. To address this issue, we propose a novel *Adaptive Curriculum Query Strategy* for AL in Medical Image Classification. During the training phase, our strategy leverages Curriculum Learning principles to initially prioritize the selection of a diverse range of samples to cover various difficulty levels, facilitating rapid model convergence. Once the distribution of the selected samples closely matches that of the entire dataset, the query strategy shifts its focus towards difficult-to-classify data based on uncertainty. This novel approach enables the model to achieve superior performance with fewer labeled samples. We perform extensive experiments demonstrating that our model not only requires fewer labeled samples but outperforms state-of-the-art models in terms of efficiency and effectiveness. The code is publicly available at https://github.com/HelenMa9998/Easy_hard_AL.

Keywords: Active learning · Query strategy · Labeling costs · Medical image classification

1 Introduction

Deep learning (DL) has achieved revolutionary progress in the field of medical image processing [22]. The success of DL models typically depends on the support of vast amounts of annotated data. However, annotating data for medical images relies on experts, which is particularly costly and time-consuming [20].

Deep Active Learning (DAL) is an iterative user interaction approach that requests experts to label only the most informative data point for training DL

models, which can significantly reduce labeling costs [19]. There are several different query strategies to select the most informative data point. The uncertainty-based query strategy [11, 17] involves selecting data points for which the model has the lowest confidence in its predictions. The diversity-based query strategy [18] selects a set of data points that are diverse in distribution or varied among themselves. The hybrid query strategy [1, 7] combines the approaches of uncertainty and diversity by selecting data points that are both highly informative due to the model's uncertainty about them and diverse to ensure a broad representation of the dataset.

Foundational research [3, 5, 14] reveals that during the training process, DL models initially focus on identifying simple, recognizable patterns and common features within data rather than memorizing individual, hard-to-classify data points. This preference assists DL models in understanding the underlying structure and general characteristics of new data, thereby enhancing their ability to generalize to unseen data effectively. However, current DAL methods often use a fixed strategy to select data points and lack the ability to adaptive adjust to the evolving learning phases of DL models [1, 7, 11, 18]. This slows the convergence of deep models while increasing the number of labeled data needed for training.

Curriculum learning and self-paced learning are training strategies for machine learning models that begin with simpler examples and progressively increase complexity, emulating the progressive learning pattern of humans. Curriculum learning follows a predefined sequence, while self-paced learning dynamically adapts sample selection based on the model's progress [12, 21]. Both strategies align with the learning patterns of DL models. However, self-paced learning require a fully labeled dataset [15], and no existing research in DAL has yet explored querying examples from easier to more difficult and enabling DL models to progressively learn from these examples in a sequence from simpler to more complex. This paper thus investigates the following research question: *Can a curriculum query strategy in DAL improve the performance of DL? How and why?*

To address these problems, we propose a novel approach named Adaptive Curriculum query strategy for Active Learning (ACAL) in medical image classification. Unlike existing DAL methods, ACAL is the first algorithm to adapt its query strategy based on the DL model's learning stages. Initially, ACAL selects a diverse range of samples to cover various difficulty levels and learn general biomarkers or lesion characteristics. Meanwhile, it monitors the difference between the distribution of the overall selected data and the entire dataset's distribution. When the distribution of selected data closely aligns with the entire dataset, ACAL transitions to choosing hard-to-classify data points adaptively, thereby training the DL model progressively from easier to harder data, mimicking an effective learning sequence in DL training. Our experiments demonstrate that our approach outperforms the state-of-the-art methods by higher classification accuracy with fewer labeled samples. This enhancement enables an effective reduction in labeling costs for different medical image classification scenerios.

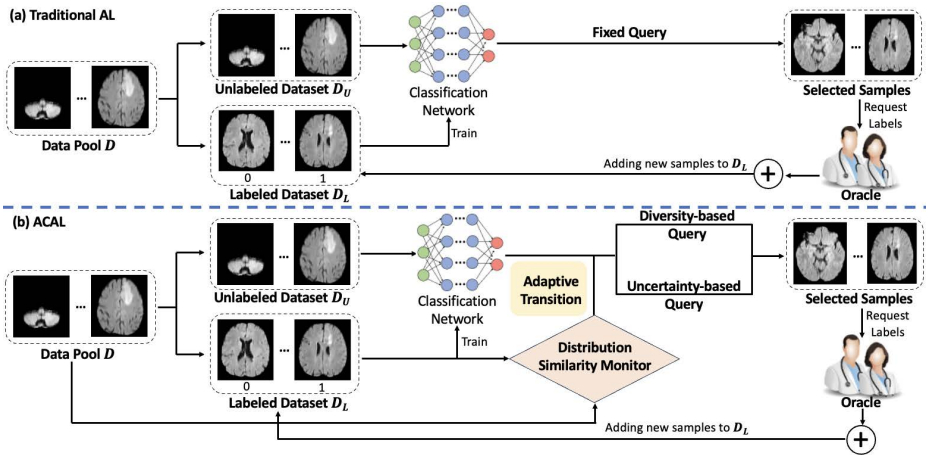


Fig. 1: Comparison of Traditional AL and ACAL: The traditional approach sticks to a fixed query strategy during the whole process, while ACAL adaptively adjusts its strategy in sync with the DL model learning process.

2 Proposed Method

Consider a dataset $\mathcal{D} = \{\mathbf{x}_i\}_{i=1}^{|\mathcal{D}|}$, where $\mathbf{x}_i \in \mathbb{R}^{h \times w}$ represent medical images with dimensions $h \times w$. Let $\mathcal{D}_U = \{\mathbf{x}_u\}_{u=1}^{|\mathcal{D}_U|} \subseteq \mathcal{D}$ and $\mathcal{D}_L = \{(\mathbf{x}_l, y_l)\}_{l=1}^{|\mathcal{D}_L|}, \forall \mathbf{x}_l \in \mathcal{D}$ denote an unlabeled and a labeled subset of $\mathcal{D}$, respectively, where $y_l \in \mathcal{Y}$ is the corresponding label for $\mathbf{x}_l$, $\mathcal{Y}$ is the set of possible labels. An *Oracle* (e.g., clinician) is available to label queried data points from $\mathcal{D}_U$. DAL involves an iterative loop that starts with a small $\mathcal{D}_L$ and a larger $\mathcal{D}_U$. In each iteration, a model $\mathcal{M}$ is trained on $\mathcal{D}_L$ and then used to identify the most informative samples from $\mathcal{D}_U$, which are subsequently labeled by an oracle and added to $\mathcal{D}_L$. This process repeats to improve the model's performance by incrementally enriching $\mathcal{D}_L$. The selection of the most informative samples is guided by a query strategy Q to optimize the performance of a DL model trained on $\mathcal{D}_L$ with the least amount of labeled data.

Differently from traditional DAL which only relies on a fixed query strategy as shown in Fig.1, our proposed method Adaptive Curriculum Query Strategy for Active Learning (ACAL) includes an additional component: the Distribution Similarity Monitor (DSM). DSM evaluates the similarity between the distribution of the labeled dataset and that of the entire dataset in each iteration. To ensure broad coverage of the data space, ACAL utilizes a diversity-based query strategy to maximize the variety of samples queried until the similarity between the two distributions meets a predefined threshold. This guarantees that during the early stages of DL model learning, the classification network has access to data that facilitates rapid learning of recognizable patterns and common features (e.g., brain tumors may exhibit an enhancement effect on MRI scans that

Algorithm 1 Learning Procedure of ACAL

Require: Dataset $\mathcal{D}$, Budget B, Deep learning model $\mathcal{M}$, Threshold τ, The number of initial query samples n_0, The number of querying samples at each iteration n

1: Initialize $\mathcal{D}_L$ by randomly querying n_0 instances from $\mathcal{D}$
2: $Q \leftarrow$ Diversity-based Query Strategy
3: **while** $|\mathcal{D}_L| < B$ **do**
4: $\quad \mathcal{D}_U \leftarrow \mathcal{D} - \mathcal{D}_L$
5: $\quad$ Train model $\mathcal{M}$ on $\mathcal{D}_L$
6: $\quad$ **for all** $\mathbf{x}_i \in \mathcal{D}$ **do**
7: $\quad\quad E \leftarrow E \cup encode(\mathbf{x}_i)$
8: $\quad\quad$ **if** $\mathbf{x}_i \in \mathcal{D}_L$ **then**
9: $\quad\quad\quad E_L \leftarrow E_L \cup encode(\mathbf{x}_i)$
10: $\quad\quad$ **end if**
11: $\quad$ **end for**
12: $\quad$ **if** $Q =$ Diversity-based Query Strategy AND $DSM(p(E)||p(E_L)) < \tau$ **then**
13: $\quad\quad Q \leftarrow$ Uncertainty-based Query Strategy
14: $\quad$ **end if**
15: $\quad \mathcal{D}_Q \leftarrow$ Query n samples from $\mathcal{D}_U$ based on Q
16: $\quad \mathcal{D}_L \leftarrow \mathcal{D}_L \cup \mathcal{D}_Q$
17: **end while**

tumor area appears brighter than the surrounding tissue). Upon reaching the similarity threshold, ACAL shifts to an uncertainty-based query strategy, focusing on samples that are challenging to classify. This curriculum approach ensures that after the DL model acquires basic patterns and features, it also learns to distinguish between individual, hard-to-classify samples, thereby mirroring the natural learning sequence in DL training.

ACAL's learning process is shown in Algorithm 1. It begins by establishing an initial labeled dataset $\mathcal{D}_L$ with a random selection of a small number of instances from $\mathcal{D}$ (line 1) and sets diversity-based query strategy as the base query strategy Q (line 2). Following this, ACAL then proceeds with a loop that runs until the size of $\mathcal{D}_L$ reaches a predefined budget B (line 3). This budget acts as a constraint on the total number of samples that can be labeled. Within each iteration, ACAL updates the unlabeled dataset $\mathcal{D}_U$ by removing samples already included in $\mathcal{D}_L$ (line 4). Subsequently, the DL model $\mathcal{M}$ is retrained to improve its learning based on the accumulated knowledge from updated $\mathcal{D}_L$ (line 5). For every instance $\mathbf{x}_i$ in the dataset $\mathcal{D}$, ACAL encodes $\mathbf{x}_i$ and adds the corresponding embedding into the encoded set E. The encoded set E_L consists of the embeddings of all the labeled samples from D_L (line 6 to 11). This algorithm employs DSM to track the differences in distributions between the full dataset and the labeled subset (line 12). It is important to note that any method capable of comparing two distributions may serve as the DSM. In this paper, we employ the Jensen-Shannon divergence [9] as the metric. If the distribution difference exceeds the threshold τ, indicating significant dissimilarity, the algorithm queries n diverse instances from $\mathcal{D}_U$ to enhance the model's generalizability. Otherwise, it focuses on querying n hard samples that are challenging for the model to

classify by shifting Q to the uncertainty-based query strategy (line 13), aiming to improve the model's accuracy on difficult instances (line 12 to line 15). The queried samples are then added to the labeled dataset $\mathcal{D}_L$, enriching it with new information for subsequent training iterations (line 16).

3 Experiments and Results

3.1 Datasets

Two well-known public medical image datasets for different scenarios are used to evaluate our proposed method.

- **BraTS 2019 Dataset** [4] comprises 3D brain MRI scan images from 335 patients, including 259 with High-Grade Glioma (HGG) and 76 with Low-Grade Glioma (LGG). Each patient has images from four modalities: T1, T2, T1ce, and FLAIR. This study specifically uses the FLAIR images of patients with LGG.
- **Breast Cancer Diagnosis Dataset** (BACH) [2] includes 400 high-resolution histopathology images of breast tissue cells. These images are evenly divided into four categories: Normal, Benign, In-situ Carcinoma, and Invasive Carcinoma, with each category containing 100 images.

The training, validation, and test datasets consist of 7,750, 1,860, and 2,170 patches for the BraTS dataset and 256, 80, and 64 patches for the BACH dataset, respectively. Online augmentation is applied throughout the process using the Albumentation library [6] (e.g., Gaussian blur, rotates, and flips).

3.2 Benchmark methods and Evaluation Metrics

We compare nine of the most well-known and commonly used methods in the AL field, covering uncertainty-based (**LeastConfidence** (LC) [16], **Margin** (Marg) [17], **Entropy** (Ent) [19], **MC-Dropout** (MC-D) [11], **BALD** [11]), diversity-based (**Random** (Rand), **K-Center-Greedy** (KCG) [18]), and hybrid AL approaches (**CDAL** [1], **ClusterMarginSampling** (CMS) [7]).

For ACAL, we explore different combinations of diversity- and uncertainty-based methods as our base query strategy. Meanwhile, we also introduce two novel diversity-based query strategies. The first, termed Stratified Uncertainty Random Sampling (SURS), groups each example in $\mathcal{D}_U$ into g categories (e.g., we choose $g = 5$ in this study) based on the uncertainty of the classification model's prediction for them. From each category, we randomly select $\frac{n}{g}$ examples to form the query set $\mathcal{D}_Q$ in each iteration. This approach enables AL to uniformly query examples across different levels of uncertainty. The second strategy involves clustering examples in $\mathcal{D}_U$ into n clusters using K-Medoids, and selecting the center of each cluster to form the query set $\mathcal{D}_Q$ in each iteration.

We use the ResNet-18 [10] as the classification model, initializing it with weights from pre-training on the ImageNet [8] dataset and employing cross-entropy loss. Using other self-supervised pretrained encoders can be investigated

Table 1: Model Performance with Varied Label Percentages for BraTS (upper section) and BACH (lower section) Datasets, with supervised learning accuracy of 0.897 and 0.713, respectively. The best approach is highlighted in red, while the approaches ranked second and third are in bold with a grey background.

Percentage of $\mathcal{D}$	0.1		0.15		0.2		0.25		0.3		**Avg.Rank**
Query Strategy	Mean+Std	Rank	Mean+Std	Rank	Mean+Std	Rank	Mean+Std	Rank	Mean+Std	Rank	
ACAL(Rand|Ent)	**0.816±0.030**	**3**	**0.855±0.006**	**1**	0.857±0.011	5	**0.868±0.013**	**1**	0.851±0.011	8	**3.6**
ACAL(Rand|LC)	0.812±0.017	5	0.848±0.011	4	**0.860±0.014**	**3**	0.847±0.005	10	0.847±0.019	10	6.6
ACAL(Rand|Marg)	0.812±0.017	5	**0.849±0.019**	**3**	0.852±0.010	8	0.855±0.009	7	**0.869±0.008**	**1**	4.8
ACAL(KMed|MC-D)	0.813±0.021	4	0.848±0.014	5	**0.860±0.025**	**2**	0.855±0.015	8	**0.864±0.012**	**3**	**4.6**
ACAL(Rand|BALD)	**0.826±0.023**	**1**	**0.854±0.016**	**2**	**0.861±0.006**	**1**	**0.860±0.011**	**2**	**0.867±0.007**	**2**	**1.8**
ACAL(KCG|Ent)	0.794±0.039	9	0.843±0.024	8	0.852±0.015	7	0.847±0.010	9	0.855±0.015	4	7.4
Ent	0.780±0.038	11	0.847±0.026	6	0.849±0.014	9	0.857±0.024	5	0.853±0.016	6	7.4
LC	0.718±0.028	13	0.686±0.020	14	0.700±0.029	14	0.697±0.032	15	0.694±0.031	15	14.2
Marg	0.743±0.028	12	0.768±0.013	12	0.798±0.008	12	0.801±0.021	12	0.787±0.024	12	12.0
MC-D	0.799±0.046	8	0.821±0.034	11	0.820±0.031	11	0.837±0.020	11	0.815±0.043	11	10.4
BALD	**0.819±0.018**	**2**	0.828±0.028	10	0.852±0.015	6	0.857±0.014	4	0.847±0.029	9	6.2
KCG	0.713±0.016	14	0.678±0.018	15	0.684±0.038	15	0.734±0.020	14	0.763±0.024	14	14.4
Rand	0.812±0.017	5	0.847±0.023	7	**0.861±0.011**	**1**	0.856±0.012	6	0.852±0.015	7	5.2
CMS	0.687±0.016	15	0.751±0.023	13	0.773±0.016	13	0.758±0.036	13	0.773±0.023	13	13.4
CDAL	0.791±0.031	10	0.830±0.014	9	0.848±0.006	10	**0.859±0.007**	**3**	0.853±0.013	5	7.4
Percentage of $\mathcal{D}$	0.15		0.3		0.45		0.6		0.75		**Avg.Rank**
Query Strategy	Mean+Std	Rank	Mean+Std	Rank	Mean+Std	Rank	Mean+Std	Rank	Mean+Std	Rank	
ACAL(Rand|Ent)	0.309±0.096	3	**0.419±0.084**	**1**	**0.603±0.032**	**2**	**0.691±0.060**	**2**	0.684±0.058	5	**2.6**
ACAL(SURS|LC)	0.300±0.080	6	0.397±0.075	8	0.563±0.077	5	0.681±0.074	4	0.666±0.050	7	6.0
ACAL(SURS|Marg)	0.291±0.018	8	0.400±0.071	6	0.528±0.119	10	0.675±0.049	6	**0.694±0.055**	**2**	6.4
ACAL(KMed|MC-D)	0.281±0.043	9	0.366±0.060	11	**0.641±0.053**	**1**	**0.706±0.045**	**1**	0.678±0.042	6	**5.6**
ACAL(SURS|BALD)	**0.331±0.072**	**1**	0.350±0.046	14	0.572±0.057	4	0.666±0.053	8	**0.716±0.030**	**1**	**5.6**
ACAL(KCG|Ent)	0.297±0.040	7	**0.406±0.077**	**3**	0.563±0.121	5	0.650±0.107	10	0.653±0.124	10	7.0
Ent	0.269±0.043	12	0.375±0.071	10	0.500±0.157	14	0.544±0.112	15	0.606±0.115	14	13.0
LC	0.303±0.024	5	0.384±0.103	9	0.525±0.097	11	0.588±0.074	14	0.666±0.073	7	9.2
Marg	0.281±0.031	9	0.403±0.034	5	0.547±0.086	7	0.672±0.077	7	**0.691±0.052**	**3**	6.2
MC-D	0.281±0.049	9	0.316±0.068	15	0.484±0.065	15	0.666±0.066	8	0.606±0.074	14	12.2
BALD	0.269±0.036	12	**0.406±0.063**	**3**	0.522±0.054	12	0.678±0.048	5	0.688±0.070	4	7.2
KCG	0.259±0.057	15	0.363±0.104	12	**0.581±0.087**	**3**	0.638±0.084	11	0.647±0.060	12	10.6
Rand	**0.309±0.096**	**3**	0.400±0.075	6	0.544±0.078	9	0.634±0.064	12	0.625±0.106	13	8.6
CMS	0.266±0.090	14	**0.409±0.040**	**2**	0.506±0.045	13	**0.684±0.052**	**3**	0.659±0.085	9	8.2
CDAL	**0.328±0.053**	**2**	0.353±0.014	13	0.547±0.119	7	0.594±0.090	13	0.650±0.039	11	9.2

as future work. The classification model is trained for a maximum of 200 epochs with an early stopping criterion of 5 epochs and uses the Adam optimizer [13] with a learning rate of 0.0001. The batch size is set to 32 for BraTS and 16 for BACH. We maintained consistent hyperparameter settings across all experiments and for all methods, with the exception of ACAL's threshold τ. A grid search was employed to tune τ for ACAL, using different base query strategies on each dataset based on preliminary runs. The range for τ varied from 0.08 to 0.49. The number of initial query samples n_0 is set to 100 and 10 for BraTS and BACH, respectively. To simulate the interactive selection process in AL, the number of querying samples at each iteration n is set to 100 in each iteration for BraTS and 10 for BACH. At every AL iteration, the model trains until the validation loss stabilizes. Subsequently, the model undergoes fine-tuning based on the model from the preceding round to expedite the learning process. For ACAL, the optimal combinations of diversity-based query strategies corresponding to each uncertainty-based query strategy are reported.

Five runs are performed for all the approaches. In each run, we measure the test accuracy at the end of each AL round. The average accuracy and standard deviation (std) across the 5 runs are reported. Each run is executed on a single NVIDIA GeForce GTX 1080 Ti GPU. The Python implementation uses PyTorch version 1.10.1.

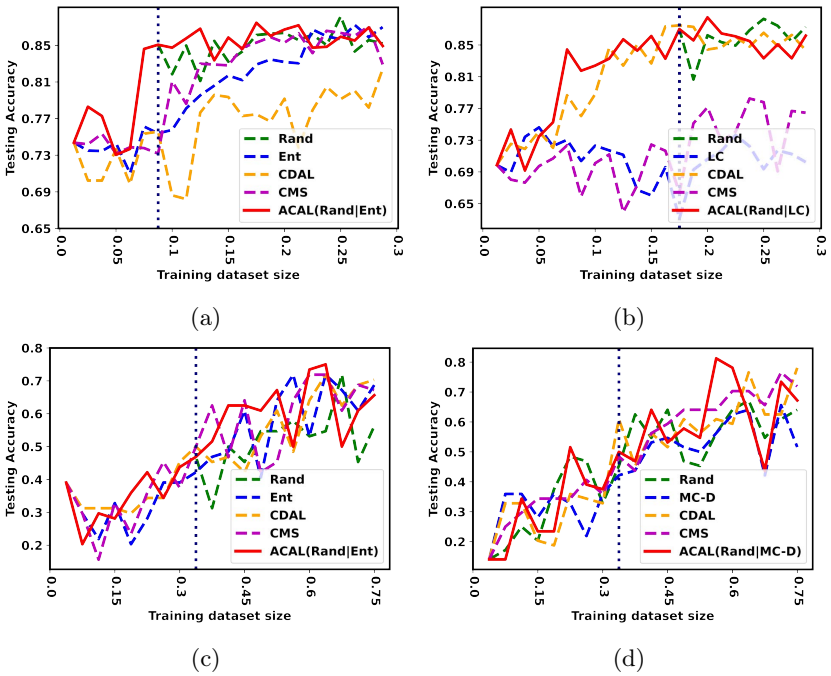
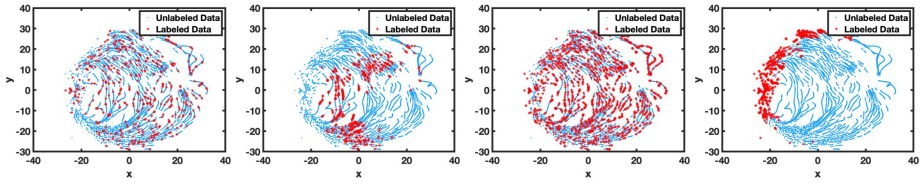


Fig. 2: Performance comparison average over five rounds using different query strategies: Ent, LC in BraTS, and Ent, MC-D in BACH. ACAL consistently achieved almost the highest scores across all methods, demonstrating its effectiveness both before and after transformation.

3.3 Performance Comparison

The experimental results for the BraTS and BACH with different percentages of labeled data are shown in Table 1. Figure 3 shows the representative results across all AL rounds. Other figures were omitted due to space restrictions.

On BraTS, our proposed approach ACAL(Rand|BALD), ACAL(Rand|Ent), and ACACL(KMed|MC-D) achieve the best, second best, and third best average ranks, respectively, across five different percentages of labeled data. From Figures 2a and 2b, it is evident that ACAL(Rand|Ent) and ACAL(Rand|LC) outperform Ent and LC prior to reaching the threshold τ. This suggests that during the early training stages of a DL model, querying diverse examples can help the model learn general biomarkers or lesion characteristics from a variety of examples, which can lead to quicker improvement in performance. After ACAL shifted the query strategy to an uncertainty-based approach, it continued to show significant advantages compared to its base query strategy. This could be because a DL model trained on data selected solely by an uncertainty-based active learning method may focus too much on difficult examples, requiring more labeled data to achieve satisfactory performance. It is worth noting that as the number of queried



(a) (b) (c) (d)

Fig. 3: Data distribution in the AL cycle prior to the transition to uncertainty-based querying in ACAL for dataset BraTS, utilizing t-SNE for two-dimensional projection. Red dots represent data queried by AL, and blue dots denote unlabeled data. The accuracy of the DL model on the test dataset is 0.851 using ACAL(Rand|Ent) in 3a, 0.754 using Ent (3b), transitioning at 700 samples (0.09). Accuracies were 0.8701 for ACAL(Rand|LC) depicted in 3c, and 0.6313 for LC sampling in 3d with transmission point of 1400 (0.18).

data increases, the performance advantage of ACAL over CDAL diminishes (see Table 1, Figures 2a and 2b). This is because, with an increase in labeled training data, every method is able to learn effectively.

ACAL(Rand|Ent) has the best average performance on BACH, while both ACAL(KMed|MC-D) and ACAL(SURS|BALD) are tied for the second-best average rank across the various labeled data percentages (Table 1). However, the smaller test size of the BACH dataset compared to BraTS leads to greater fluctuations in the performance of each method (e.g., a single incorrect prediction can significantly lower the accuracy). Also, because the training size for BACH is smaller than that for BraTS, the performance differences are not as pronounced. Nonetheless, Figures 2c and 2d clearly show the advantage of the proposed method over the base query strategy.

To further analyze the impact of different query strategies on the performance of DL models, we plot the representative distributions of variously labeled data selected by different query strategies on BraTS, as shown in Fig.3. Observations from Figures 3b and 3d, show that uncertainty-based AL primarily focuses on querying examples from specific areas. This focus may hinder the DL model's ability to identify recognizable patterns and common features. Such patterns and features are essential for the model's rapid convergence, particularly during the initial learning phases. Consequently, the DL model trained with examples from 3a has nearly 20% higher accuracy, and from 3c has nearly 24% higher accuracy, compared to the DL models trained with examples from 3b and 3d, respectively.

4 Conclusion

In this study, we address that current AL approaches do not align with the learning patterns of DL models, which prioritize easy and diverse examples before tackling difficult ones. We propose ACAL, the first AL approach that adapts

query strategies to match the learning stages of DL models. ACAL starts with a diversity-based query strategy to maximize sample diversity, accelerating DL models' understanding of the underlying structure and general characteristics of new data. As training progresses, it shifts to an uncertainty-based strategy for hard-to-learn cases.

We performed experiments with two widely used medical image classification datasets to evaluate ACAL. The results show that ACAL can achieve better classification performance with fewer query images by converging faster than fixed AL query strategies. This can significantly reduce labeling costs, which is crucial when dealing with large amounts of unlabeled data in medical imaging analysis. Future work includes enhancing performance by progressing cold-start AL, exploring different metrics to measure the distance between two distributions, and validating on larger datasets.

Acknowledgments. This research was conducted with the financial support of Science Foundation Ireland (SFI) to the Insight Centre for Data Analytics under Grant No. 12/RC/2289_P2.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Agarwal, S., Arora, H., Anand, S., Arora, C.: Contextual diversity for active learning. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVI 16. pp. 137–153. Springer (2020)
2. Aresta, G., Araújo, T., Kwok, S., Chennamsetty, S.S., Safwan, M., Alex, V., Marami, B., Prastawa, M., Chan, M., Donovan, M., et al.: Bach: Grand challenge on breast cancer histology images. Medical image analysis **56**, 122–139 (2019)
3. Arpit, D., Jastrzębski, S., Ballas, N., Krueger, D., Bengio, E., Kanwal, M.S., Maharaj, T., Fischer, A., Courville, A., Bengio, Y., et al.: A closer look at memorization in deep networks. In: International conference on machine learning. pp. 233–242. PMLR (2017)
4. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C.: Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. Scientific data **4**(1), 1–13 (2017)
5. Baldock, R., Maennel, H., Neyshabur, B.: Deep learning through the lens of example difficulty. Advances in Neural Information Processing Systems **34**, 10876–10889 (2021)
6. Buslaev, A., Iglovikov, V.I., Khvedchenya, E., Parinov, A., Druzhinin, M., Kalinin, A.A.: Albumentations: Fast and flexible image augmentations. Information **11**(2) (2020). https://doi.org/10.3390/info11020125, https://www.mdpi.com/2078-2489/11/2/125
7. Citovsky, G., DeSalvo, G., Gentile, C., Karydas, L., Rajagopalan, A., Rostamizadeh, A., Kumar, S.: Batch active learning at scale. Advances in Neural Information Processing Systems **34**, 11933–11944 (2021)

8. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. pp. 248–255. Ieee (2009)
9. Fuglede, B., Topsoe, F.: Jensen-shannon divergence and hilbert space embedding. In: International symposium onInformation theory, 2004. ISIT 2004. Proceedings. p. 31. IEEE (2004)
10. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
11. Houlsby, N., Huszár, F., Ghahramani, Z., Lengyel, M.: Bayesian active learning for classification and preference learning. stat **1050**, 24 (2011)
12. Jiang, L., Meng, D., Yu, S.I., Lan, Z., Shan, S., Hauptmann, A.: Self-paced learning with diversity. Advances in neural information processing systems **27** (2014)
13. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
14. Krueger, D., Ballas, N., Jastrzebski, S., Arpit, D., Kanwal, M.S., Maharaj, T., Bengio, E., Fischer, A., Courville, A.: Deep nets don't learn via memorization (2017)
15. Kumar, M., Packer, B., Koller, D.: Self-paced learning for latent variable models. Advances in neural information processing systems **23** (2010)
16. Lewis, D.D.: A sequential algorithm for training text classifiers: Corrigendum and additional data. In: Acm Sigir Forum. vol. 29, pp. 13–19. ACM New York, NY, USA (1995)
17. Scheffer, T., Decomain, C., Wrobel, S.: Active hidden markov models for information extraction. In: International symposium on intelligent data analysis. pp. 309–318. Springer (2001)
18. Sener, O., Savarese, S.: Active learning for convolutional neural networks: A core-set approach. In: International Conference on Learning Representations (2018)
19. Settles, B.: Active learning literature survey (2009)
20. Tajbakhsh, N., Jeyaseelan, L., Li, Q., Chiang, J.N., Wu, Z., Ding, X.: Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. Medical Image Analysis **63**, 101693 (2020)
21. Wang, X., Chen, Y., Zhu, W.: A survey on curriculum learning. IEEE Transactions on Pattern Analysis and Machine Intelligence **44**(9), 4555–4576 (2021)
22. Zhou, S.K., Greenspan, H., Shen, D.: Deep learning for medical image analysis. Academic Press (2023)