CLINICAL: Targeted Active Learning for Imbalanced Medical Image Classification

Suraj Kothawade[1(✉)], Atharv Savarkar[2], Venkat Iyer[2], Ganesh Ramakrishnan[2], and Rishabh Iyer[1]

[1] University of Texas at Dallas, Richardson, USA
suraj.kothawade@utdallas.edu
[2] Indian Institute of Technology, Bombay, Mumbai, India

Abstract. Training deep learning models on medical datasets that perform well for all classes is a challenging task. It is often the case that a suboptimal performance is obtained on some classes due to the natural class imbalance issue that comes with medical data. An effective way to tackle this problem is by using *targeted active learning*, where we iteratively add data points that belong to the rare classes, to the training data. However, existing active learning methods are ineffective in targeting rare classes in medical datasets. In this work, we propose CLINICAL (targeted a**C**tive **L**earning for **I**mbala**N**ced med**IC**al im**A**ge c**L**assification) a framework that uses submodular mutual information functions as acquisition functions to mine critical data points from rare classes. We apply our framework to a wide-array of medical imaging datasets on a variety of real-world class imbalance scenarios - namely, *binary* imbalance and *long-tail* imbalance. We show that CLINICAL outperforms the state-of-the-art active learning methods by acquiring a diverse set of data points that belong to the rare classes.

1 Introduction

Owing to the advancement of deep learning, medical image classification has made tremendous advances in the past decade. However, medical datasets are naturally imbalanced at the class level, *i.e.*, some classes are comparatively rarer than the others. For instance, cancerous classes are naturally rarer than non-cancerous ones. In such scenarios, the over-represented classes *overpower* the training process and the model ends up learning a biased representation. Deploying such biased models results in incorrect predictions, which can be catastrophic and even lead to loss of life. Active learning (AL) is a promising solution to mitigate this imbalance in the training dataset. The goal of AL is to select data points from an unlabeled set for addition to the training dataset at an additional

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-16760-7_12.

G. Zamzmi et al. (Eds.): MILLanD 2022, LNCS 13559, pp. 119–129, 2022.
https://doi.org/10.1007/978-3-031-16760-7_12

labeling cost. The model is then retrained with the new training set and the process is repeated. Reducing the labeling cost using the AL paradigm is crucial in domains like medical imaging, where labeling data requires expert supervision (*e.g.*, doctors), which makes the process extremely expensive. However, current AL methods are inefficient in selecting data points from the rare classes in medical image datasets. Broadly, they use acquisition functions that are either: i) based on the uncertainty scores of the model, which are used to select the top uncertain data points [26], or ii) based on diversity scores, where data points having diverse gradients are selected [3,25]. They mainly focus on improving the overall performance of the model, and thereby fail to target these rare yet critical classes. Unfortunately, this leads to a wastage of expensive labeling resources when the goal is to improve performance on these rare classes.

In this work, we consider two types of class imbalance that recur in a wide array of medical imaging datasets. The first scenario is *binary* imbalance, where a subset of classes is rare/infrequent and the remaining subset is relatively frequent. The second scenario is that of *long-tail* imbalance, where the frequency of data points from each class keeps *steeply* reducing as we go from the most frequent class to the rarest class (see Fig. 1). Such class imbalance scenarios are particularly challenging in the medical imaging domain since there exist subtle differences which are barely visually evident (see Fig. 1). In Sect. 3, we discuss CLINICAL, a targeted active learning algorithm that acquires a subset by maximizing the submodular mutual information with a set of *misclassified* data points from the rare classes. This enables us to focus on data points from the unlabeled set that are critical and belong to the rare classes.

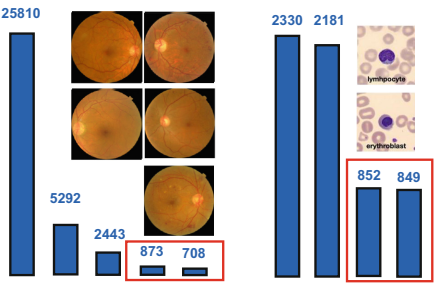

Fig. 1. Motivating examples of two main class imbalance scenarios occurring in medical imaging. **Left:** Long-tail imbalance (Diabetic retinopathy grading from retinal images in APTOS-2019 [10]). **Right:** Binary imbalance (Microscopic peripheral blood cell image classification in Blood-MNIST [1]). Red boxes in both scenario denote targeted rare classes. (Color figure online)

1.1 Related Work

Uncertainty Based Active Learning. Uncertainty based methods aim to select the most uncertain data points according to a model for labeling. The most common techniques are - 1) ENTROPY [26] selects data points with maximum entropy, 2) LEAST CONFIDENCE [28] selects data points with the lowest confidence, and 3) MARGIN [24] selects data points such that the difference between the top two predictions is minimum.

Diversity Based Active Learning. The main drawback of uncertainty based methods is that they lack diversity within the acquired subset. To mitigate this,

a number of approaches have proposed to incorporate diversity. The CORESET method [25] minimizes a coreset loss to form coresets that represent the geometric structure of the original dataset. They do so using a greedy k-center clustering. A recent approach called BADGE [3] uses the last linear layer gradients to represent data points and runs K-MEANS++ [2] to obtain centers that have a high gradient magnitude. The centers being representative and having high gradient magnitude ensures uncertainty and diversity at the same time. However, for batch AL, BADGE models diversity and uncertainty only within the batch and *not* across all batches. Another method, BATCHBALD [15] requires a large number of Monte Carlo dropout samples to obtain significant mutual information which limits its application to medical domains where data is scarce.

Class Imbalanced and Personalized Active Learning. Closely related to our method CLINICAL, are methods which optimize an objective that involves a held-out set. GRADMATCH [13] uses an orthogonal matching pursuit algorithm to select a subset whose gradient closely matches the gradient of a validation set. Another method, GLISTER-ACTIVE [14] formulates an acquisition function that maximizes the log-likelihood on a held-out validation set. We adopt GRADMATCH and GLISTER-ACTIVE as baselines that *targets* rare classes in our class imbalance setting and refer to it T-GRADMATCH and T-GLISTER in Sect. 4. Recently, [16] proposed the use of submodular information measures for active learning in realistic scenarios, while [17] used them to find rare objects in an autonomous driving object detection dataset. Finally, [19] use the submodular mutual information functions (used here) for personalized speech recognition. Our proposed method uses the submodular mutual information to target selecting data points from the rare classes via using a small set of *misclassified* data points as exemplars, which makes our method applicable to binary as well as long-tail imbalance scenarios.

1.2 Our Contributions

We summarize our contributions as follows: **1)** We emphasize on the issue of binary and long-tail class imbalance in medical datasets that leads to poor performance on rare yet critical classes. **2)** Given the limitations of current AL methods on medical datasets, we propose CLINICAL, a novel AL framework that can be applied to any class imbalance scenario. **3)** We demonstrate the effectiveness of our framework for a diverse set of image classification tasks and modalities on Pneumonia-MNIST [12], Path-MNIST [11], Blood-MNIST [1], APTOS-2019 [10], and ISIC-2018 [4] datasets. Furthermore, we show that CLINICAL outperforms the state-of-the-art AL methods by up to $\approx 6\%-10\%$ on an average in terms of the average rare classes accuracy for binary imbalance scenarios and long-tail imbalance scenarios. **4)** We provide valuable insights about the *choice* of submodular functions to be used for subset selection based on the *modality* of medical data.

2 Preliminaries

Submodular Functions: We let $\mathcal{V}$ denote the *ground-set* of n data points $\mathcal{V} = \{1, 2, 3, ..., n\}$ and a set function $f : 2^{\mathcal{V}} \rightarrow \mathbb{R}$. The function f is submodular [5] if it

satisfies diminishing returns, namely $f(j|\mathcal{X}) \geq f(j|\mathcal{Y})$ for all $\mathcal{X} \subseteq \mathcal{Y} \subseteq \mathcal{V}, j \notin \mathcal{Y}$. Facility location, graph cut, log determinants, *etc.* are some examples [9].

Submodular Mutual Information (SMI): Given a set of items $\mathcal{A}, \mathcal{Q} \subseteq \mathcal{V}$, the submodular mutual information (MI) [6,8] is defined as $I_f(\mathcal{A}; \mathcal{Q}) = f(\mathcal{A}) + f(\mathcal{Q}) - f(\mathcal{A} \cup \mathcal{Q})$. Intuitively, this measures the similarity between $\mathcal{Q}$ and $\mathcal{A}$ and we refer to $\mathcal{Q}$ as the query set. [18] extend SMI to handle the case when the *target* can come from a different set $\mathcal{V}'$ apart from the ground set $\mathcal{V}$. In the context of imbalanced medical image classification, $\mathcal{V}$ is the source set of images and the query set $\mathcal{Q}$ is the target set containing the rare class images. To find an optimal subset given a query set $\mathcal{Q} \subseteq \mathcal{V}'$, we can define $g_{\mathcal{Q}}(\mathcal{A}) = I_f(\mathcal{A}; \mathcal{Q})$, $\mathcal{A} \subseteq \mathcal{V}$ and maximize the same.

2.1 Examples of SMI Functions

For targeted active learning, we use the recently introduced SMI functions in [6,8] and their extensions introduced in [18] as acquisition functions. For any two data points $i \in \mathcal{V}$ and $j \in \mathcal{Q}$, let s_{ij} denote the similarity between them.

Graph Cut MI (GCMI): The SMI instantiation of graph-cut (GCMI) is defined as: $I_{GC}(\mathcal{A}; \mathcal{Q}) = 2 \sum_{i \in \mathcal{A}} \sum_{j \in \mathcal{Q}} s_{ij}$. Since maximizing GCMI maximizes the joint pairwise sum with the query set, it will lead to a summary similar to the query set Q. In fact, specific instantiations of GCMI have been intuitively used for query-focused summarization for videos [27] and documents [20,21].

Facility Location MI (FLMI): We consider two variants of FLMI. The first variant is defined over $\mathcal{V}$(FLVMI), the SMI instantiation can be defined as: $I_{FLV}(\mathcal{A}; \mathcal{Q}) = \sum_{i \in \mathcal{V}} \min(\max_{j \in \mathcal{A}} s_{ij}, \max_{j \in \mathcal{Q}} s_{ij})$. The first term in the min(.) of FLVMI models diversity, and the second term models query relevance.

For the second variant, which is defined over $\mathcal{Q}$ (FLQMI), the SMI instantiation can be defined as: $I_{FLQ}(\mathcal{A}; \mathcal{Q}) = \sum_{i \in \mathcal{Q}} \max_{j \in \mathcal{A}} s_{ij} + \sum_{i \in \mathcal{A}} \max_{j \in \mathcal{Q}} s_{ij}$. FLQMI is very intuitive for query relevance as well. It measures the representation of data points that are the most relevant to the query set and vice versa.

Log Determinant MI (LOGDETMI): The SMI instantiation of LOGDETMI can be defined as: $I_{LogDet}(\mathcal{A}; \mathcal{Q}) = \log\det(S_{\mathcal{A}}) - \log\det(S_{\mathcal{A}} - S_{\mathcal{A},\mathcal{Q}} S_{\mathcal{Q}}^{-1} S_{\mathcal{A},\mathcal{Q}}^T)$. $S_{\mathcal{A},\mathcal{Q}}$ denotes the cross-similarity matrix between the items in sets $\mathcal{A}$ and $\mathcal{Q}$.

3 CLINICAL: Our Targeted Active Learning Framework for Binary and Long-Tail Imbalance

In this section, we propose our targeted active learning framework, CLINICAL (see Fig. 2), and show how it can be applied to datasets with class imbalance. Concretely, we apply the SMI functions as acquisition functions for improving a

model's accuracy on rare classes at a given additional labeling cost (B instances) without compromising on the overall accuracy. The main idea in CLINICAL, is to use *only the misclassified* data points from a held-out target set $\mathcal{T}$ containing data points from the rare classes. Let $\hat{\mathcal{T}} \subseteq \mathcal{T}$ be the subset of misclassified data points. Then, we optimize the SMI function $I_f(\mathcal{A}; \hat{\mathcal{T}})$ using a greedy strategy [23].

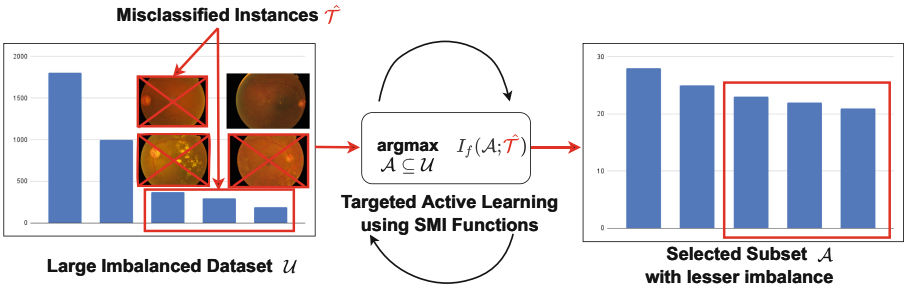

Fig. 2. The CLINICAL framework. We use a set of misclassified instances $\hat{\mathcal{T}}$ as the query set $\mathcal{Q}$ in the SMI function. We then maximize $I_f(\mathcal{A}; \hat{\mathcal{T}})$ in an AL loop to target the imbalance and gradually mine data points from the rare classes.

Note that since $\hat{\mathcal{T}}$ contains only the misclassified data points, it would contain more data points from classes that are comparatively *rarer* or the worst performing. Moreover, $\hat{\mathcal{T}}$ is updated in every AL round, this mechanism helps the SMI functions to focus on classes that require the most attention. For instance, in the long-tail imbalance scenario (see Fig. 1), CLINICAL would focus more on the tail classes in the initial rounds of AL. Next, we discuss the CLINICAL algorithm in detail:

Algorithm: Let $\mathcal{L}$ be an initial training set of labeled instances and $\mathcal{T}$ be the target set containing examples from the rare classes. Let $\mathcal{U}$ be a large unlabeled dataset and $\mathcal{M}$ be the trained model using $\mathcal{L}$. Next, we compute $\hat{\mathcal{T}}$ as the subset of data points from $\mathcal{T}$ that were misclassified by $\mathcal{M}$. Using last layer gradients as a representation for each data point which are extracted from $\mathcal{M}$, we compute similarity kernels of elements within $\mathcal{U}$, within $\hat{\mathcal{T}}$ and between $\mathcal{U}$ and $\hat{\mathcal{T}}$ to instantiate an SMI function $I_f(\mathcal{A}; \hat{\mathcal{T}})$ and maximize it to compute an optimal subset $\mathcal{A} \subseteq \mathcal{U}$ of size B given $\hat{\mathcal{T}}$ as target (query) set. We then augment $\mathcal{L}$ with labeled $\mathcal{A}$ (i.e. $L(\mathcal{A})$) and re-train the model to improve the model on the rare classes.

Algorithm 1. CLINICAL: Targeted AL for binary and long-tail imbalance

Require: Initial Labeled set of data points: $\mathcal{L}$, unlabeled dataset: $\mathcal{U}$, target set: $\mathcal{T}$, Loss function $\mathcal{H}$ for learning model $\mathcal{M}$, batch size: B, number of selection rounds: N

for selection round $i = 1 : N$ **do**
  Train $\mathcal{M}$ with loss $\mathcal{H}$ on the current labeled set $\mathcal{L}$ and obtain parameters θ_i
  Compute $\hat{\mathcal{T}} \subseteq \mathcal{T}$ that were misclassified by the trained model $\mathcal{M}$
  Use $\mathcal{M}_{\theta_i}$ to compute gradients using hypothesized labels $\{\nabla_\theta \mathcal{H}(x_j, \hat{y_j}, \theta), \forall j \in \mathcal{U}\}$ and obtain a pairwise similarity matrix X. {where $X_{ij} = \langle \nabla_\theta \mathcal{H}_i(\theta), \nabla_\theta \mathcal{H}_j(\theta) \rangle$}
  Instantiate a submodular function f based on X.
  $\mathcal{A}_i \leftarrow \text{argmax}_{\mathcal{A} \subseteq \mathcal{U}, |\mathcal{A}| \leq B} I_f(\mathcal{A}; \hat{\mathcal{T}})$
  Get labels $L(\mathcal{A}_i)$ for batch $\mathcal{A}_i$, and $\mathcal{L} \leftarrow \mathcal{L} \cup L(\mathcal{A}_i), \mathcal{U} \leftarrow \mathcal{U} - \mathcal{A}_i$
  $\mathcal{T} \leftarrow \mathcal{T} \cup \mathcal{A}_i^{\mathcal{T}}$, augment $\mathcal{T}$ with new data points that belong to target classes.
end for
Return trained model $\mathcal{M}$ and parameters θ_N.

4 Experiments

In this section, we evaluate the effectiveness of CLINICAL on binary imbalance (Sect. 4.1) and long-tail imbalance (Sect. 4.2) scenarios. We do so by comparing the accuracy and class selections of various SMI functions with the existing state-of-the-art AL approaches. In our experiments, we observe that different SMI functions outperform existing approaches depending on the modality of the medical data. We show that the choice of the SMI based acquisition function is imperative and varies based on the imbalance scenario and the modality of medical data.

Baselines in all Scenarios. We compare the performance on CLINICAL against a variety of state-of-the-art uncertainty, diversity and targeted selection methods. The uncertainty based methods include ENTROPY, LEAST CONFIDENCE (LEAST-CONF), and MARGIN. The diversity based methods include CORESET and BADGE. The targeted selection methods include T-GLISTER and T-GRADMATCH. We discuss the details of all baselines in Sect. 1.1. For a fair comparison with CLINICAL, we use the same target set of misclassified data points $\hat{\mathcal{T}}$ as the held out validation set used in T-GLISTER and T-GRADMATCH. Lastly, we compare with random sampling (RANDOM).

Experimental Setup: We use the same training procedure and hyperparameters for all AL methods to ensure a fair comparison. For all experiments, we train a ResNet-18 [7] model using an SGD optimizer with an initial learning rate of 0.001, the momentum of 0.9, and a weight decay of 5e−4. For each AL round, the weights are reinitialized using Xavier initialization and the model is trained till 99% training accuracy. The learning rate is decayed using cosine annealing [22] in every epoch. We run each experiment 5× on a V100 GPU and provide the error bars (std deviation). We discuss dataset splits for each our experiments below and provide more details in Appendix. B.

4.1 Binary Imbalance

Datasets: We apply our framework to **1)** Pneumonia-MNIST (pediatric chest X-ray) [12,29], **2)** Path-MNIST (colorectal cancer histology) [11,29], and **3)**Blood-MNIST (blood cell microscope) [1,29] medical image classification datasets. To create a more realistic medical scenario, we create a custom dataset that simulates binary class imbalance for each of these datasets for our experiments. Let $\mathcal{C}$ be the set of data points from the rare classes and $\mathcal{D}$ be the set of data points from the over-represented classes. We create the initial labeled set $\mathcal{L}$ (seed set) in AL, $|\mathcal{D}_\mathcal{L}| = \rho|\mathcal{C}_\mathcal{L}|$ and an unlabeled set $\mathcal{U}$ such that $|\mathcal{D}_\mathcal{U}| = \rho|\mathcal{C}_\mathcal{U}|$, where ρ is the imbalance factor. We use a small held out target set $\mathcal{T}$ which contains data points from the rare classes. For Path-MNIST and PneumoniaMNIST, we use $\rho = 20$, and for Blood-MNIST, we use $\rho = 7$ due to the small size of the dataset. For Pneumonia-MNIST, $|\mathcal{C}_\mathcal{L}| + |\mathcal{D}_\mathcal{L}| = 105$, $|\mathcal{C}_\mathcal{U}| + |\mathcal{D}_\mathcal{U}| = 1100$, $B = 10$ (AL batch size) and, $|\mathcal{T}| = 5$. Following the natural class imbalance, we use the 'pneumonia' class as the rare class. For Path-MNIST, $|\mathcal{C}_\mathcal{L}| + |\mathcal{D}_\mathcal{L}| = 3550$, $|\mathcal{C}_\mathcal{U}| + |\mathcal{D}_\mathcal{U}| = 56.8K$, $B = 500$ and, $|\mathcal{T}| = 20$. Following the natural class imbalance, we use two classes from the dataset ('mucus', 'normal colon mucosa') as rare classes. For Blood-MNIST, $|\mathcal{C}_\mathcal{L}| + |\mathcal{D}_\mathcal{L}| = 228$, $|\mathcal{C}_\mathcal{U}| + |\mathcal{D}_\mathcal{U}| = 1824$, $B = 20$

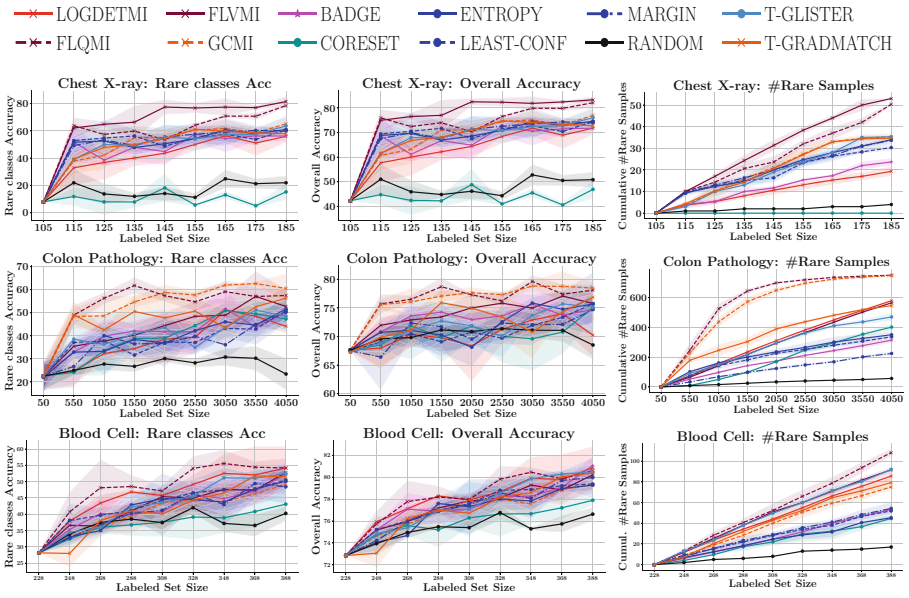

Fig. 3. AL for binary imbalanced medical image classification on Pneumonia-MNIST [12] (**first** row), Path-MNIST [11] (**second** row), and Blood-MNIST [1] (**third** row). CLINICAL outperforms the existing AL methods by $\approx 2\%-12\%$ on the rare classes acc. (**left** col.) and $\approx 2\%-6\%$ on overall acc. (**center** col.). SMI functions select the most number of rare class samples (**right** col.)

and, $|\mathcal{T}| = 20$. Following the natural class imbalance, we use four classes from the dataset ('basophil', 'eosinophil', 'lymphocyte', 'neutrophil') as rare classes.

Results: The results for the binary imbalance scenario are shown in Fig. 3. We observe that the CLINICAL consistently outperform other methods by $\approx$ 2%−12% on the rare classes accuracy (Fig. 3(left column)) and $\approx$ 2%−6% on overall accuracy (Fig. 3(center column)). This is due to the fact that the SMI functions are able to select significantly more data points that belong to the rare classes (Fig. 3(right column)). Particularly, we observe that when the data modality is *X-ray* (Pneumonia-MNIST), the facility location based SMI variants, FLVMI and FLQMI perform significantly better than other acquisition functions due to their ability to model *representation*. For the *colon pathology* modality (Path-MNIST), GCMI and FLQMI functions that model *query-relevance* significantly outperform other methods. Lastly, for the blood cell microscope modality (Blood-MNIST), we observe some improvement using FLQMI, although it selects many points from the rare classes.

4.2 Long-Tail Imbalance

Datasets: We apply CLINICAL to two datasets that naturally show a long-tail distribution: **1)** The ISIC-2018 skin lesion diagnosis dataset [4] and **2)** APTOS-2019 [10] for diabetic retinopathy (DR) grading from retinal fundus images. We evaluate all AL methods on a balanced test set to obtain a fair estimate of accuracy across all classes. We split the remaining data randomly with 20% into the initial labeled set $\mathcal{L}$ and 80% into the unlabeled set $\mathcal{U}$. We use a small held-out

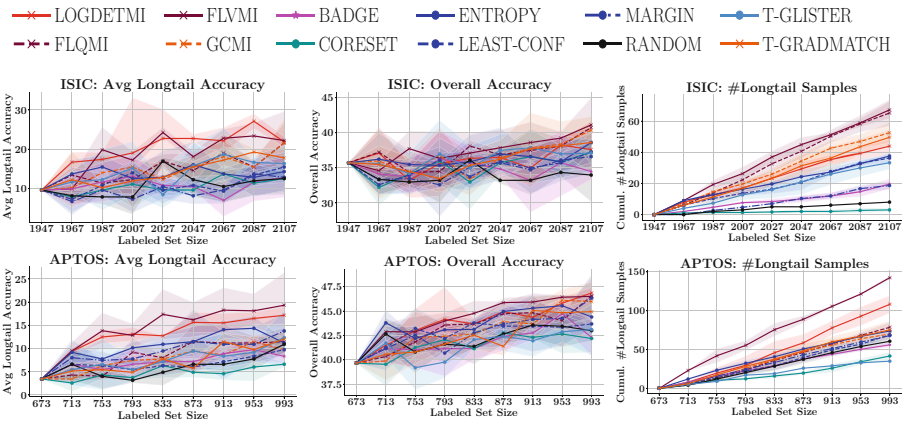

Fig. 4. Active learning for long-tail imbalanced medical image classification on ISIC-2018 [4] (**first** row) and APTOS-2019 [10] (**second** row). CLINICAL outperforms the state-of-the-art AL methods by $\approx$ 10%−12% on the average long-tail accuracy (**left** col.) and $\approx$ 2%−5% on overall accuracy (**center** col.). SMI functions select the most number of long-tail class samples (**right** col.)

target set $\mathcal{T}$ with data points from the classes at the tail of the distribution (long-tail classes, see Fig. 1). For ISIC-2018, we use the bottom three infrequent skin lesions from the tail of the distribution as long-tail classes ('bowen's disease', 'vascular lesions', and 'dermatofibroma'). We set $B = 40$ and $|\mathcal{T}| = 15$. For APTOS-2019 we use the bottom two infrequent DR gradations as long-tail classes ('severe DR' and 'proliferative DR') (see Fig. 1). We set $B = 20$ and $|\mathcal{T}| = 10$.

Results: We present the results for the long-tail imbalance scenario in Fig. 4. We observe that CLINICAL consistently outperform other methods by $\approx 10\%-12\%$ on the average long-tail classes accuracy (Fig. 4(left column)) and $\approx 2\%-5\%$ on the overall accuracy (Fig. 4(center column)). This is because the SMI functions select significantly more data points from the long-tail classes (Fig. 4(right column)). On both datasets, we observe that the functions modeling query-relevance *and* diversity (FLVMI and LOGDETMI) outperform the functions modeling *only* query-relevance (FLQMI and GCMI).

5 Conclusion

We demonstrate the effectiveness of CLINICAL for a wide range of medical data modalities for binary and long-tail imbalance. We empirically observe that the current methods in active learning cannot be directly applied to medical datasets with rare classes, and show that a targeting mechanism like SMI can greatly improve the performance on rare classes.

References

1. Acevedo, A., Merino, A., Alférez, S., Molina, Á., Boldú, L., Rodellar, J.: A dataset of microscopic peripheral blood cell images for development of automatic recognition systems. Data Brief **30** (2020). ISSN 2352-3409
2. Arthur, D., Vassilvitskii, S.: k-means++: the advantages of careful seeding. In: SODA 2007: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 1027–1035. Society for Industrial and Applied Mathematics, Philadelphia (2007)
3. Ash, J.T., Zhang, C., Krishnamurthy, A., Langford, J., Agarwal, A.: Deep batch active learning by diverse, uncertain gradient lower bounds. In: ICLR (2020)
4. Codella, N., et al.: Skin lesion analysis toward melanoma detection 2018: a challenge hosted by the international skin imaging collaboration (ISIC). arXiv preprint arXiv:1902.03368 (2019)
5. Fujishige, S.: Submodular Functions and Optimization. Elsevier, Amsterdam (2005)
6. Gupta, A., Levin, R.: The online submodular cover problem. In: ACM-SIAM Symposium on Discrete Algorithms (2020)
7. 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)

8. Iyer, R., Khargoankar, N., Bilmes, J., Asnani, H.: Submodular combinatorial information measures with applications in machine learning. arXiv preprint arXiv:2006.15412 (2020)
9. Iyer, R.K.: Submodular optimization and machine learning: theoretical results, unifying and scalable algorithms, and applications. Ph.D. thesis (2015)
10. Kaggle: Aptos 2019 blindness detection (2019). https://www.kaggle.com/c/aptos2019-blindness-detection/data
11. Kather, J.N., et al.: Predicting survival from colorectal cancer histology slides using deep learning: a retrospective multicenter study. PLoS Med. **16**(1), e1002730 (2019)
12. Kermany, D.S., et al.: Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell **172**(5), 1122–1131 (2018)
13. Killamsetty, K., Durga, S., Ramakrishnan, G., De, A., Iyer, R.: Grad-match: gradient matching based data subset selection for efficient deep model training. In: International Conference on Machine Learning, pp. 5464–5474. PMLR (2021)
14. Killamsetty, K., Sivasubramanian, D., Ramakrishnan, G., Iyer, R.: Glister: generalization based data subset selection for efficient and robust learning. In: AAAI (2021)
15. Kirsch, A., Van Amersfoort, J., Gal, Y.: Batchbald: efficient and diverse batch acquisition for deep Bayesian active learning. arXiv preprint arXiv:1906.08158 (2019)
16. Kothawade, S., Beck, N., Killamsetty, K., Iyer, R.: Similar: submodular information measures based active learning in realistic scenarios. In: Advances in Neural Information Processing Systems, vol. 34 (2021)
17. Kothawade, S., Ghosh, S., Shekhar, S., Xiang, Y., Iyer, R.: Talisman: targeted active learning for object detection with rare classes and slices using submodular mutual information. arXiv preprint arXiv:2112.00166 (2021)
18. Kothawade, S., Kaushal, V., Ramakrishnan, G., Bilmes, J., Iyer, R.: Prism: a rich class of parameterized submodular information measures for guided subset selection. arXiv preprint arXiv:2103.00128 (2021)
19. Kothyari, M., Mekala, A.R., Iyer, R., Ramakrishnan, G., Jyothi, P.: Personalizing ASR with limited data using targeted subset selection. arXiv preprint arXiv:2110.04908 (2021)
20. Li, J., Li, L., Li, T.: Multi-document summarization via submodularity. Appl. Intell. **37**(3), 420–430 (2012)
21. Lin, H.: Submodularity in natural language processing: algorithms and applications. Ph.D. thesis (2012)
22. Loshchilov, I., Hutter, F.: SGDR: stochastic gradient descent with warm restarts. arXiv preprint arXiv:1608.03983 (2016)
23. Mirzasoleiman, B., Badanidiyuru, A., Karbasi, A., Vondrák, J., Krause, A.: Lazier than lazy greedy. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 29 (2015)
24. Roth, D., Small, K.: Margin-based active learning for structured output spaces. In: Fürnkranz, J., Scheffer, T., Spiliopoulou, M. (eds.) ECML 2006. LNCS (LNAI), vol. 4212, pp. 413–424. Springer, Heidelberg (2006). https://doi.org/10.1007/11871842_40
25. Sener, O., Savarese, S.: Active learning for convolutional neural networks: a core-set approach. In: International Conference on Learning Representations (2018)
26. Settles, B.: Active learning literature survey. Technical report, University of Wisconsin-Madison, Department of Computer Sciences (2009)

27. Vasudevan, A.B., Gygli, M., Volokitin, A., Van Gool, L.: Query-adaptive video summarization via quality-aware relevance estimation. In: Proceedings of the 25th ACM International Conference on Multimedia, pp. 582–590 (2017)
28. Wang, D., Shang, Y.: A new active labeling method for deep learning. In: 2014 International Joint Conference on Neural Networks (IJCNN), pp. 112–119. IEEE (2014)
29. Yang, J., et al.: MedMNIST v2: A large-scale lightweight benchmark for 2D and 3D biomedical image classification. arXiv preprint arXiv:2008 (2021)