Yasuhiro Iida[1]
Bojan Mrazovac[1]
Yasuo Ishigure[1]

DOI10.7251/ZJF2514259I

LABEL ENTROPY WITH SIMILARITY GRAPH CLIQUE FOR ASSESSING ANNOTATION QUALITY

Abstract: We introduce simple metrics using the entropy of label distribution in local maximum cliques of similarity graph, which can assess human annotation quality for large image datasets. Since the annotation is done by a human task, it always contains potential errors hidden in a dataset due to manual fluctuation or inconsistency. This annotation error is crucial, especially in medical image multi-class classification as a label noise when we create a classifier with machine learning. In our work, we focused on how to assess the entire label quality in a large dataset. To this end, we proposed novel metrics for assessing the label quality of datasets. We also assessed existing label noise detection methodologies with our metrics and found that the transformed label propagation with label-smoothing methodology, which is first proposed here, showed a quite high accuracy among the three methodologies.

Keywords: entropy, label propagation, graph-based machine learning, annotation

ENTROPIJA OZNAKA S GRAFIKONOM SLIČNOSTI KLIKA ZA PROCENU KVALITETA ANOTACIJA

Apstrakt: Uvodimo jednostavne metrike koristeći entropiju distribucije oznaka u lokalnim maksimalnim klikovima grafikona sličnosti, koja može proceniti kvalitet ljudskih beleški za velike skupove podataka slika. Budući da se zabeleške obavljaju od strane ljudskih zadataka, uvek sadrže potencijalne greške skrivene u skupu podataka zbog ručne fluktuacije ili nedoslednosti. Ova greška u napomeni je ključna, posebno u višeklasnoj klasifikaciji medicinskih slika kao šum oznake kada kreiramo klasifikator sa mašinskim učenjem. U svom radu fokusirali smo se na to kako proceniti celokupni kvalitet oznake u velikom skupu podataka. U tu svrhu, predložili smo nove metrike za procenu kvaliteta oznaka skupova podataka. Također smo procenili postojeće metodologije za detekciju šuma na oznaci pomoću naših metrika i otkrili da je transformisano širenje oznaka s metodologijom zaglađivanja oznaka, koja je ovde prvi put predložena, pokazala prilično visoku tačnost među tri metodologije.

Ključne reči: entropija, propagacija oznaka, grafičko-bazirano mašinsko učenje, anotacija

1 Digital Transformation Office, NTT DATA Romania, Novi Sad, Serbia
iida.yasuhiro@nttdata.com, bojan.mrazovac@nttdata.com, ishigure@fun.ac.jp

1. INTRODUCTION

The annotations to multiple class image datasets are subject to fluctuation and inconsistency of human manual tasks. It is crucial, especially in medical image classification, which usually varies from 3-class to 7-class according to the stage of disease [1]. Since these low-quality annotations are known as label noise in machine learning, reducing label noise is a key element for building a high-performance classifier. Nowadays, the machine learning algorithm is more and more sophisticated, this label noise is a more important factor in ensuring the performance of the machine learning model accordingly. Nevertheless, few studies have dealt with this problem to date. There are two reasons for this, one is a lack of information about the quality of annotations. Once a label is assigned, the annotation work behind labeling is usually not recorded, making it inherently difficult to verify its reliability afterward. We have tackled this issue by expanding the existing label propagation method to unsupervised learning to detect probable low-quality labels that spread multiple categories [2, 3]. The other is a lack of indicators to assess the degree of annotation quality of the dataset. There are no supervised data to assess the annotation quality, which means that we have no ground truth label. Therefore, in order to assess the label quality of the dataset, developing a standard methodology of explainable, plausible, and applicable to any dataset with labels is one of the imminent issues.

Despite several research efforts that have been made to tackle label noise detection and to assess their performance [2, 3], it is still substantially difficult because there are no available ground truth labels. This is similar to the common recognitions when we evaluate the performance of unsupervised learning, it is necessary to develop performance metrics or indicators that should be adopted for individual unsupervised learning. For example, graph clustering, which is one of the well-known unsupervised learning, modularity [4, 5] or normalized cut [6] is conventionally used for the criteria of good clustering. In anomaly detection, we conventionally use precision and recall, even though it is only useful if we can somehow evaluate the normal and abnormal with domain experts. On the other hand, when it comes to label quality assessment, especially in the medical field, there are fewer evaluation guidelines and difficulties due to the graded labeling of medical images across multiple categories. Even if domain experts gather and evaluate the label validity, the unified consensus of determining one label for a diagnostic image is hard to acquire due to discrepancies among experts' opinions.

In this work, we propose novel and simple metrics that can assess label noise quantitatively. The concept behind this metric is a hypothesis that we can infer the quality of labels from the relationships between data. In other words, through this process, we can infer the label quality without knowledge of label quality. Our primary contribution is introducing a standard methodology to assess the entire label quality in a dataset. In addition, we introduced the label smoothing technique to detect label noise with the highest accuracy compared to existing works. Furthermore, our contribution to this work lies in pioneering a graph-based approach to deal with assessing and detecting label noise spread in multiple categories at once. In the following part of this paper, the manual classification task is referred to as an 'annotation', and the classification result by this manual annotation or result by any computation is referred to as a 'label', both of which shall be used accordingly.

2. METHOD AND DATASET

A. Problem Definition

Under the situation where labeled datasets of multi-class categories are given and no information about the reliability of the label and ground truth is given, a methodology to assess the entire label quality in the datasets and an effective methodology to find low-quality labels of each class shall be invented.

B. Basic Approach for Defining Metrics

Toward inventing these methodologies, we consider extracting the underlying principles that datasets intrinsically have. To this end, we focus on the dataset as a whole instead of individual labeled data. More specifically, we build a similarity graph among data to extract the relationship between all data with labels. The node of the graph denotes each data such as an image with its label. The edge of the graph denotes the similarity between two nodes. This graph also allows us to extract the inconsistency between data similarity and their labels simultaneously. This is the primary idea to approach assessing the label quality. We will then explain how this inconsistency can be quantified. Firstly, we find the graph cliques [7] from the similarity graph. A clique is a subgraph where every pair of distinct nodes is connected by an edge. This means that all nodes in the clique are quite similar to each other. Although the similarity graph does not always have a clique due to the feature of a complex network [8], it tends to have several cliques due to its high clustering coefficient [7]. This is one of the key points we focus on cliques here. Secondly, we represent the label distribution of each clique with entropy. If the entropy is high, the label distribution is more uniform across a clique, with a wider spread of many labels. On the other hand, if the entropy is low, a smaller variety of labels is in the clique.

Fig.1 shows our basic idea focusing on a clique with its label distribution. (A) denotes a clique of six nodes with two kinds of labels, and (B) denotes a clique of six nodes with three kinds of labels. Obviously, the more diverse the types of labels, and the closer their distribution is to a uniform distribution, the higher the entropy. Therefore, the entropy of (B) is higher than that of (A), which suggests data in (B) contains low-quality labels.

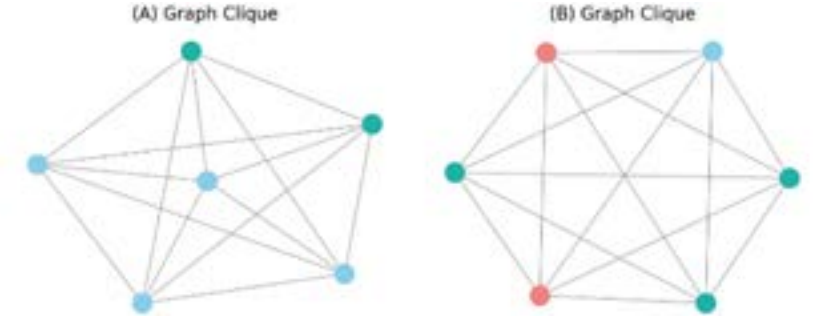

Fig. 1. Example diagram of label distribution in a graph clique.

C. Dataset

We verified our method with three types of medical image datasets, including single-cell and cluster-cell of cervical cancer: the 917 images of single-cell from Herlev datasets (7-class) [9], 950 images of cluster-cell from SIPaKMeD datasets (5-class), and 4049 images of single-cell from SIPaKMeD datasets (5-class) [10]. These 4049 images are cropped manually as isolated single-cell images from cluster-cell images. These datasets are commonly recognized in cervical cytology. The Herlev dataset is colored and provided in BMP format with approximately 100 KB. The SIPaKMeD dataset is also Pap smear images captured with a CCD camera attached to an optical microscope and classified into 5-class. This dataset is colored and provided in BMP format with approximately 1MB each and has variations in resolutions and pixel sizes. Fig.2 (a) shows the example images of the Herlev (single-cell), Fig.2 (b) shows the example images of the SIPaKMeD (cluster-cell), and Fig.2 (c) shows the example images of the SIPaKMeD (single-cell).

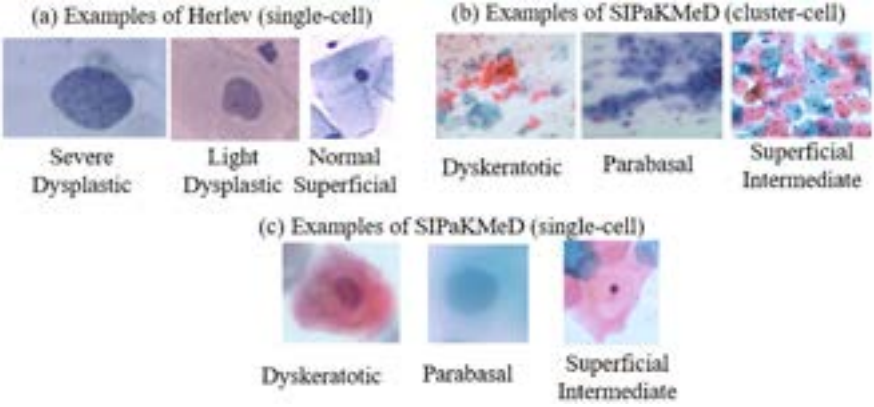

Figure 2. Example images of Herlev and SIPaKMeD.

The number of data (images) in each class is listed in Tables Ⅰ, Ⅱ, and Ⅲ. In parentheses in the class column represent the abbreviations of full class names.

Table Ⅰ: Description of Herlev Dataset(single-cell)

Category	Class	Data samples (917 in total)
Abnormal	Light Dysplastic (LD)	182
Abnormal	Moderate Dysplastic (MD)	146
Abnormal	Carcinoma in Situ (CS)	150
Abnormal	Severe Dysplastic (SD)	197
Normal	Normal Columnar (NC)	98
Normal	Normal Superficial (NS)	74
Normal	Normal Intermediate (NI)	70

Table II: Description of SIPaKMeD Dataset(cluster-cell)

Category	Class	Data samples (950 in total)
Abnormal	Dyskeratotic (DY)	223
Abnormal	Koilocytotic (KO)	232
Benign	Metaplastic (ME)	271
Normal	Parabasal (PA)	108
Normal	Superficial Intermediate (SU)	116

Table III: Description of SIPaKMeD Dataset(single-cell)

Category	Class	Data samples (4049 in total)
Abnormal	Dyskeratotic (DY)	813
Abnormal	Koilocytotic (KO)	825
Benign	Metaplastic (ME)	793
Normal	Parabasal (PA)	787
Normal	Superficial Intermediate (SU)	831

D. Proposed Method

We start by building a similarity graph with images as nodes with labels. In converting an image into a vector, we use the vision transformer which extracts the features of each image. The vision transformer develops attention technology by dividing an image into small patches of 16 x 16 pixels and vectorizing each patch by treating it as a token. Then we have a 768-dimensions feature vector of each image. Although the target resolutions and pixel sizes of these images vary from low to high, and from small to large (some images exceed 400 x 400 pixels), this vision transformer allows these images to be treated as feature vectors of the same dimension robustly. In the Herlev and SIPaKMeD datasets, ith diagnostic image (node) is expressed as a 768-dimensional vector x_i where i∈V. One of the benefits of the vision transformer is the stability in converting different image pixels into a uniform feature vector. While several methods are proposed [11, 12] to convert cervical cytology images to feature vectors, we adopted vit-base-patch16-224 [13] in our model considering its proven track record in a wider range of experiments. To calculate the similarity between feature vectors, we tried three distance functions: cosine similarity, Minkowski distance, and Hausdorff distance [14]. Among these functions, the cosine similarity showed a good similarity representation for both single-cell and cluster-cell image datasets. To build the graph, we have to decide on the algorithm and the similarity threshold. One of the commonly recognized algorithms is the cosine similarity. We adopted the cosine similarity to calculate edges, and these weighted edges were added only when the similarity was higher than 0.74. If we set higher values as the similarity, edges will decrease and if we set lower values, edges will increase. We have to note here that if we increase this threshold, isolated nodes are more likely to occur. We found the value of 0.74 as a threshold to be the most suitable for dealing with the feature vector from images. Weighted edges built between nodes based on their similarity are normalized into a numerical value between 0 and 1. Then we

have a weighted undirected graph. For example, in the Herlev dataset of single-cell images, the number of nodes is 917, which means that the adjacency matrix dimension is 917 x 917. The number of edges is 11431. The average degree is 24.9, and the average cluster coefficient is 0.347. To find cliques in the graph, we can select any existing algorithm to explore for cliques, such as the Bron-Kerbosch Algorithm [7]. In this work, we adopt a standard greedy algorithm as the most fundamental algorithm by starting from each node and sequentially adding adjacent nodes to the clique, then enumerating cliques. Table Ⅳ shows a characteristic of each graph built from the Herlev and the SIPaKMeD datasets.

Table Ⅳ: The Characteristics of Image Similarity Graph

	Herlev (single-cell)	SIPaKMeD (cluster-cell)	SIPaKMeD (single-cell)
# of nodes	917	950	4049
# of edges	11431	10929	82426
Average Degree	24.9	12.1	40.7
Average Clustering Coefficient	0.347	0.445	0.341
Maximum clique size (# of nodes)	19	30	60

The entropy of label distribution in a clique H(C) is expressed in (1). Let V_C be the set of nodes belonging to a clique C, and L be the set of labels. Now we explain how to express the metrics of label quality in (1), where l∈L.

$$H(C)=-\sum_{l} P(l|C)\log_2 P(l|C) \quad (1)$$

P(l|C) is the probability of the label l appearing in a clique C, that is,

$P(l|C)=\text{count}(l)/|V_C|$, and $|V_C|$ denotes the number of nodes in a clique C. Then, we calculate the average entropy of each clique in the graph. Thus, we have metrics of label quality assessment of the graph.

Having discussed the metrics above, we now describe our proposed method for detecting label noise with high accuracy. Our method is based on the transformed label propagation [2] and improves this with a label-smoothing technique. The transformed label propagation (T-LP) is one of the methods to detect low-quality labels and to show possible correct labels at the same time. T-LP is an unsupervised learning, which is one of the derivatives of label propagation (also known as label spreading) and is expressed in (2).

$$F=(I-\alpha S)^{(-1)}\, Y \quad (2)$$

S is a normalized adjacency matrix representing the undirected and weighted graph structure G= (V,E,L), where V denotes a set of nodes (also known as vertexes), E denotes a set of edges, and L denotes a set of labels. When we denote the number of nodes as N, S is a square matrix with the dimension of N×N. Each element contains a non-negative value between 0 and 1 if there is an edge between nodes.

If the similarity between nodes is high, the value will be closer to 1, and if there is no edge, this value equals zero. I is the identity matrix. Y is an initial label matrix. The dimensions of Y are N×c where c represents the number of classes. For example, when we classify images according to the degree of cancer progression into seven stages based on images, we will set c=7. In the initial label matrix for the Herlev dataset, the rows of this matrix correspond to the classes of LD, MD, CS, SD, NC, NS, and NI in that order from top to bottom. Likewise, the columns of this matrix correspond to the same order from left to right. Looking at this matrix row-wise, we can see that each row is a one-hot vector where only the element corresponding to the assigned label has a value of 1. α is a scalar constant that controls the label propagation; the smaller α is, the more it is affected by the given label, and the larger α is, the more it is affected by the graph structure. This formula is calculated to obtain the label estimation result F, that is, F is an estimated label score matrix. Y of the transformed label propagation (T-LP) is built by randomly selected labels (which are initial labels) from assigned labels in the matrix. Then, is a mixture of a supervised label (Y_(i,l)=1) and an unknown label (Y_(i,l)=0), where i∈V, then, we set all labels other than the selected initial label to zero (Y_(i,j)=0). F_final is the calculated result of several times selecting the initial label randomly in an ensemble manner, that is, F_final =mode(F). This means that the most frequent value among the labels given to each node is regarded as the final label within the trial number.

The label-smoothing technique is an operation L_S of the initial label matrix Y to normalize the value row-wise, Y^'=L_S (Y). Y^' is no longer a one-hot vector (0 or 1). Instead, Y_(i,l)^' has a value of $0<Y_{(i,l)}^{'}<1$ and $\sum_i Y_{(i,l)}^{'}=1$. Any function can be selected as L_S as long as it satisfies this condition.

E. Experimental Setup

To process two types of graphs in our experiment, nodes are stored in a JSON format file, in which the key is an image file name and the corresponding value is its feature vector. Since each of these datasets does not include label noise that should be used for our experiment verification, we intentionally and randomly gave pseudo-label noise to them. The label noise is defined here as the set of incorrect labels. The percentage of label noise varied from 5% to 30% of the total in 5% increments. For example, the Herlev dataset has 917 labels according to 917 nodes. If we produce 10% label noise in this dataset, it means 91 nodes are selected and changed intentionally into incorrect labels. Our experiment was executed on an Intel Core i7 1.8 GHz processor and 16 GB of memory. Any specific environment, such as a GPU and TPU, was not required for our work.

We evaluated the label quality of datasets with three patterns and one baseline and then compared the entropy. The baseline is a graph built from the Herlev and SIPaKMeD datasets as is. The three patterns consist of a graph with label noise (noise-included graph), a graph of transformed label propagation (T-LP) results from the noise-included graph, and a graph of T-LP with the label-smoothing each. We executed T-LP with the initial label ratio of 70%, α was set to 0.014, and the trial number of 11 times according to [2]. We introduced a threshold in clique size with over 10 nodes and calculated the average entropy of such cliques. This is to eliminate the

effects of excessively small cliques. As a label-smoothing technique, we set the initial label score of 0.7 in each node belonging to the class of LD, MD, CS, and SD, all of which are categorized as abnormal in the initial label matrix. As the label-smoothing technique is the operation of the initial label matrix to normalize the value row-wise, we set the initial label score of 0.1 each in the remaining elements corresponding to the abnormal category in the row.

3. RESULTS

A. Label Entropy of Several Noise Ratios

In the graph built from the Herlev dataset, we evaluated those cliques that had over 10 nodes. Each clique has nodes with labels that belong to each class in Table I. For example, the maximum size of the clique of the Herlev dataset had 19 nodes and consisted of eight SDs, seven CSs, two NCs, one MD, and one LD of label each. Then the label entropy of this clique is calculated at 1.845 according to (1). Thus the average label entropy of all cliques of over 10 nodes is calculated at 1.565. In the same way, in the graph built from the SIPaKMeD cluster-cell and single-cell dataset respectively, we evaluated those cliques that had over 10 nodes. The maximum size of the clique was 30 nodes and 60 nodes each. The average label entropy of these cliques is calculated at 0.718 and 0.298, respectively. These values are the baseline as the average label entropy represented in the blue lines respectively in Fig. 3 of (a) Herlev, (b) SIPaKMeD cluster-cell, and (c) SIPaKMeD single-cell. We have to note here that these baselines are not down to zero. This is because these datasets are also considered to inherently contain annotation uncertainties. The average label entropy represented in the orange lines in Fig. 3 corresponds to the graph with label noise (noise-included graph). Because the label noise was intentionally and randomly added to the graph, the orange lines always show higher entropy than the baselines. The average label entropy represented in the green lines in Fig. 3 corresponds to the graph of T-LP results. Because T-LP encourages label reassignment from the original noise-added graphs by utilizing the similarity graph structure [2], the green lines tend to show lower entropy than the orange lines. It is of particular importance to emphasize here that the average label entropy represented in the red lines in Fig. 3, which corresponds to the T-LP with the label-smoothing, shows almost always lower entropy than green lines. This suggests our proposed method of T-LP with label-smoothing is the most effective method to detect the low-quality labels from the dataset with label nose.

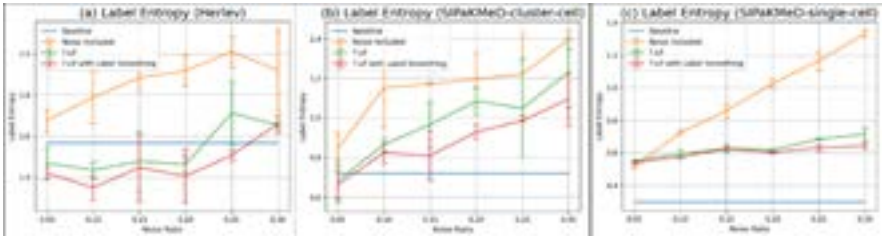

Figure 3. Label entropy of several noise ratios.

We have to note again that the performance of T-LP with label-smoothing (red lines) should be compared to T-LP (green lines), not to baseline (blue lines). This is because this experiment is always executed under noise-included situations.

B. Ablation Studies

In a graph-related process, both the threshold of cosine similarity of images and the control parameter of transformed label propagation α are tested for how these affect our results. We compared cosine similarity with a threshold from 0.6 to 0.8 for the building graph. The precision of the noise detection is sensitive to this threshold. When we tighten this threshold from 0.74 to 0.75, edges reduce from 11431 to 8612 in the Herlev single-cell dataset, which means that isolated nodes are more likely to remain, which will compromise the subsequent transformed label propagation. α is another important parameter that decides the label and is usually set around 0.2, however, we set much lower values varied from 0.01 to 0.02 for controlling the label decision process in the transformed label propagation aiming at minimizing the influence of initial labels which might have low-quality labels and maximizing the benefit of graph structures. Therefore, we found that 0.014 is the most suitable for all three datasets.

4. DISCUSSIONS

We explain several analyses here regarding each process, that is, graph building, graph clique search, and label-smoothing techniques for considering the robustness of our experiment results.

For graph building, we adopt cosine similarity for graph construction. This is a general method that builds weighted edges based on the similarity between nodes that have a high degree of similarity. We also used locally linear embedding (LLE) [15]. The LLE embeds a high-dimensional vector into a low-dimensional space while preserving the local linear relationship of data points. This is also a well-known approach to building a graph, however, we did not observe any noticeable improvement in lowering label entropy.

The graph clique is one of the key points of our idea because it represents the tight connection of similar data, and all labels of the nodes should substantially be the same accordingly. However, these labels are not always the same due to the fluctuation of annotations or inconsistencies in manual tasks. This is why we focus on this discrepancy as a good indicator to measure the label quality. Because the graph we built is a weighted graph, even if the same sized two cliques might have a difference in that they have different edge weights. Therefore, we evaluated how the edge weights affect the label quality assessment. To this end, we multiply the label entropy by the inversed value of the summation of edge weights in the clique. This is because a more weighted clique is more significant in the graph because of higher node similarity in the clique, compared to the less weighted clique. Fig. 4 shows the average label entropy similar to Fig. 3. The only difference here is that each label entropy of the clique is weighted by the edge weights in each clique in Fig. 4. Similar to Fig.3, we observed that T-LP with the label-smoothing (red lines) shows almost always lower entropy than T-LP (green lines) and both lines are always under graph with label noise (orange line).

The label-smoothing technique is another key point of our work to lower the label entropy. In our experiment, we distributed the initial score of the label only within abnormal categories, however, our interest here is how label entropy will change if the initial label score validation changes.

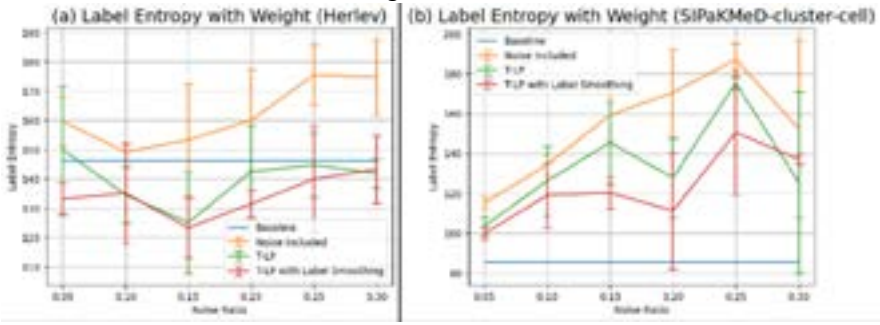

Figure 4. Label entropy with weighted cliques.

In our experimental setup, the initial label score is spread across each node belonging to the class of LD, MD, CS, and SD, all of which are categorized as abnormal. Here, we also distributed the initial label score across not only abnormal categories but also normal categories and then evaluated how label entropy changed with the Herlev dataset. Therefore, in addition to label-smoothing in abnormal categories, we set the initial label score of 0.8 in each node belonging to the class of NC, NS, and NI, all of which are categorized as normal. Following this, we set the initial label score of 0.1 each in the remaining elements corresponding to the normal category in the row of the initial label matrix. Fig. 5 shows both label-smoothing only in abnormal categories (red line) and label-smoothing in abnormal and normal categories (purple line). As an overall view, we observed that applying label-smoothing to both abnormal and normal categories tends to slightly lower label entropies across all label noise ratios.

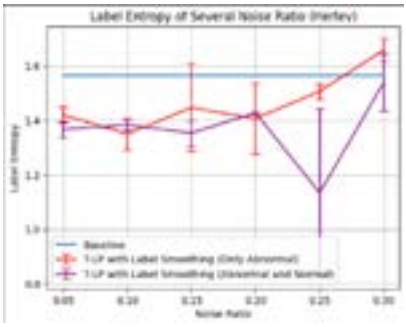

Figure 5. Label entropy with two types of label smoothing.

5. CONCLUSION

We focused on the label noise that appears in any dataset because of human annotation error and proposed standard metrics of label quality assessment as our primary contribution. In particular, since multiple class datasets such as cancer images are subject to manual fluctuation or inconsistency, it is crucial when we rely on machine learning for automated image classification in the medical field. Nevertheless, measurement and detection of low-quality labels have been open issues to be addressed to date, and few studies have tackled these issues due to the substantial difficulty caused by the lack of information in tracing the label quality. The essential point of our idea is that it does not require any supervised data, and more specifically, we focus on the discrepancy between data relations in similarity and assigned labels. If they are inconsistent with each other, it means that the data contains low-quality labels. Thus, the key point of our method is utilizing unsupervised learning since there is no supervised data regarding the assessment of the label qualities. We solved this problem with a series of processes of similarity graph building, finding graph cliques, and introducing label entropy calculation in each graph clique.

The secondary contribution of our work is introducing the label-smoothing technique which extends the existing transformed label propagation method and demonstrates its effectiveness in good detection of label noise. In light of our label entropy assessment, the label-smoothing technique that applies to all categories showed the highest performance among the three methodologies of transformed label propagation (T-LP), T-LP with the label-smoothing technique that applies only to abnormal category, and T-LP with the label-smoothing technique that applies to all categories.

We demonstrated that our unique graph-based approach to measure and detect label noise in large datasets is quite effective, and this approach has a huge potential to pioneer any expert system research including a human annotation support system [16]. In addition, these pioneering works are applicable to several industries such as quality control in manufacturing and health monitoring in agriculture, not limited to the diagnostics in the medical field as a part of a machine learning process for images.

APPENDIX

Our Python code of label entropy and label smoothing for transformed label propagation is publicly available from 'https://github.com/iida0yasuhiro/Experiment'

Acknowledgment

This work was achieved with the multifaceted advice from several domain experts such as Tasuku Mariya of Sapporo Medical University, and Yuta Nambu of NTT Human Informatics Laboratory.

REFERENCES

1. B. S. Deo, M. Pal, Prasanta K. Panigarhi, and A. Pradhan, "CerviFormer: A Pap-smear based cervical cancer classification method using cross attention and latent transformer," International Journal of Imaging Systems and Technology 34(2), February 2024.
2. Y. Iida, B. Mrazovac, and Y. Ishigure, "The label noise detection with unsupervised graph-based reliability estimation in Pap smear image annotation," IEEE Engineering in Medicine and Biology Society (EMBC), July 2025, To appear (Unpublished)
3. Y. Iida, B. Mrazovac, and Y. Ishigure, "Unsupervised Learning with Label Commonization for Uncovering Low-Quality Labels," IEEE 2025 International Conference on Advanced Machine Learning and Data Science, July 2025, To appear (Unpublished)
4. Blondel, V., Guillaume, J., Lambiotte, R. and Lefebvre, E.: Fast unfolding of communities in large networks, J. Statistical Mechanics: Theory and Experiment, Vol. 2008, P1000, 2008.
5. Aggarwal, C. C.: Social Network Data Analytics, 1st edition, Springer Publishing Company, Incorporated, 2011.
6. Shi, J. and Malik, J.: Normalized cuts and image segmentation, IEEE Trans. Pattern Anal. Mach. Intell., Vol. 22, No. 8, pp. 888-905, 2000.
7. D. Eppstein, M. Löffler, and D. Strash, "Listing All Maximal Cliques in Large Sparse Real-World Graphs," ACM Journal of Experimental Algorithmics, Vol. 18, 28 November 2013.
8. A. L. Barabási and R. Albert, "Emergence of scaling in random networks," Science 286, pp. 509-512, 1999.
9. J. Jantzen, J. Norup, G. Dounias, and B. Bjerregaard, "Pap-smear Benchmark Data for Pattern Classification," Nature inspired Smart Information Systems (NiSIS 2005), pp. 1–9, 2005.
10. M. E. Plissiti, P. Dimitrakopoulos, G. Sfikas, C. Nikou, O. Krikoni, and A. Charchanti, "SIPAKMED: A New Dataset for Feature and Image Based Classification of Normal and Pathological Cervical Cells in Pap Smear Images," 2018 IEEE Int. Conf. Image Processing.
11. I. Pacal, "MaxCerVixT: A novel lightweight vision transformer-based Approach for precise cervical cancer detection," Knowledge-Based Systems, vol. 289, 8 April 2024.
12. A. Halder et al., "Implementing vision transformer for classifying 2D biomedical images," Nature, Scientific Reports vol. 14, num 12567, 31 May 2024.
13. A. Dosovitskiy, et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," ICLR 2021.
14. M. Skobel, M. Kowal, and J. Korbicz, "Breast Cancer Computer-Aided Diagnosis System Using k-NN Algorithm Based on Hausdorff Distance," Proceedings of the 21st Polish Conference on Biocybernetics and Biomedical Engineering, pp. 179-188, 23 August 2019.
15. S. Roweis and L. Saul, "Nonlinear Dimensionality Reduction by Locally Linear Embedding," Science, vol. 290, no. 5500, pp. 2323–2326, 2000.
16. Y. Iida, B. Mrazovac, and Y. Ishigure, "Graph-based Similarity Search for Robust Annotation," IEEE 2025 International Conference on Advanced Machine Learning and Data Science, July 2025, To appear (Unpublished)