Multi-view Learning and Deep Learning for Microscopic Neuroblastoma Pathology Image Diagnosis

Yuhan Liu[1], Minzhi Yin[2], and Shiliang Sun[1]

[1] Department of Computer Science and Technology, East China Normal University, Shanghai, China

[2] Department of Pathology, Shanghai Children's Medical Center, Shanghai Jiao Tong University School of Medicine, Shanghai, China

slsun@cs.ecnu.edu.cn

Abstract. Automated pathology image diagnosis is one of the most crucial research in the computer-aided medical field, and many studies on the recognition of various cancers are currently actively conducted. However, neuroblastoma, the most common extracranial solid tumor of childhood, has not got enough attention in the computer-aided diagnosis research. Accurate diagnosis of this cancer requires professional pathologists with sufficient experience, which makes lack of experts lead to misdiagnosis. In this paper, we apply multi-view and single-view maximum entropy discrimination, with traditional image representations and deep neural network representations respectively. The diagnosis is performed in three neuroblastoma subtypes, undifferentiated subtype (UD), poorly differentiated subtype (PD), differentiating subtype (D), and the normal type unneoplasm tissues (UN). The best classification performance (94.25%), which far exceeds the diagnosis accuracy (56.5%) of a senior resident in the corresponding field, demonstrates the potential of neural network representations in analyzing microscopic pathology images of neuroblastoma tumors.

Keywords: Compter-aided diagnosis · Pathology image · Multi-view learning · Deep learning · Maximum entropy discrimination.

1 Introduction

Peripheral neuroblastic tumors (pNTs) are a group of embryonal tumors arising from primitive sympathetic ganglia and containing variably differentiated neural elements and variable amounts of Schwannian stromal cells. They commonly affect children and are the most frequently extracranial solid tumors in childhood. There is one case in every 7000 live children, accounting for $8 \sim 10\%$ of all childhood cancers [23]. The treatment and management of pNTs depend on the pathological diagnosis as well as the disease stage and some molecular information. The accurate pathological diagnosis plays the most important role in the whole treatment plan of patients. At this stage, the international neuroblastoma pathology classification (INPC) is recommended by the World Health Organization (WHO) for pNTs morphological categorization [18]. Based on morphologic criteria, mainly the differentiation of the neural elements and the variable amounts of Schwannian stromal cells, pNTs are classified into four basic

morphological categories (shown in Fig. 1): 1) neuroblastoma (NB); 2) ganglioneuroblastoma, intermixed (GNBi); 3) ganglioneuroma (GN); and 4) ganglioneuroblastoma, nodular (GNBn).

The tumor usually varies from one microscopic field to another in the same tumor or from one tumor to another. Due to lack of pathologists especially pediatric pathologists, and the rise of patients in China, pathologists nowadays have to go over a large number of slides every day. Thus, it is not surprising that even the same pathologist would give different diagnoses for the same case, just because the slides were too much to examine. For each slide, the pathologists pick up several fields to observe, making the final decision about the entire slide based on these sampled regions. It was reported that there is a 20% discrepancy between central and institutional reviewers [24] because of experience, subjectivity and so on. Hence, it is necessary to find some ways to help the pathologist alleviate this time-consuming, experience-needed and non-objective work.

In all the cancer categories in pNTs, NB is the most common tumor type, composed of neoplastic neuroblasts in various differentiations with no or limited Schwannian stromal cells. By definition, the proportion of tumor tissues with Schwannian stromal cells should not exceed 50%. According to the INPC, NB can be further classified into three subtypes: undifferentiated (UD), poorly differentiated (PD), and differentiating (D). In this paper, we focus on the above three subtypes of the NB category as well as the un-neoplasm type (UN) which indicates the areas with no tumor cells, such as fibrous adipose tissue, blood vessels.

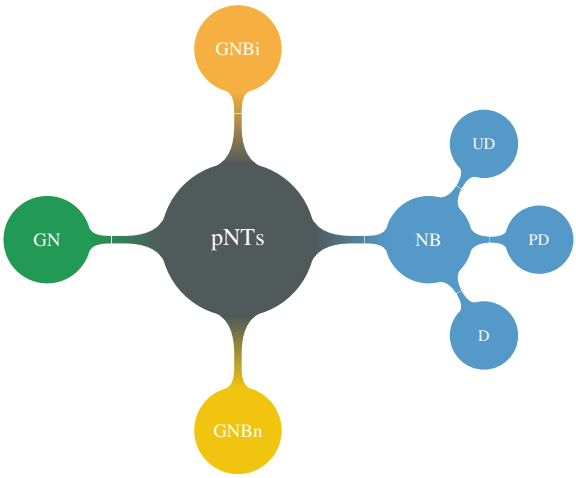


Fig. 1. The morphological categorization of pNTs

Different from regular natural images, microscopic tissue images have many unique characteristics. In this research, finding targeted kinds of image representations for microscopic pathology classification is one of the essential problems that we need to explore. Many traditional features have their distinctive superiorities and strong ro-

bustness, e.g., LBP [15] has been proved to be an efficient texture operator, aimed at describing the textural characteristics of target images. Pathology images, in a sense, show texture-like structures and are intuitively more suitable for texture representations, whereas other possible proper representations should also be considered to avoid the excessive deficiency of contributive information. Thus, different traditional features are supposed to be taken into consideration.

The uncertainty of proper representation and the limitation of information that one kind of feature contains both obstruct the further improvement of pathology image diagnosis. If several aspects of information can be utilized to diagnose at the same time and complement each other, which is called multi-view learning, then the above problems can be naturally solved. Multi-view learning, in recent years, is a rapidly developing research direction, having great theoretical basis and enormous practical success [26] [20] [30]. We consider performing multi-view maximum entropy discrimination (MVMED) [21] to combine the contributive information of different representations, which can be various features of the same pathology image. Consequently, we can utilize a kind of typical feature that only emphasizes texture information to adapt the characteristics of tissue images and combine the texture feature with other types of features to make up for the limitation of information that it can contribute.

Despite traditional representations, deep neural network representations are also worthy of experimenting with. In recent years, many kinds of CNN networks, e.g., AlexNet [11], OverFeat [17], achieve excellent accuracy in natural image classification tasks, with large-scale datasets published in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) Competition. These trained networks can well extract semantic level features of natural images and show outstanding performance in many other kinds of natural image classification tasks [11] [19] [22]. Although the pattern of pathology images is far different from natural ones, the basic edge and figure information, which the deep networks have already been trained to gain, can still be reused. Also, the complexity and flexibility of CNN could provide the semantic information extraction of microscopic pathology image. Until now, there is no work about the application of deep learning in pNTs, and it was recently illustrated that it could be a future direction for research [4].

We highlight the main contributions as follows. 1) We are the first to successfully apply multi-view learning and deep learning technology on the microscopic neuroblastoma image diagnosis task, fully proving the effectiveness of the multi-view learning and deep learning algorithm. 2) We observe that deep neural network representations exceed traditional image representations, which illustrates the excellent potential of the application of deep neural network in computer-aided pNTs diagnosis. 3) From the experimental results, we achieve much better diagnosis performance comparing with a human senior resident, with both higher efficiency and accuracy.

2 Related Work

Pathological diagnosis is usually achieved by a professional human pathologist through observing a stained specimen on the slide glass with a microscope. Recently, thanks to the rapid progress in digitization technology of information, digital pathological im-

ages, which are called whole slide image (WSI) [16], can be accumulated by utilizing a scanner to capture the entire slide of the specimen. However, the tremendous amount of information that one WSI contained (more than a billion pixels) and the high morphology variability of one same disease both make it difficult to remember and diagnose with the human brain. Digital image analysis based on machine learning algorithms can assist pathologists to diagnose with lower misjudgment, regarding the detection efficiency and accuracy.

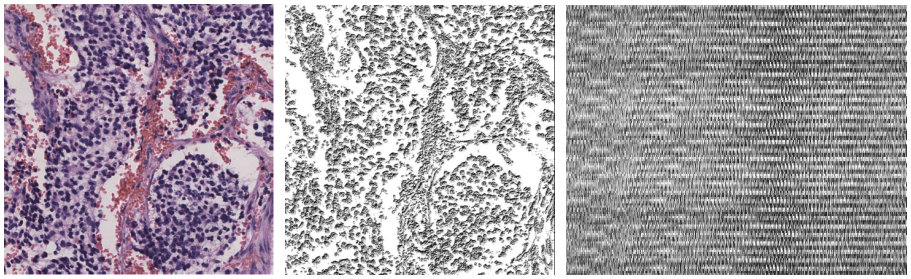
In Aug. 2016, research from Stanford University School of Medicine illustrated that computers could be trained to assess lung cancer pathological specimens with higher accuracy compared with pathologists [28]. In 2017, scientists from Google, Google Brain and Verily developed an artificial intelligence technique that can be used to diagnose breast cancer [12]. They segmented a single WSI into tens of thousands of 128×128-pixel patches for training. Ultimately, this algorithm learned to identify pixels within a single small patch that is labeled as a "tumor," effectively distinguishing tumor tissues from healthy tissues. Then in comparison with human pathologists, based on the sensitivity (how many correct tumors were found) and the false positive (how many normal tissues were diagnosed as tumors), the accuracy rate for human pathologists was 73.3% and the algorithm is 88.5%. In addition, artificial intelligence has been successfully applied in intestinal polyposis, prostate cancer, and lymph node metastasis [8] [25] [10]. There are also some papers predicting that artificial intelligence will be embedded in routine pathological workflows in the future, helping with repetitive tasks that require quantitative evaluation and counting, and reducing the time spent by pathologists for diagnosis. In this case, humans can undertake higher-level diagnosis and consultation tasks, e.g., integrating information on molecular changes, pathological diagnosis, and clinical manifestations, assisting clinical treatment planning and providing individualized health management for patients [5]. Relevant work has been carried out mainly in the diagnosis of adult cancer, such as the diagnosis of breast cancer, cervical cancer screening. However, no reports have been reported on children's tumors, and it is urgent to carry out relative researches.

3 Traditional Representations and Deep Representations

As mentioned above, the pattern of histopathology image is visually closer to texture rather than objects, and thus we consider choosing a classical kind of efficient texture operator: local binary pattern (LBP) [15]. LBP labels the pixels of an image by thresholding the neighborhood of each pixel and considers the result as a binary number, first proposed by Ojala et al. in 1994 [14]. It describes the local texture construction of an image, having many remarkable strengths such as rotational and gray-scale invariance. A visualization example is shown in Fig. 2.

In spite of texture features, we also consider other traditional features. In this paper, we choose dense scale-invariant feature transform (DSIFT) [27] descriptor, which is often utilized in image matching or retrieval tasks. Average-DSIFT extracts SIFT [13] features densely, with a specific step size, from each image on a specified grid. Fig. 2 displays the visualization result. SIFT features are detected through a staged filtering approach, which guarantees scale invariance of the algorithm. It recognizes stable

extreme points in the scaled space and then extracts invariant values of position, rotation, and scale. Because image classification tasks require unified and fixed feature structures, DSIFT features are applied to get the normalized input of models. Since the dimension of the extracted feature is unnecessarily large, we pool it with the pooling size of 19×19. Two different mainstream ways are applied: one way is max-pooling, that is, keeping only the largest value (or the smallest) of the pool, and discarding the others; the other way is average-pooling, that is, averaging all the values in the pool. Max-pooling, in general, retains more texture information, while average-pooling retains more background information [2].



(a) *original* (b) *LBP* (c) *DSIFT*

Fig. 2. An example for LBP and DSIFT visualization

For deep neural network presentation, we choose the AlexNet architecture [11], which has won the ILSVRC12. This architecture consists of 5 convolutional layers, some of which are followed by the max-pooling layers, and three fully-connected layers. Each convolution layer contains 96 to 384 convolution kernels, each with a size range of 3×3 to 11×11. It was trained on the subset of the ImageNet dataset and achieved the best accuracy. Inside the architecture, the original images are center cropped and resized to 256×256. After these simple image preprocessing operations, they are treated as input data for network training. Data augmentation and dropout [6] are also utilized during the training process to avoid overfitting phenomenon. At the same time, in order to further improve the classification accuracy, local response normalization (LRN)[9] is also applied. We extract the representation output of the 7th layer in the network, which is the topmost feature extraction layer beneath the final classification layer inside the model and obtain features of 4096 dimensions. Also, we discard the ReLU part of the 7th layer to avoid the data sparsity caused by it. Since there is only one deep representation for MVMED, the classifier will be the degraded version MED instead.

4 Multi-view Learning with Maximum Entropy Discrimination

Multi-view learning focuses on machine learning problems that can be characterized by a variety of individual feature sets. Instead of using a single representation or sim-

ply concatenating multiple representations into one representation, multi-view learning explicitly uses different representations of the data and models the relationships between them or subsequent operations that they induce. The canonical correlation analysis (CCA) [7] and the co-training algorithm [1] are two representative work for the early study of multi-view learning. This type of learning mechanism comes from the fact that real-world objects can often be represented by many different views or features in various scenarios. For example, when understanding multimedia content, multimedia clips can be described with both image and audio. In the task of pathological diagnosis of cell or tissue images, the possible influence factors on the classification result are various, e.g., texture information may be an important basis, but at the same time, the color information that the doctor diagnoses as a part of the foundation cannot be ignored. The information contained in one set of features is often targeted and not comprehensive enough. In this problem, we need to obtain better accuracy by combining the advantages of different features and multi-view learning provides a natural solution for this purpose.

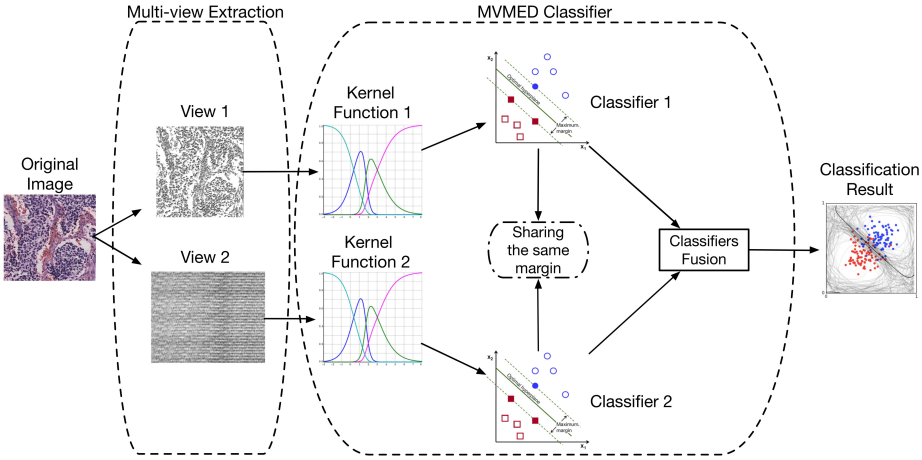


Fig. 3. The main architecture of MVMED

We consider using the MVMED algorithm [21] to combine multiple sets of features extracted from the original image and the main architecture of the model is shown in Fig. 3. In this algorithm, we have a multi-view dataset $\mathcal{D} = \{x_t^1, x_t^2, y_t\}_{t=1}^N$ with N examples, each of which has two views x_t^1 and x_t^2. For simpler expression, we let $v \in \{1, 2\}$ denote the index of views. For one pair $\{x_t^v, y_t\}$, x_t^v represents the tth input from view v, and $y_t \in \{1, -1\}$ denotes the corresponding label. Each view corresponds to a discriminant function $L_v(x^v|\mathbf{\Theta}_v)$ parameterized by a set of parameters $\mathbf{\Theta}_v$, satisfying $y_t L_v(x_t^v|\mathbf{\Theta}_v) \geq \gamma_t^v, t \in [N], v \in [V]$. The discriminant functions from different views are regularized so that they classify the training samples with the same confidence. Margin consistency theory [20] is applied in this model, i.e., classification margins from

different views are forced to be the same, namely, $\gamma_t \ddot{=} \gamma_t^1 = \ldots = \gamma_t^V$, $t \in [N]$. Considering the case that specified margins may be not reasonable for some examples, the margins $\boldsymbol{\Theta} = \{\boldsymbol{\Theta}_v\}_{v\in[V]}$ are as random variables, assigned with prior $p_0(\boldsymbol{\gamma})$ as in the MED framework. The following optimization is solved for the posterior $p(\boldsymbol{\Theta}, \boldsymbol{\gamma})$.

$$\begin{aligned} &\min_{p(\boldsymbol{\Theta},\boldsymbol{\gamma})} \mathrm{KL}(p(\boldsymbol{\Theta},\boldsymbol{\gamma}) \,||\, p_0(\boldsymbol{\Theta},\boldsymbol{\gamma})) \\ &\text{s.t.} \begin{cases} \int p(\boldsymbol{\Theta},\boldsymbol{\gamma})[y_t L_v(x_t^v|\boldsymbol{\Theta}_v) - \gamma_t] d\boldsymbol{\Theta}_v d\boldsymbol{\gamma} \geq 0, \\ t \in [N], v \in [V]. \end{cases} \end{aligned} \tag{1}$$

Afterwards, $p(\boldsymbol{\Theta}) = \int p(\boldsymbol{\Theta}, \boldsymbol{\gamma}) d\boldsymbol{\gamma}$ can be recovered, and the following formula can be treated as decision rules

$$\hat{y} = \mathrm{sign}\left(\frac{1}{V}\sum_{v=1}^{V} \int p(\boldsymbol{\Theta})\, L_v(x^v|\boldsymbol{\Theta}_v)\, d\boldsymbol{\Theta}\right).$$

Also, the idea of sequential minimum optimization (SMO) [29] is adapted to the dual problem for efficient training, which is inspired by the original SMO for the standard SVM dual problem. This coordinate ascent algorithm decomposes the original quadratic programming problem for solving N parameters into amounts of subquadratic programming problems, each of which requires only two parameters to be addressed, saving time cost and memory requirements. It is guaranteed to converge to the global maximum because the optimization problem (1) is strictly convex and smooth. Especially, if there is only one view, MVMED will degenerate to MED.

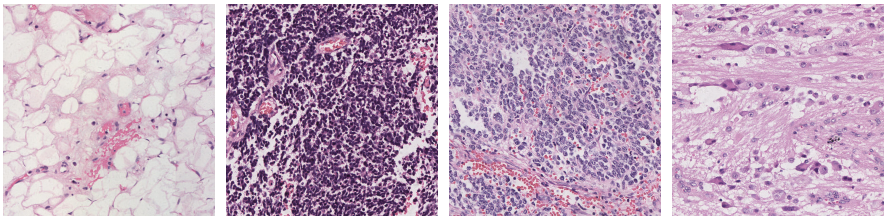
5 Experiments

In this section, we will first introduce the detailed information about our pathology dataset and the specific partition configuration for training, validation, and testing. Then, we observe and compare the experimental results of deep neural representations and traditional representations through performing binary classification tasks, getting the characteristics of two kinds of features respectively. Finally, we use the one-vs-rest method to get the multi-class classification results and compare the results with human doctor diagnosis.

5.1 Dataset Description

Between Jan. 1st, 2014 and Dec. 31st, 2015, 163 slides with a diagnosis of Peripheral neuroblastic tumors from 72 patients were seen at a third-level grade-A children's hospital, included ganglioneuroma (34 slides); ganglioneuroblastoma, intermixed (35 slides); and neuroblastoma (94 slides). Among the slides in the neuroblastoma category, there were 11 UD subtypes, 52 PD subtypes, and 31 D subtypes. These specimens were prepared according to the standard histopathologic protocol and subsequently digitized using ScanScope T2 digitizer (Aperio, Leica, Germany) at $40\times$ magnification. The whole slide images are quite large, varying from 1 to 4.5 GB. To make the image analysis more tractable, we snapshot each histology slide image into multiple non-overlapping image tiles of the size 768×768 in pixels. All the snapshot images were

reviewed histologically by a senior pediatric pathologist following the histological criteria according to INPC and classified into UD, PD, D, and UN. We name this dataset after neuroblastoma tumors (NBT).



(a) *UN* (b) *UD* (c) *PD* (d) *D*

Fig. 4. Example for UN and three subtypes of NB.

As briefly mentioned in Section 2, one WSI contains enormous amounts of information and could consist of as many as tens of billions of pixels, which usually make it hard to analyze. Thus, sampling local mini patches is a useful method so that feature extraction and classification can be performed in each local patch. In this experiment, we choose to sample smaller patches from WSIs with size 768×768, since images of this size can contain just enough information for doctors to judge it and also comfortable for human vision.

Furthermore, since the real patient cases that can be collected are limited, and labeling requires a lot of labor, available training data that we can get are insufficient. In the four classes, there are only 100 patches in the class UN and around 300 patches in the other three classes, which are far not enough for general digital image analysis tasks. Therefore, data augmentation is necessary for us to perform. To achieve the data balance between four classes, we sample patches with size 740×740 from original patches to generate new samples to make the data amount of every class be 300 patches, and these samples can be considered independent ones. Afterwards, for both traditional and deep representations, we resize all the images to size 128×128, which is big enough for our model to see the details of one patch. Then, we rotate each patch by 90 degrees for four times, flip the original patch horizontally and rotate again. This makes the image number in all classes increase from 300 to 2400. Also, we utilize the one-vs-rest method in the experiment, the training data of the class with the smaller number are copied by three times to avoid the unbalanced problem. The ratio of the partition in our dataset is $Training : Validation : Testing = 4 : 1 : 1$, i.e., the patch number of training data is 1600, and either of the other two is 400 for each class. To compare with the human doctor, 50 patches are sampled from the test sets of four classes, composing 200 test patches in total. All the labels of the human test patches are hidden and utilized to be compared with the diagnosis results of the human doctor.

5.2 Binary Classification

To analyze the classification difficulty of each class pair, we list the random combinations of two classes and treat each of them as one binary classification task respectively. Since we have four classes (UN, UD, PD, and D) in our dataset, there are $6 = 4 \times 3/div2$ binary pairs, which are UN-UD, UN-PD, UN-D, UD-PD, UD-D, and PD-D. As mentioned in Section 3, we utilize LBP, average-DSIFT and max-DSIFT representations as the traditional feature sets and AlexNet representation as the deep neural network feature. In the experiment, we obtain four results of each binary classification pair for traditional representations, which are gained from single LBP, single max-DSIFT or average-DSIFT, the concatenation of the two feature sets, and the combination of them with MVMED method separately. The experiment results are shown in Table 1.

Table 1. Results (%) of MVMED with traditional representations on NBT dataset.

Dataset	aver-DSIFT LBP	aver-DSIFT	LBP	aver-DSIFT LBP concat.	max-DSIFT LBP	max-DSIFT	max-DSIFT LBP concat.
UN - UD	97.15±0.45	86.84±0.95	96.74±0.49	88.28±0.80	**97.35±0.53**	84.99±0.54	84.48±0.89
UN - PD	**97.03±0.67**	84.13±0.98	96.52±0.57	89.26±0.97	96.81±0.78	84.15±1.30	87.27±1.29
UN - D	**93.62±0.80**	84.58±1.16	89.75±1.25	86.27±0.58	93.52±0.62	84.06±0.90	86.04±1.09
UD - PD	74.00±1.17	60.61±1.46	**76.71±0.94**	68.21±1.16	75.03±0.98	60.80±1.04	64.99±1.40
UD - D	90.36±0.96	61.15±1.91	**93.94±0.77**	74.45±1.06	90.66±1.10	59.76±1.75	67.75±1.57
PD - D	83.36±1.19	60.90±1.06	**87.90±1.16**	69.68±1.13	83.54±1.81	59.48±1.45	62.04±1.46

From Table 1, where aver-DSIFT means average-DSIFT and concat. means concatenation for simplicity, we can clearly see that utilizing LBP feature achieves much better performance than DSIFT features for every binary classification pair, which may potentially confirm that pathology images tend to contain the texture-like structures. Furthermore, it is not necessarily the case that one of max-DSIFT or average-DSIFT is better than the other, which means there is no apparent difference between the feature extraction effectiveness of the two. Also, if the two views can both well capture meaningful information of the image and achieve nice accuracy, the combination of them will perform better than either of them; if not so, the performance of the multi-view will be relatively unstable. In addition, the performance of the naive concatenation of two views cannot exceed either of them and is always between the two. At the same time, no matter the single representation is LBP, average-DSIFT or max-DSIFT, when it comes to the classification of UD-PD, the accuracy declines considerably. Thus, it can be illustrated that the classification between UD and PD is tougher than any other tasks, which also confirms the real diagnosis situations (the boundary between the two differentiating types is relatively blurred and usually influenced by the subjectivity of individual pathologists).

Table 2. Results (%) of MVMED with deep neural network representations on NBT dataset.

View	UN - UD	UN - PD	UN - D	UD - PD	UD - D	PD - D
AlexNet	99.80±0.15	99.80±0.15	98.93±0.29	92.43±0.74	99.81±0.14	97.76±0.35

Table 2 shows the classification accuracy of the deep representation in binary classification tasks. High accuracies in all case prove this neural network representation can efficiently extract useful features from pathology images. Although in the UD-PD classification case, the accuracy is still inferior a little comparing with other cases, it remains above 90%.

5.3 Multi-class Classification

Binary classification results can only convey us the concept the difficulty between random pairs from the four classes and our final goal is to classify one pathology patch into UN, UD, PD or D. We choose a typical multi-class classification method one-versus-rest (OvR) [3] to make the prediction. An OvR strategy requires the training of a single classifier for each category, samples of which are treated positive and all other samples are negative. This strategy requires the underlying classifier to generate real-valued confidence values for its decisions, not just class labels. Since when multiple classes are predicted to be a single sample, individual discrete class labels can cause ambiguities. The corresponding classifier is namely: $\hat{y} = \underset{k\in\{1...K\}}{\operatorname{argmax}} f_k(x)$, where $\hat{y} \in \{1, ..., K\}$ is the predicted label for the sample x and f_k is the confidence value of label k.

Table 3 shows the intermediate result of the classifier of each class. The deep representation is still superior to traditional ones and achieves brilliant performance. For traditional representations, the most cases illustrate that the multi-view learning helps the intermediate classification accuracy increase, which may further improve the final multi-class classification result when comparing with single views. Furthermore, the lowest accuracy of class PD in most cases demonstrates that this type of NB tumor is harder to recognize from other types. In contrast, the type UN is the easiest one to recognize, whose reason might be that healthy tissue is easy to distinguish from tumor tissue regarding morphology. Through the above OvR classifier, we obtain the final multi-view classification results afterwards, and the accuracies are shown in Table 4.

Table 3. Intermediate results (%) for one-vs-rest method on NBT dataset.

View	UN - UD PD D	UD - UN PD D	PD - UN UD D	D - UN UD PD
average-DSIFT LBP	**96.40±0.40**	82.59±0.55	79.50±0.81	82.52±0.70
average-DSIFT	90.54±0.54	76.34±0.20	69.86±0.64	71.53±0.90
LBP	93.68±0.44	**85.64±0.75**	80.33±0.66	81.40±1.08
average-DSIFT LBP concat	92.17±0.41	78.65±0.46	77.06±0.91	76.36±0.57
max-DSIFT LBP	96.24±0.40	82.43±0.81	**80.49±0.77**	**82.77±0.53**
max-DSIFT	90.89±0.77	75.69±0.25	70.29±0.85	81.29±1.07
max-DSIFT LBP concat	92.09±0.50	77.10±0.39	74.04±0.62	73.35±0.99
AlexNet	**99.70±0.25**	**96.13±0.41**	**93.88±0.56**	**98.41±0.35**

Table 4. Results (%) of multi-class classification via one-vs-rest method

View	accuracy
AlexNet	**94.25±0.75**
average-DSIFT LBP	72.59±1.02
average-DSIFT	50.73±0.91
LBP	72.52±1.27
average-DSIFT LBP concat	60.97±1.14
max-DSIFT LBP	**72.83±1.37**
max-DSIFT	49.81±0.85
max-DSIFT LBP concat	55.92±0.55

From the final accuracies in Table 4, deep feature AlexNet achieves the best performance overall, exceeding the best result in traditional features by up to around 22%. We invite a doctor, who is a senior resident with eight years of experience from a general hospital, to try to diagnose the human test dataset. After four months reviewing all 163 slides with his supervisor via multi-microscope, he had been familiar with INPC and finished the image diagnosis himself. Four and a half hours were spent in total on human diagnosing and the time spent by the algorithm can almost be ignored, which confirms the efficiency of the model. The accuracy of AlexNet feature is far higher than the diagnosis results (56.5%) of the doctor (96.5% for the UN-UD, PD, D case). Since deep features show more excellent performance and robustness than traditional ones in this experiment, it is reasonable to believe that they have more potentials worthy of exploring. Also, they may be able to achieve higher accuracy when combined with other kinds deep features or traditional features through the end-to-end training, which will be a future work for us to do. For traditional representations, the multi-view learning combination of LBP and max-DSIFT features make the accuracy higher than either of the two of them, which demonstrates the effectiveness of MVMED in terms of utilizing the complementary information of different views.

6 Conclusion

Our work yields state-of-the-art sensitivity on the challenging task of diagnosing neuroblastoma tumors through digitally classifying pathology images with the multi-view learning method. We make explorations in the multi-view learning and single-view learning with traditional representations and deep representations on the NBT dataset. The binary classification results with both kinds of features help us analyze the task difficulty of random two classes and some primary characteristics of MVMED. Then, the intermediate and final results of OvR multi-class classification method demonstrates the excellence of multi-view learning and deep learning, especially when compared with the professional human diagnosis.

The number of the authoritative professional pathologists is nowadays far below the social requirement, whereas machine learning methods can “learn” the abundant medical human experience through the sample data labeled by these exports, achieving far higher accuracy than human doctors. This kinds of computer-aided diagnosis methods

can possibly assist young pathologists who lack years of work experience. The potential shown by deep neural network representations provides our future work more possibilities, e.g., the end-to-end multi-view deep learning may the direction that we will focus on. Also, we will also attempt to apply pre-training through other hand-annotated tumor datasets with similar image conditions and further extend the size of our pathology dataset.

Acknowledgements

This work is supported by the National Natural Science Foundation of China under Project 61673179. The corresponding author is Prof. Shiliang Sun.

References

1. Blum, A., Mitchell, T.: Combining labeled and unlabeled data with co-training. In: Proceedings of the Eleventh Annual Conference on Computational Learning Theory. pp. 92–100. ACM (1998)
2. Boureau, Y.L., Bach, F., LeCun, Y., Ponce, J.: Learning mid-level features for recognition. In: Proceedings of Computer Vision and Pattern Recognition Conference. pp. 2559–2566. IEEE (2010)
3. Christopher, M.B.: Pattern Recongnition and Machine Learning. Springer-Verlag, New York (2016)
4. Gheisari, S., Charlton, A., Catchpoole, D.R., Kennedy, P.J.: Computers can classify neuroblastic tumours from histopathological images using machine learning. Pathology **49**, S72–S73 (2017)
5. Granter, S.R., Beck, A.H., Papke Jr, D.J.: Alphago, deep learning, and the future of the human microscopist. Pathology & Laboratory Medicine **141**(5), 619–621 (2017)
6. Hinton, G.E., Srivastava, N., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.R.: Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint, arXiv:1207.0580 (2012)
7. Hotelling, H.: Relations between two sets of variates. Biometrika **28**(3/4), 321–377 (1936)
8. Hou, L., Samaras, D., Kurc, T.M., Gao, Y., Davis, J.E., Saltz, J.H.: Patch-based convolutional neural network for whole slide tissue image classification. In: Proceedings of the Conference on Computer Vision and Pattern Recognition. pp. 2424–2433. IEEE (2016)
9. Jarrett, K., Kavukcuoglu, K., LeCun, Y., et al.: What is the best multi-stage architecture for object recognition? In: Proceedings of the 12th International Computer Vision Conference. pp. 2146–2153. IEEE (2009)
10. Korbar, B., Olofson, A.M., Miraflor, A.P., Nicka, C.M., Suriawinata, M.A., Torresani, L., Suriawinata, A.A., Hassanpour, S.: Deep learning for classification of colorectal polyps on whole-slide images. Journal of Pathology Informatics **8** (2017)
11. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Proceedings of Advances in Neural Information Processing Systems. pp. 1097–1105 (2012)
12. Liu, Y., Gadepalli, K., Norouzi, M., Dahl, G.E., Kohlberger, T., Boyko, A., Venugopalan, S., Timofeev, A., Nelson, P.Q., Corrado, G.S., et al.: Detecting cancer metastases on gigapixel pathology images. arXiv preprint, arXiv:1703.02442 (2017)
13. Lowe, D.G.: Object recognition from local scale-invariant features. In: Proceedings of the Computer Vision International Conference. vol. 2, pp. 1150–1157. IEEE (1999)

14. Ojala, T., Pietikainen, M., Harwood, D.: Performance evaluation of texture measures with classification based on kullback discrimination of distributions. In: Proceedings of the 12th International Association for Pattern Recognition Conference. vol. 1, pp. 582–585. IEEE (1994)
15. Ojala, T., Pietikäinen, M., Harwood, D.: A comparative study of texture measures with classification based on featured distributions. Pattern recognition **29**(1), 51–59 (1996)
16. Pantanowitz, L.: Digital images and the future of digital pathology. Journal of Pathology Informatics **1** (2010)
17. Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., LeCun, Y.: Overfeat: Integrated recognition, localization and detection using convolutional networks. arXiv preprint, arXiv:1312.6229 (2013)
18. Shimada, H., Ambros, I.M., Dehner, L.P., Hata, J.i., Joshi, V.V., Roald, B., Stram, D.O., Gerbing, R.B., Lukens, J.N., Matthay, K.K., et al.: The international neuroblastoma pathology classification (the shimada system). Cancer **86**(2), 364–372 (1999)
19. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint, arXiv:1409.1556 (2014)
20. Sun, S.: A survey of multi-view machine learning. Neural Computing and Applications **23**(7-8), 2031–2038 (2013)
21. Sun, S., Chao, G.: Multi-view maximum entropy discrimination. In: Proceedings of International Joint Conferences on Artificial Intelligence. pp. 1706–1712 (2013)
22. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., et al.: Going deeper with convolutions. In: Proceedings of the Conference on Computer Vision and Pattern Recognition (2015)
23. Tang, J., Li, Z., Chen, j., et al.: Children's cancer diagnosis and treatment (2011)
24. Teot, L.A., Sposto, R., Khayat, A., Qualman, S., Reaman, G., Parham, D.: The problems and promise of central pathology review: development of a standardized procedure for the children's oncology group. Pediatric and Developmental Pathology **10**(3), 199–207 (2007)
25. Wang, D., Khosla, A., Gargeya, R., Irshad, H., Beck, A.H.: Deep learning for identifying metastatic breast cancer. arXiv preprint, arXiv:1606.05718 (2016)
26. Xu, C., Tao, D., Xu, C.: A survey on multi-view learning. arXiv preprint, arXiv:1304.5634 (2013)
27. Yang, J., Yu, K., Gong, Y., Huang, T.: Linear spatial pyramid matching using sparse coding for image classification. In: Proceedings of the Conference on Computer Vision and Pattern Recognition. pp. 1794–1801. IEEE (2009)
28. Yu, K.H., Zhang, C., Berry, G.J., Altman, R.B., Ré, C., Rubin, D.L., Snyder, M.: Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. Nature communications **7**, 12474 (2016)
29. Zeng, Z.Q., Yu, H.B., Xu, H.R., Xie, Y.Q., Gao, J.: Fast training support vector machines using parallel sequential minimal optimization. In: Proceedings of Intelligent System and Knowledge Engineering International Conference. vol. 1, pp. 997–1001. IEEE (2008)
30. Zhao, J., Xie, X., Xu, X., Sun, S.: Multi-view learning overview: Recent progress and new challenges. Information Fusion **38**, 43–54 (2017)