Applied Intelligence (2023) 53:7186–7200
https://doi.org/10.1007/s10489-022-03918-1

A multi-view deep learning model for pathology image diagnosis

Wenbo Dong[1] · **Shiliang Sun**[1,2] · **Minzhi Yin**[3]

Accepted: 20 June 2022 / Published online: 18 July 2022

Abstract
Pathology image diagnosis plays a critical role in cancer diagnosis and treatment. However, due to the serious lack of experienced pathologists, computer-aided pathological diagnosis has become extremely important. In addition, while machine learning technologies have been successfully and widely used in other medical fields, there is still a lack of computer intervention in the basic process of diagnosis in pathology images. This paper proposes a multi-view deep learning model for pathology image diagnosis (named MvPID), which combines image features and multi-view deep learning networks. Specifically, first, the whole slide image is segmented into different non-overlapping sub-slices. Then, we extract different image features from sub-slices as different views for multi-view learning. Subsequently, we propose to use the view-specific deep Gaussian processes to extract the unique information of different views and the view-common autoencoder (AE) network to integrate the information of different views into a common representation. The common representation is put into the downstream classifier to realize automatic pathological diagnosis. The experimental results on real pathological data show that the proposed approach is effective. The best classification performance far exceeds the diagnosis accuracy of pathologists, which proves the application potential of the proposed MvPID.

Keywords Pathology image diagnosis · Multi-view learning · Deep learning · Computer-aided diagnosis

1 Introduction

Pathological diagnosis is a process that the pathologist fixes and stains the surgically excised pathological specimens, and then performs a histological examination under the microscope, and finally makes the diagnosis according to the classification criteria [1]. The accurate pathological diagnosis plays the most important role in the whole treatment plan of patients, which is known as the "gold standard" of disease (tumor) diagnosis [2]. However, the lack of pathologists is an evident issue in most parts of the world. Pathologists need to check and deal with a large number of pathological slides every day, and urgently need a method to assist themselves to reduce time-consuming and labor-consuming diagnosis work [3, 4].

Recently, with the rapid progress in digitization technology, digital pathology has emerged as a potential new diagnostic standard, in which digital slide scanners are used to digitized glass slides into whole slide images (WSIs) [5, 6]. As scanner technology becomes more and more reliable, a large amount of available WSIs is accumulated. The emergence of pathology image diagnosis promotes computer-aided diagnosis and provides a digital workflow for pathologists [7, 8]. The results of pathology image diagnosis can be used to improve the diagnostic efficiency and accuracy of pathologists, and finally provide better diagnostic experience for patients [9]. However, the large amount of information contained in a WSI and the high morphological variability of the same disease make pathology image diagnosis challenging. For each slide image, the pathologists need to skim through the whole slide and pick up

✉ Shiliang Sun
slsun@cs.ecnu.edu.cn

Wenbo Dong
dongwb7@163.com

Minzhi Yin
yinminzhi@scmc.com.cn

[1] School of Computer Science and Technology, East China Normal University, 3663 North Zhongshan Road, Shanghai, 200062, China

[2] College of Mathematics and Computer Science, Zhejiang Normal University, 688 Yingbin Road, Jinhua, 321004, China

[3] Department of Pathology, Shanghai Jiao Tong University School of Medicine, 1678 Dongfang Road, Shanghai, 200127, China

several fields to observe the details, and make the final decision about the entire slide based on these sampled regions. If computers are used to automatically recognize pathology images for pathology diagnosis, we can greatly relieve the diagnosis pressure of pathologists [10, 11]. However, a concern when analyzing pathology images is color inconsistency that may significantly affect the image analysis. The inconsistencies could be caused by different reasons, such as the use of different chemicals for staining, variations in color concentrations, or differences in scanners from different vendors. Hence, the stain normalization is needed to deal with these variations.

Different algorithms have been suggested for stain normalization [12]. Even though they have obtained good image diagnosis results, they ignore a large amount of image information. These neglected image information is also conducive to computer automatic recognition of images for pathological diagnosis. In image processing, different image features, such as color, texture and shape, can be extracted according to the appearance of the image. Different features clearly describe the image information from different perspectives. These image features have their unique discrimination information and strong robustness. Therefore, these image features should also be considered to avoid the excessive deficiency of contributive information. Although a combination of features is used and some works on representation learning are given to learn a suitable representation in various medical tasks [13], these features are concatenated directly or weighted in the classification system, which can not make full use of the information between features, especially the discriminant information. It is a key factor to improve the classification accuracy to synthesize the information provided by various features, whereas how to use these features comprehensively is a challenging task. If different features are regarded as multiple views of pathology images, multi-view learning can synthesize the complementary information provided by different views to improve the performance of pathology image diagnosis [14, 15].

Multi-view learning is an attractive and emerging direction of machine learning, which considers learning with multiple views to improve the generalization performance [16] and makes the downstream tasks easier [17]. So far, multi-view learning has achieved encouraging results [18] and has been successfully applied in many research fields, including web page classification, natural language processing, and computer vision [19]. Although recently several approaches have sprung up, how to respect the consistency and complementarity simultaneously in multi-view data is still challenging. Meanwhile, many approaches can only deal with two view data, while the number of views frequently exceeds two in real applications. Finally, multi-view learning is rarely used in medical diagnosis, especially pathology image diagnosis. Multi-view medical image fusion brings many benefits to clinical diagnosis and analysis because it creates favorable conditions for diagnostic imaging practitioners to make a more accurate diagnosis [20].

To address the aforementioned problems, we propose a multi-view deep learning model for pathology image diagnosis (named MvPID) to help pathologists reduce time-consuming and laborious diagnostic tasks and improve work efficiency. In proposed approach, the whole slide image is firstly segmented into different non-overlapping sub-slices. Then, we extract different image features from sub-slices as different views for multi-view learning. We propose a view-specific deep Gaussian processes to extract the unique information of different views and a view-common autoencoder (AE) network to integrate the information of different views into a common representation, which flexibly balances the complementarity and consistency among multiple views. The common representation is put into the downstream classifier to realize automatic pathological diagnosis. The experimental results on real pathological data demonstrate the effectiveness of the proposed MvPID. The best classification performance far exceeds the diagnosis accuracy of pathologists, which proves the application potential of the proposed MvPID. Moreover, in some special environments where all features cannot be used simultaneously, the proposed approach can provide information to select better feature combination.

We highlight the main contributions as follows.

(1) This paper proposes a multi-view deep learning model for pathology image diagnosis, which combines the different features of pathology images for multi-view learning.
(2) The proposed approach can comprehensively use the information contained in the pathology image to avoid the excessive deficiency of contributive information caused by stain normalization.
(3) The experimental results show that the proposed approach is effective, and the best classification performance far exceeds the diagnosis accuracy of pathologists, which shows the application potential of multi-view learning in pathology image diagnosis.

The remainder of this paper is organized as follows. Section 2 introduces related work. Section 3 presents the proposed multi-view deep learning model for pathology image diagnosis. Section 4 reports the data set information, implementation details and comprehensive experiments and results. Finally, conclusions are presented in Section 5.

2 Related work

This section briefly elaborates the related work, including the artificial intelligence in pathology diagnosis and multi-view learning, which lays the foundation for the following research on pathology image diagnosis with multi-view deep learning model.

2.1 Artificial intelligence in pathology diagnosis

Recent reviews have highlighted that deep learning has been applied to a wide range of medical image analysis tasks (segmentation, classification, detection, image reconstruction, enhancement, etc.) across a wide range of anatomical sites (brain, heart, lung, breast, prostate, abdomen, musculature, etc.) [21–24]. Although computer processing have been successfully and widely used in other medical fields, there is still a lack of computer intervention in the basic process of diagnosis in pathology images. The latest development of digital pathology imaging has initiated the transition to digital pathology practice [25, 26].

Pathology image diagnosis plays a critical role in cancer diagnosis and treatment [27]. Recently, some enlightening studies have been proposed. These studies applied artificial intelligence methods to pathology diagnosis, which makes pathology diagnosis more convenient and efficient [28, 29]. Wang et al. [30] reviewed the current applications of artificial intelligence methods in lung cancer pathology image analysis, and pointed out that deep learning can potentially affect the development of pathology image analysis. Generally, pathology images are tremendously large, which requires extensive annotations by expert pathologists. Therefore, pathology datasets are usually small [31]. Data augmentation can be used to generate a large number of image tiles, which can greatly increase the size of training set, reduce over-fitting and improve generalization ability. Sirazitdinov et al. [32] investigated the necessity and contribution of different augmentation approaches in computerized lung pathology diagnosis. Wright et al. [33] presented an in-depth perspective on the effect of digital pathology image quality on downstream analysis. It is worth noting that data augmentation is not the only key factor to get good results. Several studies have shown that designing a unique task specific architecture can also achieve better results. Dietterich et al. [34] proposed a two-step method, which first uses multi-instance learning to train the classifier at the tile level, and then aggregates the prediction scores of each tile of the whole slide images to make the final diagnosis. Zhu et al. [35] proposed a weakly supervised model to extract gastric cancer pathological image features to classify the cancer region. Experiments on gastric dataset showed that their model achieved better performance in classification accuracy. In order to visual analysis of computational pathology and clinical data, Corvo et al. [36] developed a visual analysis method that enables clinical researchers to create computational pathology pipelines. Koohbanani et al. [37] presented a self-supervised convolutional neural network framework (termed Self-PATH) to extract generalizable representations in unlabeled pathology images. Experimental results showed the effectiveness of the proposed framework. At the same time, they showed that the framework can be extended to other computational pathology problems. Lu et al. [38] reported a clustering-constrained-attention multiple-instance learning approach, which uses attention-based learning to accurately classify whole slides and instance-level clustering to constrain the feature space. Experimental results demonstrated that their approach outperforms standard weakly supervised classification algorithms. Informative representations are crucial for histopathological image classification tasks. In order to learn an effective whole-slide-level representation, Li et al. [39] presented a embedded-space multi-instance learning model (named DT-MIL) based on transformer and convolutional layers.

However, the existing approaches do not consider how to use the information contained in different image features of pathology images to improve the diagnostic accuracy [40]. Different features clearly describe the image information from different perspectives. These image features should also be considered to improve diagnosis accuracy. Different from the above studies and inspired by multi-view learning, we developed a multi-view deep learning model for pathology image diagnosis, which aims to improve the efficiency and accuracy of diagnosis by combining the different features of images for multi-view learning. The proposed approach can comprehensively use the information contained in the pathology image to avoid the excessive deficiency of contributive information caused by stain normalization. The learned low dimensional representation can be put into the classifier to realize automatic pathology diagnosis, so as to help doctors reduce time-consuming and laborious diagnosis.

At the end of this subsection, we discuss the main differences between this research and the current commonly used methods based on multi-instance learning. The differences from the current commonly used methods based on multi-instance learning are as follows. Firstly, multi-instance learning and multi-view learning are two different research hotspots in machine learning. The purpose of multi-instance learning is to segment the pathology image into different small blocks, and then label the data or identify the location of cancer in the pathology image by processing several different small blocks at each time. Multi-view learning mainly considers learning with

multiple views to avoid the excessive deficiency of information, so as to improve the prediction performance. Secondly, in the pathology image classification task, the current methods based on multi-instance learning only uses one view feature of pathology image, and they are weakly supervised learning methods. The multi-view learning method proposed in this paper can synthesize the complementary information provided by different views to improve the performance of pathology image diagnosis.

2.2 Multi-view learning

Multi-view is a research hotspot in machine learning, which considers learning with multiple views to improve the prediction performance. Unlike using a single view or simply concatenating multiple views into one view, multi-view learning can model the relationship between views and integrate the information of different views to improve learning performance [41]. The existing multi-view research results have made encouraging achievements in many fields, such as web page classification, natural language processing, and computer vision [42].

At present, there are two popular types of multi-view learning algorithms that have been extensively studied. The first type is multi-view learning based on view correlation, which is mainly to find common subspace according to the correlation between views. Li et al. [43] exploited Gaussian process latent variable model to represent multiple views in a common subspace. Yin et al. [44] proposed multi-view uncorrelated locality preserving projection (MULPP), which considers pairwise correlation and high-order correlation simultaneously. Meanwhile, they preserved the local structure of all views and made the extracted features of different views uncorrelated. The second type is multi-view learning based on deep models, which mainly models the consistency and complementarity of views through deep neural networks. Yang et al. [45] developed a end-to-end deep multi-view collaborative learning model, which utilizes multiple autoencoder networks to embed multi-view data into various latent spaces and employs a heterogeneous graph module to fuse the latent representations adaptively. Similarly, Zhang et al. [46] presented an autoencoder in autoencoder network, which integrates information from multiple views into a representation by the nested autoencoder framework. Inspired by the excellent performance of deep Gaussian processes in data representation learning, Sun et al. [47] extended the deep Gaussian processes to multi-view learning scenario and proposed a multi-view deep Gaussian processes model. They explored the complementary information contained in two views to learn more comprehensive representations.

Although these methods integrate deep learning in the multi-view learning framework, new chances to develop multi-view deep learning still exist in terms of methodologies and applications. Specifically, how to respect the consistency and complementarity simultaneously in multi-view data is still challenging. Meanwhile, many approaches can only deal with two view data, while the number of views frequently exceeds two in real applications. Finally, multi-view learning is rarely used in medical diagnosis. In this paper, we propose a multi-view deep learning model for pathology image diagnosis. The model uses view-specific deep Gaussian processes to model different views to respect the complementarity principle of views. The consistency principle of views is followed by optimizing the view-common AE network. At the same time, the model allows the number of views to exceed two, which is more flexible to adapt to practical problems.

3 The proposed multi-view pathology image diagnosis model

In this section, we present the proposed MvPID model in detail. Pathology images are usually as large as gigapixels. However, pathology image datasets are usually much smaller because of patient privacy issues and the need for expert annotation. One approach to alleviate the lack of pathology images is data augmentation that not change their corresponding labels. This is a common approach in digital histopathology using deep learning. To fully leverage the scale of pathology image dataset, data enhancement based on cropping is adopted, which divides the WSI into different sub-slices.

Below we present the proposed model. Figure 1 illustrates the framework of MvPID model. Overall, first, the whole slide image is segmented into different non-overlapping sub-slices. This practice can also be potentially used as a data enhancement method. Second, we extract different features from each sub-slices image as different views for multi-view learning. Then, we propose a view-specific deep Gaussian processes to extract unique information from different views. View-common AE network is used to integrate the unique information of different views and learn a compact common representation. Finally, the common representation is put into the downstream classifier to realize automatic pathological diagnosis. The components in the model are introduced as follows.

3.1 View-specific deep Gaussian processes

Deep Gaussian process model is a deep extended version of the Gaussian process, which has many flexible properties and has shown excellent performance in artificial intelligence applications. Single Gaussian process is essentially

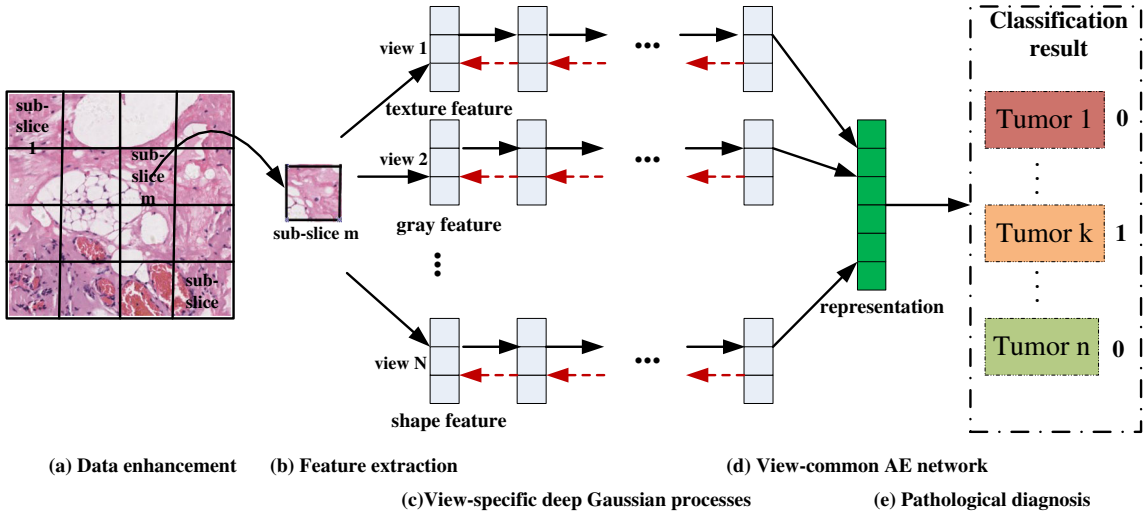

Fig. 1 Overview of the proposed multi-view deep learning model for pathology image diagnosis. We divided the pathological image into 64 × 64 non-overlapping sub-slices. For each sub-slice, we extract different image features as different views, such as gray features, texture features. The view-specific deep Gaussian process extracts the unique information of different views, where red arrow represents the optimization process of the deep Gaussian process. View-common AE network integrates information from different views into a compact common representation. The representation is put into the downstream classifier for pathological diagnosis. '1' and '0' respectively indicate whether the tumor type is diagnosed

an approximate single layer neural network with infinite neurons. Therefor, the feature expression ability of deep Gaussian processes is more powerful. DGPs provide a Bayesian non-parametric alternative to standard parametric deep learning models. Recent studies have provided convincing evidence to show that deep Gaussian processes are powerful to extract data information [48]. In this paper, we first modify the deep Gaussian processes to adapt to multi-view learning, and propose a view-specific deep Gaussian process.

The proposed view-specific deep Gaussian processes are composed of multiple layers of latent variables, and employ a hierarchical structure of Gaussian process mappings. Formally, we present the view-specific deep Gaussian processes with a set of multi-view samples $X = \{X^1, \cdots, X^N\}$, where $X^n \in R^{M \times d_n}$ is the feature matrix of the nth view, N is the number of views, M is the number of samples and d_n is the dimensionality of feature space of nth view. Given a view matrix X^n, DGPs define H layers of hidden variables $X_h^n \in R^{M \times Q_h}, h = 1, \cdots, H$ through the following generative process,

$$X^n = f_1(X_1^n) + \epsilon_1, \epsilon_1 \sim \mathcal{N}(0, \sigma_1^2 I), \tag{1}$$

where X_1^n is the latent variable of first layer, f_1 is a Gaussian process with the covariance function k_1, i.e., $f_1(X) \sim \mathcal{GP}(0, k_1(X, X'))$, and ϵ_1 is Gaussian noise.

$$X_h^n = f_{h+1}(X_{h+1}^n) + \epsilon_{h+1}, \epsilon_{h+1} \sim \mathcal{N}(0, \sigma_{h+1}^2 I), \tag{2}$$

where $h = 1, \cdots, H-1$, X_h^n is the latent variable at h layer, f_{h+1} is a Gaussian process with the covariance function k_{h+1}, i.e., $f_{h+1}(X) \sim \mathcal{GP}(0, k_{h+1}(X, X'))$, and ϵ_{h+1} is Gaussian noise.

$$X_H^n \sim \mathcal{N}(0, I), \tag{3}$$

where X_H^n is the parent node, which is imposed a prior distribution to constrain the whole latent space through propagation of the prior density.

The expression power of the model comes from the nonlinear mapping ability of the kernel function. In this paper, we use the automatic relevance determination squared exponential kernel as the covariance function, which is defined as follows,

$$k(x, x') = \sigma_f^2 \exp\left(-\frac{1}{2}\sum_{d=1}^{D} \alpha_d (x_d - x_d')^2\right), \tag{4}$$

where σ_f^2 and α_q are hyper-parameters, which can be explored in a Bayesian training framework to automatically search relevant features and weaken irrelevant features, and D represents the dimension of x.

It is worth mentioning that because each view may have a very different statistical property, MvPID employs the relatively independent deep structure to model different multi-view data, which is more reasonable and flexible.

3.2 View-common AE network

When the features of each view are extracted by the view-specific deep Gaussian processes, we use the view-common AE network to integrate different view features into a common compact representation. This part follows the consistency principle between views, and AE network

ensures that each single view could be reconstructed from the common representation [46]. For simplicity, the AE network is denoted as $f(X_H; \Theta_{ae})$, where $X_H = \{X_H^n\}_{n=1}^N$ represents the view features extracted by the view-specific deep Gaussian processes, $\Theta_{ae} = \{W_{ae}^l, b_{ae}^l\}_{l=1}^L$ is the parameter set for all layers, L is the number of layers. Specifically, the first $L/2$ hidden layers encode the input as a common representation, and the last $L/2$ layers decode the representation to reconstruct the input.

Let $X_H^{(0,n)}$ denote the input of the nth view. Then, the output of the lth layer is,

$$X_H^{(l,n)} = a(W_{ae}^{(l,n)} X_H^{(l-1,n)} + b_{ae}^{(l,n)}), l = 1, 2, \cdots, L, \quad (5)$$

where $W_{ae}^{(l,n)}$ and $b_{ae}^{(l,n)}$ denote the weights and noise associated with the lth layer, respectively. $a(\cdot)$ is a nonlinear activation function.

Then, given the view-specific feature matrix $X_H = \{X_H^1, \cdots, X_H^N\}$, the corresponding reconstruct representation is denoted as $Y_L = \{Y_L^1, Y_L^2, \cdots, Y_L^N\}$. We should minimize the following reconstruction loss,

$$\min_{\{\Theta_{ae}\}} \mathcal{L} = \frac{1}{2} \sum_{n=1}^{N} \|X_H^n - Y_L^n\|_F^2, \quad (6)$$

where Y_L^n is the reconstructed representation for the nth view.

4 Model optimization

This section gives the optimization process of the proposed model. Specifically, we divide the optimization process into the following two steps: learning the view-specific representation with deep Gaussian processes and learning view-common representation with AE network. The optimization process of each of the above steps will be described in detail below.

4.1 Learning view-specific representation with DGPs

Here, we first use the deep Gaussian processes to learn and optimize the view-specific representation. For view X^n, we assume that $\bar{Z} = \{X_h^n\}_{h=1}^H$ denotes the set of the hidden variables. According to the Bayesian method, the joint distribution of observation data X^n and hidden variables $\bar{Z}$ can be written in the following form,

$$p(X^n, \bar{Z}) = \prod_{h=1}^{H-1} [p(X_h^n \mid X_{h+1}^n)] p(X_H^n). \quad (7)$$

The central objective is the posterior distribution of hidden variables, $p(\bar{Z} \mid X^n) = p(X^n, \bar{Z}) / \int p(X^n, \bar{Z}) \mathrm{d}\bar{Z}$. Because this integral is high dimensional and the integrand is complex, we use variational inference to approximate the posterior with a simpler distribution.

According to the Bayesian training procedure, the optimization of the model evidence is required, i.e., optimization of $\log p(X^n)$.

$$\begin{aligned} \log p(X^n) &= \log \int p(X^n, \bar{Z}) \mathrm{d}\bar{Z} \\ &= \log \int \prod_{h=1}^{H-1} [p(X_h^n \mid X_{h+1}^n)] p(X_H^n) \mathrm{d}\bar{Z}. \end{aligned} \quad (8)$$

However, it can be seen from [48] and [47] that the solution of the above model evidence is not feasible. We obtain the variational lower bound of the model evidence $\mathcal{F}_\upsilon$ through Jensen's inequality, so as to transform the above problem into an optimization problem for approximate solution. The specific derivation process is given below,

$$\begin{aligned} \log p(X^n) &= \log \int \frac{p(X^n, \bar{Z}) Q(\bar{Z})}{Q(\bar{Z})} d\bar{Z} \\ &= \log \mathbb{E}_{Q(\bar{Z})} [\frac{p(X^n, \bar{Z})}{Q(\bar{Z})}] \\ &\geq \mathbb{E}_{Q(\bar{Z})} [\log \frac{p(X^n, \bar{Z})}{Q(\bar{Z})}] \\ &= \mathcal{F}_\upsilon, \end{aligned} \quad (9)$$

where $Q(\bar{Z})$ is the variational distribution. We employ the mean field approximation and assume that latent variables are independent, which simplifies derivations. Meanwhile, in order to make the problem solvable, we apply the "data augmentation" principle. we expand the joint probability space with K auxiliary pseudo inputs $\widetilde{\mathcal{X}} = \{\widetilde{X}_h\}_{h=1}^{H-1}$, where $\widetilde{X}_h \in R^{K \times Q_h}$, and correspond to a collection of function values $\mathcal{U} = \{U_h\}_{h=1}^{H-1}$ respectively. Following this approach, we obtain the augmented probability space,

$$\begin{aligned} p(X^n, \bar{Z}, \widetilde{\mathcal{X}}, \mathcal{U}) = {} & p(X^n \mid f_1) \cdots p(X_{H-1}^n \mid f_{H-1}) \\ & p(f_{H-1} \mid U_{H-1}, \widetilde{X}_{H-1}) \\ & p(U_{H-1} \mid \widetilde{X}_{H-1}) p(X_H^n). \end{aligned} \quad (10)$$

Here, we are able to define the variational distribution $Q(\bar{Z})$ which combined with the new expressions for the augmented GP priors.

$$Q(\bar{Z}) = \prod_{h=2}^{H} p(f_{h-1} \mid U_{h-1}, X_h^n) q(U_{h-1}) q(X_h^n). \quad (11)$$

We select $q(U_{h-1})$ to be free-form variational distributions, while $q(X_h^n)$ is chosen to be Gaussian,

$$q(X_h^n) = \prod_{j=1}^{Q_h} \mathcal{N}(\mu_j^h, S_j^h). \quad (12)$$

By substituting (11) back to (9) while also replacing the original joint distribution with its augmented version in (10), the problem can be computed analytically.

4.2 Learning view-common representation with AE network

To update the view-common representation, we should minimize (6). By applying the chain rule to calculate the gradient of above equation w.r.t. $W_{ae}^{(l)}$ and $b_{ae}^{(l)}$, we have $\frac{\partial \mathcal{L}_{ae}}{\partial W_{ae}^{(l)}} = \Delta^{(l)}(Y_{l-1}^{n})^{T}$, $\frac{\partial \mathcal{L}_{ae}}{\partial b_{ae}^{(l)}} = \Delta^{(l)}$, where $\Delta^{(l)}$ is defined as,

$$\Delta^{(l)} = \begin{cases} -(X_H^n - Y_l^n) \circ a'(y_n^l) & , l = L \\ (W_{ae}^{(l+1)})^T \Delta^{(l+1)} \circ a'(y_n^l) & , otherwise \end{cases} \tag{13}$$

where $a'(\cdot)$ is the derivative of the activation function $a(\cdot)$, $\circ$ denotes the element-wise multiplication, and $y_n^{(l)} = W_{ae}^{(l)} y_n^{(l-1)} + b_{ae}^{(l)}$. Then we can update the parameters $\{W_{ae}^{(l)}, b_{ae}^{(l)}\}$ with gradient descent as, $W_{ae}^{(l)} = W_{ae}^{(l)} - \mu \frac{\partial \mathcal{L}_{ae}}{\partial W_{ae}^{(l)}}$, $b_{ae}^{(l)} = b_{ae}^{(l)} - \mu \frac{\partial \mathcal{L}_{ae}}{\partial b_{ae}^{(l)}}$, where $\mu > 0$ is the learning rate which is usually set to a small positive value. In this paper, $\mu = 0.001$.

Finally, we get the bottleneck representation of the middle layer of AE network. The bottleneck representation is the common representation of multiple views, which is input to the classifier for the final pathological diagnosis. In the next section, we will evaluate the effectiveness of the proposed MvPID model from different perspectives.

5 Experiments and analysis

In this section, comprehensive experiments on real world datasets are conducted to evaluate the performance of the proposed approach. Specifically, we first present the dataset description and experimental settings, and then validate the superiority and benefit of MvPID over single-view learning and multi-view learning methods in pathological image diagnosis. Finally, we performed ablation experiments to verify the effects of different components in the model.

5.1 Dataset description and experimental settings

Dataset collection and description Two pathology image data sets are used to evaluate the effectiveness of the proposed approach.

The first pathology image dataset used in this paper is provided by Shanghai Children's Medical Center, which is introduced as follows: peripheral neuroblastic tumors (pNTs) are a group of embryonal tumors originated from primitive sympathetic ganglia, which contain variably differentiated neural elements and variable amounts of Schwannian stromal cells [49]. They commonly affect children and are the most frequently extracranial solid tumors in childhood. 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. pNTs are classified into four basic morphological categories: (1) neuroblastoma (NB); (2) ganglioneuroblastoma, intermixed (GNBi); (3) ganglioneuroma (GN); (4) ganglioneuroblastoma, nodular (GNBn). According to the international neuroblastoma pathology classification (INPC), NB can be further classified into three subtypes: undifferentiated (UD), poorly differentiated (PD), and differentiating (D). Besides, there are un-neoplasm type (UN) which indicates the areas with no NB tumor cells, such as fibrous adipose tissue, blood vessels. Ewing sarcoma (EWS) is a typical small round blue cell tumor, which is the name given to a group of highly malignant neoplasms that occur mostly in childhood. EWS presents a relatively monotonous pattern of sheets of small blue cells with round to oval hyperchromatic nuclei, modest amounts of cytoplasm, and inconspicuous cellular boundaries. Not otherwise specified (NOS) refers to the rare, but observed tumor cells due to the different reasons could not be classified. For example, with crush or extensive calcification, hemorrhage. The pathological slide images of these eight categories are shown in Fig. 2. For data protection, all publicly shared WSIs are de-identified and do not contain any protected health information. Depending on the downstream analyses that will be performed, manual annotation of the digitalized slides is required by a trained pathologist. Annotating pathology image data sets has high costs, which requires high levels of expertise from clinicians. The corresponding label is included in the dataset. For each tumor type, the pathologist divided the original pathology image into 64×64 non-overlapping sub-slices. Finally, we have a total of 43200 pathology sub-slices.

The second pathology image dataset comes from public medical data, which is called BREAST-LNM [39]. This dataset is mainly used to predict whether lymph node metastasis is based on biopsy slides of breast cancer patients. The dataset is annotated by two experienced pathologists and contains 1840 positive samples and 2117 negative samples.

Comparison methods and implementation details To validate the performance of our proposed approach in pathology image diagnosis, the following classifiers are used: supervised deep Gaussian processes (SDGPs), support vector machine (SVM), k-nearest neighbor (KNN), NaiveBayes, Adaboost, deep convolutional neural networks (DCN). DCN is composed of two convolutional layers and two fully connected layers. We have applied ReLU activation function

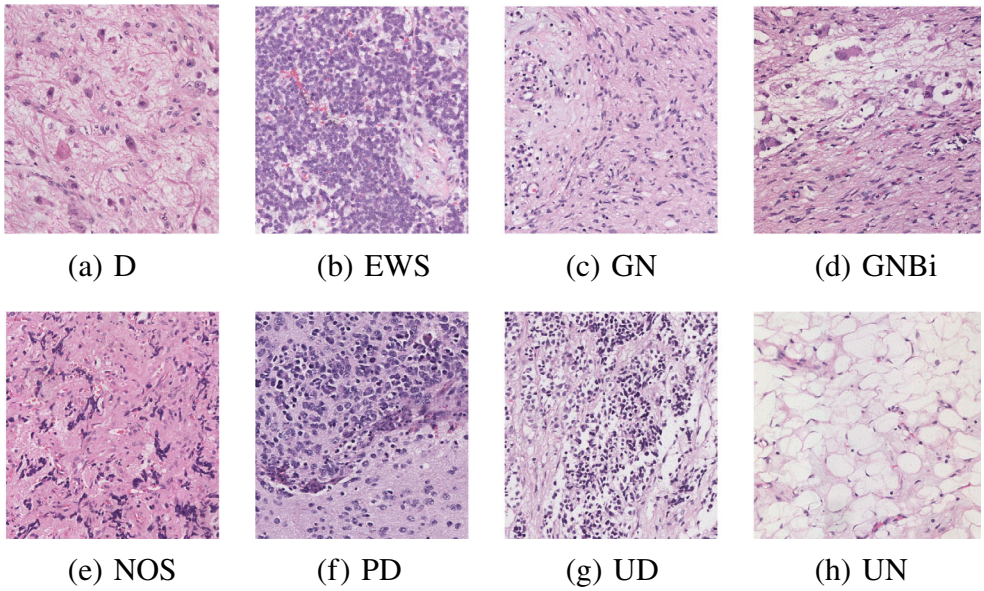

Fig. 2 Slides of different images from the categories

to all convolutional and fully connected layers. The standard gradient descend is used in training, learning rate is set to 0.01, and dropout rate is set to 50%. The number of neighbors used in KNN is 3 and we use Euclidean distance to calculate the distance.

Meanwhile, in order to compare with other algorithms, the following single-view learning methods are used: unsupervised deep Gaussian processes (UDGPs), principal component analysis (PCA), linear discriminant analysis (LDA), Isomap, t-distributed stochastic neighbor embedding (t-SNE), Gaussian process latent variable model (GPLVM), locally linear embedding (LLE). The following multi-view learning methods are used: CCA [50], KCCA [50], DCCA [51], MULPP [44], MvDGPs [47], AE^2-Nets [46]. Meanwhile, we also compared the recently published pathology image diagnosis methods, such as Self-PATH [37] and DT-MIL [39]. However, these methods are essentially single view learning methods. In the experiment, we concatenate the features of different views as the input of these methods. We use the code released by the corresponding paper of these methods.

Since the evaluation task belongs to classification problems, we employ the mean and standard deviation (std) of classification accuracy (Acc) of multiple experiments as evaluation metric. At the same time, in order to compare with other commonly used methods more comprehensively, we also add F1-score, precision and recall as evaluation metrics. These evaluation metrics are often employed in classification tasks, and their calculation formulas are widely recorded in the corresponding papers. Therefor, their calculation formulas are omitted in this paper to save more space for experiments and discussion. Because pathological diagnosis is a time-consuming and labor-consuming thing, pathologists do not have to give all the comparison results, but only give the comparison results in several typical comparison experiments.

5.2 The rationality of multi-view deep learning in pathology diagnosis

When pathologists diagnose pathological slices, they usually only use stained slices, which are usually processed as gray image in computer-aided automatic diagnosis. Therefore, a lot of image information is ignored. In order to verify the rationality of the proposed model in pathological diagnosis, we first use two views to conduct experiments. In the experiment, we mainly performed pathological diagnosis of UD and EWS. The data set is randomly divided into 60% training set and 40% test set. In addition to the original gray image, we use the MATLAB deep learning toolbox to extract texture features as the second view. The learned representation is put into the downstream Adaboost classifier to obtain the final classification results. Experiments are performed several times independently, in which the classification Acc and F1-score are used as the evaluation metrics. The experimental results are shown in Fig. 3.

As shown in Fig. 3, even if we use different classifiers to make better use of gray features, the performance of directly using gray features for pathological diagnosis is relatively poor. On the contrary, the multi-view diagnosis model proposed in this paper significantly improves the performance of pathology diagnosis. At the same time, the model also surpasses the diagnosis results of experienced

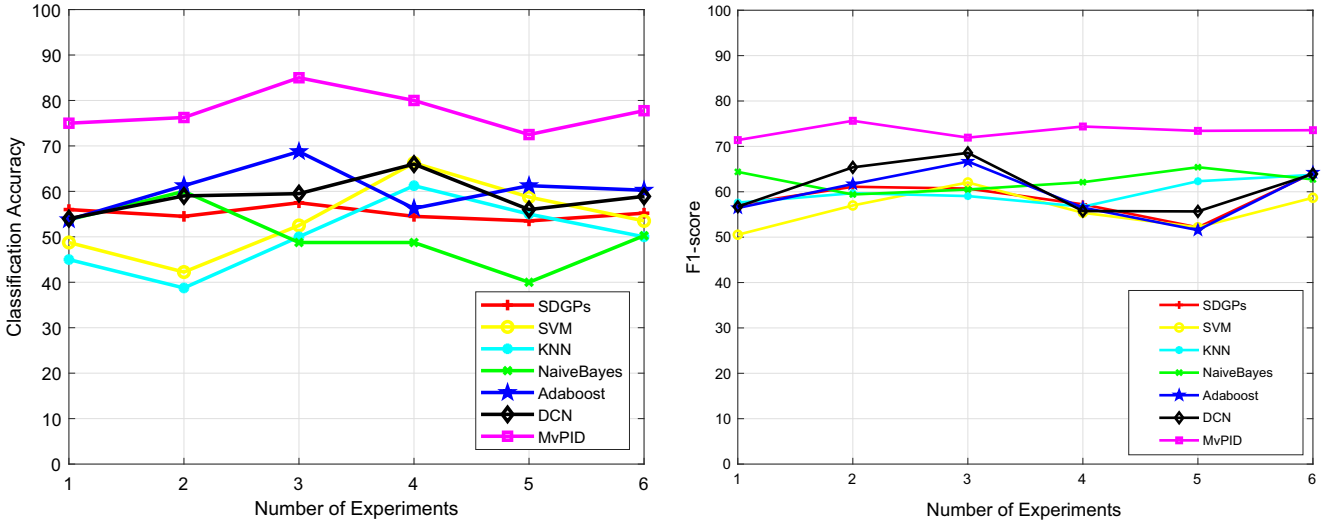

Fig. 3 The rationality of the proposed MvPID in pathology diagnosis

pathologists, whose average diagnosis result is 65%. These results prove the rationality and effectiveness of the proposed model.

5.3 The superiority of the MvPID in pathological diagnosis

For different view features, the most straightforward method is to concatenate them directly. However, this method may cause dimensional disaster, which is not conducive to downstream classification. In this experiment, we compare the proposed MvPID with single view learning methods and multi-view learning methods respectively. In addition to the original gray image, we use the MATLAB deep learning toolbox to extract texture features, HOG (histogram of oriented gradient) features, and binary features as different views, respectively. The experiment data sets are randomly divided into 80% training sets and 20% test sets, and five experiments are conducted respectively. The learned representation vector is put into Adaboost classifier to get the classification results. Finally, the mean and standard deviation (std) of five experiments with different methods are reported as evaluation criteria. CCA, KCCA, DCCA and MvDGPs can only use the information of two views at the same time. Therefore, gray features and texture features are used for experiments. Other methods use the information of all views. The experimental results are shown in Table 1.

As can be seen from Table 1, in most scenarios, the multi-view learning methods are better than the single view learning methods and the methods of directly concatenating different views. The results obtained by MvPID are generally better than those obtained by the comparative methods. Specifically, compared with the single view learning methods, MvPID can integrate the information of different views of pathological images, and make full use of these information to make up for the deviation caused by the lack of information in a single view. Compared with CCA, KCCA and MULPP, MvPID uses asymmetric deep Gaussian processes to model different views and AE network to integrate different view information into a common representation. It not only respects the different structures and statistical characteristics of different views, but also deeply excavates the complementarity between views. Compared with DCCA and AE^2-Nets, our model adopts multi-layer Gaussian process mappings, which inherits the advantages of deep model and Gaussian process. Compared with MvDGPs, the approach proposed in this paper can model more than two views at the same time. The performance of MvPID also outperforms DT-MIL model based on multi-instance learning and Self-PATH model based on deep learning based. Meanwhile, the model also surpasses the diagnosis results of experienced pathologists, which verifies the effectiveness of the proposed model.

We also conduct experiments on BREAST-LNM dataset to verify the superiority of the MvPID in pathology image diagnosis. In order to compare with the existing methods more reasonably, we conduct experiments according to the implementation details of [39]. The experimental results are shown in Table 2. Since DCN, LDA, DT-MIL and Self-PATH are essentially single view learning methods, we concatenate features of different views as the input of these methods.

As can be seen from Table 2, our method achieves the best diagnostic results in multiple evaluation metrics, which indicates that our method can synthesize features provided by multiple views to improve the diagnostic accuracy. The experimental results also show that the methods of directly concatenating the features of different views cannot

Table 1 The comparison with single-view learning and multi-view learning methods. Best performances are shown in bold

Methods	D/NOS		GN/UN		GNBi/UN		PD/UN		UD/UN	
	Mean	std	Mean	std	Mean	std	Mean	std	Mean	std
Concate	78.25	0.36	76.25	0.74	70.00	0.40	78.50	0.65	76.25	0.44
PCA	84.75	0.03	86.25	0.03	73.25	0.02	85.50	0.02	86.25	0.05
Isomap	85.75	0.03	50.25	0.05	70.50	0.03	72.75	0.02	90.00	0.03
t-SNE	75.75	0.04	72.00	0.07	71.75	0.10	85.25	0.02	89.25	0.04
LLE	80.25	0.10	79.25	0.08	81.50	0.09	73.00	0.04	79.25	0.04
GPLVM	67.50	0.11	55.75	0.19	58.75	0.11	62.25	0.13	79.75	0.03
LDA	82.25	0.03	81.50	0.04	68.50	0.02	79.50	0.04	75.00	0.06
UDGPs	85.75	0.15	86.50	0.28	78.25	0.43	83.00	0.25	85.75	0.55
CCA	87.00	0.04	91.75	0.07	90.50	0.05	90.75	0.06	76.75	0.08
KCCA	88.75	0.65	92.50	0.66	91.25	0.04	89.75	0.64	80.00	0.75
DCCA	90.50	0.53	94.75	0.65	91.75	0.86	90.75	0.75	90.50	0.64
MULPP	88.25	0.07	96.75	0.01	86.75	0.05	86.00	0.06	88.50	0.02
AE^2-Nets	90.75	0.74	95.25	0.65	90.25	0.68	90.50	0.36	91.25	0.86
MvDGPs	90.75	1.47	96.50	1.24	91.25	1.36	**91.75**	1.54	90.75	1.36
DT-MIL	88.75	1.24	94.75	1.18	90.25	1.12	88.75	1.24	88.75	1.04
Self-PATH	89.25	0.61	95.75	0.37	91.75	0.63	90.75	0.58	90.75	0.61
Pathologist	80.56	0.00	85.64	0.00	87.86	0.00	89.75	0.00	88.57	0.00
MvPID	**92.00**	0.73	**97.50**	1.33	**92.50**	0.65	**91.75**	1.89	**92.00**	0.42

effectively utilize these features, even if they are used as the input of the deep models. Compared with the multi-view learning methods, MvPID also achieves competitive results.

5.4 Comparison of different view combinations

Different views provide different information, hence we explore the impact of different view combinations on classification results. We take the GN/UN binary classification as an example. For the original pathology image, gray features (v1), texture features (v2), HOG (histogram of oriented gradient) features [52] (v3), and binary features (v4) are extracted as different views, respectively. We use the five-fold crossover method to select the 80% training set and the 20% test set, and the mean and standard deviation are recorded. The performance of the proposed MvPID on different view combinations is shown in Table 3.

Table 2 Pathology image diagnosis results of different methods on BREAST-LNM dataset. Best performances are shown in bold

Methods	BREAST-LNM			
	Accuracy	F1-score	Precision	Recall
DCN-Concate	60.27	52.18	50.94	54.39
LDA	58.53	51.67	50.48	55.84
CNN-MIL [39]	52.86	53.97	48.20	61.30
DT-MIL	70.63	63.93	61.10	67.05
MULPP	66.48	60.29	54.85	64.58
AE^2-Nets	68.83	63.27	58.69	66.47
Self-PATH	60.59	62.69	54.28	64.92
MvPID	**76.37**	**70.66**	**68.82**	**73.45**

The experimental results in Table 3 show that different view combinations have different effects on classification results. For GN/UN classification tasks, it is better to select grays features and texture features. In particular, in some special environments where all features cannot be used simultaneously, the proposed framework can provide information to select better feature combination. In addition, the experimental results also show that different classifiers have different performances on the representations learned by MvPID.

The proposed model can use different network depths to extract different view features. In order to verify the influence of different view features on the diagnosis results, we conducted experiments to present the diagnosis results of different views in different depths. The experimental results are shown in Fig. 4. We use (layer 1, layer 2, layer 3) to represent the deep structure of the network, where layer 1 represents the deep structure of texture view, layer 2 represents the deep structure of gray view, and layer 3 represents the deep structure of shape view. The experimental results show that the view specific deep Gaussian processes can model different views differently. Therefore, we can adjust different depths to achieve better diagnostic results.

Table 3 The performance of the model on different view features. Best performances are shown in bold

Views		MvPID+SVM		MvPID+KNN		MvPID+NaiveBayes		MvPID+Adaboost	
		Mean	std	Mean	std	Mean	std	Mean	std
v1	v2	92.00	0.0467	**96.50**	0.5473	**96.50**	0.4564	**98.50**	0.8786
v1	v3	93.00	0.0268	95.00	0.8543	94.75	0.8378	97.25	0.8544
v1	v4	**94.50**	0.0953	95.75	0.9432	93.50	0.3895	95.00	0.8754
v2	v3	91.75	0.0643	95.50	0.6438	95.75	0.7553	**98.50**	0.7543
v2	v4	92.00	0.0954	**96.50**	0.8569	86.75	0.5478	96.75	0.4585
v3	v4	90.50	0.0503	93.00	0.2486	93.25	0.6536	95.25	0.6549

5.5 The performance of MvPID in different tumor classification tasks

In order to further verify the rationality and efficiency of the proposed MvPID in pathology image diagnosis, we conduct experiments in different tumor classification tasks as shown in Table 4. We use the five-fold crossover method to select the 60% training set and the 40% test set, and the mean and standard deviation are recorded.

The experimental results in Table 4 show that the proposed model is effective in pathological diagnosis, and the diagnostic results are relatively stable. At the same

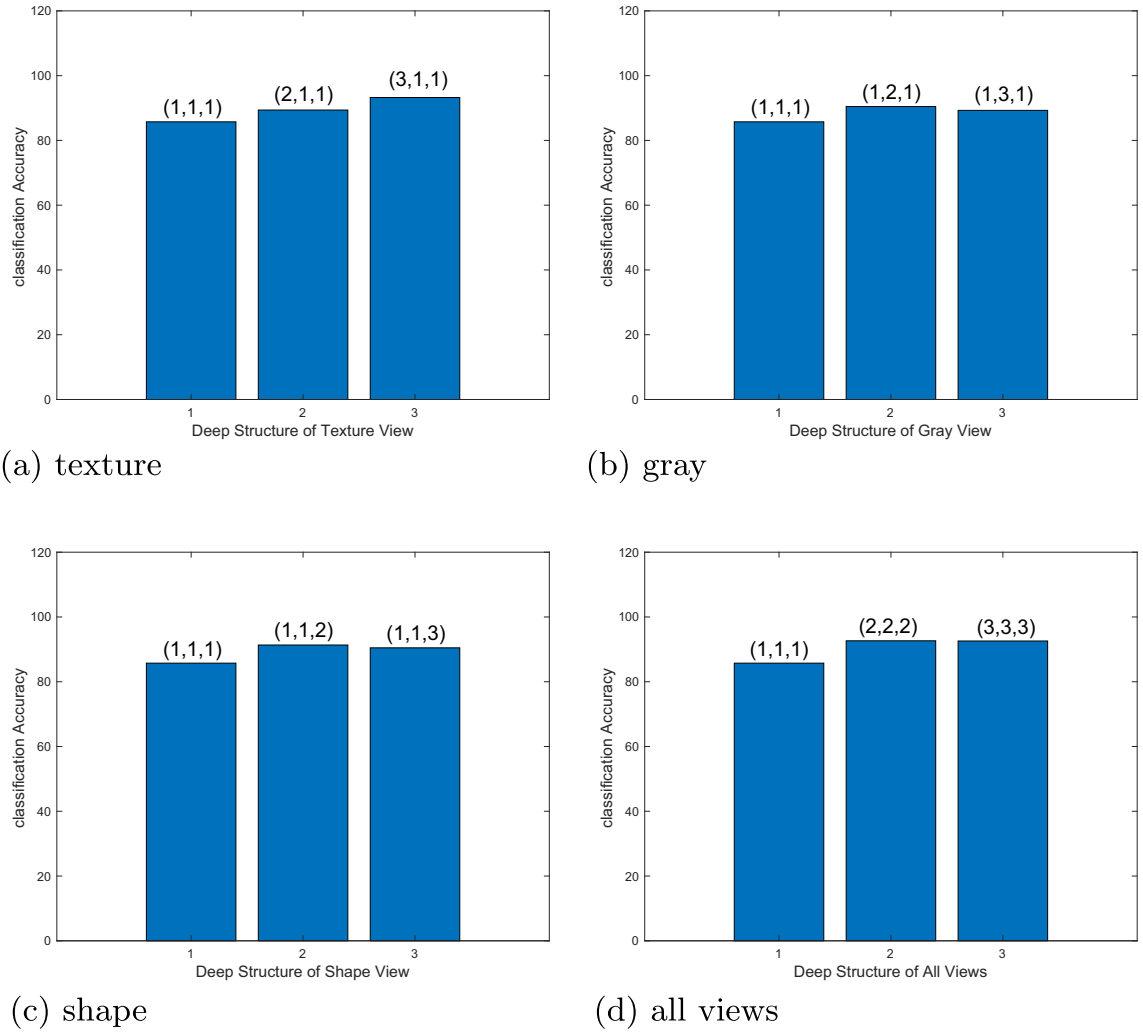

Fig. 4 The diagnosis results of different views in different depths

Table 4 The performance of MvPID+ different classifiers in tumor classification tasks. Best performances are shown in bold

Tumors		SVM		NaiveBayes		KNN		Adaboost	
		Mean	std	Mean	std	Mean	std	Mean	std
D	EWS	63.50	0.5634	62.25	4.8643	**80.75**	1.6457	75.50	0.5483
D	GN	62.75	0.5732	60.00	4.5479	78.50	1.2356	**80.50**	0.4356
D	GNBi	56.50	0.9734	58.00	4.3742	60.75	1.4634	**65.75**	0.3577
D	PD	62.25	0.5383	58.00	4.2477	65.00	1.8434	**75.25**	0.5438
D	UD	72.75	0.6538	68.75	4.5683	**85.00**	1.2456	82.25	0.6568
D	UN	88.50	0.9453	82.00	3.4267	**93.75**	1.4365	90.75	0.5437
EWS	UN	92.25	0.4573	88.00	4.1043	**97.25**	1.4574	96.75	0.3588
EWS	PD	63.75	0.8328	61.75	3.4338	**65.50**	1.3468	62.50	0.2654
EWS	NOS	88.25	0.5793	87.75	3.4578	**95.00**	1.7654	92.25	0.7643
EWS	GNBi	65.00	0.5683	67.75	2.4773	80.00	1.5664	**83.50**	0.5632
EWS	GN	78.50	0.5665	76.75	4.6583	82.00	1.7644	**86.25**	0.3547
GN	GNBi	65.50	0.3358	68.00	3.7727	73.25	1.5689	**76.50**	0.6832
GN	NOS	94.00	0.7653	88.25	3.7638	97.25	1.6545	**97.50**	0.7594
GN	PD	65.75	0.5683	69.25	3.7538	83.75	1.7649	**87.00**	0.6554
GN	UD	78.00	0.5458	77.50	3.6562	88.50	1.7643	**91.50**	0.6538
GNBi	NOS	85.50	0.3869	82.75	3.6538	**92.25**	1.7354	89.25	0.7649
GNBi	PD	69.00	0.7434	67.50	3.6563	68.50	1.7648	**72.50**	0.6549
GNBi	UD	65.50	0.6739	69.25	4.1357	81.00	1.7643	**84.50**	0.6757
NOS	UN	95.00	0.5638	80.50	2.7643	95.00	1.3586	**98.00**	0.7684
NOS	UD	80.50	0.6709	78.50	3.7653	**92.25**	1.7638	89.75	0.7644
NOS	PD	85.50	0.7843	89.75	2.8656	**90.50**	1.5663	88.50	0.6553
PD	UD	62.75	0.3659	60.25	2.5678	65.00	1.6548	**66.25**	0.5649

time, the experimental results show that the selection of downstream classifiers also has an impact on the performances of the model, but this impact is relatively small, which proves that the representation learned by MvPID has strong robustness.

In order to make the proposed model more practical, we test the diagnostic performance of the proposed model according to different patients. We collated the collected PNTs pathology image dataset, and used 2100 pathology images of 16 patients to evaluate the effectiveness of the proposed method. Specifically, we divide the training set and the test set according to the leave-one-out method, that is, we use the pathology images of one patient for testing, and use the pathology images of other patients for training. The experimental results are shown in Fig. 5. In Fig. 5, the numbering rule of the patient is (patient-tumor types-patient ID).

As can be seen from Fig. 5, when we test different patients, the diagnosis results obtained by the proposed model exceed the diagnosis results of pathologists in most scenarios. However, when we tested patients P-11-1223 and P-12-2669, the diagnostic results of the model were lower than those of the pathologists. The reason is that when we test the two patients, the same cases did not appear in the training set. Therefore, the model did not learn the corresponding information, which led to the relatively unsatisfactory diagnostic results. The experimental results show that when the training data set is sufficient and diverse, the MvPID can obtain better diagnostic results than pathologists, which can reduce the workload of pathologists.

5.6 Ablation study

In this section, we conducted ablation experiments to verify the effects of different components in the model, mainly including the effects of view-specific deep Gaussian processes and view-common AE network.

The influence of view-specific deep Gaussian processes In order to explore the impact of view-specific deep Gaussian process on model performance, we use PCA, LDA, t-SNE, LLE and GPLVM as different view-specific feature extraction methods for comparison. The experiment

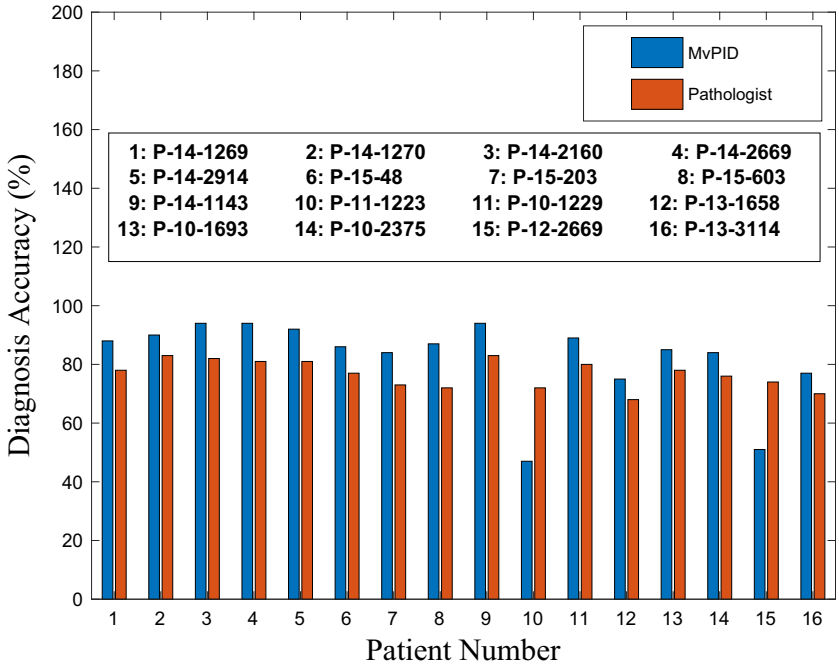

Fig. 5 Comparative diagnosis results of different patients

is a classification task with eight categories, and the classification accuracy is used as the evaluation criterion. The experimental results are shown in Table 5.

As can be seen from Table 5, compared with other methods, the view-specific deep Gaussian process can better extract the unique information of different views. The reason is that the deep Gaussian process is a deep structure with Gaussian mapping, and the Gaussian precess is equivalent to the single-layer infinite width neural network. Therefore, the ability of feature expression has been improved. Meanwhile, the view-specific deep Gaussian processes we designed can use the deep Gaussian process with different deep structure for different views, so that we can extract the unique information of the view more pertinently.

The influence of view-common AE network In order to explore the impact of view-common AE network on model performance, we integrate the features of different views by concatenating and weighting. The experimental setup is similar to the previous one. The experimental results are shown in Table 6.

It can be seen from Table 6 that AE network can better integrate the feature information of different views, and then learn a common representation, which is better than directly concatenating or weighting the feature information of different views.

6 Conclusion

In computer-aided medicine, it is a potential and inevitable trend to use artificial intelligence technology to simulate the doctor's analysis and diagnosis process by comprehensively using image processing methods and combining the characteristics of human body and diseases. In this paper, we applied multi-view learning to pathology image diagnosis, which can reduce the workload of pathologists and improve the efficiency and accuracy of tumor recognition. For pathology images, we extracted different image features as different views for multi-view learning. Subsequently, we propose to use the view-specific deep Gaussian processes to extract the unique information of different views and the view-common AE network to integrate the information of different views into a common representation. The common representation is put into the downstream classifier to realize automatic pathological diagnosis. Experimental results on real data show the effectiveness and rationality of the proposed model. In the future work, we will focus on improving the calculation speed of the model to develop

Table 5 The impact of view-specific deep Gaussian processes.

Methods	PCA	LDA	t-SNE	LLE	GPLVM	View-specific DGPs
Classification ACC	70.3	74.8	81.2	78.4	82.5	86.6

Table 6 The impact of AE network

Methods	Concatenating	Weighting	AE network integration
Classification ACC	75.8	80.3	86.6

specific medical software. In addition, we will explore multi-view fusion between pathology images and medical text information (i.e. immunohistochemical data), and consider using a pre-trained deep model to extract features from pathology images to further improve the diagnostic performance.

Acknowledgments This work is supported by the National Natural Science Foundation of China under Project 62076096, Shanghai Municipal Project 20511100900, and Shanghai Knowledge Service Platform Project (No. ZF1213).

Declarations

Competing interests The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. **Wenbo Dong:** Conceptualization, Methodology, Software, writing-original draft. **Shiliang Sun:** Supervision, writing-review & editing, Funding acquisition. **Minzhi Yin:** Data curation, Supervision, Formal analysis.

References

1. Subramanian H, Subramanian S (2022) Improving diagnosis through digital pathology: proof-of-concept implementation using smart contracts and decentralized file storage. J Med Internet Res 24(3):207–217
2. Liu S, Zhang B, Liu Y, Han A, Shi H, Guan T, He Y (2021) Unpaired stain transfer using pathology-consistent constrained generative adversarial networks. IEEE Trans Med Imaging 40(8):1977–1989
3. Liu Y, Yin M, Sun S (2018) Multi-view learning and deep learning for microscopic neuroblastoma pathology image diagnosis. In: Pacific rim international conference on artificial intelligence, Nanjing, China, pp 545–558
4. Li Z, Zhang P, Xie N, Zhang G, Wen C-F (2020) A novel three-way decision method in a hybrid information system with images and its application in medical diagnosis. Eng Appl Artif Intell 92:1–16
5. Fink O, Wang Q, Svensen M, Dersin P, Lee W-J, Ducoffe M (2020) Potential, challenges and future directions for deep learning in prognostics and health management applications. Eng Appl Artif Intell 92:1–15
6. Qu J, Hiruta N, Terai K, Nosato H, Murakawa M, Sakanashi H (2018) Gastric pathology image classification using stepwise fine-tuning for deep neural networks. J Healthcare Eng 2018:1781–1792
7. Campanella G, Hanna MG, Geneslaw L, Miraflor A, Silva VWK, Busam KJ, Brogi E, Reuter VE, Klimstra DS, Fuchs TJ (2019) Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nat Med 25(8):1301–1309
8. Retamero JA, Aneiros-Fernandez J, Del Moral RG (2020) Complete digital pathology for routine histopathology diagnosis in a multicenter hospital network. Arch Pathol Lab Med 144(2):221–228
9. Xu G, Song Z, Sun Z, Ku C, Yang Z, Liu C, Wang S, Ma J, Xu W (2019) Camel: a weakly supervised learning framework for histopathology image segmentation. In: IEEE International conference on computer vision, Seoul, Korea, pp 10682–10691
10. Tang C, Ji J, Tang Y, Gao S, Tang Z, Todo Y (2020) A novel machine learning technique for computer-aided diagnosis. Eng Appl Artif Intell 92:1–14
11. Vivaldi N, Caiola M, Solarana K, Ye M (2021) Evaluating performance of EEG data-driven machine learning for traumatic brain injury classification. IEEE Trans Biomed Eng 68(11):3205–3216
12. Avendaño-Valencia LD, Yderstræde KB, Nadimi ES, Blanes-Vidal V (2021) Video-based eye tracking performance for computer-assisted diagnostic support of diabetic neuropathy. Artif Intell Med:1–12
13. Nateghi R, Danyali H, Helfroush MS (2021) A deep learning approach for mitosis detection: application in tumor proliferation prediction from whole slide images. Artif Intell Med:1–13
14. Wang X, Lei Z, Guo X, Zhang C, Shi H, Li SZ (2019) Multi-view subspace clustering with intactness-aware similarity. Pattern Recogn 88:50–63
15. Qu G, Xiao L, Hu W, Wang J, Zhang K, Calhoun VD, Wang Y-P (2021) Ensemble manifold regularized multi-modal graph convolutional network for cognitive ability prediction. IEEE Trans Biomed Eng 68(12):3564–3573
16. Zhao Y, You X, Yu S, Xu C, Yuan W, Jing X-Y, Zhang T, Tao D (2018) Multi-view manifold learning with locality alignment. Pattern Recogn 78:154–166
17. Zhang C, Fu H, Hu Q, Cao X, Xie Y, Tao D, Xu D (2020) Generalized latent multi-view subspace clustering. IEEE Trans Pattern Anal Mach Intell 42(1):86–99
18. Zhu W, Lu J, Zhou J (2019) Structured general and specific multi-view subspace clustering. Pattern Recogn 93:392–403
19. Li Y, Yang M, Zhang Z (2018) A survey of multi-view representation learning. IEEE Trans Knowl Data Eng 31(10):1863–1883
20. Dinh P-H (2021) Multi-modal medical image fusion based on equilibrium optimizer algorithm and local energy functions. Appl Intell 50:8416–8431
21. Wang S, Yang DM, Rong R, Zhan X, Xiao G (2019) Pathology image analysis using segmentation deep learning algorithms. Am J Pathol 189(9):1686–1698
22. Robertson S, Azizpour H, Smith K, Hartman J (2018) Digital image analysis in breast pathology from image processing techniques to artificial intelligence. Transl Res 194:19–35
23. Niazi MKK, Parwani AV, Gurcan MN (2019) Digital pathology and artificial intelligence. Lancet Oncol 20(5):253–261
24. Bera K, Schalper KA, Rimm DL, Velcheti V, Madabhushi A (2019) Artificial intelligence in digital pathology new tools for diagnosis and precision oncology. Nat Rev Clin Oncol (11):703–715
25. Parwani AV (2019) Next generation diagnostic pathology: use of digital pathology and artificial intelligence tools to augment a pathological diagnosis. Diagn Pathol 14(1):1–3
26. Echle A, Rindtorff NT, Brinker TJ, Luedde T, Pearson AT, Kather JN (2021) Deep learning in cancer pathology: a new generation of clinical biomarkers. Br J Cancer 124(4):686–696
27. Sharma H, Zerbe N, Klempert I, Hellwich O, Hufnagl P (2017) Deep convolutional neural networks for automatic classification of gastric carcinoma using whole slide images in digital histopathology. Comput Med Imaging Graph 61:2–13

28. Yuan Y, Xun G, Jia K, Zhang A (2018) A multi-view deep learning framework for eeg seizure detection. IEEE J Biomed Health Inform 23(1):83–94
29. Pinckaers H, Bulten W, van der Laak J, Litjens G (2021) Detection of prostate cancer in whole-slide images through end-to-end training with image-level labels. IEEE Trans Med Imaging 40(7):1817–1826
30. Wang S, Yang DM, Rong R, Zhan X, Fujimoto J, Liu H, Minna J, Wistuba II, Xie Y, Xiao G (2019) Artificial intelligence in lung cancer pathology image analysis. Cancers 11(11):1673–1683
31. Gandomkar Z, Brennan PC, Mello-Thoms C (2018) Mudern: multi-category classification of breast histopathological image using deep residual networks. Artif Intell Med 88:14–24
32. Sirazitdinov I, Kholiavchenko M, Kuleev R, Ibragimov B (2019) Data augmentation for chest pathologies classification. In: IEEE international symposium on biomedical imaging, Venice, Italy, pp 1216–1219
33. Wright AI, Dunn CM, Hale M, Hutchins GGA, Treanor DE (2021) The effect of quality control on accuracy of digital pathology image analysis. IEEE J Biomed Health Inform 25(2):307–314
34. Dietterich TG, Lathrop RH, Perez TL (1997) Solving the multiple instance problem with axis parallel rectangles. Artif Intell 89(1):31–71
35. Zhu Z, Ding X, Zhang D, Wang L (2020) Weakly-supervised balanced attention network for gastric pathology image localization and classification. In: IEEE international symposium on biomedical imaging, Iowa City, USA, pp 1–4
36. Corvo A, Caballero HSG, Westenberg MA, van Driel MA, van Wijk JJ (2021) Visual analytics for hypothesis-driven exploration in computational pathology. IEEE Trans Visual Comput Graph 27(10):3851–3866
37. Koohbanani NA, Unnikrishnan B, Khurram SA, Krishnaswamy P, Rajpoot N (2021) Self-path: self-supervision for classification of pathology images with limited annotations. IEEE Trans Med Imaging 40(10):2845–2856
38. Lu MY, Williamson DF, Chen TY, Chen RJ, Barbieri M, Mahmood F (2021) Data-efficient and weakly supervised computational pathology on whole-slide images. Nat Biomed Eng 5(6):555–570
39. Li H, Yang F, Zhao Y, Xing X, Zhang J, Gao M, Huang J, Wang L, Yao J (2021) Dt-mil: Deformable transformer for multi-instance learning on histopathological image. In: International conference on medical image computing and computer-assisted intervention, Strasbourg, France, pp 206–216
40. Wang H, Hu J, Song Y, Zhang L, Bai S, Yi Z (2021) Multi-view fusion segmentation for brain glioma on CT images. Appl Intell 1–15
41. Mi Y, Ren Z, Mukherjee M, Huang Y, Sun Q, Chen L (2021) Diversity and consistency embedding learning for multi-view subspace clustering. Appl Intell 1–14
42. Koohzadi M, Charkari NM, Ghaderi F (2020) Unsupervised representation learning based on the deep multi-view ensemble learning. Appl Intell 50(2):562–581
43. Li J, Li Z, Lu G, Xu Y, Zhang B, Zhang D (2021) Asymmetric 5gaussian process multi-view learning for visual classification. Inf Fusion 65:108–118
44. Yin J, Sun S (2020) Multiview uncorrelated locality preserving projection. IEEE Trans Neural Netw Learn Syst 31(9):3442–3455
45. Yang X, Deng C, Dang Z, Tao D (2021) Deep multiview collaborative clustering. IEEE Trans Neural Netw Learn Sys 1–11
46. Zhang C, Liu Y, Fu H (2019) AE^2-Nets: autoencoder in autoencoder networks. In: IEEE conference on computer vision and pattern recognition, Los Angeles, USA, pp 2577–2585
47. Sun S, Dong W, Liu Q (2021) Multi-view representation learning with deep Gaussian processes. IEEE Trans Pattern Anal Mach Intell 43(12):4453–4468
48. Damianou A, Lawrence N (2013) Deep Gaussian processes. In: International conference on artificial intelligence and statistics, Scottsdale, Arizona, USA, pp 207–215
49. Shimada H, Ambros IM, Dehner LP, Hata J-I, Joshi VV, Roald B (1999) Terminology and morphologic criteria of neuroblastic tumors: recommendations by the international neuroblastoma pathology committee. Cancer: Interdiscip Int J Am Cancer Soc 86(2):349–363
50. Hardoon DR, Szedmak S, Shawe-Taylor J (2004) Canonical correlation analysis: an overview with application to learning methods. Neural Comput 16(12):2639–2664
51. Andrew G, Arora R, Bilmes J, Livescu K (2013) Deep canonical correlation analysis. In: International conference on machine learning, Atlanta, USA, pp 1247–1255
52. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: IEEE Computer society conference on computer vision and pattern recognition, San Diego, CA, USA, pp 886–893