Medical Image Analysis 62 (2020) 101640

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

SDAE-GAN: Enable high-dimensional pathological images in liver cancer survival prediction with a policy gradient based data augmentation method

Hejun Wu[a], Rong Gao[a], Yeong Poh Sheng[b], Bo Chen[c], Shuo Li[d,e,*]

[a] Department of Computer Science, Sun Yat-sen University, Guangzhou, China
[b] Singapore General Hospital, Singapore
[c] School of Health Science, Western University, Canada
[d] Department of Medical Imaging and Medical Biophysics, Western University, Canada
[e] Digital Imaging Group of London, Canada

ARTICLE INFO

Article history:
Received 5 September 2019
Revised 9 December 2019
Accepted 9 January 2020
Available online 11 February 2020

Keywords:
Computer-aided diagnosis
Imaging and non-Imaging biomarkers
Integration
Cancer survival prediction

ABSTRACT

High-dimensional pathological images produced by Immunohistochemistry (IHC) methods consist of many pathological indexes, which play critical roles in cancer treatment planning. However, these indexes currently cannot be utilized in survival prediction because joining them with patients' clinicopathological features (e.g., age and tumor size) is challenging due to their high dimension and sparse features. To address this problem, we propose a novel two-stage survival prediction model named ICSPM to join the IHC images and clinicopathological features. For the first stage, our proposed SDAE-GAN compresses high-dimensional IHC images to flat, compact and representative feature vectors by compressing and reconstructing them. For the first time, SDAE-GAN integrates dense blocks, the stacked auto-encoder and the GAN architecture to maximize the ability to detect patterns in IHC images. In addition, we propose a novel policy gradient based data augmentation method to involve the diversity in IHC images without breaking patterns inside them. For the second stage, ICSPM adopts a DenseNet to join feature vectors and clinicopathological features for survival prediction. Experimental results demonstrate that ICSPM reached a state-of-the-art prediction accuracy of 0.72 on the five-year survival. ICSPM is the first work to enable high-dimensional IHC images in cancer survival prediction. We prove that high-dimensional IHC images and clinicopathological features provide valuable and complementary information in survival prediction.

1. Introduction

Immunohistochemistry(IHC) plays a critical role in cancer diagnosis, treatment planning, and pathological decisions by showing the status and development of cancer in a very low-level perspective. IHC methods produce pathological indexes such as DAPI, Glypican, and CD3, which reveal the low-level information about the immune system or liver cancer. Therefore, IHC methods can be supplementarily utilized in cancer survival prediction. However, the previous studies only used clinicopathological features such as the tumor size, genetic data and pathological images (Chen et al., 2011; Vanneschi et al., 2011; Zhu et al., 2016). Table 1 lists the clinicopathological features and pathological indexes used in this paper.

These pathological indexes contain seven groups related to the immune system or liver cancer. PD-1(Programmed cell death protein 1) prevents the immune system from killing cancer cells (Syn et al., 2017). Glypicans play a vital role in developmental morphogenesis (Filmus et al., 2008). Abnormal expression of glypicans has been noted in multiple types of cancer (Li et al., 2017). ECAD(E-cadherin) increases in cellular motility. It allows cancer cells to cross the basement membrane and invade surrounding tissues (Weinberg, 2013). DAPI is used extensively in fluorescence microscopy to stain both live and fixed cells. CD103 (cluster of differentiation 103) is expressed widely on different types of T cells and a subset of dendritic cells (Lehmann et al., 2002; Aziz et al., 2005; Johansson-Lindbom et al., 2005). It plays a key role in the immune system and fighting cancer, controlling cancer progression (Broz et al., 2014; Binnewies et al., 2019). CD3 (cluster of differentiation 3) and CD8 (cluster of differentiation 8) are both T cell co-receptors (Gao and Jakobsen, 2000). In immunology, the CD3 T cell co-receptor helps to activate both the cytotoxic T cells (CD8+

* Corresponding author.
E-mail address: sli287@uwo.ca (S. Li).

https://doi.org/10.1016/j.media.2020.101640
1361-8415/

Table 1
Used features and indexes .

Clinicopathological features	Immunohistochemistry indexes
HepB	PD1 (nucleus, cytoplasm, membrane, entire cell)
HepC	Glypican (nucleus, cytoplasm, membrane, entire cell)
AFP level	Ecad (nucleus, cytoplasm, membrane, entire cell)
Histo largest tumor size	DAPI (nucleus, cytoplasm, membrane, entire cell)
Histo Grade	CD103 (nucleus, cytoplasm, membrane, entire cell)
Patho Stage	CD3 (nucleus, cytoplasm, membrane, entire cell)
T Satge	CD8 (nucleus, cytoplasm, membrane, entire cell)
N Stage	(28 indexes in total)
Age when operated	
Gender	

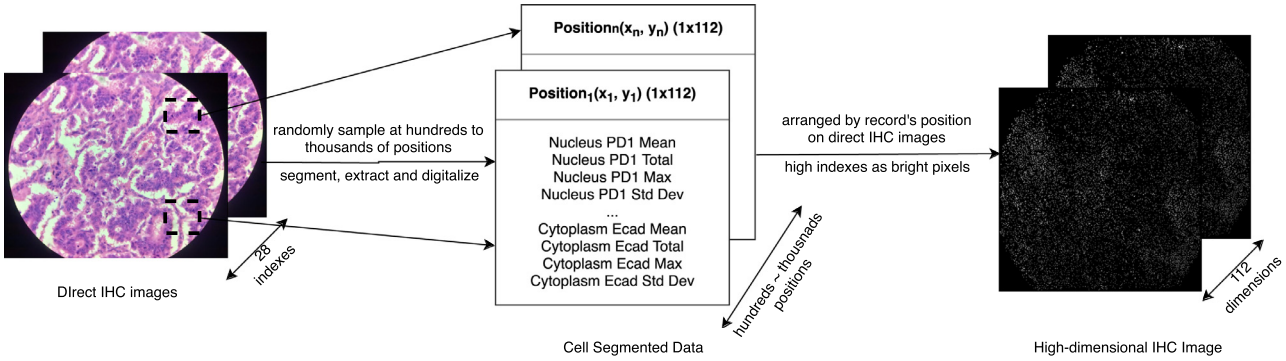

Fig. 1. Direct IHC images do not have numeric marks. They are segmented, extracted and digitalized as hundreds to thousands of record. Then this numeric record is converted to gray scale images in which bright pixels represent high pathological indexes, named high-dimensional IHC image.

naive T cells) and also T helper cells (CD4+ naive T cells). Considering these seven groups of indexes on three different components (nucleus, cytoplasm and membrane) of the cell and the entire cell, we got 28 pathological indexes.

IHC cannot be directly used in traditional classification or regression models. The output of the IHC is direct IHC images without numeric marks, shown in Fig. 1. They are extracted and digitalized to cell-segmented data with numeric marks by some automatic medical methods. Though cell-segmented data obtained from direct IHC images has numeric pathological indexes, it is a great challenge to use cell-segmented data directly due to three reasons: 1) The number of features is too large. Four values (mean value, maximum value, standard deviation and total value) of 28 indexes make 112 statistics for each sample position. Patients have hundreds to thousands of sample positions containing all these 112 statistics. Overall, each patient has hundreds of thousands of features. 2) Patients have different numbers of samples at different positions. Such variances in samples and positions result in significantly different feature spaces for individual patients. Moreover, 3) These indexes are very sparse because most of them are at very low-value levels. Only parts of these indexes have meaningful values. These three factors make it inapplicable to use these pathological indexes for survival prediction direct.

Representing pathological indexes as grayscale images is an intuitive and better way to utilize them instead of directly using them as structureless features. As shown in Fig. 1, these indexes are represented as grayscale images (200*200*112) in which bright pixels mean high pathological indexes values. This process brings three benefits: 1) Position relations among sample positions are preserved, 2) Image processing models like CNN can be adopted to detect patterns in these indexes and 3) The dimension of images reduces to a lower extent. These images are named high-dimensional IHC images, exemplified in Fig. 2. Since five years is a long period in survival prediction, there are no directly and easily distinguishable differences between images of deceased and alive patients. Nevertheless, both of them have some salient clusters of bright or dark pixels and noticeable shapes existing in images. This fact indicates that these pathological indexes contain meaningful information about cells and liver cancer instead of just noises.

Models to join image input and scalar input have shown their advantages. In reinforcement learning, (Levine et al., 2018) proposed an intuitive and universal approach. To concatenate scalar input and images, (Levine et al., 2018) used scalar input as an additional input to a fully connected layer of the CNN, which processed the image input. Larsen et al. (2015) and Masci et al. (2011) proposed two auto-encoders to use an image-embedding way to extract features from images. After image-embedding, these features from images can be regarded as numeric and flat feature vectors to work with other scalar inputs. In cancer survival prediction, (Zhu et al., 2016) proposed a model to extract geometry and texture descriptors from direct pathological images by CellProfiler (Carpenter et al., 2006) and joined them to genetic data for survival prediction.

Nevertheless, these previous models are incapable of joining the IHC images and scalars due to the three differences between IHC images and direct pathological images as follows: IHC images have 1) fewer continuous lines; 2) fewer noticeable shapes and 3) too much more channels. Traditional deep learning architectures (Xue et al., 2017a; 2017b; 2017c; Zhen et al., 2015) which can handle MRI images may fail to detect all their irregular image patterns because convolution layers gradually discover lines, curves, and shapes layer by layer. Therefore, these models are not applicable to join high-dimensional IHC images and regular clinicopathological features.

Furthermore, traditional data augmentation methods are inapplicable to high-dimensional IHC images.Data augmentation is regularly adopted in image processing models to enhance their robustness, especially for medical image processing models due to the lack of samples. However, traditional data augmentations are not usable for high-dimensional IHC images. As can be seen in Fig. 2, these images have sparse bright pixels representing clinical indexes with high values. For high-dimensional IHC images, tra-

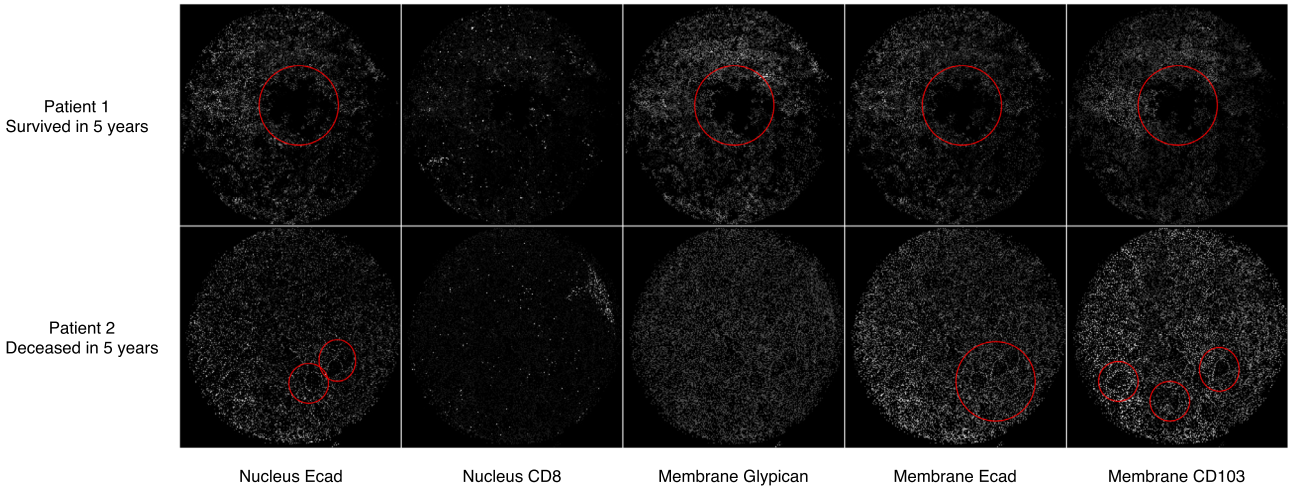

Fig. 2. IHC images of different pathological indexes of two patients. One survived in five years while the other not. IHC images contain salient clusters of bright or dark pixels, e.g., some recognizable circles in visualized Membrane CD103 of the deceased patient. Nevertheless, no decisive patterns can be seen directly to predict a patient's five-year survival, which demands the junction of these images and clinicopathological features.

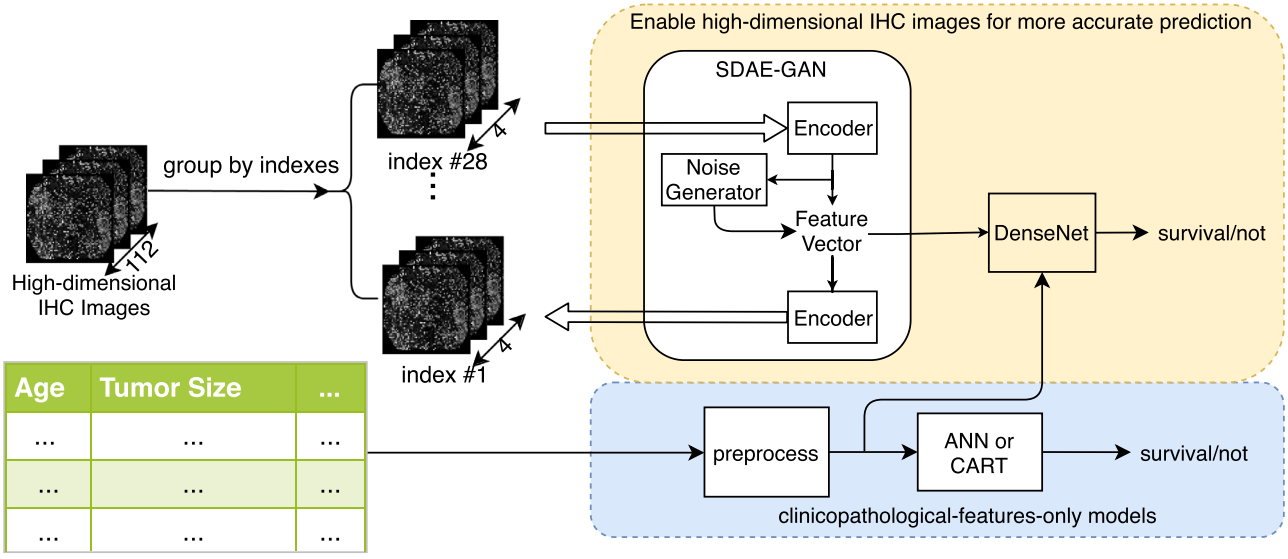

Fig. 3. Illustration of ICSPM. Unlike existing methods focusing on the classification model using clinicopathological features, ICSPM utilized high-dimensional IHC images for the first time to utilize low-level clinical indexes inside cells.

ditional data augmentation methods may fall into a dilemma of bringing too many noises to break weak patterns or too little not to involve enough diversity to enhance the robustness of classification models. Therefore, a new method for data augmentation of high-dimensional IHC images is needed to hold the right balance of the involved diversity and preserved patterns in these images.

We propose an image-compression-based survival prediction model (ICSPM) shown in Fig. 3 to join high-dimensional IHC images and patients' clinicopathological features for survival prediction for the first time. It includes two stages: 1) stage 1: Our proposed SDAE-GAN (Stacked Dense Auto-encoder GAN) compresses high-dimensional IHC images to 1-dimension compact and representative feature vectors. 2) stage 2: A DenseNet (Huang et al., 2017) takes feature vectors and patients' clinicopathological features as two parts of its input to predict a five-year survival status. Because of the high-dimensional IHC images' weakness, models using them directly for survival prediction may fail to detect patterns in images. By compressing and reconstructing IHC images while preserving salient patterns via minimizing reconstruction loss, SDAE-GAN ensures that compact feature vectors contain the essential discriminative information useful for survival prediction.

In stage 1, we use the following innovative approaches to make high-dimensional IHC images usable in liver cancer survival prediction:

1) SDAE-GAN involves dense blocks (Huang et al., 2017) to replace traditional convolution layers in the stacked auto-encoder (Vincent et al., 2010). Dense blocks maximize the information flow between layers and alleviate the vanishing-gradient problem. Therefore, dense blocks can enhance the ability of auto-encoders to detect patterns in high-dimensional IHC images and better reserve them in feature vectors.

2) We add a discriminator to make up a GAN architecture. Involving of GAN makes reconstructed images more realistic than using traditional auto-encoders (Larsen et al., 2015), which tend to produce blurry reconstructed images. Since patterns in images are contributing to survival prediction, more precise image reconstruction means more representative feature vectors.

3) In data augmentation, we propose a policy gradient (Sutton et al., 2000) based trainable method to add a noise gen-

erator to SDAE-GAN. Our designed reward function for the policy gradient method encourages the noise generator to maximize noises added to compact feature vectors without increasing the reconstruction loss. This scheme ensures to maximize the diversity involved without breaking patterns in IHC images.

In stage 2, a DenseNet is adopted for its new connection fashion's ability to alleviate the over-fitting on smaller data sets (Huang et al., 2017). Overall, underlying information of cells hidden in IHC images is utilized. This contributes to a higher accuracy than using clinicopathological features only.

The contribution of this paper is summarized as follows:

1) Our proposed SDAE-GAN enables Immunohistochemistry to work with patients' clinicopathological features in liver cancer survival prediction.

2) For the first time, our proposed policy gradient based data augmentation method uses a trainable way to involve adequate diversity without breaking patterns in high-dimensional IHC images. This method can easily be adopted in other types of medical images.

3) We prove that high-dimensional IHC images and clinicopathological features are complementary to each other for cancer survival prediction by both correlation analysis and experiments.

2. Image-compression-based survival prediction model (ICSPM)

As introduced, ICSPM uses a two-stage approach to join high-dimensional IHC images and clinicopathological features because of the limited ability of existing models to utilize this type of images. In stage 1, our proposed SDAE-GAN compresses high-dimensional IHC images to represent them as 1-dimensional feature vectors and perform data augmentation by a novel policy gradient based method. In stage 2, a DenseNet uses these compact and representative feature vectors together with patients' clinicopathological features to predict their five-year survival.

2.1. SDAE-GAN for feature extraction by image compression and data augmentation

SDAE-GAN compresses high-dimensional IHC images into compact feature vectors for more effective representation and then applies data augmentation to feature vectors, as shown in Fig. 4. Images are grouped according to pathological indexes to avoid mutual interferences among them. SDAE-GAN leverages the strengths of two unsupervised learning methods, the auto-encoder and the GAN architecture, to recognize salient patterns of these images better. It uses a stacked auto-encoder as the generator of the GAN and uses dense convolution blocks and deconvolution layers to constitute the auto-encoder. The GAN architecture is involved to train the stacked auto-encoder to output more realistic images. This design is due to the fact that the higher quality of output images leads to the more accurate feature representation of original IHC images. Furthermore, SDAE-GAN augments data by using a noise generator to involve some additional noises to feature vectors, similar to the way in which VAE-GAN (Kingma and Welling, 2013) brings diversity to its generated images.

2.1.1. Stacked dense auto-encoder for compressing images to compact feature vectors

SDAE uses convolution auto-encoder (CAE) as its base component to compress images. By compressing original images to 1-dimensional feature vectors and reconstructing them to minimize the reconstruction loss, CAE retains the essential information of images in a much lower-dimensional and more effective representation. However, CAE shows high instability on high-dimensional IHC images because of the disadvantages of this type of images mentioned in Section I. Our proposed SDAE replaces the convolution layers with dense convolution blocks and stacked them for a better stability and a stronger ability to detect features. Therefore, SDAE is able to obtain a more accurate representation of features.

In SDAE, every single auto-encoder uses a dense convolution block as an encoder to compress the original input to a lower dimension. Since pixels in IHC images are sparse, traditional convolution kernel may be inefficient. A dilated convolution layer (Yu and Koltun, 2015) is applied at the beginning of each dense block for a larger receptive field. Subsequently, it uses a deconvolution layer to reconstruct the original input and minimize the reconstruction loss. SDAE stacks dense auto-encoders for a layer-by-layer compression. Each encoder uses the output of its previous encoder and reconstructs it. The encoders are greedily trained one by one to minimize their reconstruction losses. The output of the last encoder is the feature vectors, which are also the output of the entire SDAE, used in survival prediction. This architecture has the following advantages: 1) Dense blocks and dilated convolution layers enhance the ability to detect patterns in high-dimensional IHC images. 2) Greedily trained layers of the auto-encoder ensure the convergence of training. These two designs make SDAE more efficient and effective in processing high-dimensional IHC images.

Suppose that x is the input of layer l while θ^l_{conv} and θ^l_{deconv} are the parameters of dense convolution block and deconvolution layer of layer l respectively. The encoding process, the decoding process and the loss function are given as:

$$f(x) = Conv(\theta^l_{conv}, x) \tag{1}$$

$$g(f(x)) = Deconv(\theta^l_{deconv}, Conv(\theta^l_{conv}, x)) \tag{2}$$

$$L(\theta^l_{conv}, \theta^l_{deconv}, x) = \frac{1}{2}(g(f(x)) - x)^2 \tag{3}$$

Similar to traditional CAE (Masci et al., 2011), SDAE has many stacked auto-encoders that are greedily trained layer by layer. Every single auto-encoder compresses and reconstructs the feature map of its previous auto-encoder. When training a layer, parameters of all its previous layers are frozen. After greedily training all the layers, these trained dense convolution blocks and deconvolution layers constitute the encoder and the decoder of SDAE respectively. An incoming image first passes through dense convolution blocks. Then it is converted to the feature vector z. Next, the code z passes through deconvolution layers to reconstruct the original image. Supposing $\tilde{x}$ is the reconstructed image, a reconstruction loss is given as:

$$L_r(x, \tilde{x}) = \frac{1}{2}(\tilde{x} - x)^2 \tag{4}$$

To fine-tune the whole SDAE, L_r is minimized on all parameters of the encoder and the decoder after greedily pre-training.

2.1.2. Policy gradient based noise generator for data augmentation

An additional module called noise generator is proposed to provide adequate but controllable data augmentation. The noise generator takes original images as input and outputs the noises added to them. For space and time efficiency, it reuses the convolution layers of the encoder and attaches two fully connected layers, as shown in Fig. 4. For each image s input to SDAE, the noise generator G, with θ as its parameters, outputs the mean of a Gaussian distribution. Then it samples some noise vectors n on the given Gaussian distribution and adds them to the corresponding feature vector to perform data augmentation. The probability density of the noise samples is shown in Eq. 5:

$$\pi_{\theta,s}(n) = N(G(\theta, s), \sigma^2) \tag{5}$$

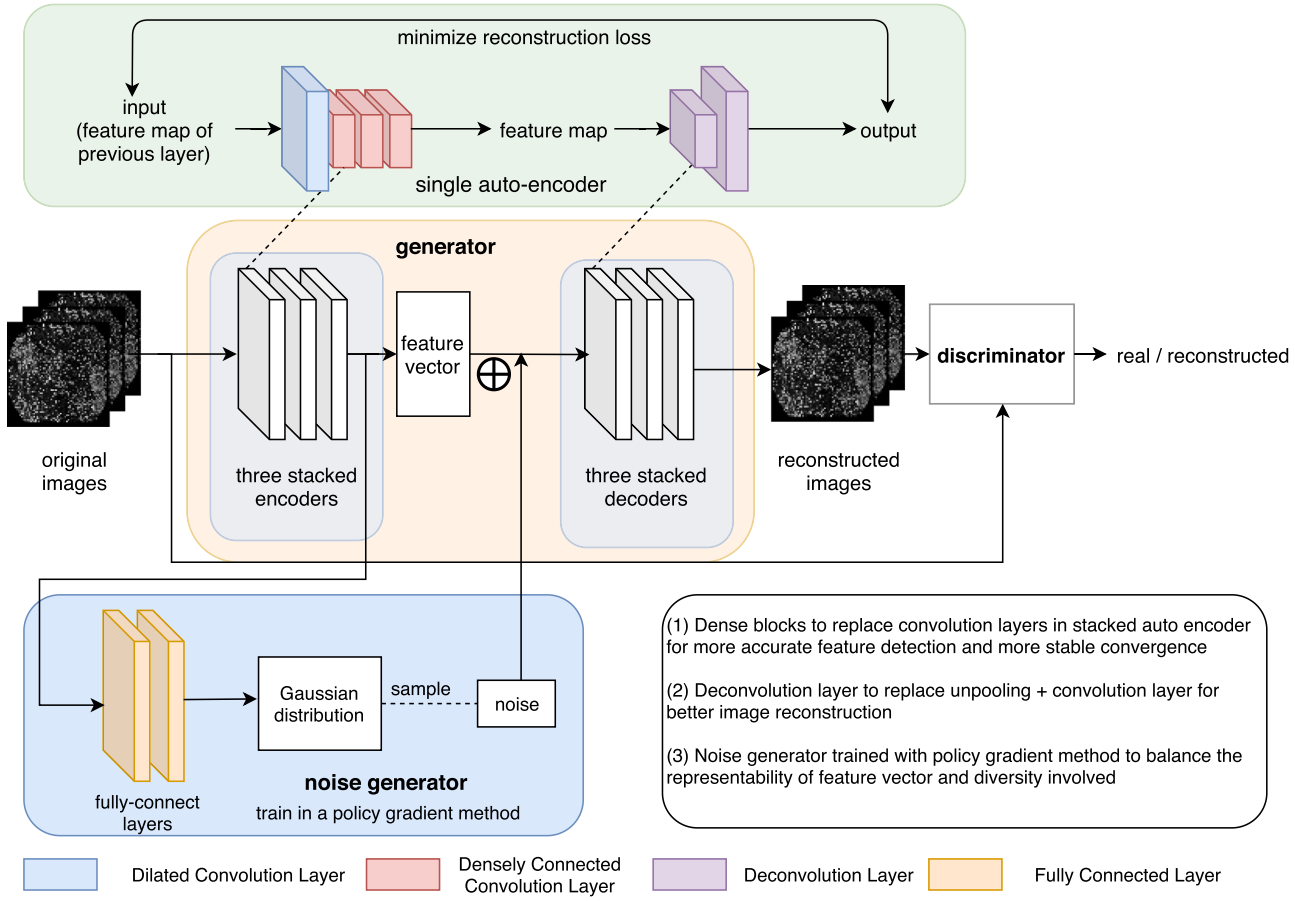

Fig. 4. Architecture of SDAE-GAN. Every single auto-encoder compresses its input and reconstructs it to train an encoder and a decoder. Several single auto-encoders are stacked and use their previous encoder's feature map as input. Then single auto-encoders are trained layer by layer greedily. At last, encoders and decoders of all the single auto-encoders make up the encoder and decoder, respectively. This training process ensures features to be preserved in compressing. In addition, a noise generator is used to apply data augmentation on feature vectors. GAN architecture is used for producing more realistic reconstructed images (Dosovitskiy and Brox, 2016) to obtain more representative feature vectors.

The noise generator needs to do adequate data augmentation without destroy the sparse features inside the IHC images. Therefore, it intends to maximize an reward function considering both the representability of feature vectors and the diversity involved by adding noises. Therefore, we propose the reward function as:

$$r(L_r, n) = \frac{\log(1 + \|n\|_2)}{e^{L_r}} \tag{6}$$

where L_r denotes the reconstruction loss and n denotes the generated noise, measuring the representability of feature vectors and adequacy of involved diversity respectively.

Fig. 5 visualized different ways to combine L_r and n. Eq. 6 shows the best mathematical properties. Eq. 6 avoids the tendency to output a high norm of the noise vector or output a zero one to minimize reconstruction loss. Therefore, it is not too sharp or too flat. As a result, Eq. 6 holds the best balance of representability of feature vectors and adequacy of involved diversity.

The noise generator is trained using a policy gradient method (Sutton et al., 2000). Since the noise generator outputs a Gaussian distribution, it has an objective function consisting of the reward functions of all single noise vectors.

$$\begin{aligned} J(\theta) &= E_{s\sim S} E_{n\sim\pi_{\theta,s}(n)}[r(L_r, n)] \\ &= E_{s\sim S} \int \pi_{\theta,s}(n) r(L_r, n) dn \end{aligned} \tag{7}$$

where S denotes the data set of all the original IHC images.

Then, the noise generator is trained to maximize J. The policy gradient is

$$\begin{aligned} \nabla_\theta J &= E_{s\sim S} \int \nabla_\theta \pi_{\theta,s}(n) r(L_r, n) dn \\ &= E_{s\sim S} \int \pi_{\theta,s}(n) \nabla_\theta \log \pi_{\theta,s} r(L_r, n) dn \\ &= E_{s\sim S} E_{n\sim\pi_{\theta,s}}[\nabla_\theta \log \pi_{\theta,s} r(L_r, n)] \end{aligned} \tag{8}$$

For the balance of exploration and stabilization, the standard deviation σ of the Gaussian distribution keeps decaying during the training process. A higher σ encourages exploration at the beginning of the training. At the end of the training, a decayed σ ensures the convergence of the training process and stabilization of the noise generator.

2.1.3. Adversarial training for more accurate feature vectors by making reconstructed images more realistic

Adversarial training is involved to further improve the accuracy of feature vectors produced by SDAE. Since IHC images have some salient patterns, the ability to reconstruct clearer and more recognizable images measures the representability of the feature vectors. VAE networks tend to generate blurry images (Dosovitskiy and Brox, 2016), which indicates that some salient patterns are not preserved in feature vectors. Because of the advantage of GAN architecture to produce clearer and realistic images (Larsen et al., 2015), we use SDAE as the generator and involve an additional discriminator to make up a novel GAN architecture,

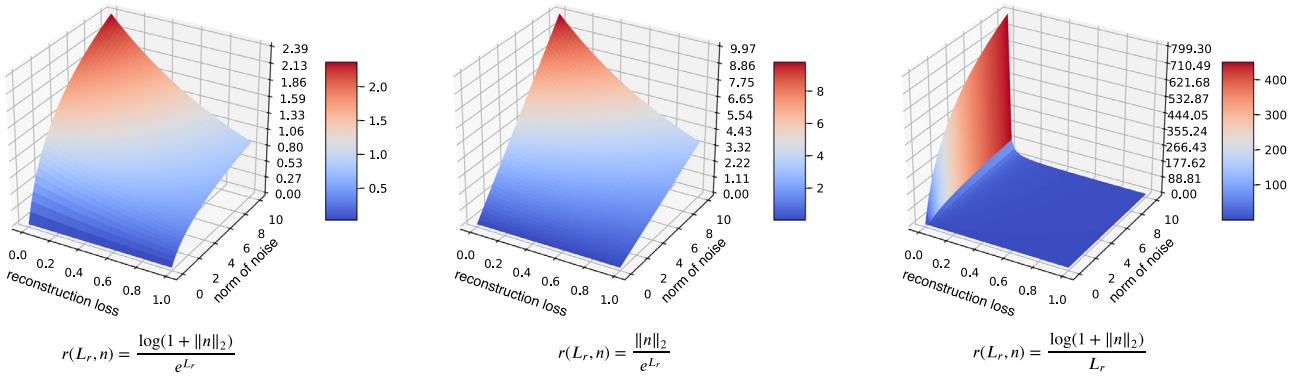

Fig. 5. Eq. 6 is the most applicable one among the three reward functions combining the norm of noises and the reconstruction loss. It holds the best balance of accurate image reconstruction and involving diversity to enhance the robustness of the classification model. In the second reward function, the norm of noises has overwhelming importance than the reconstruction loss. In the last one, the reward becomes huge when reconstruction losses come to zero, causing a tendency to add no noises.

SDAE-GAN. SDAE-GAN has the advantage to produce more accurate feature vectors by improving the clarity of generated images. Suppose G is the generator and D is the discriminator. For the training sample x, we got the binary cross-entropy:

$$L_{GAN}(x) = \log(D(x)) + \log(1 - D(G(x))) \tag{9}$$

After training and fine-tuning the SDAE, $L_r(x, \tilde{x})$ in Eq. 4 and $L_{GAN}(x)$ are used together to optimize the whole SDAE-GAN. $L_r(x, \tilde{x}) + L_{GAN}(x)$ is minimized by the generator and $L_{GAN}(x)$ is maximized by the discriminator. Finally, output of the encoder is feature vectors.

2.2. DenseNet for survival prediction

Finally, a DenseNet is used to predict patients' five-year survival using feature vectors and clinicopathological features. It connects each layer to every other layer in a feed-forward fashion, which strengthens feature propagation, encourages feature reuse, and reduces the number of parameters (Huang et al., 2017). Also, the dense connection scheme has a regularizing effect to reduces overfitting on tasks with a smaller size of training sets (Huang et al., 2017). This enables the dense connection scheme to be a suitable choice for medical applications. For clinicopathological features, we cleaned and normalized them as preprocessing. After being compressed by SDAE-GAN, there is still a feature dimension imbalance between feature vectors and patients' clinicopathological features. To solve this problem, DenseNet takes two parts of data into two different dense blocks respectively before concatenating them. The architecture of our dense net is shown in Fig. 6.

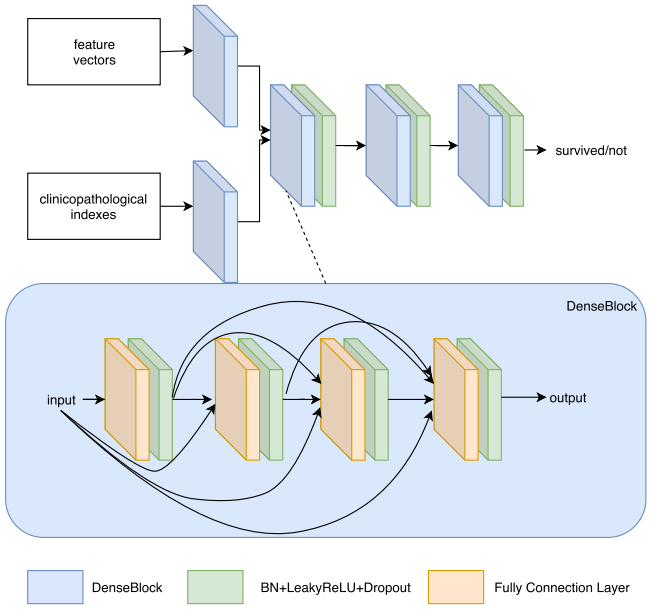

Fig. 6. Architecture of our dense net for survival prediction. Two parts of the input are passed into two dense blocks and concatenated together to predict patients' five-year survival.

3. Experimental results

Experimental results indicate that our proposed ICSPM succeeded in joining high-dimensional IHC images and clinicopathological features for a more accurate five-year survival prediction than using each of them separately. Also, ICSPM outperforms existing models on cancer liver survival prediction using only clinicopathological features. It got an AUC score at 0.73 and an accuracy of 0.72. For sensitivity and specificity, it earned 0.75 and 0.66 respectively.

3.1. Dataset and experiment settings

The study consists of data from 111 patients. Seventy-four of the patients were alive five years after treatment while others were not. For the lack of samples, we used leave-one-out cross-validation on all experiments to utilize most information. Our experiments consist of two parts: 1) Experiments on the whole ICSPM to predict patients' five-year survival. We joined the feature vectors and clinicopathological features to train a DenseNet to show it has higher accuracy, specificity, sensitivity, f1-score, and AUC score than existing models. In addition, we performed experiments to evaluate whether IHC images have useful information for survival prediction. 2) We also did experiments on SDAE-GAN to compress and reconstruct images to show its advantage against existing models. The quality of its reconstructed images measures the performance of SDAE-GAN.

To evaluate the performance of ICSPM on joining record and pathological images and predicting patients' survival, we compared it with three other models. We trained a CART and an ANN model (Chen et al., 2011) for survival prediction based on only clinicopathological features. We also implemented a DenseNet using the existing method (Levine et al., 2018) to join images and scalars in one neural network. In addition, to evaluate the information contained in feature vectors, we tested our model only using them.

Table 2
Detailed architecture of SDAE-GAN.

layer name	output size	
input	200x200x4	
encoder_1	100x100x16	[3x3 conv, stride=1, dilation_rate=(2,2)] [3x3 conv, stride=2] [3x3 conv, stride=1]x2
encoder_2	50x50x32	[3x3 conv, stride=1, dilation_rate=(2,2)] [3x3 conv, stride=2] [3x3 conv, stride=1]x2
encoder_3	25x25x64	[3x3 conv, stride=1, dilation_rate=(2,2)] [3x3 conv, stride=2] [3x3 conv, stride=1]x2
feature vector	25x25x64	[flatten] [fully connection 128 as feature vector] [**add noise to feature vector**] [fully connection 40000]
decoder_1	50x50x32	[reshape] [3x3 deconv, strides=2] [3x3 deconv]
decoder_2	100x100x16	[3x3 deconv, strides=2] [3x3 deconv]
decoder_3	200x200x4	[3x3 deconv, strides=2] [3x3 deconv]
discriminator_layer_1	100x100x8	[5x5 conv, strides=1, dilation_rate=(2, 2)] [batch normalization] [2x2 max pooling]
discriminator_layer_2	50x50x16	[3x3 conv, strides=1, dilation_rate=(2, 2)] [batch normailization] [2x2 max pooling]
discriminator_layer_3	25x25x32	[3x3 conv, strides=1, dilation_rate=(2, 2)] [batch normailization] [2x2 max pooling]
discriminator_layer_4	13x13x32	[3x3 conv, strides=1, dilation_rate=(2, 2)] [batch normailization] [2x2 max pooling]
discriminator_layer_5	7x7x32	[3x3conv, strides=1, dilation_ rate=(2, 2)] [batch normailization] [2x2 max pooling] [flatten]
classify layer	1	[fully connection 1024] [batch normalization] [fully connection 1]

For SDAE-GAN, we compared its reconstructed images with other models (Larsen et al., 2015; Masci et al., 2011) and plotted the reconstruction loss curve of them to show a numeric comparison. Results show that ICSPM has a higher accuracy and AUC score than existing models, and SDAE-GAN shows better reconstruction results, which indicates it produces a more accurate presentation of images in 1-dimension vectors.

We performed all the experiments using Tensorflow 1.13 on Ubuntu 16.04 with Intel i5-7500, 16GB RAM and Nvidia 1080Ti. It took 10 hours to run the first stage to extract features from images. Using SDAE-GAN to extract features does not need a leave-one-out cross-validation since it is unsupervised learning. It took 130 min to perform the leave-one-out cross-validation of survival prediction. Because the DenseNet for survival prediction is not very complicated and we did not run too many epochs, it took about one minute to perform each test.

The detailed implementation of SDAE-GAN for image compression and DenseNet for survival prediction is shown in Tables 2 and 3. On SDAE-GAN, we use a batch size at 64, learning rate at 0.0003, leaky ReLU as activation function and a decay rate at 0.999 to train the noise generator. On DenseNet, we use Adadelta (Zeiler, 2012) as the optimizer, drop out rate at 0.25, learning rate at 0.01 and binary cross-entropy as the loss function.

3.2. Experiments on survival prediction

Joining pathological indexes and clinicopathological features showing the superior performance of the state-of-art method

First of all, we performed experiments to prove our assumptions: 1) Feature vectors compressed from high-dimensional images contain discriminative information about survival status. 2) Joining high-dimensional IHC images and clinicopathological features reaches a higher accuracy of survival prediction than using each of them and 3) SDAE-GAN has advantages than traditional CAE in extracting features from high-dimensional IHC images. The result shown in Fig. 7 indicates that feature vectors are truly a representative compression of original high-dimensional IHC images and have useful information for survival prediction. In addition, joining IHC images and clinicopathological features reaches a higher score than using each of them separately. It proves our hypothesis that they have discriminative information for survival prediction in complementary to each other.

Table 3
Detailedarchitecture of dense net.

layer name	output	feature vector	pathological feature
input		3584	1000 (repeated from 10)
dense_block_1	512	[fc layer 512, batch normalization, leaky ReLU] x 3 [dropout 0.25]	[fc layer 512, batch normalization, leaky ReLU] x 3 [dropout 0.25]
concatenate_1	1024	concatenate	
fully_connected_1	512	[fully connected layer 512, batch normalization, leaky ReLU] [dropout 0.25]	
dense_block_2	256	[fully connected layer 256, batch normalization, leaky ReLU] x 3 [dropout 0.25]	
dense_block_3	128	[fully connected layer 128, batch normalization, leaky ReLU] x 3 [dropout 0.25]	
dense_block_4	64	[fully connected layer 64, batch normalization, leaky ReLU] x 3 [dropout 0.25]	
fully_connected_2	32	[fully connected layer 32, batch normalization, leaky ReLU] [dropout 0.25]	
fully_connected_3	1	[fully connected layer 1, batch normalization, leaky ReLU] [dropout 0.25]	

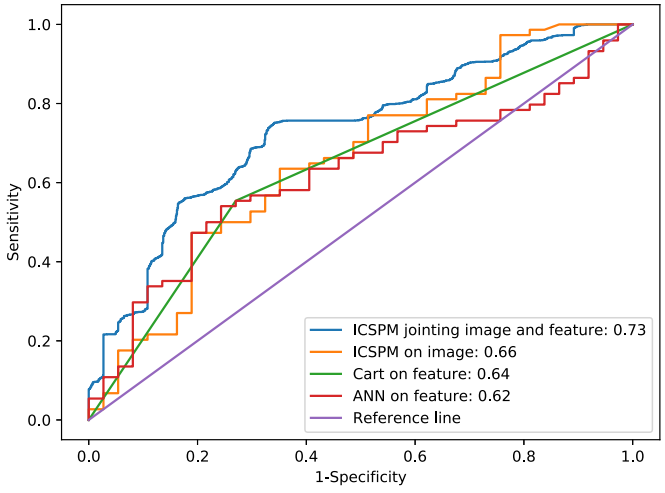

Fig. 7. Joining images and features makes a more accurate survival prediction than using each of them separately. Using each of the two parts of data has a higher score than the baseline, indicating that they both have useful information for survival prediction. In addition, joining them achieves a higher AUC score verifies the complementarity between them.

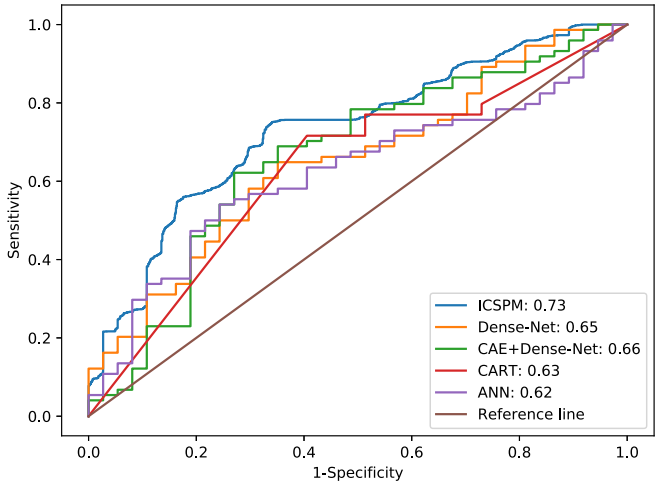

Fig. 8. ROC curves and AUC scores of different models indicate that ICSPM outperforms existing models. Meanwhile, Dense-Net joining images and record has a higher AUC score than CART or ANN which only used clinicopathological features.

Comparison of ICSPM and existing models showing that ICSPM outperforms them in the accuracy of five-year survival prediction ICSPM earned an AUC score at 0.73, an accuracy at 0.72 and a sensitivity at 0.74, outperforming existing models. Results show that ICSPM has better AUC score and accuracy than the existing way to join images and scalars (Levine et al., 2018) or only using regular clinicopathological features (Chen et al., 2011). The comparison of these models is shown in Fig. 8 and Table 4.

Experiments illustrating benefits of the additional noise generator and GAN architecture Also, we did some experiments to indicate that the additional noise generator and the added discriminator to SDAE contribute to survival prediction. We used three meth-

Table 4
Performance of different models.

Model	Accuracy	AUC score	Sensitivity	Specificity	F1-score
ICSPM	0.72	0.73	0.75	0.66	0.72
ICSPM(feature vectors)	0.65	0.63	0.64	0.62	0.64
ICSPM(traditional DA)	0.65	0.65	0.65	0.64	0.66
ICSPM(no DA)	0.69	0.71	0.74	0.64	0.71
ICSPM(no GAN)	0.68	0.68	0.70	0.65	0.69
CART	0.65	0.63	0.71	0.54	0.66
ANN	0.60	0.62	0.52	0.75	0.61
DenseNet	0.65	0.65	0.63	0.64	0.65
CAE+DenseNet	0.67	0.66	0.65	0.65	0.66

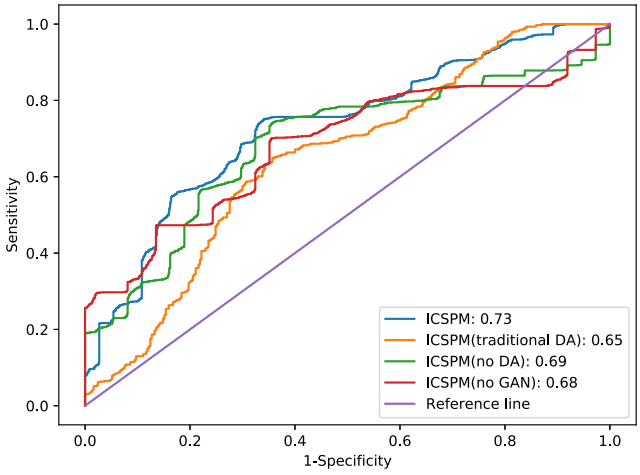

Fig. 9. AUC scores indicate the benefits of our proposed noise generator and additional discriminator. AUC scores decline without the noise generator or the GAN architecture. Using traditional data augmentation methods shows a low AUC score since it breaks the patterns inside high-dimensional IHC images.

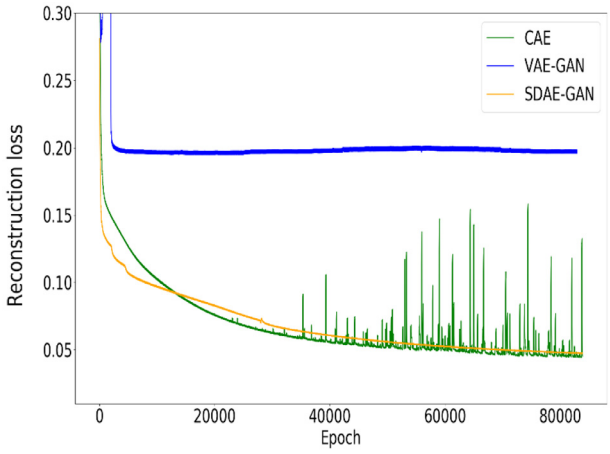

Fig. 10. SDAE-GAN has the best loss curve. VAE-GAN has a very high loss and keeps oscillating while CAE shows a high instability at the end of training.

ods to perform the traditional data augmentation: 1) We cropped images from each side by 0 to 8 pixels (randomly chosen); 2) We flipped the images horizontally with a probability of 0.3. and 3) We added Gaussian noises to blur images with a standard deviation of 0.03. Involving the noise generator and GAN architecture reaches a higher AUC score, shown in Fig. 9. It indicates that our SDAE-GAN with the noise generator produces more representative and accurate feature vectors. As for tradition data augmentation, it makes a decline on AUC score since it breaks the weak patterns inside high-dimensional IHC images.

3.3. Results on image reconstruction

As can be seen in Fig. 11, SDAE-GAN preserved all the salient patterns in original IHC images, which indicates that it retained the essential information in them. Because SDAE-GAN produced clearer

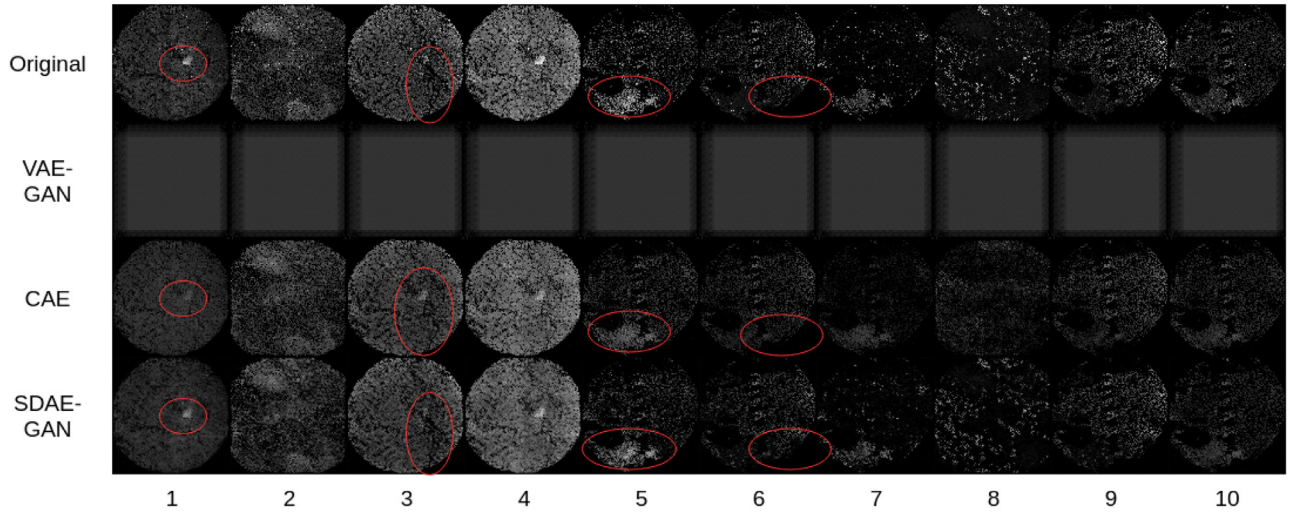

Fig. 11. Reconstruction results of the three models. SDAE showed the best reconstructed images. Traditional VAE-GAN could not reconstruct the original images. In comparison with CAE, SDAE-GAN produced clearer and sharper images and preserved more details in them. In detail, main differences of reconstructed images are marked on Image 1, 3, 5 and 6. For Image 7 and 8, SDAE-GAN preserved some details which were lost by CAE.

images than the other two models, its generated feature vectors are more accurate and representative.

Fig. 10 shows the reconstruction losses of the three models, CAE, VAE-GAN and SDAE-GAN. SDAE-GAN has the most stable training process, although it cannot reach as low loss as CAE because it does not use reconstruction loss as its loss for minimization only. Traditional VAE-GAN's performance shows that it is not suitable for high-dimensional IHC images. This inability is because VAE-GAN tries to minimize the loss of hidden layers representing feature maps. However, IHC images have different patterns against regular medical images directly observed under a microscope or generated by CT or MRI. This difference shortens traditional convolution layers' ability to capture these patterns inside IHC images.

4. Conclusion

For the first time, we combined high-dimensional IHC images with patients' clinicopathological features in liver cancer survival prediction by a two-stage model, ICSPM. The result of this integration proved that high-dimensional these two parts of data have information about cancer in complementary to each other. In future work, our method will be extended to other medical image applications to join images and scalars.

Declaration of Competing Interest

Conflicts of interest: none

Acknowledgments

This work was partially supported by the National Natural Science Foundation of China (NSFC) (grant no. 61672552 and U1611461).

Appendix A. Correlation analysis

We performed the correlation analysis to prove that pathological indexes and clinicopathological features are complementary to each other in survival prediction. Fig. 12 shows Pearson correlation coefficients between pathological indexes, clinicopathological features and patients' survival. Fig. 13 shows the coefficient matrix of pathological indexes and clinicopathological features. Pearson correlation coefficients in Fig. 12 and Fig. 13 indicate that high-dimensional IHC images and traditionally used clinicopathological features are both contributing to survival prediction and complementary to each other.

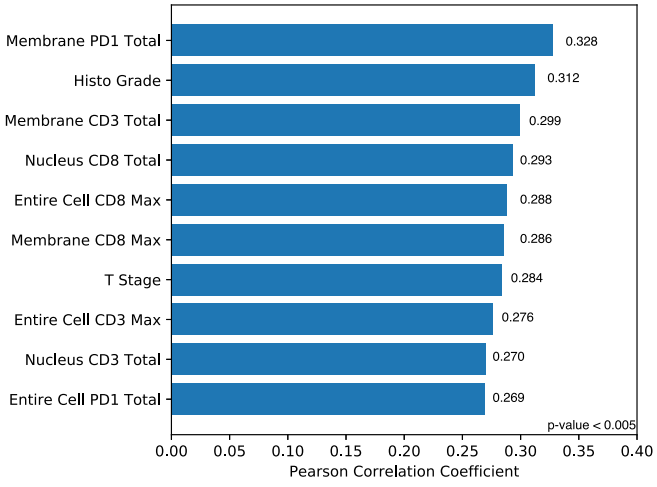

Fig. 12. Pearson correlation coefficients between the ten most contributing indexes or features and survival status. Pathological indexes show a similar contribution(eight out of ten) to survival prediction with regular clinicopathological features(two out of ten, Histo Grade and T Stage).

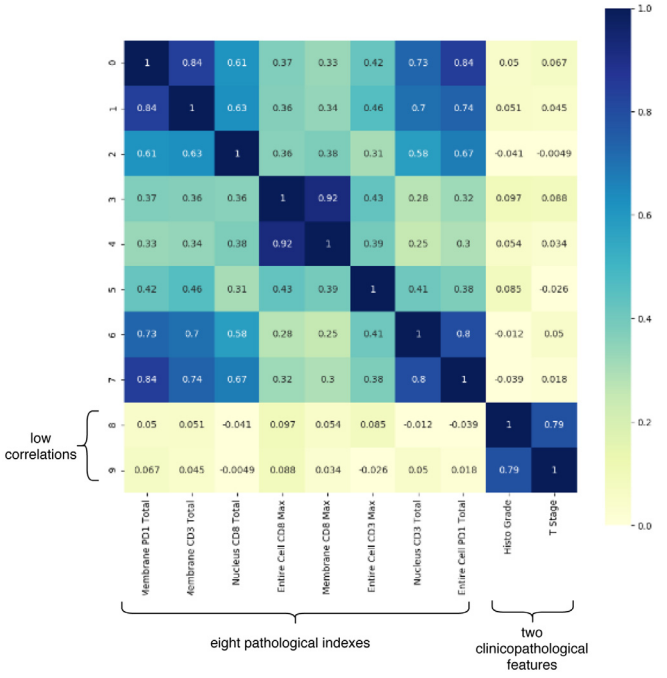

Fig. 13. Pearson correlation coefficients of pairs of ten most contributing indexes or features. Both pathological indexes and regular clinicopathological features have high coefficients with patients' survival. Besides, they have low coefficients among themselves. These two facts strongly suggest the complementarity of pathological indexes and regular clinicopathological features in survival prediction.

The details are as follows: 1) Fig. 12 shows that they have similar correlation coefficients with a five-year survival status. Eight out of ten most contributing indexes or features are pathological indexes represented by IHC images. The other two (Histo Grade and T Stage) are regular clinicopathological features. They all have Pearson correlation coefficients about 0.3 with p-values under 0.005, which can be regarded as weak correlations. Since the period of five years is long in survival prediction, features having weak correlations with survival status are still beneficial. 2) Correlation coefficient matrix in Fig. 13 verifies the complementarity between pathological indexes and clinicopathological features. Two clinicopathological features (Histo Grade and T Stage) have very low correlation coefficients with eight pathological indexes. This fact indicates that features and indexes are complementary instead of just redundant to each other.

References

Aziz, S., Fackler, O.T., Meyerhans, A., Müller-Lantzsch, N., Zeitz, M., Schneider, T., 2005. Replication of m-tropic hiv-1 in activated human intestinal lamina propria lymphocytes is the main reason for increased virus load in the intestinal mucosa. JAIDS J. Acq. Immune Def.Syndr. 38 (1), 23–30.

Binnewies, M., Mujal, A.M., Pollack, J.L., Combes, A.J., Hardison, E.A., Barry, K.C., Tsui, J., Ruhland, M.K., Kersten, K., Abushawish, M.A., et al., 2019. Unleashing type-2 dendritic cells to drive protective antitumor cd4+ t cell immunity. Cell 177 (3), 556–571.

Broz, M.L., Binnewies, M., Boldajipour, B., Nelson, A.E., Pollack, J.L., Erle, D.J., Barczak, A., Rosenblum, M.D., Daud, A., Barber, D.L., et al., 2014. Dissecting the tumor myeloid compartment reveals rare activating antigen-presenting cells critical for t cell immunity. Cancer Cell 26 (5), 638–652.

Carpenter, A.E., Jones, T.R., Lamprecht, M.R., Clarke, C., Kang, I.H., Friman, O., Guertin, D.A., Chang, J.H., Lindquist, R.A., Moffat, J., Others, 2006. Cellprofiler: image analysis software for identifying and quantifying cell phenotypes. Genome Biol. 7 (10), R100.

Chen, C.-M., Hsu, C.-Y., Chiu, H.-W., Rau, H.-H., 2011. Prediction of survival in patients with liver cancer using artificial neural networks and classification and regression trees. In: 2011 Seventh International Conference on Natural Computation, 2. IEEE, pp. 811–815.

Dosovitskiy, A., Brox, T., 2016. Generating images with perceptual similarity metrics based on deep networks. In: Advances in Neural Information Processing Systems, pp. 658–666.
Filmus, J., Capurro, M., Rast, J., 2008. Glypicans. Genome Biol. 9 (5), 224.
Gao, G.F., Jakobsen, B.K., 2000. Molecular interactions of coreceptor cd8 and mhc class i: the molecular basis for functional coordination with the t-cell receptor. Immunol. Today 21 (12), 630–636.
Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4700–4708.
Johansson-Lindbom, B., Svensson, M., Pabst, O., Palmqvist, C., Marquez, G., Förster, R., Agace, W.W., 2005. Functional specialization of gut cd103+ dendritic cells in the regulation of tissue-selective t cell homing. J. Exp. Med. 202 (8), 1063–1073.
Kingma, D.P., Welling, M., 2013. Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114.
Larsen, A.B.L., Sønderby, S.K., Larochelle, H., Winther, O., 2015. Autoencoding beyond pixels using a learned similarity metric. arXiv preprint arXiv:1512.09300.
Lehmann, J., Huehn, J., de la Rosa, M., Maszyna, F., Kretschmer, U., Krenn, V., Brunner, M., Scheffold, A., Hamann, A., 2002. Expression of the integrin $\alpha e \beta 7$ identifies unique subsets of cd25+ as well as cd25- regulatory t cells. Proc. Natl. Acad. Sci. 99 (20), 13031–13036.
Levine, S., Pastor, P., Krizhevsky, A., Ibarz, J., Quillen, D., 2018. Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection. Int. J. Rob. Res. 37 (4–5), 421–436.
Li, N., Fu, H., Hewitt, S.M., Dimitrov, D.S., Ho, M., 2017. Therapeutically targeting glypican-2 via single-domain antibody-based chimeric antigen receptors and immunotoxins in neuroblastoma. Proc. Natl. Acad. Sci. 114 (32), E6623–E6631.
Masci, J., Meier, U., Cire\csan, D., Schmidhuber, J., 2011. Stacked convolutional auto-encoders for hierarchical feature extraction. In: International Conference on Artificial Neural Networks. Springer, pp. 52–59.
Sutton, R.S., McAllester, D.A., Singh, S.P., Mansour, Y., 2000. Policy gradient methods for reinforcement learning with function approximation. In: Advances in neural information processing systems, pp. 1057–1063.
Syn, N.L., Teng, M.W., Mok, T.S., Soo, R.A., 2017. De-novo and acquired resistance to immune checkpoint targeting. Lancet Oncol. 18 (12), e731–e741.
Vanneschi, L., Farinaccio, A., Mauri, G., Antoniotti, M., Provero, P., Giacobini, M., 2011. A comparison of machine learning techniques for survival prediction in breast cancer. BioData Min. 4 (1), 12.
Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., Manzagol, P.-A., 2010. Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. J. Mach. Learn. Res. 11 (Dec), 3371–3408.
Weinberg, R.A., 2013. The Biology of Cancer: Second International Student Edition. WW Norton & Company.
Xue, W., Islam, A., Bhaduri, M., Li, S., 2017. Direct multitype cardiac indices estimation via joint representation and regression learning. IEEE Trans. Med. Imag. 36 (10), 2057–2067.
Xue, W., Lum, A., Mercado, A., Landis, M., Warrington, J., Li, S., 2017. Full quantification of left ventricle via deep multitask learning network respecting intra-and inter-task relatedness. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 276–284.
Xue, W., Nachum, I.B., Pandey, S., Warrington, J., Leung, S., Li, S., 2017. Direct estimation of regional wall thicknesses via residual recurrent neural network. In: International Conference on Information Processing in Medical Imaging. Springer, pp. 505–516.
Yu, F., Koltun, V., 2015. Multi-scale context aggregation by dilated convolutions. arXiv preprint arXiv:1511.07122.
Zeiler, M.D., 2012. ADADELTA: An adaptive learning rate method. arXiv preprint arXiv:1212.5701.
Zhen, X., Wang, Z., Yu, M., Li, S., 2015. Supervised descriptor learning for multi-output regression. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1211–1218.
Zhu, X., Yao, J., Luo, X., Xiao, G., Xie, Y., Gazdar, A., Huang, J., 2016. Lung cancer survival prediction from pathological images and genetic data An integration study. In: 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI). IEEE, pp. 1173–1176.