2017 International Conference on Security, Pattern Analysis,
and Cybernetics (SPAC)

Cancer Classification with Multi-task Deep Learning

Qing Liao†, Lin Jiang†, Xuan Wang†, Chunkai Zhang†, Ye Ding‡
† *Department of Computer Science and Technology, Harbin Institute of Technology (Shenzhen)*
‡ *Guangzhou HKUST Fok Ying Tung Research Institute, The Hong Kong University of Science and Technology*
liaoqing@hit.edu, zoeljiang@gmail.com, wangxuan@cs.hitsz.edu.cn, ckzhang812@gmail.com, yeding@ust.hk

***Abstract*—Microarray technique can generate a large amount of gene expression profiles for thousands of genes simultaneously. The gene expression data has been widely used in disease diagnosis and deep learning approach has achieved great successes in this task. However, the deep learning approach may fail when the expression data for a particular tumor is insufficient for training an effective model. In this paper, we propose a novel multi-task deep learning (MTDL) to overcome the aforementioned deficiency by leveraging the knowledge among multiple expression data of related cancers. MTDL learns local features from each task with some private neurons, and learns shared features for all tasks simultaneously with some shared neurons, and learns to inference for each task separately in the end layer. Since MTDL leverages the expression data of multiple cancers, it can learn more stable representation for each cancer even its expression profiles are inadequate. The experimental results show that MTDL significantly improves the performance of diagnosing each type of cancer when it jointly learns from the expression data of twelve cancer datasets.**

***Index Terms*—Multi-task learning, Deep learning, Cancer diagnosis,**

I. Introduction

With the rapid development of microarray technologies, it becomes possible to monitor the expression levels of tens of thousands genes in parallel. The gene expression data have been widely used to classify tumors as it greatly saves both expenses and time, and have been used to delineate the disease pathology as its ability to precisely detect the physiological process. However, as the progress of genomic research, the volume of gene expression data has continuously outstripped the ability of human beings to analyze it. Towards this end, the machine learning methods have been successfully applied to analyze gene expression data and automatically classify tumors.

In the past decades, several machine learning algorithms have been applied to classify different types of cancers by microarray gene expression data. The earliest machine learning methods is decision tree (DT) [1] which uses distinctive sequence features of known diseases proteins compared to all human proteins. K-nearest neighbor (k-NN) classifier and Bayesian classifier (NB) [2] are also used to identify human diseases by classifying multiple types of genomic. In this direction, Bharathi [3] used Analysis of Variance (ANOVA) rank scheme on important genes and tested the classification capability using Support Vector Machine (SVM). Hu *et al.* [4] proposed Maximally Diversified Multiple Trees (MDMT) algorithm which focuses on ensemble a set of unique trees in the decision committee. Halder *et al.* [5] proposed active learning based on fuzzy k-nearest neighbor (ALFKNN) which first applies unlabeled samples to get the labels from experts, then the labeled "informative samples" can be iteratively added to the training samples to improve the prediction accuracy. Begum *et al.* [6] incorporated AdaBoost and linear SVM (ADASVM) as a component classifier, and shown that it has better performance than the state-of-the-art classifiers.

However, the above techniques are difficult to learn a good representation of cancer when the tissue sample of a specific cancer are insufficient. The insufficiency of tissue samples has two situations in the modeling learning process: (1) Samples of rare cancers (some only occurring a few people in each year) are much difficult to obtain than common cancers. Moreover, the number of patients tissues is much fewer than the normal peoples tissues. The insufficiency problem of cancer samples leads classifiers to have not enough data to learn a stable representation of cancer pattern. (2) Cancer samples from different data sources are hard to integrate which aggravates the insufficiency problem. For example, Leukemia is one of the most common type of cancer in the world which has caused 353,500 deaths in 2015 [7]. Therefore, many research institutions pay their attentions to study the Leukemia for human beings health. However, the data sources from different research institutions cannot be integrated together, because different research institutions have different experimental conditions, and they choose different gene features and even tag different cancer labels. In the tested cancer datasets, there are three Leukemia datasets from three data sources. The first Leukemia dataset classifies Leukemia samples into two types (NPMI1 + and NPMI1 -) while the second Leukemia dataset classifies samples into four types (MP, HDMTX, HDMTX+MP and LDMTX+MP) because the first dataset classifies based on gene point and second dataset classifies based on drug responses in human beings leukemia cells. Moreover, the first Leukemia dataset has 54,675 gene features and the third Leukemia dataset has 54,613 gene features because different research institutions may choose different gene features to investigate the same cancer according to their domain knowledge. Because the above restrictions, the insufficiency problem usually occurs in common cancer types even the cancer has many datasets.

Most traditional classification algorithms cannot achieve satisfactory performance when there are only a few samples. Dimension reduction [8], [9] and feature selection methods [10]–[16] are proposed to indirectly solve the reverse impact of the insufficiency problem in the datasets. Fakoor *et al.*

978-1-5386-3016-7/17/$31.00 ©2017 IEEE

[10] and Liu *et al.* [11] firstly utilized PCA to project gene expression data onto a low-dimensional subspace to relieve the imbalance problem between gene features and samples, and then utilized deep learning method to enhance cancer diagnosis and classification. Moreover, Isabelle *et al.* [16] applied the support vector machine (SVM) to select gene features to improve cancer classification. Relieving imbalance via dimension reduction and feature selection can greatly reduce thousands of features so that the model can choose the most important genes during learning and can indirectly relieve the reverse influence of the insufficiency of samples. However, all existed methods can only indirectly relieve the insufficiency problem. It is still challenging to fully utilize small-scale datasets from different cancers in a model.

In order to directly solve the insufficiency problem, we propose a novel Multi-task Deep Learning (MTDL) algorithm which is inspired by multi-task learning and deep learning. MTDL can not only utilize the small-scale datasets from different cancers simultaneously but also employ the closely related datasets to help learning a better representation and enhance the classification performance. Specifically, MTDL simultaneously learns local features from each task (cancer) and shared features from multiple tasks via deep neural network to incorporate both local features and shared features to boost the performance. Our contributions are summarized as below:

1) We propose a MTDL algorithm which can directly solve the insufficiency problem of data and largely enhance the performance of classification.
2) MTDL can integrate datasets of closely related cancer from various data sources to enhance the classification performance because the hidden representation of the same cancer types are similar.
3) MTDL can simultaneously utilize closely related cancer datasets so that other hidden cancer representation can provide more information to small-scale cancer datasets to enhance the classification performance.

The rest of the paper is organized as follow. Section II briefly reviews the related techniques and Section III presents the multi-task deep learning (MTDL) algorithm. Section IV evaluates the classification performance of MTDL and the representative algorithms on twelves real-world cancer datasets. Finally, the concluding remark is given in Section V.

II. Related Work

This section will briefly reviews multi-task learning (MTL) and its related works. Multi-task learning [17] has been thoroughly proven to improve the generalization performance significantly when there is not enough number of samples to train individual task [18]. Comparing with single task learning, multi-task learning can use related tasks to learn information to help the tasks when the number of samples is rather insufficient. Figure 1 demonstrates the differences between single-task learning model (STL) and multi-task learning model (MTL). In the Figure 1 (a) all four datasets have no connection and each task will be trained separately,

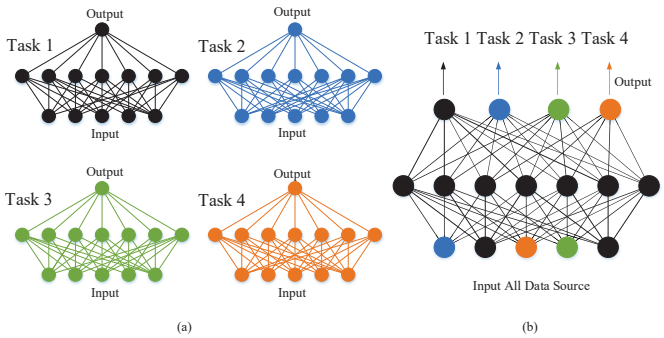

Fig. 1. (a) Single-Task Learning, (b) Multi-Task Learning of four tasks.

because single-task learning assumes that the training samples are drawn independently from a particular distribution. Hence the single-task learning totally ignores the relationship among related tasks. In contrast to single-task learning, the multi-task learning in the Figure 1 (b) assumes that some tasks may be highly correlated which implies that the information learned from one task can be leveraged to another. Hence multi-task learning will jointly train four models. Multi-task learning is not simply pooling all tasks and treating them as a single task due to the strategy of isolating each task. Therefore, multi-task learning uses a shared layer (the second layer in the Figure 1 (b)) to transfer tasks representation and learn the classifiers of each task in parallel to output each tasks classification result. The transferred representation from related tasks is particularly important when the training data of a specialized task only contains a limited amount of samples for learning its classifier.

Recently, multi-task deep learning methods which cooperate multi-task learning and neural network together are shown successfully in computer vision [19]–[25], bioinformatics [26]–[28], climate analytics [29], etc. Zhang *et al.* [19] proposed a tasks-constrained deep convolution network (TCDCN) model to jointly optimal facial landmark detection with a set of related tasks, e.g., head pose estimation and facial attribute inference. Rejeev *et al.* [20] proposed a HyperFace architecture based on CNN which can simultaneously conduct face detection, facial landmark localization, head pose estimation and gender recognition from a given image. The multi-task model consists of two components including pose regression and body-part detection via sliding-window classifiers. Abrar *et al.* [21] proposed a multi-task CNN model to better predict attributes in images, for example, whether wearing necktie or wearing a blue dress, using deep convolutional neural networks (CNN). Tao et al. proposed a deep model based on transfer learning and multi-task learning for biological image analysis [26], [27] on the domain-specific biological images. The model first uses multiple CNN models to pre-train a set of parameters on the ImageNet dataset and then transfers the learned parameters with the mouse brain images. At last, multi-task model is used to output each tasks result. Although these DNN models based on multi-task learning shows its success in computer vision and biomedical image

analysis in the past five years, all the existed methods use deep neural network and multi-task learning model separately. Specifically, all these method applies neural network to learn shared feature maps in the low layers, then splits features and outputs classification performance of each task via multi-task learning in the top layers. To remedy the above deficiencies, we propose a multi-task deep learning (MTDL) algorithm which cooperates multi-task learning in each hidden layers via the deep neural network rather than split these two models in the low and high layers separately. The advantage of MTDL is that cooperating multi-task learning during deep neural network can learn more shared representations than using only multi-task learning in the high layers so that it can boost the classification performance.

III. Multi-task Deep Learning

In this paper, we propose a novel multi-task deep learning (MTDL) algorithm for cancer classification. The structure of the network is shown in Figure 2. The proposed MTDL shares information across different tasks by setting a so-called "shared hidden units". In Figure 2, the red shapes signify shared hidden units of all the task sources in each layer, and the triangle, square and pentagon of rest colors signify local hidden units of each task source in different layers. In this work, we design two hidden layers and one soft-max output layer. It receives n groups of input units and each group corresponds to one task. In the first hidden layer, there are n groups of local hidden units correspond to n task source and a single group of shared units, which receives from all n groups of input units. Similar to the first hidden layer, the second hidden layer contains n groups of local hidden units and one group of shared units. In contrast to the first hidden layer, each group of local hidden units in the second hidden layer receives not only the activations of corresponding local hidden units in the first hidden layer, but also the activations of the shared units in the first hidden layer. The shared hidden units in the second hidden layer receive activations of the whole units in the first hidden layer, including local hidden units and shared hidden units.

Let $x_1, x_2, \cdots, x_n$ denote the inputs of n tasks. For the first hidden layer, the activations of each group of local hidden units a_i^1 are calculated by

$$a_i^1 = \sigma(W_{li}^1 x_i + b_i^1), i = 1, \cdots n \tag{1}$$

where the upscript of a_i^1 denotes the index of layer and the subscript of a_i^1 denotes the index of task source. Activation function is the rectified linear unit (ReLU), i.e., $\sigma(x) = max(0, x)$, and W_{li}^1 is the local weight of edge between the $\#1$ task source and the local hidden units a_i^1. Moreover b_i^1 is the bias for the i-th group of local hidden units. The activations of the first shared hidden units s^1 are calculated by

$$s^1 = \sigma(\sum_{i=1}^{n} W_{si}^1 x_i + b_s^1), \tag{2}$$

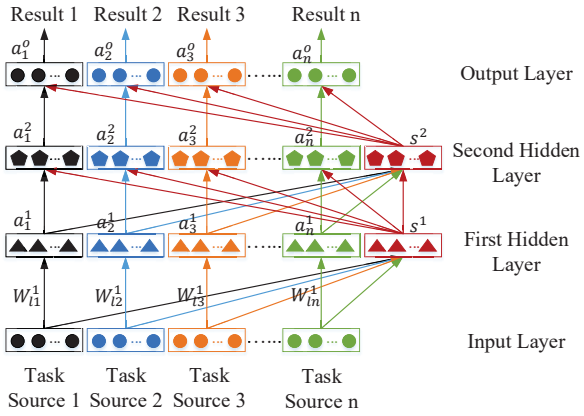

Fig. 2. The proposed multi-task deep neural network structure. In the two hidden layers, the red units denote the shared hidden units and the units of the rest color denote the local hidden units.

where W_{si}^1 is the shared weight of edge between a_i^1 and s^1. The b_s^1 is the bias of the shared hidden units s^1, and the activation function is ReLU.

For the second hidden layer, the activations of each group of local hidden units a_i^2 are calculated by

$$a_i^2 = \sigma(W_{li}^2 a_i^1 + W_{si}^2 s^1 + b_i^2), i = 1, \cdots n \tag{3}$$

where W_{li}^2 is local weight of edge between a_i^1 and a_i^2. The W_{si}^2 is the share weight of edge between all the first local unit and the s^2. The b_i^2 is the bias for the i-th group of local hidden units in the second layer. The activations of the shared hidden units s^2 are calculated by

$$s^2 = \sigma(\sum_{i=1}^{n} W_{si}^2 a_i + W_s^2 s^1 + b_s^2), \tag{4}$$

where W_s^2 is the weight of edge between s^1 and s^2.

For the output layer, the output a_i^o of each task is calculated by

$$a_i^o = sigmoid(W_{li}^o a_i^2 + W_{si}^o s^2 + b_i^o), i = 1, \cdots n, \tag{5}$$

where W_{li}^o is the local weight of edge between a_i^2 and a_i^o. W_{si}^o is the weight of edge between s^2 and a_i^o. Moverover, b_i^o is the bias for the output units of the i-th task, and the activation function is defined as $sigmoid(x) = 1/(1 + e^{(-x)})$.

The advantages of setting local units and the shared units are that each task can learn private representation from its local units perform classification. Different task learns separate private representation, because the local units preserve the feature of each separate task. On the other hand, the shared units learn shared representation from the whole datasets to leverage the information obtained from the microarray system. It is the shared units that boost the performance of each task because they leverage information through all tasks. In summary, MTDL can not only to preserve each tasks local features but also utilize the shared knowledge to provide stable features for all tasks. The following experiments confirm this point.

TABLE I
SUMMARIZATION OF THE USED GENE EXPRESSION DATA.

Cancer	Description	#Features	#Samples	Labels
Task 1	AML [30]	54613	2341	1=AML, 2=MDS
Task 2	Adenocarinoma [31]	34749	193	1=adenocarcinoma, 2=squamous cell carcinoma
Task 3	Breast Cancer [32]	30006	1047	1=non-IBC, 2=IBC
Task 4	Leukemia [33]	54675	2284	1=NPM1+, 2=NPM1-
Task 5	Leukemia [34]	12600	658	1=MP, 2=HDMTX, 3=HDMTX+MP, 4=LDMTX+MP
Task 6	AML [35]	12625	625	1=Complete Remission, 2=Relapse
Task 7	Seminoma [36]	12625	618	1=state I, 2=state II and III
Task 8	Ovarian Cancer [37]	15154	153	1=cancer, 2=normal
Task 9	Colon Cancer [38]	2000	32	1=cancer, 2=non-cancer
Task 10	Medulloblastoma [39]	7129	30	1=class 0, 2=class 1
Task 11	Prostate Cancer [40]	12600	102	1=tumor, 2=normal
Task 12	Leukemia [41]	54613	2389	1=NPM1+, 2=NPM1-

TABLE II
THE ACCURACIES OF CLASSIFYING 12 CANCERS BY USING THE NEURAL NETWORK IN FIGURE 2 WITHOUT THE SHARED HIDDEN LAYERS, I.E., SEPARATELY TRAINING A DNN ON EACH TASK.

Cancer	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8	k=9	k=10	Mean	Std. Dev.	Median
Task 1	0.6996	0.6996	0.747	0.6996	0.6552	0.7552	0.697	0.6496	0.7552	0.6947	0.70527	0.03743	0.6996
Task 2	1	1	1	1	1	1	1	1	1	1	1	0	1
Task 3	1	1	1	1	1	1	1	1	1	1	1	0	1
Task 4	0.5817	0.5817	0.5817	0.5817	0.6234	0.5817	0.5817	0.5817	0.6298	0.5836	0.59087	0.0189	0.5817
Task 5	0.2335	0.2502	0.2335	0.2169	0.2501	0.2335	0.2501	0.2335	0.2169	0.2169	0.23351	0.01357	0.2335
Task 6	0.9	0.9	0.9	0.9	0.8	0.9	0.9	0.9	0.9	0.8	0.88	0.04216	0.9
Task 7	0.7335	0.7668	0.8001	0.8001	0.7001	0.6335	0.6001	0.7001	0.6001	0.7668	0.71012	0.07706	0.7168
Task 8	0.68	0.68	0.68	0.68	0.68	0.68	0.68	0.68	0.68	0.68	0.68	0	0.68
Task 9	0.9667	1	1	1	1	0.9667	0.9667	1	1	1	0.99001	0.01609	1
Task 10	0.9	0.9	0.8667	0.8667	0.8001	0.9333	0.9333	0.9333	0.9	0.9333	0.89667	0.04285	0.9
Task 11	1	1	1	1	1	1	1	1	1	1	1	0	1
Task 12	0.6871	0.6871	0.6871	0.6871	0.6871	0.6871	0.6871	0.6871	0.7176	0.6828	0.68972	0.00989	0.6871

TABLE III
THE ACCURACIES OF CLASSIFYING 12 CANCERS WITH THE SHARED HIDDEN LAYERS BY USING MTDL.

Cancer	k=1	k=2	k=3	k=4	k=5	k=6	k=7	k=8	k=9	k=10	Mean	Std. Dev.	Median
Task 1	1	1	0.9526	1	1	1	1	1	0.95	1	0.99026	0.02054	1
Task 2	1	1	1	1	1	1	1	1	1	1	1	0	1
Task 3	1	1	1	0.9	1	0.95	1	1	1	1	0.985	0.03375	1
Task 4	1	0.9462	0.923	1	0.9462	1	1	1	0.95	1	0.97654	0.03112	1
Task 5	1	0.9333	0.9333	0.8666	1	0.8666	0.95	1	1	0.9333	0.94831	0.05241	0.94165
Task 6	1	1	1	1	1	1	1	1	1	1	1	0	1
Task 7	1	1	1	1	0.95	1	1	1	1	1	0.995	0.01581	1
Task 8	1	1	0.88	0.94	1	0.96	0.96	1	0.99	1	0.973	0.03945	0.995
Task 9	1	0.9667	1	1	1	1	1	1	1	1	0.99667	0.01053	1
Task 10	1	0.9667	1	0.9333	1	1	1	1	1	1	0.99	0.02250	1
Task 11	1	1	1	1	1	1	1	0.9667	0.9333	1	0.99	0.02250	1
Task 12	1	1	0.9696	0.9652	0.9392	1	0.9696	0.9696	1	0.9652	0.97784	0.02103	0.9696

IV. EXPERIMENTS

We evaluate the effectiveness of MTDL comparing with the two basic deep learning methods i.e., deep neural network (DNN) and sparse auto-encoder for twelve cancers with the 10 folds leave-one-out cross-validation. The used gene expression data are summarized in Table I.

There are two characteristics in the Table I. First, we can find that there are three Leukemia datasets (Task 4, 5, 12). The Task 4 and Task 5 cannot be utilized at the same time in a traditional computation model, because they have different Leukemia labels under different domain knowledge. Moreover, we cannot simply integrated Task 4 and Task 5 to enlarge the cancer datasets because the number of features are different (54675 vs. 54613) because different research institute choose different gene features to investigate samples even they study the same cancer. Unlike the traditional model, MTDL can deal with these cancer datasets simultaneously, leverage information among them to learn a better representation of each cancer and output classification result of each cancer, simultaneously. For example, MTDL can classify Task 4 into

TABLE IV
THE ACCURACIES OF CLASSIFYING 12 CANCERS BY USING MTDL, SEPARATE DNN, AND SPARSE AUTO-ENCODER.

	MTDL	Separate DNN	Separate auto-encoder
Task 1	**0.99±0.021**	0.705±0.037	0.744±0.062
Task 2	**1±0**	**1±0**	0.912±0.18
Task 3	0.985±0.034	**1±0**	0.867±0.219
Task 4	**0.977±0.031**	0.591±0.019	0.561±0.024
Task 5	**0.948±0.052**	0.234±0.014	0.468±0.23
Task 6	**1±0**	0.88±0.042	0.817±0.298
Task 7	**0.995±0.016**	0.710±0.077	0.350±0.337
Task 8	**0.973±0.039**	0.68±0	0.755±0.135
Task 9	**0.997±0.011**	0.990±0.016	0.667±0
Task 10	**0.990±0.021**	0.897±0.043	0.667±0
Task 11	0.990±0.021	**1±0**	0.975±0.079
Task 12	**0.978±0.021**	0.690±0.010	0.692±0.108

two labels and classify Task 5 into four labels at the same time. Second, it shows that the number of features distinguished significantly among different tasks, and the number of samples also distinguished significantly among tasks. For example, the AML dataset (Task 1) has 2341 samples and the Medulloblastoma dataset (Task 11) has only 32 samples. MTDL model use multiple shared hidden layers to learn more representation information from 12 cancer datasets, and the representations from the shared hidden layers can help Medulloblastoma dataset to achieve better classification performance.

Table II gives the classification accuracies of twelve cancers by a traditional DNN on the gene expression data of each cancer tissue. We choose the 10-fold leave-one-out method to evaluate DNN, each dataset is divided into ten folds, where nine folds for training and one fold for testing in each trial. k is the index of the fold and each fold runs 10 trials to obtain an averaged accuracy. The DNN model receives twelve cancer datasets separately and output classification results one by one. Since traditional DNN model ignores the similarity information between related cancer datasets and the classification performance will be rather poor if the cancer samples are insufficient. We can find the DNN has very high accuracy results in Task 2, 3, 9, 10 and 11, because there are only two labels in these cancer datasets and these samples are very easy to assign to each labels. By contrary, Leukemia datasets (Task 4, 5 and 12) do not achieve good classification performances comparing with other cancer dataset. There are two reasons: (1) it is still hard to accurately diagnosis Leukemia because the pattern of each Leukemia label are ambiguous, and (2) Task 5 has more labels, i.e., four labels, than other tasks so that its accuracy is much lower than other tasks.

The accuracy results of twelve cancers by sparse auto-encoder of each cancer tissue are similar because the sparse auto-encoder also ignore to use shared knowledge to improve the classification performance. Due to the limited space, we do not depicted the details in paper.

Table III gives the classification accuracies of twelve cancers by MTDL on the gene expression data of each cancer tissue. Unlike DNN and sparse auto-encoder, MTDL processes multiple cancer datasets (tasks) simultaneously so that MTDL can fully take advantages of similar hidden information between all the datasets (cancers) to improve each tasks performance. Table III shows that the accuracies of Leukemia datasets (Task 4, 5 and 12) has largely improvement (more than 20%) compared with DNN and sparse auto-encode, because MTDL can utilized these datasets at the same time to relieve the insufficiency problem of the Leukemia dataset. Moreover, MTDL also achieves more than 20% improvement in the Task 7 and Task 8, because other cancer dataset provides more representation information via shared layer to help Task 7 and Task to learn a better representation and enhance the classification result. At last, MTDL has satisfactory performance on the rest cancer datasets, some datasets have clean pattern to classify so that the classification result of all the three methods are good. However, MTDL still has the highest performance in most datasets because MTDL simultaneously utilizes all the cancer datasets and learning shared representations through the shared layers can enhance the classification performance of most tasks. Table V summarizes the accuracies of classifying twelve cancers by using MTDL, separate DNN, and sparse auto-encoder. The experimental results show that the proposed MTDL algorithm can leverage information among multiple cancer classification tasks without neither pre-training nor loss of information.

Table IV summarizes the accuracies of classifying twelve cancers by using MTDL, separate DNN, and sparse auto-encoder, the experimental results show that the proposed MTDL approach can leverage information among multiple cancer classification tasks without neither pre-training nor dimension reduction.

V. CONCLUSION

The gene expression data plays an important role in precise medicine because it simultaneously measures the expression levels of thousands of genes. However, for cancer classification, the expression data of a particular cancer might be limited. We propose a novel multi-task deep learning method (MTDL) to classify multiple cancers simultaneously and enhance the classification performance of each cancer by leveraging knowledge through shared layers. MTDL method can process multiple datasets without pre-integrate at the same time even if each datasets has different class labels, features or samples. With the help of knowledge transfer, the classification accuracies of twelve cancers with few samples per cancer are significantly improved.

ACKNOWLEDGMENTS.

This work is supported by National Natural Science Foundation of China under grant No.61702134. And this work was also supported in part by Natural Science Foundation of Guangdong Province, China (No.2014A030313697).

REFERENCES

[1] M. B. A. Snousy, H. M. El-Deeb, K. Badran, I. A. A. Khlil, M. B. A. Snousy, H. M. El-Deeb, K. Badran, and I. A. A. Khlil, "Suite of decision tree-based classification algorithms on cancer gene expression data," *Egyptian Informatics Journal*, vol. 12, no. 2, pp. 73–82, 2011.

[2] P. Helman, R. Veroff, S. R. Atlas, and C. Willman, "A bayesian network classification methodology for gene expression data." *Journal of Computational Biology A Journal of Computational Molecular Cell Biology*, vol. 11, no. 4, pp. 581–615, 2004.

[3] A. Bharathi and A. M. Natarajan, "Cancer classification of bioinformatics datausing anova," vol. 2, no. 3, pp. 369–373, 2010.

[4] H. Hu, J. Li, H. Wang, G. Daggard, and M. Shi, "A maximally diversified multiple decision tree algorithm for microarray data classification," 2006, pp. 35–38.

[5] A. Halder, S. Dey, and A. Kumar, *Active Learning Using Fuzzy k-NN for Cancer Classification from Microarray Gene Expression Data*. Springer India, 2015.

[6] S. Begum, D. Chakraborty, and R. Sarkar, "Cancer classification from gene expression based microarray data using svm ensemble," in *International Conference on Condition Assessment Techniques in Electrical Systems*, 2016, pp. 13–16.

[7] T. Vos, C. Allen, M. K. Arora, R. M. Barber, Z. A. Bhutta, and A. C. Brown, "Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 19902015: a systematic analysis for the global burden of disease study 2015," *The Lancet*, vol. 388, no. 10053, pp. 1545–1602, 2016.

[8] N. Guan, D. Tao, Z. Luo, and B. Yuan, "Online nonnegative matrix factorization with robust stochastic approximation," *IEEE Transactions on Neural Networks & Learning Systems*, vol. 23, no. 7, p. 1087, 2012.

[9] N. Guan, D. Tao, Z. Luo, and B. Yuan, "Nenmf: an optimal gradient method for nonnegative matrix factorization," *IEEE Transactions on Signal Processing*, vol. 60, no. 6, pp. 2882–2898, 2012.

[10] R. Fakoor, F. Ladhak, A. Nazi, and M. Huber, "Using deep learning to enhance cancer diagnosis and classification," in *The International Conference on Machine Learning*, 2013.

[11] J. X. Liu, Y. Xu, C. H. Zheng, H. Kong, and Z. H. Lai, "Rpca-based tumor classification using gene expression data," *IEEE/ACM Transactions on Computational Biology & Bioinformatics*, vol. 12, no. 4, pp. 964–970, 2015.

[12] D. Tao, X. Li, X. Wu, and S. J. Maybank, "Geometric mean for subspace selection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 260–274, 2009.

[13] D. Tao, X. Tang, X. Li, and X. Wu, "Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 7, pp. 1088–1099, 2006.

[14] T. Liu and D. Tao, "Classification with noisy labels by importance reweighting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 3, pp. 447–461, 2016.

[15] D. Tao, X. Li, X. Wu, and S. J. Maybank, "General tensor discriminant analysis and gabor features for gait recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 10, pp. 1700–1715, 2007.

[16] I. Guyon, J. Weston, S. D. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Machine Learning*, vol. 46, pp. 389–422, 2002.

[17] T. Liu, D. Tao, M. Song, and S. J. Maybank, "Algorithm-dependent generalization bounds for multi-task learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2016.

[18] R. Caruana, *Multitask learning*. Kluwer Academic Publishers, 1998.

[19] Z. Zhang, P. Luo, C. L. Chen, and X. Tang, "Facial landmark detection by deep multi-task learning," in *European Conference on Computer Vision*, 2014, pp. 94–108.

[20] R. Ranjan, V. M. Patel, and R. Chellappa, "Hyperface: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition," 2016.

[21] A. H. Abdulnabi, G. Wang, J. Lu, and K. Jia, "Multi-task cnn model for attribute prediction," *IEEE Transactions on Multimedia*, vol. 17, no. 11, pp. 1949–1959, 2016.

[22] Q. Yao, X. Jiang, M. Gong, X. You, Y. Liu, and D. Xu, "Efficient group learning with hypergraph partition in multi-task learning," *Pattern Recognition*, pp. 9–16, 2012.

[23] Z. He, X. Li, X. You, D. Tao, and Y. Y. Tang, "Connected component model for multi-object tracking," *IEEE transactions on image processing*, vol. 25, no. 8, pp. 3698–3711, 2016.

[24] X. You, Q. Li, D. Tao, W. Ou, and M. Gong, "Local metric learning for exemplar-based object detection," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 24, no. 8, pp. 1265–1276, 2014.

[25] Z. He, S. Yi, Y.-M. Cheung, X. You, and Y. Y. Tang, "Robust object tracking via key patch sparse representation," *IEEE transactions on cybernetics*, vol. 47, no. 2, pp. 354–364, 2017.

[26] W. Zhang, R. Li, T. Zeng, Q. Sun, S. Kumar, J. Ye, and S. Ji, "Deep model based transfer and multi-task learning for biological image analysis," in *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2015, pp. 1475–1484.

[27] T. Zeng and S. Ji, "Deep convolutional neural networks for multi-instance multi-task learning," in *IEEE International Conference on Data Mining*, 2015, pp. 579–588.

[28] C. Sotiriou, S.-Y. Neo, L. M. McShane, E. L. Korn, P. M. Long, A. Jazaeri, P. Martiat, S. B. Fox, A. L. Harris, and E. T. Liu, "Breast cancer classification and prognosis based on gene expression profiles from a population-based study," *Proceedings of the National Academy of Sciences*, vol. 100, no. 18, pp. 10 393–10 398, 2003.

[29] A. R. Gonçalves, F. J. Von Zuben, and A. Banerjee, "Multi-task sparse structure learning with gaussian copula models," *The Journal of Machine Learning Research*, vol. 17, no. 1, pp. 1205–1234, 2016.

[30] K. I. Mills, A. Kohlmann, P. M. Williams, L. Wieczorek, W.-m. Liu, R. Li, W. Wei, D. T. Bowen, H. Loeffler, J. M. Hernandez *et al.*, "Microarray-based classifiers and prognosis models identify subgroups with distinct clinical outcomes and high risk of aml transformation of myelodysplastic syndrome," *Blood*, vol. 114, no. 5, pp. 1063–1072, 2009.

[31] T. Fujiwara, M. Hiramatsu, T. Isagawa, H. Ninomiya, K. Inamura, S. Ishikawa, M. Ushijima, M. Matsuura, M. H. Jones, and M. Shimane, "Ascl1-coexpression profiling but not single gene expression profiling defines lung adenocarcinomas of neuroendocrine nature with poor prognosis." *Lung Cancer*, vol. 75, no. 1, pp. 119–125, 2012.

[32] W. A. Woodward, S. Krishnamurthy, H. Yamauchi, R. Elzein, O. Dai, E. Kitadai, S. Niwa, M. Cristofanilli, P. Vermeulen, and L. Dirix, "Genomic and expression analysis of microdissected inflammatory breast cancer," *Breast Cancer Research and Treatment*, vol. 138, no. 3, pp. 761–72, 2013.

[33] H. U. Klein, C. Ruckert, A. Kohlmann, L. Bullinger, C. Thiede, T. Haferlach, and M. Dugas, "Quantitative comparison of microarray experiments with published leukemia related gene expression signatures." *BMC Bioinformatics*, vol. 10, no. 1, pp. 1–11, 2009.

[34] M. H. Cheok, W. Yang, C. H. Pui, J. R. Downing, C. Cheng, C. W. Naeve, M. V. Relling, and W. E. Evans, "Treatment-specific changes in gene expression discriminate in vivo drug response in human leukemia cells," *Nature Genetics*, vol. 34, no. 1, p. 85, 2003.

[35] T. Yagi, A. Morimoto, M. Eguchi, S. Hibi, M. Sako, E. Ishii, S. Mizutani, S. Imashuku, M. Ohki, and H. Ichikawa, "Identification of a gene expression signature associated with pediatric aml prognosis," *Blood*, vol. 102, no. 5, p. 1849, 2003.

[36] I. Gashaw, R. Grmmer, L. Klein-Hitpass, O. Dushaj, M. Bergmann, R. Brehm, R. Grobholz, S. Kliesch, T. P. Neuvians, and K. W. Schmid, "Gene signatures of testicular seminoma with emphasis on expression of ets variant gene 4," *Cellular and Molecular Life Sciences Cmls*, vol. 62, no. 19-20, pp. 2359–2368, 2005.

[37] E. F. Petricoin, A. M. Ardekani, B. A. Hitt, P. J. Levine, V. A. Fusaro, S. M. Steinberg, G. B. Mills, C. Simone, D. A. Fishman, and E. C. Kohn, "Use of proteomic patterns in serum to identify ovarian cancer." *Lancet*, vol. 359, no. 9306, pp. 572–7, 2002.

[38] U. Alon, N. Barkai, D. A. Notterman, K. Gish, S. Ybarra, D. Mack, and A. J. Levine, "Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays," *Proceedings of the National Academy of Sciences*, vol. 96, no. 12, pp. 6745–6750, 1999.

[39] S. L. Pomeroy, P. Tamayo, M. Gaasenbeek, L. M. Sturla, M. Angelo, M. E. McLaughlin, J. Y. Kim, L. C. Goumnerova, P. M. Black, C. Lau *et al.*, "Prediction of central nervous system embryonal tumour outcome based on gene expression," *Nature*, vol. 415, no. 6870, pp. 436–442, 2002.

[40] D. Singh, P. G. Febbo, K. Ross, D. G. Jackson, J. Manola, C. Ladd, P. Tamayo, A. A. Renshaw, A. V. D'Amico, J. P. Richie *et al.*, "Gene expression correlates of clinical prostate cancer behavior," *Cancer cell*, vol. 1, no. 2, pp. 203–209, 2002.

[41] R. G. Verhaak, B. J. Wouters, C. A. Erpelinck, S. Abbas, H. B. Beverloo, S. Lugthart, B. Löwenberg, R. Delwel, and P. J. Valk, "Prediction of molecular subtypes in acute myeloid leukemia based on gene expression profiling," *haematologica*, vol. 94, no. 1, pp. 131–134, 2009.