CS6536 – Final Report

Title: Using deep transfer learning to reduce health care inequities between ethnic groups

Table of Contents

Abstract

In biomedical research, artificial intelligence is now frequently applied. Artificial intelligence (AI) is being utilized in medicine to help with automated consultations and clinical judgments. The development of high-quality AI models depends on a large amount of standardized data, and the data skewing among races is becoming one of the most important factors affecting the intelligence of AI models at present. According to recent statistics, about 91 percent of the samples collected by the Cancer Genomics Research Project were Caucasian, followed by 6 percent of Asians, and others[1]. Data distribution biases can affect the predictive accuracy and robustness of AI models, and have profound negative implications on clinical decisions for statistically disadvantaged populations. Therefore, a new machine learning approach to overcome data inequities between ethnic groups is necessary. In this project, we used the combination of deep learning and transfer learning to fine-tune the previous research results. Compared with the previous baseline experiment that only used a single target domain, we took multi-ethnic groups as the target domain of transfer learning and optimized the model. In our work, the model performance has been improved compared with the baseline experiment.

Introduction

In the case of the current COVID-19 pandemic, observations of critical infection and mortality data show that immunization outcomes vary widely among ethnic groups, with significantly higher critical mortality rates among Africans and Hispanics, for example. The same observation applies to cancer patients. With the development of bioinformatics, how to implement personalized medicine for different groups is a potential development direction[2].

The rise of artificial intelligence technology promotes the rapid development of biomedical research. Compared with the subjective decision-making of traditional medical system, the modern medical system based on machine learning algorithms shows higher efficiency and more accurate results in diagnosis and treatment decisions. The development of high-quality artificial intelligence models depends on high-quality data sources. In the field of medical research, the accuracy of the medical decision-making model is limited by the medical data skewness caused by ethnic population differences and data collection biases. The majority of genetic and genomic data is acquired from people of European heritage, according to data from current cancer genomics research initiatives, and the ethnic diversity of the research population has fallen in recent years. This will inevitably lead to the disadvantage of data accumulation for other ethnic groups except white people for a long time, which will affect the accuracy of the medical decision-making model and cause differences in health care. Data skew can be understood as the uneven distribution of the proportions of different ethnic groups in the current medical data set, which is caused on the one hand by population differences and on the other hand by differences between medical systems in various countries.

Currently, the widely adopted machine learning scheme is the mixture learning scheme, that is, the medical data of all races are mixed and directly used in the training and testing of the machine learning model (Figure 1)[3]. Data skew has a significant impact on this method's model performance, and it tends to generate models with lower accuracy for those races with inferior data volumes, resulting in poor model performance among races. Another independent learning method seems to solve the problem of data skew by training models separately for data sets of different races, but the trained models lack connection and cannot integrate the correlation between racial data well, and the performance of models trained for races with inferior data volume is still poor.

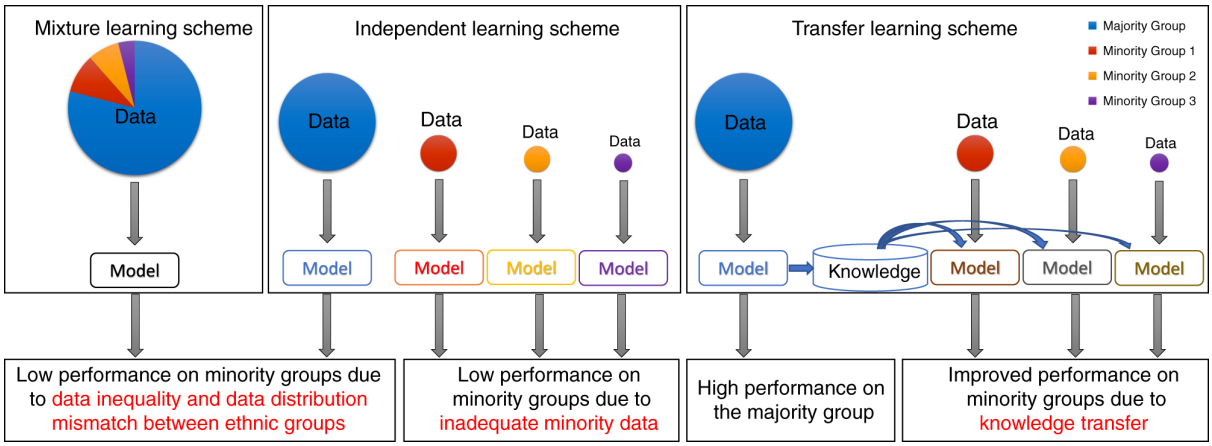

Figure 1. Comparison of machine learning schemes

Based on previous studies, we find that transfer learning can improve the performance of the machine learning model and compensate for the impact of data volume disadvantage. Transfer learning firstly trains the model on the large ethnic dataset, then transfers the knowledge to other small-scale datasets to improve the performance continuously. We believe that our research will provide an inspiration for the research field of multi-ethnic machine learning. By reducing the impact of data distribution differences through transfer learning methods, we can break through the health care differences caused by biomedical data inequality.

Related work

Artificial intelligence helps us learn features from data and make a range of predictions. With enough data, Artificial intelligence can be used more widely. Machine learning, a subset of artificial intelligence, has found widespread application in biomedicine, particularly in cancer diagnosis. Machine learning is based on statistics and probability, enabling computers to pick up features and learn from complex data sets, which makes it more suitable for complex proteome and genomic measurements[4]. It is obvious from examining several

types of machine learning approaches and their effectiveness in cancer prediction that machine learning methods considerably enhance cancer prediction accuracy. Machine learning has considerably increased our investigation and understanding of cancer on a fundamental level. Machine learning also has a subfield called deep learning, which consists of multiple layers. The model may learn the feature representation of incoming data automatically and progress to increasingly advanced and abstract features. As a result, deep learning can be used to evaluate the internal structure of data and extract valuable characteristics while processing complex medical training data[567].

Although deep learning is capable of solving the problem, it is unable to deliver good model performance under the conditions of significant data sample deviation in the multi-ethnic study described in this paper. Traditional machine learning algorithms are best for solving isolated tasks, whereas transfer learning uses knowledge transfer to improve traditional machine learning in multi-source tasks[8] (see from Figure 2[15]). Using data from the target domain to fine-tune the pre-trained model in the source domain is a common approach of using deep neural networks for transfer learning[9]. Transfer learning can be understood as the extension of a learned distribution feature to the unlearned data set. Therefore, transfer learning is suitable for solving problems with multiple domains. If the data set has only a single domain, transfer learning cannot be well applied. Transfer learning has been increasingly significant in the medical industry in recent years. The transfer learning fine-tuning approach based on CNN developed by Zhenghao Shi et al. successfully minimizes the false positive rate of lung cancer in the CAD system[10]. Guillermo et al. 's transfer learning fine-tuning algorithm based on the convolutional neural network also effectively solved the accuracy problem of the tumor prediction model caused by unbalanced genetic data samples[11].

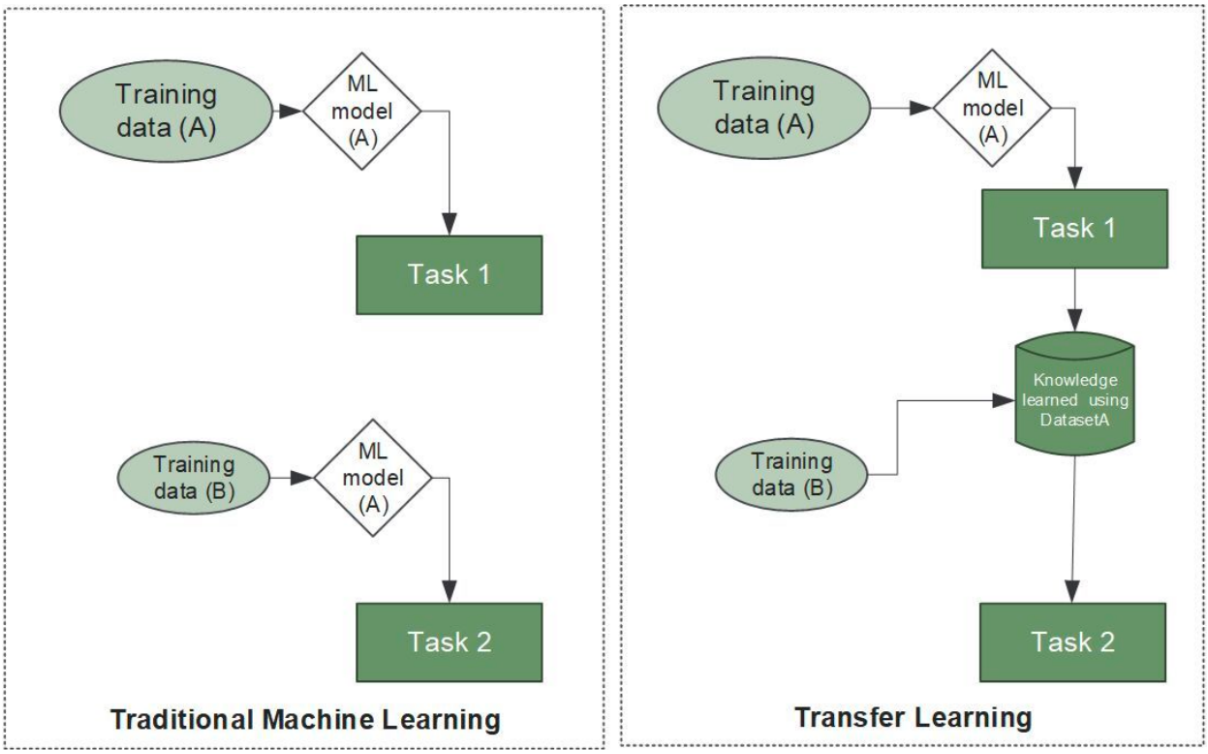

Figure 2. Traditional ML and Transfer Learning

Benchmark experiment

In the research project of this paper, data from different cancer genome projects were collected as the main research data, and the data distribution diagram shows the data skew problem in the cancer data set (see Figure 3).

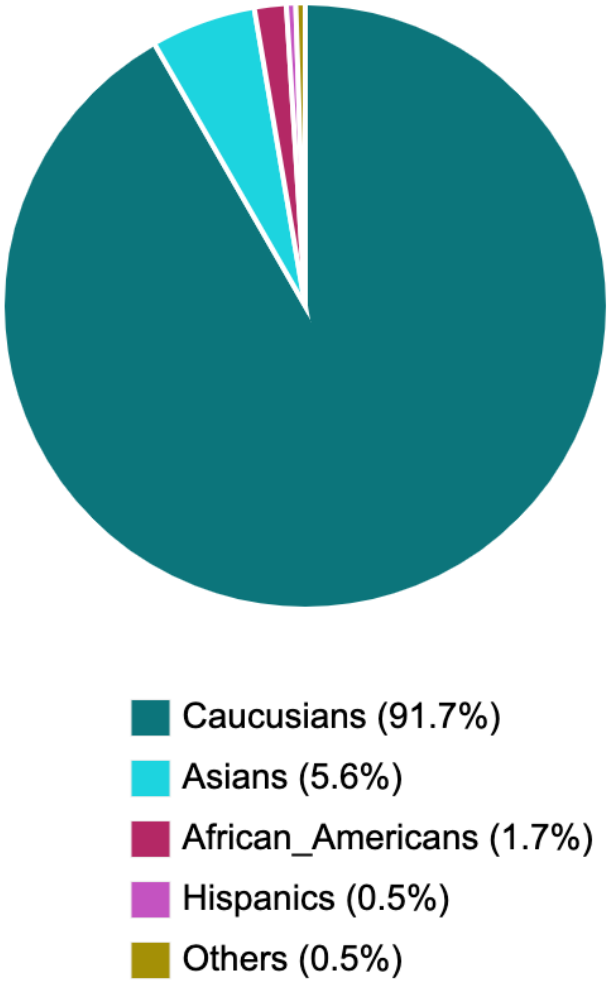

Figure 3. Data skew for different ethnic groups

This data difference between ethnic groups of cancer results in poor model accuracy for ethnic groups with insufficient data when we use machine learning algorithms for training. For example, most of the ethnic data comes from Europeans, but only a small percentage of Hispanics, so the model we train will be much more adaptable to Europeans than to Hispanics, which leads to serious medical differences. In order to solve this problem, the deep transfer learning algorithm proposed by the paper "Deep transfer learning for reducing health care disparities arising from biomedical data inequality" creates a solution. The author first compares the performance of various machine learning algorithms

in source domain and target domain, and finally defines a new domain adaptive method of transfer learning to optimize and adjust the original deep learning model. We will use this as our benchmark experiment and try to improve and innovate on it.

1. Overview

We selected two races AA and EA from TCGA data to assemble machine learning tasks. The following four factors were used to create 1600 machine learning tasks: (1) 40 cancer types (see Table 1); (2) mRNA and protein expression characteristics; (3) 4 clinical outcome endpoints; (4) 5 event time thresholds related with clinical outcome endpoints. As the proportion of group AA is less than 10%, it is necessary to filter out the learning tasks that cannot carry out reliable machine learning due to insufficient AA data. We trained a DNN model for each filtered learning task to use the mixture learning scheme for classification, of which 224 learning tasks achieved good baseline performance (AUROC>0.65). Subsequently, we conducted machine learning experiments for each of the 224 baseline tasks to compare the three learning methods' performance (Table 2).

Cancer category	**Cancer types**
GBMLGG	GBM, LGG
COADREAD	COAD, READ
KIPAN	KIRC, KICH, KIRP
STES	ESCA, STAD
PanGI	COAD, STAD, READ, ESCA
PanGyn	OV, CESC, USC, UCEC
PanSCCs	LUSC, HNSC, ESCA, CESC, BLCA
PanPan	ACC, BLCA, BRCA, CESC, COAD, DLBC, ESCA, GBM, HNSC, KICH, KIRC, KIRP, LAML, LGG, LIHC, LUAD, LUSC, MESO, OV, PAAD, PCPG, PRAD, READ, SARC, SKCM, STAD, TGCT, THCA, THYM, UCEC, UCS, UVM

Table 1. Cancer types

Machine learning scheme	Experiment	Training data ethnic composition	Testing data ethnic composition	AUROC	
				Median	Mean
Mixture learning	Mixture 0	AA+EA	AA+EA	0.71	0.72
	Mixture 1		EA	0.71	0.73
	Mixture 2		AA	0.68	0.67
Independent learning	Independent 1	EA	EA	0.70	0.71
	Independent 2	AA	AA	0.59	0.58
Transfer learning	Transfer learning	EA (source domain) AA (target domain)	AA	0.70	0.69

Table 2. The machine learning experiments

2. Dataset

The cancer genome atlas (TCGA) and MMRF CoMMpass data can be downloaded from GDC (https://gdc.cancer.gov), which contains over 10,000 patient data records, including 80.5 percent of European Americans (EA), 9.3 percent of African Americans (AA), 6.2 percent of East Asian Americans (EAA), and others. The genetic ancestry data can be downloaded from (http://52.25.87.215/TCGAA) and (https://portal.gdc.cancer.gov), which contains about 40 types of cancer and can determine the ethnic groups of patients. As can be seen from these datasets, the data skew between ethnic groups is widespread (Figure 4)[3].

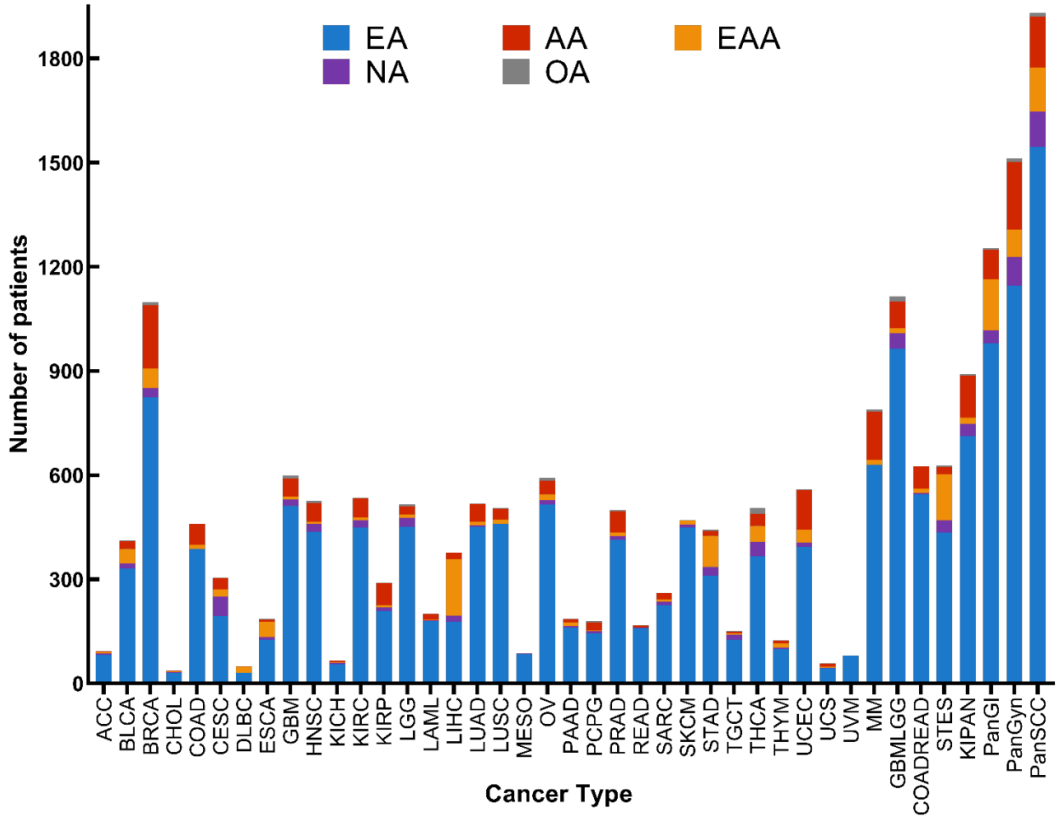

Figure 4. Distribution of ethnic groups for each cancer type

3. Methods

For the TCGA, 189 protein features and 17,176 mRNA features were used as input features for machine learning, while for MMRF CoMMpass, 600 mRNA features which have the most average deviation were selected. In the data processing stage, samples with missing values of more than 20% were deleted and samples lacking genetic ancestry and clinical endpoints were filtered out.

DNN modeling

We built the DNN model using a pyramid architecture (see from Figure 5) consisting of 1 input layer, 4 hidden layers, and 1 logistic regression output layer. The input layer is responsible for processing mRNA and protein features

and has about 389 nodes. A 128-node full connection layer, a dropout layer, a 64-node full connection layer, and a dropout layer make up the hidden layer. To avoid forward propagation of training error caused by the gradient disappearance problem produced by the classic sigmoid function, we adopt the ReLU function f(x) = max(0, x) and set dropout_value = 0.5 for each dropout layer to reduce collinearity. Meanwhile, we increased the batch size to 20 to speed up the training process. To fit the model, stochastic gradient descent with a learning rate LR of 0.01 was used. We set max_iter to 100 and lr_decay to 0.03. Both regularization items $\lambda 1$ and $\lambda 2$ are set to 0.001. We use logistic regression to model conditional distributions. For the computation of regression parameters of source domain and target domain, we fit the following two logistic regression models for AA and EA groups:

$$Y^{AA} = 1/(1 + e^{-\beta^{AA} \cdot X^{AA}})$$
$$Y^{EA} = 1/(1 + e^{-\beta^{EA} \cdot X^{EA}})$$

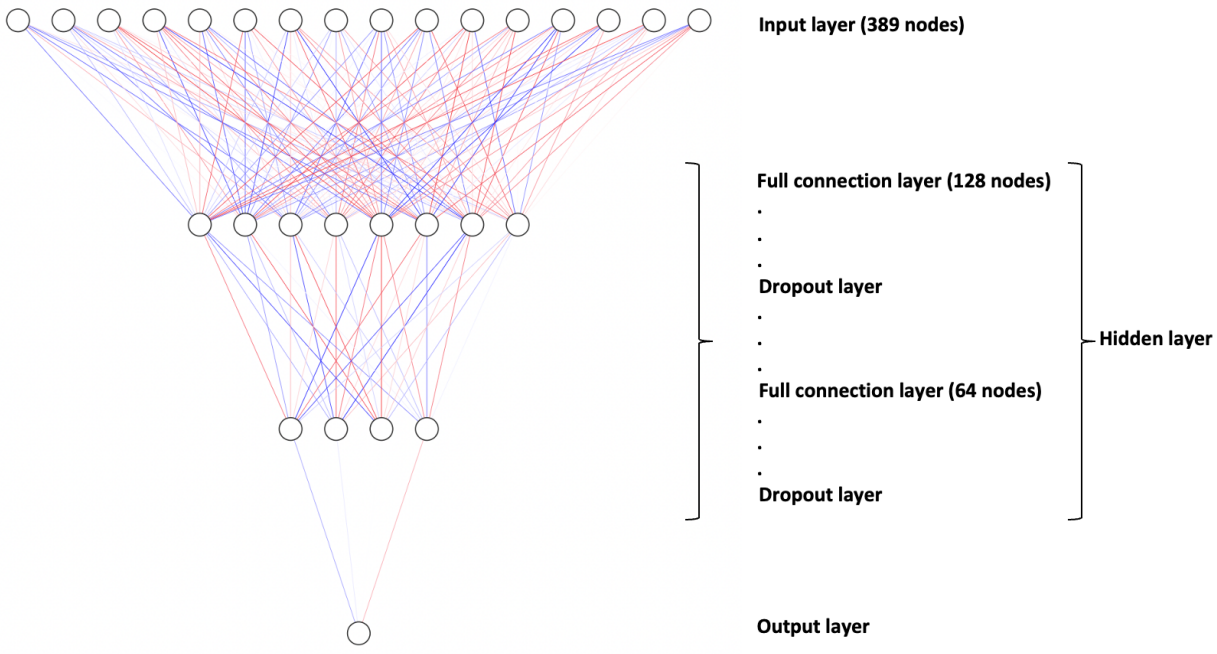

Figure 5. DNN model Structure

Transfer learning

Assume that the source domain is EA group and the target domain is AA group. We proposed 2 fine-tuning algorithms and 1 domain adaptation algorithm, and use AUROC as a performance indicator to evaluate these three transfer learning schemes:

1) Fine-tuning algorithm 1

First, a DNN model with the pyramid structure mentioned earlier is trained using the source domain data. After obtaining the preliminary training model, we then fine-tune the DNN model by using the backpropagation of target domain data to obtain the final model M' = fine_tune ($M|Y_{target}$, X_{target}). The parameters are set as follows:

Parameter	**Value**
batch_size	20
learning_rate	0.01
dropout_value	0.5
max_iter	100

2) Fine-tuning algorithm 2

The stacked denoising autoencoder[12] (consisting of one input layer and one output layer, 128 nodes encoding layer, 64 nodes bottleneck layer, and 128 nodes decoding layer) is first trained using source domain data, then the decoding layer is removed and for each hidden layer added a dropout layer, and finally is a logistic regression layer. The result is a model similar to the DNN model mentioned earlier with fine-tuning of the target domain data. The parameters are set as follows:

Parameter	Value
batch_size	32
learning_rate	0.01
p	0.5
max_iter	500

3) Domain adaptation

Domain adaptation in transfer learning is suitable for dealing with data bias. Domain adaptation in transfer learning is suitable for dealing with data bias. The idea of domain adaptation is to map the data from different source and target domains into a feature space and make it as close as possible in this space. In this way, the target function trained for source domain in feature space can be transferred to target domain to improve the accuracy of target domain. Domain adaptation can improve the machine learning performance of the model for the target domain by adjusting the distribution differences between different groups of domains. We use the CCSA method for domain adaptation, which can train and improve the model accuracy of the target domain in the case of little labeled data. Meanwhile, the method can deal with the domain differences in edge distribution and conditional distribution.

We can assume source domain as $Ds = \{X_s, P(Xs)\}$, and target domain as $Dt = \{X_t, P(Xt)\}$. Domain is composed of feature space X and edge probability distribution P(X). Task is defined as $T = \{Y, P(Y|X)\}$, which consists of label space Y and conditional probability distribution P(Y|X). P(X) means marginal probability distribution[13]. If given source domain Ds and target domain Dt, and machine learning tasks Ts and Tt, the purpose of transfer learning is to use the knowledge learned in source domain Ds to help improve the learning of predictive functions F in target domain Dt[14]. In this study, the most suitable

method for data distribution deviation is to use domain adaptation. For example, we can learn data distribution from EA and extend it to AA.

Cross-validation and performance evaluation

We use 3-fold cross-validation to contrast these three machine learning solutions respectively. For mixture learning, the overall data were stratified according to clinical results and ethnic categories to make it evenly distributed. Model training was carried out using AA and EA data samples, with Mixture 0 performance assessed throughout the full test set, Mixture 1 performance assessed across the test set by EA, and Mixture 2 performance assessed by AA. EA and AA data were segregated and stratified by clinical outcome for independent learning, and carry out cross-validation on each of two races. EA and AA data are split for transfer learning, and EA is utilized for model training, followed by AA for model fine-tuning or domain adaptation, and lastly AA test set for performance evaluation. We use the area under the ROC curve (AUROC) to evaluate the model in order to effectively compare the relative performance of several machine learning approaches.

4. Results

Mixture learning and independent learning approaches have great model performance for EA, but poor (with p value of 6.72×10^{-11} and 1.29×10^{-26}) for the AA group with inferior data volume, according to machine learning experiments. If AA and EA are not evaluated independently, the overall good performance of the dataset will hide the performance discrepancy (Figure 6, Mixture 0)[3]. Compared with mixture learning and independent learning, transfer learning improves model performance for AA group with inferior data volume and reduces the model performance gap $\tilde{G} = \overline{AUROC}_{EA} - A_{Transfer}$, which shown in Figure 7[3]. These results suggest that data skew between races causes

the model performance difference between mixture learning and independent learning, while transfer learning can effectively compensate for this performance difference.

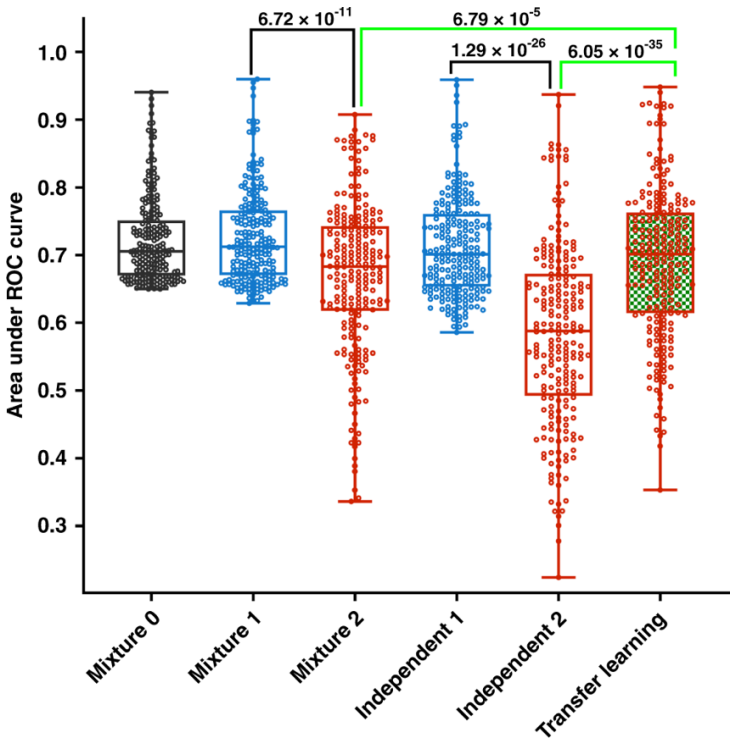

Figure 6. Comparison of AUROC

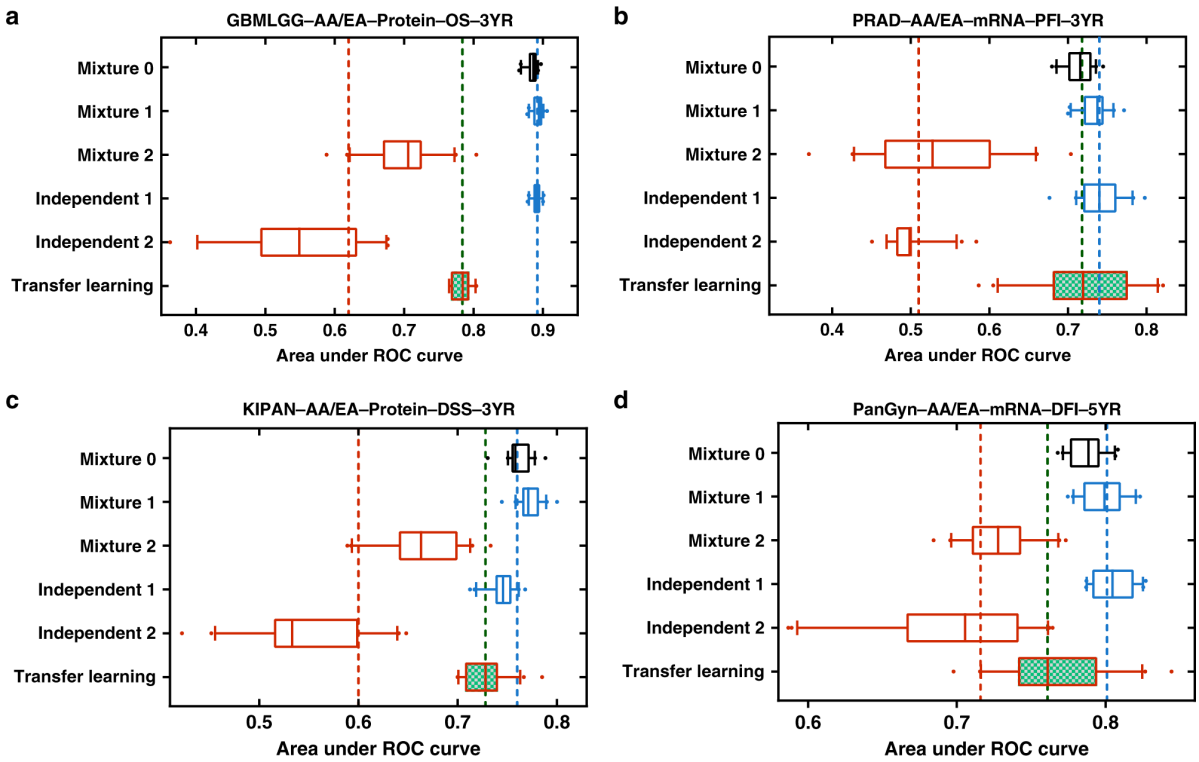

Figure 7. Comparison of machine learning Algorithms for ethnic group

System modeling and structure

The proposed system structure can see from Figure 8. In the data preparation phase, we kept the baseline data source unchanged and performed some data bias analysis on the original basis, attempting to delineate the severity of data skew in the ethnic cancer medical data set. Secondly, in the machine learning phase, feature combinations were made according to feature types, cancer types, clinical outcome endpoints and event points, and an independent machine learning task was assigned to each feature combination for training. After data cleaning, a total of 224 machine learning tasks were run in the baseline experiment. The results of each task are analysed to compare the performance of various machine learning schemes.

Due to the large dimension of feature data set, we have to consider using this feature combination method for analysis, instead of analysing on the whole feature. The advantage of this method is that the trained model has stronger

interpretability. However, due to the diversity of feature combination, it is necessary to pay more computational cost for machine training. At the same time, different feature factors may have a great impact on the performance of machine learning models, and no particularly effective method has been found to analyse the influence of feature factors, which may be a possible research direction in the future.

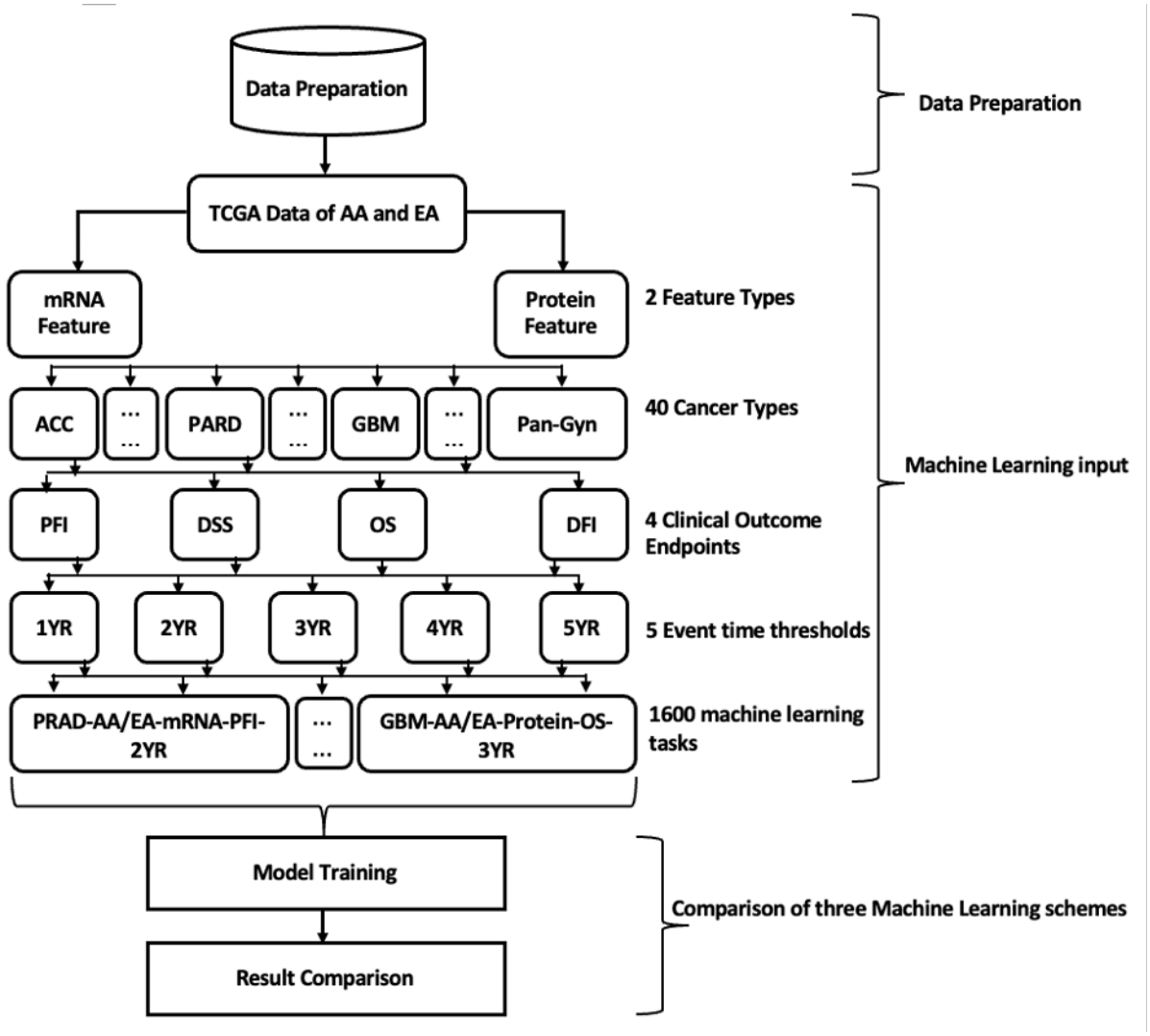

Figure 8. System structure overview

Methodology

The deep migration model in the benchmark experiment has achieved good results, but the coverage of the results is not wide, resulting in poor interpretability. In the baseline experiment, only EA was used as the source domain and was generalized to the AA dataset for evaluation. The baseline study was not well analysed for characteristic factors such as cancer type, ethnic cohort, and clinical outcome, which may have a significant impact on machine learning programs. We hope to carry out a new feature combination based on the baseline experiment, analyse the model performance of the baseline experiment and fine-tune and optimize our deep migration model under the new feature combination. And extend our target domains to other ethnic groups for analysis, such as ASA and NAT-A. The results show that under the new feature combination space, the performance of the baseline machine learning model is significantly reduced, which indicates that we have a large room for improvement to optimize the model. We will further analyse and put forward new attempts. The main purpose of this study is to try and analyse the methods and conclusions of the baseline experiment, propose a new improved scheme to compare with the original scheme and try to improve the performance of the deep transfer learning model.

We focus on fine-tuning the deep transfer learning model to accommodate generalization across the target domains of different ethnic groups. Based on the baseline experiment, we supplemented the model training with different feature combinations and fine-tuned the model. We kept EA as the source domain and added ASA and NAT-A as the target domain. The experimental process is shown in Figure 9. In the first stage, we first reviewed the baseline experiment, and then carried out supplementary experiments using new feature combinations and model hyperparameters. The supplementary experiments were also carried out based on the baseline experiment. We then use the same

new features and fine-tuned model to experiment on the new target domain data. The new data distribution can see from Table 3.

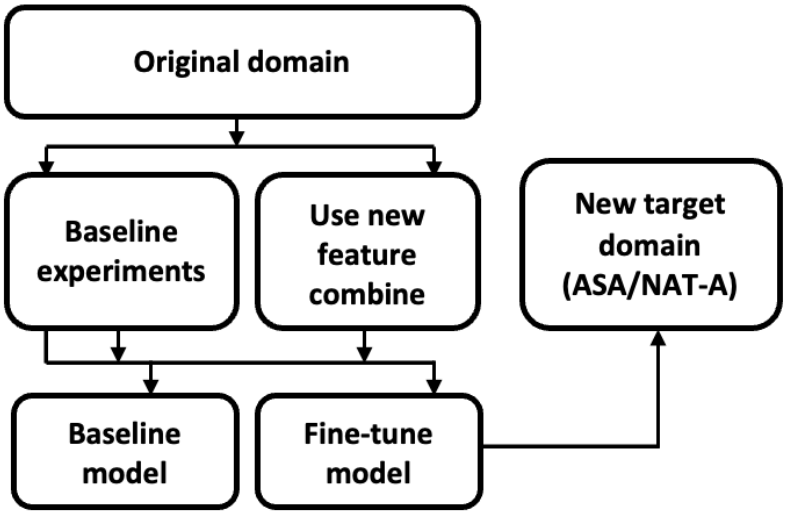

Figure 9. Experimental flow chart

Machine learning scheme	Experiment	Training data ethnic composition	Testing data ethnic composition	Training data ethnic composition	Testing data ethnic composition
Mixture learning	Mixture 0	EA+ASA	EA+ASA	EA+NAT-A	EA+NAT-A
	Mixture 1		EA		EA
	Mixture 2		ASA		NAT-A
Independent learning	Independent 1	EA	EA	EA	EA
	Independent 2	ASA	ASA	NAT-A	NAT-A
Transfer learning	Transfer learning	EA (source domain) ASA (target domain)	ASA	EA (source domain) NAT-A (target domain)	NAT-A

Table 3. New data distribution based on fine-tuned model

Result analysis

In this section, we introduce model-related Settings and analyze the results. Our project was run in Python 3.6 environment and used third-party libraries like tensorflow and sklearn. Due to the lack of high-performance machines for experiments, we had to reduce the number of experiments, which may not obtain higher precision model parameters. Secondly, in terms of the selection of model parameters and analysis methods, we refer to the paper published by Muhammad[15] et al., with which some of the work in this paper overlaps. In order to evaluate the performance of the model, we still adopted AUROC method for analysis. AUROC is used to measure the total area of the model under the true positive and false positive parameter curves, as shown in the following formula:

$$TPR = \frac{TruePositive}{TruePositive + FalseNegative}$$

$$FPR = \frac{FalsePositive}{TrueNegative + FalsePositive}$$

In the baseline experiment, the authors performed all experiments using EA and AA (as shown in Table 2). Using protein and mRNA as characteristic types and 40 cancer categories respectively, the authors trained and tuned the results of three machine learning schemes using 3x cross-validation. The three machine learning schemes include hybrid learning, independent learning and transfer learning. For mixed learning, EA and AA are used for deep learning model training, and EA and AA are used as test data for model performance test to obtain the results. For independent learning, EA and AA are used for data cross model training and testing to obtain the respective test results on the two data sets. For domain adaptive transfer learning, EA is used as the source domain for training and extended to the AA target domain for model testing. The results

show that when EA is used as the source domain, the performance of the three machine learning schemes is generally high, but the performance of AA is poor, indicating that data skew seriously affects the performance of the model.

Based on the baseline experiment, we adopted the new data distribution for the experiment, as shown in Table 3, and the experimental process was shown in Figure 9.

In order to ensure the fairness of the results, we first used the hyperparameter pairs consistent with the baseline experiment for model training. The specific experimental steps can be summarized as follows:
1) First use baseline method to evaluate the new feature combination and analyze the cause of low performance.
2) Secondly, the model is fine-tuned and optimized and the new results are evaluated.
3) Then use the new ethnic data distribution and apply our fine-tuning model to compare with the baseline model.
We ensured that there were at least 5 cases in the prognostic categories of EA-AA, EE-ASA, and EA-NAT-A data distributions to ensure the reliability of our model. Due to the limitation of computation, we only compare the transfer learning scheme and baseline, and we do not discuss the mixed learning scheme and independent learning scheme.

Transfer learning result on EA/AA

We follow the baseline experiment, using source domain EA and target domain AA. The analysis results showed that the overall model performance of transfer learning was not significantly different from that of independent learning, which indicated that the performance of the adaptive deep transfer learning model in the application field was not as good as expected. We analyzed the reasons for

the low performance and adjusted parameters according to Muhammad's paper[15] (see Table 4), and then used the optimized model to conduct the experiment again. The result can be seen from Table 5. Due to the limitation of computation, only four feature combinations were selected for calculation. It can be seen from the results that the AUROC index of the fine-tuned model has been significantly improved compared with the baseline experiment (see figure 10).

Parameter	**Value**
Batch_size	16
Dropout	0.5
Epochs	20
Learning_rate	0.00001
L1 reg	0.0001
L2 reg	0.0001

Table 4. Hyperparameter values for fine-tuning model

		AUROC	
Feature Combine	Data Distribution	Baseline	Fine-tune model
SARC-mRNA-OS-2YR	EA->AA	0.54	0.68
BLCA-Protein-PFI-4YR	EA->AA	0.79	0.86
UCEC-Protein-DFI-3YR	EA->AA	0.59	0.63
SARC-mRNA-DSS-3YR	EA->AA	0.63	0.68

Table 5. AUROC results on EA/AA

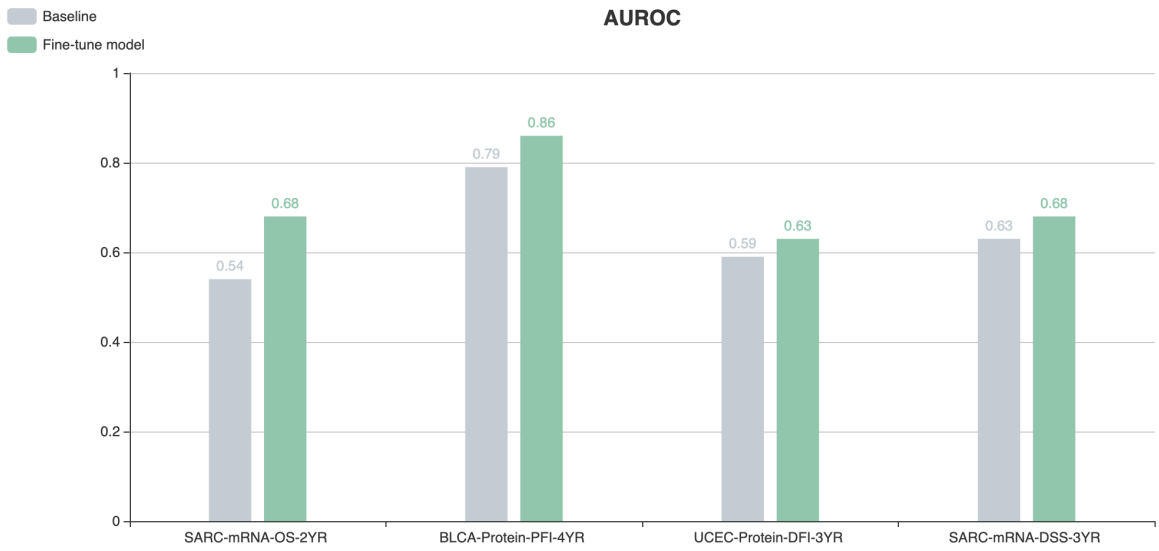

Figure 10. AUROC comparison with baseline

Transfer learning result on EA/ASA and EA/NAT-A

In this stage, we will use ASA and NAT-A, two new data distributions, to carry out the experiment. In order to ensure the reliability of the results, we filtered out the cancer types with less than 5 cases. The experimental results are shown in Table 6, from which we can see that our model also has a great improvement compared with the baseline (see from Figure 11).

		AUROC	
Feature Combine	Data Distribution	Baseline	Fine-tune model
BRCA-mRNA-PFI-3YR	EA->ASA	0.60	0.64
STAD-mRNA-OS-1YR	EA->ASA	0.61	0.64
UCEC-mRNA-DSS-4YR	EA->ASA	0.65	0.63
CESE-mRNA-OS-2YR	EA->NAT-A	0.56	0.61
CESE-mRNA-DSS-1YR	EA->NAT-A	0.44	0.51
CESE-mRNA-DSS-2YR	EA->NAT-A	0.54	0.6
PanGyn-mRNA-OS-2YR	EA->NAT-A	0.42	0.51

Table 6. AUROC results on EA/ASA and EA/NAT-A

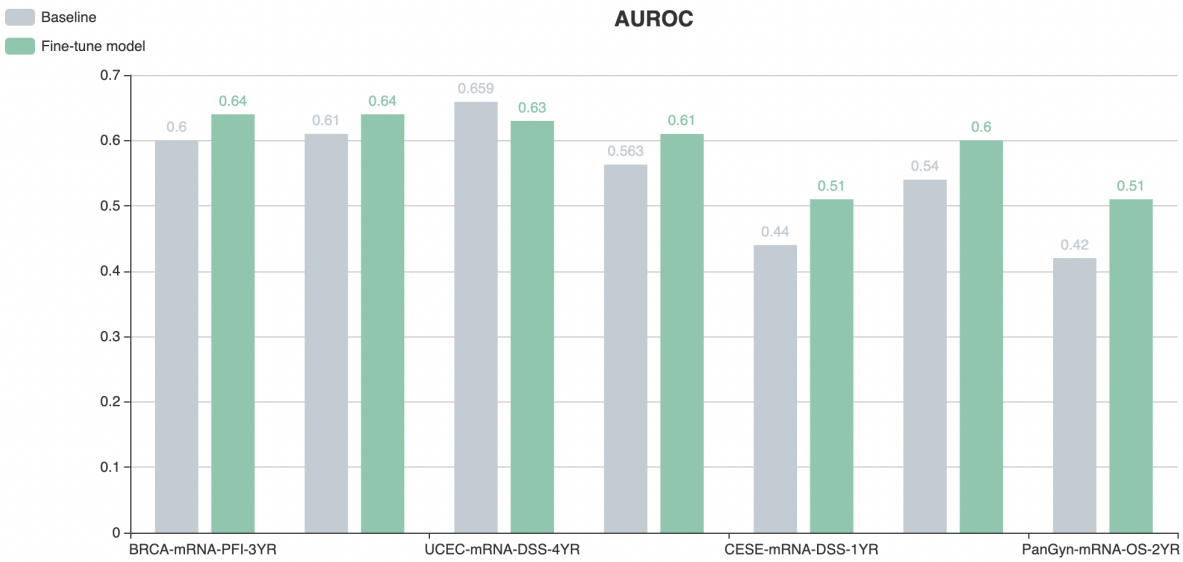

Figure 11. AUROC comparison with baseline

Conclusion

In this study, we initially touched on the application of transfer learning. We fine-tuned the deep transfer learning model based on the baseline experiment, and used the fine-tuned model to test and compare the baseline data distribution and the new ethnic data group. Compared with the single target domain of the baseline, we tested and analyzed the model performance in the new race target domain. We did not switch the source domain, because the source domain accounts for the most data volume, which can help the model to learn the best feature distribution. The results show that there is still a lot of room for improvement in transfer learning. Migration study is the hot research direction, especially in the field of medical information, for medical analysis of ethnic groups is a very promising direction, I personally in machine learning knowledge reserves is not particularly deep, met a lot of algorithms in the experiments on the understanding of the difficulties, and large scale of data brought me calculation difficulties. In the future, I hope to have an in-depth grasp of relevant theories of machine learning and apply them to do some original data analysis work.

Future Work

Due to the high dimension of protein expression and mRNA characteristics, it is difficult to dig out important influencing factors. The potential future research direction is to explore the influence range of protein, mRNA and other characteristic factors on cancer, as well as the influence of other combination of characteristics, such as clinical endpoint and year, on model performance. It may be possible to reduce the dimension of feature space through data dimension reduction to further improve the accuracy of the model. In addition, if we make significant progress in predicting multiracial cancers in models, it could be extended to other medical applications in the future, such as the current COVID-19 pandemic.

Reference

[1] Guerrero, S. et al. Analysis of racial/ethnic representation in select basic and applied cancer research studies. Sci. Rep. 8, 13978 (2018).

[2] Binhua Tang, Zixiang Pan, Kang Yin, and Asif Khateeb. Recent advances of deep learning in bioinformatics and computational biology. Frontiers in genetics, 10:214, 2019.

[3] Gao, Y., Cui, Y. Deep transfer learning for reducing health care disparities arising from biomedical data inequality. Nat Commun 11, 5131 (2020).

[4] Cruz JA, Wishart DS. Applications of Machine Learning in Cancer Prediction and Prognosis. Cancer Informatics. January 2006. doi:10.1177/117693510600200030

[5] Daniel Berrar, Werner Dubitzky, Deep learning in bioinformatics and biomedicine, Briefings in Bioinformatics, Volume 22, Issue 2, March 2021, Pages 1513–1514.

[6] Wainberg, M., Merico, D., Delong, A. et al. Deep learning in biomedicine. Nat Biotechnol 36, 829–838 (2018).

[7] R. K. Sevakula, V. Singh, N. K. Verma, C. Kumar and Y. Cui, "Transfer Learning for Molecular Cancer Classification Using Deep Neural Networks," in IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 16, no. 6, pp. 2089-2100, 1 Nov.-Dec. 2019, doi: 10.1109/TCBB.2018.2822803.

[8] Torrey L, Shavlik J. Transfer learning[M]//Handbook of research on machine learning applications and trends: algorithms, methods, and techniques. IGI global, 2010: 242-264.

[9] Yunhui Guo, Honghui Shi, Abhishek Kumar, Kristen Grauman, Tajana Rosing, Rogerio Feris; Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 4805-4814

[10] Shi, Z., Hao, H., Zhao, M. et al. A deep CNN based transfer learning method for false positive reduction. Multimed Tools Appl 78, 1017–1033 (2019).

[11] Lopez-Garcia G, Jerez J M, Franco L, et al. Transfer learning with convolutional neural networks for cancer survival prediction using gene-expression data[J]. PloS one, 2020, 15(3): e0230536.

[12] Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y. & Manzagol, P.-A. Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. J. Mach. Learn. Res. 11, 3371–3408 (2010).

[13] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. IEEE Transactions on knowledge and data engineering, 22(10):1345–1359, 2009.

[14] Pan, Sinno Jialin, and Qiang Yang. "A survey on transfer learning." IEEE Transactions on knowledge and data engineering 22.10 (2009): 1345-1359.

[15] Toseef, Muhammad, Xiangtao Li, and Ka-Chun Wong. "Reducing healthcare disparities using multiple multiethnic data distributions with fine-tuning of transfer learning." Briefings in Bioinformatics (2022).