2023 IEEE Symposium Series on Computational Intelligence (SSCI)
Mexico City, Mexico. December 5-8, 2023

Classification Using Deep Transfer Learning on Structured Healthcare Data

1st Ayda Farhadi
Department of Computer Sciences
University of Georgia
GA,USA,
aydafarhadi14@gmail.com

2nd David Chen
Department of Health Sciences Research
Mayo Clinic
MN,USA
chen.david@mayo.edu

3rd Rozalina McCoy
Department of Health Sciences Research
Mayo Clinic
MN,USA
mccoy.rozalina@mayo.edu

4th Christopher Scott
Department of Health Sciences Research
Mayo Clinic
MN,USA
scott.christopher@mayo.edu

5th Ping Ma
Department of Statistics
University of Georgia
GA,USA
pingma@uga.edu

6th Celine M. Vachon
Department of Health Sciences Research
Mayo Clinic
MN,USA
vachon.celine@mayo.edu

7th Jingyi Zhang
University Center for Statistical Science
Tsinghua University
Beijing, China
jingyizhang@tsinghua.edu.cn

8th Che Ngufor
Department of Health Sciences Research
Mayo Clinic
MN,USA
ngufor.che@mayo.edu

9th John A. Miller
Department of Computer Sciences
University of Georgia
GA,USA
jamill@uga.edu

***Abstract*—In healthcare, building a supervised learning system faces the challenge of access to a large, labeled dataset. To overcome this problem, we propose a deep transfer learning method that addresses imbalanced data problems in healthcare, focusing on structured data. We use publicly available breast cancer datasets to generate a source model and transfer learned concepts to predict high-grade malignant tumors in patients diagnosed with breast cancer at Mayo Clinic. The diabetes dataset is then used to generalize the transfer learning idea. We compare our results with state-of-the-art techniques and demonstrate the superiority of our proposed methods. Our experiments on breast cancer data under simulated class imbalanced settings further demonstrate the proposed method's ability to handle different degrees of class imbalance. We conclude that deep transfer learning on structured data can efficiently address imbalanced class and poor performance learning on small dataset problems in clinical research.**

***Index Terms*—Breast cancer, Deep transfer learning, Deep learning, Class imbalance, SMOTE**

I. Introduction

The problem of imbalanced data is a challenge for machine learning algorithms as they tend to generalize patterns observed over the majority class and ignore the minority class. One of the most popular techniques to handle class imbalance is the Synthetic Minority Over-Sampling Technique (SMOTE). However, SMOTE and variants such as RUSBoost may cause information deficiency problems. To mitigate these issues, SMOTE has been combined with ensemble learning techniques, but hybrid techniques can still generate sub-optimal results for severely imbalanced datasets.

Identify applicable funding agency here. If none, delete this.

Data dependence is another problem in machine learning, especially in healthcare, where it is difficult to construct a large-scale annotated dataset due to the complexity or rarity of diseases and heterogeneity of clinical data sources. The transfer learning methodology can be used to mitigate insufficient training data and class imbalance problems in a given target domain. Deep transfer learning has been well studied in the context of image classification, but its application to structured healthcare data is relatively unknown.

In this paper, we propose a deep transfer learning (DTL) approach on structured data to address the imbalance and insufficient data classification problem of healthcare data. We hypothesize that DTL enables structured-data-based solutions that can improve the early detection and classification of disease. Two case studies are presented to illustrate the importance of DTL.

A. Deep Learning

Deep learning is increasingly being used to analyze electronic health records (EHRs) for various clinical tasks. It aids in disease diagnosis, cancer identification, genotype-phenotype relationships, and disease prediction. By discovering complex relationships between features, deep learning assists clinicians in decision-making without requiring extensive data pre-processing. This machine learning algorithm utilizes multiple hidden layers to learn healthcare data representations. Deep neural networks with more hidden layers are considered "deep," while the final output layer acts as a nonlinear classifier using abstract features from the hidden layers.

978-0-7381-4408-5/23/$31.00 ©2023 IEEE

Feature transformations can be shared across related datasets, while the classifier varies based on the dataset.

II. Challenges in Applying Classification in Healthcare

Applying deep learning to health informatics presents challenges such as high-dimensional and heterogeneous data, data scarcity leading to missing values and class imbalance [1]. Additionally, addressing data credibility, model interpretability, reliability, feasibility, security, and scalability is crucial. Health's unique characteristics, including limited disease understanding, human interventions, and fragmented data, require thoughtful application of machine learning in healthcare.

A. Data Scarcity Problem in Healthcare

Healthcare data has several unique characteristics that differentiate them from data in other areas. Healthcare data are difficult to access due to privacy concerns, and structured due to the extraction process. Additionally, healthcare data are collected in safety-critical conditions and may be affected by several sources of uncertainty. Furthermore, the insufficiency of labeled data is a significant issue in machine learning and data mining for healthcare. Semi-supervised learning can address this problem by utilizing small sets of labeled data and much larger unlabeled data collections. However, the scarcity of labeled data renders many statistical approaches, such as deep learning, unusable. The solutions to this issue are to collect more data or develop techniques that can handle smaller datasets.

B. Class Imbalance Learning

An imbalanced dataset presents a challenge to learning, as standard classification algorithms assume that test data comes from the same distribution as training data. Most standard classifiers optimize a loss function based on minimizing error, leading to a misleading performance metric in imbalanced learning. Existing methods for handling class imbalance can be categorized into three groups: data sampling, algorithmic modification, and cost-sensitive learning. Imbalanced learning studies show that performance loss is mainly due to the skewed class distribution, but feature selection may also be beneficial in handling high-dimensional imbalanced datasets. indicates that the original dataset may lack density and information or that the classes may be overlapped, but feature selection provides the right information needed to discriminate between the classes [2].

III. Techniques to Address the Challenges

A. Re-sampling Imbalanced Data

Class imbalance in datasets can be addressed by adjusting the class distribution using random oversampling (ROS) and random undersampling (RUS). However, these techniques can lead to overfitting or loss of majority samples. Synthetic Minority Over-Sampling TEchnique (SMOTE) generates synthetic minority examples using a k-nearest neighbor algorithm, overcoming these limitations. Combining SMOTE with ensemble methods like SMOTEBoost and RUSBoost has been successful, but it can result in less informative training data and suboptimal models for severely imbalanced datasets.

B. Transfer learning

Transfer learning is a technique that leverages knowledge from a source domain ($D_s = (X_s, Y_s)$) to improve the learning of a target domain ($D_t = (X_t, Y_t)$) with partially overlapping data distributions and feature spaces [3] . This approach helps address the data dependence problem in scenarios with limited or unbalanced training data [4]. Consider n_s and n_t observations for the source and target, respectively. Figure 1 illustrates the transfer learning process. Let L_s and L_t be the layers for the source and target tasks. Suppose there are M transferable hidden layers, denoted as $\mathbf{v}s^{k_j} j = 1^M$ and $\mathbf{v}t^{l_j} j = 1^M$, where $k_j, l_j \in 1, ..., \min(L_s, L_t)$. The transferable hidden layers satisfy certain conditions related to the dimension of the latent space and the distance between source and target distributions. $\mathcal{D}\left(p(\mathbf{v}_s^{l_j}), p(\mathbf{v}_t^{k_j})\right) < \epsilon$, where $p(\mathbf{v}_s^l)$ and $p(\mathbf{v}_t^l)$ denote the distributions of the source hidden layer and the target hidden layer, respectively; $\mathcal{D}(\cdot, \cdot)$ denotes the "distance" between two distributions, such as the Kullback-Leibler (KL) divergence. Transfer m layers from the M transferable hidden layers to improve the learning of the target predictive function f_t.

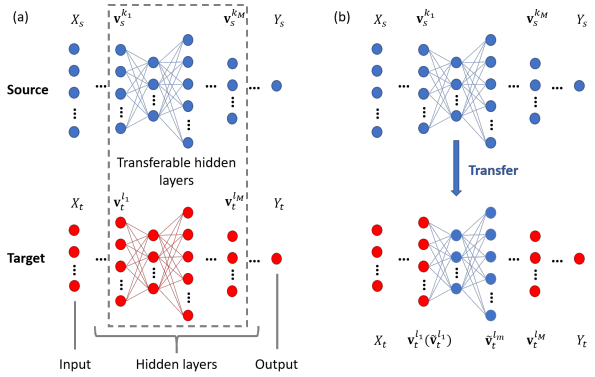

Fig.1: Illustration of transfer learning. Left: locating transferable hidden layers. Right: transferring layers that can improve the learning of the target predictive function.

Notice that the value of M is based on the predetermined threshold ϵ. A large ϵ means that more information can be borrowed from source data. However, if the borrowed information is too specific with the source data, it may results in an inefficient model for the target data. If ϵ is small, we turn to transfer hidden layers that contain less "detailed" information, i.e., the transferred information could be too general. As a result, the threshold ϵ need to be carefully chosen in transfer learning

IV. Case Study 1

A. Data

1) Source Data:

*a) **Breast Cancer Datasets**:* In this paper, two publicly available breast cancer datasets from the UCI machine learning repository were utilized to implement the proposed DTL on structured data. The datasets used were the Wisconsin Diagnostic Breast Cancer (WDBC) with 357 benign and 212 malignant cases, and the Wisconsin Prognostic Breast Cancer (WPBC) data which represented follow-up data for patients seen at the University of Wisconsin Hospital at Madison from 1984 until 1995, and included 47 breast cancer recurrence and 151 nonrecurrence cases. The 30 features in each dataset were calculated for each image and were used to create a merged dataset consisting of 767 samples.

2) Target Data:

*a) **The Mayo Mammography Health Study (MMHS)**:* To demonstrate the effectiveness of the proposed transfer learning method on structured data, a primary dataset consisting of 15,386 women aged 37 or older diagnosed with breast cancer between January 15, 2009, and January 15, 2016 was used. This dataset is a subset of the Mayo Mammography Health Study (MMHS) Cohort, which included 19,936 women enrolled at the Mayo Clinic and had a screening mammogram performed between 2003 and 2006. Table 1 presents the distribution of features selected for this study stratified by diagnosis status. The dataset contains patient demographics, BMI, family history, and clinical breast density measurements. During the follow-up period, 487 (3.2%) women were diagnosed with invasive breast cancer, resulting in severely imbalanced data (IR = 30.6). The dataset was challenging for standard classification methods to learn due to the limited number of features and severe class imbalance. The MMHS was approved by the Mayo Clinic institutional review board. Additional details about the MMHS can be found in [5].

TABLE I
STATISTICAL ANALYSIS OF TARGET MMHS DATA (LOGISTIC REGRESSION)

Features	Number	Benign(N=14899)	Malignant(n=487)	Total(N=15386)	*p*-value	coef	OR (30.5)
BI-RADS Breast Density Changes	1						
First Month					0.000	0.017	
1		3652(24.5%)	74(15.2%)	3726 (24.2%)			0.020
2		6175(41.4%)	215(44.1%)	6390(41.5%)			0.034
3		4339(29.1%)	179(36.8%)	4518(29.4%)			0.041
4		733(4.9%)	19(3.9%)	752(4.9%)			0.025
Second Month	2				0.000	0.008	
1		3434 (23.0%)	73(15.0%)	3507(22.8%)			0.021
2		6437(43.2%)	226(46.4%)	6663(43.4%)			0.035
3		4364(29.3%)	166(34.1%)	4530(29.4%)			0.038
4		664(4.5%)	22(4.5%)	686(4.5%)			0.033
Third Month	3				0.83	0.007	
1		3193(21.4%)	86(17.7%)	3279(21.3%)			0.026
2		6459(43.4%)	200(41.1%)	6659(43.4%)			0.030
3		4633(31.1%)	178(36.6%)	4811(31.3%)			0.038
4		614(4.1%)	23(4.7%)	637(4.1%)			0.037
Fourth Month	4				0.78	-0.009	
1		2990(20.1%)	89(18.3%)	3079(20.0%)			0.029
2		6638(44.6%)	201(41.3%)	6839(44.4%)			0.030
3		4702(31.6%)	177(36.3%)	4879(31.7%)			0.037
4		569(3.8%)	20(4.1%)	589(3.8%)			0.035
Family History of BC	7	2698(18.1%)	119(24.4%)	2817(18.3%)	0.007	0.009	
Age at Enrollment	8	56.66(11.31)	58.82(10.30)	56.73(11.29)	0.000	0.037	
Autofluorescence Bronchoscopy Group	9				0.000	0.005	
0		1970(13.2%)	65 (13.3%)	2035(13.2%)			0.032
1		2267(15.2%)	69(14.2%)	2336(15.2%)			0.030
2		5801(38.9%)	193(39.6%)	5994(39.0%)			0.033
3		3308(22.2%)	107(22.0%)	3415(22.2%)			0.032
4		1445(9.7%)	47(9.7%)	1492(9.7%)			0.032
Age at Menarche	10				0.001	0.002	
1		2503(16.8%)	80 (16.4%)	2583(16.8%)			0.031
2		7954 (53.4%)	283(58.1%)	8237(53.5%)			0.035
3		3481 (23.4%)	93 (19.1%)	3574(23.3%)			30.026
Menopause at Enrollment	11	9081(61.0%)	346(71.0%)	9427(61.3%)	0.20	0.004	
BMI at Enrollment	12	28.14(6.42%)	28.39(6.43%)	28.14(6.42%)	0.000	0.004	
BI-RADS at Enrollment	13				0.000	0.0151	
1		3268(21.9%)	73(15.9%)	3341(21.7%)			0.022
2		5865(39.4%)	194(39.8%)	6059(39.4%)			0.033
3		4731(31.8)	176(36.1%)	4907(31.9%)			0.037
4		1035(6.9%)	44(9.0%)	1079(7.0%)			0.042
Age at Diagnosis	14				0.000	-0.035	
0	30	352(2.3%)	0	352			0
1	40	4253(28.5%)	28(5.74%)	4281(27.8%)			0.006
2	50	4335(29.09%)	110(22.5%)	4445 (28.8%)			0.025
3	60	3492(23.4%)	167 (34.2%)	3659(23.7%)			0.047
4	70	2031(13.63%)	131 (26.8%)	2161 (14%)			0.064
5+	80+	435(2.8%)	51 (9.8%)	485(3%)			0.11

For the missing values in the data, we used two imputation techniques. The first technique, LOCF (Last Observation Carried Forward) and NOCB(Next Observation Carried Backward), was used to impute the missing breast density measurements. This involved carrying forward the last valid non-missing level. Afterward, we transformed the serial values from a long to a wide format, with individual values at each time point as feature vectors. MissForest was then applied to impute any remaining missing features.

*b) **UCI Mammographic Mass Data**:* The publicly available mammographic mass dataset from the UCI machine learning repository was used in this study. The data was used in this study to predict the classification (benign or malignant) of a mammographic mass lesion. It contains 516 benign and 445 malignant cases. Because the mammographic mass dataset was relatively balanced (IR = 1.2), the baseline classifiers were expected to perform well. As such, we also used the data to investigate the effect of imbalanced learning by training the models under different IRs in the training data. Specifically, we carried out a controlled experiment by simulating class imbalance in the training data, whereby samples from the malignant class were removed randomly. We created training datasets where the proportion of observations in the minority class was: 10% (IR = 5.4), 5% (IR = 10.7), and 2% (IR = 26.8).

B. Method

1) Proposed Deep Transfer Learning for Breast Cancer: For our proposed deep transfer learning method, we utilized the DNN as described previously. The objective of this method is to transfer knowledge from a structured source domain data with potentially large and balanced class distribution to a structured target domain data that is imbalanced. The model architecture is composed of an input layer, an output layer, and three hidden layers, with a total of 8334 parameters. The initial layers of a DNN capture general features of the disease, while later layers focus on specific characteristics. With transfer learning, one can freeze the initial layers of a pre-trained model and retrain the remaining layers on a new dataset, using the same DNN architecture. This allows for faster and more efficient training on new data. Any of the layers of the DNN may be frozen or unfrozen during transfer learning, and are referred to as transferred hidden layers. We

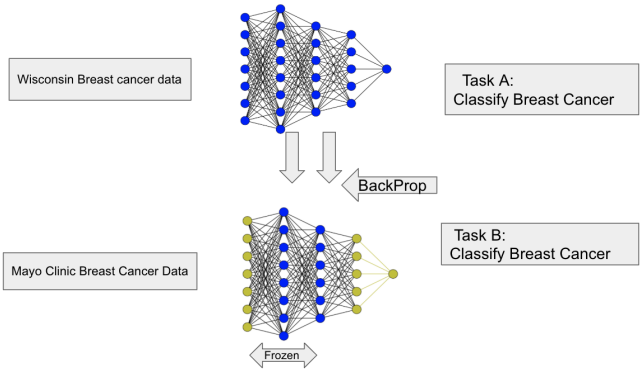

Fig.2: Deep transfer learning for breast cancer classification

also develop our target DNN using 5 layered DNN and 7 layered DNN based on the Mayo Clinic Breast cancer data. We implemented both freeze transferred hidden layers and unfreeze transferred layers, with a different number of layers.

Specifically, we implemented and compared the performance of several DTL models. Figure 2 illustrates the implemented DTL network architecture, where the source model (shown in blue) is a DNN. To derive the 5L-2T-Freeze model, for example, we transferred the top two hidden layers of the source model to target model and froze their weights.

C. Results

*a) **The performance of classifiers on Breast Cancer dataset**:* Basic descriptive statistics of the features in the MMHS target breast cancer data stratified by breast cancer diagnosis is presented in Table 1. Each of the 4 BI-RADS levels represents gradations of the likelihood that a cancer exists, from lowest to highest probability. We also include a transformed variable with levels (0, -1, and 1) indicating no change, decrease, and increase in the BI-RADS respectively over the four time periods.

1) Performance of Source DNN on Source Data: We used AUC to select the best model parameters during training and validation in predicting breast cancer on the source data. The corresponding validation AUC on breast cancer source model was 0.96(0.10).

2) Performance of DTL and Baseline Models: We equally trained the DTL models: 5L-2T-Freeze, 7L-2T-Freeze, 5L-2T-UNFreeze, and 5L-1T-UNFreeze as previously described to classify breast cancer on the MMHS data and the UCI Mammographic mass datasets.

TABLE II
PERFORMANCE OF MODELS ON HOLD OUT FOLD OF MMHS

Classifiers	AUC-cv	Sens-cv	F1-cv
LR	0.59	0.20	0.30
RF	0.61	0.09	0.05
XGBoost	0.68	0.10	0.15
DNN-5L	0.65	0.32	0.37
DNN-7L	0.66	0.32	0.37
LR-sampling	0.65	0.20	0.31
RF-sampling	0.53	0.47	0.07
SMOTEBoost	0.67	0.36	**0.44**
RUSBoost	0.76	0.74	0.17
5L-2T-Freeze	0.80	0.75	0.29
5L-2T-UNFreeze	**0.81**	0.75	0.28
5L-1T-UNFreeze	0.80	0.71	0.27
7L-2T-Freeze	0.76	**0.81**	0.19

Table 2 presents performance results of the models on the MMHS data. With respect to the AUC metric, resampling the data with SMOTE marginally improved the performance of LR, while the performance of RF deteriorated. The performance of RUSBoost, which creates a balanced training set by undersampling rather than generating synthetic examples, was significantly better than all baseline models including SMOTEBoost. On the other hand, all the DTL models except 7L-2T-Freeze outperformed the other comparator methods. The 5L-2T-UNFreeze showed the best AUC performance. However, the 7L-2T-Freeze model was the most sensitive model in detecting malignant cases.

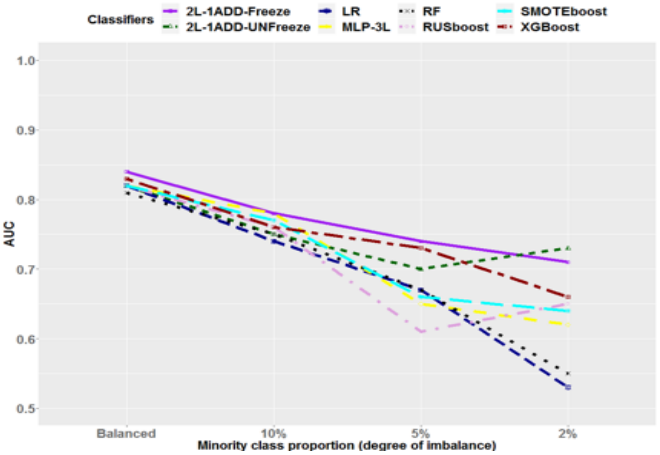

Fig.3: Performance of models on the UCI Mammographic mass data with simulated imbalanced training data distribution

Figure 3 presents performance results of the models on the UCI Mammographic Mass data. As expected, the performance of the models deteriorated as the proportion of minority classes in the data decreased (or increased in IR). The AUC values in the figure represents the performance of the models on the original dataset (IR = 1.2). The 2L-1ADDFreeze DTL model showed slightly better performance compared to other methods, and remained dominant throughout the different simulated levels of IRs. Finally, the 2L-1ADDFreeze and 2L-1ADD-UNFreeze DTL models significantly outperformed all the other models for the very severe imbalanced (IR = 26.8) case. Based on the comparison results, it can be concluded that the DTL is an efficient methodology for standard classification tasks with balanced or imbalanced structured healthcare data.

D. Discussion

Breast cancer poses a substantial burden on global and US morbidity and mortality rates. Timely detection of high-grade malignant breast cancers plays a pivotal role in enhancing survival rates and patient outcomes. Nevertheless, the accurate differentiation between malignant and benign lesions using digital mammography, the primary screening modality for breast cancer, is hindered by the limited representation of cases in the overall population. Conventional approaches for predicting rare events lack the required accuracy. By leveraging transfer learning, significant advancements can be made in early breast cancer detection. Specifically, the study shows how transfer learning can facilitate the prediction of breast cancer occurrence in a general screening cohort using features learned from differentiating between malignant and benign or recurrent and non-recurrent cancers in publicly available datasets (WDBC and WPBC).

V. CASE STUDY 2

A. Data

1) Source Data:

*a) **Diabetes datasets**:* To generalize our transfer learning models on structured datasets, we utilized a dataset encompassing 10 years (1999-2008) of clinical care for diabetes patients from 130 US hospitals and integrated delivery networks, sourced from the University of California Irvine. The dataset met specific criteria, including being an inpatient encounter, a diabetic encounter, having a length of stay between 1 and 14 days, including laboratory tests and medication administration. The target variable for our binary classification problem was readmission, categorized as '< 30' (readmitted within 30 days) or '> 30' (readmitted after 30 days), which we defined as 'yes' and 'no' respectively. With over 49 features capturing patient and hospital outcomes, the dataset comprised 100,000 instances.

2) Target Data:

*a) **Pima Indians Diabetes Dataset**:* The dataset used in this study was sourced from the UCI machine learning repository. The dataset contained 768 cases of Pima Indian diabetes and included various features such as the number of times pregnant, plasma glucose concentration 2 hours in an oral glucose tolerance test, diastolic blood pressure (mm Hg), triceps skinfold thickness (mm), 2-hour serum insulin (mu U/ml), body mass index (weight(kg)/$(height(m))^2$), diabetes pedigree function, and age (years). A total of 768 cases were available in the Pima Indians diabetes dataset.

B. Method

1) Proposed Deep Transfer Learning for Diabetes: In this study, we investigate two types of transfer learning - feature-based and network-based. For feature-based transfer learning, we develop autoencoder-decoders for the Diabetes readmission and Pima Indian diabetes datasets. The bottleneck layers of these models are saved for transfer to the target model for classification. We also create a 5-layered deep neural network based on the Pima Indian diabetes dataset for classification. Additionally, we employ network-based transfer learning on the Breast cancer dataset, aiming to transfer knowledge from a structured source domain to a target domain with less data. We use backpropagation, batch normalization, and freeze initial layers for retraining, saving the model only when accuracy improves.

C. Results

*a) **KL divergence between source and target features in Diabetes dataset**:* For our study, we aimed to evaluate the difference between two datasets using KL divergence, which is a widely accepted measure of distance between two distributions. However, calculating the KL divergence was challenging due to differences in the dimensions of the source and target datasets. To address this issue, we used Principal Component Analysis (PCA) to remove redundancy and simplify the high-dimensional data while preserving trends and patterns. We then extracted Principle Components (PCs) and encoded features on both datasets to compare the KL divergence between PCs and encoded features of source and target data. This allowed us to investigate the potential for reducing the KL divergence between source and target data. While autoencoders may have nonlinear encoder/decoders, we opted for PCA as a simpler alternative. Then we compared KL divergence between PCs of source and target data, and encoded features of source and target data using a method as explained below:

Let KL_{before} denote the KL-divergence before encoding, and KL_{after} denote the KL-divergence after encoding. Consider the following hypothesis:

$H_0 : KL_{after} = KL_{before}.$
$H_1 : KL_{after} < KL_{before}.$

To test our hypothesis that implementing the autoencoder-decoder could reduce the KL-divergence between source and target and enable proper domain adaptation for transfer learning, we randomly selected 200 samples from source and target principal components and calculated the KL-divergence. Based on the empirical distributions, we conducted a two-sample test and found that the p-value was nearly 0, indicating that autoencoder-decoder reduced the KL-divergence between source and target. We then applied the feature-based transfer learning method to this data and chose δ to be $D_{KL}(x_s, x_t)/8$, which confirmed the similarity between the source and target datasets and made transfer learning possible. The corresponding ACC on diabetes source model was 0.91(0.03).

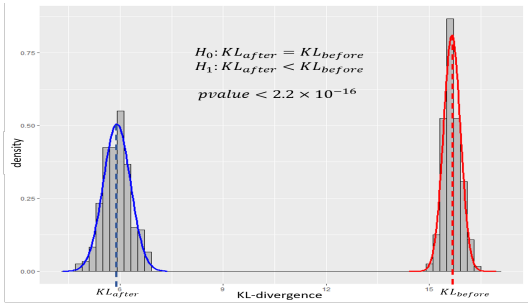

Fig.4:Two-sample KL-divergence testing.

*b) **The performance of classifiers on Diabetes dataset**:*

1) Performance of Source DNN on Source Data: The ACC and cross entropy loss function were utilized to select the optimal model parameters during the training and validation stages of predicting diabetes on the source data. The corresponding ACC for the diabetes source model was 0.91(0.03).

Initially, we trained the baseline models on the target datasets. Next, we trained the DTL models - 6L-2T-Freeze, 7L-2T-Freeze, 6L-2T-UNFreeze, and 7L-2T-UNFreeze - equally, as described earlier, to classify breast cancer on the Pima Indians Diabetes dataset. Finally, we trained the feature-based DTL model. Figure5,A displays the model loss for target data before transfer learning, as training epochs progress. Figure5, B showcases the model loss after transfer learning, as training epochs of the target DNN in predicting diabetes on target data continue.

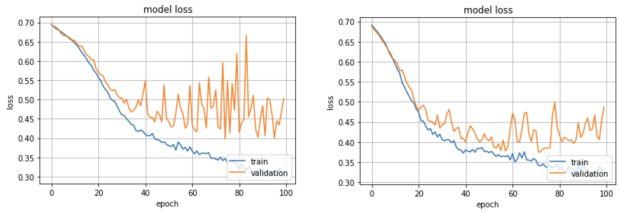

Fig.5 A.BeforeTransfer, B.AfterTransfer

Prior to transfer, the training loss reduces rapidly throughout training epochs, whereas the validation loss initially drops rapidly to 0.45, stabilizes around 0.41 with significant variation, and then slowly starts to increase around 60 epochs. However, following the transfer, both training and validation loss decrease gradually to 0.40 with less variation throughout the training epochs. The losses level off at 0.38 before slowly beginning to increase.

TABLE III
PERFORMANCE OF MODELS ON DIABETES DATASET

Classifiers	AUC	Sens	Acc	F1	AUC-cv	Sens-cv	Acc-cv	F1-cv
Logistic-regression	0.72	0.78	0.77	0.77	0.70	0.77	0.75	0.76
Random-Forest	0.91	0.93	0.93	0.93	0.67	0.73	0.74	0.73
XGBoost	0.97	0.98	0.97	0.98	0.71	0.75	0.77	0.74
DNN-5L	0.78	0.70	0.82	0.72	0.75	0.68	0.79	0.71
DNN-6L	0.80	0.74	0.83	0.75	0.79	0.72	0.81	0.73
DTL-5L-Feature	0.86	0.79	0.88	0.82	**0.85**	**0.80**	**0.86**	**0.81**
6L-2T-Freeze	0.82	0.77	0.83	0.76	0.81	0.76	0.83	0.75
6L-2T-UNFreeze	0.82	0.79	0.85	0.77	0.82	0.75	0.85	0.75
7L-2T-Freeze	0.80	0.75	0.82	0.75	0.80	0.73	0.82	0.74
7L-2T-UNFreeze	0.82	0.78	0.85	0.77	0.80	0.72	0.82	0.74

2) Performance of DTL and Baseline Models: Results indicate that the use of a source model on the diabetes readmission dataset improved target data accuracy to 0.85 via DTL-5L-Feature. Table 3 displays model performance results, with the feature-based DTL model with 5L outperforming all other classification models. Additionally, the 6L-2T-UNfreeze DTL model outperforms baseline classification models.

D. Discussion

The paper highlights the significance of diabetes as a widespread disease with serious health complications that could be avoided through early detection. Despite the extensive research on this topic, the prevalence of diabetes continues to increase, which indicates the need for innovative approaches in predicting and managing this disease. The authors demonstrate that transfer learning on structured data is an effective learning technique that can provide accurate classification when labeled datasets in healthcare are not readily available. They illustrate this approach through a case study on diabetes prediction by transferring features learned from differentiating between diabetic and non-diabetic patients. The unsupervised representation-learning phase of the autoencoder is shown to capture some of the essential items that explain the relationship between input and output, which is useful for predicting different classes from the target dataset.

The study's limitations are highlighted, and they are similar to those of Case Study 1. Overall, the study provides insights into how transfer learning can be leveraged in healthcare to address challenges in predicting and managing diseases like diabetes.

VI. CONCLUSION

Healthcare faces significant challenges in dealing with limited labeled data and imbalanced datasets, demanding urgent attention. Unlike traditional machine learning methods that require manual feature selection and encoding, deep learning allows models to automatically learn relevant features from the data. This study emphasizes the importance of transfer learning, which bridges the gap between source and target domains by leveraging invariant feature representations learned from the source. By applying transfer learning in healthcare, specifically for imbalanced and small labeled datasets, we achieve improved classification performance, as demonstrated in breast cancer and diabetes prediction tasks. Our approach outperforms other machine learning algorithms and provides valuable insights for researchers and professionals in the field, offering state-of-the-art knowledge and guidelines for developing and applying transfer learning in structured healthcare data. The software code for the DTL model implemented in this study can be downloaded from GitHub [1].

ACKNOWLEDGMENT

This work was supported by Mayo Clinic and multiple funding sources including National Institute of Diabetes and Digestive and Kidney Diseases, National Science Foundation grants and National Institutes of Health grant. The authors are solely responsible for the content and it may not represent the official views of the National Institutes of Health.

REFERENCES

[1] Kwak GH, Hui P. DeepHealth: Deep Learning for Health Informatics reviews, challenges, and opportunities on medical imaging, electronic health records, genomics, sensing, and online communication health. arXiv preprint arXiv:1909.00384. 2019 Sep.

[2] Alqaaidi SK, Bozorgi E, Kochut KJ. Multiple Relations Classification using Imbalanced Predictions Adaptation. arXiv preprint arXiv:2309.13718. 2023 Sep 24.

[3] Farhadi A, Chen D, McCoy R, Scott C, Miller JA, Vachon CM, Ngufor C. Breast cancer classification using deep transfer learning on structured healthcare data. In2019 IEEE International Conference on Data Science and Advanced Analytics (DSAA) 2019 Oct 5 (pp. 277-286). IEEE.

[4] Wasikowski M, Chen XW. Combating the small sample class imbalance problem using feature selection. IEEE Transactions on knowledge and data engineering. 2009 Oct 2;22(10):1388-400.

[5] Olson JE, Sellers TA, Scott CG, Schueler BA, Brandt KR, Serie DJ, Jensen MR, Wu FF, Morton MJ, Heine JJ, Couch FJ. The influence of mammogram acquisition on the mammographic density and breast cancer association in the Mayo Mammography Health Study cohort. Breast Cancer Research. 2012 Dec;14:1-2.

[1] https://github.com/AydaFarhadi/TransferLearning