Transfer learning and data augmentation for mortality predictive models in kidney disease

Edwar Macias*, Jose Ibeas †, Javier Serrano *, Jose Lopez Vicario*, Antoni Morell*
*Wireless Information Networking (WIN) group **
Nephrology Department, Institut de Investigació i Innovació Parc Taulí I3PT †
Universitat Autònoma de Barcelona (UAB)
08193 Bellaterra, Spain
{edwar.macias, javier.serrano, jose.vicario, antoni.morell}@uab.cat
{jibeas@tauli.cat}

***Abstract*—Deep learning is becoming a fundamental piece for the paradigm shift from evidence-based medicine to data-based medicine. However, its learning capacity is rarely exploited when working with small data sets. This issue, along with data imbalance, affects the performance in predictive models of mortality using the follow-up of patients in end-stage renal disease (ESRD). Such drawbacks can be addressed by integrating a transfer learning approach to transfer knowledge from an auxiliary domain. We transfer information from patients with acute kidney injury (AKI) from the massive MIMIC-III database to ESRD in the proposed method. Increasing samples in ESRD allows to benefit from the predictive capacity of DL-based models and reduce the effect of data imbalance. In the proposed approach, autoencoders are trained in both domains. Then, latent data representations are extracted. Both domains are then linked through a mapping matrix that relates their latent representation. With this matrix, it is possible to transfer samples from AKI to ESRD. The proposed method is evaluated in several scenarios in which both the latent spaces and the percentage of data imbalance are modified. The experiments have shown that increasing the number of samples implies a significant improvement in the predictive models when the class with imbalance is included. Finally, the proposed approach is compared with the Constructive Covering Algorithm, an improved version of SMOTE that is the most common strategy to deal with data imbalance. The proposed method offers better performance.**

***Index Terms*—Transfer learning, deep learning, autoencoder, mortality prediction, kidney disease**

I. Introduction

In the Big Data era, deep learning (DL) is becoming a fundamental piece in the paradigm shift from evidence-based medicine to data-based medicine [1]. DL exploits complex relationships through a latent representation of data and support decisions in the medical field [2]. DL has significantly impacted medical applications supported by large amounts of data [3]. However, in pathologies with a small volume of data, DL performance is not exploited. This effect usually occurs in specialized hospital units with few patients and whose volume of information is small compared to the hospital itself or other medical institutions. Another obstacle that DL challenges is the generalization problem in learning tasks with imbalanced data. Examples of these tasks include pathology prediction [4], rare event detection [5], or mortality prediction [7]. In such studies, a learning task is usually performed to predict outcomes of patients where there are many more samples for one class than the other ones. For instance, in a previous work of mortality prediction in patients in end-stage renal disease (ESRD) [7], only the last samples from the follow-up of the patients were part of the deceased class, generating a class imbalance in the range of 76-94%. Some of the alternatives to reduce the impact of this issue are based on the generation of samples of the unbalanced class or using oversampling of the minority class and downsampling of the majority one [8], [9]. Another alternative that tackles both issues: the amount of data and the data imbalance, is to support the learning task with related diseases from other information sources. This transfer of information from one domain to another is known as transfer learning (TL) [10].

There is a growing interest in TL in medicine [6]. For instance, in medical image analysis, artificial neural networks (ANN) are trained in the source domain, then fine-tuned with data in the target domain and perform a learning task [11]–[13]. Although a pre-trained ANN is extremely useful, this approach is not suitable for augmenting samples in a data sets with samples from another domain. On the other hand, specialized ANNs can extract complex relationships in latent representations of data. Such relationships are extracted in the hidden layers of ANNS. Thus, latent representations could be employed to find a bridge between domains, and this bridge could be used to transfer knowledge from one domain to another. Several approaches from the literature have used autoencoders (AE) to carry out this task [14], [15]. Minimization between latent representations may be the bridge between transferring samples from one domain to another.

In this work, we propose to transfer medical information from the medical information mart for intensive care III (MIMIC III) database [16], to a data set from a nephrology unit. Information about in-hospital mortality from patients with acute kidney injury (AKI) is transferred to support the prediction of mortality in ESRD patients [7]. The transfer mechanism used in this work is a reinterpretation of the proposed work in [15]. In our work, we use a mapping matrix

This work is supported by the Spanish Government under Project TEC2017-84321-C4-4-R co-funded with European Union ERDF funds and also by the Catalan Government under Project 2017 SGR 1670.

between domains and adjust the approach in [15] to transfer information between domains. We change the learning task and take it to a clinical setting. The main contributions of our work are the following: (1) tackle the problem of data imbalance through a solution based on TL and (2) improve the learning capacity of the DL-based models and enhance the prediction of mortality in ESRD patients by incorporating knowledge from a massive data source.

In the next section, the background of the proposed method is established. Section III presents the details of the proposed approach. In Section IV, the approach is applied using the clinical data sets, and Section V presents the remarks and conclusion of this work.

II. BACKGROUND

This section contains the necessary components to support the proposed method. Initially, the elements of TL, along with the problem definition, are explained. Then, the knowledge extraction mechanism from AEs is presented and finally, the mechanism on which we have relied the proposed method is described.

A. Problem definition

Given labelled data from the source and target domains, $\mathbf{D}_S = \{(\mathbf{x}_{S_i}, y_{S_i})\}_{i=1}^{n_1}$ and $\mathbf{D}_T = \{(\mathbf{x}_{T_i}, y_{T_i})\}_{i=1}^{n_2}$, respectively, where $\mathbf{x}_{S_i} \in \mathbb{R}^{d_S,1}$ and $\mathbf{x}_{T_i} \in \mathbb{R}^{d_T,1}$ are the data and y_{S_i} and y_{T_i} their labels. n_1 and n_2 refer to the total of samples and d_S and d_T their features. We aim to improve the learning task in $\mathbf{D}_T$ by increasing the number of samples and tackling the data imbalance problem. This goal is carried out through transferring knowledge from $\mathbf{D}_S$ to $\mathbf{D}_T$. Knowledge transfer is achieved through a transfer of samples from one domain to another computing a feature mapping matrix $\mathbf{G}$. $\mathbf{G}$ maps latent representations from one domain to the other one. Thus, a sample $\mathbf{x}_S^*$ can be transferred to $\mathbf{D}_T$ through $\mathbf{G}(\mathbf{h}_S^*)$, where $\mathbf{h}_S^*$ is the latent representation of $\mathbf{x}_S^*$. This increase in samples and class balance may reinforce the learning task in $\mathbf{D}_T$.

B. Autoencoders

An AE is a type of ANN that replicates input data $\mathbf{x}$ to the output of the network $\mathbf{x}'$ with a minimum error. This mechanism allows extracting the most representative relationships from the data in its latent space, the so-called code. AEs have an encoding function, $e(\cdot)$, which is the portion of the ANN that extracts knowledge in its code $\mathbf{h} = e(\mathbf{x})$, and the decoding function, $d(\cdot)$, in charge of reconstructing the input, $\mathbf{x}' = d(\mathbf{h})$. The components of an AE can be appreciated in Fig. 1.

To find the minimum error, the input of the AE is forward propagated through the network. Each unit combines the outputs of the previous layer linearly and its output is modified by a non-linear function, in other words,

$$a_j^l = f\left(\sum_{i=0}^{N_{l-1}} w_{i,j}^{l-1} a_i^{l-1}\right), \tag{1}$$

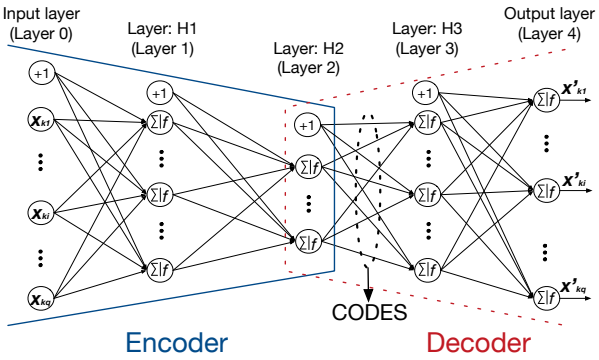

Fig. 1. Structure of an autoencoder with three hidden layers.

with an ANN with L layers ($l = 1, \ldots, L$). N_l are the units in layer l. a_j^l represents the activation of the unit j in layer l and $w_{i,j}^{l-1}$ the weight that connect it with unit i.

Once the propagations reach the output layer, a cost function $\mathcal{L}$ is computed, the weights of the ANN are updated with the gradient of the error through the network following the back-propagation algorithm [17]. In this work mean squared error is used as a lost function,

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{x}_i - \mathbf{x}_i'\right)^2, \tag{2}$$

where $\mathbf{x}_i$ represents a sample i and N is the total samples in a dataset.

C. Hybrid heterogeneous transfer learning

The so-called Hybrid Heterogeneous Transfer Learning (HHTL) proposed in [15], uses the labelled data from the $\mathbf{D}_T = \{(\mathbf{x}_{T_i}, \mathbf{y}_{T_i})\}_{i=1}^{n_2}$ to assign labels to the unlabelled data from the $\mathbf{D}_S = \{\mathbf{x}_{S_i}\}_{i=1}^{n_1}$. They solve this learning task in two stages. In the first one, they train AEs with k ($k = 1, \ldots, K$) layers in both domains. Then they extract latent representations of each hidden layer of the AEs, $\mathbf{H}_{S,1}, \ldots, \mathbf{H}_{S,K}$ and $\mathbf{H}_{T,1}, \ldots, \mathbf{H}_{T,K}$. Finally, they compute a mapping matrix of latent representations $\mathbf{G}_k$ between the domains, minimizing the objective:

$$\underset{\mathbf{G}_k}{min} \left\|\mathbf{H}_S - \mathbf{G}_k\mathbf{H}_T\right\|^2 + \lambda\left\|\mathbf{G}_k\right\|^2. \tag{3}$$

In the second stage they generate a new feature space with the latent representations of the target data $\mathbf{Z}_T = \left[\mathbf{H}_{T,1}^\top \ldots \mathbf{H}_{T,K}^\top\right]^\top$, and train a classifier with $\{(\mathbf{Z}_T, \mathbf{y}_T)\}$. Then use $\mathbf{G}_k$ to transfer samples $\mathbf{X}_S^*$ from $\mathbf{D}_S$ to the latent spaces of $\mathbf{D}_T$, $\mathbf{H}_{S\to T}^*$. With the latent transferred representations, a feature space $\mathbf{Z}_{S\to T} = \left[\left(\mathbf{G}_1\mathbf{H}_{S\to T,1}^*\right)^\top \ldots \left(\mathbf{G}_k\mathbf{H}_{S\to T,k}^*\right)^\top\right]^\top$ is created. Finally, with the trained classifier they predict over $\mathbf{Z}_{S\to T}$ the labels for $\mathbf{D}_S$ samples, where $S \to T$ refers to the transfer from $\mathbf{D}_S$ to $\mathbf{D}_T$.

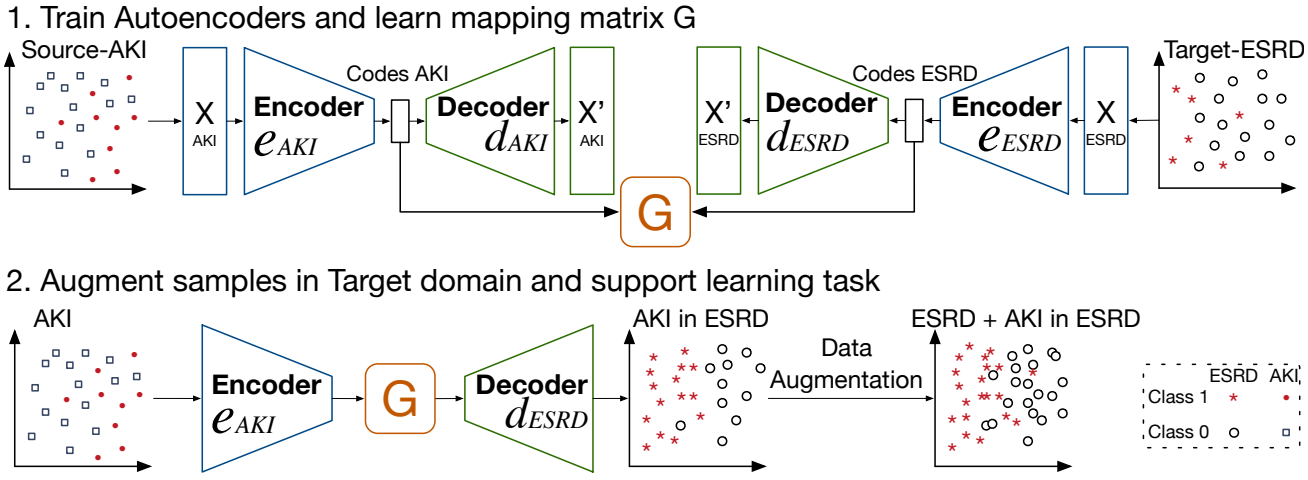

Fig. 2. Scheme of proposed method for transfer of samples between domains and the support of a learning task in the target domain.

III. PROPOSED METHOD

In this work, we propose to apply a TL approach to increase the number of samples in $\mathbf{D}_T$ using $\mathbf{D}_S$. This mechanism is carried out with the twofold purpose of tackling data imbalance and improving the predictive capacity of DL models with the augmented dataset in $\mathbf{D}_T$. Both domains, $\mathbf{D}_S$ and $\mathbf{D}_T$, contain labelled data. Fig. 2 presents an overview of the proposed method. The approach is divided into two stages. In the first one, unlike HHTL, we have labelled data. Thus, inspired by HHTL, we compute $\mathbf{G}$ using the codes from trained AEs in each domain. In these codes, the latent representations of the input data of each domain are extracted. In the second stage, $\mathbf{G}$ is used to transfer codes, $\mathbf{H}_S^*$, produced by data $\mathbf{X}_S^*$ in $\mathbf{D}_S$, to codes in $\mathbf{D}_T$. Then, the decoder function in $\mathbf{D}_T$ reconstructs the transferred codes in such domain. Thus, the samples in $\mathbf{D}_T$ are increased. As mentioned previously, this last step allows to improve the capacity of the DL-based learning model and tackle the imbalance issue. The steps to perform the proposed method are presented in Algorithm 1.

Algorithm 1: Proposed method

Input: Data from both domains, $\lambda = 0.001$:
$\mathbf{D}_S = \{(\mathbf{x}_{S_i}, y_{S_i})\}_{i=1}^{n_1}$, $\mathbf{D}_T = \{(\mathbf{x}_{T_i}, y_{T_i})\}_{i=1}^{n_2}$

1 Train AEs with $\mathbf{X}_S$ and $\mathbf{X}_T$. Extract encoder (e) and decoder (d) functions from both domains, and the latent representations-codes $\mathbf{H}$:
$\mathbf{H}_S = e_S(\mathbf{X}_S)$, $\mathbf{X}'_S = d_S(\mathbf{H}_S)$
$\mathbf{H}_T = e_T(\mathbf{X}_T)$, $\mathbf{X}'_T = d_T(\mathbf{H}_T)$;

2 Learn heterogeneous feature mapping $\mathbf{G}$:
$\min_{\mathbf{G}} \|\mathbf{H}_S - \mathbf{G}\mathbf{H}_T\|^2 + \lambda \|\mathbf{G}\|^2$;

3 Augment samples in $\mathbf{D}_T$ with samples from $\mathbf{D}_S$:
$\mathbf{X}^*_{S\to T} = \mathbf{G}^\top \mathbf{X}^*_S$
Note: $S \to T$ means the transfer from $\mathbf{D}_S$ to $\mathbf{D}_T$.
$\mathbf{X}^*_T = [\mathbf{X}_T \quad \mathbf{X}^*_{S\to T}]$, $\mathbf{y}^*_T = [\mathbf{y}_T \quad \mathbf{y}_S]$

4 Train a classifier f with $\{(\mathbf{X}^*_T, \mathbf{y}^*_T)\}$

Output: Classifier f

IV. EXPERIMENTAL RESULTS

In this section, the performance of the proposed approach is evaluated in a clinical domain. Both domains are related to the mortality of patients with kidney diseases. The task to support in the target domain is the prediction of mortality at 1, 2, 3 and 6 months. Details about the used datasets, the experiments and the results are presented in the rest of this section. All the reported experiments were repeated 10 times using 5-folds for cross-validation. Thus, the test data of each fold is only used when evaluating the mortality prediction models trained with the augmented data.

A. Datasets

Information for the target domain is part of a previous study of prediction of mortality in ESRD patients [7]. The study cohort contains 8229 samples with 53 variables from the monthly follow-up of 261 patients during the evolution of their disease until the deceased event. The dataset is a mixture of categorical and continuous variables that include information about demographics, laboratory tests, diagnoses and variables measured during the haemodialysis sessions.

The dataset for the source domain has been extracted from MIMIC-III database [16]. From this massive database, those patients with acute kidney injury (AKI) were filtered based on the kidney disease improving global outcomes (KDIGO) clinical practice guideline [18]. The total cohort contains 4152 samples and 23 features. These features are also a mixture of categorical and continuous measurements of the health condition of patients in intensive care units (ICU). Their follow-up includes demographics, diagnoses, laboratory tests, physiological measurements during the ICU stay and the in-hospital mortality label.

In terms of data imbalance, AKI contains 1565 samples from patients who deceased in ICU. For ESRD data, the label varies based on the mortality range to be predicted.

Information on how this label is computed can be found in the previous work [7]. Table. I shows the information on the class imbalance in ESRD.

TABLE I
IMBALANCE OF SAMPLES FOR THE PREDICTION OF MORTALITY IN PATIENTS IN ESRD. **CLASS 0** AND **CLASS 1** REFER TO SAMPLES IN ALIVE AND DECEASED CLASSES, RESPECTIVELY.

Mortality	Class 0	Class 1	Imbalance (%)
1	7734	495	93.6
2	7488	741	90.1
3	7251	978	86.5
6	6632	1597	75.9

B. AEs training and TL mechanism

Initially, an AE is trained in each domain. From both AEs, their latent representations are extracted. Two AEs with two hidden layers were trained for both datasets. The hyperbolic tangent (Tanh) activation function was used for the hidden layers and the sigmoid for the output layer in AKI. For the ESRD dataset, rectified linear unit (ReLU) activation function for hidden layers and Sigmoid at the output layer were used. For both AEs, a dropout of 0.1, and batch normalization were applied in the hidden layers of the networks to avoid overfitting. Once the AEs are trained, the mapping matrix $\mathbf{G}$ is generated using the latent representations from both domains. Then, the latent representation of AKI data is transferred to the latent space of the ESRD domain using $\mathbf{G}$. Finally, transformation is reconstructed using the decoding function of the trained AE in ESRD.

C. Experiments

The performance of the proposed method is evaluated on the learning task in ESRD. The area under the receiver operating characteristic (AUROC) curve is measured. AUROC relates the sensitivity and specificity of a classifier. Its values are between 0 and 1, with 1 being the perfect classifier and 0.5 a random one. The baseline performance and classifiers used in this work are based on the ones implemented in [7]. Three experiments have been defined to determine the performance of the proposed method.

1) Tunning latent representation dimensions: In the first experiment, the dimensions of the latent representations in both domains are evaluated. Thus, the combination of dimensions that presents the best overall performance for the prediction task is empirically found. In Fig. 3 , S_* and T_* refer to the dimensions of the codes in AKI and in ESRD, respectively. It can be appreciated that most of the combinations present a higher performance than the baseline one. However, the best combination of dimensions in the codes is 30 and 80 units in AKI and ESRD, respectively.

2) Increasing samples in ESRD: This experiment evaluates how the increase of samples in the training set affects the predictive models of mortality in ESRD. For this experiment, three possible scenarios were defined. In the first scenario, the data imbalance in ESRD is intentionally increased. Thus, only AKI Class 0 samples are transferred to the ESRD training set. This transfer is carried out to evaluate whether there is a negative effect linked to the increase in data imbalance. In the second scenario, the training set samples are increases, but only those that belong to AKI Class 1 are transferred. In this case, the aim is to balance the imbalanced class. Finally, both cases are combined in a third scenario. Therefore, we seek to evaluate both the effect of the increase in samples and the reduction of the data imbalance in the predictive models. Table. II shows how the data imbalance varies for each scenario.

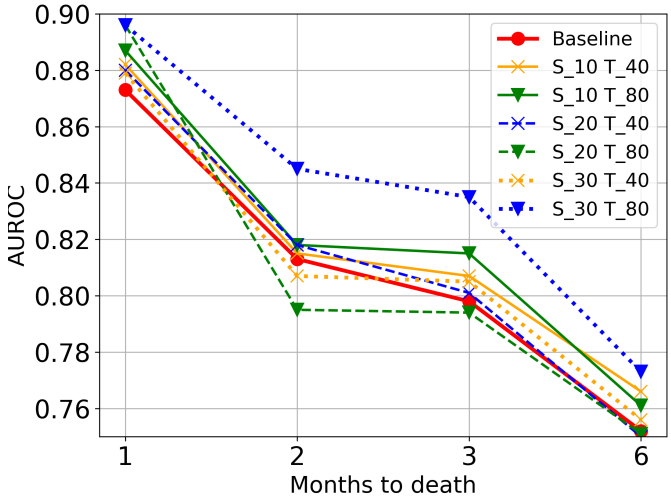

Fig. 3. Mortality prediction varying the dimension of latent representations in source (S) and target (T) domain.

TABLE II
IMBALANCE IN ESRD GENERATED BY INCREASING TRAINING SAMPLES IN ESRD FROM AKI. SCENARIO 1, 2 AND 3 REFER TO THE TRANSFER OF SAMPLES FROM THE CLASS 0, 1 AND COMBINING BOTH CLASSES, RESPECTIVELY.

Mortality	Generated data imbalance (%)		
	Scenario 1	Scenario 2	Scenario 3
1	95.2	73.4	80.0
2	92.6	69.2	77.1
3	90.1	74.9	74.2
6	82.7	52.3	65.7

In Fig. 4 it can be appreciated that increasing samples in the training set of the ESRD data does not imply, in any of the scenarios, a deterioration in the predictive models. On the other hand, when the number of samples is increased, considering the imbalance issue, the learning models present a better predictive capacity.

3) Comparison with CCA-SMOTE: Finally, the proposed method is compared with the so-called Constructive Covering Algorithm (CCA) [9]. A technique that improves the widely used Synthetic Minority Oversampling Technique (SMOTE) [8], incorporating a mechanism based on ANN to delete the hard to learn samples. The imbalance of training samples for CCA-SMOTE is adjusted following the third scenario in Table. II for a fair comparison. It can be appreciated in Fig. 5 that for

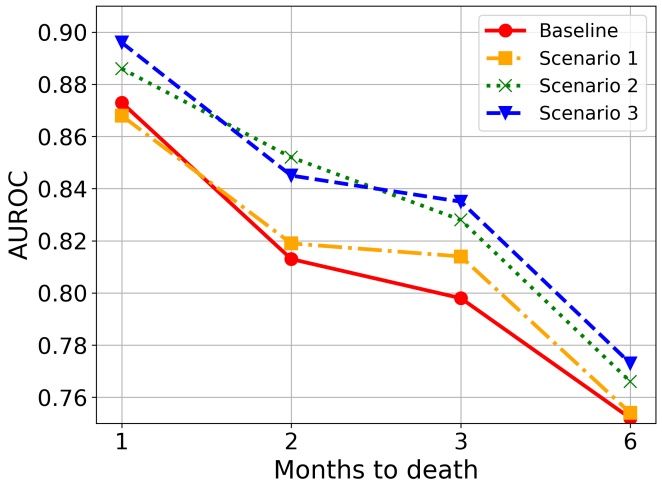

Fig. 4. Imbalance performance by increasing the samples in individual and both classes.

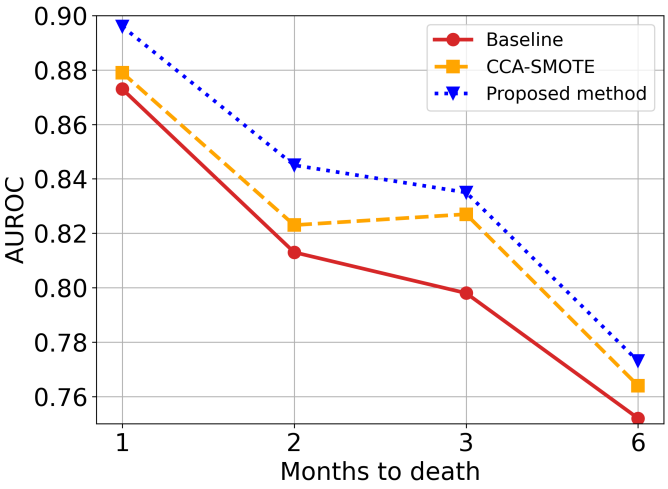

Fig. 5. Comparison of proposed method and CCA-SMOTE.

most of the predictors, CCA-SMOTE performs better than the baseline. However, the proposed method outperforms CCA-SMOTE for all the predictive models.

V. CONCLUSION

In this paper, an approach to support the predictive learning task of mortality in ESRD patients based on TL was presented. The TL mechanism made it possible to reduce the imbalance problem, and with the increase in samples, the DL-based models exhibited a better predictive capacity than previous predictive models. The experiments showed that the increase of samples in the target domain positively influences predictive performance. Moreover, when this increase of samples reduces the data imbalance, the improvement in predictions becomes more significant. Finally, as data imbalance is an inherent problem in many chronic pathologies, we recommend applying the proposed approach to transfer massive clinical data to data from relatively small units and support learning tasks in such domains.

REFERENCES

[1] Travers Ching, Daniel S Himmelstein, Brett K Beaulieu-Jones, Alexandr A Kalinin, Brian T Do, Gregory P Way, Enrico Ferrero, Paul-Michael Agapow, Michael Zietz, Michael M Hoffman, et al., "Opportunities and obstacles for deep learning in biology and medicine," *Journal of The Royal Society Interface*, vol. 15, no. 141, pp. 20170387, 2018.

[2] Edwar Macias, Javier Serrano, Jose Lopez Vicario, and Antoni Morell, "Novel imputation method using average code from autoencoders in clinical data," in *2020 28th European Signal Processing Conference (EUSIPCO)*, 2021, pp. 1576–1579.

[3] Francesco Piccialli, Vittorio Di Somma, Fabio Giampaolo, Salvatore Cuomo, and Giancarlo Fortino, "A survey on deep learning in medicine: Why, how and when?," *Information Fusion*, vol. 66, pp. 111–137, 2021.

[4] E Macias, G Boquet, J Serrano, JL Vicario, J Ibeas, and A Morel, "Novel imputing method and deep learning techniques for early prediction of sepsis in intensive care units," in *2019 Computing in Cardiology (CinC)*, 2019, pp. 1–4.

[5] E Macias, A Morell, J Serrano, and JL Vicario, "Knowledge extraction based on wavelets and dnn for classification of physiological signals: Arousals case," in *2018 Computing in Cardiology Conference (CinC)*, 2018, vol. 45, pp. 1–4.

[6] Guo Haixiang, Li Yijing, Jennifer Shang, Gu Mingyun, Huang Yuanyue, and Gong Bing, "Learning from class-imbalanced data: Review of methods and applications," *Expert Systems with Applications*, vol. 73, pp. 220–239, 2017.

[7] Edwar Macias, Antoni Morell, Javier Serrano, Jose Lopez Vicario, and Jose Ibeas, "Mortality prediction enhancement in end-stage renal disease: A machine learning approach," *Informatics in Medicine Unlocked*, vol. 19, pp. 100351, 2020.

[8] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer, "Smote: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.

[9] Yuanting Yan, Ruiqing Liu, Zihan Ding, Xiuquan Du, Jie Chen, and Yanping Zhang, "A parameter-free cleaning method for smote in imbalanced classification," *IEEE Access*, vol. 7, pp. 23537–23548, 2019.

[10] Sinno Jialin Pan and Qiang Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.

[11] Muazzam Maqsood, Faria Nazir, Umair Khan, Farhan Aadil, Habibullah Jamal, Irfan Mehmood, and Oh-young Song, "Transfer learning assisted classification and detection of alzheimer's disease stages using 3d mri scans," *Sensors*, vol. 19, no. 11, 2019.

[12] Michal Byra, Mei Wu, Xiaodong Zhang, Hyungseok Jang, Ya-Jun Ma, Eric Y Chang, Sameer Shah, and Jiang Du, "Knee menisci segmentation and relaxometry of 3d ultrashort echo time cones mr imaging using attention u-net with transfer learning," *Magnetic resonance in medicine*, vol. 83, no. 3, pp. 1109–1122, 2020.

[13] Daniel S Marcus, Anthony F Fotenos, John G Csernansky, John C Morris, and Randy L Buckner, "Open access series of imaging studies: longitudinal mri data in nondemented and demented older adults," *Journal of cognitive neuroscience*, vol. 22, no. 12, pp. 2677–2684, 2010.

[14] Long Wen, Liang Gao, and Xinyu Li, "A new deep transfer learning based on sparse auto-encoder for fault diagnosis," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 49, no. 1, pp. 136–144, 2019.

[15] Joey Tianyi Zhou, Sinno Jialin Pan, and Ivor W. Tsang, "A deep learning framework for hybrid heterogeneous transfer learning," *Artificial Intelligence*, vol. 275, pp. 310–328, 2019.

[16] Alistair EW Johnson, Tom J Pollard, Lu Shen, H Lehman Li-wei, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark, "Mimic-iii, a freely accessible critical care database," *Scientific data*, vol. 3, pp. 160035, 2016.

[17] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams, "Learning representations by back-propagating errors," *nature*, vol. 323, no. 6088, pp. 533–536, 1986.

[18] Lesley A Inker, Brad C Astor, Chester H Fox, Tamara Isakova, James P Lash, Carmen A Peralta, Manjula Kurella Tamura, and Harold I Feldman, "Kdoqi us commentary on the 2012 kdigo clinical practice guideline for the evaluation and management of ckd," *American Journal of Kidney Diseases*, vol. 63, no. 5, pp. 713–735, 2014.