Spatial Knowledge Transfer with Deep Adaptation Network for Predicting Hospital Readmission

Ameen Abdel Hai[1], Mark G. Weiner[2], Alice Livshits[3], Jeremiah R. Brown[4], Anuradha Paranjape[3], Zoran Obradovic[1], Daniel J. Rubin[3]

[1] Computer and Information Sciences, Temple University, Philadelphia, PA
[2] Weill Cornell Medicine, New York, NY
[3] Lewis Katz School of Medicine, Temple University, Philadelphia, PA
[4] Department of Epidemiology and Biomedical Data Science, Geisel School of Medicine at Dartmouth, Hanover, NH
zoran.obradovic@temple.edu, daniel.rubin@tuhs.temple.edu

Abstract. A hospital readmission risk prediction model based on electronic health record (EHR) data can be an important tool for identifying high-risk patients in need of additional support. Performant readmission models based on deep learning approaches require large, high-quality training datasets to perform optimally. Utilizing EHR data from a source hospital system to enhance prediction on a target hospital using traditional approaches might bias the dataset if distributions of the source and target data are different. There is a lack of an end-to-end readmission model that can capture cross-domain knowledge. Herein, we propose an early readmission risk temporal deep adaptation network, ERR-TDAN, for cross-domain spatial knowledge transfer. ERR-TDAN transforms source and target data to a common embedding space while capturing temporal dependencies of the sequential EHR data. Domain adaptation is then applied on a domain-specific fully connected linear layer. The model is optimized by a loss function that combines distribution discrepancy loss to match the mean embeddings of the two distributions and the task loss to optimize predicting readmission at the target hospital. In a use case of patients with diabetes, a model developed using target data of 37,091 patients from an urban academic hospital was enhanced by transferring knowledge from high-quality source data of 20,471 patients from a rural academic hospital. The proposed method yielded a 5% increase in F1-score compared to baselines. ERR-TDAN may be an effective way to increase a readmission risk model's performance when data from multiple sites are available.

Keywords: Transfer Learning, Readmission Prediction, Electronic Health Records data, Machine Learning, Deep Learning.

1 Introduction

Hospital readmission is an undesirable outcome and a driver of high financial costs. Approximately, 20% of Medicare discharges had readmission within 30-days, corresponding to $20+ billion in hospital costs. [1]. Identifying patients with higher risk of readmission would enable the targeting of interventions to those at greatest need, optimizing the cost-benefit ratio.

Previously, we published a risk deep learning (DL) model based on electronic health records (EHR) data collected from an urban academic hospital that predicts the risk of unplanned, 30-day readmission among patients with diabetes. We used a sequential model, long short-term memory (LSTM). Performance was adequate (F-1 Score 0.80), and results showed that this LSTM model can capture temporal dependencies of the EHR data [2].

Performant readmission models based on DL techniques require large, high-quality training data to perform optimally. Utilizing EHR data from a source hospital system to enhance prediction on a target hospital using traditional approaches enlarge dataset bias which might deteriorate performance due to distributional difference of the source and target datasets, resulting in statistically unbounded risk for the target tasks [3]. Traditional approaches are designed for a specific data type, and not capable of generalizing to other temporal data.

Transfer learning approaches have been explored for hospital readmission with the objective to improve learning at the target population by exploiting information from a related source population. In [4, 5], classical transfer learning was employed to address data scarcity using a relevant source dataset. In [6], classical transfer learning techniques were explored as to what extent can transfer learning benefit learning on target tasks by fine-tuning pre-trained models in the healthcare domain. However, there is still a need for an end-to-end model to perform cross-domain spatial knowledge transfer and predictive learning in a unified learning framework while capturing temporal dependencies for hospital readmissions.

In this paper, we propose an early readmission risk temporal deep adaptation network, ERR-TDAN, to perform cross-domain spatial knowledge transfer from EHR data of different sites and perform predictive learning. Deep Adaption Network (DAN) utilizes deep convolutional neural network (CNN) and generalizes it to the domain adaptation setting through learning transferable latent features between source and target domains for computer vision tasks [3, 7]. Motivated by the success of DAN in numerous transfer learning tasks in computer vision, we employed the idea of learning transferable features of temporal data by matching the source and target domain distributions in the latent feature space. We tailored it for hospital readmission using EHR data and optimized for the target task.

The aims of this study were as follows: 1) To develop a hospital readmission framework using EHR data that transfers knowledge between a rural academic hospital and an urban academic hospital to enhance predictions on the urban academic hospital. 2) To study the optimal amount of retrospective EHR data needed for future predictions. 3) To study the duration of optimal performance. Experiments conducted show that ERR-TDAN can enhance hospital readmission prediction.

2 Deep Adaptation Network

Domain adaptation is a form of transfer learning commonly used in computer vision to address the problem of learning using data from two related domains but under different distributions [3, 8]. Domain adaptation can help improve the performance of a model by learning transferable features to minimize the gap between the source and target domains in an isomorphic latent feature space. DAN generalizes deep CNN for computer vision applications to utilize domain adaptation techniques to learn transferable feature representation in the latent embedding space [7]. Motivated by the success of DAN in various computer vision tasks [9-11], we utilized the idea of DAN for transferring cross-domain spatial knowledge tailored for predicting hospital readmission on EHR data and optimized to enhance predictions on the target, rather than generalizing on both domains. A direct comparison to DAN is not applicable since DAN is modified for computer vision tasks using CNN. CNNs capture spatial correlations and are unable to capture temporal correlations of EHR data [3]. Thus, we employed the idea of DAN and tailored it for hospital readmission on EHR data to capture temporal dependencies using LSTM layers, establish cross-domain knowledge transfer, and optimized it for the target task using a customized loss function.

3 The Proposed ERR-TDAN Framework

An early readmission risk framework based on temporal deep adaptation network was developed to enhance prediction on the target data collected from Temple University Hospital System (TUHS) by establishing spatial knowledge transfer from a source data with higher quality features collected from Penn State University Hospital System (PSUHS). The model was developed using data as defined by the National Patient-Centered Clinical Research Network (PCORnet) Common Data Model (CDM) [12].

We applied a hospital readmission LSTM model that we previously published using EHR data collected from TUHS [2]. When trained on TUHS data and tested on the following year TUHS data this model F-1 score was 0.80. We trained and tested the same method on EHR data collected from PSUHS, where performance was better (F1-score 0.91). The 11% increase in F-1 score was achieved since EHR data from PSUHS contained fewer missing data, denser features, and less erroneous data. However, training and evaluating the same method on data from both domains affected the performance (F-1 score 0.79) since the model struggled to generalize and converge due to training data drawn from different distributions. To address this limitation, we employed the idea of DAN, tailored for hospital readmission on EHR data that captures temporal correlations and enhances target prediction through learning transferable features via domain-specific fully connected linear layers to

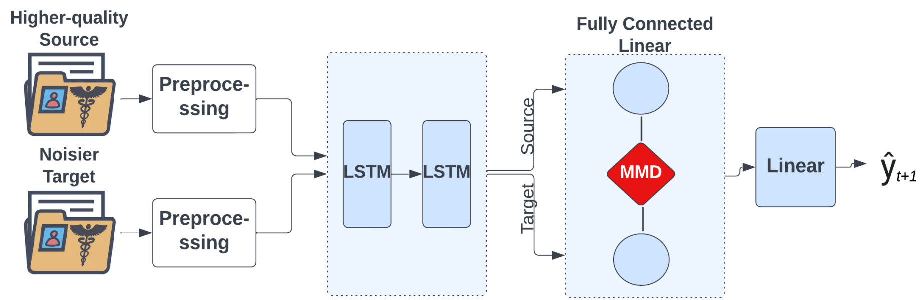

Fig. 1. The proposed method framework, ERR-TDAN. It comprises of three main processes. 1) LSTM layers learn hidden representation of the input of both source and target domains. 2) Deep adaption network structure with fully connected layers is constructed to match the mean embeddings of different domains drawn from different distributions. 3) The matched embeddings are then passed to a fully connected layer with sigmoid function for binary classifications. The model is optimized through a customized loss function that penalizes on domain discrepancy of both source and target, and task loss to optimize for the target task.

explicitly reduce the domain discrepancy. DAN generalizes on both domains for computer vision tasks, whereas in our study we tailored this technique for hospital readmission using EHR data and optimized on the target task, instead of generalizing on both domains. To accomplish this, the hidden embeddings of the domain-specific layers are embedded to a reproducing kernel Hilbert space through maximum mean discrepancy (MMD), to match the mean embeddings of two domain distributions. The model was optimized via a customized loss function.

Fig. 1 presents the proposed framework, ERR-TDAN which consists of the following main processes: 1) LSTM's input was data from both source and target to learn hidden representation to map source and target data to a common embedding while capturing temporal dependencies of the EHR data. 2) To match the embedding distributions of the source and target domains, deep adaptation network scenario is established through fully connected linear layers constructed to match the mean embeddings of different domain distributions. The hidden representation is embedded through a reproducing kernel Hilbert space to transfer knowledge and bridge the gap between two distributions via MMD to reduce domain discrepancy. 3) The matched embeddings are then passed to a fully connected layer with a sigmoid function to classify if a patient is likely to be readmitted or not. In backpropagation, we optimize the model on the target domain using a customized loss function that combines the domain discrepancy loss and binary cross entropy loss. The following sections illustrate the framework in more detail.

3.1 Representation Learning of Temporal EHR Data with LSTM

Initially, we utilized LSTM with two recurrent layers to form a stacked LSTM to learn hidden data representation embedded to a common latent feature space of the temporal EHR data of the source and target domains. LSTM, a sequential model capable of capturing temporal correlations, is commonly used for sequential tasks and is proven to be effective for hospital readmission using EHR data [2, 13]. LSTM takes as an input a 3-dimensional tensor of stacked source and target data. LSTM is structured based on basic neural network, but neurons of the same layer are connected, enabling a neuron to learn from adjacent layers, in addition to learning from outputs of the previous layers and the input data. Hence, neurons include two sources of inputs, the recent past and the present. A dropout of 0.1 was applied between the first and second LSTM layers. To add nonlinearity, we utilized ReLU activation function on the output of LSTM (embeddings), formulated as follows:

$$b^t = \mathrm{ReLU}\,(b + Wh^{t-1} + Ux^t) \tag{1}$$

3.2 Learning Transferable Features and Predictions

The output b^t is then fed to domain-specific fully connected linear layers with deep adaptation network setting. Domain discrepancy is reduced by matching the mean embeddings of the source and target distributions. Hidden representation of the linear layers embedded through a reproducing kernel Hilbert space to bridge the gap between two distributions and transfer knowledge via MMD. MMD measures the distance of the source and target distributions in the embedding space. MMD distance measure was originally used to determine whether two samples are drawn from the same distribution and measures how distant the samples are [14]. In this study, we utilized MMD to learn transferable features between source and target domains to enhance prediction on the target. MMD was utilized as one of the two components of the loss function to minimize the domain discrepancy. The loss function is explained in more detail in the next section. MMD is defined as:

$$MMD_{loss}(\mathcal{D}^{\mathcal{S}}, \mathcal{D}^{\mathcal{T}}) = \left\| \frac{1}{n} \sum_{i=1}^{n} \phi\left(d_i^{\mathcal{S}}\right) - \frac{1}{m} \sum_{j=1}^{m} \phi\left(d_j^{\mathcal{T}}\right) \right\|_{\mathcal{H}}^{2}, \tag{2}$$

where $\mathcal{D}^{\mathcal{S}}$ and $\mathcal{D}^{\mathcal{T}}$ denote source and target data respectively, ϕ denotes the Gaussian kernel function, $\mathcal{H}$ denotes the Hilbert space, and n and m denote the number of observations of the source and target sets, respectively. The temporal embeddings of the LSTM are then fed into fully connected layers with MMD loss to measure the distance between two distributions and reduce domain discrepancy.

Prediction. The matched embeddings are then fed into a linear layer with output of 1 with sigmoid activation function for predictions $\hat{y}$ [2].

3.3 Model Optimization via a Customized Loss Function

We tailored the loss function for hospital readmission on the target domain by combining Binary Cross Entropy (BCE) loss to measure the error of reconstruction, applied on the target task only, and MMD loss applied on both source and target to reduce domain discrepancy. Since the aim of this study is to enhance prediction on the target domain using higher-quality source data, we reduced the weight of the MMD loss via the penalty parameter γ and optimized the loss on the target domain. Loss function used in the proposed ERR-TDAN model is defined as follows:

$$
\begin{aligned}
BCE_{loss} = (x, y) = L = \{l_1, \dots, l_N\}^{\top}, l_n \\
= -w_n[y_n \cdot \log x_n + (1 - y_n) \cdot \log(1 - x_n)], \\
TOTAL_{loss} = \frac{1}{L^N} \sum_{t=1}^{L^N} (BCE_{loss}(x, y) + \gamma \, MMD_{loss}(d^{\mathcal{S}}, d^{\mathcal{T}})),
\end{aligned}
\tag{3}
$$

where x and y are the predictions and ground truth for a given batch respectively. L denotes loss. N is the batch size, w is a rescaling weight given to the loss of each batch element, γ is the penalty parameter of domain discrepancy. To optimize for the target task, we determined empirically that 0.5 value of γ is appropriate.

4 Data

We collected data from an urban academic hospital, TUHS, and a rural academic hospital, PSUHS, between July 1st, 2010, and December 31st, 2020. We extracted data on encounters, demographics, diagnosis, laboratory tests, medication orders, procedures, and vital signs. in the cohort of patients with diabetes was defined as previously described [2]. Data preprocessing, handling of missingness of data, different number of recordings per encounter, learning embeddings to reduce dimensionality, address sparse feature vectors, and data representation were performed as presented in [2]. Additional features were aggregated to assist with learning temporal dependencies, including duration of stay in days, and number of days since the prior encounter.

We obtained a total of 1,421,992 encounters corresponding to 20,471 patients for PSUHS, and a total of 3,023,267 encounters corresponding to 37,091 patients for TUHS. The class distributions were as follows. TUHS: 28,107 for the negative class (no readmission), and 8,984 for the positive class (readmitted within 30-days); PSUHS: 18,775 for the negative class and 1,696 for the positive class.

The characteristics of the samples from the two sites were different. For instance, 4.9% of patients were Hispanic at PSUHS, whereas TUHS contained large Hispanic population of 22%. Other differences included race and tobacco use. The numbers of unique ICD-9 and ICD-10 codes, and vital recordings at PSUHS were larger than that at TUHS.

Patient encounters were sequentially ordered by admission date and represented in a 3-dimensional tensor for the LSTM model, where each patient's data is represented as a 2-

dimensional matrix in which features of each encounter are represented in a 1-dimensional array while a second dimension represents different hospitalizations of that patient. The third dimension is used to encode hospitalization information of different patients.

5 Experimental Setup and Results

We hypothesize that it is feasible to enhance readmission predictions on target data of TUHS using a source data from a relevant domain under different distribution. In this section, we conduct extensive experiments to evaluate the performance of the proposed model, ERR-TDAN and compare it to baselines. F-1 score, precision, recall (sensitivity), specificity, and accuracy were used to evaluate the model's performance [15]. We randomly selected different patients for training and testing. Experiments were iterated 10 times; results were presented based on the mean and two-sided 95% confidence interval (CI). Moreover, we address the following research questions to evaluate optimal performance of the model.

5.1 Can we enhance readmission risk prediction for a target hospital by utilizing data from another hospital?

We randomly split TUHS and PSUHS data to 70% training, 10% validation, and 20% testing. Then, we concatenated training data of both domains, and fed to the ERR-TDAN. We tested the model on TUHS using 7,418 patients, of whom 1,557 had a readmission.

Table 1 presents a comparative analysis to evaluate the proposed method, ERR-TDAN compared to alternative baselines. Table 1 shows that ERR-TDAN yielded a 5% increase in F1-score when compared to a model we previously published for hospital readmission on EHR data collected from TUHS, and 3% increase using a generalized version of ERR-TDAN (G-ERR-TDAN) of the domain adaptation framework with MMD loss without optimizing on the target task. G-ERR-TDAN results provide evidence that optimizing on the target task enhances target's predictions is superior to generalizing on both domains.

Table 1. Performance of the proposed method, ERR-TDAN and three alternatives tested on the target domain (TUHS) enhanced by a related source data (PSUHS). The Average F1, Recall/Sensitivity, Specificity, and accuracy and their corresponding two-sides 95% confidence interval (CI) on 10 experiments on training and testing patients' data selected completely at random.

Model	Train	F1-score	Recall	Specificity	Accuracy
[2]	TUHS	0.80 ±0.003	0.81 ±0.002	0.94 ±0.010	0.81 ±0.002
LSTM	TUHS + PSUH	0.79 ±0.007	0.81 ±0.006	0.95 ±0.008	0.81 ±0.005
G-ERR-TDAN	TUHS + PSUH	0.82 ±0.001	0.81±0.001	0.92 ±0.002	0.81 ±0.001
ERR-TDAN	TUHS + PSUH	**0.85** ±0.002	0.84 ±0.002	0.91 ±0.003	0.84 ±0.001

5.2 What is the retrospective optimal amount of EHR data needed for future predictions?

We conducted extensive experiments to find the optimal amounts of patient's historical data needed for the model to perform optimally. Our objective was to determine a size of training data so that further enlargements do not improve predictions of hospitalization risk. The model was trained on varying t and tested on $t + x$, where t denotes a period in the past and $t + x$ denotes a period in the future. For a fair comparison, $t + x$ was a fixed test dataset of 2020, and trained on varying training sets of t, including 6 months (July-December of 2019), 1 year (2019), 2 years (2018-2019), 3 years (2017-2019), 4 years (2016-2019), and 5 years (2015-2019) look-back time. For instance, training on 2019, and testing on 2020 (1 year look-back) to test if learning on 1 year of historical EHR data from the past is sufficient to perform optimally.

Fig. 2 (left) shows that 1 year of historical data are optimal to predict readmission since it yielded the highest F1-score with least amounts of data required.

5.3 How often do we need to retrain the model to achieve optimal performance?

Concept and covariate shifts are one of the major reasons model performances degrade overtime. Monitoring data drift helps avoid performance degradation. Thus, we conducted experiments to study the lifetime of the proposed model. Based on the optimal look-back time of question 2, we trained the model on EHR data collected in 2015 and tested it with 1, 2, 3, 4, and 5 future gaps. For instance, training on data collected in 2015 and testing in

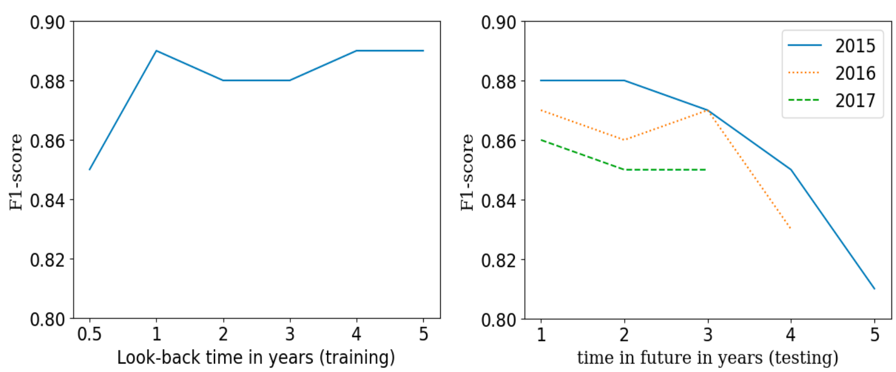

Fig 2. (left) Presents the retrospective optimal amount of EHR data needed for future predictions. Results show that 1 year of historical data are sufficient to predict hospital readmission from the leading year. (right) Presents the lifetime of the model to maintain and achieve optimal performance. Three different models were developed on data collected from 2015, 2016, and 2017 to predict future instances with 1 to 5 years gap. Results show that to maintain optimal performance, the proposed framework may benefit from retraining every three years.

2020 to experiment if the model's performance would degrade after 5 years. We iterated this over various models trained on 1 year of data collected in 2015, 2016, and 2017 and tested for readmissions on future instances.

Fig. 2 (right) shows that F1-score decreased over time due to data drift. Performance was relatively stable when tested on 1 and 2 years in the future. There was a significant decrease in F1-score when used to predict readmissions with 3 years gap between training and testing. On overage, F1-score degraded 0.6% when used 3 years later, 3.5% when used 4 years later, and 7% when used 5 years later. Therefore, to maintain optimal performance of hospital readmission models on EHR data, retraining the model every 3 years may avoid model degradation and maintain optimal performance.

6 Discussion and Conclusion

We examined the hypothesis that it is feasible to enhance hospital readmission risk predictions on EHR data using data collected from a related source domain. ERR-TDAN model trained on joint TUHS, and PSUHS data yielded a 5% increase in F1-score when compared to an LSTM model trained on TUHS only, 6% increase in F1-score when compared to LSTM model trained on both TUHS and PUSH, and 3% increase when compared to a generalized version of ERR-TDAN (G-ERR-TDAN) aimed to generalize on both domains. Furthermore, conducted experiments showed that one year of historical data is sufficient to predict readmission. We studied the lifetime of the model to avoid performance degradation due to data drift over time. Experiments suggest that retraining the ERR-TDAN framework every three years avoids performance degradation.

We propose a framework, ERR-TDAN that establishes spatial knowledge transfer based on a temporal deep adaptation network tailored for hospital readmission on EHR data and optimized for the target task. ERR-TDAN can enhance readmission predictions of the target task using higher quality data from a related source domain under different distributions by matching the mean embeddings to reduce domain discrepancy. This is the first end-to-end transfer learning framework based on domain adaptation for hospital readmission. A deployment challenge for the proposed framework is that it requires training data from both source and target domains which might be difficult to obtain. In a planned follow up study we will evaluate applicability of the proposed method for prospective applications. In addition, we will compare the proposed hospital readmission method to alternatives aimed to learn from integrated data with explanatory variables of various quality.

7 Acknowledgments

This research was supported by the National Health Institute (NIH) under grant number R01DK122073.

References

1. McIlvennan CK., Eapen ZJ., Allen LA.: Hospital readmissions reduction program. Circulation, 131(20):1796-803 (2015 May 19). doi: 10.1161/CIRCULATIONAHA.114.010270. PMID: 25986448; PMCID: PMC4439931.
2. Hai, A.A., Weiner, M.G., Paranjape, A.P., Livshits, A., Brown, J.R., Obradovic, Z., Rubin, D.J.: Deep learning vs traditional models for predicting hospital readmission among patients with diabetes. Proc. AMIA 2022 Annual symposium, Washington, D.C. (2022).
3. Long, M., Cao, Y., Wang, J., Jordan, M.: Learning transferable features with deep adaptation networks. Proceedings of the 32nd International Conference on Machine Learning. Proceedings of Machine Learning Research, 37:97-105 (2015).
4. Helm JE., Alaeddini A., Stauffer JM., Bretthauer KM., Skolarus TA.: Reducing hospital readmissions by integrating empirical prediction with resource optimization. Production and Operations Management, 25(2):233-57 (2016 Feb).
5. Desautels T., Das R., Calvert J., Trivedi M., Summers C., Wales DJ., Ercole A.: Prediction of early unplanned intensive care unit readmission in a UK tertiary care hospital: A cross-sectional machine learning approach. BMJ Open, 7(9):e017199 (2017 Sep 1).
6. Gupta, P., Malhotra, P., Narwariya, J. et al.: Transfer learning for clinical time series analysis using deep neural networks. J Healthc Inform Res 4, 112–137 (2020). https://doi.org/10.1007/s41666-019-00062-3.
7. Long, M., Cao, Y., Wang, J., Jordan, M.: Learning transferable features with deep adaptation networks. In Proc. 32nd Int. Conf. Mach. Learn, pp. 97–105 (2015).
8. Wang, M., Deng, W.: Deep visual domain adaptation: A survey. Neurocomputing, vol. 312, pp. 135–153 (Oct. 2018).
9. Deng, W., Zheng, L., Ye, Q., Kang, G., Yang, Y., Jiao, J.: Image- image domain adaptation with preserved self-similarity and domain- dissimilarity for person re-identification. In Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., pp. 994–1003 (Jun 2018).
10. Chen, Y., Li, W., Sakaridis, C., Dai, D.,Van Gool, L.: Domain adaptive faster R-CNN for object detection in the wild. In Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., pp. 994–1003 (Jun 2018).
11. Hosseini-Asl, E., Keynton, R., El-Baz, A.: Alzheimer's disease diagnostics by adaptation of 3D convolutional network. In Proc. IEEE Int. Conf. Image Process. (ICIP), pp. 126–130 (Sep 2016).
12. Elixhauser A SC PL. Clinical classifications software (ccs): Agency for healthcare research and quality (2014), http://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp, last accessed 12/27/2021.
13. Hochreiter, S., Schmidhuber, J.: Long short-term memory, Neural Comput., vol. 9, no. 8, pp. 1735–1780 (1997).
14. Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B., Smola, A.: A kernel two-sample test. J. Mach. Learn. Res., vol. 13, pp. 723–773 (Mar 2012).
15. Powers DMW. Evaluation: From precision, recall and f-measure to roc., informedness, markedness & correlation. Journal of Machine Learning Technologies. 2:37-63 (2011).