BMC Medical Informatics and Decision Making

RESEARCH Open Access

Treatment effect prediction with adversarial deep learning using electronic health records

Jiebin Chu[1], Wei Dong[2], Jinliang Wang[3], Kunlun He[2*] and Zhengxing Huang[1*]

From The 4th International Workshop on Semantics-Powered Data Analytics
Auckland, New Zealand. 27 October 2019

Abstract

Background: Treatment effect prediction (TEP) plays an important role in disease management by ensuring that the expected clinical outcomes are obtained after performing specialized and sophisticated treatments on patients given their personalized clinical status. In recent years, the wide adoption of electronic health records (EHRs) has provided a comprehensive data source for intelligent clinical applications including the TEP investigated in this study.

Method: We examined the problem of using a large volume of heterogeneous EHR data to predict treatment effects and developed an adversarial deep treatment effect prediction model to address the problem. Our model employed two auto-encoders for learning the representative and discriminative features of both patient characteristics and treatments from EHR data. The discriminative power of the learned features was further enhanced by decoding the correlational information between the patient characteristics and subsequent treatments by means of a generated adversarial learning strategy. Thereafter, a logistic regression layer was appended on the top of the resulting feature representation layer for TEP.

Result: The proposed model was evaluated on two real clinical datasets collected from the cardiology department of a Chinese hospital. In particular, on acute coronary syndrome (ACS) dataset, the proposed adversarial deep treatment effect prediction (ADTEP) (0.662) exhibited 1.4, 2.2, and 6.3% performance gains in terms of the area under the ROC curve (AUC) over deep treatment effect prediction (DTEP) (0.653), logistic regression (LR) (0.648), and support vector machine (SVM) (0.621), respectively. As for heart failure (HF) case study, the proposed ADTEP also outperformed all benchmarks. The experimental results demonstrated that our proposed model achieved competitive performance compared to state-of-the-art models in tackling the TEP problem.

Conclusion: In this work, we propose a novel model to address the TEP problem by utilizing a large volume of observational data from EHR. With adversarial learning strategy, our proposed model can further explore the correlational information between patient statuses and treatments to extract more robust and discriminative representation of patient samples from their EHR data. Such representation finally benefits the model on TEP. The experimental results of two case studies demonstrate the superiority of our proposed method compared to state-of-the-art methods.

Keywords: Treatment effect prediction, Deep learning, Adversarial learning, Electronic health records

* Correspondence: kunlunhe@plagh.org; zhengxing.h@gmail.com
[2]Department of Cardiology, Chinese PLA General Hospital, Beijing, China
[1]College of Biomedical Engineering and Instrumental Science, Zhejiang University, Hangzhou, China
Full list of author information is available at the end of the article

Background

Defined as the operations and medication delivered during hospitalization, treatments have a significant impact on the prognosis of patients. Are the treatment interventions appropriate to be conducted on an individual patient given his or her specific clinical status? Will the delivered treatments achieve the expected effects on patients during their hospitalization? Traditional approaches to addressing such questions have mostly relied on evidence-based medicine [1], which urges healthcare professionals to make treatment decisions according to the best evidence from systematic research on both the efficacy and efficiency of various therapeutic alternatives [2]. Ideally, healthcare professionals compare different treatment options by referring to randomized, double-blind, head-to-head clinical trials [1], evaluate the resulting treatment effects in a prospective manner, and then select the best one to be conducted on individuals according to their specific clinical status [3].

Although valuable, there are two typical limitations to randomized controlled trial (RCT) studies [1, 4–8]. The first is that participants in RCTs are strictly selected and tend to be a “pretty rarefied population”, which is not representative of the real-world population that the scheduled treatments will eventually target [5, 6]. The second is that existing approaches are almost from a reactive perspective, in that they allow healthcare professionals to identify inappropriate interventions only after they have occurred, rather than supporting them in preventing unexpected treatment effects in advance [7, 8].

Electronic health records (EHRs), with their increasingly widespread adoption in clinical practice, provide a comprehensive source for treatment effect analysis to augment traditional RCT studies [9–15]. An EHR contains large amounts of clinical data generated as a byproduct of treatment activities [10]. A wide variety of data types are available in EHRs, including patient demographics, symptoms, vital signs, laboratory test results, and other data types that can be used to describe a patient’s clinical status, and therefore subsequent treatments (for example, drugs, injections, surgery, and care activities) performed on the patient conditioned on his or her clinical status [14, 15]. In this regard, the different aspects of medical information recorded in EHR data are highly correlated and thus provide significant potential for exploitation, for example, to extract representative and discriminative features for treatment effect prediction (TEP), which is the main objective of this study.

TEP is vital for efficiently managing disease care and therapy, owing to its usefulness in capturing actionable knowledge to assist healthcare professionals in selecting among the many therapies claimed to be efficacious for treating a patient within a specific clinical status [16–19]. As a fundamental problem of precision medicine with a wide range of applications, such as treatment recommendation [20, 21] and medical error avoidance [22], TEP can generate nontrivial knowledge with dual benefits. Not only can it demonstrate comprehension regarding patient treatment adoption, but it can also serve as an efficient and proactive indicator of medical errors before they actually occur.

To address the challenges of TEP, EHR-driven models are generally required to be capable of capturing representative and discriminative features of patient characteristics and subsequent treatments in an integrated manner and from a large volume of EHR data. In this study, we use deep learning tactics to leverage the potential of EHR data to anticipate treatment effects. Specifically, we propose a novel adversarial deep learning model for treatment effect prediction (ADTEP) based on the auto-encoder (AE) [23, 24] and adversarial learning [25]. In detail, we employ two AEs, which encode the physical condition and treatment information of patient samples into latent robust representations. In addition to the treatment decoder, treatments can be generated based on the latent representation of the patient status, under the manipulation of the actual treatment effect, so as to regularize the latent features and capture correlations between patient characteristics and treatments. To align the generated treatments with the actual performed treatments, we adopt an adversarial learning scheme and use a discriminator to differentiate the fake generated treatments from the real performed treatments documented in the EHR data. With this adversarial learning strategy, not only the patient characteristics and subsequent treatments, but also the correlational information between them are encoded in the latent representation, making the generated features sufficiently representative to convey the essential and critical information in the EHR data. Note that the latent representations of patient samples and the treatment effect predictor are jointly trained, making the representations discriminative and optimized for TEP. We conducted experiments to evaluate the effectiveness of the proposed model on two real clinical data sets collected from the Cardiology Department of the Chinese PLA General hospital. The experimental results demonstrate that our proposed model outperforms other state-of-the-art models.

The remainder of this paper is organized as follows. We review the related work in Section 2. Section 3 formulates the problem and presents our proposed approach in detail. The experimental setup and results using a real clinical dataset are presented in Section 4. Finally, we conclude the paper in Section 5.

Related work

TEP models [16–18, 26–33] have been proposed to predict the treatment effects of patient individuals following the performed treatment. The gold standard approach to addressing the problem of TEP is clinical RCTs, which aim to avoid bias when testing new treatments [5]. Although valuable, RCTs exhibit several serious limitations [6–8]; for example, they require strict inclusion and exclusion criteria, the causal conclusions from RCT studies cannot be applied to other localities automatically, and most importantly, RCTs are sometimes infeasible owing to ethical issues.

In recent years, the increased availability of EHRs has demonstrated significant potential for improving the performance of various clinical applications [9–15, 34, 35]. EHRs regularly document various care and treatment behaviors, such as procedures, diagnoses, and laboratory tests and measurements, of patients within the context of large healthcare systems [9, 10], capture the characteristics of heterogeneous populations of patients receiving care in their current clinical setting [11], and therefore form a large volume of clinical observational data sources. As an essential source of clinical observational data, and an efficient and alternative channel for TEP, EHR data have been gradually incorporated into estimating treatment effects [36, 37]. For example, Rosenbaum and Rubin proposed a classical propensity score matching model to reduce selection bias for estimating treatment effects [2]. Wager and Athey [26] proposed a variant of random forests, known as causal forests, to measure the propensity scores for treatment effect estimation.

Although valuable, two main limitations exist when using EHR data for TEP: (1) treatment selection bias inevitably exists in clinical practice [16, 17], that is, similar patients always receive the same treatments based on the recommendations from certain pre-existing clinical guidelines or protocols, and thus, EHR data are typically biased as they faithfully documents the actual treatment behavior and do not contain all possible outcomes for all treatments; (2) only the factual outcomes of the assigned treatments are observed, and counterfactual outcomes of alternative treatments are not observed [26–29, 33, 36, 37]. Note that the treatment outcomes of patients are never the same, and therefore, the learning process must provide an understanding of how the current patient is similar to previous patients [30]. This learning problem is further complicated by the fact that the data include only the received treatment outcomes, and - not the potential outcomes of the alternative treatments, namely the counterfactuals [27].

To overcome these limitations, numerous studies have proposed creating a balance by re-weighting samples with their inverse propensity score (IPS) and formulating the problem of counterfactual inference as the domain adaption problem [28, 29]. For example, Swaminathan and Joachims proposed a direct estimation model to minimize the "corrected" loss function, using IPS corrected by a regularization term over the linear stochastic policy class [28]. As a further study, Swaminathan and Joachims developed a variant of the IPS estimator, that is, a self-normalizing estimator, to learn the counterfactuals [30]. Jordan and Schaar proposed combining the direct and IPS methods and generate more robust counterfactual estimates [30]. In particular, they used a novel AE network to reduce bias by learning a representation map to control the trade-off between the bias reduction and information loss [30].

In recent years, deep learning has attracted considerable interest in various research fields for achieving impressive performance. Shifting to the clinical domain, deep learning tactics have been receiving increased attention for solving the TEP problem [17, 27, 30, 32]. For example, Louizos et al. [16] proposed the causal effect variational AE to learn the latent variables for estimating individual treatment effects. Atan, Jordan and Schaar [30] proposed a deep-treat model to estimate the treatment policies on the transformed data learned from an AE. Lee et al. [31] developed a novel adversarial learning framework to conduct unbiased treatment effect estimation using noisy proxies. Yoon et al. [17] employed a generative adversarial network (GAN) to estimate individual treatment effects. Alaa et al. [33] proposed multi-task deep counterfactual networks for treatment effect estimation by learning shared representations for treated and control outcomes and reducing the impact of selection bias in observational data by means of a propensity-dropout regularization scheme. Although valuable, it must be mentioned that most of these deep learning models have assumed that only binary actions or a few treatment options exist, namely treat and do not treat, while in most situations, various treatment combinations are possible.

In comparison with state-of-the-art models that simply tackle binary or several treatment options, our proposed ADTEP elegantly deals with various treatment combinations by extracting representative and discriminative features from observational data. Moreover, the proposed model is capable of extracting correlational information between patient characteristics and treatments from EHR data, which is essential for treatment effect estimation but somehow neglected by numerous existing models.

Methods

($\boldsymbol{x}$: patient feature vector, y: outcome, $\boldsymbol{a}$: treatment vector, $\boldsymbol{h}_x$: latent feature vector of patient features, $\boldsymbol{h}_a$: latent feature vector of treatments, $\boldsymbol{x}'$: reconstructed feature vector of patient features, $\boldsymbol{a}'$: reconstructed feature vector of treatments, $\tilde{\boldsymbol{a}}$: fabricated vector of treatments, $\mathrm{E_x}$: patient feature encoder, $\mathrm{E_a}$: treatment intervention encoder, $\mathrm{G_x}$: patient feature decoder, $\mathrm{G_a}$: treatment decoder, $\mathrm{G_{xa}}$: treatment generator, $\mathrm{D_a}$: treatment discriminator, $\mathrm{C_y}$: logistic regression layer for TEP, l_x: patient feature reconstruction loss, l_a: treatment reconstruction loss, l_{GAN}: adversarial loss, l_{pred}: treatment outcome prediction loss.)

We consider a typical clinical study of TEP, in which the EHR data record patient features, treatment interventions, and achieved treatment outcomes. For each patient sample u, we observe a set of patient features $\boldsymbol{x}_{\mathrm{u}}$, a set of treatment interventions $\boldsymbol{a}_{\mathrm{u}}$ conditioned on $\boldsymbol{x}_{\mathrm{u}}$, and the achieved treatment outcome y_u. The EHR dataset can be described as,

$$\mathcal{D} = \{(\boldsymbol{x}_{\mathrm{u}}, \boldsymbol{a}_{\mathrm{u}}, y_u) | \mathrm{u} = 1, \cdots, \mathrm{N}_{\mathcal{D}}\} \tag{1}$$

We propose the ADTEP model to address the aforementioned problem. The ADTEP inherits the loss function of traditional classification models, and takes advantage of the adversarial learning scheme to extract representative and discriminative features, which not only semantically encode the essential and critical information contained in the patient EHR, but also provide the benefit of achieving high accuracy for TEP.

As illustrated in Fig. 1(A), during the training process, the proposed ADTEP contains seven components: a patient feature encoder $\mathrm{E_x}$, a treatment intervention encoder $\mathrm{E_a}$, a patient feature decoder $\mathrm{G_x}$, a treatment intervention decoder $\mathrm{G_a}$, a treatment intervention generator $\mathrm{G_{xa}}$, a treatment intervention discriminator $\mathrm{D_a}$, and a logistic regression layer for TEP $\mathrm{C_y}$. In detail, given a patient sample $(\boldsymbol{x}, \boldsymbol{a}, y)$, two encoder layers $\mathrm{E_x}$ and $\mathrm{E_a}$ are first employed to extract the latent features $\boldsymbol{h}_x$ and $\boldsymbol{h}_a$ from $\boldsymbol{x}$ and $\boldsymbol{a}$, respectively. The reconstructed features $\boldsymbol{x}^{'}$ and $\boldsymbol{a}^{'}$ can then be estimated from the latent features $\boldsymbol{h}_x$ and $\boldsymbol{h}_a$, using the decoders $\mathrm{G_x}$ and $\mathrm{G_a}$. Note that $\mathrm{E_x}$ and $\mathrm{G_x}$ form an AE for patient feature observations, and for $\mathrm{E_a}$ and $\mathrm{G_a}$ to reconstruct treatment interventions. Both AEs $\mathrm{E_x}$- $\mathrm{G_x}$ / $\mathrm{E_a}$- $\mathrm{G_a}$ are adopted to capture robust and discriminative patient feature/treatment representations in the latent feature vector $\boldsymbol{h}_{\mathrm{x}}$ / $\boldsymbol{h}_{\mathrm{a}}$. Consequently, the latent feature vectors $\boldsymbol{h}_{\mathrm{x}}$ and $\boldsymbol{h}_{\mathrm{a}}$ are concatenated to form the input of $\mathrm{C_y}$ for TEP.

As treatment interventions are performed conditioned on patient features in clinical practice, we feed the latent patient features $\boldsymbol{h}_x$ into another generator $\mathrm{G_{xa}}$ to yield treatment interventions $\tilde{\boldsymbol{a}}$**,** conditioned on the treatment outcome y of the patient sample $\tilde{\boldsymbol{a}} = \mathrm{G}_{xa}(\boldsymbol{h}_x, y)$, and then use a discriminator to distinguish whether or not the generated treatment interventions $\tilde{\boldsymbol{a}}$ and original ones $\boldsymbol{a}$ originate from the same treatment distributions. The use of the generator $\mathrm{G_{xa}}$ allows us to learn the latent correlations between patient features and treatments. This learning strategy can regularize the latent features $\boldsymbol{h}_x$ to encode most of the information shared between the patient characteristics and subsequent treatments. The details are as follows.

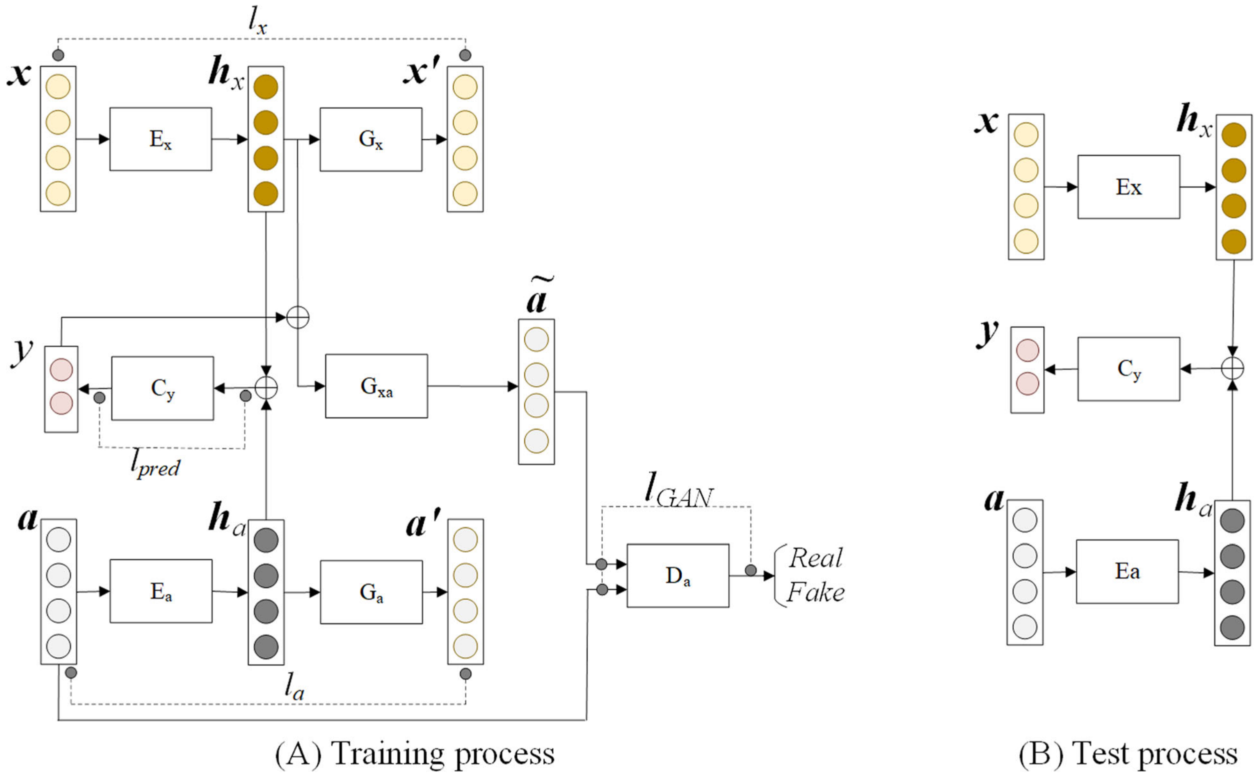

Fig. 1 Adversarial deep treatment effect prediction (ADTEP) model

Encoder-decoder

We employ two AEs, namely $\mathrm{E_x}$- $\mathrm{G_x}$, and $\mathrm{E_a}$- $\mathrm{G_a}$, to learn the latent representations of patient characteristics and treatments, respectively. A simple form of an AE is a feed-forward and non-recurrent neural network [24, 38], consisting of an input layer, an output layer and one or multiple hidden layers in between. The AE attempts to reconstruct the input from the corrupted data. Formally, given an M-dimensional input patient feature vector $\boldsymbol{x}\in\mathbb{R}^{\mathrm{M}}$, it is mapped to the code vector $\boldsymbol{h}_x$ with the encoding function $\mathrm{E}_x(\mathrm{W_e}\boldsymbol{x}+\mathrm{b_e})$, and subsequently during the decoding step, it maps the code vector $\boldsymbol{h}_x$ to the output vector $\boldsymbol{x}^{'}$, which reconstructs the input vector with the decoding function $\mathrm{G}_x(W_d\boldsymbol{h}_x+b_d)$, where $\mathrm{W_e}\in\mathbb{R}^{\mathrm{K\times M}}$ and $\mathrm{W_d}\in\mathbb{R}^{\mathrm{M\times K}}$ are weighted matrices, $b_{\mathrm{e}}\in\mathbb{R}^{\mathrm{K}}$ and $b_d\in\mathbb{R}^{\mathrm{M}}$ are the corresponding bias terms, $\mathrm{E}_x(\cdot)$ and $\mathrm{G}_x(\cdot)$ are nonlinear activation functions, and K is the number of nodes in the hidden layer.

Similar to the AE $\mathrm{E_x}$- $\mathrm{G_x}$, the treatment encoder $\mathrm{E_a}$ takes the treatment vector $\boldsymbol{a}$ as input and generates the latent treatment vector $\boldsymbol{h}_a$, which is subsequently fed into decoder $\mathrm{G_a}$ to generate the reconstructed treatment $\boldsymbol{a}^{'}$. Both $\mathrm{E_a}$ and $\mathrm{G_a}$ constitute a treatment AE, which aims at reconstructing the treatment behavior from the patient EHR data.

It is very challenging to generate the treatment vector $\tilde{\boldsymbol{a}}$ of a patient sample from his or her clinical status representation $\boldsymbol{x}{:}P(\tilde{\boldsymbol{a}}|\boldsymbol{x})$, owing to the large appearance variations in the treatment selections given the patient characteristics in clinical settings. To address this problem, we use the patient feature encoder E_x and treatment generator G_{xa} to form an AE. Specifically, given a patient feature vector $\boldsymbol{x}$ and the known treatment effect y, E_x is adopted to extract the latent feature vector $\boldsymbol{h}_{\mathrm{x}}=\mathrm{E}_x(\boldsymbol{x})$. The feature vector $\boldsymbol{h}_{\mathrm{x}}$ is expected to encode the correlational information between the patient characteristics and treatments after adversarial training, and the treatment vector $\tilde{\boldsymbol{a}}$ can be estimated from the latent feature vector $\boldsymbol{h}_{\mathrm{x}}$, using G_{xa} conditioned on the obtained treatment effect y : $\tilde{\boldsymbol{a}}{=}\mathrm{G}_{xa}(\boldsymbol{h}_x, y)$.

Patient feature reconstruction loss

In this study, we measure the reconstruction performance for patient feature $\boldsymbol{x}$ conducted by the encoder $\mathrm{E_x}$ and decoder $\mathrm{G_x}$. For efficient learning of the encoder-decoder, standard practice is to use the Euclidean distance between the input and the generated output to minimize the patient feature reconstruction loss, that is,

$$\mathcal{L}_x = \mathbb{E}_{\boldsymbol{x},\boldsymbol{a},y\sim P_{data}(\boldsymbol{x},\boldsymbol{a},y)}||\boldsymbol{x}-\mathrm{G}_x(\mathrm{E}_x(\boldsymbol{x}))||_2^2 \qquad (2)$$

Here, the encoder $\mathrm{E_x}$ maps the input patient feature vector $\boldsymbol{x}$ into the latent one $\boldsymbol{h}_x$, and then, the decoder $\mathrm{G_x}$ reconstructs the feature $\boldsymbol{x}^{'}$ from $\boldsymbol{h}_x$.

Treatment reconstruction loss

The reconstruction performance for treatment vector $\boldsymbol{a}$ is measured by means of the encoder $\mathrm{E_a}$ and decoder $\mathrm{G_a}$. Similarly to the patient feature reconstruction loss $\mathcal{L}_x$, the treatment reconstruction loss $\mathcal{L}_a$ can be measured as follows:

$$\mathcal{L}_a = \mathbb{E}_{\boldsymbol{x},\boldsymbol{a},y\sim P_{data}(\boldsymbol{x},\boldsymbol{a},y)}||\boldsymbol{a}-\mathrm{G}_a(\mathrm{E}_a(\boldsymbol{a}))||_2^2 \qquad (3)$$

Minimizing Eqs. (2) and (3) aids us in determining a representative latent feature space for the patient clinical characteristics and subsequent treatments.

Discriminator

As a popular learning formulation for deep learning, adversarial learning is similar to a competition game, in which a discriminator judges a data sample as real or fake; in contrast, a generator attempts to produce indistinguishable samples without being detected [17, 25, 39]. Inspired by adversarial learning and based on the common sense whereby treatments are conditioned on patient characteristics in a clinical context [10], we encourage the reconstruction of treatments from discriminative patient features that are similar to real ones, so that the prediction performance can be enriched.

To this end, we design a treatment discriminator $\mathrm{D_a}$ to differentiate the reconstructed treatment vector $\tilde{\boldsymbol{a}}$ from the true observed treatment $\boldsymbol{a}$. In particular, we employ a binary classifier to categorize the given input as "real" if the input is the actual treatment vector performed on patients, and "fake" otherwise. $\mathrm{D_a}$ enables the proposed model to learn a hidden treatment representation $\boldsymbol{h}_a$ from the EHR data. Meanwhile, $\mathrm{D_a}$ causes the latent patient features $\boldsymbol{h}_x$ to be treatment specific. As a result, it improves the discriminative capability of the learned features, and makes them particularly optimized for TEP.

Adversarial loss

$\mathcal{L}_{GAN}$ is optimized to train the encoder $\mathrm{E_x}$, decoder G_{xa}, and discriminator D_a. The encoder $\mathrm{E_x}$ is trained to generate the treatment-specific patient feature $\boldsymbol{h}_x$, while the decoder G_{xa} is trained to generate treatments conditioned on $\boldsymbol{h}_x$ manipulated by the treatment outcome label y. The discriminator D_a attempts to distinguish the actual treatment vector $\boldsymbol{a}$ as real and the reconstructed one $\tilde{\boldsymbol{a}}$ as fake. We define the adversarial loss $\mathcal{L}_{GAN}$ as:

$$\begin{aligned}\mathcal{L}_{GAN} = {} & \mathbb{E}_{\boldsymbol{x},\boldsymbol{a},y\sim p_{data}(\boldsymbol{x},\boldsymbol{a},y)}[\ \log\mathrm{D}_a(\boldsymbol{a})] \\ & + \mathbb{E}_{\boldsymbol{x},\boldsymbol{a},y\sim p_{data}(\boldsymbol{x},\boldsymbol{a},y)}[\ \log(1-\mathrm{D}_a(\mathrm{G}_{xa}(\mathrm{E}_x(\boldsymbol{x}),y)))].\end{aligned} \qquad (4)$$

Treatment outcome predictor

Given a testing patient sample with patient feature vector $\boldsymbol{x}$, treatment vector $\boldsymbol{a}$ conditioned on $\boldsymbol{x}$, and an unknown treatment outcome label y, we can learn the representative and informative features $\boldsymbol{h}_\mathrm{x}$ and $\boldsymbol{h}_\mathrm{a}$ with respect to the patient characteristics, and subsequently the treatments performed on the patient, respectively, and then concatenate these as $[\boldsymbol{h}_\mathrm{x}, \boldsymbol{h}_\mathrm{a}]$ to be fed into the treatment effect predictor C_y, so that treatment effects can be estimated for the target patient.

Treatment outcome prediction loss

$\mathcal{L}_{pred}$. In this study, we employ a logistic regression layer for the treatment outcome prediction, in which the input is the concatenation of the latent patient feature vector $\boldsymbol{h}_x$ and treatment vector $\boldsymbol{h}_\mathrm{a}$. This is used to estimate the treatment effect of patient samples given their clinical conditions and performed treatment interventions. The loss can be measured using cross-entropy as follows:

$$\begin{aligned}\mathcal{L}_{pred} &= \frac{1}{N_D}\sum\nolimits_{u=1}^{N_D} L\left(W_{pred}, b_{pred}; \boldsymbol{x}_u, \boldsymbol{a}_u, y_u\right)\\ &= \frac{1}{N_D}\sum\nolimits_{u=1}^{N_D}\left(y_u \ \log y'_u + (1-y_u)\ \log\left(1-y'_u\right)\right),\end{aligned} \tag{5}$$

where y' is the predicted treatment outcome.

Model learning

As demonstrated in the section above, our training is defined by four loss functions: 1) loss of GAN $\mathcal{L}_{GAN}$, loss of patient feature reconstruction $\mathcal{L}_x$, loss of treatment reconstruction $\mathcal{L}_a$, and loss of treatment outcome prediction $\mathcal{L}_{pred}$. In summary, the objective function of the ADTEP is expressed as:

$$\begin{aligned}&\min_{E_x,E_a,G_x,G_a,G_{xa},C_y}\ \max_{D_a} L_{pred} + \alpha(L_x + L_a)\\ &\quad + \beta L_{GAN},\end{aligned} \tag{6}$$

where α and β are trade-off parameters for balancing the importance of the corresponding components.

The learning algorithm of the proposed model can be formulated as follows:

1. Update the parameters of the patient feature encoder and decoder $\{\Theta_{E_x}, \Theta_{G_x}\}$ by minimizing the patient feature reconstruction loss $\mathcal{L}_x$. Note that the encoder E_x and decoder G_x are trained to reconstruct patient characteristics. Moreover, the encoder E_x is regularized to generate treatment-specific patient characteristics, as it also needs to generate treatments, as discussed previously.
2. Update the parameters of the treatment encoder and decoder $\{\Theta_{E_a}, \Theta_{G_a}\}$ by minimizing the treatment reconstruction loss $\mathcal{L}_a$.
3. Update the discriminator parameter $\{\Theta_d\}$ to optimize $\mathcal{L}_{GAN}$ by maximizing the adversarial loss $\max\limits_{D} \mathcal{L}_{GAN}$.
4. Update the treatment effect predictor $\{\Theta_c\}$ by minimizing the prediction loss $\min\limits_{C} \mathcal{L}_{pred}$.

Note that the above objectives are optimized in an iterative manner. Specifically, E_x, E_a, G_x, G_a, G_{xa}, D_a, and C_y improve one another during the alternative training process. With D_a being more capable of distinguishing the generated fake treatment vector and real one, G_{xa} encourages the generation of fake treatments which based on patient feature to compete with the discriminator D_a. To this end, the encoder E_x and decoder G_x are driven to encode the representative patient features into the latent feature vector $\boldsymbol{h}_x$. Thereafter, the treatment generator G_{xa} learns how to map the latent patient feature $\boldsymbol{h}_x$ to conditioned treatments $\tilde{\boldsymbol{a}}$ corresponding to the input patient feature $\boldsymbol{x}$. This process makes the features particularly optimized for TEP.

Fig. 1 (B) presents the flowchart of the TEP test process. In particular, the AEs E_x- G_x and E_a- G_a are used to generate the latent feature representations $\boldsymbol{h}_x$ and $\boldsymbol{h}_a$, which are then concatenated as the input of C_y to predict treatment outcomes for the test patient samples

Treatment effect analysis for target outcome

To analyze the association between the treatment and clinical outcome in an interpretable manner, we compute the effect of each treatment for the target outcome following training. We firstly compute the mean loss $\mathcal{L}_{pred}$ over the training samples. Thereafter, for each treatment k, $1 \le k \le \mathrm{K}$, and for each patient sample u, $1 \le u \le N_D$, we let $\hat{\boldsymbol{a}}^{(u)}=\boldsymbol{a}^{(u)}$ and then set $\hat{\boldsymbol{a}}_k^{(u)} = 0$. Based on the adjusted $\hat{\boldsymbol{a}}^{(u)}$, we compute the mean loss, as follows:

$$\begin{aligned}\mathcal{L}^k_{pred} &: \mathcal{L}_{pred}\\ &= \frac{1}{N_D}\sum\nolimits_{u=1}^{N_D} L\left(W_{pred}, b_{pred}; \boldsymbol{x}_u, \hat{\boldsymbol{a}}^{(u)}, y_u\right),\end{aligned} \tag{7}$$

and then compute the effect of treatment k for the target outcome:

$$\mathrm{eff}^k = \mathcal{L}^k_{pred} - \mathcal{L}_{pred} \tag{8}$$

Note that the calculated value of eff^k discloses the relevant treatment for the target variation, which is helpful for physicians to understand whether the performed treatment has an effect on the target outcome, and the means by which the black-box deep learning-based TEP model operates in a reasonable and trustworthy manner. We argue that the analysis results can provide certain insights for the formation of treatment effects on the target clinical outcomes.

Experiments

We conducted two clinical case studies in cooperation with the Cardiology Department of the Chinese PLA General Hospital. The first investigated major adverse cardiac event (MACE) prediction after acute coronary syndrome (ACS), while the second focused on one-year readmission prediction for heart failure (HF) patients, as detailed in the following subsections.

Note that categorical features, such as gender, operation, medicine and complication, are represented as binary values. Meanwhile, continuous features, such as age, BMI and lab test values, are categorized into three levels: lower than normal, normal and higher than normal, according to the clinical protocol adopted by the hospital, and represented as one-hot vectors with three dimensions.

All experiments were conducted on a Microsoft Surface Pro 5 Compatible PC with an Intel Core i7-7660U CPU 2.50 GHz and 8 GB of main memory, running on Microsoft Windows 10. The proposed model was implemented in Python, and the source code is available at https://github.com/ZJU-BMI/treatment. Prior approval for conducting the study was obtained from the data protection committee of the hospital. We wish to make it clear that the patient data were anonymized in this study and in this paper.

Performance comparisons

To demonstrate the effectiveness of our proposed model, we compare the proposed ADTEP with: the proposed model without adversarial learning, namely the DTEP model. For the DTEP, we use AEs to generate the latent representations of both the patient characteristics and the subsequent treatments, concatenate the derived latent features, and then feed the obtained feature vector into a logistic regression layer, yielding a TEP model. Note that DTEP does not consider the correlations between the patient state and the treatment. Moreover, we compare the proposed model to benchmark models using the experimental datasets, including logistic regression (LR) and the support vector machine (SVM). L2-regularization is used in LR, DTEP and ADTEP. We search the best values of hyper-parameters with grid search strategy and all the results shown in this paper are obtained on the condition of the best settings.

Evaluation metrics

The performance was evaluated by the Area Under the receiver operating characteristic (ROC) curve (AUC), accuracy, precision, recall and F1 score. To estimate the performance of the treatment effect estimation in a less biased manner than single-round testing, we repeated the experiments five times to validate the performance of each model on the experimental dataset. Furthermore, the five-fold cross-validation strategy was applied in each run of the experiment. As a result, we obtained a group of experimental results for each model, on which the mean value and confidence intervals were calculated.

ACS case study

Data description

ACS refers to a group of conditions resulting from decreased blood flow in the coronary arteries, whereby that part of the heart muscle is unable to function properly or dies [40]. The basic treatment principles are the same for all types of ACS; however, several important aspects of treatment depend on the specific characteristics of ACS patients. For example, the comorbidities of ACS patients, presence or absence of elevation of the ST segment on the electrocardiogram, and different treatment interventions may result in varying treatment effects [41–43]. To this end, the ability to leverage a quantitative paradigm for alleviating adverse treatment effects and improving patient outcomes, in terms of both prediction and prevention could potentially deliver significant benefits to both patients and their families, as well as society. Regarding the indicators of treatment effects for ACS patient samples, we select the MACE after ACS as the label for treatment effects. MACE is a typical indicator of the treatment effect, and it often occurs suddenly, resulting in high mortality and morbidity [12, 44]. In clinical practice, MACE has a significant impact on clinical decision-making for ACS patient care and treatment.

To conduct the ACS case study, we collaborated with the clinicians of the cardiology department, and extracted a collection of 3463 ACS patient samples from the hospital EHR system. The dataset documented 326 patient features including demographics, operations, medications, laboratory values and diagnosis, etc. Specifically, for features with multiple measurement, like laboratory values, we kept the initial measurement on admission. Preprocessing was conducted on the collected ACS dataset. In particular, both patient samples and variables with more than 30% of missing values were excluded from the analysis. Other than this, no further efforts were made to handle the missing data in the experiments. As a result, 2930 patient samples with a median age of 62.27 years were obtained, among which 2080 (71%) were female. A summary of the statistics of the dataset is provided in Table 1, where shows the information of several important patient characteristics selected by our clinical collaborates. Note that the *P*-values of features with continues values were calculated by Mann-Whitney U test, while the P-values of features with binary values were calculated by Chi-squared test.

Experimental results and analysis

Table 2 presents the TEP performance achieved on the experimental ACS dataset. As can be observed from

Table 2, the proposed model achieved superior performance compared to benchmark models on the experimental dataset. ADTEP performed slightly better than DTEP in terms of both the AUC and F1. Although DTEP outperformed ADTEP in terms of the average accuracy, the performance gain was marginal. These findings indicate that the incorporation of correlational information between patient characteristics and treatments by means of the adversarial learning strategy was useful in predicting the treatment effects of ACS patient samples. Figure 2 illustrates the ROC curves for MACE prediction after ACS, also demonstrating that the proposed ADTEP achieved comparative performance with benchmark models. In particular, ADTEP exhibited 1.4, 2.2, and 6.3% performance gains for MACE prediction in terms of AUC over DTEP, LR, and SVM, respectively.

Figure 3 displays the measured Eff^k values of treatments on the ACS dataset. Two of the three most relevant treatments for MACE prediction were found to be: antiplatelet and lipid lowering therapy, which are consistent with existing medical knowledge as major interventions for ACS [45, 46]. The most irrelevant treatment for MACE was found to be coronary angiography. This finding is also reasonable, because coronary angiography is not a specific treatment for relieving the symptoms of ACS, but rather a procedure to determine how blood flows through the arteries in the hearts, and thus, is less relevant to influencing the occurrence of MACE after ACS. Surprisingly, we found that hypoglycemic therapy had the strongest correlation with MACE, while nitroglycerin had a less significant correlation. This is inconsistent with clinical guidelines as hypoglycemic therapy is mainly adopted for the treatment of type II diabetes, while nitroglycerin is recognized as a major treatment for preventing ischemic events after ACS. These findings may contain suggestive hypotheses that could be validated by further clinical investigations.

HF case study

Experimental setup

HF is a complex clinical syndrome that affects at least 40 million people globally and is increasing in prevalence

Table 1 Baseline characteristics of experimental ACS dataset

Characteristics	No. of participants (*n* = 2930)	MACE (*n* = 752)	Non-MACE (*n* = 2178)	P-value
Age (years), mean (SD)	62.27 ± 12.11	67.12 ± 11.95	60.60 ± 11.71	< 0.001
Female sex (T/F)	2080/850	528/225	1552/625	0.573
Hypertension (T/F)	1981/949	537/215	1444/734	0.011
Diabetes mellitus (T/F)	1986/803	482/224	1504/439	< 0.001
Hypercholesterolemia (T/F)	2362/568	623/129	1739/439	0.082
Previous PCI (T/F)	816/2114	214/538	602/1576	0.701
Previous CABG (T/F)	86/2844	38/714	48/2130	< 0.001
ST-segment elevations ECG (T/F)	106/2824	27/725	79/2099	0.947
BMI (kg/m^2), mean (SD)	25.90 ± 11.30	25.50 ± 12.73	26.03 ± 10.76	0.333
CCR (ml/min/ m^2), mean (SD)	78.73 ± 38.19	85.36 ± 48.30	76.41 ± 33.65	< 0.001
CKMB (umol/L), mean (SD)	9.49 ± 14.95	9.30 ± 11.45	9.56 ± 16.09	0.715
Treatment				
Coronary angiography (T/F)	993/1937	270/482	723/1455	0.191
Nitroglycerin (T/F)	904/2026	292/460	612/1566	< 0.001
Vasodilator (T/F)	951/1979	305/447	646/1532	< 0.001
Antihypertensive therapy (T/F)	1375/1555	385/367	990/1188	< 0.001
Hypoglycemic therapy (T/F)	451/2479	127/625	324/1854	0.208
Lipid lowering therapy (T/F)	511/2419	129/623	382/1796	0.854
Blood transfusion (T/F)	91/2839	28/724	63/2115	0.312
Quick-acting rescue (T/F)	796/2134	236/516	560/1618	< 0.001
Aspirin (T/F)	730/2200	204/548	526/1652	0.114
Antiarrhythmia (T/F)	114/2816	45/707	69/2109	< 0.001
Anti-angina (T/F)	1216/1714	376/376	840/1338	< 0.001
Antiplatelet (T/F)	933/1997	255/497	678/1500	0.172

BMI: body mass index; CABG: coronary artery bypass grafting; CCR: Creatinine clearance; CKMB: creatine kinase MB; ECG: electrocardiogram; PCI: percutaneous coronary intervention; SD: standard deviation

Table 2 Experimental results for accuracy, AUC, precision, recall and F1 score on ACS experimental dataset

Method	Accuracy (mean ± SD)	AUC (mean ± SD)	Precision (mean ± SD)	Recall (mean ± SD)	F1 score (mean ± SD)
LR	0.744 ± 0.016	0.648 ± 0.026	0.505 ± 0.078	0.198 ± 0.034	0.284 ± 0.044
SVM	0.716 ± 0.010	0.621 ± 0.014	0.402 ± 0.032	**0.219** ± 0.026	0.283 ± 0.027
DTEP	**0.747** ± 0.010	0.653 ± 0.021	**0.524** ± 0.056	0.181 ± 0.025	0.268 ± 0.031
ADTEP	0.746 ± 0.012	**0.662** ± 0.020	0.515 ± 0.058	0.210 ± 0.036	**0.297** ± 0.042

[47]. Although not all conditions leading to HF can be reversed, treatments can improve the signs and symptoms of HF and help patients to live longer. Usually, several HF-specific treatments are available, such as angiotensin converting enzyme inhibitor (ACEI)/angiotensin receptor blocker (ARB), beta-blockers and aldosterone antagonists, and it is meaningful to select appropriate treatments for an individual HF patient according to his or her clinical conditions and the desired treatment effects. The objective of this case study was to analyze the effects of treatments on the one-year readmission of HF patients.

The experimental dataset consisted of 736 HF patients with one-year follow up information (461 readmitted, 275 not readmitted). Each patient sample contained 105 features including demographics (such as age, gender), vital signs (including blood pressure and heart rate), laboratory tests (for example, creatinine kinase (CK), cardiac troponin T (cTnT)), echocardiography (such as ejection fraction), comorbidities (for example, diabetes and renal insufficiency), and treatments (including ACEI, ARB, and beta-blockers) adopted for these patients. Specifically, for features with multiple measurement, like vital signs and laboratory values, we kept the initial measurement on admission. Table 3 lists the information of several important patient characteristics suggested by our clinical collaborators based on their knowledge about HF. As the same with Table 1, the *P*-values of features with continues values were calculated by Mann-Whitney U test, while the P-values of features with binary values were calculated by Chi-squared test.

Experimental results and analysis

Table 4 reports the experimental results on the HF dataset. It can be observed that the proposed ADTEP outperformed benchmark models in terms of both accuracy and AUC. Specifically, ADTEP exhibited boosted performance compared to the benchmark models. This finding indicates that the proposed model can extract more discriminative representations from EHR data for predicting the treatment effects of HF patients, by using deep learning tactics. Moreover, by introducing the

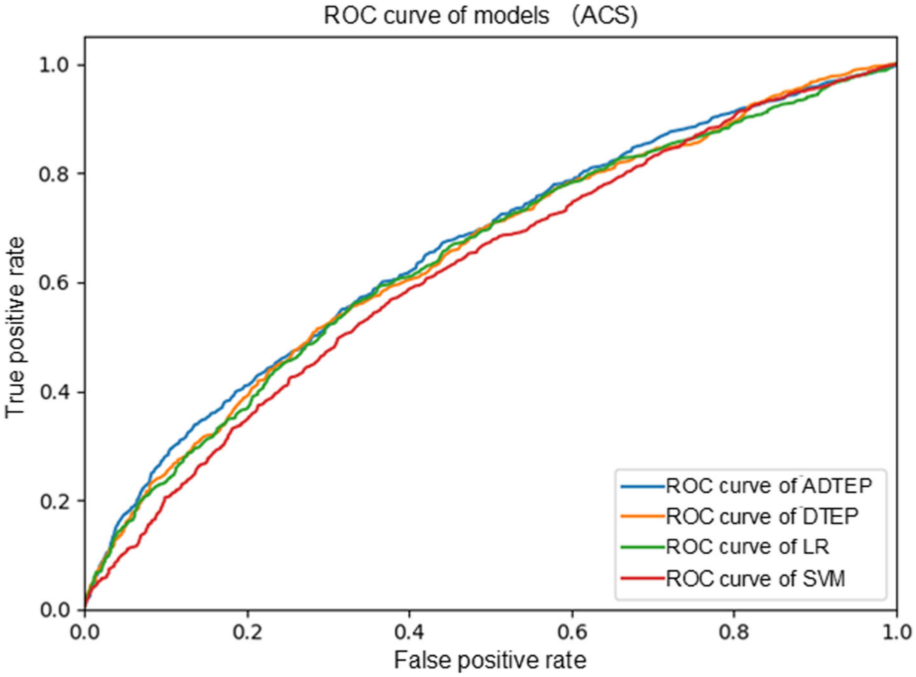

Fig. 2 ROC curves for MACE prediction after ACS

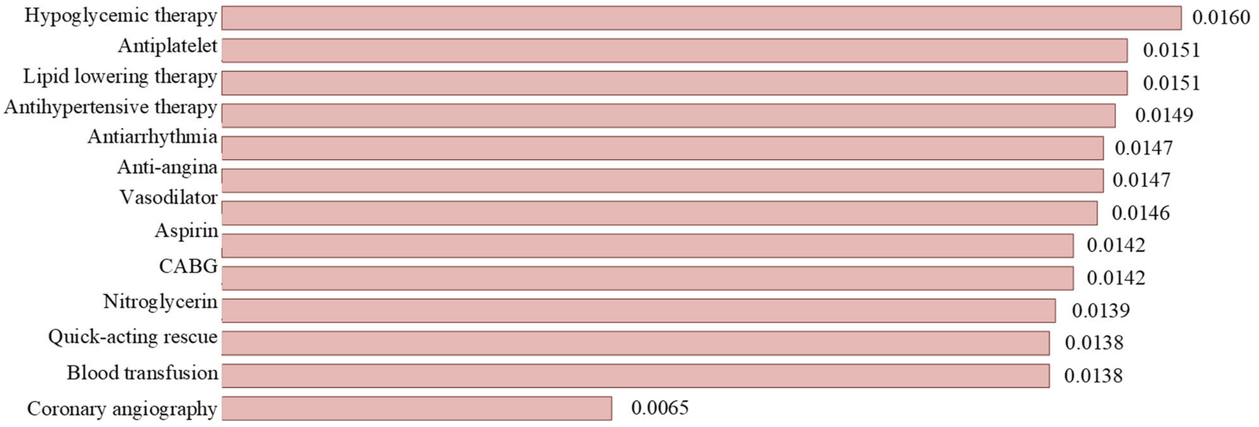

Fig. 3 Achieved Eff^k values of treatments on ACS dataset

Table 3 Baseline characteristics of experimental HF dataset

Characteristics	No. of participants (*n* = 736)	Readmission in one year (*n* = 461)	Non-readmission in one year (*n* = 275)	P-value
Age (years), mean (SD)	64.29 ± 13.55	63.66 ± 13.55	65.34 ± 13.53	0.104
Female sex (T/F)	508/227	331/130	177/97	0.050
Hypertension (T/F)	526/210	323/138	203/72	0.314
Diabetes mellitus (T/F)	466/270	283/178	183/92	0.185
Renal insufficiency (T/F)	592/144	359/102	233/42	0.030
SBP (mmHg), mean (SD)	133.41 ± 20.41	130.33 ± 20.02	138.57 ± 20.04	< 0.001
DBP (mmHg), mean (SD)	77.13 ± 13.66	76.15 ± 13.86	78.76 ± 13.19	0.012
Heart rate (b.p.m) mean (SD)	79.98 ± 16.37	81.17 ± 17.01	78.00 ± 15.06	0.011
Creatinine (umol/L), mean (SD)	100.35 ± 64.5	106.77 ± 72.85	89.61 ± 45.50	< 0.001
LVEF (%), mean (SD)	43.74 ± 11.86	41.92 ± 12.12	46.80 ± 10.76	< 0.001
CK (umol/L), mean (SD)	87.79 ± 82.04	89.71 ± 80.56	84.60 ± 84.50	0.414
cTnT (ng/ml), mean (SD)	0.058 ± 0.38	0.077 ± 0.47	0.025 ± 0.057	0.068
Treatment				
Diuretics (T/F)	536/200	344/117	202/73	< 0.001
ACEI (T/F)	442/294	279/182	163/112	0.797
ARB (T/F)	480/256	296/165	184/91	0.507
Beta-blocker (T/F)	588/148	367/94	221/54	0.879
CCB (T/F)	454/282	307/154	147/128	< 0.001
Statin (T/F)	536/200	322/139	214/61	0.023
Digoxin (T/F)	457/279	257/204	200/75	< 0.001
Nitrates (T/F)	454/282	274/187	180/95	0.122
Aspirin (T/F)	513/223	314/147	199/76	0.258
Clopidogrel (T/F)	379/357	244/217	140/135	0.650
Warfarin (T/F)	638/98	399/62	239/36	0.979
Spironolactone (T/F)	402/334	288/173	161/114	0.328
Antibiotics (T/F)	713/23	446/15	267/8	0.967
Antiacid (T/F)	589/147	367/94	222/53	0.786

ACEI: angiotensin-converting enzyme inhibitor; ARB: angiotensin receptor blocker; CCB: calcium channel blocker; cTnT: cardiac troponin T; CK: creatinine kinase; DBP: diastolic blood pressure; LVEF: left ventricular ejection fraction; SBP: systolic blood pressure

Table 4 Experimental results for accuracy, AUC, precision, recall and F1 score on experimental HF dataset

Method	Accuracy (mean ± SD)	AUC (mean ± SD)	Precision (mean ± SD)	Recall (mean ± SD)	F1 score (mean ± SD)
LR	0.647 ± 0.030	0.682 ± 0.039	0.677 ± 0.022	0.836 ± 0.037	0.748 ± 0.021
SVM	0.642 ± 0.034	0.633 ± 0.027	0.669 ± 0.018	**0.849** ± 0.050	0.748 ± 0.028
DTEP	0.624 ± 0.034	0.661 ± 0.038	0.679 ± 0.055	0.830 ± 0.149	0.721 ± 0.064
ADTEP	**0.654** ± 0.025	**0.688** ± 0.040	**0.680** ± 0.019	0.848 ± 0.034	**0.754** ± 0.019

adversarial learning strategy, the proposed ADTEP obtained performance gains of 4.8 and 4.1% in terms of accuracy and AUC, respectively, compared to DTEP. This demonstrates that discriminative representations can be obtained for efficient treatment effect estimation by extracting correlational information between patient characteristics and subsequent treatments.

Figure 4 illustrates the ROC curves achieved by both the proposed model and baseline approaches on the HF dataset. As can be observed from Fig. 4, the proposed ADTEP performed better than benchmark models. In particular, the proposed ADTEP exhibited performance gains of over 4.1, 0.9, and 8.7% in terms of the AUC in comparison with DTEP, LR, and SVM, respectively, on the experimental dataset, although LR curve closely approached the ADTEP curve. These observations indicate that deep learning tactics can indeed extract representative and discriminative features from data and therefore aid in achieving comparable TEP performance compared to state-of-the-art models. When comparing ADTEP and DTEP, it was observed that ADTEP outperformed DTEP in terms of the ROC curve. This indicates that incorporating adversarial learning into the TEP can extract more representative features to improve the TEP performance.

Moreover, to analyze the correlations between the treatments and clinical outcomes, we used Eq. (8) to measure the Eff^k values of the treatments on the target outcome (that is, one-year readmission), based on the HF dataset. As can be observed from Fig. 5, the most relevant treatments for the target outcome were: diuretics, AECI, and Warfarin, which is consistent with existing medical domain knowledge, as these are the main adopted medications for HF [47]. In contrast, the least relevant treatment for the target outcome of HF was Digoxin. Note that this finding is also consistent with the newly published clinical guidelines because Digoxin is a type of obsolete medications for HF therapy and may increase the risk of bleeding of HF patients [48]. This finding may contain suggestive hypotheses that could be validated by further clinical investigations.

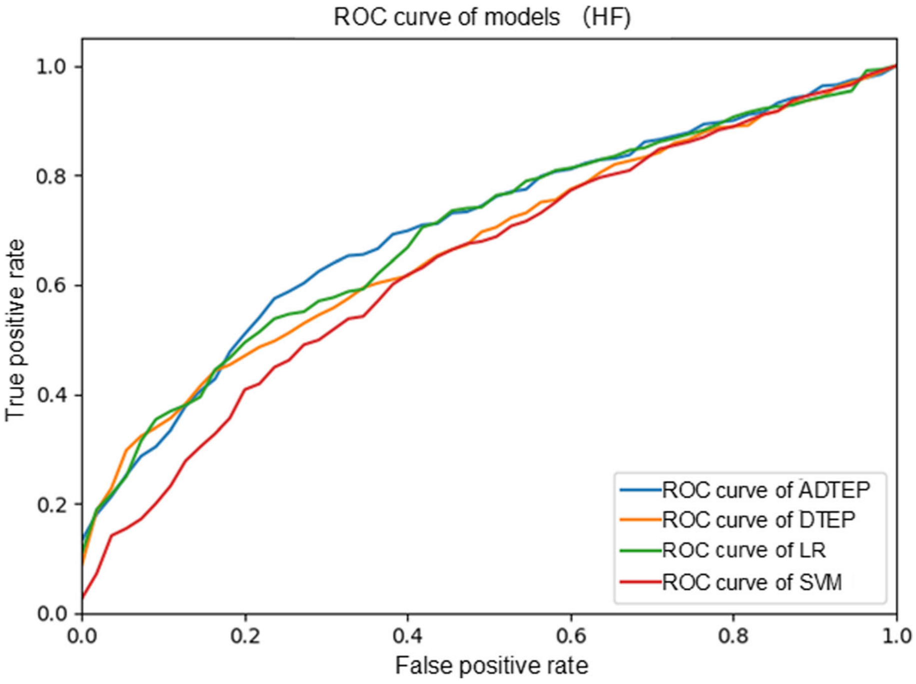

Fig. 4 ROC curves for one-year readmission prediction for HF patients

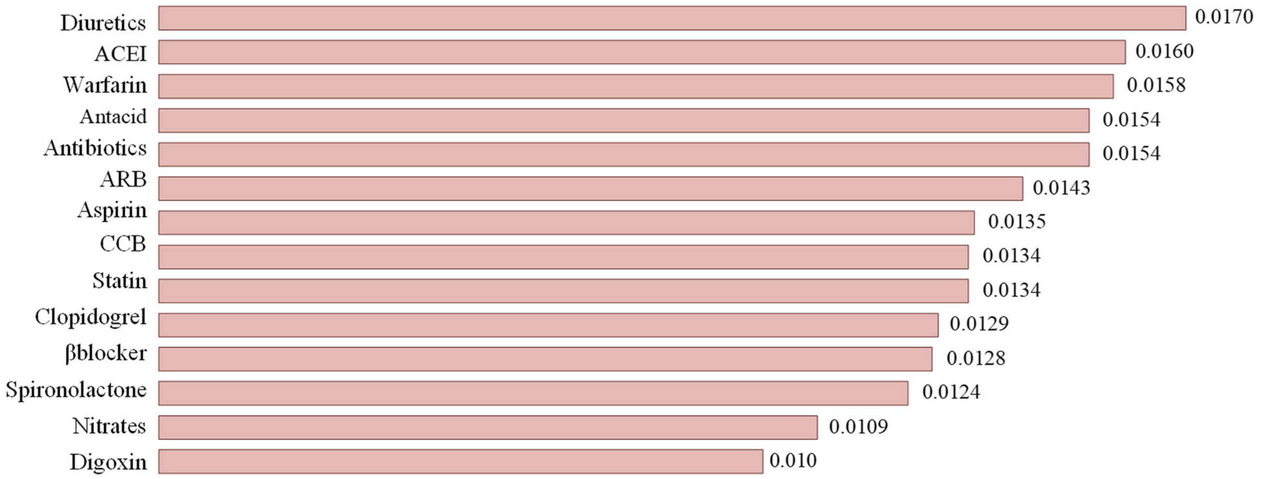

Fig. 5 Achieved Eff^k values of treatments on HF dataset

Discussion

Overall, compared to benchmark approaches, our model can improve the TEP performance in terms of two aspects. Firstly, we use deep learning models to generate latent representations of patient features and treatments. This can extract deep information from heterogeneous EHR data. Secondly, the expression of adversarial learning extracts abundant latent and nonlinear correlations between patient status and corresponding treatments, so that precisely representative features can be extracted from the data. With such ability, our proposed model shows superiority against other models on experimental results. Moreover, the results validate our assumption that the correlational information between patient characteristics and treatments can indeed improve the TEP performance. Furthermore, our model can extract informative treatments given the target outcome. Several of these extracted treatments are not only consistent with existing medical knowledge, but also contain suggestive hypotheses that could be validated by further investigations in the medical domain.

The experimental results were evaluated by hospital managers and clinical experts at the Chinese PLA General Hospital, who understand the beneficial effects of the proposed model. They indicated the potential of applying the proposed model in clinical practice for efficient treatment selection and improvement. Specifically, the proposed model can be utilized to support clinical decision-making and aid in treatment adoption. For example, the patient characteristics can be analyzed to aid healthcare professionals in scheduling individual treatment interventions for patients, in order to achieve the expected treatment effects. The method is also applicable to clinical decision support systems that recommend appropriate treatment interventions matching the specific patient statuses. This could guide healthcare professionals to schedule appropriate treatment interventions based on the measurement of the target patient statuses and the desired treatment effects, by meaningfully employing a large volume of EHR data to derive non-trivial knowledge explaining the treatment intentions and behaviors. In this regard, our clinical collaborators advocate us to develop and deploy a TEP service in the EHR system. Such a service will not only predict treatment effects nearly at run-time in the treatment processes of patients, but also essentially assist healthcare professionals to schedule appropriate treatment behaviors in a continuous and predictive manner.

Although our study has revealed that the proposed model is effective in predicting treatment effects, even more complex analysis and evaluation tasks remain to be addressed. In this study, patient characteristics are generated using the data collected at a single time point. However, the dynamic nature of patient characteristics is often essential in the adoption of treatment interventions. In treatment processes, a patient status may be changed dynamically, and new evidence often becomes available at certain time points, which inevitably influences physician decisions on treatment selection. To address this challenge, our model should incorporate richer execution information into the learning, so as to be more intelligent in terms of treatment adoption and treatment effect improvement.

Moreover, the proposed work simply uses one treatment property, namely the treatment type, as features. This is not entirely consistent with clinical practice. In actual clinical settings, medications with different dosages and frequencies may be grouped into many treatment variants according to the physical conditions of individual patients. To address this problem, the significant potential of EHR data is required to be exploited for treatment effect estimation in a fine-grained manner.

A further limitation of our proposed model is that the causal interactions between patient status and treatments are not considered. Causal interaction analysis may be useful to identify unexpected changes in patient characteristics and explain why scheduled treatments are changed to guarantee the expected treatment effects in an interpretable manner. That is, such an approach may provide interpretable prediction on treatment effects given a specific patient status. Note that this is an open medical problem and could be addressed by mining a large amount of EHR data in a maximum-informative manner. Substantial research is still necessary to make such mining both effective and efficient.

Conclusions

In this work, we have addressed quite a challenging problem in medical informatics, namely utilizing a large volume of observational data for TEP. We have proposed a novel model for extracting robust and discriminative representations of patient samples from their EHR data. We further improved the representation and discrimination power of the features by using adversarial loss to explore the correlational information between patient statuses and treatments. Our proposed model was evaluated on two real clinical datasets pertaining to ACS and HF, and collected from the cardiovascular department of a Chinese hospital. The experimental results demonstrate significant improvements in TEP compared to state-of-the-art methods. An interesting finding is that treatments are conditioned on patient clinical statuses and may result in varying outcomes. This inspires us to explore the correlations between patient characteristics and treatments further for promptly and accurately predicting treatment effects in our future work.

Abbreviations

ACEI: Angiotensin converting enzyme inhibitor; ACS: Acute coronary syndrome; ADTEP: Adversarial deep learning model for treatment effect prediction; AE: Auto-encoder; ARB: Angiotensin receptor blockers; AUC: Area Under the receiver operating characteristic Curve (AUC); BMI: Body mass index; CABG: Coronary artery bypass grafting; CCB: Calcium channel blocker; CCR: Creatinine clearance; CK: Creatinine kinase; CKMB: Creatine kinase MB; cTnT: Cardiac troponin T; DBP: Diastolic blood pressure; DTEP: Deep learning model for treatment effect prediction; ECG: Electrocardiogram; EHR: Electronic health records; GAN: Generative adversarial network; HF: Heart failure; IPS: Inverse propensity score; LR: Logistic regression; LVEF: Left ventricular ejection fraction; MACE: Major adverse event prediction; PCI: Percutaneous coronary intervention; RCT: Randomized controlled trial; ROC: Receiver operating characteristic; SBP: Systolic blood pressure; SD: Standard deviation; SVM: Support vector machine; TEP: Treatment effect prediction

Acknowledgments

The authors would like to extend special thanks to all experts who cooperated in the evaluation of the proposed method. The authors are particularly thankful for the positive support received from the cooperative hospitals as well as to all medical staff involved.

About this supplement

This article has been published as part of *BMC Medical Informatics and Decision Making Volume 20 Supplement 4 2020:* Selected articles from the *Fourth International Workshop on Semantics-Powered Data Analytics (SEPDA 2019)*. The full contents of the supplement are available at https://bmcmedinformdecismak.biomedcentral.com/articles/supplements/volume-20-supplement-4.

Authors' contributions

JC was involved in the design of the study, implementation of experiments, analysis, and drafting and critical review of the manuscript. WD was involved in data collection and provided clinical interpretation for result. JW and KH was involved in the design of the study and data collection. ZH conceived, and designed the study, critically reviewed the manuscript, and prepared the final version for publication. All authors read and approved the final manuscript.

Funding

This work is partially supported by the National Natural Science Foundation of China under Grant No. 61672450 and the National Key Research and Development Program of China under Grant No. 2018YFC2001204. Especially, publication costs are funded by the National Natural Science Foundation of China under Grant No. 61672450.

Availability of data and materials

The datasets used during current study are not available due to the data protection policy of our collaborative hospital.

Ethics approval and consent to participate

Approval by the Data Protection Committee at the Chinese PLA General Hospital was obtained prior to initiation of the study. Informed consent was waived because of the presence of deidentified data and lack of feasibility of obtaining informed consent from all participants in the experimental dataset. The ethics committee at the Chinese PLA General Hospital ruled that no formal ethics approval was required in this particular case.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]College of Biomedical Engineering and Instrumental Science, Zhejiang University, Hangzhou, China. [2]Department of Cardiology, Chinese PLA General Hospital, Beijing, China. [3]Cardiocloud Medical Technology, Beijing, China.

Received: 25 May 2020 Accepted: 8 June 2020
Published: 14 December 2020

References

1. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. New Engl. J. Med. 2000; 342(25):1887–92.
2. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika. 1983;70(1):41–55.
3. D'Agostino RB. Estimating treatment effects using observational data. JAMA. 2007;297(3):314–6.
4. Ellenberg JH. Selection bias in observational and experimental studies. Stat Med. 1994;13(5–7):557–67.
5. Ioannidis JP, Haidich AB, Pappa M, Pantazis N, Kokori SI, Tektonidou MG, Contopoulos-Ioannidis DG, Lau J. Comparison of evidence of treatment effects in randomized and nonrandomized studies. JAMA. 2001;286(7):821–30.
6. Sanson-Fisher RW, Bonevski B, Green LW, D'Este C. Limitations of the randomized controlled trial in evaluating population-based health interventions. Am J Prev Med. 2007;33(2):155–61.
7. Cartwright N, Munro E. The limitations of randomized controlled trials in predicting effectiveness. J Eval Clin Pract. 2010;16(2):260–6.
8. Deaton A, Cartwright N. Understanding and misunderstanding randomized controlled trials. Soc Sci Med. 2018;210:2–21.
9. Xiao C, Choi E, Sun J. Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review. JAMIA. 2018;25(10):1419–28.
10. Huang Z, et al. On mining latent treatment patterns from electronic medical records. Data Min Knowl Disc. 2015;29(4):914–49.

11. Huang Z, et al. A probabilistic topic model for clinical risk stratification from electronic health records. J Biomed Inform. 2015;58:28–36.
12. Huang Z, et al. MACE prediction of acute coronary syndrome via boosted resampling classification using electronic medical records. J Biomed Inform. 2017;66:161–70.
13. Huang Z, et al. A regularized deep learning approach for clinical risk prediction of acute coronary syndrome using electronic health records. IEEE Trans Biomed Eng. 2018;65(5):956–68.
14. Hruby GW, et al. Facilitating biomedical researchers' interrogation of electronic health record data: ideas from outside of biomedical informatics. J Biomed Inform. 2016;60:376–84.
15. Wu PY, et al. Omic and electronic health record big data analytics for precision medicine. IEEE Trans Biomed Eng. 2017;64(2):263–73.
16. Shalit U, Johansson FD, Sontag D. Estimating individual treatment effect: generalization bounds and algorithms. In Proc. 34th Int. Conf. Mach Learn. 2017:3076–85.
17. Yoon J. Jordon J, van der Schaar M. Estimation of individualized treatment effects using generative adversarial net. In Int. Conf. Learning Representations: GANITE; 2018.
18. Feng P, Zhou XH, Zou QM, Fan MY, Li XS. Generalized propensity score for estimating the average treatment effect of multiple treatment. Stat Med. 2012;31(7):681–97.
19. Becker SO, Ichino A, et al. Estimation of average treatment effects based on propensity scores. Stata J. 2002;2(4):358–77.
20. Huang Z, Lu X, Duan H. Using recommendation to support adaptive clinical pathways. J Med Syst. 2012;36(3):1849–60.
21. Lu X, Huang Z, Duan H. Supporting adaptive clinical treatment processes through recommendations. Comput Methods Prog Biomed. 2012;107(3): 413–24.
22. Zhang Y, Chen R, Tang J, Stewart WF, Sun J. LEAP: learning to prescribe effective and safe treatment combinations for multimorbidity. In Proceedings of KDD. 2017:1315–24.
23. Doersch C. Tutorial on Variational Autoencoders, arXiv:1606.05908. 2016.
24. Vincent P, Larochelle H, Lajoie I, Bengio Y, Manzagol PA. Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. J Mach Learn Res. 2010;11:3371–408.
25. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. Generative Adversarial Networks, arXiv:1406.2661. 2014.
26. Wager S, Athey S. Estimation and inference of heterogeneous treatment effects using random forests. ArXivpreprint arXiv:1510.04342. 2015.
27. Johansson F, Shalit U, Sontag D. Learning representations for counterfactual inference. In International Conference on Machine Learning (ICML). 2016.
28. Swaminathan A, Joachims T. Batch learning from logged bandit feedback through counterfactual risk minimization. J Mach Learn Res. 2015;16:1731–55.
29. Swaminathan A, Joachims T. The self normalized estimator for counterfactual learning. In Advances in Neural Information Processing Systems. 2015:3231–9.
30. Atan O, Jordan J, van der Schaar M. Deep-treat: learning optimal personalized treatments from observational data using neural networks. Intell: In Proc. Assoc. Adv. Artif; 2018.
31. Lee C, Mastronarde N, van der Schaar M. Estimation of individual treatment effect in latent confounder models via adversarial learning. In Proc. Mach. Learn. Health (ML4H) Workshop at NeurIPS. 2018.
32. Louizos C, Shalit U, Mooij JM, Sontag D, Zemel R, Welling M. Causal effect inference with deep latent-variable models. In Proc. Adv. Neural Inf. Process. Syst. (NIPS). 2017;6446–6456.
33. Alaa AM, Weisz M, van der Schaar M. Deep counterfactual networks with propensity-dropout. ICML Workshop Principled Approaches Deep Learn: In Proc; 2017.
34. Richter AN, Khoshgoftaar TM. A review of statistical and machine learning methods for modeling cancer risk using structured clinical data. Artif Intell Med. 2018;90:1–14.
35. Zhao C, Jiang J, Guan Y, Guo X, He B. EMR-based medical knowledge representation and inference via Markov random fields and distributed representation learning. Artif Intell Med. 2018;87:49–59.
36. Athey S and Imbens GW. Recursive partitioning for heterogeneous causal effects. arXiv preprint arXiv:1504.01132. 2015.
37. Hill JL. Bayesian nonparametric modeling for causal inference. Journal of Computational and Graphical Statistics. 2011;20(1).
38. Kingma DP, Welling M. Auto-encoding variational bayes, arXiv:1312.6114. 2013.
39. Huang Z. Dong W. IEEE Journal of Biomedical and Health Informatics: Adversarial MACE Prediction after Acute Coronary Syndrome using Electronic Health Records; 2018.
40. Mega JL, Eugene DB, Stephen W, et al. Rivaroxaban in patients with a recent acute coronary syndrome. New Engl J Med. 2012;366(1):9–9.
41. Antman EM, Cohen M, Bernink PJLM, et al. The TIMI risk score for unstable angina/non-ST elevation MI: a method for prognostication and therapeutic decision making. J Am Med Assoc. 2000;284(7):835–42.
42. Goodman SG, Huang W, Yan AT, et al. The expanded global registry of acute coronary events: baseline characteristics, management practices, and hospital outcomes of patients with acute coronary syndromes. Am Heart J. 2009;158(2):193–201.
43. D'Agostino R, Vasan R, Pencina M, Wolf A, Cobain M, Massaro JK. General cardiovascular risk profile for use in primary care: the Framingham heart study. Circulation. 2008;117:743–57.
44. Huang Z, Lu Y, Dong W. Utilizing electronic health records to predict multi-type major adverse cardiovascular events after acute coronary syndrome. Knowl Inf Syst. 2019;60(3):1725–52.
45. The GRACE. Investigators. Rationale and design of the GRACE (global registry of acute coronary events) project: a multinational registry of patients hospitalized with acute coronary syndromes. Am Heart J. 2001; 141(2):190–9.
46. Subherwal S, Bach RG, Chen AY, et al. Baseline risk of major bleeding in non–ST-segment–elevation myocardial infarction: the CRUSADE (can rapid risk stratification of unstable angina patients suppress ADverse outcomes with early implementation of the ACC/AHA guidelines) bleeding score. Circulation. 2009;119:1873–82.
47. Coronel R, de Groot JR, van Lieshout JJ. Defining heart failure. Cardiovasc Res. 2001;50(3):419.
48. 2017 ACC/AHA/HFSA Focused Update of the 2013 ACCF/AHA guideline for the Management of Heart Failure. http://wwwonlinejaccorg/content/accj/70/6/776fullpdf?_ga=2158896255451869789155538622519085087151555237127, Last access on 2019-04-16.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.