Removing Confounding Factors Associated Weights in Deep Neural Networks Improves the Prediction Accuracy for Healthcare Applications

Haohan Wang[1], Zhenglin Wu[2], Eric P. Xing[1,3]
[1]*School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA*
[2]*School of Information Sciences, University of Illinois Urbana-Champaign Champaign, IL, USA*
[3]*Petuum Inc. Pittsburgh, PA, USA*
E-mail: haohanw@cs.cmu.edu

The proliferation of healthcare data has brought the opportunities of applying data-driven approaches, such as machine learning methods, to assist diagnosis. Recently, many deep learning methods have been shown with impressive successes in predicting disease status with raw input data. However, the "black-box" nature of deep learning and the high-reliability requirement of biomedical applications have created new challenges regarding the existence of confounding factors. In this paper, with a brief argument that inappropriate handling of confounding factors will lead to models' sub-optimal performance in real-world applications, we present an efficient method that can remove the influences of confounding factors such as age or gender to improve the across-cohort prediction accuracy of neural networks. One distinct advantage of our method is that it only requires minimal changes of the baseline model's architecture so that it can be plugged into most of the existing neural networks. We conduct experiments across CT-scan, MRA, and EEG brain wave with convolutional neural networks and LSTM to verify the efficiency of our method.

Keywords: neural networks, healthcare, confounding factor correction

1. Introduction

The increasing amount of data has led healthcare to a new era where the diagnosis can be made directly from raw data such as CT-scan or MRI with data-driven approaches. Machine learning methods, especially deep learning methods, have achieved significant successes in biomedical and healthcare applications, such as classifying lung nodule,[1] breast lesions,[2] or brain lesions[3] from CT-scans, segmentation of brain regions with MRI,[4,5] or emotion classification with EEG data.[6,7]

However, different from how deep learning has revolutionized many other applications, the "black-box" nature of deep learning and the high-reliability requirement of healthcare industry have created new challenges.[8] One of these challenges is about removing the false signals extracted by deep learning methods due to the existence of confounding factors. Acknowledging the recognition mistakes made by neural networks[9–11] and empirical evidence that deep neural networks can learn signals from confounding factors,[12] it is likely that a well-trained deep learning model will exhibit limited predictive performance on external data sets despite its high predictive power on lab collected data sets. The hazard of inappropriate control of

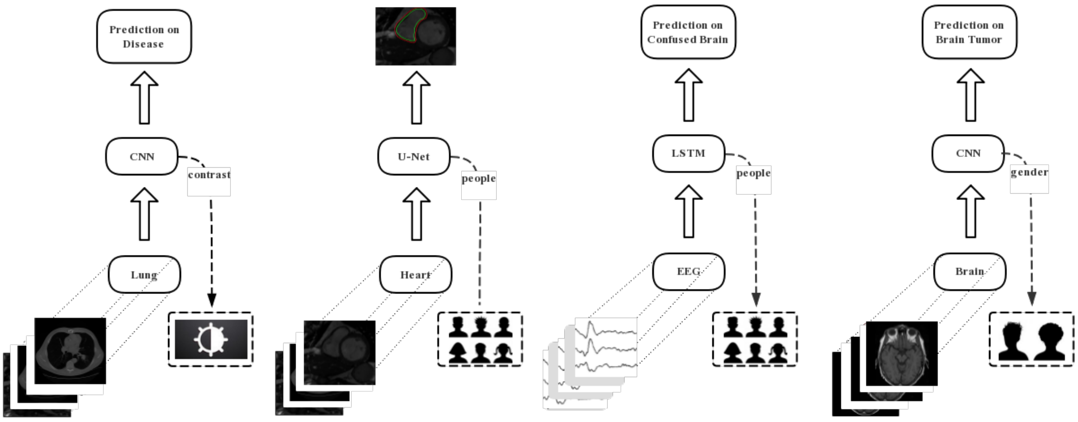


Fig. 1: An illustration of the empirical contribution of this paper. From left to right, 1) lung adenocarcinoma prediction from CT-scan with CNN, where contrast material is the confounding factor, 2) heart right ventricle segmentation from CT-scan with U-net, where subject identification is the confounding factor, 3) students' confusion status prediction from EEG signals with Bidirectional LSTM, where the students' demographic information is the confounding factor, 4) brain tumor prediction from CT-scan/MRA with CNN, where gender associated information is the confounding factor.

confounding factors in healthcare-related science has been discussed extensively,[13–15] but these discussions are mainly in the scope of causal analyses or association studies.

In addition to a very recent result showing that confounding factors can adversely affect the predictive performance of neural network models,[16] we offer a straightforward example as another motivation: a neural network predictive model for Hodgkin lymphoma diagnosis is trained on a data set collected from young volunteers with high predictive performance, but when the model is applied to the entire society, it may report more false positives than expected. One of the reason could be that the gender ratio reverses toward adolescence in Hodgkin lymphoma,[17] and a model trained over data collected from young volunteers is very likely to learn a different gender bias than what is expected in a data collected different age groups. In fact, even if the gender ratio does not change along the aging process, it is still inappropriate for a model to predict based on features related to gender because these features are not directly associated with disease status. As another example, skin cancer[18] and colorectal cancer[19] are also observed with gender bias, and it is already observed that there is a higher false negative rate in colorectal cancer diagnosis for women[19] with traditional methods. Confounding factors do not just exist in the forms of gender. Also, it is observed that other factors, such as age,[20] or demographic information,[21] will affect the model's performance if not handled appropriately. Considering that the generalization theory of neural networks is still an open research topic and people are unsure of how neural networks predict, it is particularly important to design methods to handle the influence of these confounding factors explicitly.

In this paper, inspired by previous de-confounding techniques applied to deep learning models,[12] we propose a Confounder Filtering (CF) method. A distinct advantage of our method is that CF directly builds upon the original confounded neural network with a minimal change that replaces the original top layer with a layer that predicts the confounding factors. Further, we apply our methods to a broad spectrum of related tasks, such as:

- improved lung adenocarcinoma prediction with convolutional neural networks (CNN) by removing contrast material as confounding factors.
- improved heart right ventricle segmentation with U-net by removing subject identifications as confounding factors.
- improved students' confusion status prediction with Bidirectional LSTM by removing students' demographic information as confounding factors.
- improved brain tumor prediction with CNN by removing gender associated information as confounding factors.

We have observed consistent improvements in predictive performance by removing the confounding factors. These four empirical contributions have been conveniently summarized in Figure 1, which illustrates the experiments we perform in this paper, including the predictive task, the model we use, the data, and the confounding factors.

The remainder of this paper is organized as follows. In Section 2, we first briefly discuss the related work of this paper, mainly in the methodological perspective. In Section 3, we formally introduce our method, namely Confounder Filtering. Then in Section 4, we apply our method to a wide spectrum of experiments to show the effectiveness of our method and report relevant analysis. Finally, we conclude this paper with discussion of limitations and future directions in Section 5.

2. Related Work

The recent boom of deep learning techniques has allowed a large number of neural network methods developed for healthcare applications rapidly. Readers can refer to comprehensive reviews on how the deep learning can be applied to healthcare and biomedical areas.[8,22–24] In this section, we will mainly discuss the related work of our paper in the methodological perspective.

To the best of our knowledge, there are not many deep learning works that control the effects of confounding factors explicitly. Wang *et al* presented a two-phase algorithm named Select-Additive Learning.[12] In the first phase, the model uses information of confounding factors to select which components of the representation learned by neural networks are associated with confounding factors, and then in the addition phase, the algorithm forces the neural networks to discard these components by adding noises. Zhong *et al* also discussed how confounding factors affect the predictive performance of neural networks. They presented an augment training framework that requires little additional computational costs.[25] The idea is to add another neural classifier that predicts confounding factors while predicting original labels, and gradient descent optimizes both of these classifiers. The general additional structure is very similar to the Confounding Filtering method that we are going to present, but our method trains the network in differently so that we can differentiate the weights associated with confounding factors and filter them out explicitly.

In a broader view, correcting confounding factors is related to reducing the representations learned by neural networks through some components of the raw data that are not related to the predictive task. In this perspective, there is a significant amount of neural network methods that can be considered as related work, covering the fields such as domain adaptation,[26] transfer learning,[27,28] and domain generalization.[29] Readers can refer to the survey papers cited and the references therein if interested. Within the scope of this paper, we do not discuss with these methods for two reasons: 1) these methods are not designed for correcting confounding factors explicitly, therefore they may or may not be applicable in this specific situation, 2) even if our CF method behave similar to, or slightly shy of the performance of these methods, there is still a distinct advantage: CF is simple enough to be plugged into any neural networks with almost no changes of the architecture.

3. Confounder Filtering (CF) Method

In this section, we will formally introduce the Confounder Filtering (CF) method. CF method's goal is to reduce the effects of confounders, therefore improves the generalizability of deep neural networks. We first offer an intuitive overview of the main idea of CF, then we formalize our method, which is followed by a discussion of the availability of the implementation.

3.1. *Overview*

CF method is aimed to remove the effects of confounding factors by removing the weights that are associated with them. Therefore, the core step is to identify such weights. We first train a model, namely G, conventionally for the predictive task. Then we replace the top model layer with another classifier that predicts the labels of confounding factors, and we continue to train the model. During this training phase, we keep track of the updates of weights. Finally, we filter out all the weights that are frequently updated during this training phase out of G by replacing these weights with zeros, leading to a new confounder-free model. This process is illustrated in Fig. 2.

3.2. *Method*

We continue to formalize our method. For the convenience of discussion, we split a deep neural network architecture into two components: representation learner component and classification component, denoted by $g(\cdot;\theta)$ and $f(\cdot;\phi)$ respectively, where θ and ϕ stand for the corresponding parameters. Therefore, the complete neural network classifier is denoted as $f(g(\cdot;\theta);\phi)$. Given data $< y, X >$, the classical training process of the neural network is achieved via solving the following equation:

$$\hat{\theta}, \hat{\phi} = \underset{\theta,\phi}{\operatorname{argmin}} \; c(y, f(g(X;\theta);\phi)) \tag{1}$$

where $c(\cdot,\cdot)$ stands for the cost function, with famous examples such as mean-squared-error loss or cross-entropy loss.

Ideally, to effectively remove the effects of confounding factors, a method needs the labels of the confounding factors. In other words, we need data in the form of $< X, y, s >$, where s

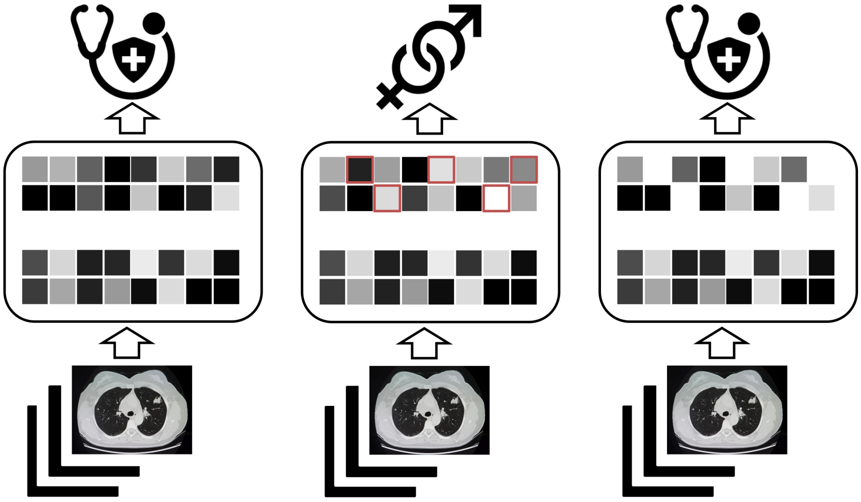


Fig. 2: This figure shows the overview of the CF method. From left to right: 1) Train the neural network conventionally. 2) Train the neural network to predict confounding factors (*e.g.* gender information) and inspect the changes of weights each iteration to locate the ones with largest changes. 3) Remove the located weights, then the model is ready for confounder-free prediction.

stands for the label of the confounding factors (*e.g.* age, gender, physical factors of medical devices *etc.*). This is also required by similar previous work.[12,25] However, our method does not require full correspondence between X, y, and s. For example, later in our experiment, we will show that with two independently collected data sets $< X_1, y_1 >$ and $< X_2, s_2 >$(*i.e.* we only have correspondence between X_1 and y_1, and between X_2 and s_2, but not between y_1 and s_2), we are able to correct the confounding factors between X_1 and y_1 with help of X_2 and s_2. For simplicity, we still present our method with $< X, y, s >$.

After we train the neural network following the conventional manner as showed in Equation 1 with $< X, y >$ and get $\hat{\theta}$ and $\hat{\phi}$, we continue to identify the weights associating with confounding factors through tuning the classification component via $< X, s >$. Formally, we solve the following problem:

$$\tilde{\phi} = \underset{\phi}{\text{argmin}}\ c(s, f(g(X; \hat{\theta}); \phi))$$

During the optimization, our method inspects how the gradient of the cost function with respect to $< X, s >$ updates the previous trained weights (*i.e.* $\hat{\phi}$) with $< X, y >$. For the i^{th} value of ϕ (denoted as ϕ_i), we calculate the frequency of updating it during the entire training process (denoted as π_i). Formally, we have:

$$\pi_i = \frac{1}{n}\sum_{t=1}^{n} |\Delta\phi_{i,t}|$$

where n is the number of total steps, t stands for the index of step.

Further, we construct a masking matrix/tensor M of the same shape as ϕ, and M_i is constructed according to π_i. For example, common choices could be either through a Bernoulli sampling

$$M_i = \mathrm{Ber}(\pi_i)$$

or a straightforward thresholding procedure:

$$M_i = \begin{cases} 0, & \pi_i > \tau \\ 1, & \text{otherwise} \end{cases}$$

In the following experiment, we choose to use the thresholding procedure with τ, whose value lies between top 20% and top 25% of π_i's values.

Finally, we have $\hat{\phi}' = \hat{\phi} \otimes M$, where $\otimes$ stands for element-wise product, and the final trained neural network after confounding factor associated weights filtered out is as following:

$$f(g(X; \hat{\theta}); \hat{\phi}')$$

which is ready for confounder-free prediction.

3.3. *Availability*

The implementation of our method in TensorFlow is available online[a] with a simple example that trains a CNN for Cifar10 dataset, onto which we add some image patterns as confounding factors. Users can follow the online instruction to apply CF to their own customized neural networks.

4. Experiments

In this section, we will verify the performance of our CF method on four different tasks by adding CF towards the current baseline models. For each task, we will first introduce the data set, and then introduce the methods we compare and the results. After discussions of these four tasks, we will introduce some analyses of the model behaviors to further validate the performance of our method.

4.1. *lung adenocarcinoma prediction*

4.1.1. *Data*

We construct a data set to test the model performance in classifying adenocarcinomas and healthy lungs from CT-scans. Our experimental data set is a composition of three data sets:

- **Data Set 1:** The CT-images from healthy people are collected from ELCAP Public Lung Image Database[b]. The CT scans have obtained in a single breath hold with a 1.25 mm slice thickness that consists of 1310 DICOM images from 25 persons.
- **Data Set 2:** The CT-scans of diseased lungs are collected from 69 different patients by Grove *et al.*[30] These scans are diagnostic contrast-enhanced CT scans, being done at diagnosis and prior to surgery and slice thickness at variable from 3 to 6 mm.

[a] https://github.com/HaohanWang/CF

[b] http://www.via.cornell.edu/lungdb.html

- **Data Set 3:** Since these two data sets are collected differently, and one of them is a collection of contrast-enhanced CT scans. The contrast material will likely serve as the confounding factor in prediction. To correct the confounding factor. We noticed a processed version[c] of **Data Set 2**, which consists of explicit labels of contrast information. The data set contains 475 series from 69 different patients selected 50% with contrast and 50% without contrast.

Therefore, we use the 1290 healthy images from 20 persons in **Data Set 1** and 1214 diseased lung images from 61 patients in **Data Set 2** as the training set, and the rest from these two data sets as the testing set. We use the images from **Data Set 3** with corresponding contrast labels to correct confounding factors.

4.1.2. *Results*

We experiment with the most popular architectures of CNNs, including AlexNet,[31] CifarNet,[32] LeNet,[33] VGG16,[34] and VGG19.[34] We first sufficiently train these baseline models with appropriate learning rate until the training accuracy converges, and then use our CF method to correct the confounding factors. We test the prediction accuracy of both vanilla CNNs and CF-improved CNNs. Fig. 3 shows the results. We can see that CF can consistently improve the predictive results over a variety of different CNNs.

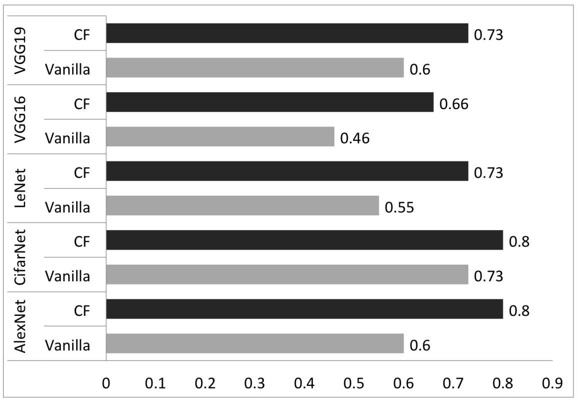


Fig. 3: Prediction accuracy of CNN in comparison with CF-CNN

4.2. *Segmentation on right ventricle(RV) of Heart*

4.2.1. *Data*

The data set[35] contains 243 physician-segmented CT images (216×256 pixels) from 16 patients. Data augmentation techniques, such as random rotations, translations, zooms, shears and elastic deformations (locally stretch and compress the image), are used to increase the number of samples. More information regarding the data set, including how the training/testing data sets are split, can be found online[d].

4.2.2. *Results*

The main baseline in this experiment is U-net, which is a convolutional network architecture for fast and precise segmentation of images. Previous experiments show that U-net can behave well

[c] https://www.kaggle.com/kmader/siim-medical-images/home

[d] https://blog.insightdatascience.com/heart-disease-diagnosis-with-deep-learning-c2d92c27e730

even with a small dataset.[36] We first test U-net following previous setting[35] and interestingly, we achieve a higher accuracy that what was reported. Vanilla U-net achieves an accuracy of 0.9477. Then, we use CF method to remove the subject identities as confounding factors and improve the accuracy from 0.9477 to **0.9565**.

4.3. *Students' confusion status prediction*

4.3.1. *Data*

The data set[37] contains EEG brainwave data from 10 college students while they watch MOOC video clips[e]. The EEG data is collected rom MindSet equipment wore by college students when watch ten video clips, five out of which are confusing ones. The students' identities are considered as confounding factors in this experiment.

Following previous work,[38] we normalize the training data in a feature-wise fashion (*i.e.*, each feature representation is normalized to have a mean of 0 and standard deviation of 1 across each batch of samples). The batch size is set to 20.

Table 1: Comparison with average accuracy for 5-fold cross validation[38]

Methods	Accuracy(%)
SVM	67.2
KNN	51.9
CNN	64.0
DBN	52.7
RNN-LSTM	69.0
BiLSTM	73.3
CF-BiLSTM	**75.0**

4.3.2. *Results*

We use the state-of-the-art method applied to this data set,[38] namely a Bidirectional LSTM, as the baseline method to compare with. The model is configured as following: the LSTM layer has 50 units, with *tanh* as activation function. The output is connected to a fully connected layer with a sigmoid activation. We compare five-fold-cross-validated results from CF-improved Bidirectional LSTM with results reported previously.[38] The results are shown in Table 1. As we can see, CF method helps improve the predictive performance once plugged in.

4.4. *Brain tumor prediction*

4.4.1. *Data*

We construct another data set for the last experiment of this paper. We test our method in predicting brain tumors with MRA scans of healthy brain[f] and CT-scans with tumor brain.[39] The healthy data set consists of images of the brain from 100 healthy subjects, in which 20 patients were scanned per decade and each group are equally divided by sex. The tumor data set is collected with 120 patients. The gender information is regarded as confounding factors in this experiment.

[e] https://www.kaggle.com/wanghaohan/confused-eeg/home

[f] http://insight-journal.org/midas/community/view/21

4.4.2. *Results*

Similar to the lung adenocarcinoma prediction experiment, we compare with the set of popular CNNs. The results are shown in Fig. 4. As we can see that, CF helps improve the prediction performance in most cases, except that in the VGG19 cases, when the model's performance deteriorates after CF is plugged in.

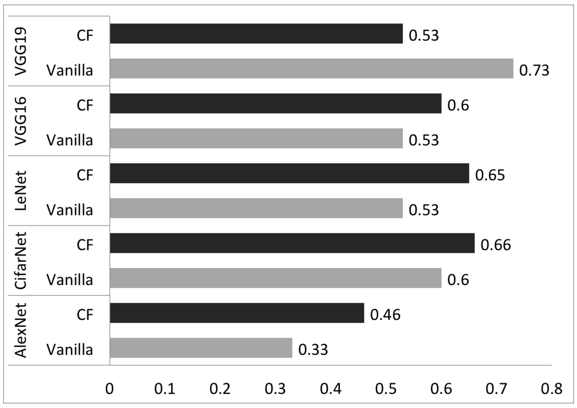


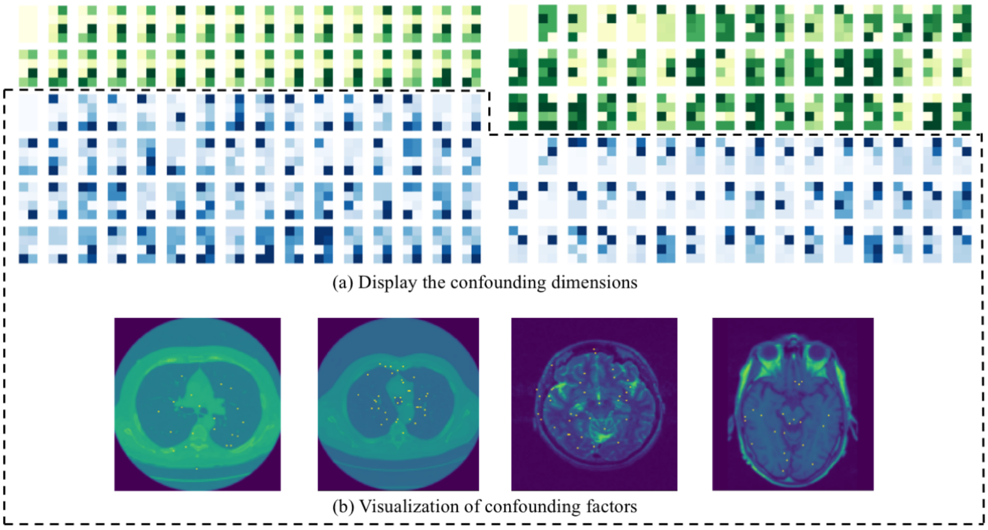
Fig. 4: Prediction accuracy of CNN in comparison with CF-CNN

4.5. *Analyses of the method behaviors*

To further understand the process of CF in identifying the weights that are associated with the confounding factors. We inspect how the weights are updated during the training process and visualize which part of the input data is related to confounding factors.

Fig. 5(a) visualizes the weights during each epoch. The figure splits into two panels, and the left panel is for lung adenocarcinoma prediction experiment, and the right panel is for brain tumor prediction experiment. The figure only shows eight weights of the top layer (in a 4×2 rectangle), and visualizes how the weights in the layer change as the training epoch increases. This figure visualizes 96 epochs for lung adenocarcinoma prediction and brain tumor prediction each. The blue dots visualize the weights when the model is trained during the first phase, and the green dots visualize the weights when the model is trained in the second phase for prediction confounding factors. The darker each dot is, the more frequent it gets updated in that epoch. As we can see, for the same 4×2 layer, the frequencies of the weights get updated are different between the training during the first phase and training during the second phase. This differences of updating frequencies verify the primary assumption of our method, that the weights associated with the task and the weights associated with the confounding factors are different. Therefore, we can remove the effects of confounding factors by removing the weights associated with them.

Further, we try to investigate which parts of the input data are corresponding to the confounding factors. With the help of Deep Feature Selection[40] method, we select the pixels of the image that are associated with the confounding factors. Fig 5 visualizes these pixels with yellow dots. From left to right, these four images are examples for healthy lung, diseased lung, healthy brain, tumorous brain respectively. Interestingly, we do not see clear patterns on the images that are related to the confounding factors. This observation further verify the importance of our CF method because these results indicate that it is barely possible to first exclude the information from raw images by conventional methods since these yellow dots do not form into any clear pattern.


(a) Display the confounding dimensions

(b) Visualization of confounding factors

Fig. 5: (a) Display of trained weights and (b) the visualization of confounding factors.

5. Conclusion

In this paper, we proposed a straightforward method, named Confounder Filtering, which aims to reduce the effects of confounders and improve the generalizability of deep neural networks, to achieve a confounding-factor-free predictive model for healthcare applications. One distinct advantage of our method is that we only require minimal changes to the existing network model to adopt our method. There are still limitations of our method: despite our method only requires a minimal changes of the network architecture, it needs a repeated training process (the second phase training with confounding factors). Another limitation is that our method still requires the switching of the top classification layer from a label predictor to a confounder predictor, which may lose the one-to-one correspondence of weights at the top layer. In the future, in the methodological perspective, we look forward to further improving the training process of our method. On the practical side, as we have released our code, we hope to help the community to increase the performance of other predictive models for healthcare application by removing the confounding factors.

6. Acknowledgement

The authors would like to thank Mingze Cao and Yin Chen for discussions and creation of Fig 1 and Fig 5. This work is funded and supported by the Department of Defense under Contract No. FA8721-05-C-0003 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center. This work is also supported by the National Institutes of Health grants R01-GM093156 and P30-DA035778. The MR brain images from healthy volunteers used in this paper were collected and made available by the CASILab at The University of North Carolina at Chapel Hill and were distributed by the MIDAS Data Server at Kitware, Inc

References

1. K.-L. Hua, C.-H. Hsu, S. C. Hidayati, W.-H. Cheng and Y.-J. Chen, Computer-aided classification of lung nodules on computed tomography images via deep learning technique, *OncoTargets and therapy* **8** (2015).
2. J.-Z. Cheng, D. Ni, Y.-H. Chou, J. Qin, C.-M. Tiu, Y.-C. Chang, C.-S. Huang, D. Shen and C.-M. Chen, Computer-aided diagnosis with deep learning architecture: applications to breast lesions in us images and pulmonary nodules in ct scans, *Scientific reports* **6**, p. 24454 (2016).
3. X. W. Gao, R. Hui and Z. Tian, Classification of ct brain images based on deep learning networks, *Computer methods and programs in biomedicine* **138**, 49 (2017).
4. A. Işın, C. Direkoğlu and M. Şah, Review of mri-based brain tumor image segmentation using deep learning methods, *Procedia Computer Science* **102**, 317 (2016).
5. F. Milletari, S.-A. Ahmadi, C. Kroll, A. Plate, V. Rozanski, J. Maiostre, J. Levin, O. Dietrich, B. Ertl-Wagner, K. Bötzel *et al.*, Hough-cnn: deep learning for segmentation of deep brain regions in mri and ultrasound, *Computer Vision and Image Understanding* **164**, 92 (2017).
6. S. Jirayucharoensak, S. Pan-Ngum and P. Israsena, Eeg-based emotion recognition using deep learning network with principal component based covariate shift adaptation, *The Scientific World Journal* **2014** (2014).
7. W.-L. Zheng, J.-Y. Zhu, Y. Peng and B.-L. Lu, Eeg-based emotion classification using deep belief networks, in *Multimedia and Expo (ICME), 2014 IEEE International Conference on*, 2014.
8. R. Miotto, F. Wang, S. Wang, X. Jiang and J. T. Dudley, Deep learning for healthcare: review, opportunities and challenges, *Briefings in bioinformatics* (2017).
9. C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow and R. Fergus, Intriguing properties of neural networks, *arXiv preprint arXiv:1312.6199* (2013).
10. A. Nguyen, J. Yosinski and J. Clune, Deep neural networks are easily fooled: High confidence predictions for unrecognizable images, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015.
11. H. Wang, B. Raj and E. P. Xing, On the origin of deep learning, *arXiv preprint arXiv:1702.07800* (2017).
12. H. Wang, A. Meghawat, L. P. Morency and E. P. Xing, Select-additive learning: Improving generalization in multimodal sentiment analysis, in *IEEE International Conference on Multimedia and Expo*, 2017.
13. M. A. Brookhart, T. Stürmer, R. J. Glynn, J. Rassen and S. Schneeweiss, Confounding control in healthcare database research: challenges and potential approaches, *Medical care* **48**, p. S114 (2010).
14. A. C. Skelly, J. R. Dettori and E. D. Brodt, Assessing bias: the importance of considering confounding, *Evidence-based spine-care journal* **3**, 9 (2012).
15. M. Nørgaard, V. Ehrenstein and J. P. Vandenbroucke, Confounding in observational studies based on large health care databases: problems and potential solutions–a primer for the clinician, *Clinical epidemiology* **9**, p. 185 (2017).
16. J. R. Zech, M. A. Badgeley, M. Liu, A. B. Costa, J. J. Titano and E. K. Oermann, Confounding variables can degrade generalization performance of radiological deep learning models, *arXiv preprint arXiv:1807.00431* (2018).
17. M. T. Dorak and E. Karpuzoglu, Gender differences in cancer susceptibility: an inadequately addressed issue, *Frontiers in genetics* **3**, p. 268 (2012).
18. D. N. Syed and H. Mukhtar, Gender bias in skin cancer: role of catalase revealed, *Journal of Investigative Dermatology* **132**, 512 (2012).
19. S.-E. Kim, H. Y. Paik, H. Yoon, J. E. Lee, N. Kim and M.-K. Sung, Sex-and gender-specific disparities in colorectal cancer risk, *World journal of gastroenterology: WJG* **21**, p. 5167 (2015).

20. R. Guerreiro and J. Bras, The age factor in alzheimers disease, *Genome medicine* **7**, p. 106 (2015).
21. C. Fincher, J. E. Williams, V. MacLean, J. J. Allison, C. I. Kiefe and J. Canto, Racial disparities in coronary heart disease: a sociological view of the medical literature on physician bias., *Ethnicity & disease* **14**, 360 (2004).
22. C. Angermueller, T. Pärnamaa, L. Parts and O. Stegle, Deep learning for computational biology, *Molecular systems biology* **12**, p. 878 (2016).
23. D. Ravı, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo and G.-Z. Yang, Deep learning for health informatics, *IEEE journal of biomedical and health informatics* **21**, 4 (2017).
24. T. Yue and H. Wang, Deep learning for genomics: A concise overview, *arXiv preprint arXiv:1802.00810* (2018).
25. Y. Zhong and G. Ettinger, Enlightening deep neural networks with knowledge of confounding factors, in *Computer Vision Workshop (ICCVW), 2017 IEEE International Conference on*, 2017.
26. M. Wang and W. Deng, Deep visual domain adaptation: A survey, *Neurocomputing* (2018).
27. K. Weiss, T. M. Khoshgoftaar and D. Wang, A survey of transfer learning, *Journal of Big Data* **3**, p. 9 (2016).
28. S. Moon, S. Kim and H. Wang, Multimodal transfer deep learning with applications in audio-visual recognition, *arXiv preprint arXiv:1412.3121* (2014).
29. K. Muandet, D. Balduzzi and B. Schölkopf, Domain generalization via invariant feature representation, in *International Conference on Machine Learning*, 2013.
30. O. Grove, A. E. Berglund, M. B. Schabath, H. J. Aerts, A. Dekker, H. Wang, E. R. Velazquez, P. Lambin, Y. Gu, Y. Balagurunathan *et al.*, Quantitative computed tomographic descriptors associate tumor shape complexity and intratumor heterogeneity with prognosis in lung adenocarcinoma, *PloS one* **10**, p. e0118261 (2015).
31. A. Krizhevsky, I. Sutskever and G. E. Hinton, Imagenet classification with deep convolutional neural networks, in *International Conference on Neural Information Processing Systems*, 2012.
32. J. Hosang, M. Omran, R. Benenson and B. Schiele, Taking a deeper look at pedestrians, in *Computer Vision and Pattern Recognition*, 2015.
33. O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein *et al.*, Imagenet large scale visual recognition challenge, *International Journal of Computer Vision* **115**, 211 (2015).
34. K. Simonyan and A. Zisserman, Very deep convolutional networks for large-scale image recognition, *Computer Science* (2014).
35. C.-H. Yee, Heart disease diagnosis with deep learning: State-of-the-art results with 60x fewer parameters `https://blog.insightdatascience.com/heart-disease-diagnosis-with-deep-learning-c2d92c27e730`.
36. O. Ronneberger, P. Fischer and T. Brox, U-net: Convolutional networks for biomedical image segmentation, in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015.
37. H. Wang, Y. Li, X. Hu, Y. Yang, Z. Meng and K.-m. Chang, Using eeg to improve massive open online courses feedback interaction., in *AIED Workshops*, 2013.
38. Z. Ni, A. C. Yuksel, X. Ni, M. I. Mandel and L. Xie, Confused or not confused?: Disentangling brain activity from eeg data using bidirectional lstm recurrent neural networks, in *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology,and Health Informatics*, ACM-BCB '17 (ACM, New York, NY, USA, 2017).
39. L. Scarpace *et al.*, Data from rembrandt. the cancer imaging archive (2015).
40. Y. Li, C.-Y. Chen and W. W. Wasserman, Deep feature selection: Theory and application to identify enhancers and promoters., in *RECOMB*, 2015.