Improving Outcome Prediction of Pulmonary Embolism by De-biased Multi-modality Model

Zhusi Zhong[1,2,3], Jie Li[1], Shreyas Kulkarni[2,3], Yang Li[4], Fayez H. Fayad[2,3], Helen Zhang[2,3], Sun Ho Ahn[2,3], Harrison Bai[5], Xinbo Gao[1], Michael K. Atalay[2,3], and Zhicheng Jiao[2,3](✉)

[1] School of Electronic Engineering, Xidian University, Xi'an, China
[2] Warren Alpert Medical School, Brown University, Providence, USA
zhicheng_jiao@brown.edu
[3] Department of Diagnostic Imaging, Rhode Island Hospital, Providence, USA
[4] School of Computer Science and Engineering, Central South University, Changsha, China
[5] Department of Radiology and Radiological Sciences, Johns Hopkins University School of Medicine, Baltimore, USA

Abstract. Bias in healthcare negatively impacts marginalized populations with lower socioeconomic status and contributes to healthcare inequalities. Eliminating bias in AI models is crucial for fair and precise medical implementation. The development of a holistic approach to reducing bias aggregation in multimodal medical data and promoting equity in healthcare is highly demanded. Racial disparities exist in the presentation and development of algorithms for pulmonary embolism (PE), and deep survival prediction model can be de-biased with multimodal data. In this paper, we present a novel survival prediction (SP) framework with demographic bias disentanglement for PE. The CTPA images and clinical reports are encoded by the state-of-the-art backbones pretrained with large-scale medical-related tasks. The proposed de-biased SP modules effectively disentangle latent race-intrinsic attributes from the survival features, which provides a fair survival outcome through the survival prediction head. We evaluate our method using a multimodal PE dataset with time-to-event labels and race identifications. The comprehensive results show an effective de-biased performance of our framework on outcome predictions.

Keywords: Pulmonary Embolism · Deep Survival Prediction · De-Bias learning · Multi-modal learning

1 Introduction

Bias in medicine has demonstrated a notable challenge for providing comprehensive and equitable care. Implicit biases can negatively affect patient care, particularly for marginalized populations with lower socioeconomic status [30]. Evidence

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-43904-9_50.

H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14224, pp. 515–525, 2023.
https://doi.org/10.1007/978-3-031-43904-9_50

has demonstrated that implicit biases in healthcare providers could contribute to exacerbating these healthcare inequalities and create a more unfair system for people of lower socioeconomic status [30]. Based on the data with racial bias, the unfairness presents in developing evaluative algorithms. In an algorithm used to predict healthcare costs, black patients who received the same health risk scores as white patients were consistently sicker [21]. Using biased data for AI models reinforces racial inequities, worsening disparities among minorities in healthcare decision-making [22].

Within the radiology arm of AI research, there have been significant advances in diagnostics and decision making [19]. Along these advancements, bias in healthcare and AI are exposing poignant gaps in the field's understanding of model implementation and their utility [25,26]. AI model quality relies on input data and addressing bias is a crucial research area. Systemic bias poses a greater threat to AI model's applications, as these biases can be baked right into the model's decision process [22].

Pulmonary embolism (PE) is an example of health disparities related to race. Black patients exhibit a 50% higher age-standardized PE fatality rate and a twofold risk for PE hospitalization than White patients [18,24]. Hospitalized Black patients with PE were younger than Whites. In terms of PE severity, Blacks received fewer surgical interventions for intermediate PE but more for high-severity PE [24]. Racial disparities exist in PE and demonstrate the inequities that affect Black patients. The Pulmonary Embolism Severity Index (PESI) is a well-validated clinical tool based on 11 clinical variables and used for outcome prediction measurement [2]. Survival analysis is often used in PE to assess how survival is affected by different variables, using a statistical method like Kaplan-Meier method and Cox proportional-hazards regression model [7,12,14].

However, one issue with traditional survival analysis is bias from single modal data that gets compounded when curating multimodal datasets, as different combinations of modes and datasets create with a unified structure. Multimodal data sets are useful for fair AI model development as the bias complementary from different sources can make de-biased decisions and assessments. In that process, the biases of each individual data set will get pooled together, creating a multimodal data set that inherits multiple biases, such as racial bias [1,15,23]. In addition, it has been found that creating multimodal datasets without any de-biasing techniques does not improve performance significantly and does increase bias and reduce fairness [5]. Overall, a holistic approach to model development would be beneficial in reducing bias aggregation in multimodal datasets. In recent years, Disentangled Representation Learning (DRL) [4] for bias disentanglement improves model generalization for fairness [3,6,27].

We developed a PE outcome model that predicted mortality and detected bias in the output. We then implemented methods to remove racial bias in our dataset and model and output unbiased PE outcomes as a result. Our contributions are as follows: (1) We identified bias diversity in multimodal information using a survival prediction fusion framework. (2) We proposed a de-biased survival prediction framework with demographic bias disentanglement. (3) The multimodal CPH learning models improve fairness with unbiased features.

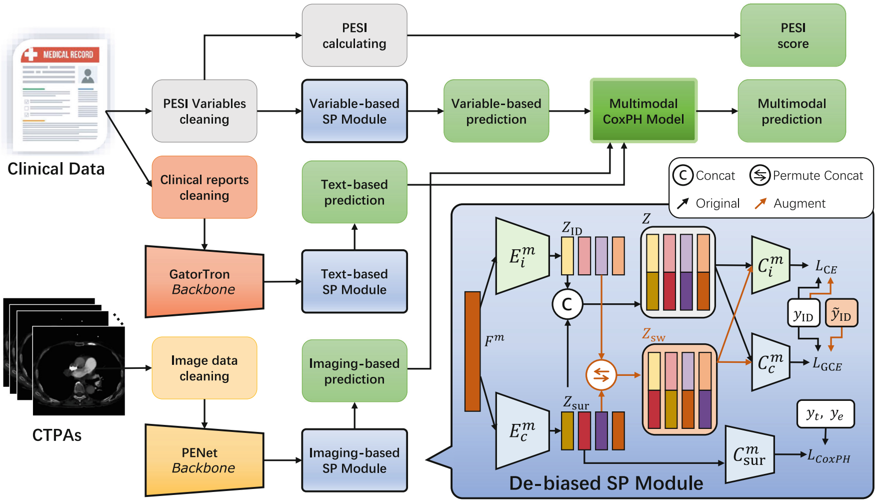

Fig. 1. Overview of the Survival Prediction (SP) framework and the proposed de-biased SP module (lower right). ID branch (E_i;C_i) and Survival branch (E_c;C_c) are trained to disentangle race-intrinsic attributes and survival attributes with the feature swapping augmentation, respectively. The survival head predicts the outcomes based on the de-biased survival attributes.

2 Bias in Survival Prediction

This section describes the detail of how we identify the varying degrees of bias in multimodal information and illustrates bias using the relative difference in survival outcomes. We will first introduce our pulmonary embolism multimodal datasets, including survival and race labels. Then, we evaluate the baseline survival learning framework without de-biasing in the various racial groups.

Dataset. The Pulmonary Embolism dataset used in this study from 918 patients (163 deceased, median age 64 years, range 13–99 years, 52% female), including 3978 CTPA images and 918 clinical reports, which were identified via retrospective review across three institutions. The clinical reports from physicians that provided crucial information are anonymized and divided into four parts: medical history, clinical diagnosis, observations and radiologist's opinion. For each patient, the race labels, survival time-to-event labels and PESI variables are collected from clinical data, and the 11 PESI variables are used to calculate the PESI scores, which include age, sex, comorbid illnesses (cancer, heart failure, chronic lung disease), pulse, systolic blood pressure, respiratory rate, temperature, altered mental status, and arterial oxygen saturation at the time of diagnosis [2].

Diverse Bias of Multimodal Survival Prediction Model. We designed a deep survival prediction (SP) baseline framework for multimodal data as shown in Fig. 1, which compares the impact of different population distributions. The frameworks without de-basing are evaluated for risk prediction in the test set by performing survival prediction on CTPA images, clinical reports, and clinical variables, respectively. First, we use two large-scale data-trained models as backbones to respectively extract features from preprocessed images and cleaned clinical reports. A state-of-the-art PE detecting model, PENet [11] is used as the backbone model for analyzing imaging risk and extracting information from multiple slices of volumetric CTPA scans to locate the PE. The feature with the highest PE probability from a patient's multiple CTPAs is considered as the most PE-related visual representation. Next, the GatorTron [29] model is employed to recognize clinical concepts and identify medical relations for getting accurate patient information from PE clinical reports. The extracted features from the backbones and PESI variables are represented as $F^m, m \in [\mathrm{Img}, \mathrm{Text}, \mathrm{Var}]$. The survival prediction baseline framework, built upon the backbones, consists of three multi-layer perceptron (MLP) modules named Imaging-based, Text-based and Variable-based SP modules. To encode survival features Z^m_{sur} from image, text and PESI variables, these modules are trained to distinguish critical disease from non-critical disease with Cox partial log-likelihood loss (CoxPHloss) [13]. The framework also consists of a Cox proportional hazard (CoxPH) model [7] that is trained to predict patient ranking using a multimodal combination of risk predictions from the above three SP modules. These CoxPH models calculate the corresponding time-to-event evaluation and predict the fusion of patients' risk as the survival outcome. We evaluate the performance of each module with concordance probability (C-index), which measures the accuracy of prediction in terms of ranking the order of survival times [8]. For reference, the C-index of PESI scores is additionally provided for comparative analysis.

In Table 1 (Baseline), we computed the C-index between the predicted risk of each model and time-to-event labels. When debiasing is not performed, significant differences exist among the different modalities, with the image modality exhibiting the most pronounced deviation, followed by text and PESI variables. The biased performance of the imaging-based module is likely caused by the richness of redundant information in images, which includes implicit features such as body structure and posture that reflect the distribution of different races. This redundancy leads to model overfitting on race, compromising the fairness of risk prediction across different races. Besides, clinical data in the form of text reports and PESI variables objectively reflect the patient's physiological information and the physician's diagnosis, exhibiting smaller race biases in correlation with survival across different races. Moreover, the multimodal fusion strategy is found to be effective, yielding more relevant survival outcomes than the clinical gold standard PESI scores.

3 De-biased Survival Prediction Model

Based on our SP baseline framework and multimodal findings from Sect. 2, we present a feature-level de-biased SP module that enhances fairness in survival outcomes by decoupling race attributes, as shown in the lower right of Fig. 1. In the de-biased SP module, firstly, two separate encoders E_i^m and E_c^m are formulated to embed features F^m into disentangled latent vectors for race-intrinsic attributes z_{ID} or race-conflicting attributes z_{sur} implied survival information [16]. Then, the linear classifiers C_i^m and C_c^m constructed to predict the race label y_{ID} with concatenated vector $z = [z_{\mathrm{ID}}; z_{\mathrm{sur}}]$. To disentangle survival features from the race identification, we use the generalized cross-entropy (GCE) loss [31] to train E_c^m and C_c^m to overfit to race label while training E_i^m and C_i^m with cross-entropy (CE) loss. The relative difficulty scores W as defined in Eq. 1 reweight and enhance the learning of the race-intrinsic attributes [20]. The objective function for disentanglement shown in Eq. 2, but the parameters of ID or survival branch are only updated by their respective losses:

$$W(z) = \frac{CE\left(C_c(z), y_{\mathrm{ID}}\right)}{CE\left(C_c(z), y_{\mathrm{ID}}\right) + CE\left(C_i(z), y_{\mathrm{ID}}\right)} \tag{1}$$

$$L_{\mathrm{dis}} = W(z) CE\left(C_i(z), y_{\mathrm{ID}}\right) + GCE\left(C_c(z), y_{\mathrm{ID}}\right) \tag{2}$$

To promote race-intrinsic learning in E_i^m and C_i^m, we apply diversify with latent vectors swapping. The randomly permuted $\tilde{z}_{\mathrm{sur}}$ in each mini-batch concatenate with z_{ID} to obtain $z_{\mathrm{sw}} = [z_{\mathrm{ID}}; \tilde{z}_{\mathrm{sur}}]$. The two neural networks are trained to predict y_{ID} or $\tilde{y}_{\mathrm{ID}}$ with CE loss or GCE loss. As the random combination are generated from different samples, the swapping decreases the correlation of these feature vectors, thereby enhancing the race-intrinsic attributes. The loss functions of swapping augmentation added to train two neural networks is defined as:

$$L_{\mathrm{sw}} = W(z) CE\left(C_i(z_{\mathrm{sw}}), y_{\mathrm{ID}}\right) + GCE\left(C_c(z_{\mathrm{sw}}), \tilde{y}_{\mathrm{ID}}\right) \tag{3}$$

The survival prediction head C_{sur}^m predicts the risk on the survival feature z_{sur}. CoxPH loss function [13], which optimizes the Cox partial likelihood, is used to maximize concordance differentiable and update model weights of the survival branch. Thus, CoxPH loss and overall loss function are formulated as:

$$L_{\mathrm{CoxPH}} := -\sum_{i: y_e^i = 1} \left(C_{\mathrm{sur}}(z_{\mathrm{sur}}^i) - \log \sum_{j: y_t^j > y_t^i} e^{C_{\mathrm{sur}}(z_{\mathrm{sur}}^j)} \right) \tag{4}$$

$$L_{\mathrm{overall}} = L_{\mathrm{dis}} + \lambda_{\mathrm{sw}} L_{\mathrm{sw}} + \lambda_{\mathrm{sur}} L_{\mathrm{CoxPH}} \tag{5}$$

where Y_t and Y_e are survival labels including the survival time and the event, respectively. The weights λ_{sw} and λ_{sur} are assigned as 0.5 and 0.8, respectively, to balance the feature disentanglement and survival prediction.

4 Experiment

We validate the proposed de-biased survival prediction frameworks on the collected multi-modality PE data. The data from 3 institutions are randomly split

Table 1. Performance comparison of the proposed de-biased SP framework and baseline using C-index values on multiple modal outcomes. The larger C-index value is better and the lower bias is fairer.

Method	Baseline				De-biased SP model			
Dataset	Overall	White	Color	Bias	Overall	White	Color	Bias
Imaging	**0.662**	0.736	0.422	0.314	0.646	0.656	0.622	0.035
Text	0.657	0.642	0.714	0.071	**0.719**	0.689	0.746	0.057
Variable	0.668	0.669	0.741	0.072	**0.698**	0.683	0.778	0.095
Multimodal	0.709	0.692	0.816	0.124	**0.759**	0.756	0.768	0.012

into training, validation, and testing sets, with a ratio of 7:1:2, the same ratio of survival events is maintained in each institution. We apply race-balanced resampling to the training and validation sets to eliminate training bias caused by minority groups.

The lung region of CPTA images is extracted with a slice thickness of 1.25 mm and scaled to $N \times 512 \times 512$ pixels [10]. Hounsfield units (HU) of all slices are clipped to the range of $[-1000, 900]$ and applied with zero-centered normalization. The PENet-based imaging backbone consists of a 77-layer 3D convolutional neural network and linear regression layers. It takes in a sliding window of 24 slices at a time, resulting in a window-level prediction that represents the probability of PE for the current slices [11]. The PENet is pre-trained on large-scale CTPA studies and shows excellent PE detection performance with an AUROC of 0.85 on our entire dataset. The 2048 dimensional features from the last convolution with the highest probability of PE, are designated as the imaging features.

The GatorTron [29] uses a transformer-based architecture to extract features from the clinical text, which was pre-trained on over 82 billion words of de-identified clinical text. We used the Huggingface library [28] to deploy the 345m-parameter cased model as the clinical report feature extractor. The outputs from each patient's medical history, clinical diagnosis, observations, and radiologist impression are separately generated and concatenated to form the 1024×4 features.

We build the encoders of the baseline SP modules and de-biased SP modules with multi-layer perceptron (MLP) neural networks and ReLu activation. The MLPs with 3 hidden layers are used to encode image and text features, and another MLPs with 2 layers encodes the features of PESI variables. A fully connected layer with sigmoid activation acts as a risk classifier $C_{\text{sur}}^{m}\left(z_{\text{sur}}^{m}\right)$ for survival prediction, where z_{sur}^{m} is the feature encoded from single modal data. For training the biased and de-biased SP modules, we collect data from one modality as a batch with synchronized batch normalization. The SP modules are optimized using the *AdamW* [17] optimizer with a momentum of 0.9, a weight decay of 0.0005, and a learning rate of 0.001. We apply early stopping when validation loss doesn't decrease for 600 epochs. Experiments are conducted on an Nvidia GV100 GPU.

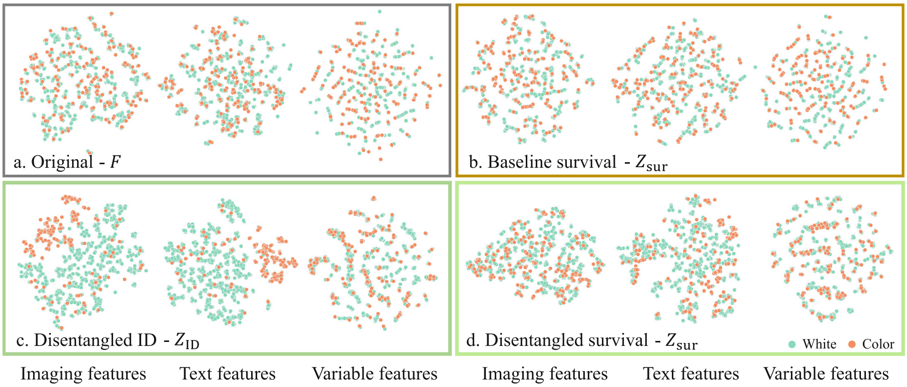

Fig. 2. tSNE visualizations of the features from multimodal data. Based on the comparison between the ID features and others, it is observed that the clusters containing race obtained from the same class are more compact.

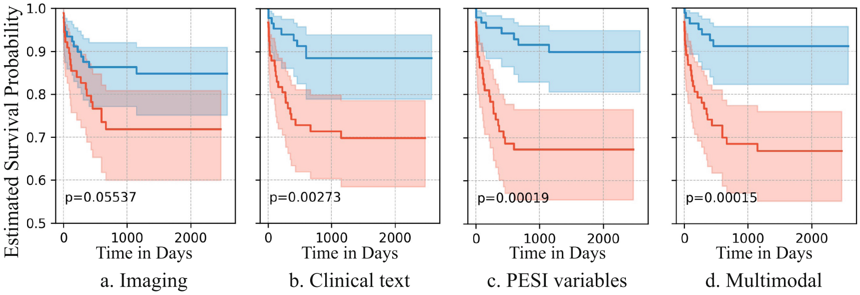

Fig. 3. Kaplan-Meier survival curves of our 3 de-biased SP modules and the multimodal CoxPH model. High-risk and low-risk groups are plotted as red and blue lines, respectively. The x-axis shows the time in days, and y-axis presents the estimated survival probability. Log-rank p value is shown on each figure. (Color figure online)

4.1 Results

Table 1 shows the quantitative comparisons of the baseline and de-biased frameworks with the C-indexes of the multimodal survival predictions. In general, our framework including de-biased SP modules shows significantly better predictions in testing set than the PESI-based outcome estimation with C-indexes of 0.669, 0.654, 0.697, 0.043 for the overall testset, White testset, Color testset and race bias. The de-biased results outperform the baseline in overall survival C-index and show a lower race bias, especially in imaging- and fusion-based predictions. The results indicate the effectiveness of the proposed de-biasing in mitigating race inequity. The results also prove the observations for the different biases present in different modalities, especially in the CTPA images containing more abundant race-related information. It also explains the limited effectiveness of de-biasing the clinical results, which contain less racial identification. The pre-

Table 2. Results of ablation studies. Every 2 columns (overall performance of Testing and Bias) represent a training setting.

Swapping	×		✓		×		✓	
Resampling	×		×		✓		✓	
Dataset	Testing	Bias	Testing	Bias	Testing	Bias	Testing	Bias
Imaging	0.666	0.062	0.641	0.014	0.649	0.050	0.622	0.035
Text	0.684	0.090	0.711	0.123	0.698	0.102	0.709	0.057
Variable	0.702	0.095	0.701	0.052	0.697	0.082	0.699	0.095
Multimodal	**0.716**	0.025	**0.737**	0.041	**0.741**	0.011	**0.743**	0.012

diction performance based on multiply modalities is significantly better than the PESI-based outcome estimation. The disentangled representations, transformed from latent space to a 2D plane via tSNE and color-coded by race [9], are shown in Fig. 2. We observe the disentanglement in the visualization of the ID features z_{ID}, while the survival features z_{sur} eliminate the race bias. The lack of apparent race bias observed in both the original features and those encoded in the baseline can be attributed to the subordinate role that ID features play in the multimodal information. The Kaplan-Meier (K-M) survival curve [14], as shown in Fig. 3, is used to compare the survival prediction between high-risk and low-risk patient groups. The p-values in the hypothesis test were found to be less than 0.001, which is considered statistically significant difference. In addition, the predictions of the de-biased framework show favorable performance, and our multimodal fusion demonstrates a more pronounced discriminative ability in the K-M survival analysis compared to the single-modal results.

We conducted ablation studies to examine the effect of the two key components, including swapping feature augmentation and race-balance resampling. As shown in Table 2, the different training settings show significant differences in survival prediction performance across modalities. The swapping augmentation provides a strong bias correction effect for image data with obvious bias. For clinical data, the resampling generally improves performance in most cases. Overall, multimodal fusion approaches are effective in all training settings, and the CoxPH model can actively learn the optimal combination of multimodal features to predict survival outcomes.

5 Discussions and Conclusions

In this work, we developed a de-biased survival prediction framework based on the race-disentangled representation. The proposed de-biased SP framework, based on the SOTA PE detection backbone and large-scale clinical language model, can predict the PE outcome with a higher survival correlation ahead of the clinical evaluation index. We detected indications of racial bias in our dataset and conducted an analysis of the multimodal diversity. Experimental results illustrate that our approach is effective for eliminating racial bias while resulting in an overall improved model performance. The proposed technique is clinically relevant as it can address the pervasive presence of racial bias in healthcare systems and offer a solution for minimizing or eliminating bias without pausing to evaluate their affection for the models and tools. Our study is significant as it highlights and evaluates the negative impact of racial bias on deep learning models. The proposed de-biased method has already shown the capacity to relieve them, which is vital when serving patients with an accurate analysis. The research in our paper demonstrates and proves that eliminating racial biases from data improves performance, and yields a more precise and robust survival prediction tool. In the future, these de-biased SP modules can be plugged into other models, offering a fairer method to predict survival outcomes.

References

1. Acosta, J.N., Falcone, G.J., Rajpurkar, P., Topol, E.J.: Multimodal biomedical AI. Nat. Med. **28**, 1773–1784 (2022). https://doi.org/10.1038/s41591-022-01981-2
2. Aujesky, D., et al.: Derivation and validation of a prognostic model for pulmonary embolism. Am. J. Respir. Crit. Care Med. **172**, 1041–1046 (2005). https://doi.org/10.1164/rccm.200506-862OC
3. Bahng, H., Chun, S., Yun, S., Choo, J., Oh, S.J.: Learning de-biased representations with biased representations. In: International Conference on Machine Learning, pp. 528–539. PMLR (2020)
4. Bengio, Y., Courville, A., Vincent, P.: Representation learning: a review and new perspectives. IEEE Trans. Pattern Anal. Mach. Intell. **35**(8), 1798–1828 (2013)
5. Booth, B.M., Hickman, L., Subburaj, S.K., Tay, L., Woo, S.E., D'Mello, S.K.: Bias and fairness in multimodal machine learning: a case study of automated video interviews. In: Proceedings of the 2021 International Conference on Multimodal Interaction, pp. 268–277 (2021). https://doi.org/10.1145/3462244.3479897
6. Creager, E., et al.: Flexibly fair representation learning by disentanglement. In: International Conference on Machine Learning, pp. 1436–1445. PMLR (2019)
7. Fox, J., Weisberg, S.: Cox proportional-hazards regression for survival data. An R and S-PLUS companion to applied regression 2002 (2002)
8. Harrell, F.E., Jr., Lee, K.L., Califf, R.M., Pryor, D.B., Rosati, R.A.: Regression modelling strategies for improved prognostic prediction. Stat. Med. **3**(2), 143–152 (1984)
9. Hinton, G., van der Maaten, L.: Visualizing data using t-SNE journal of machine learning research (2008)
10. Hofmanninger, J., Prayer, F., Pan, J., Röhrich, S., Prosch, H., Langs, G.: Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem. Eur. Radiol. Exp. **4**(1), 1–13 (2020). https://doi.org/10.1186/s41747-020-00173-2
11. Huang, S.C., et al.: PENet-a scalable deep-learning model for automated diagnosis of pulmonary embolism using volumetric CT imaging. NPJ Digit. Med. **3**(1), 61 (2020)
12. Kaplan, E.L., Meier, P.: Nonparametric estimation from incomplete observations. J. Am. Stat. Assoc. **53**(282), 457–481 (1958)

13. Katzman, J.L., Shaham, U., Cloninger, A., Bates, J., Jiang, T., Kluger, Y.: Deepsurv: personalized treatment recommender system using a cox proportional hazards deep neural network. BMC Med. Res. Methodol. **18**(1), 1–12 (2018)
14. Klok, F.A.: Patient outcomes after acute pulmonary embolism a pooled survival analysis of different adverse events. Am. J. Respir. Crit. Care Med. **181**, 501–506 (2009)
15. Lahat, D., Adali, T., Jutten, C.: Multimodal data fusion: an overview of methods, challenges, and prospects. Proc. IEEE **103**, 144–1477 (2015). https://doi.org/10.1109/JPROC.2015.2460697
16. Lee, J., Kim, E., Lee, J., Lee, J., Choo, J.: Learning debiased representation via disentangled feature augmentation. Adv. Neural. Inf. Process. Syst. **34**, 25123–25133 (2021)
17. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. In: International Conference on Learning Representations
18. Martin, K.A., McCabe, M.E., Feinglass, J., Khan, S.S.: Racial disparities exist across age groups in illinois for pulmonary embolism hospitalizations. Arterioscler. Thromb. Vasc. Biol. **40**, 2338–2340 (2020). https://doi.org/10.1161/ATVBAHA.120.314573
19. Perera, N, Perchik, J.D., Perchik, M.C., Tridandapani, S.: Trends in medical artificial intelligence publications from 2000–2020: where does radiology stand? Open J. Clin. Med. Images **2** (2022)
20. Nam, J., Cha, H., Ahn, S., Lee, J., Shin, J.: Learning from failure: de-biasing classifier from biased classifier. In: Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., Lin, H. (eds.) Advances in Neural Information Processing Systems, vol. 33, pp. 20673–20684. Curran Associates, Inc. (2020), https://proceedings.neurips.cc/paper/2020/file/eddc3427c5d77843c2253f1e799fe933-Paper.pdf
21. Obermeyer, Z., Powers, B., Vogeli, C., Mullainathan, S.: Dissecting racial bias in an algorithm used to manage the health of populations. Science **366**, 447–453 (2019). https://doi.org/10.1126/science.aax234
22. Parikh, R.B., Teeple, S., Navathe, A.S.: Addressing bias in artificial intelligence in health care. JAMA **322**, 2377–2378 (2019). https://doi.org/10.1001/jama.2019.18058
23. Pena, A., Serna, I., Morales, A., Fierrez, J.: Bias in multimodal AI: testbed for fair automatic recruitment. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 129–137 (2022). https://doi.org/10.1109/CVPRW50498.2020.00022
24. Phillips, A.R., et al.: Association between black race, clinical severity, and management of acute pulmonary embolism: a retrospective cohort study. J. Am. Heart Assoc. **10** (2021). https://doi.org/10.1161/JAHA.121.021818
25. Rajpurkar, P., Chen, E., Banerjee, O., Topol, E.J.: AI in health and medicine. Nat. Med. **28**, 31–38 (2022). https://doi.org/10.1038/s41591-021-01614-0
26. Rouzrokh, P., et al.: Mitigating bias in radiology machine learning: 1. data handling. Radiol. Artif. Intell. **4**, 1–10 (2022). https://doi.org/10.1148/ryai.210290
27. Song, J., Kalluri, P., Grover, A., Zhao, S., Ermon, S.: Learning controllable fair representations. In: The 22nd International Conference on Artificial Intelligence and Statistics, pp. 2164–2173. PMLR (2019)
28. Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., et al.: Huggingface's transformers: State-of-the-art natural language processing. arXiv preprint arXiv:1910.03771 (2019)
29. Yang, X., et al.: A large language model for electronic health records. NPJ Digit. Med. **5**(1), 194 (2022)

30. Zestcott, C.A., Blair, I.V., Stone, J.: Examining the presence, consequences, and reduction of implicit bias in health care: a narrative review. Group Process. Intergroup Relat. **19**, 528–542 (2016). https://doi.org/10.1177/1368430216642029
31. Zhang, Z., Sabuncu, M.: Generalized cross entropy loss for training deep neural networks with noisy labels. In: Advances in Neural Information Processing Systems, vol. 31 (2018)