Knowledge-Guided Multiview Deep Curriculum Learning for Elbow Fracture Classification

Jun Luo[1], Gene Kitamura[2], Dooman Arefan[2], Emine Doganay[2],
Ashok Panigrahy[2,3], and Shandong Wu[1,2,4](✉)

[1] Intelligent Systems Program, School of Computing and Information,
University of Pittsburgh, Pittsburgh, PA, USA
jul117@pitt.edu

[2] Department of Radiology, School of Medicine, University of Pittsburgh,
Pittsburgh, PA, USA

[3] University of Pittsburgh Medical Center Children's Hospital of Pittsburgh,
Pittsburgh, PA, USA

[4] Department of Biomedical Informatics and Department of Bioengineering,
University of Pittsburgh, Pittsburgh, PA, USA
wus3@upmc.edu

Abstract. Elbow fracture diagnosis often requires patients to take both frontal and lateral views of elbow X-ray radiographs. In this paper, we propose a multiview deep learning method for an elbow fracture subtype classification task. Our strategy leverages transfer learning by first training two single-view models, one for frontal view and the other for lateral view, and then transferring the weights to the corresponding layers in the proposed multiview network architecture. Meanwhile, quantitative medical knowledge was integrated into the training process through a curriculum learning framework, which enables the model to first learn from "easier" samples and then transition to "harder" samples to reach better performance. In addition, our multiview network can work both in a dual-view setting and with a single view as input. We evaluate our method through extensive experiments on a classification task of elbow fracture with a dataset of 1,964 images. Results show that our method outperforms two related methods on bone fracture study in multiple settings, and our technique is able to boost the performance of the compared methods. The code is available at https://github.com/ljaiverson/multiview-curriculum.

Keywords: Multiview learning · Deep learning · Curriculum learning · Elbow fracture · Clinical knowledge

1 Introduction

Human's cognitive ability relies deeply on integrating information from different views of the objects. This is particularly the case for elbow fracture diagnosis

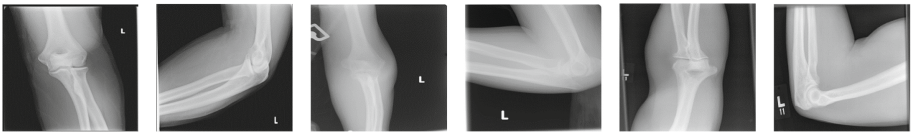
C. Lian et al. (Eds.): MLMI 2021, LNCS 12966, pp. 555–564, 2021.
https://doi.org/10.1007/978-3-030-87589-3_57

where patients are often required to take both the frontal view (i.e. Anterior-Posterior view) and lateral view of elbow X-ray radiographs for diagnosis. This is because some fracture subtypes might be more visible from a certain perspective: the frontal view projects the distal humerus, the proximal ulna and the radius [7, 21, 22], while the lateral view shows the coronoid process and the olecranon process [9, 18, 22]. In practice, it is also common that some patients only have a single view radiograph acquired, or have a missing view for various reasons.

In recent years, the advance of deep learning has been facilitating the automation of bone fracture diagnosis [3, 10, 12] through multiple views of X-ray images, which shows faster speed and decent accuracy compared to human experts [13, 14, 17]. However, few methods leverage multiview information, which provide more visual information from different perspectives for elbow fracture diagnosis.

In this work, we propose a novel multiview deep learning network architecture for elbow fracture subtype classification that takes frontal view and lateral view elbow radiographs as input. While the proposed model is a dual-view (frontal and lateral) architecture, it is flexible as it does not strictly require a dual-view input during inference. Furthermore, our training strategy for the multiview model takes advantage of transfer learning by first training two single-view models, one for frontal view and the other for lateral view, and then transferring the trained weights to the corresponding layers in the proposed multiview network architecture. In addition, we investigate the utilities of integrating medical knowledge of different views into the training via a curriculum learning scheme, which enables the model to first learn from "easier" samples and then transition to "harder" samples to reach better performance.

To evaluate our method, we conduct experiments on a classification task of three classes of elbow fractures that shown in Fig. 1. We compare our method to multiple settings including the single-view models, different combinations of the transfer learning strategy and the knowledge-guided curriculum learning. Our method is also compared to a previous method [11]. Results show that our proposed method outperforms the compared methods, and our method functions seamlessly on a multiview and a single-view settings.



(a) (b) (c) (d) (e) (f)

Fig. 1. Example images from the three categories from our dataset for classification task: (a) and (b) show the frontal and lateral non-fracture category respectively; (c) and (d) show the frontal and lateral ulnar fracture category respectively; (e) and (f) show the frontal and lateral radial fracture category respectively.

2 Related Work

Multiview learning [23] takes advantage of data with multiple views of the same objects. Co-training [2,16,20] style algorithms were a group of traditional multiview learning algorithms originally focusing on semi-supervised learning, where multiple views of data were iteratively added to the labeled set and learned by the classifier. Another group of multiview learning algorithms explore Multiple Kernel Learning (MKL), which was originally proposed to restrict the search space of kernels [4,6]. Recent work on multiview learning based modeling shows promising effects for medical fields such as bone fracture and breast cancer detection [8,13,17].

Curriculum learning is also an area of active research. It was first introduced by Bengio et al. in [1] to enable the machine learning to mimic human learning by training a machine learning model first with “easier” samples and then transition to “harder” samples. Some existing work focus on integrating domain knowledge into the training process through curriculum learning. For example, [11,15] integrate domain knowledge by using the classification difficulty level of different classes.

3 Methods

3.1 Multiview Model Architecture

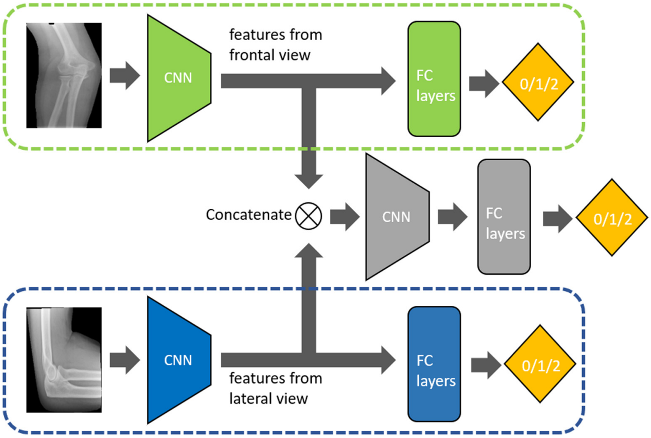


Fig. 2. The proposed multiview model architecture. The green and blue dotted line box represent the frontal and lateral view modules, respectively. Yellow diamonds are the predicted labels, 0, 1, 2 corresponding to non-fracture, ulnar fracture, radial fracture respectively (Color figure online)

To incorporate information from both frontal and lateral view for the elbow X-ray images while maintaining the flexibility of being able to output predictions with one view as input, we propose a novel multiview model architecture shown in Fig. 2. In this architecture, during training, pairs of frontal and lateral view images are fed into their corresponding modules for feature extraction by the convolutional neural networks (CNNs). After the feature extraction, the model splits into three branches as shown in Fig. 2. The top and bottom branches take the corresponding single-view features to the fully connected (FC) layers for classification, while the middle branch takes the concatenated features from both views as input to further extract features and then conducts classification.

Consider a data sample triplet $\mathcal{D}_i = \{x_i^{(F)}, x_i^{(L)}, y_i\}$ where $\mathcal{D}_i$ represents the i-th data sample, $x_i^{(F)}$, and $x_i^{(L)}$ are its images from the frontal and lateral view, and $y_i \in \{0, 1, 2\}$ is its ground truth label with 0, 1, 2 corresponding to non-fracture, ulnar fracture, radial fracture respectively. We denote the three predicted labels from the three branches of our multiview model as $\mathcal{F}(x_i^{(F)})$, $\mathcal{L}(x_i^{(L)})$, and $\mathcal{M}(x_i^{(F)}, x_i^{(L)})$, where $\mathcal{F}$, $\mathcal{L}$, $\mathcal{M}$ represent the *frontal view module*, the *lateral view module*, and the "*merge module*" that contains the two CNN blocks from the frontal and lateral module, the CNN as well as the FC layers in the middle branch.

During training, we minimize the objective function over the i-th data sample computed by Eq. (1) where θ, $\theta_{\mathcal{F}}$, $\theta_{\mathcal{L}}$, and $\theta_{\mathcal{M}}$ represent the parameters in the entire model, the frontal view module, the lateral view module, and the merge module. As shown in Eq. (1) (with C being the number of classes), for each module, the loss is computed with cross entropy loss over the corresponding predicted label and ground truth y_i in a one-hot representation.

$$\begin{aligned} J_\theta(x_i^{(F)}, x_i^{(L)}, y_i) &= J_{\theta_{\mathcal{F}}}(x_i^{(F)}, y_i) + J_{\theta_{\mathcal{L}}}(x_i^{(L)}, y_i) + J_{\theta_{\mathcal{M}}}(x_i^{(F)}, x_i^{(L)}, y_i) \\ &= -\sum_{c=1}^{C} \left(y_{i,c} \left(\log(\mathcal{F}(x_i^{(F)})_c) + \log(\mathcal{L}(x_i^{(L)})_c) + \log(\mathcal{M}(x_i^{(F)}, x_i^{(L)})_c) \right) \right) \end{aligned} \tag{1}$$

During test phase, if a frontal view image and a lateral view image are both presented, the default final predicted label is the one predicted from the merge module, i.e. $\mathcal{M}(x_i^{(F)}, x_i^{(L)})$. Alternatively, if there is only one view, the model will still output a predicted label from the module of the corresponding view credited to the designed architecture of our model.

3.2 Transfer Learning from Pretrained Single-View Models

In most medical applications with deep learning, researchers use the ImageNet [5] pretrained model as a way of transfer learning. However, a great number of deep learning models do not have publicly available pretrained weights, especially for self-designed models. Here, we investigate a homogeneous way of transfer learning as shown in Fig. 3: we first train two single-view models (using the same training set as the one for the multiview model) that have identical structure as the frontal view and lateral view module in the multiview architecture. Then, we transfer the trained weights of the CNNs and FC layers from the single view models to the counterparts of the multiview model (refer to the links in Fig. 3). For the middle branch (the gray CNN and LC layers blocks in Fig. 2) in the merge module, we randomly initialize their weights. We make all weights trainable in the multiview model.

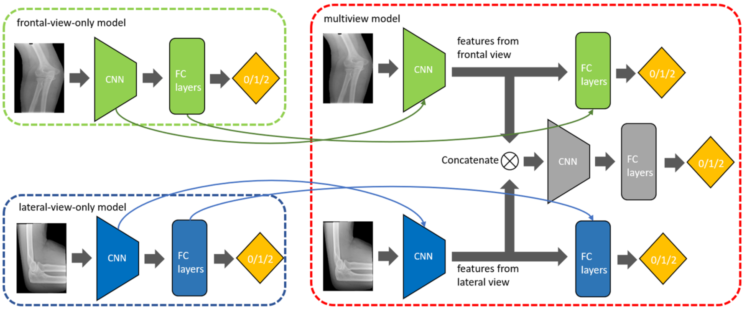


Fig. 3. Transfer learning from pretrained single-view models.

3.3 Knowledge-Guided Curriculum Learning

For the model training, we propose a knowledge-guided curriculum learning to enhance learning effects. The idea of curriculum learning is to enable the training process to follow an "easy-to-hard" order, where the easier samples will be fed into the model for training earlier than the harder samples. To do so, we implemented a multiview-based curriculum learning by adapting the method from [15]. We quantify and integrate medical knowledge by scoring the classification difficulty levels of each category of elbow fracture with board-certified radiologist's expertise. Table 1 shows the quantitative scores reflecting the classification difficulty based on experience of expert radiologists. Note that we use the "Both views" scores to train the multiview model, and use "Frontal/Lateral view only" for homogeneous transfer learning.

Table 1. Quantitative classification difficulty levels for each category of elbow fracture (1-hardest; 100-easiest), which enables the integration of medical knowledge into curriculum learning.

	Non-fracture (normal)	Ulnar fracture	Radial fracture
Frontal view only	30	30	30
Lateral view only	35	60	45
Both views	45	65	55

These scores are used to initialize the sampling probability for each training data point according to Eq. (2) with $e = 1$, where $p_i^{(1)}$ is the initial sampling probability for data point $\mathcal{D}_i$, s_i is its score, s_k is the score of the data point $\mathcal{D}_k$, and N is the number of data points in the dataset. Using the sampling probabilities, at the beginning of every epoch, we permute the training set by sampling all the data points without replacement.

$$p_i^{(e)} = \begin{cases} \frac{s_i}{\sum_{k=1}^{N} s_k} & e = 1, \\ p_i^{(e-1)} \cdot \sqrt[E']{\frac{1/N}{p_i^{(0)}}} & 2 \leq e \leq E', \\ 1/N & E' < e \leq E \end{cases} \quad (2)$$

This enables the easier samples to have a higher chance of being presented before the harder samples. This chance will be exponentially reduced by updating the sampling probabilities for each data point according to Eq. (2). In this equation, e is the current epoch, E' is the last epoch that we update the sampling probabilities. For the rest of the training ($E' < e \leq E$) the sampling probabilities will be fixed to $1/N$.

4 Experiments and Results

4.1 Experiment Settings

Dataset and Implementation Details. This study includes a private dataset of 982 subjects of elbow fractures in an Institutional Review Board-approved retrospective study. The subjects are categorized into three classes: 500 non-fracture (normal) cases, 98 ulnar fracture cases, and 384 radial fracture cases. Each subject includes one frontal and one lateral elbow X-ray image, which makes it a total of 1,964 elbow X-ray images. To increase the robustness of our results, we conduct 8-fold cross validation. For each split of the entire dataset, one fold was used as the hold-out test set. Within the remaining seven folds, we randomly select one fold as the validation set for hyperparameter tuning. The remaining folds are used as the training set. All separations of the dataset are in a stratified manner, which maintains the ratio over different classes. The reported results are averages over the 8 disjoint held-out test sets.

VGG16 [19] is used as the backbone for the two single-view models, and the frontal and lateral modules in the multiview model. We customize the middle branch two $3 \times 3 \times 512$ convolutional layers with max pooling layers, followed by VGG16's classifier for the FC layers. The hyperparameters are selected based on the best validation AUCs. We use the following hyperparameters for the proposed model: batch size 64, learning rate 10^{-4} for the Adam optimizer, and after 16 epochs every sample is treated as having an equal difficulty score. All models were trained on an NVIDIA Tesla V100 GPU. The code is available at https://github.com/ljaiverson/multiview-curriculum.

Metrics. The metrics for the 3-class classification task include accuracy and area under receiver operating characteristic curve (AUC). We also compute a

Table 2. Model performance with both views. The bold numbers correspond to the highest value for each metric (TL: proposed transfer learning from single view models; CL: proposed knowledge-guided curriculum learning).

Model	Accuracy	AUC	Balanced accuracy	Binary task accuracy	Binary task AUC
Single-view-frontal	0.683	0.807	0.570	0.732	0.813
Single-view-lateral	0.856	0.954	0.807	0.895	0.959
Multiview	0.854	0.958	0.796	0.884	0.964
Multiview + TL	**0.891**	0.966	0.847	**0.916**	0.973
Multiview + [11]	0.818	0.939	0.746	0.864	0.952
Multiview + [11] + TL	0.870	0.961	0.811	0.898	0.973
Multiview + CL	0.889	0.970	0.847	0.908	**0.978**
Multiview + CL + TL	0.889	**0.974**	**0.864**	0.910	0.976

balanced accuracy by averaging the ratios between the number of true positives and the total number of samples with respect to each class, which reduces the effect induced by data imbalance. In addition, we evaluate the models' overall ability to distinguish fracture against non-fracture images. This is done by binarizing the ground truth and predicted labels by assigning 0 to them if they originally are 0, and assigning 1 otherwise. We compute the binary task accuracy and the AUC as two additional measures.

4.2 Results

As shown in Table 2, we compare our proposed multiview model with curriculum learning method (CL) and transfer learning (TL) with the following six types of models: 1) two single-view models (frontal/lateral view only), referred as Single-view-frontal/lateral; 2) multiview model with regular training, referred as Multiview; 3) multiview model with only transfer learning strategy, referred as Multiview + TL; 4) multiview model with a previous curriculum training method [11], referred as Multiview + [11]; 5) multiview model with [11] and our proposed transfer learning strategy, referred as Multiview + [11] + TL; and 6) multiview model with only our curriculum learning method, referred as Multiview + CL. We use the output from the middle branch, as the predicted label.

Attributed to the multiple branches of our model and the customized loss function, our model has the flexibility of generating the prediction with a single view as input. In Table 3, we show the results of the performance from the frontal view module and lateral view module separately. Different from [11], our curriculum updates the difficulty score of every sample after every epoch, which benefits the multiview model. Table 2 shows that with both views presented in the test phase, our method achieves the highest AUC and balanced accuracy with a margin of up to 0.118 compared to the state-of-the-art performance. In settings with missing views, however, our strategy does not always perform the

Table 3. Model performance with a single view as input

Model	Input view	Accuracy	AUC	Balanced accuracy	Binary task accuracy	Binary task AUC
Single-view	Frontal	0.720	0.828	0.593	0.761	0.844
Single-view + CL [15]	Frontal	0.683	0.807	0.570	0.732	0.813
Multiview	Frontal	0.658	0.749	0.514	0.702	0.766
Multiview + TL	Frontal	0.738	0.827	0.617	0.774	0.829
Multiview + [11]	Frontal	0.566	0.675	0.396	0.575	0.648
Multiview + [11] + TL	Frontal	0.737	0.815	0.605	0.773	0.831
Multiview + CL	Frontal	0.723	0.814	0.602	0.761	0.823
Multiview + CL + TL	Frontal	**0.756**	**0.829**	**0.636**	**0.786**	**0.846**
Single-view	Lateral	0.856	0.954	0.807	**0.895**	0.959
Single-view + CL [15]	Lateral	0.840	0.946	0.809	0.872	0.948
Multiview	Lateral	0.844	0.951	0.800	0.870	0.956
Multiview + TL	Lateral	0.848	0.954	0.804	0.876	0.961
Multiview + [11]	Lateral	0.837	0.945	0.779	0.870	0.949
Multiview + [11] + TL	Lateral	**0.857**	**0.960**	**0.819**	0.885	**0.969**
Multiview + CL	Lateral	0.838	0.956	0.807	0.867	0.956
Multiview + CL + TL	Lateral	0.840	0.955	0.794	0.874	0.960

best. Table 3 shows that with frontal view as the only input view, our method outperforms all the compared methods per each metric, but with the lateral view as the only input view, our method achieves slightly lower performance than the best results.

5 Conclusion

In this work, we propose a novel multiview deep learning method for elbow fracture subtype classification from frontal and lateral view X-ray images. We leverage transfer learning by first pretraining two single-view models. Meanwhile, medical knowledge was quantified and incorporated in the training process through curriculum learning. The results show that our multiview model outperforms the compared methods, and we achieved improved results over the previously published curriculum training strategies. As future work, we plan to further integrate other domain knowledge with respect to different views and explore curriculum learning in the output space.

Acknowledgements. This work was supported in part by National Institutes of Health grants (1R01CA193603 and 1R01CA218405), the Stimulation Pilot Research Program of the Pittsburgh Center for AI Innovation in Medical Imaging and the associated Pitt Momentum Funds of a Scaling grant from the University of Pittsburgh (2020), and an Amazon Machine Learning Research Award.

References

1. Bengio, Y., Louradour, J., Collobert, R., Weston, J.: Curriculum learning. In: Proceedings of the 26th Annual International Conference on Machine Learning, pp. 41–48 (2009)
2. Blum, A., Mitchell, T.: Combining labeled and unlabeled data with co-training. In: Proceedings of the Eleventh Annual Conference on Computational Learning Theory, pp. 92–100 (1998)
3. Cheng, C.T., et al.: A scalable physician-level deep learning algorithm detects universal trauma on pelvic radiographs. Nat. Commun. **12**(1), 1–10 (2021)
4. Cortes, C., Mohri, M., Rostamizadeh, A.: Learning non-linear combinations of kernels. In: Bengio, Y., Schuurmans, D., Lafferty, J., Williams, C., Culotta, A. (eds.) Advances in Neural Information Processing Systems, vol. 22. Curran Associates, Inc. (2009)
5. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255. IEEE (2009)
6. Duffy, N., Helmbold, D.P.: Leveraging for regression. In: COLT, pp. 208–219 (2000)
7. El-Khoury, G.Y., Daniel, W.W., Kathol, M.H.: Acute and chronic avulsive injuries. Radiol. Clin. North Am. **35**(3), 747–766 (1997)
8. Geras, K.J., et al.: High-resolution breast cancer screening with multi-view deep convolutional neural networks. arXiv preprint arXiv:1703.07047 (2017)
9. Goldfarb, C.A., Patterson, J.M.M., Sutter, M., Krauss, M., Steffen, J.A., Galatz, L.: Elbow radiographic anatomy: measurement techniques and normative data. J. Shoulder Elbow Surg. **21**(9), 1236–1246 (2012)
10. Guan, B., Zhang, G., Yao, J., Wang, X., Wang, M.: Arm fracture detection in x-rays based on improved deep convolutional neural network. Comput. Electr. Eng. **81**, 106530 (2020)
11. Jiménez-Sánchez, A., et al.: Medical-based deep curriculum learning for improved fracture classification. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11769, pp. 694–702. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32226-7_77
12. Kalmet, P.H., et al.: Deep learning in fracture detection: a narrative review. Acta Orthopaedica **91**(2), 215–220 (2020)
13. Kitamura, G., Chung, C.Y., Moore, B.E.: Ankle fracture detection utilizing a convolutional neural network ensemble implemented with a small sample, de novo training, and multiview incorporation. J. Digit. Imaging **32**(4), 672–677 (2019)
14. Krogue, J.D., et al.: Automatic hip fracture identification and functional subclassification with deep learning. Radiol. Artif. Intell. **2**(2), e190023 (2020)
15. Luo, J., Kitamura, G., Doganay, E., Arefan, D., Wu, S.: Medical knowledge-guided deep curriculum learning for elbow fracture diagnosis from x-ray images. In: Medical Imaging 2021: Computer-Aided Diagnosis, vol. 11597, p. 1159712. International Society for Optics and Photonics (2021)
16. Nigam, K., Ghani, R.: Analyzing the effectiveness and applicability of co-training. In: Proceedings of the Ninth International Conference Information Knowledge Management, pp. 86–93 (2000)
17. Rayan, J.C., Reddy, N., Kan, J.H., Zhang, W., Annapragada, A.: Binomial classification of pediatric elbow fractures using a deep learning multiview approach emulating radiologist decision making. Radiol. Artif. Intell. **1**(1), e180015 (2019)

18. Sandman, E., Canet, F., Petit, Y., Laflamme, G.Y., Athwal, G.S., Rouleau, D.M.: Effect of elbow position on radiographic measurements of radio-capitellar alignment. World J. Orthop. **7**(2), 117 (2016)
19. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
20. Sindhwani, V., Niyogi, P., Belkin, M.: A co-regularization approach to semi-supervised learning with multiple views. In: Proceedings of ICML Workshop on Learning with Multiple Views, vol. 2005, pp. 74–79. Citeseer (2005)
21. Stevens, M.A., El-Khoury, G.Y., Kathol, M.H., Brandser, E.A., Chow, S.: Imaging features of avulsion injuries. Radiographics **19**(3), 655–672 (1999)
22. Whitley, A.S., Jefferson, G., Holmes, K., Sloane, C., Anderson, C., Hoadley, G.: Clark's Positioning in Radiography 13E. CRC Press Boca Raton (2015)
23. Xu, C., Tao, D., Xu, C.: A survey on multi-view learning. arXiv preprint arXiv:1304.5634 (2013)