Medical-based Deep Curriculum Learning for Improved Fracture Classification

Amelia Jiménez-Sánchez, D. Mateus, S. Kirchhoff, C. Kirchhoff, P. Biberthaler, N. Navab, MA. González Ballester, G. Piella

Medical-based Deep Curriculum Learning for Improved Fracture Classification [1]

[1] Jiménez-Sánchez, A., Mateus, D., Kirchhoff, S., Kirchhoff, C., Biberthaler, P., Navab, N., González Ballester, MÁ. & Piella, G. (2019, October). Medical-based Deep Curriculum Learning for Improved Fracture Classification. In International Conference on Medical Image Computing and Computer-Assisted Intervention – MICCAI (pp. 694-702). Springer, Cham.

Index

[1] Jiménez-Sánchez, A., Mateus, D., Kirchhoff, S., Kirchhoff, C., Biberthaler, P., Navab, N., González Ballester, MÁ. & Piella, G. (2019, October). Medical-based Deep Curriculum Learning for Improved Fracture Classification. In International Conference on Medical Image Computing and Computer-Assisted Intervention – MICCAI (pp. 694-702). Springer, Cham.

Introduction

Learning

- Educational system - learning relies in a curriculum.
- *Starting small* concept: systematic and gradual learning [2].

[2] Elman, J.L. (1993). Learning and development in neural networks: the importance of starting small. Cognition, 48, 71-99.

Learning

- Educational system - learning relies in a curriculum.
- *Starting small* concept: systematic and gradual learning [2].
- **Curriculum learning [3]:**
    - 2-step schedule.
   - Improved accuracy and faster convergence.
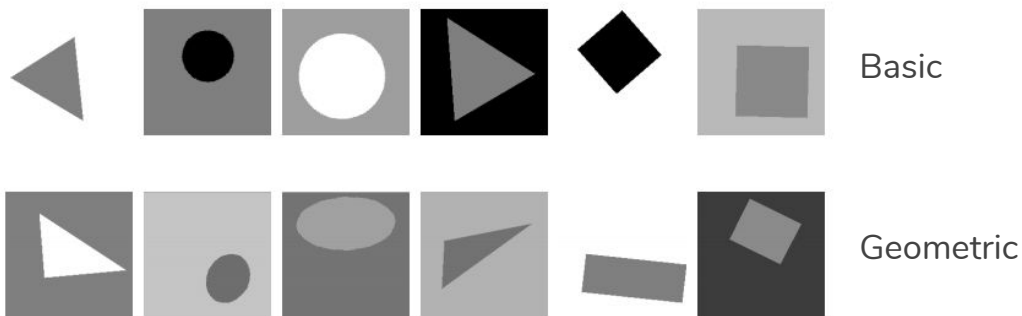
Basic

Geometric

[2] Elman, J.L. (1993). Learning and development in neural networks: the importance of starting small. Cognition, 48, 71-99.

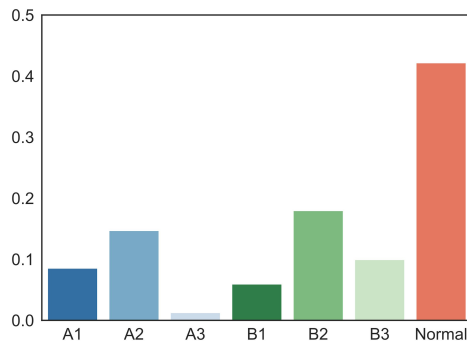[3] Bengio, Y., Louradour, J., Collobert, R., & Weston, J. (2009). Curriculum learning. ICML.

Learning

- Educational system - learning relies in a curriculum.
- *Starting small* concept: systematic and gradual learning [2].
- Curriculum learning [3].

What is **easy**? And what is **difficult**?

[2] Elman, J.L. (1993). Learning and development in neural networks: the importance of starting small. Cognition, 48, 71-99.
[3] Bengio, Y., Louradour, J., Collobert, R., & Weston, J. (2009). Curriculum learning. ICML.
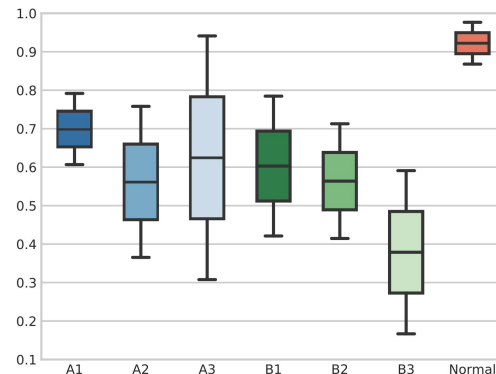
Medical Knowledge
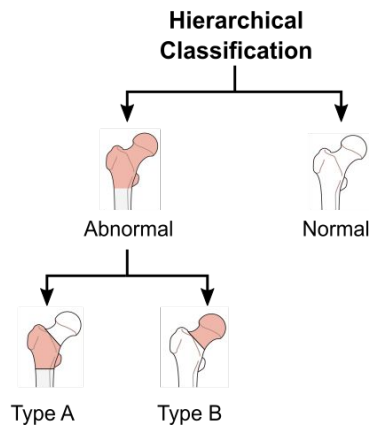
Intra-/Inter- rater variability

Decision Trees

Frequency of diseases

Can medical knowledge determine a meaningful curriculum?

Clinical Motivation

- Treatment options for femur fracture [4]:
  - Gold standard: **surgery**
  - Depends on **fracture type**

[4] Tan, L. T. J., Wong, S. J., & Kwek, E. B. K. (2017). Inpatient cost for hip fracture patients managed with an orthogeriatric care model in Singapore. Singapore medical journal, 58(3), 139.
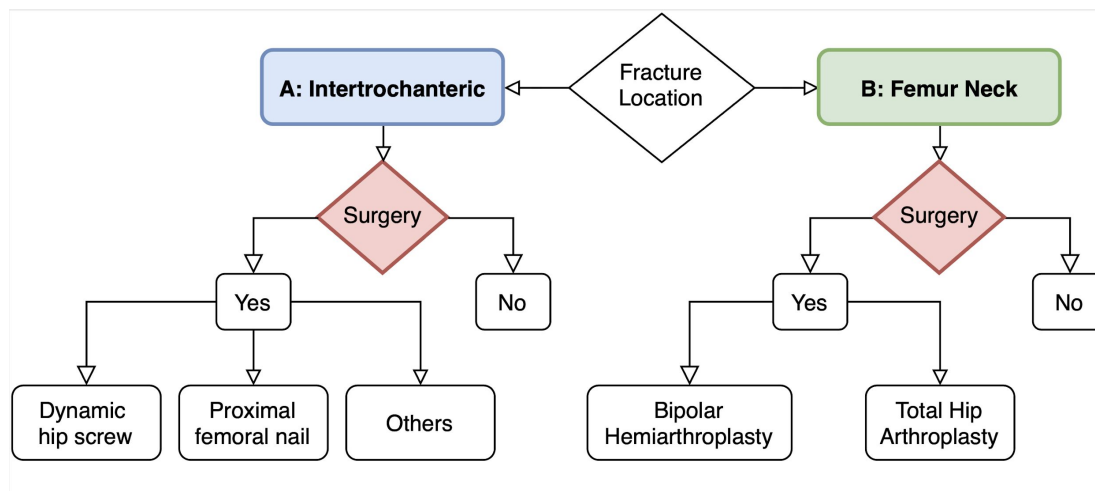
Clinical Motivation

- Classification according to the
  Arbeitsgemeinschaft Osteosynthese (**AO**) Standard [5]

- Required **years of experience** for reliable classification:
  (5-10 years)

- **Variability:** Inter-expert agreement on subclasses:
  68% kappa correlation (71% experts, 66% residents) [6]

[5] Kellam, J., Meinberg, E., Agel, J., Karam, M., Roberts, C. (2018). Introduction: Fracture and Dislocation Classification
Comprehensive Classification of Fractures and Dislocations Committee. Journal of Orthopaedic Trauma. 32 Suppl 1. S1-S10.
[6] Embden, Daphne van et al. "The comparison of two classifications for trochanteric femur fractures: the AO/ASIF classification and the Jensen
classification." Injury 41 4 (2010): 377-81 .

AO Classification: Proximal Femur Fractures

[5] Kellam, J., Meinberg, E., Agel, J., Karam, M., Roberts, C. (2018). Introduction: Fracture and Dislocation Classification Compendium-2018: International Comprehensive Classification of Fractures and Dislocations Committee. Journal of Orthopaedic Trauma. 32 Suppl 1. S1-S10.

AO Classification: Proximal Femur Fractures

[5] Kellam, J., Meinberg, E., Agel, J., Karam, M., Roberts, C. (2018). Introduction: Fracture and Dislocation Classification Compendium-2018: International Comprehensive Classification of Fractures and Dislocations Committee. Journal of Orthopaedic Trauma. 32 Suppl 1. S1-S10.

Method

Problem Statement: 7-classes

Abnormal

Normal

Type A

Type B

3 classes

Normal

A1   **A2**   **A3**

B1   **B2**   **B3**

Normal

7 classes

Conventional Training

input: *X* (X-ray images), *Y* (classification labels),
B (mini-batch size), *E* (expected training epochs)

Conventional Training

input: *X* (X-ray images), *Y* (classification labels),
        B (mini-batch size), *E* (expected training epochs)

for *each epoch* **e do**:
        Random permutation of training set $f^{(e)}$: $\{X, Y\} \rightarrow \{X, Y\}^r$;

end

1	4
2	1
3	5
4	9
5	12
6	2
7	11
8	8
9	10
10	6
11	3
12	7

Conventional Training

input: X (X-ray images), Y (classification labels),
B (mini-batch size), E (expected training epochs)

for *each epoch* **e do**:
Random permutation of training set $f^{(e)}: \{X, Y\} \to \{X, Y\}^r$;
for *each training round* **do**:
Get the **next** mini-batch from $\{X, Y\}^r$: $\{x_b, y_b\}_{b=1}^{B}$;
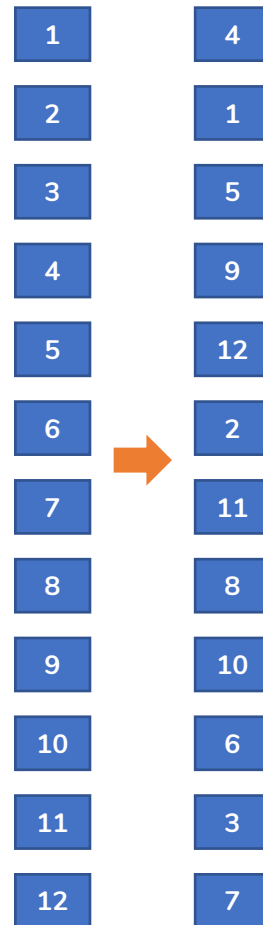
end
end

Conventional Training

input: *X* (X-ray images), *Y* (classification labels),
 B (mini-batch size), *E* (expected training epochs)

for *each epoch* **e do***:*
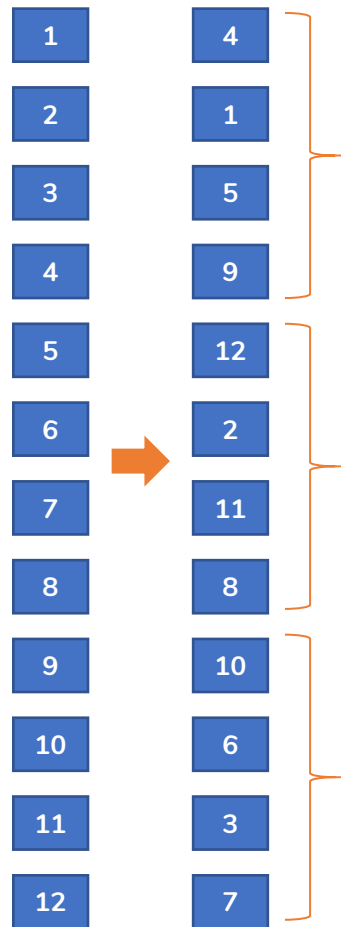 Random permutation of training set $f^{(e)}$: $\{X, Y\} \rightarrow \{X, Y\}^r$;
 for *each training round* **do***:*
  Get the **next** mini-batch from $\{X, Y\}^r$: $\{x_b, y_b\}_{b=1}^{B}$;
  Calculate cross-entropy loss $L(y_{b,} \hat{y}_b)$;

 end
end

Conventional Training

input: *X* (X-ray images), *Y* (classification labels),
    B (mini-batch size), *E* (expected training epochs)

for *each epoch* **e do**:
    Random permutation of training set $f^{(e)}$: $\{X, Y\} \rightarrow \{X, Y\}^r$
    for *each training round* **do**:
        Get the **next** mini-batch from $\{X, Y\}^r$: $\{x_b, y_b\}_{b=1}^{B}$;
        Calculate cross-entropy loss $L(y_b, \hat{y}_b)$;
        Compute gradients and update model weights ;
    end
end

Conventional Training

input: *X* (X-ray images), *Y* (classification labels),
    B (mini-batch size), *E* (expected training epochs)

for *each epoch* **e do***:
    Random permutation of training set $f^{(e)}$: $\{X, Y\} \rightarrow \{X, Y\}^r$
    for *each training round* **do***:
        Get the **next** mini-batch from $\{X, Y\}^r$: $\{x_b, y_b\}_{b=1}^{B}$;
        Calculate cross-entropy loss $L(y_b, \hat{y}_b)$;
        Compute gradients and update model weights ;
    end
end

What changes with a **curriculum**?

Training with Curriculum Learning (CL)

input: X (X-ray images), Y (classification labels), $c \in C$ (curriculum)
$\quad\quad$ B (mini-batch size), E (expected training epochs)

for *each epoch* **e do***:*
$\quad\quad$ Random permutation of training set $f^{(e)}:\ \{X, Y\} \rightarrow \{X, Y\}^r$
$\quad\quad$ **for** *each training round* **do***:*
$\quad\quad\quad\quad$ Get the **next** mini-batch from $\{X, Y\}^r: \{x_b, y_b\}_{b=1}^{B}$;
$\quad\quad\quad\quad$ Calculate cross-entropy loss $L(y_{b,} \hat{y}_b)$;
$\quad\quad\quad\quad$ Compute gradients and update model weights ;
$\quad\quad$ **end**
end

Training with Curriculum Learning (CL)

input: *X* (X-ray images), *Y* (classification labels), *c* ∈ *C* (curriculum)
     B (mini-batch size), *E* (expected training epochs)

for *each epoch* **e do**:
     ~~Random permutation of training set $f^{(e)}: \{X, Y\} \rightarrow \{X, Y\}^f$~~
     for *each training round* **do**:
          Get the **next** mini-batch from $\{X, Y\}^r: \{x_b, y_b\}_{b=1}^{B}$;
          Calculate cross-entropy loss $L(y_b, \hat{y}_b)$;
          Compute gradients and update model weights ;
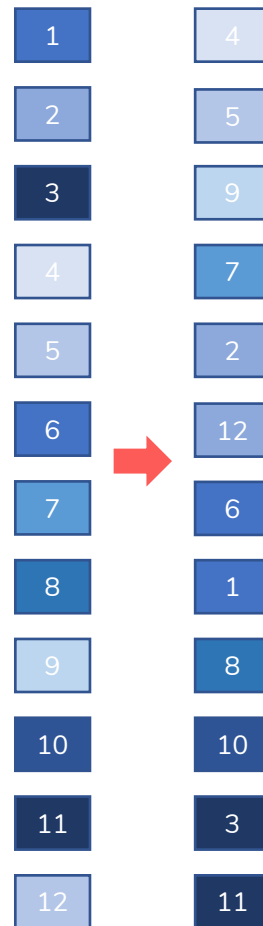   **end**
end

Training with Curriculum Learning (CL)

for *each epoch* **e do**

~~Random permutation of training set $f^{(e)}$: {X, Y} → {X, Y}^{f}~~

if *first epoch* **then**

Define initial probabilities: $p_i^{(0)} = w_m^{\ c}$

else

Update probabilities with Eqs. (1-2)

for *each training round* **do**

Get the **next** mini-batch from $\{X, Y\}^r$: $\{x_b, y_b\}_{b=1}^{B}$;

Calculate cross-entropy loss $L(y_b, \hat{y}_b)$;
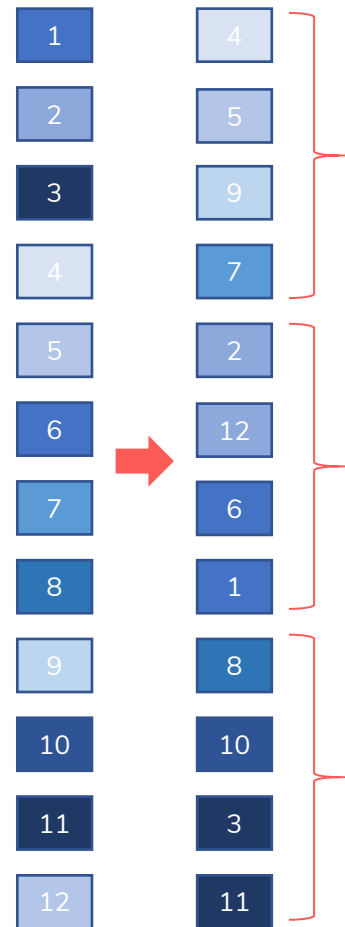
Compute gradients and update model weights ;

end

end

1
2
3
4
5
6
7
8
9
10
11
12

Training with **Curriculum Learning (CL)**

for *each epoch* **e do**

~~Random permutation of training set $f^{(e)}$: {X, Y} → {X, Y}^{f}~~

if *first epoch* **then**

Define initial probabilities: $p_i^{(0)} = w_m^c$

else

Update probabilities with Eqs. (1-2)

Get reordering function $f^{(e)}$ *by sampling {X, Y} according to* $p^{(e)}$;

Permute training set $f^{(e)}$: {X, Y} → {X, Y}^{c} ;

for *each training round* **do**

Get the **next** mini-batch from {X, Y}^{r}: {x_b, y_b}$_{b=1}^B$;

Calculate cross-entropy loss $L(y_b, \hat{y}_b)$;

Compute gradients and update model weights ;

end

end

1	4
2	5
3	9
4	7
5	2
6	12
7	6
8	1
9	8
10	10
11	3
12	11

Training with Curriculum Learning (CL)

for *each epoch* **e do**

~~Random permutation of training set $f^{(e)}$: $(X, Y) \rightarrow (X, Y)^f$~~

**if** *first epoch* **then**

Define initial probabilities: $p_i^{(0)} = w_m^c$

else

Update probabilities with Eqs. (1-2)

Get reordering function $f^{(e)}$ *by sampling* $\{X, Y\}$ according to $p^{(e)}$;

Permute training set $f^{(e)}$: $\{X, Y\} \rightarrow \{X, Y\}^c$;

for *each training round* **do**

Get the **next** mini-batch from $\{X, Y\}^c$: $\{x_b, y_b\}_{b=1}^B$;

Calculate cross-entropy loss $L(y_{b,} \hat{y}_b)$;

Compute gradients and update model weights ;

end

end

Training with **Curriculum Learning (CL)**

curriculum

for *each epoch* **e do**

~~Random permutation of training set~~ $f^{(e)}$~~: (X, Y) → (X, Y)~~^{f}

if *first epoch* **then**

Define initial probabilities: $p_i^{(0)} = w_m^c$

else

Update probabilities with Eqs. (1-2)

$$q_i^{(e)} = p_i^{(e-1)} \cdot \exp(-cn_i^2/10) \quad \forall e > 0, \tag{1}$$

$$p_i^{(e)} = \frac{q_i^{(e)}}{\sum_{i=1}^{N} q_i^{(e)}}, \tag{2}$$

uniform

end
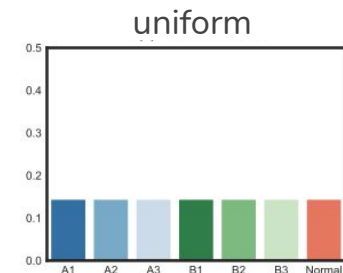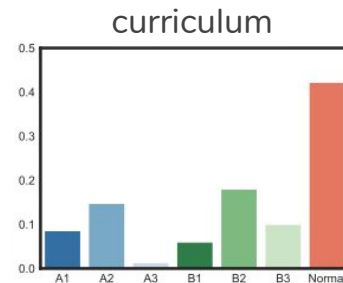
Defining the curriculum

How to assign the **initial curriculum probabilities**?

$$p^{(0)}(y_i = m) = w_m^c$$

where $m \in [1, 2, \dots, M]$ serves as index of the classes.

Defining the curriculum

How to assign the **initial curriculum probabilities**?

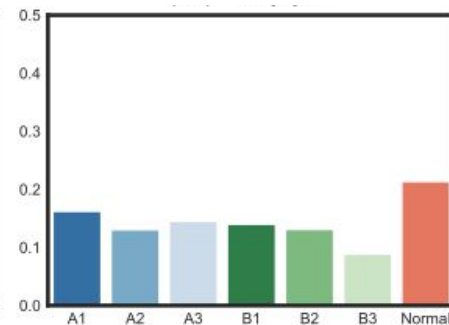$$p^{(0)}(y_i = m) = w_m^c$$

We propose four strategies:

(i) uniform (ii) frequency (iii) AO (iv) kappa
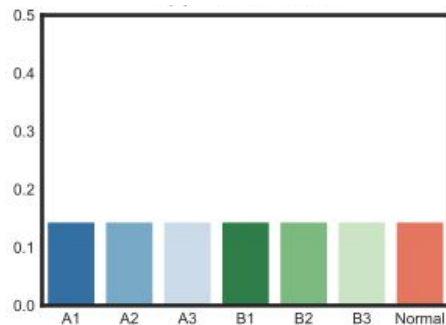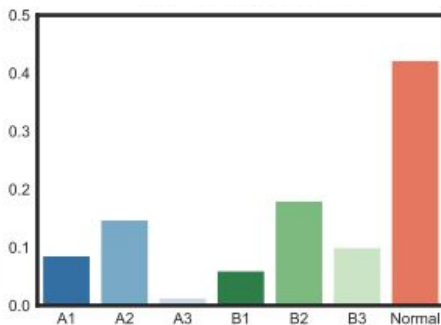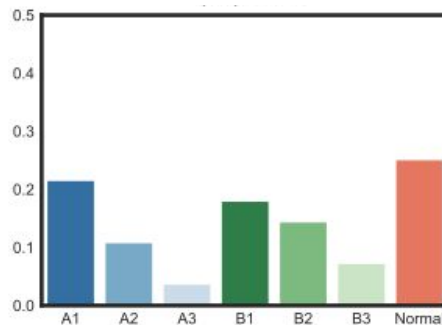
Defining the curriculum

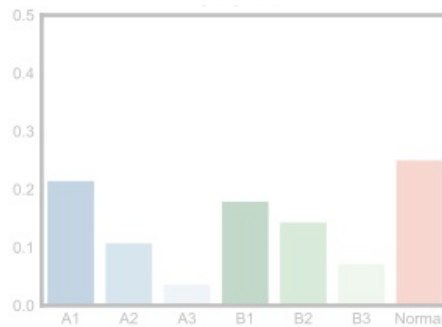(i)   c : **uniform**: balanced over classes.

$$p^{(0)}(y_i = m) = w_m^c \qquad\qquad w_m = 1/M$$
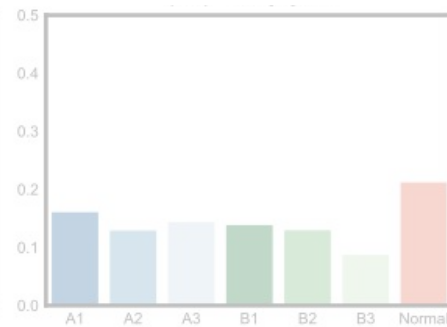

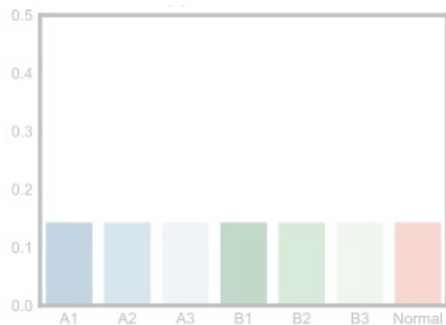(i) uniform   (ii) frequency   (iii) AO   (iv) kappa

29
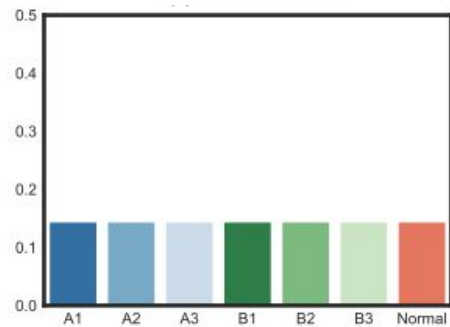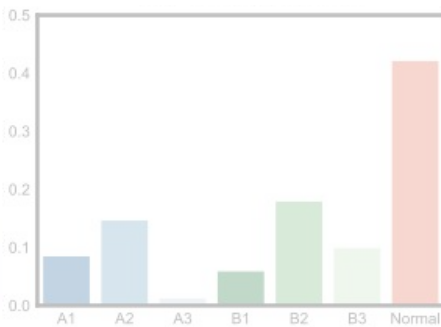
Defining the curriculum

(ii)   c : **frequency**: probabilities are proportional to their original frequency in the dataset.
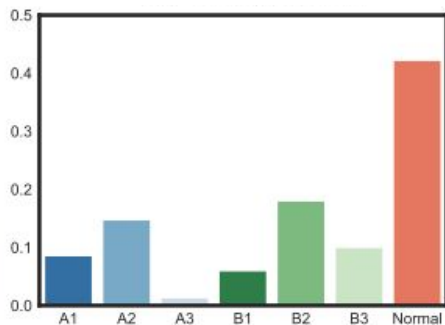
$$p^{(0)}(y_i = m) = w_m^c$$

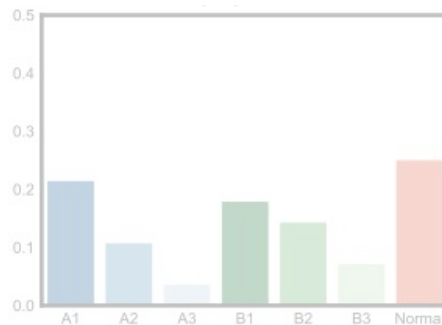$$w_m = \frac{1}{N} \sum_{i=1}^{N} \delta_{y_i, m}$$
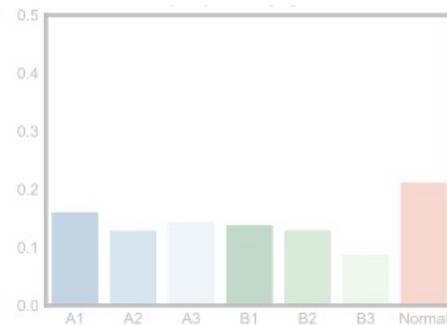

(i) uniform   (ii) frequency   (iii) AO   (iv) kappa
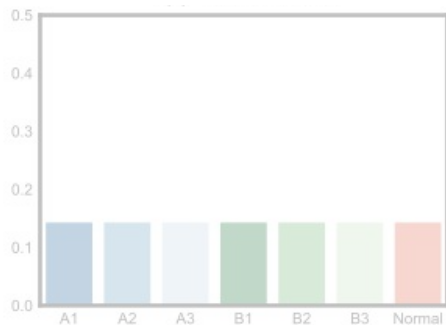
Defining the curriculum

(iii)   c : **AO**: an experienced radiologist ranked the difficulty of the AO classes.
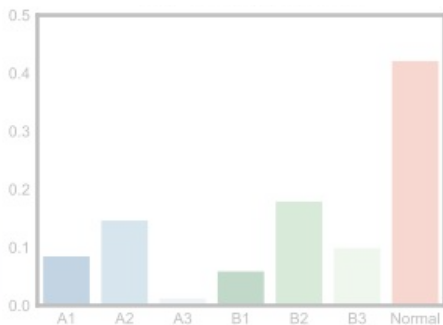
RANK

$$p^{(0)}(y_i = m) = w_m^c$$

$$w_m = \frac{k}{\sum_{m=1}^{M} m}$$

(i) uniform   (ii) frequency   (iii) AO   (iv) kappa

Defining the curriculum

(iv) c : **kappa**: according to intra-rater Cohen's kappa agreement.

$$p^{(0)}(y_i = m) = w_m^c$$

$w_m =$ kappa statistics
Ratio between observed and chance agreement.
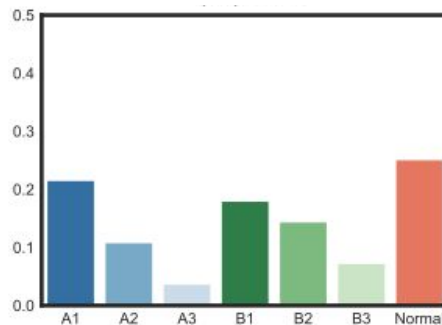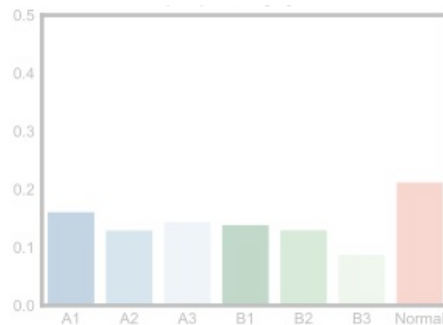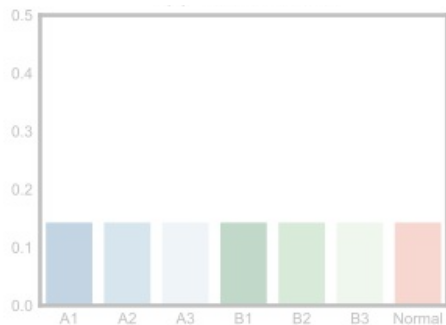
(i) uniform   (ii) frequency   (iii) AO   (iv) kappa
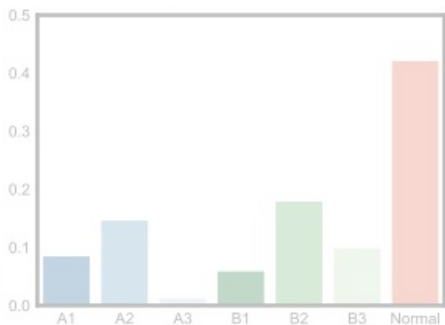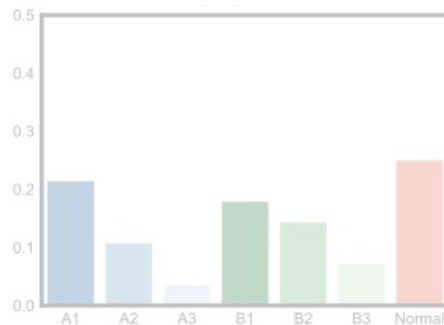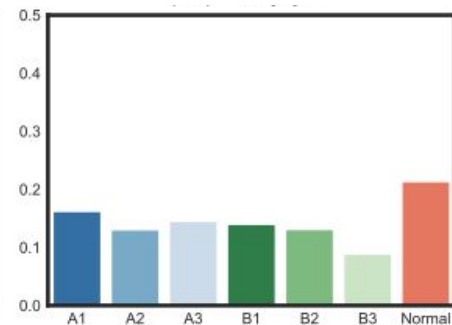
Defining the curriculum

(iv) c : **kappa**: according to intra-rater Cohen's kappa agreement.

(iv) kappa

[6] Jiménez-Sánchez, A., Kazi, A., Albarqouni, S., Kirchhoff, C., Biberthaler, P., Navab, N., Mateus, D. and Kirchhoff, S., (2019) Towards an Interactive and Interpretable CAD System to Support Proximal Femur Fracture Classification. arXiv preprint arXiv:1902.01338v1
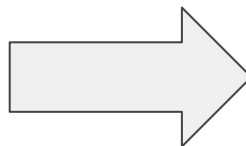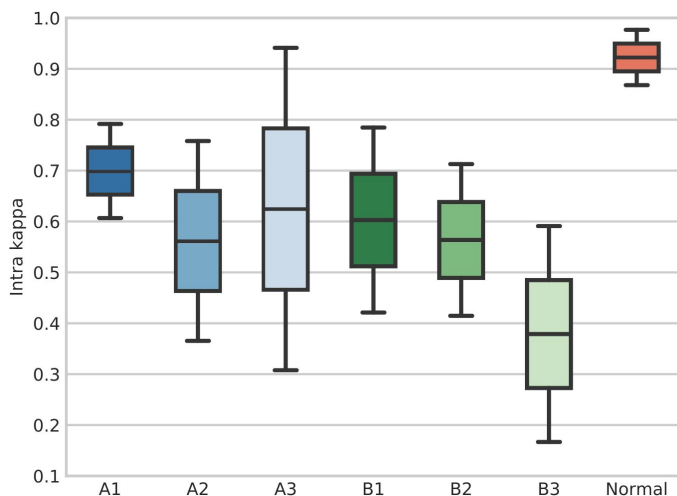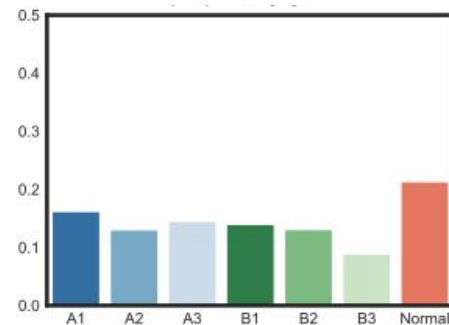
Results

Experimental Validation

- **Clinical dataset**

    ~1300 X-ray images,
    from 780 patients.

    Offline data augmentation:
    translation, scale and rotation.

    Split into three parts with the ratio
            70% - training
            10% - validation
            20% - test

Experimental Validation

- **Clinical dataset**

  ~1300 X-ray images,
  from 780 patients.

  Offline data augmentation:
  translation, scale and rotation.

  Split into three parts with the ratio
  - 70% - training
  - 10% - validation
  - 20% - test

- **Evaluation**

  curriculum: difficulty is gradually
  increased (*C: easy → hard*).

  anti-curriculum: difficulty is gradually
  decreased (*C: hard → easy*).

Experimental Validation

- **Clinical dataset**

    ~1300 X-ray images,
    from 780 patients.

    Offline data augmentation:
    translation, scale and rotation.

    Split into three parts with the ratio
        70% - training
        10% - validation
        20% - test

- **Evaluation**

    f1-score

    10 runs each model

Classification Performance
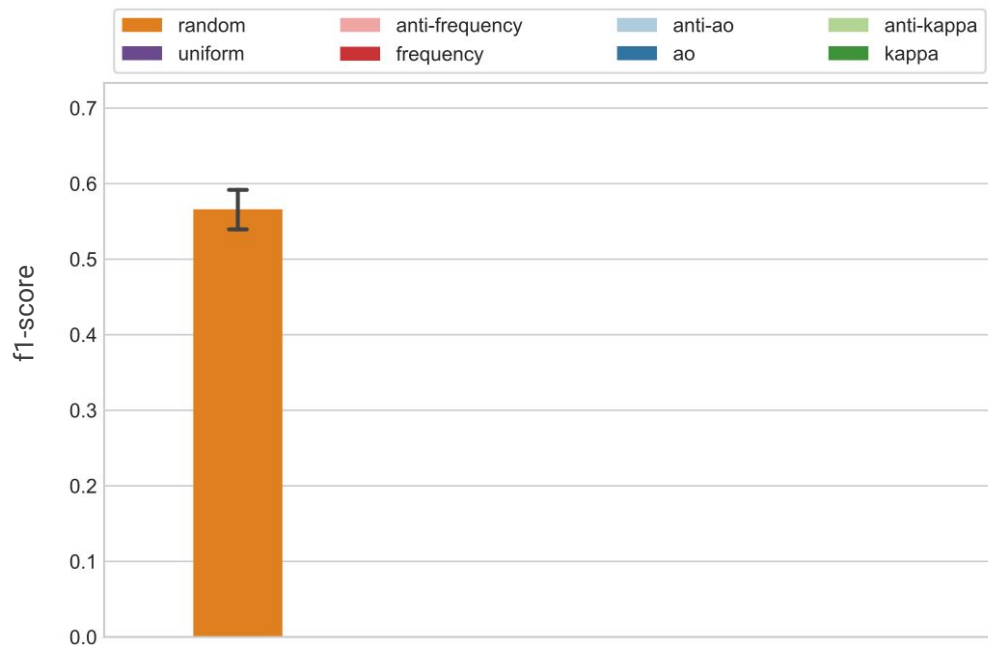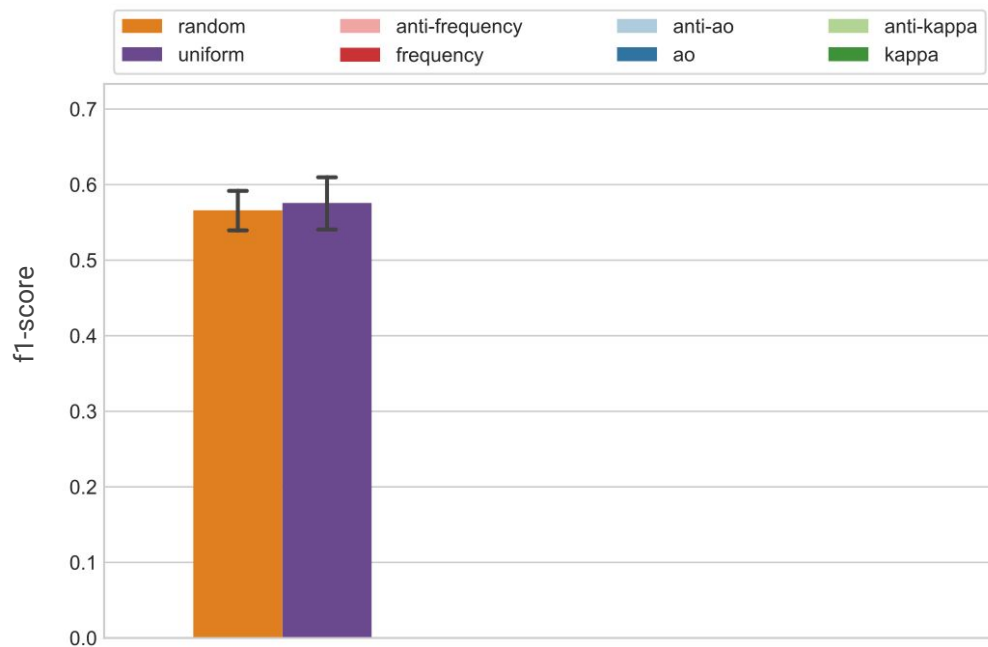
Classification Performance

Classification Performance
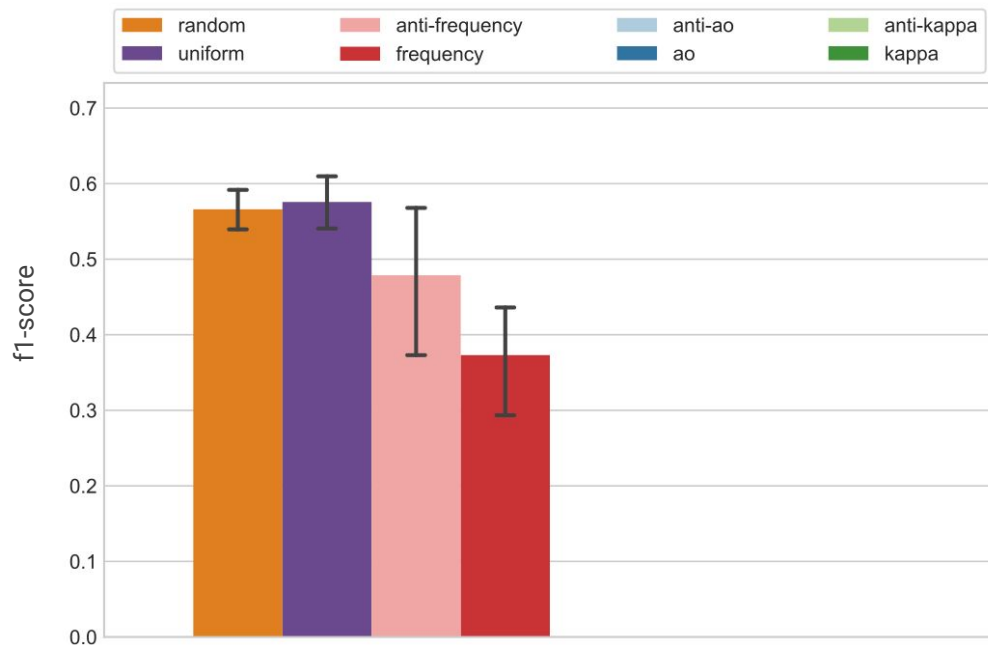
Classification Performance
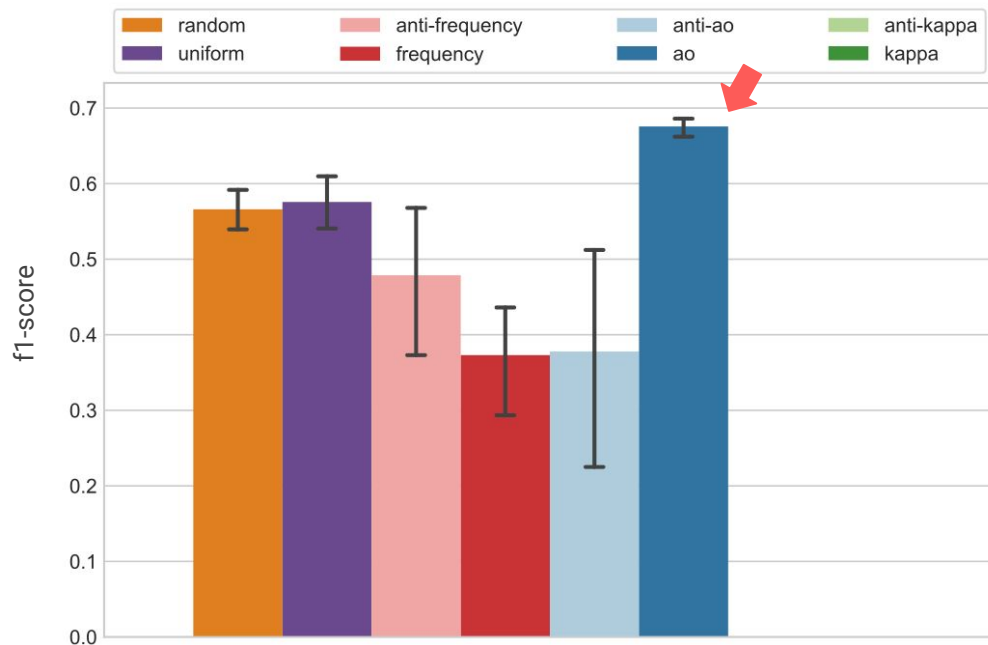
Classification Performance
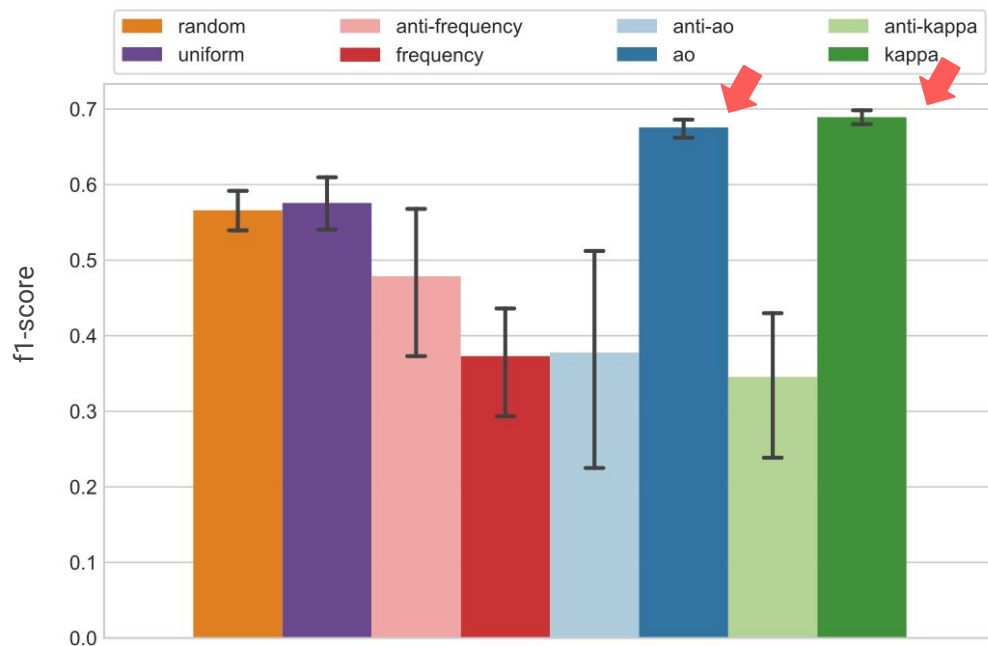
Classification Performance
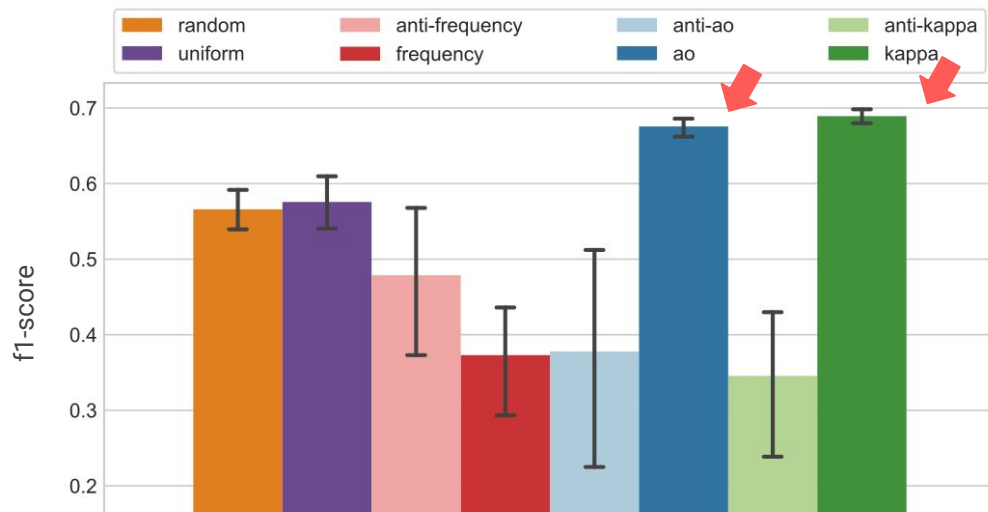
1. Random ~ uniform-curriculum.
2. Significance difference between curriculum & anti-curriculum.
3. Frequency-curriculum suggests that the imbalance scenario is easier.
4. **AO- and kappa-curriculum improves** median f1-score by **15%.**
5. **Reach a performance comparable to experienced trauma surgeons (66-71% agreement).**

Classification Performance

F_1-score	7 classes			3 classes		
	Mean	Median	SD	Mean	Median	SD
Random	0.5662	0.5731	0.0423	0.8063	0.8171	0.0337
Uniform	0.5757	0.5923	0.0590	0.8011	0.7971	0.0399
AO	0.6757	**0.6783**	0.0197	0.8651	**0.8657**	0.0172
Kappa	0.6893	**0.6900**	0.0150	0.8623	**0.8657**	0.0146

6. **State-of-the art results for 3 classes** (Normal, type A and type B), aggregating output probabilities

Classification Performance

F_1-score	7 classes			3 classes		
	Mean	Median	SD	Mean	Median	SD
Random	0.5662	0.5731	0.0423	0.8063	0.8171	0.0337
Uniform	0.5757	0.5923	0.0590	0.8011	0.7971	0.0399
AO	0.6757	**0.6783**	0.0197	0.8651	**0.8657**	0.0172
Kappa	0.6893	**0.6900**	0.0150	0.8623	**0.8657**	0.0146
AO - 60%	0.6325	0.6188	0.0302	0.8457	0.8486	0.0191
Kappa - 60%	0.6352	0.6500	0.0398	0.8446	0.8457	0.0222

6. **State-of-the art results for 3 classes** (Normal, type A and type B), aggregating outputs
7. **AO- and kappa-curriculum training on only 60% training data,** performs better than random and uniform-curriculum using 100% training data

Conclusions & Future Work

Medical-based Deep Curriculum Learning

- **Medical knowledge**:
    - can be integrated as a **curriculum** strategy,
    - **improved up to 15%** the classification score,
    - helps against **small datasets.**

- Extend to **other applications:**
    - Where medical decision trees are available, e.g. malignancy grading.
    - Whenever intra-, inter-expert agreement is available.

- **Future work:** combination with **uncertainty** of the model.

Thank you for your attention!

Amelia Jiménez-Sánchez
amelia.jimenez@upf.edu

website

https://www.epicentrofestival.com/

Jiménez-Sánchez, A., Mateus, D., Kirchhoff, S., Kirchhoff, C., Biberthaler, P., Navab, N., González Ballester, MÁ. & Piella, G. (2019, October). Medical-based Deep Curriculum Learning for Improved Fracture Classification. In International Conference on Medical Image Computing and Computer-Assisted Intervention – MICCAI (pp. 694-702). Springer, Cham.