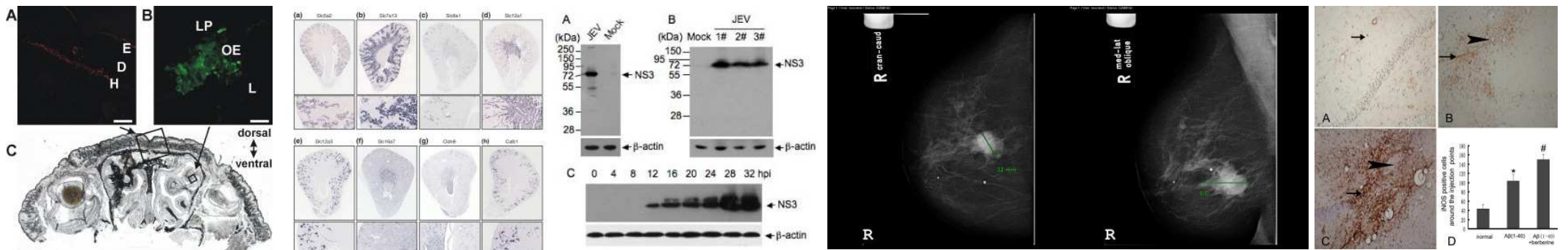


Overview of the medical task of ImageCLEF 2016



Alba G. Seco de Herrera
Stefano Bromuri
Roger Schaar
Henning Müller



Tasks in ImageCLEF 2016

- Automatic image annotation
- **Medical image classification**
 - Sub-tasks in compound figure identification, separation, multi-label classification, modality classification and caption prediction
- Handwritten scanned document retrieval
 - New task in 2016

Motivation

- The medical literature contains large amounts of images
 - The majority are compound figures that require to have subfigures treated separately
- Image content and caption text can be used for the analysis
- Making the **compound figure content accessible** is the main goal

Changes in medical task in 2016



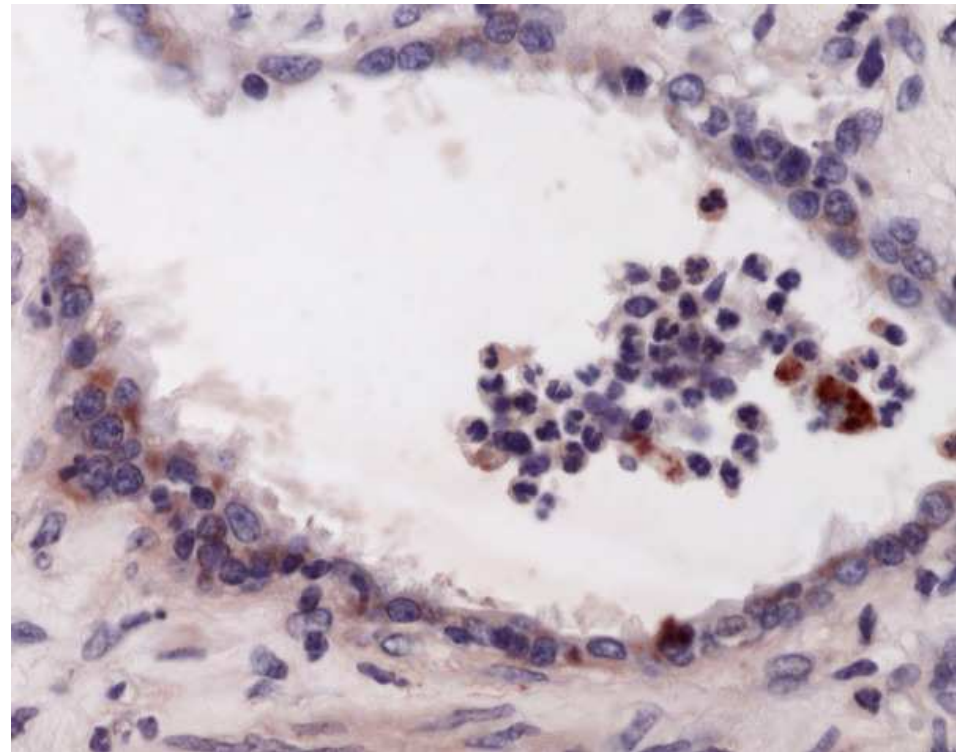
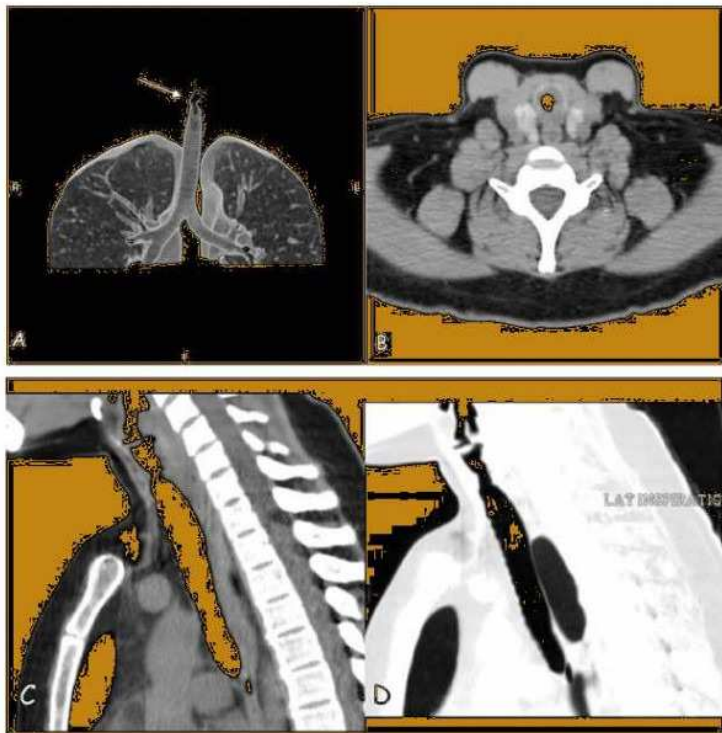
- More figures for all subtasks
- New subtask
 - **Caption prediction**: given a medical image, produce a caption associated with the image that is then compared to the real captions

ImageCLEFmed 2016 subtasks

- Compound figure detection
- Compound figure separation
- Multi-label classification
- Subfigure classification
- Caption prediction

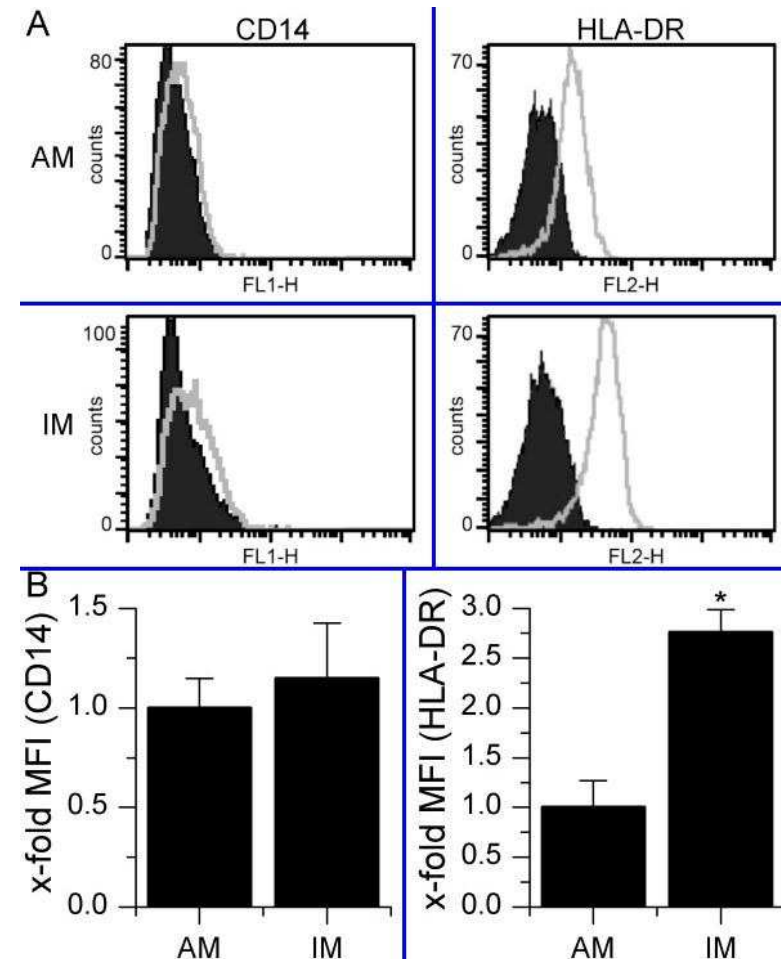
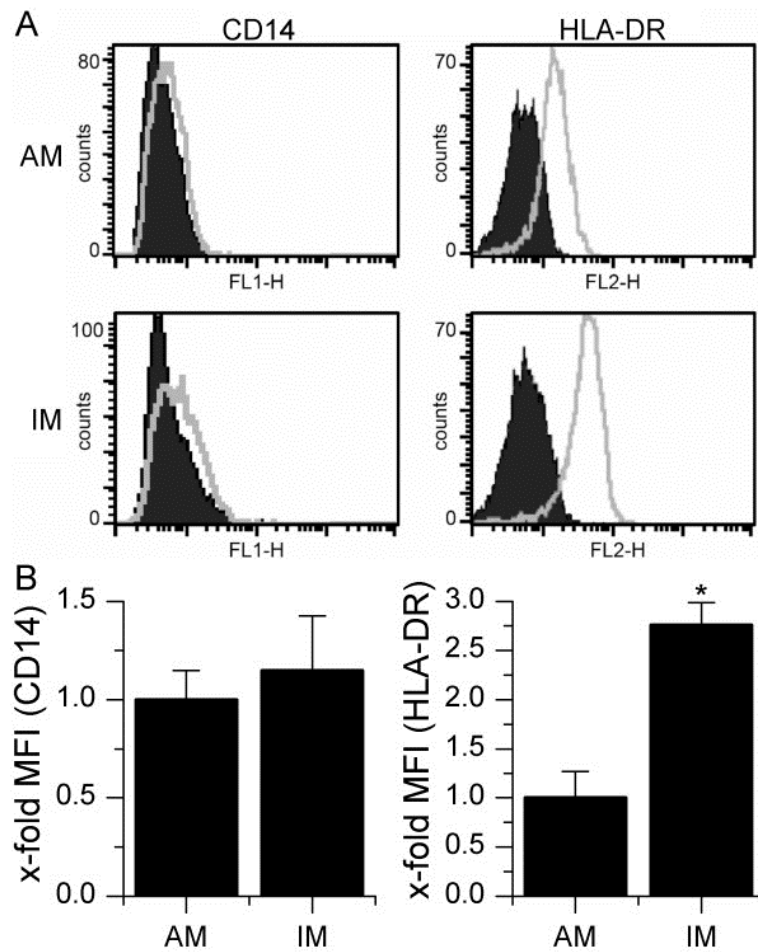
Compound figure detection

- To **identify** if a figure is a compound figure or has a single content



Compound figure separation

- To **separate** the compound figures into subfigures



Multi-label classification

- To **label** compound figures with each of the modalities of the subfigures
- Without previously separating them

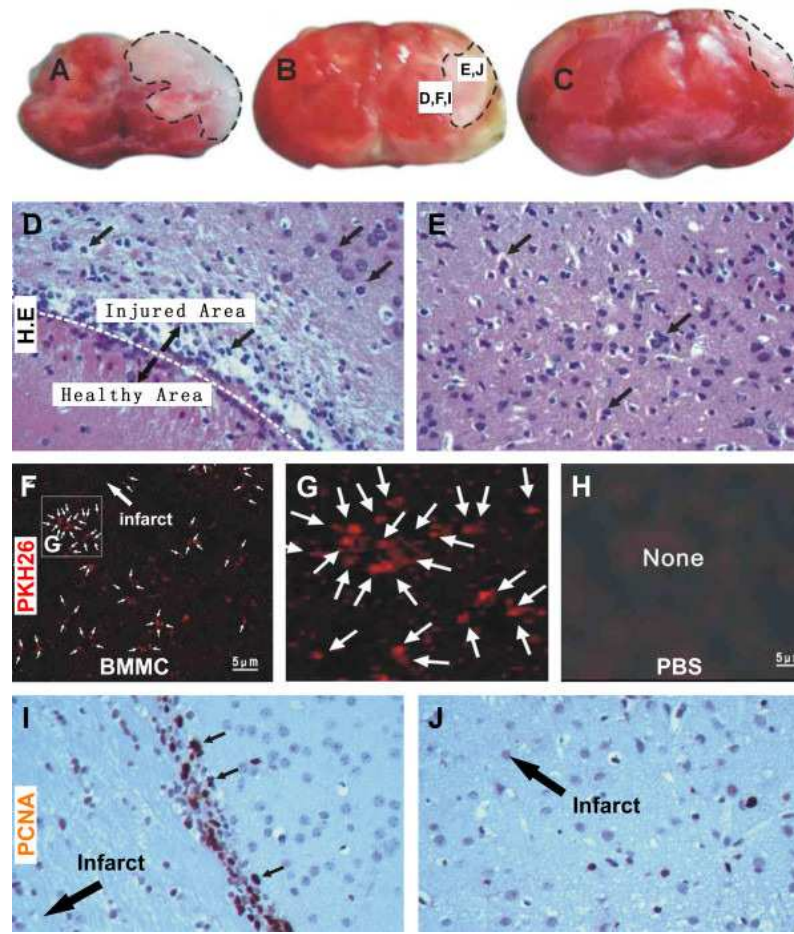
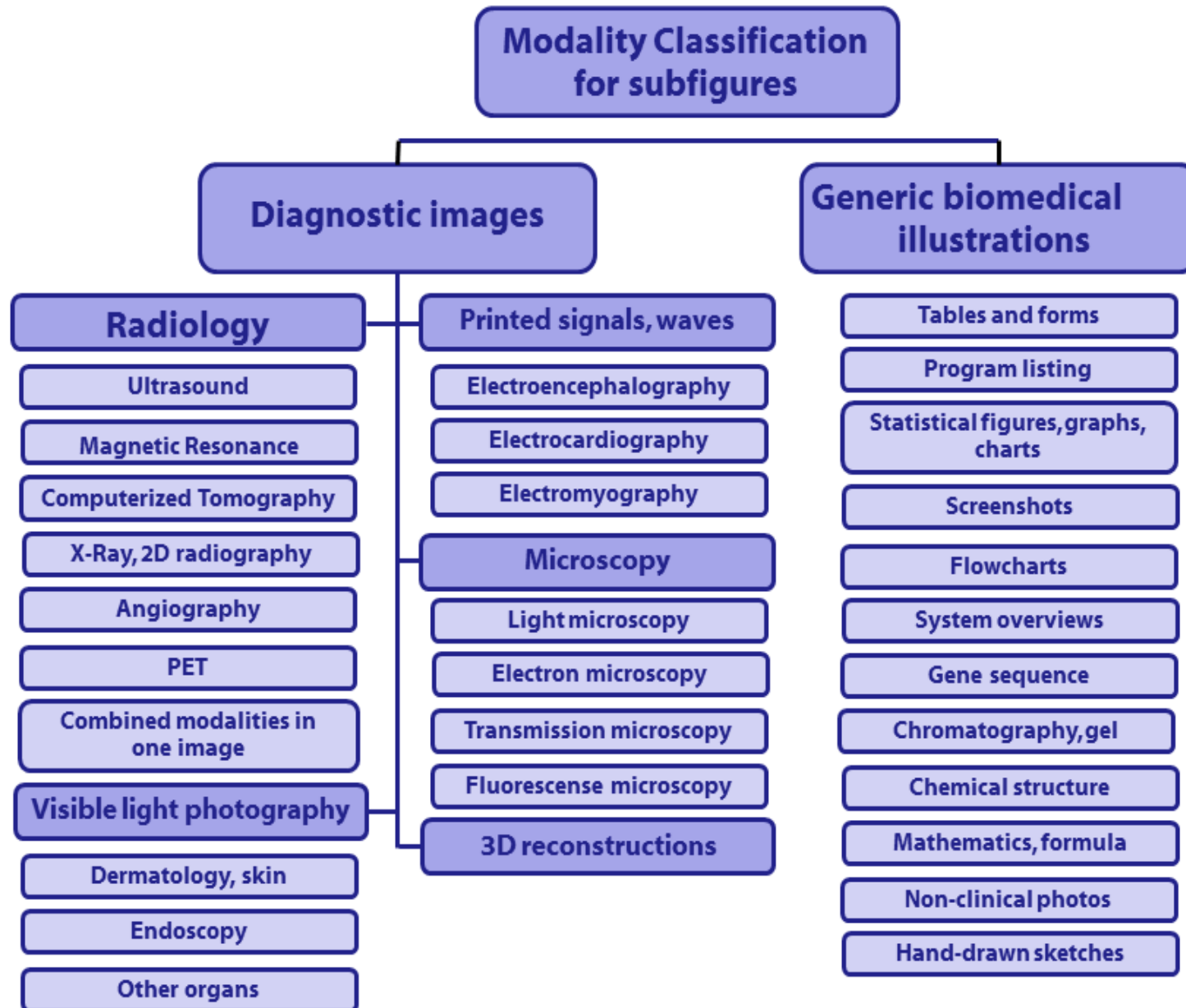
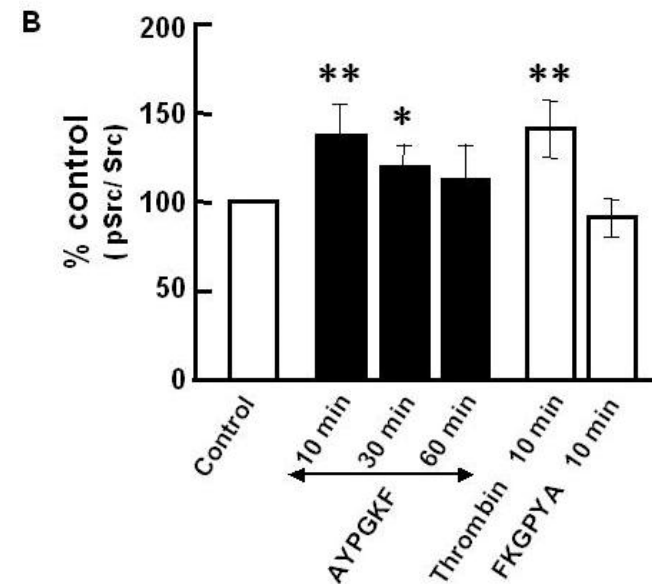
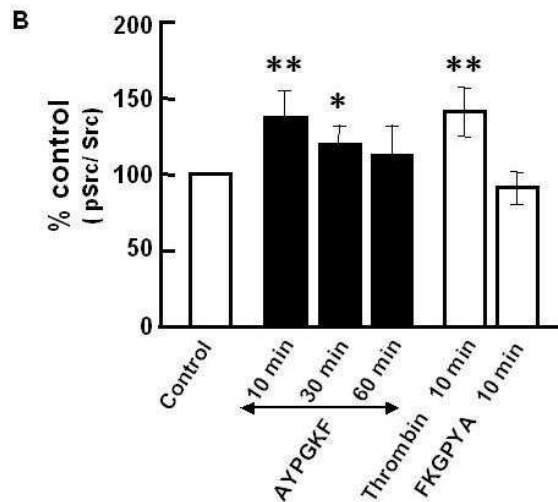
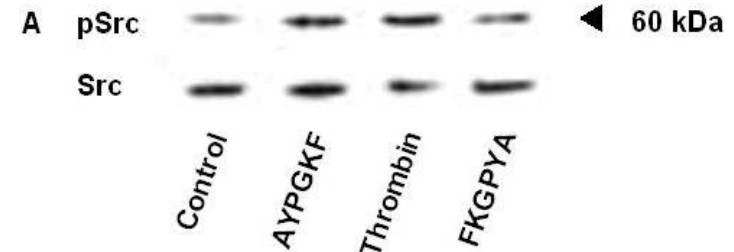
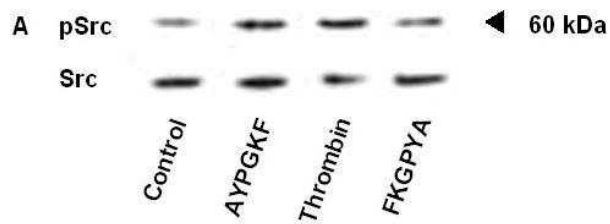


Image type hierarchy



Subfigure classification

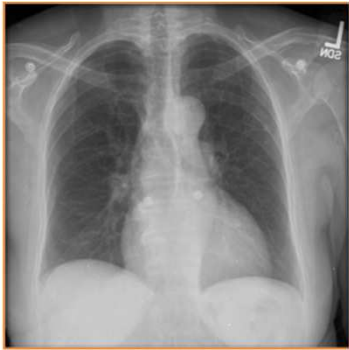

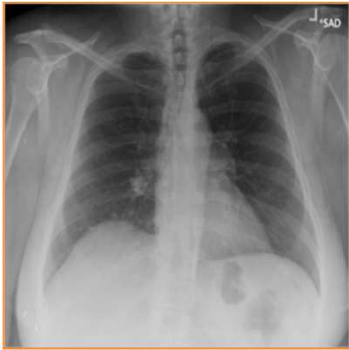




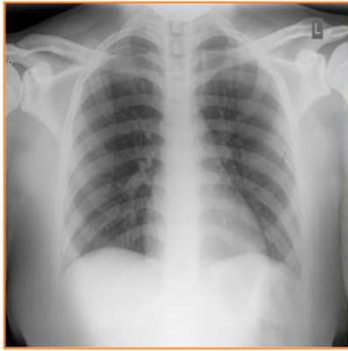
- To classify subfigures into the 30 classes



Caption detection



- To generate a caption that is a close as possible to the real one

| | input image | generated annotation | true annotation |
|--|---|--|--|
| |  | <p>aorta_thoracic / tortuous / mild</p> <p>aorta_thoracic / tortuous</p> | |
| |  | <p>opacity / lung / middle_lobe / right / aorta_thoracic / tortuous</p> <p>opacity / lung / base / left</p> | |
| |  | <p>calcified_granuloma / lung / middle_lobe / right / multiple</p> <p>calcified_granuloma / lung / hilum / right</p> | |
| |  | <p>opacity / lung / middle_lobe / right / blood_vessels</p> <p>calcified_granuloma / lung / middle_lobe / right</p> | |
| |  | | <p>airspace_disease / lung / hilum / right / lung / hilum</p> <p>nodule / lung / hilum / right</p> |
| |  | | <p>thoracic_vertebrae_degenerative / mild</p> <p>aorta_tortuous / thoracic_vertebrae_degenerative / mild</p> |
| |  | <p>normal</p> <p>normal</p> | |
| |  | <p>normal</p> <p>normal</p> | |

Caption prediction task

- Predicting the caption text based on the image content (and training data)
 - Using a word similarity metric to evaluate success
 - Can also help to make the image content accessible
 - Training data are available



- ImageCLEFmed 2016
 - 26,456 figures
 - Distributed in training and test sets
- Subset of PubMed Central
 - Over 4 million images of over 1,000,000 articles (2016)



Dataset by subtasks

- Compound figure detection:
 - Full dataset: 26,456 compound and single figures
- Compound figure separation
 - Subset containing 8,397 compound figures
- Multi-label classification
 - Subset containing 2,651 compound figures
- Subfigure classification
 - 10,942 subfigures of compound figures
- Caption prediction
 - 20,000 diagnostic figures (non-compound)

Compound figures and subfigures



- 2,651 figures are:
 - Labeled with all subfigure types
 - Separated into subfigures
 - Using Crowdsourcing
- Figure ID:
 - “1297-9686-42-10-3”
- Subfigure IDs:
 - “1297-9686-42-10-3-1”, “1297-9686-42-10-3-2”, ..., “1297-9686-42-10-3-4”

Participation

- Over 72 groups registered
- 8 groups from 4 continents submitted results
- 69 runs submitted



Results: compound figure detection



- **Multimodal** approaches achieve best results
- Best results using deep convolutional neural networks (**CNN**)

| Group | Run type | Accuracy |
|----------|----------|----------|
| DUTIR | mixed | 92.70 |
| CIS UDEL | mixed | 90.74 |
| CIS UDEL | mixed | 90.39 |
| CIS UDEL | mixed | 90.39 |
| CIS UDEL | mixed | 85.47 |
| CIS UDEL | mixed | 69.06 |
| CIS UDEL | mixed | 52.25 |
| MLKD | textual | 88.13 |
| DUTIR | textual | 87.03 |
| DUTIR | textual | 86.05 |
| CIS UDEL | textual | 85.47 |
| DUTIR | visual | 92.01 |
| CIS UDEL | visual | 89.64 |
| CIS UDEL | visual | 89.29 |
| DIS UDEL | visual | 69.82 |

Results: compound figure separation



- CIS UDEL applied:
 - Connected component analysis
 - Post-processing to avoid over-fragmentation

| Group | Run type | Accuracy |
|----------|----------|----------|
| CIS UDEL | visual | 84.43 |
| CIS UDEL | visual | 84.08 |
| CIS UDEL | visual | 84.03 |
| CIS UDEL | visual | 83.04 |
| CIS UDEL | visual | 81.23 |
| CIS UDEL | visual | 75.27 |
| CIS UDEL | visual | 74.83 |
| CIS UDEL | visual | 74.30 |
| CIS UDEL | visual | 73.57 |

Results: multi-label classification

- BMET uses **CNN** and **deep learning**
- MLKD uses a **textual** approach with a **random forest** classifier

| Group | Hamming Loss | F-Measure |
|-------|--------------|-----------|
| BMET | 0.0131 | 0.295 |
| BMET | 0.0135 | 0.320 |
| MLKD | 0.0294 | 0.320 |

Results: compound figure separation



- **Multimodal** approaches achieve best result
- CNN, feature selection, multiple visual descriptors are used

| Group | Run type | Accuracy |
|------------|----------|----------|
| BCSG | mixed | 88.43 |
| ... | mixed | ... |
| MLKD | textual | 72.22 |
| BCSG | textual | 58.37 |
| BCSG | visual | 85.38 |
| IPL | visual | 84.01 |
| BMET | visual | 77.55 |
| NWPU | visual | 76.38 |
| CIS UDEL | visual | 72.46 |
| NovaSearch | visual | 65.31 |
| ... | visual | ... |

Main tendencies

- **CNNs** (convolutional neural networks) are prominent in 2016
- **Multimodal** approaches achieve generally best results
- **Multiple features** used for **visual classification**
- **Connected component** analysis for figure separation

Conclusions

- Participants present a large **variety** of techniques
- The trend is towards the use of **neural networks**
- There were no submissions in the caption prediction subtask
- The subfigure classification subtask was the most popular task



Thank you for your attention!!!

Questions?

<http://imageclef.org/2016/medical>

albagarcia@nih.gov