Contact and more information:
albagarcia@nih.gov

Using crowdsourcing for multi-label biomedical compound figure annotation

Alba G. Seco de Herrera[1], Roger Schaer[2], Sameer Antani[1] and Henning Müller[2]
[1] National Library of Medicine, Bethesda, USA
[2] University of Applied Sciences Western Switzerland (HES–SO), Sierre, Switzerland;

Summary

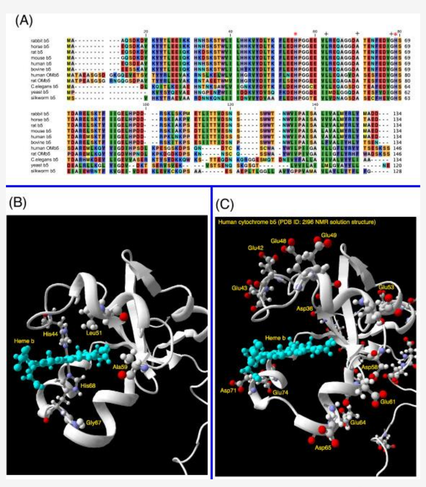
1. Probably more than **50%** of the **figures in the biomedical literature** in PubMed Central are **compound figures.**
2. Compound figures are annotated with figure type through a crowdsourcing process in the work described.
3. As a result **2,651 compound figures**, containing **8,397 subfigures**, were annotated with figure type and all figures were made available for the ImageCLEFmed 2016 multi–label task.

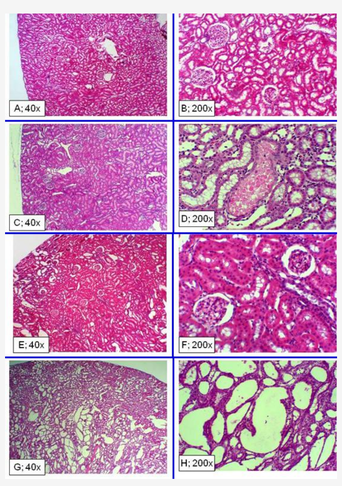
Introduction

- **Information analysis/retrieval** in the biomedical literature deals with a large amount of **compound figures**
- The **ImageCLEFmed** benchmark proposed a **multi-label classification task** in 2015 and 2016
- A methodology for multi-label annotation is needed

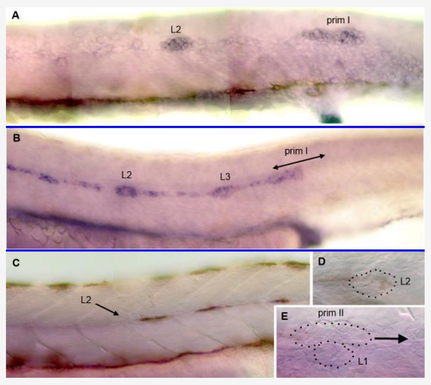
Methods

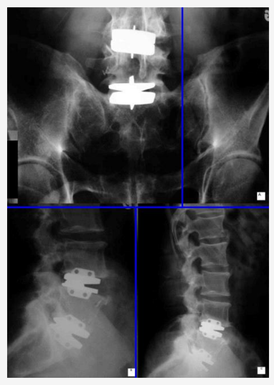
1. Automatic compound figure detection
   - Decide whether a figure is a compound or non-compound figure
2. Automatic compound figure separation
   - find the lines that cut compound figures into their parts





3. Manual compound figure separation verification
   - Check whether images were correctly separated

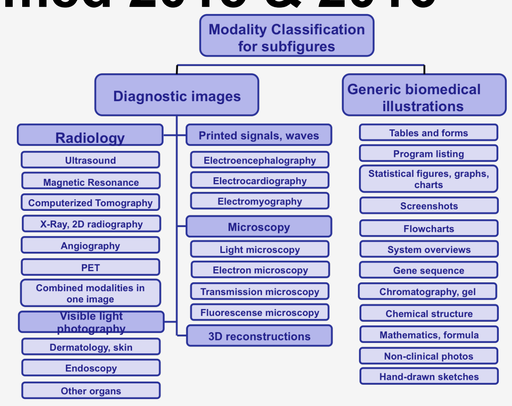




4. Automatic subfigure classification:
   - Automatic determination of the type of image in a subfigure
5. Manual subfigure classification verification
   - Validate the results of the previous step
6. Manual subfigure classification
   - Manually classify the images incorrectly classified in the automatic step
7. Manual class balancing
   - assure that all classes are represented
8. Compound figure multi–label assignment.

Image collection: ImageCLEFmed 2015 & 2016

- Subset of 231,000 figures from PubMed Central (over 4,200,000 figures)
- A hierarchy of figures types was used



Quality Control (QC)

- **Crowdflower** was used for crowdsourcing
- **QC** needed to ensure the success of the annotation
- QC during **design-time**
  - Design easy to understand
  - Automatic steps to limit manual tasks
  - Detailed and unambiguous description
  - Feedback from Crowdflower
- QC during **runtime**
  - Output agreement (at least from two contributors)
  - Control with known ground truth
  - Monitor answer patterns
- **Expert review** to finalize

Results

- **15,403** compound figures were selected and automatically separated
- ~57% of the figures were correctly separated based on a manual validation
- A subset was separated into subfigures and automatically classified
- ~56% were correctly classified based on a manual validation
- The incorrectly classified subfigures were manually classified
- An expert reviewed and solved any subfigure classification mistakes
- 122 figures contained rare subfigures types were added
- **~625 hours** were invested with a cost of **~870$**
- **~175,000 crowdsourced judgments** were performed by contributors

Conclusions

- Labels for **2,651** compound figures, containing **8,397 subfigures**, were generated using a **crowdsourcing** approach
- Annotations available for **ImageCLEFmed**