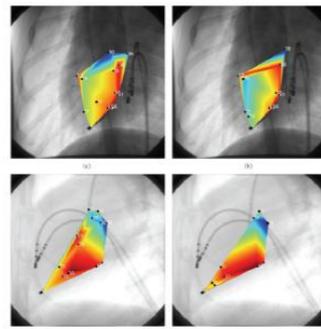
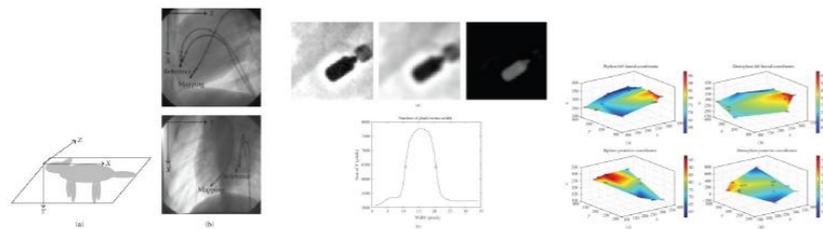




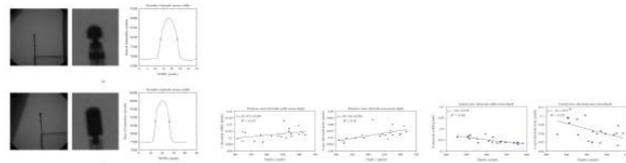
# Determining the importance of figures in journal articles to find representative images



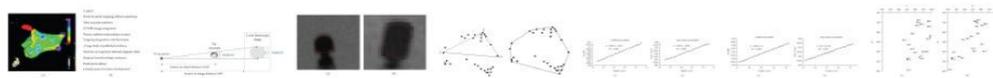
(a) Most important figure



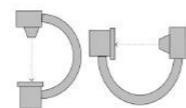
(b) Essential figures



(c) Important but not essential figures



(d) Not important figures



(e) Totally unimportant figures

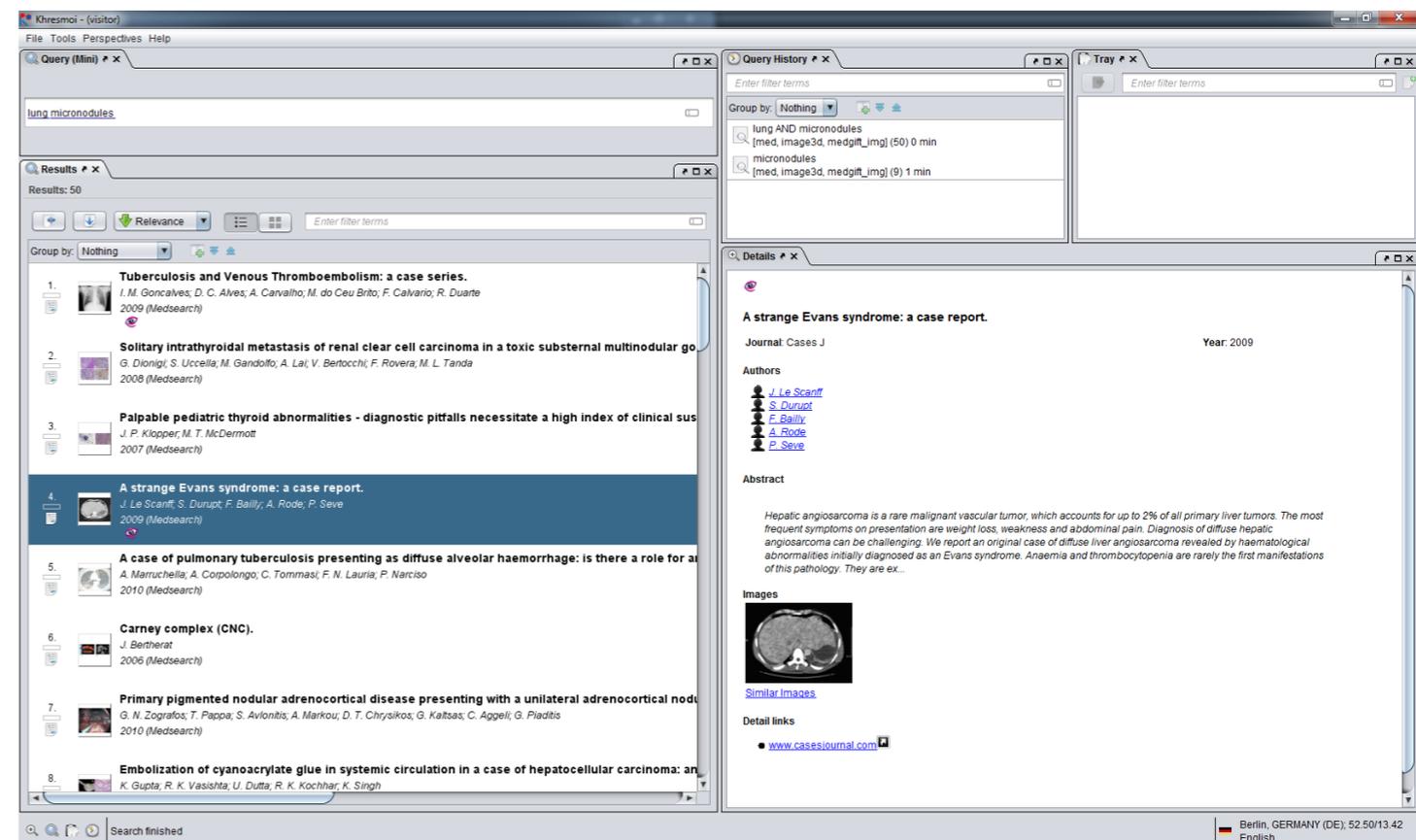
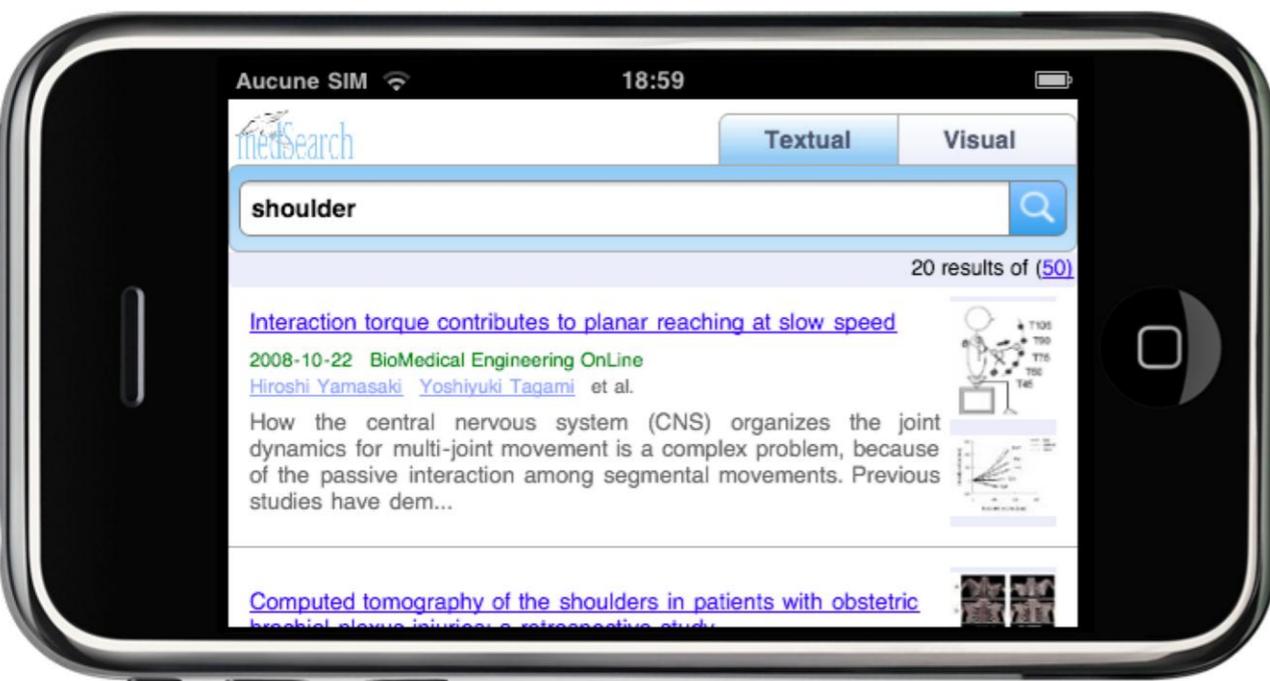
Henning Müller, Antonio Foncubierto-Rodriguez, Chang Lin, Ivan Egge

- Images are **essential for diagnosis** and treatment planning
- **Diversity** of modalities and protocols is growing
- Medical images represent **largest amount** of medical information produced
- Medical literature carries much of the medical knowledge
  - Images are essential for medical articles
  - Images can help understand content of an article quickly
    - Much quicker than reading a text to determine relevance
  - Search in the literature is frequent

- Image **retrieval** and **classification** make medical visual content accessible for search
  - Images from PACS or from the medical literature (as in the ImageCLEF image retrieval campaign)
- When searching the image needs to be put into context
  - Text and visual data to understand context
  - Often more than a single image
- Identification of **important figures**
  - Use of visual content to determine relevance quickly
  - Limit search to important figures

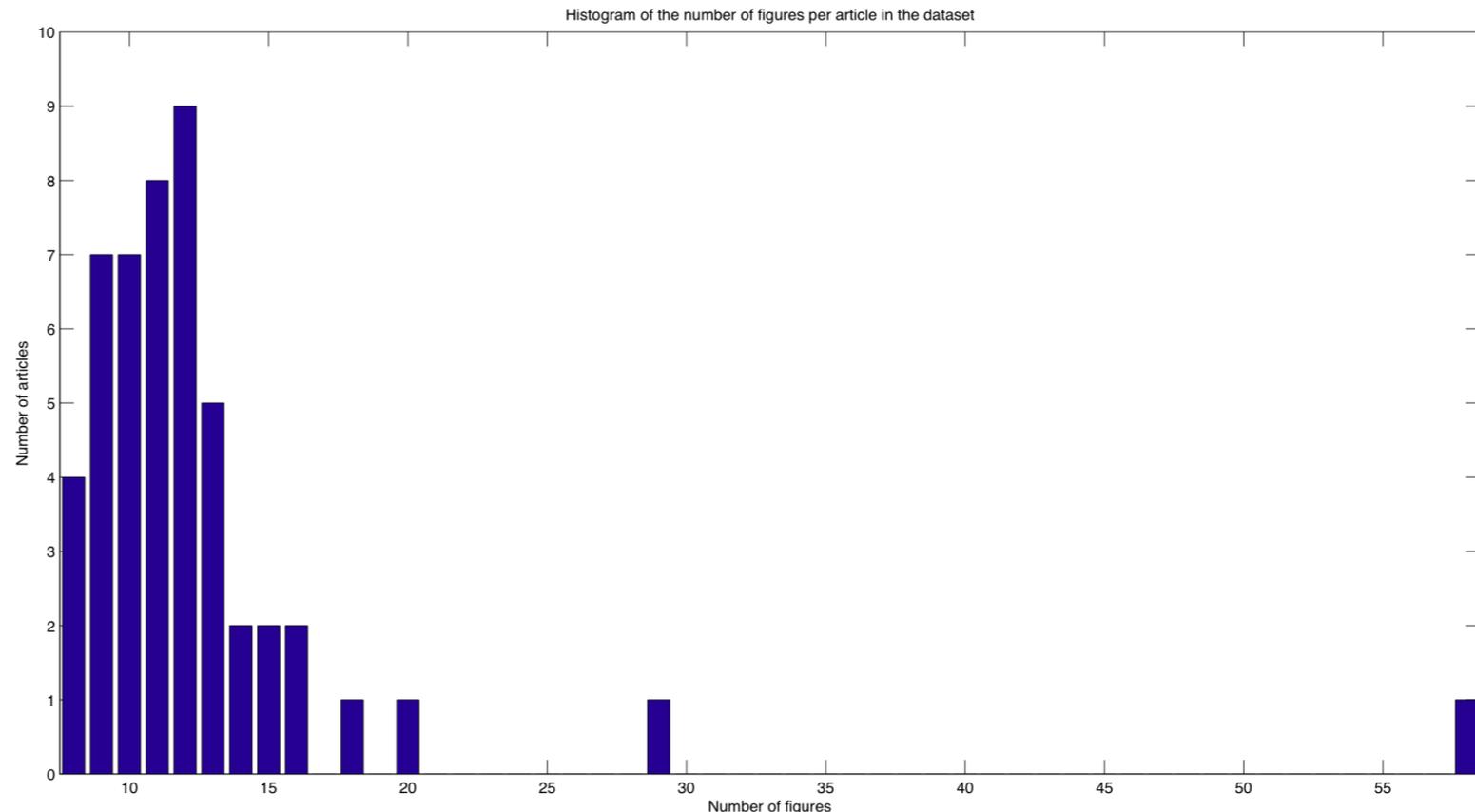
# Examples for using figure importance

- **Interfaces** and **visualization** require choices
- Mobile information search has small interface
- Creation of ground truth and rules for importance ranking to select in the best way
- Often the first N are taken



# Database used

- **ImageCLEF 2012** dataset
  - ~75,000 articles of **PubMed Central**, ~300,000 images
- Subset was used
  - ~800 articles contained more than 8 images
  - Final choice of 50 articles containing 641 diagnostic images



# Types of figure importance

- Five categories – sorted by **importance**
- **Most important** figure – 1 figure that best describes the article
- **Essential** figures – 0-3 figures
- **Important** figures – 0-3 figures
- Little importance figures – 0+ figures
- Totally unimportant figures – 0+ figures

- User test participant profile:
  - 45-year-old physician, surgeon
  - Working with medical images on a daily basis
- User test scenario:
  - Searching for relevant articles on the topic of the text
  - Which figures contain the most important information about the article?
- Requires to read and understand article
- All systematic findings or rules for the ranking were written down

- All 50 articles were **analyzed** and all 641 images **manually ranked** by importance
  - Usually over 1 hour per article
- **Rules** and findings were noted and discussed
- Statistics on figure importance and places in the text were made
- **Compound figures** were difficult to judge
  - Information content is high as they contain different subfigures and links between them
  - Compound figure separation new sub task in ImageCLEF

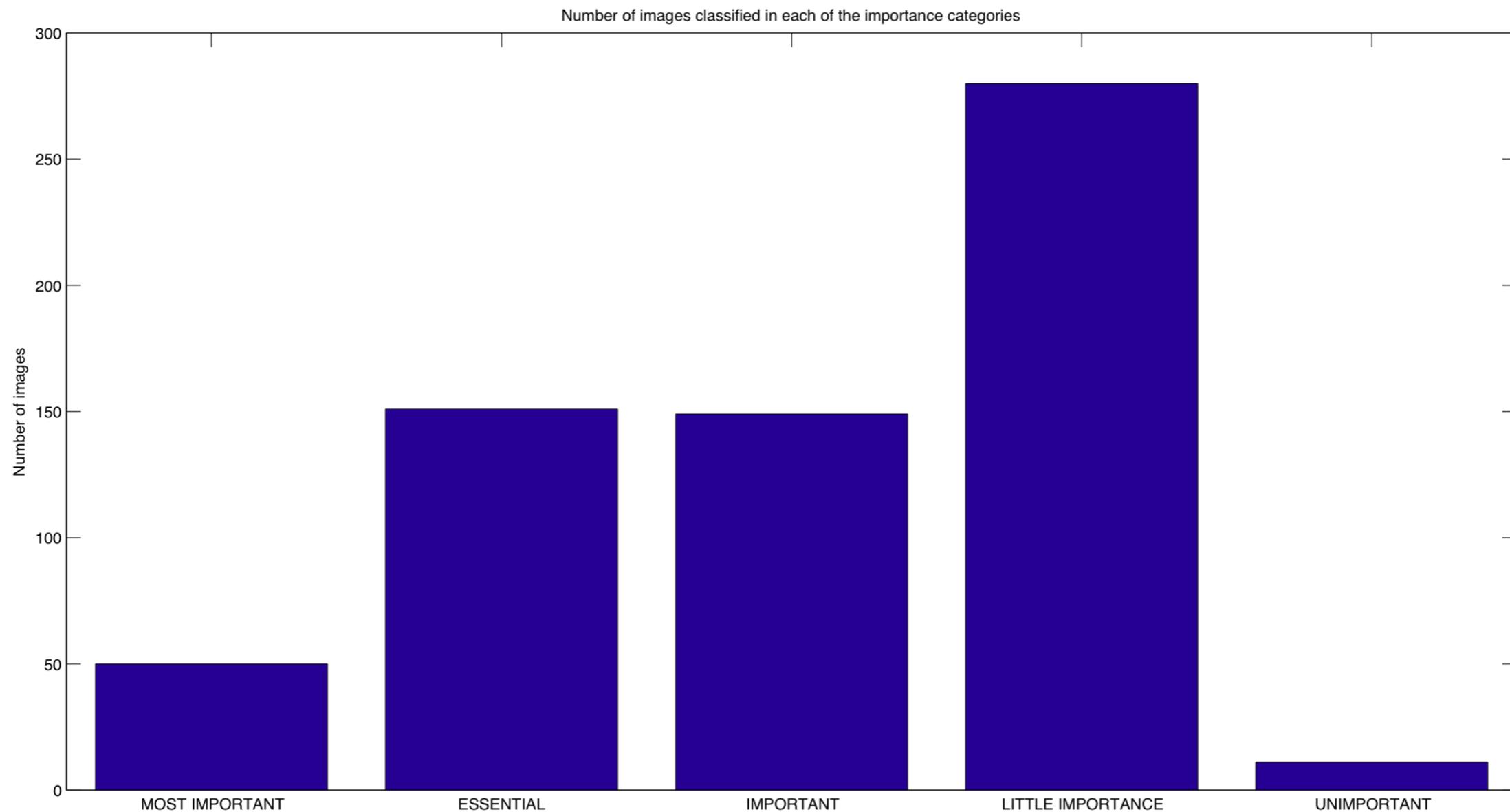
# Rules for determining importance

- Diagnostic images are more important than graphs/system overviews
- **Diversity** in the results set is important
  - Not several times the same modality if others exist
- Size/resolution matters
  - Image quality is importance
- **Position** in article
  - Most importance in the conclusions, then results
- **Common** vs. rare modality
  - Common modalities are often easier to understand

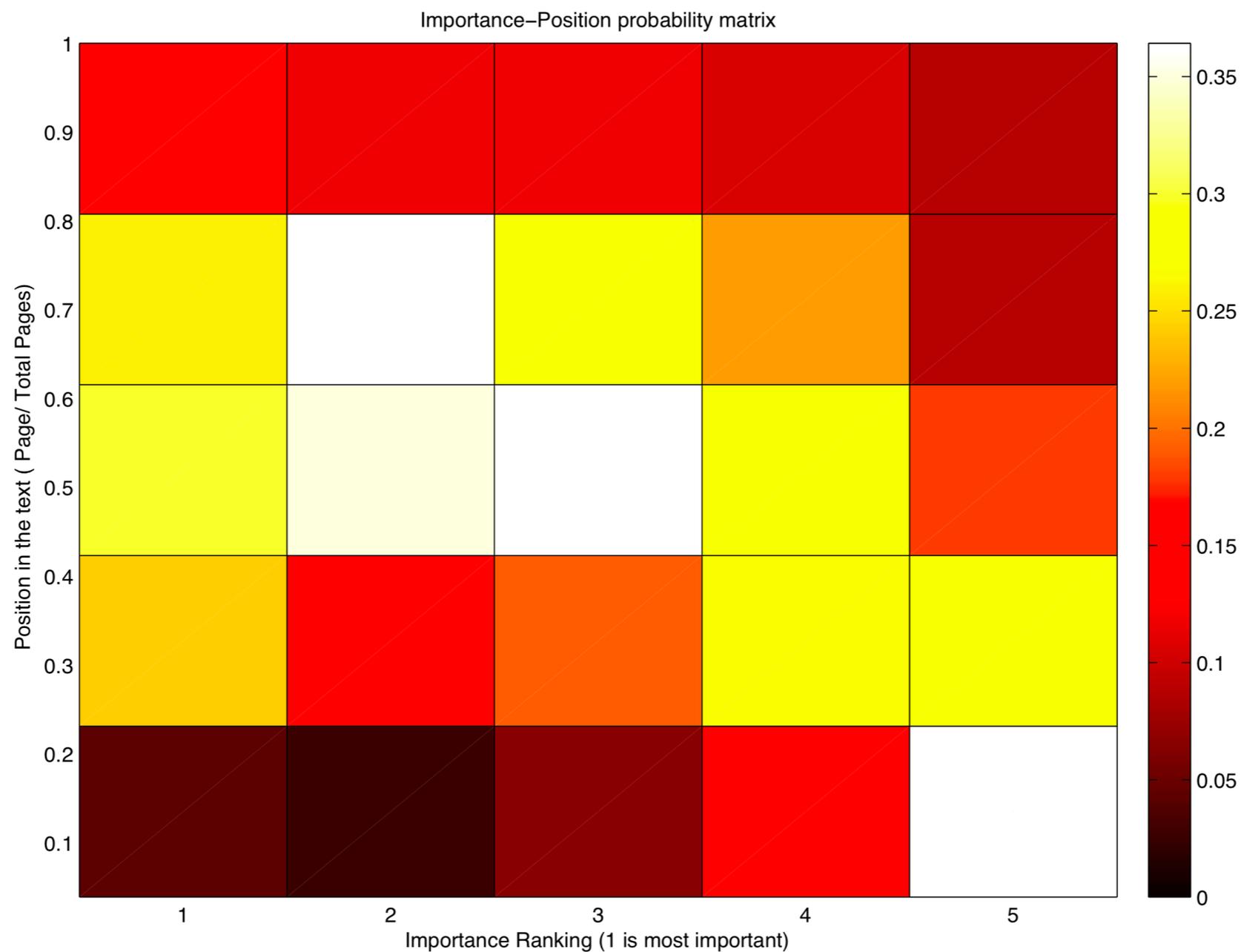
# More rules for the importance

- Magnification
  - Highest magnification most important
- MRI for soft tissue diseases, CT for bone related ones
  - **Modality adapted** to the disease, can be derived from RadLex terms
- Compound vs. other
  - Compound figure contains much information
- In non-diagnostic images, statistical graphs most important

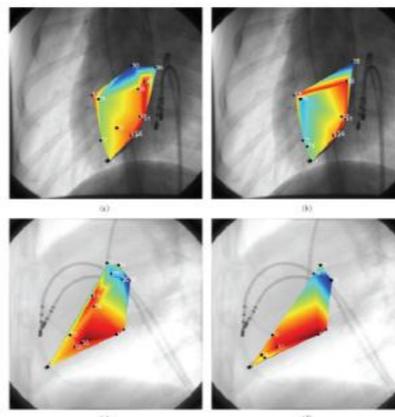
## ■ Distribution of figure importance



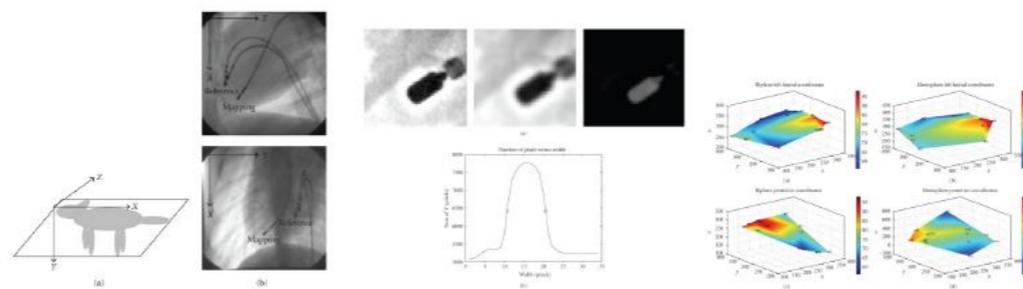
- **Importance vs. position** of the figure in the article based on relative position in the text



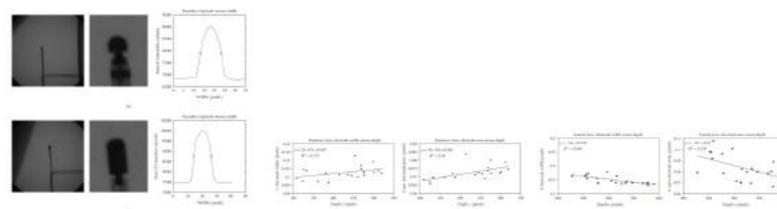
# Example for figure importance



(a) Most important figure



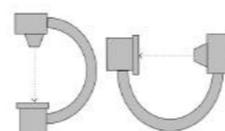
(b) Essential figures



(c) Important but not essential figures



(d) Not important figures



(e) Totally unimportant figures

- **Relative importance** of visual content in scientific articles is important for many applications
  - First study to investigate this in a **user test**
- Can improve IR system **performance**
  - Rank images based on position or remove some images to search a smaller sub space
- Can improve **visualization**
  - For small interfaces such as mobile phones and tablets
  - To combine text and images in an optimal way

- Repeat the same test with a second person to measure **inter-rater disagreement**
- Quantify the subjectivity of the task
- Model the rules and optimize an automatic selection strategy
- Modality classification to filter out unwanted images
- Clustering and use of cluster centers to increase diversity
- Find the place of figures in the text and rank based on this

Thank you for your attention!

Further information:

[henning.mueller@hevs.ch](mailto:henning.mueller@hevs.ch)

<http://medgift.hevs.ch/>

<http://www.khresmoi.eu/>

<http://www.mitime.org/width/>

