# Image CLEF Liver CT Image Annotation Task

Neda B. Marvasti[1], Nadin Kökciyan[2], Rüştü Türkay[3], Abdulkadir Yazıcı[2], Pınar Yolum [2], Suzan Üsküdarlı [2], Burak Acar[1]

[1] Boğazici Univ., EE Dept., Istanbul, Turkey

[2] Boğazici Univ., CmpE Dept., Istanbul, Turkey

[3]Istanbul Univ. Medical School, Çapa, Istanbul, Turkey

Sep 16, 2014

# Outline

- CES concept
- CaReRa project
- imageCLEF: Liver CT annotation task
- Task definition: datasets, evaluation methods
- Participants
- Results
- Conclusion

# Clinical Experience Sharing (CES)

- Clinical Experience Sharing (CES) refers to a searchable collective medical knowledge base that enables experience sharing among large community of medical professionals, for clinical and educational purposes.
- A CES platform would
  - Empower comparative diagnosis in the clinical use by presenting past cases that are relevant to a query case from a diagnostic point of view.
  - Assist medical students in the educational use by allowing them to browse past cases with similar/dissimilar symptoms and clinical observation but dissimilar/similar diagnoses.
- A CES platform can be implemented in the form of a Content-Based Case Retrieval (CBCR) system.

# CaReRa: Case Retrieval in Radiology

- CaReRa is a prototype CBCR implementation of the CES concept, which focuses on liver cases.
- Given a query case with incomplete representation, CaReRa searches and retrieves past cases relevant to the query case.
- CaReRa content analysis is context-free and is driven by an underlying ontology, user preferences and user relevance feedback.
- CaReRa (liver) case representation involves
  - *Demographics*
  - *Clinical history (ICD-10 codes)*
  - *Drugs (ATC codes)*
  - *Laboratory results*
  - *Physical examination*
  - *Semantic radiological (CT) observations (ONLIRA ontology): UsE*
  - *Low-level image (CT) features: CoG*

# imageCLEF: Liver CT Annotation Task

- **Motivation:**
  - The query cases in CaReRa are likely to be incomplete (missing the semantic UsE features).
  - It has been shown that using semantic (UsE) features give a better retrieval perform than low-level computer generated CoG features.[1]
  - Prediction of UsE features from a given CT volume and/or CoG features is required to build a query for CaReRa.
  - Besides, an automated semantic annotation using low-level computer generated features would be operational in standardized radiology reporting and CAD systems.

[1] Neda Barzegar Marvasti, Ceyhun Burak Akgül, Burak Acar, Nadin Kökciyan, Suzan Üsküdarl, Pnar Yolum, Rüstü Türkay, and Bars Bakr, Clinical experience sharing by similar case retrieval, in Proceedings of the 1st ACM international workshop on Multimedia indexing and information retrieval for healthcare. ACM, 2013, pp. 6774

## Liver CT annotation task:

# Task definition and Datasets

- **Task definition:**
  - Given a cropped CT volume enclosing the liver, the ONLIRA[1] ontology and a rich set of CoG features (for optional use), the task is to fill in a standardized radiology report that is composed of UsE features.
- **Datasets:**
  - 50 training and 10 test datasets.
  - Each training dataset is represented as:
    - *A cropped CT volume of the liver.*
    - *A liver mask, which defines liver in the image.*
    - *ROI, which defines lesion area in the image.*
    - *A set of 60 CoG image descriptors of size 454.*
    - *A set of 73 UsE features annotated using ONLIRA.*
  - Test sets has the same format but UsE features are missing, which are asked to be predicted.

[1] N. Kokciyan, R. Turkay, S. Uskudarli, P. Yolum, B. Bakir, and B. Acar, Semantic description of Liver CT images: An ontological approach, 2014.

## Liver CT annotation task:
# Computer generated features (CoG)

- CoG features describe the characteristics of liver, vessels and lesions of a CT image.
- Separated in two groups:
  - Global features (of size 93)

| Group | Descriptor | Size |
|---|---|---|
| **Liver** | Volume<br>Intensity mean and variance | 3 |
| **Vessels** | Volume<br>Ratio in the liver | 2 |
| **All lesions** | Intensity mean and variance<br>Number of lesions<br>Volume of biggest and smallest lesions<br>Ratio in the liver<br>Histogram<br>Histogram based features | 88 |

VAVlab

## Liver CT annotation task:

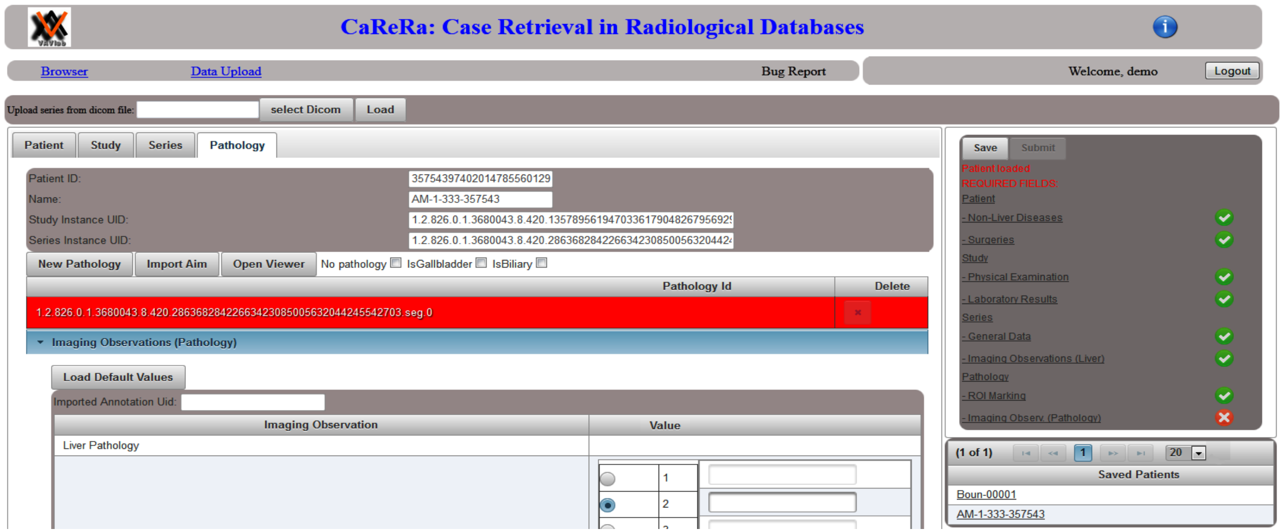
# Computer generated features (CoG)

- Individual features (of size 361)

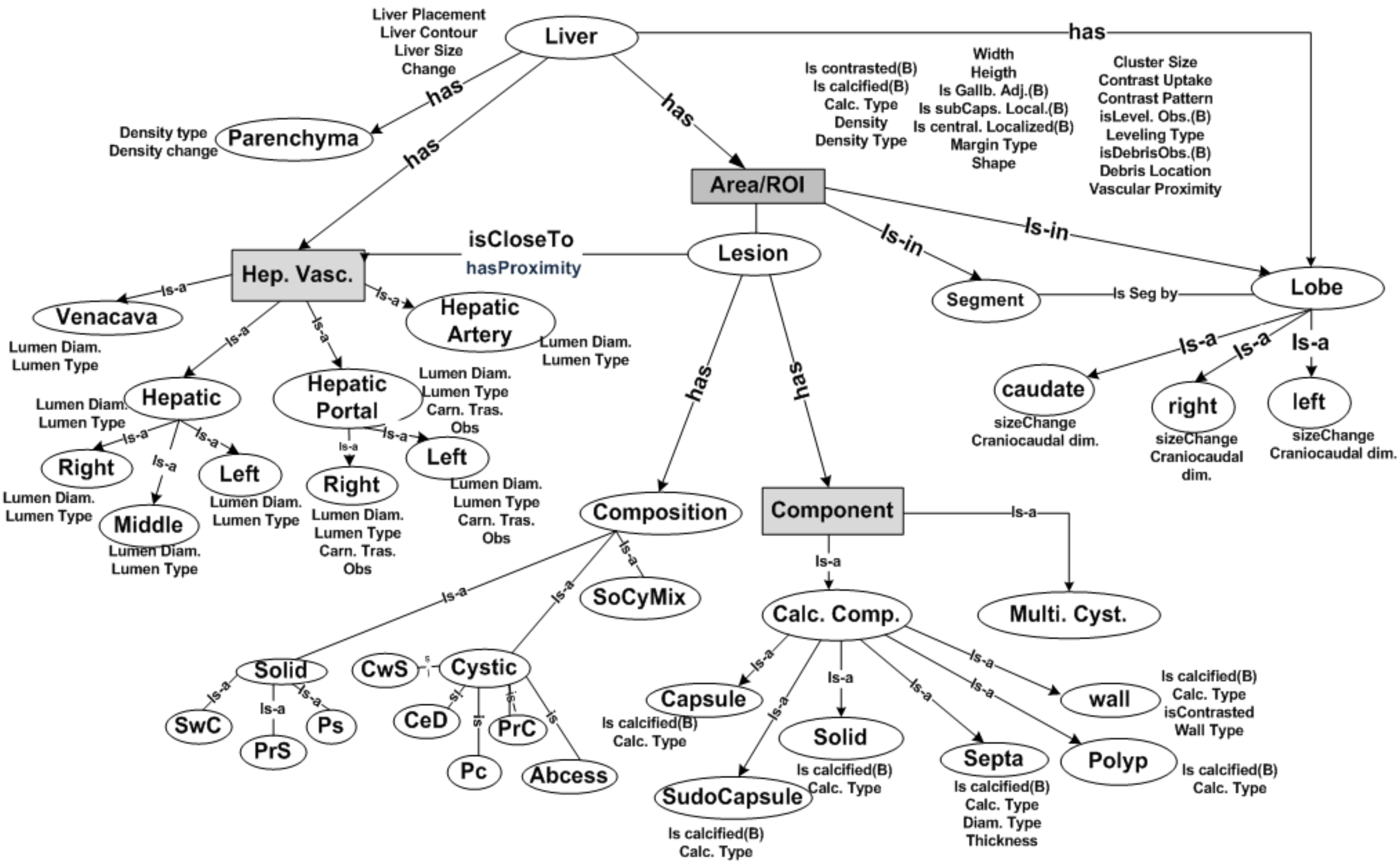
| Desc. Type | Desc. Name | Size |
|---|---|---|
| **Geometric** | Volume, area,<br>sphericity, convexity, solidity, compactness, max. extent, aspect ratio | 8 |
| **Locational** | Anatomical location<br>Convex area ratios<br>Touch area ratio<br>Proximity to vessels | 5<br>2<br>1<br>1 |
| **Boundary** | Scale, Window<br>Fourier descriptor | 33,3<br>3,20 |
| **GrayScale** | Histogram<br>Histogram based features | 84 |
| **Texture** | Haralick<br>Tamura<br>Gabor<br>Hu moments | 24<br>83<br>64<br>3 |

# Liver CT annotation task:
# User expressed features (UsE)

- ONLIRA (**ON**tology of the **LI**ver for **RA**diology) [1]
  - models the imaging observations of the liver domain.
  - represents properties and relationships between liver, hepatic veins and lesions.
- Imaging observations of a radiologist for the liver domain are represented with ONLIRA and collected in CaReRa-Web.
- CaReRa-Web is a web based data application which can be accessed at https://vavlab.ee.boun.edu.tr:5904/CareraWeb2



# ONLIRA


Liver
Liver Placement
Liver Contour
Liver Size
Change
has
Parenchyma
Density type
Density change
Area/ROI
Is contrasted(B)
Is calcified(B)
Calc. Type
Density
Density Type
Width
Heigth
Is Gallb. Adj.(B)
Is subCaps. Local.(B)
Is central. Localized(B)
Margin Type
Shape
Cluster Size
Contrast Uptake
Contrast Pattern
isLevel. Obs.(B)
Leveling Type
isDebrisObs.(B)
Debris Location
Vascular Proximity
Hep. Vasc.
isCloseTo
hasProximity
Lesion
Is-in
Segment
Is Seg by
Lobe
Is-a
caudate
sizeChange
Craniocaudal dim.
right
sizeChange
Craniocaudal
dim.
left
sizeChange
Craniocaudal dim.
Venacava
Lumen Diam.
Lumen Type
Hepatic
Artery
Lumen Diam.
Lumen Type
Hepatic
Lumen Diam.
Lumen Type
Hepatic
Portal
Lumen Diam.
Lumen Type
Carn. Tras.
Obs
Right
Lumen Diam.
Lumen Type
Middle
Lumen Diam.
Lumen Type
Left
Lumen Diam.
Lumen Type
Right
Lumen Diam.
Lumen Type
Carn. Tras.
Obs
Left
Lumen Diam.
Lumen Type
Carn. Tras.
Obs
Composition
Component
Solid
SwC
PrS
Ps
CwS
Cystic
CeD
Pc
PrC
Abcess
SoCyMix
Calc. Comp.
Multi. Cyst.
Capsule
Is calcified(B)
Calc. Type
SudoCapsule
Is calcified(B)
Calc. Type
Solid
Is calcified(B)
Calc. Type
Septa
Is calcified(B)
Calc. Type
Diam. Type
Thickness
wall
Is calcified(B)
Calc. Type
isContrasted
Wall Type
Polyp
Is calcified(B)
Calc. Type

## Liver CT annotation task:
# Evaluation methodology

- Evaluation is based on completeness and accuracy of the predicted annotations with reference to the manual annotations of the test dataset.

$$Completeness = \frac{\text{number of predicted UsE features}}{\text{total number of UsE features}}$$

$$Accuracy = \frac{\text{number of predicted UsE features}}{\text{total number of predicted UsE features}}$$

- For answers, which allow multiple values to a question, the correct prediction of a single value is considered as the correct annotation.
- 7 out of 73 UsE features were excluded from the evaluation due to their unbounded labels(numeric continuous values).

# Liver CT annotation task: Participants

- 20 groups registered, however 3 of them submitted their results.
- The number of runs per group was limited to ten.

| Group name | Affiliation | runs |
| --- | --- | --- |
| BMET | University of Sydney, Australia | 8 |
| CASMIP | The Hebrew University of Jerusalem, Israel | 1 |
| piLab | Bogaziçi University, Turkey | 1 |

- BMET group, achieved the best results using an image retrieval technique with total score of 0.947.
- A classifier-based method was employed by CASMIP group.
- piLab used a Generalized couple tensor factorization (GCTF) method.

VAVlab

## Liver CT annotation task:
# BMET group results [1]

- They submitted 8 runs:
  - 4 using classier-based approach (RBF and linear kernels)
  - 4 using image retrieval algorithm (with feature selection, without feature selection)
- Experiments with 2 different feature sets:
  - CoG features
  - CoG features and bag of visual words (BoVW)
- The best result is achieved in the experiments with an image retrieval approach and by using the CoG features only.

1: Ashnil Kumar, Shane Dyer, Changyang Li, Philip H. W. Leong, and Jinman Kim, Automatic annotation of liver ct images: the submission of the bmet group to imageclefmed 2014, in CLEF 2014 Labs and Workshops, Notebook Papers. CEUR Workshop Proceedings (CEUR-WS.org), September 2014.

## Liver CT annotation task:
# BMET group results

- In Classifier-based approach
  - They used 2-stage support vector machine (SVM) classification to annotate every UsE feature.
  - 1st stage is done using 1-vs-all SVM classifier.
  - 2nd stage is done using 1-vs-1 SVM classifier.
- 2nd stage is activated if the result of 1st step is more than one label, which is applied to the results of 1st step followed by a majority voting.
- This approach is employed using two different kernels and on two different feature sets.

## Liver CT annotation task:
# BMET group results

- In image retrieval based approach
  - The most similar images from the training set to the current image are selected.
  - Then, a weighted voting scheme is applied to assign labels to each of the UsE features.
  - Similarity measure is defined as Euclidean distance.
- A sequential feature selection method is applied to use the most distinct features for each question during the similarity calculation.
- This approach is done with and without feature selection on two different feature sets.

VAVlab

## Liver CT annotation task:

# BMET group results

| Group | Run | Completeness | Accuracy | Score | Method | Feature |
|---|---|---|---|---|---|---|
| BMET | 1 | 0.98 | 0.89 | 0.935 | SVM-linear | CoG |
| BMET | 2 | 0.98 | 0.90 | 0.939 | SVM-linear | CoG+ |
| BMET | 3 | 0.98 | 0.89 | 0.933 | SVM-RBF | CoG |
| BMET | 4 | 0.98 | 0.90 | 0.939 | SVM-RBF | CoG+ |
| BMET | 5 | 0.98 | 0.91 | 0.947 | IR-noFS | CoG |
| BMET | 6 | 0.98 | 0.87 | 0.927 | IR-noFS | CoG+ |
| BMET | 7 | 0.98 | 0.91 | 0.947 | IR-FS | CoG |
| BMET | 8 | 0.98 | 0.87 | 0.926 | IR-FS | CoG+ |

## Liver CT annotation task:
# CASMIP group results [1]

- They achieved the 2nd best performance.
- They tried 4 different classifiers in the learning phase:
  - Linear discriminant analysis (LDA)
  - Logistic regression (LR)
  - K-nearest neighbors (KNN)
  - SVM
- In learning phase, for every UsE feature the best classifier and CoG feature sets are learned via leave-one-out cross validation method.
- CoG features with dimensionality more than one are ignored, which reduces the number of employed CoG features to 39.

1: Assaf B. Spanier and Leo Joskowicz, Towards content-based image retrieval: From computer generated features to semantic descriptions of liver ct scans, in CLEF 2014 Labs and Workshops, Notebook Papers. CEUR Workshop Proceedings (CEUR-WS.org), September 2014.

# Liver CT annotation task:
# CASMIP group results

- Nine additional low level image features describing the gray level properties of liver and lesion and also boundary properties of the lesion are used.
- Three UsE features including: cluster size, segment and lobe were extracted directly from the image features.
- For most of UsE features, they observed the same performance using any classifier and any combination of CoG features.
- For 6 of them related to density, contrast and location of the lesion, one of the LDA or KNN is chosen with their selected features.

| Group | Run | Completeness | Accuracy | Score | Method | Feature |
|---|---|---|---|---|---|---|
| CASMIP | 1 | 0.95 | 0.91 | 0.93 | LDA + KNN | CoG+ |

VAVlab

# Liver CT annotation task: piLab group results [1]

- They considered the dataset as a heterogeneous data and applied coupled matrix factorization models using GCTF framework.
- Both KL divergence and Euclidean distance based cost functions are applied.
- They considered two groups of UsE features: the 1st group includes UsE features, which have values vary from 0 to 3 and the 2nd group contains UsE features that have binary values.
- Following matrices are provided:
  - X1: UsE features of first group(60*21)
  - X2: UsE features of second group (60*13)
  - Z1: CoG features (60*447)

1: Beyza Ermis and A. Taylan Cemgil, Liver ct annotation via generalized coupled tensor factorization, in CLEF 2014 Labs and Workshops, Notebook Papers. CEUR Workshop Proceedings (CEUR-WS.org), September 2014.

## Liver CT annotation task:
# piLab group results

- Then the latent matrices Z2 and Z3 are estimated as:

  X1 = Z1*Z2  X2 = Z1*Z3
- The UsE features of test cases can be predicted using Z2 and Z3 via GCTF.
- Since the predicted values are not discrete values, a binary thresholding is applied.
- This group submitted three runs:

| Group | Run | Completeness | Accuracy | Score | Method | Feature |
|---|---|---|---|---|---|---|
| piLab | 1 | 0.51 | 0.39 | 0.45 | GCTF-KL | CoG |
| piLab | 2 | 0.51 | 0.89 | 0.677 | GCTF-EUC | CoG |
| piLab | 3 | 0.51 | 0.88 | 0.676 | GCTF-KL | CoG |

## Liver CT annotation task:
# Discussion

- The BMET group achieved the highest score with completeness of %98.
- In terms of accuracy, BMET group has also attained the best performance by using an image retrieval method.
- In terms of classier-based methods, BMET and CASMIP groups both obtained the total score of %93.

| Group | Completeness | Accuracy | Score | Method | Feature |
|---|---|---|---|---|---|
| BMET | 0.98 | 0.91 | 0.947 | Image Retrieval | CoG |
| CASMIP | 0.95 | 0.91 | 0.93 | LDA + KNN | CoG+ |
| piLab | 0.51 | 0.89 | 0.677 | GCTF | CoG |

# Liver CT annotation task: Discussion

- Results of different runs in predicting different groups of UsE features are as:

| Group | Liver | | Vessel | | Lesion-Area | | Lesion-Lesion | | Lesion-Component | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Cmp. | Acc | Cmp | Acc | Cmp. | Acc | Cmp. | Acc | Cmp | Acc |
| BMET | 1.00 | 0.93 | 1.00 | 1.00 | 0.92 | 0.79 | 1.00 | 0.83 | 1.00 | 0.94 |
| CASMIP | 1.00 | 0.93 | 1.00 | 1.00 | 0.85 | 0.81 | 0.90 | 0.82 | 1.00 | 0.94 |
| piLab | 0.62 | 0.88 | 1.00 | 1.00 | 0.46 | 0.77 | 0.20 | 1.00 | 0.12 | 0.15 |

## Liver CT annotation task:
# Discussion

- Results show that all groups have completed the vessel UsE features with high accuracy.
- The BMET and CASMIP groups completed liver features in full with accuracy more than %80.
- None of the groups can completely annotate the concepts related to lesion-area.
- Concepts related to lesion-components are fully completed and annotated with accuracy higher than %72 by both BIMET and CASMIP groups.
- Concepts related to lesion-lesion are annotated completely by only BIMET group with accuracy more than %72.

# Liver CT annotation task: Conclusion

- This was the first time that liver CT annotation task was proposed.
- The challenge was to predict UsE features of patient records, given the CoG features.
- The main challenge of the task was due to the unbalanced dataset. Participants tried to overcome this issue with different methods.
- Among all methods, image retrieval scored the best performance.
- It was observed that feature selection is important for the best performance of the prediction method.

# Thank you

## Any question?