

Overview of the ImageCLEF 2015 liver CT annotation task

Neda B.Marvasti¹, María del Mar Roldán García³, Suzan Uskudarli², José F. Aldana³, and Burak Acar¹

¹ Boğaziçi University, Electrical and Electronics Department, Istanbul, Turkey

² Boğaziçi University, Computer Engineering Department, Istanbul, Turkey

³ University of Malaga, Computer Languages and Computing Science Department
Malaga 29071, Spain

{neda.marvasti@boun.edu.tr, acarbu@boun.edu.tr}
vavlab.ee.boun.edu.tr

Abstract. The second Liver CT (Computed tomography) annotation challenge was organized during the 2015 Image-CLEF workshop, hosted by the Institut de Recherche en Informatique de Toulouse (IRIT), University of Toulouse, France. This challenge entailed the annotation of Liver CT scans to generate structured reports. This paper describes the motivations for this task, training and test datasets, the evaluation methods, and discusses the approaches of the participating groups.

Keywords: ImageCLEF, Liver CT annotation task, Automatic annotation

1 INTRODUCTION

ImageCLEF [7] was part of the Cross Language Evaluation Forum (CLEF) 2015 consisting of four main tasks: Image Annotation, Medical Classification, Medical Clustering, and Liver CT Annotation. It was the second time that the automatic annotation of Liver CT images was provided as a challenge. However, there are two changes compared to the last edition of the challenge. First, the format of the given Use (user expressed) features has been changed. Furthermore, there is no CoG (Computer generated) features provided this year. In the term of Use features, LiCO (liver case ontology) is used instead of ONLIRA (ontology of liver for radiology)[4], which has additional patient and study information.

The purpose of the Liver CT annotation task is to automatically generate structured reports with use of computer generated features of liver CT volumes. Structured reports are highly valuable in medical contexts due to the processing opportunities they provide, such as reporting, image retrieval, and computer-aided diagnosis systems. However, structured report generation is cumbersome and time consuming. Furthermore, their creation requires administration of the domain expert, who is time constrained. Consequently, such structured medical reports are often not found or are incomplete in practice. This challenge has

been designed to aid the fill a pre-prepared structured report using the imaging information derived from CT images.

The data provided for this challenge consists of 50 training and 10 test datasets. Participants were asked to answer a fixed set of multiple-choice questions about livers. The questions were automatically generated from an open-source liver case ontology (LiCO) [4] and provided in files with RDF (resource description framework) format. The answers to the questions describe the properties of the liver, the hepatic vasculature of the liver, and a specific lesion within it. During this task, the user is presented with the following training data: (1) data from a CT scan, (2) a liver mask, (3) a volume-of-interest that highlights the selected lesion, and (4) a rich set of imaging observations (annotations) provided in RDF format. The imaging observations are LiCObased annotations that were manually entered by radiologists. Participants need to extract their own image features from the CT data and use them to automatically annotate the liver CT volumes. The results have been evaluated in terms of the completeness and accuracy of the generated annotations.

The rest of the paper is organized as follows, Section 2 gives a detailed description of the task and introduces the participants. Section 3 presents the main results of the task and the results of the participants, and finally, Section 4 concludes the paper.

2 The Liver CT Annotation Challenge

This section describes the task and introduces the participants.

2.1 Task Definition and Datasets

The Liver CT annotation task is proposed towards the generation of structured reports describing the semantic features of the liver, its vascularity, and the types of its lesions. The goal of this task is to develop automated mechanisms to assist medical experts in difficult and practically infeasible task of annotating medical records.

The training dataset includes 50 cases, each consisting of:

- a cropped CT image of the liver – a 3D matrix with the same size as cropped CT image,
- a liver mask that specifies the part corresponding to the liver – a 3D matrix indicating the liver areas with a 1 and nonliver areas with a 0,
- a bounding box (ROI) corresponding to the region of the selected lesion within the liver – as a vector of 6 numbers corresponding to the coordinates of two opposite corners,
- An RDF file generated using LiCOrepresenting manually entered imaging observations by a radiologist. In total, there are 73 Use features. If a feature is not applicable for a case, it will not be represented in the corresponding RDF file.

In the training set, 50 ".mat" files, each containing the first three of above data, as well as an RDF file representing the imaging observations, have been given to the participants. In the test set, there is no RDF file and imaging observations are missing and participants are asked to predict them. The participants have been asked to extract and use their own image features to complete the task. RFD files include information of patient, study, and imaging observation. Participants are expected to predict the only imaging information, same as the last year's challenge.

The resolution of CT images may vary in the range of (x : 190 – 308 pixels, y : 213 – 238 pixels, and z : 41 – 588 slices). The spacing may also vary in the range of (x, y : 0.674 – 1.007 mm, $slice$: 0.399 – 2.5 mm).

It is important to note that, this dataset is partially available through image-CLEF2015 system (<http://medgift.hevs.ch:8080/CLEF2015/faces/Login.jsp>), for academic use only. If you are interested in using this dataset, you need to properly cite this paper.

User Expressed Features Imaging observations of a radiologist for the liver domain are represented with LiCO. A web-based data collection application, called CaReRa-Web (case retrieval in radiological databases), which can be accessed for academic use from the CaReRa project website <http://www.vavlab.ee.boun.edu.tr>.

For each case, there are 73 UsE features represented in RDF format. These features clinically characterize the liver, hepatic vascularity, and liver's lesions. In the training set, the UsE features are manually entered by an expert radiologist. Every UsE feature corresponds to a question answered by a radiologist. Some UsE features may take on more than one values. Such features are represented with a multi-choice answers. Features with value marked as "NA" are not included in the RDF file.

In the test phase, the participants are expected to predict the UsE features in the following format (73×4 UsE data):

Column Annotation Features Type

1	Group	string
2	Concept	string
3	Properties	string
4	Values	bar separated list of strings

The "Group" and "Concept" are the LiCO-based concepts. Each concept may have several properties. Each property may have multiple values, whose indices and meaning are given in Table 4 under "Possible values" column. Properties deemed irrelevant are marked as NA by the radiologist. Note that, UsE features are grouped as: Liver, Vessel, General, and Lesion. Table 4 lists every group and its corresponding concepts, properties, possible values, and their assigned indices.

In every RDF file, there is a patient which has a study and a set of data properties same as Name, age and gender, as well as a set of object properties same as disease and drugs. Each study has a set of data properties including ID

and dates as well as a set of object properties same as laboratory results and physical examinations. Also each study has a liver, which has a set of imaging observation. Each liver also has a lesion, which has relevant imaging observation.

2.2 Evaluation methodology

The evaluation is performed on the basis of completeness and accuracy of the predicted annotations with reference to the manual annotations of the test dataset. Completeness is defined as the number of predicted features divided by the total number of features, while accuracy is the number of correctly-predicted features divided by the total number of predicted features.

For answers that allow multiple values to a question, the correct prediction of a single feature is considered as the correct annotation.

$$Completeness = \frac{\text{number of predicted } UsE \text{ features}}{\text{Total number of } UsE \text{ features}} \quad (1)$$

$$Accuracy = \frac{\text{number of correctly predicted } UsE \text{ features}}{\text{Number of predicted } UsE \text{ features}} \quad (2)$$

$$TotalScore = \sqrt{Completeness \times Accuracy} \quad (3)$$

2.3 Participation

Among 32 groups, which registered for the task and signed the license agreement to access the datasets, only 1 of them submitted his results. The group name is "CREDOM", from Biomedical Engineering Laboratory, Tlemcen University, Algeria. They have submitted three runs to the task.

3 Results

3.1 Runs submitted in 2014

Last year in 2014 [1], three groups submitted their results. Their prediction were based on classifiers, image retrieval, and generalized-coupled tensor factorization (GCTF). Last year a set of computer generated (CoG) features were also provided for the participants as an optional data.

The best performance was achieved by the BMET group [5] submitting 8 runs using two different methods: classifier-based approach using SVM (support vector machine) and image retrieval algorithm. They also used two different sets of feature: the prepared CoG features from the database and a bag of visual words (BoVW). Their classification methods outperformed the other methods, when they employed their expanded feature set. However, their retrieval method gave the best results, when the given CoG features were employed. This observation suggests that the nature of feature sets are important for utilizing different methods.

The second best performance was achieved by CASMIP group [2] submitting only one run to the task. They tried four different classifiers in the learning phase: linear discriminant analysis (LDA), logistic regression (LR), K-nearest neighbors (KNN), and finally SVM to predict UsE features. For each UsE feature the best classifier and its related features were learnt by using exhaustive search. They used only a certain part of provided CoG features, as well as 9 additional features extracted in the lesion ROI. As a result, for most of the UsE features, they achieved the same performance using any classifier and any combination of image features.

piLabVAVlab group [3] considered the dataset as heterogeneous data and GCTF approach was applied to predict UsE features. They considered both KL-divergence and Euclidean-distance-based cost functions, as well as the coupled matrix factorization models using GCTF framework.

The BMET group achieved the highest scores with completeness of %98 (See Table 3. In terms of accuracy, BMET group has also attained the best score by using an image retrieval method.

3.2 Runs submitted this year in 2015

This year in 2015, three runs has been submitted by the "CREDOM" group [6]. They used two different methods: classification by using random forest (RF) classifier, and retrieval by considering the specific signature of the liver (See Table 1.

For the classification-based method (run 1 and run 2), they have employed two different sets of features. The first set contains 115 liver texture features in addition to 9 lesion geometric features, and the second set includes 214 lesion texture features, in addition to 9 lesion geometric features. Classification is performed by using supervised multi-class RF classifier. In this method, they divided the UsE features into two groups. For the first group, they have used the RF classifier, but for the second group, they have used a retrieval based method with their proposed similarity metric. The reason of proposing this separation is the unbalanced dataset.

Second method (run 3) is a retrieval-based method. Basically, they have encoded the 2D image extracted from the central slice of the lesion by applying 1D Log-Gabor filter, and then break the output of the filter into small blocks and quantize the dominant angular direction of each block to four levels by using Daugman method. Afterward, the Hamming distance has been employed as the similarity metric to retrieve the five most similar images to the test image. Finally, for each UsE feature, they have used majority voting between retrieved images.

Among 73 UsE features, 7 of them were excluded from the evaluation because of their unbounded labels (numeric continuous values). Table 1, compares the results of three submitted runs in terms of completeness, accuracy and total scores.

Table 2 compares the results of different runs in predicting different groups of UsE features. We divide UsE features into 5 groups: liver, vessels and three

Table 1: Results of the runs of Liver CT annotation task 2015.

Group name	Run	Completeness	Accuracy	Total Score	method used
CREDOM	run1	0.99	0.825	0.904	RDF-feature1
CREDOM	run2	0.99	0.822	0.902	RDF-feature2
CREDOM	run3	0.99	0.836	0.910	Image Retrieval

lesion groups with area, lesion and component concepts. Results show that all runs have completed UsE features of the liver and vessel with high accuracy. None of the runs can completely annotate the component-related concepts of lesions. Lesion-related concepts of lesion are fully completed, while showing a very low accuracy. Results show that the third run (retrieval based method), outperforms the other runs.

Table 2: Completeness(complete.) and Accuracy(acc.) for five different groups of UsE features

Group name	Run	Liver		Vessel		LesionArea		LesionLesion		LesionComponent	
name	run	complete.	acc.	complete.	acc.	complete.	acc.	complete.	acc.	complete.	acc.
CREDOM	run1	1.00	0.925	1.00	1.00	1.00	0.730	1.00	0.47	0.96	0.87
CREDOM	run2	1.00	0.925	1.00	1.00	1.00	0.746	1.00	0.47	0.96	0.84
CREDOM	run3	1.00	0.925	1.00	1.00	1.00	0.753	1.00	0.48	0.96	0.89

Table 3 compares the results of both liver CT annotation task 2014 and 2015 participants. As can be seen, results indicate that the BMET group from 2014 has the best performance in this task.

Table 3: Results of the runs of Liver CT annotation task 2014 and 2015.

Group name	Run	Completeness	Accuracy	Total Score	method used
BMET	run5	0.98	0.91	0.947	IR
CASMIP	run1	0.95	0.91	0.93	LDA+KNN
piLabVAVlab	run2	0.51	0.89	0.677	MF-EUC
CREDOM	run3	0.99	0.836	0.910	IR

4 Conclusion

This was the second time that the liver CT annotation task was organized. We provided liver patient data collected via a hybrid patient information entry system, whose liver characteristics are based on the LICO ontology. The challenge was to predict UsE features of patient records, given in RDF format. This year

in 2015, in contrast to last year's edition, users have not been provided with CoG features. Also, they were free to use any set of image features to perform the task. As this was the first time that UsE features were provided in RDF formats, it was not surprising that few groups were finally able to submit their runs for this complex task. Out of 32 teams, 1 team submitted its runs. The approaches and results were reviewed and documented in this paper. Since the dataset was exactly the same as last year's, comparison of the results of all runs submitted to the liver CT annotation task in both 2014 and 2015 is also provided in this paper. The main challenge of the task was due to the unbalanced dataset and participants tried to overcome this issue with different methods. Among all methods, image retrieval obtained the best performance. It was observed that feature selection is important for the best performance of the prediction method.

Acknowledgments The Liver CT Annotation task is supported by TÜBİTAK Grant # 110E264 (CaReRa project), and in part by COST Action IC1302 (KEYSTONE).

References

1. B. Marvasti, N., Kökciyan, N., Türkay, R., Yazıcı, A., Yolum, P., Üsküdarlı, S., Acar, B.: imageclef liver ct image annotation task 2014
2. B. Spanier, A., Joskowicz, L.: 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)
3. Ermis, B., Cemgil, A.T.: Liver ct annotation via generalized coupled tensor factorization. In: CLEF 2014 Labs and Workshops, Notebook Papers. CEUR Workshop Proceedings (CEUR-WS.org) (September 2014)
4. Kokciyan, N., Turkay, R., Uskudarli, S., Yolum, P., Bakir, B., Acar, B.: Semantic description of liver ct images: An ontological approach (2014)
5. Kumar, A., Dyer, S., Li, C., H. W. Leong, P., Kim, J.: 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)
6. Nedjar, I., Mahmoudi, S., Chikh, A., Abi-yad, K., Bouafia, Z.: Automatic annotation of liver ct image: Imageclefmed 2015. In: CLEF2015 Working Notes. CEUR Workshop Proceedings, CEUR-WS.org, Toulouse, France (September 8-11 2015)
7. Villegas, M., Müller, H., Gilbert, A., Piras, L., Wang, J., Mikolajczyk, K., de Herrera, A.G.S., Bromuri, S., Amin, M.A., Mohammed, M.K., Acar, B., Uskudarli, S., Marvasti, N.B., Aldana, J.F., del Mar Roldán García, M.: General Overview of ImageCLEF at the CLEF 2015 Labs. Lecture Notes in Computer Science, Springer International Publishing (2015)

Group	Concept	Properties	Possible values(assigned indices)
Liver	Liver	Liver Placement	downward displacement(0), normal placement(1), leftward displacement(2), upward displacement(3), other(4)
		Liver Contour	irregular(0), lobulated(1), nodular(2), regular(3), other(4)
		Liver Size Change	decreased(0), increased(1), normal(2), other(3)
		Liver Craniocaudal Dimension(mm)	The amount change in size of liver(mm)
		Density Type	heterogeneous(0), homogeneous(1), other(2)
		Density Change	decreased(0), increased(1), normal(2), other(3)
	Right Lobe	Right Lobe Craniocaudal Dimension(mm)	The amount change in size of right lobe(mm)
		Right Lobe Size Change	decreased(0), increased(1), normal(2), other(3)
	Left Lobe	Left Lobe Craniocaudal Dimension(mm)	The amount change in size of left lobe(mm)
		Left Lobe Size Change	decreased(0), increased(1), normal(2), other(3)
	Caudate Lobe	Caudate Lobe Craniocaudal Dimension(mm)	The amount change in size of caudate lobe(mm)
		Caudate Lobe Size Change	decreased(0), increased(1), normal(2), other(3)
Vessel	Hepatic Artery	Hepatic Artery Lumen Diameter	decreased(0), increased(1), normal(2), other(3)
		Hepatic Artery Lumen Type	obliterated(0), open(1), partially obliterated(2), other(3)
	Hepatic Portal Vein	Hepatic Portal V. Lumen Diam.	decreased(0), increased(1), normal(2), other(3)
		Hepatic Portal V. Lumen Type	obliterated(0), open(1), partially obliterated(2), other(3)
		is Cavernous Transformation Observed?(Hepatic Portal Vein)	NA(-1),True(1),False(0),NA(-1)
	Left Portal Vein	Left Portal V. Lumen Diam.	decreased(0), increased(1), normal(2), other(3)
		Left Portal V. Lumen Type	obliterated(0), open(1), partially obliterated(2), other(3)
		is Cavernous Transformation Observed?(Left Portal Vein)	NA(-1),True(1),False(0),NA(-1)
	Right Portal Vein	Right Portal V. Lumen Diam.	decreased(0), increased(1), normal(2), other(3)
		Right Portal V. Lumen Type	obliterated(0), open(1), partially obliterated(2), other(3)
		is Cavernous Transformation Observed?(Right Portal Vein)	NA(-1),True(1),False(0),NA(-1)
	Hepatic Vein	Hepatic V. Lumen Diam.	decreased(0), increased(1), normal(2), other(3)
		Hepatic V. Lumen Type	obliterated(0), open(1), partially obliterated(2), other(3)

Table 4: List of Use features

Group	Concept	Properties	Possible values(assigned indices)
Vessel	Left Hepatic Vein	Left Hepatic V. Lumen Diam.	decreased(0), increased(1), normal(2), other(3)
		Left Hepatic V. Lumen Type	obliterated(0), open(1), partially obliterated(2), other(3)
	Middle Hepatic Vein	Middle Hepatic V. Lumen Diam.	decreased(0), increased(1), normal(2), other(3)
		Middle Hepatic V. Lumen Type	obliterated(0), open(1), partially obliterated(2), other(3)
	Right Hepatic Vein	Right Hepatic V. Lumen Diam.	decreased(0), increased(1), normal(2), other(3)
		Right Hepatic V. Lumen Type	obliterated(0), open(1), partially obliterated(2), other(3)
General	Patient	Diagnosis	Diagnosis of given image using ICD10 codes (bar separated) and in the free text MD's comments are written (bar separated).
Lesion	Lesion	Cluster Size	1(1), 2(2), 3(3), 4(4), 5(5), multiple(6) For simple cases this value shows number of lesions inside the ROI, but in case of having more than one lesions of a certain type, the biggest lesion is annotated as a sample of that cluster and number of lesions with same properties is written here
		Contrast Uptake	NA(-1), dense(0), heterogeneous(1), homogeneous(2), minimal(3), moderate(4), other(5)
		Contrast Pattern	NA(-1), central(0), early uptake then wash out(1), fixing contrast in late phase(2), heterogeneous(3), homogeneous(4), peripheric(5), peripheric nodular(6), spokes wheel(7), undecided(8), other(9)
		Lesion Composition	SolidCysticMix(0), Solid(1), SolidWithCystic(2), PureSolid(3), PredominantSolid(4), Cystic(5), PureCystic(6), PredominantCystic(7), CysticWithSolidComponent(8), CysticWithDebris(9), Abcess(10)
		is Leveling Observed?	True(1),False(0)
		Leveling Type	NA(-1), fluid fluid(0), fluid gas(1), fluid solid(2), gas solid(3), other(4)
		is Debris observed?	True(1),False(0),NA(-1)
		Debris Location	NA(-1), floating inside(0), located on dependent position(1),other(2)
		is Close to Vein	NA(-1), HepaticArtery(0), HepaticPortalVein(1), RightPortalVein(2), LeftPortalVein(3), HepaticVein(4), RightHepaticVein(5), MiddleHepaticVein(6), LeftHepaticVein(7), VenacavaInferior(8), PosteriorBranchOfRightPortalVein(9), AnteriorBranchOfRightPortalVein(10), other(11)
		Vasculature Proximity	NA(-1), adjacent(0), adjunct to contact(1), bended(2), circumscribed(3), invaded(4), other(5)
	Area	Lobe	LeftLobe(0), CaudateLobe(1), RightLobe(2)
		Segment	SegmentI(1), SegmentII(2), SegmentIII(3), SegmentIV(4), SegmentV(5), SegmentVI(6), SegmentVII(7), SegmentVIII(8)
		width	a number in mm which represents width of the lesion
		height	a number in mm which represents heigth of the lesion
		is Gallbladder Adjacent?	True(1),False(0)
		is Peripheral Localized?	True(1),False(0)
is Subcapsular Localized?	True(1),False(0)		
is Central Localized	True(1),False(0)		

Group	Concept	Properties	Possible values(assigned indices)
	Area	Margin Type	geographical(0), ill defined(1), irregular(2), lobular(3), serpiginous(4), spiculative(5), well defined(6), other(7)
		Shape	band(0), fusiform(1), irregular(2), linear(3), nodular(4), ovoid(5), round(6), serpiginous(7), other(8)
		is Contrasted	True(1),False(0),NA(-1)
		is Calcified? (Area)	True(1),False(0),NA(-1)
		Area Calcification Type	NA(-1), coarse(0), focal(1), millimetric-fine(2), punctate(3), scattered(4), other(5)
		Density	NA(-1), hyperdense(0), hypodense(1), isodense(2), other(3)
		Density Type	NA(-1), heterogeneous(0), homogeneous(1), other(2)
	Capsule	is Calcified? (Capsule)	True(1),False(0),NA(-1)
		Capsule Calcification Type	NA(-1), coarse(0), focal(1), millimetric-fine(2), punctate(3), scattered(4), other(5)
	Polyp	is Calcified? (Polyp)	True(1),False(0),NA(-1)
		Polyp Calcification Type	NA(-1), coarse(0), focal(1), millimetric-fine(2), punctate(3), scattered(4), other(5)
	Pseudocapsule	is Calcified?(Pseudocapsule)	True(1),False(0),NA(-1)
		Pseudocapsule Calc. Type	NA(-1), coarse(0), focal(1), millimetric-fine(2), punctate(3), scattered(4), other(5)
	Septa	is Calcified? (Septa)	True(1),False(0),NA(-1)
		Septa Calcification Type	NA(-1), coarse(0), focal(1), millimetric-fine(2), punctate(3), scattered(4), other(5)
		Diameter Type	NA(-1), complete(0), incomplete(1), other(2)
		Thickness	NA(-1), thick(0), thin(1), other(2)
	Solid Component	is Calcified? (Solid Component)	True(1),False(0),NA(-1)
		Solid Component Calcification Type	NA(-1), coarse(0), focal(1), millimetric-fine(2), punctate(3), scattered(4), other(5)
	Wall	is Calcified? (Wall)	True(1),False(0),NA(-1)
		Wall Calcification Type	NA(-1), coarse(0), focal(1), millimetric-fine(2), punctate(3), scattered(4), other(5)
		Wall Type	NA(-1), thick(0), thin(1), other(2)
		is Contrasted?(Wall)	True(1),False(0),NA(-1)