Overview of the ImageCLEF 2015 medical clustering task

M. Ashraful Amin[1] and Mahmood Kazi Mohammed[1, 2]

[1]Computer Vision and Cybernetics Group, CSE, Independent University, Bangladesh
[2]Sir Salimullah Medical College, Dhaka, Bangladesh
aminmdashraful@iub.edu.bd, mkmohammed86@gmail.com

Abstract. There are thousands of unlabeled x-ray images available and in the third world countries more is generated every day. With the advancement of digital technology now a day's digital x-ray imaging techniques are available, however due to the high cost of the machines it is not popular in the third world countries. Moreover, old school x-ray plates are still there. The medical clustering task of ImageCLEF 2015 addresses the issue of automated organization of x-ray images. The challenge is that there are x-ray images containing different parts of human body and the participant have to device a mechanism to identify that body part. The main challenge is that an image could contain several body part and the classifier has to identify all of them separately or as many as possible. Body parts are divided in to four major larger groups: head-neck, upper-limb, body, and lower-limb. The secondary goal of this task is farther partitioning the initial clusters into sub-clusters, for example the upper-limb cluster can be farther divided into: Clavicle, Scapula, Humerus, Radius, Ulna, and Hand. However, due to the time constrain and difficulty level of the task this year we decided to go with the primary objective. Data was collected by 71 groups from all around the world, however 8 groups submitted the final test results and working note papers were submitted by 6. Interestingly, this 6 groups explored the discriminating ability of 27 different types of feature extraction method and also many different types of classifiers were used. Three different performance measurement is used. Best result for exact match was 0.752; for any match was 0.864; and for Hamming similarity was 0.895.

Keywords: medical imaging, x-ray, information organization, feature extraction, classifiers, convolution neural network.

1 Introduction

Automatic identification of body parts in x-ray image has many application. Organizing and retrieving or searching x-rays with specific body part with or without anomaly from a large database, automated diagnostic system assistance, creating education tools for medical students. We are trying to develop a diagnostic imaging teaching and learning system for medical students of Bangladesh [2]. Thus we collected a large digital x-ray image data set from a local hospital. We are using this data set to build our teaching and learning system, however during this development process automatically archiving and retrieving x-ray image from the large database seemed to be a challenging task. Thus we decided to seek help through ImageCLEF.

CLEF* is a competent and very useful platform to share and seek support for various information collection, organization and retrieval issues. Especially from the multidisciplinary and multiplatform point of view. ImageCLEF [1] is a major part within CLEF and mainly focuses the issue of different types of image data collection, organization or archiving, and retrieval.

The primary objective of this task (as part of ImageCLEF) is to group digital x-ray images into four major clusters: head-neck, upper-limb, body, and lower-limb. The secondary goal of this task is farther partitioning the initial clusters into sub-clusters, for example the upper-limb cluster can be farther divided into: Clavicle, Scapula, Humerus, Radius, Ulna, and Hand.

Our x-ray image clustering task is running for almost a year. Participant's registration started on 1st November 2014 and training dataset was released on 19th November 2014. Five months' time was given to develop the systems based on the training dataset. Then, test dataset was released on 18th April 2015 and they were given a month (till 18th May 2015) to submit the predictions of their developed systems on the test dataset. On 19th May 2015 task organizers published the result of all the submission. The participants were then asked to submit a working note paper for each group describing the approach taken to solve the problem of clustering x-ray images into four groups by 7th June 2015. On 30th June 2015 the participants were given feedback on their submissions and final camera ready submission was set to be 15th July 2015. The ImageCLEF conference as part of the CLEF conference was organized during 8th to 11th October 2015.

In rest of the paper first we provide a description on the dataset, then we provide a description on the participants, then we discuss the features and classification techniques used by the participants and final we discuss the results on the test dataset.

2 The Data

During modernization a hospital authority in Dhaka, Bangladesh acquired a high resolution digital x-ray machine in 2011. We convinced hospital authority to provide us some x-ray image so that we can work on development of a diagnostic imaging teaching and learning system. We collected data for about a year and acquired about 5000 digital x-ray image taken by that newly acquired x-ray machine.

The dataset contains images of various parts of human body. That means, some image can contain just the figures, others can contain the palm, hand, or entire arm including shoulder, some contains feet, leg or the entire lower limb and so on. X-ray images of both male and female are present in the dataset. Age of the patients range from 6 months to 72 years and there are images of small children and that shows the skeleton of entire body. Images contain various bone related pathology such as, broken bones, disjoint bones, hairline fractures, some images are also missing vital body

* http://clef2015.clef-initiative.eu/CLEF2015/

parts and also there are images that does not contain any form of pathology. Some images contains foreign parts attached with the bone, this is because of the metallic dentures and fixtures are used to fix broken or disjoint bones. To calibrate the x-ray machine operators took some x-ray image of non-body parts such as keys, mobile phones, pen, etc., so the dataset also contains some x-ray image that is not body part.

The digital x-ray machine takes very high resolution images and saves them as DICOM[†] format with .dcm extension. The header of this file type can hold many information including the classification information of the x-ray image, patient name, ID and etc. The x-ray images are taken with a special resolution of 2136x2136 pixels and with default gray level color depth of 16 bits. However, due to large size and obligation of keeping patient and hospital information anonymous we made smaller size high resolution .jpeg images available for the ImageCLEF 2015 task.

All together there are 500 digital x-ray image in the training dataset, of which 100 from each of the four desired clusters: head-neck, upper-limb, body, and lower-limb, and there are 100 true negative images that are taken by the same digital x-ray camera for calibration purpose. Some example images are given in figure 1.

250 test images are made available to the participants to check the performance of their system. At this moment we could make 750 data available, however, all 5000 image data in high resolution will be made available for non-commercial uses from our research group website[‡] soon after the CLEF 2015 conferences.

Fig.1. Example images from the training data set

3 The participants

71 groups from all 6 continents of the world participated in the initial level and acquired data from ImageCLEF website. In table 1 and 2 we have provided statistics about participants based on region. Though it is primarily a European event, 15 groups from EU, 14 from North America, 6 from Australia and 29 from Asia participated in the initial event. Among all EU countries there were 5 German groups and in Asia, China had 5 groups which was the highest from that region. Finally, participants were given the test data and a month time to submit their results on the test data. Only, 8 groups submitted their final results. There were, 2 submissions from Australia, 2

[†] DICOM meaning digital imaging and communications in medicine

[‡] www.cvcrbd.org

from USA, 1 from each of the countries Republic of Korea, Israel, Egypt, China and none from the EU. One group has withdrawn their runs (submitted results) as their method was semi-automatic. 7 groups submitted 29 runs (table 4) and the best results for each group is selected and provided in table 5. Finally, 6 groups were able to submit working note papers describing methods used to implement their x-ray clustering system.

Table 1. Participants statistics based on the country of residence

Country	Number of Groups
Australia	6
Bangladesh	3
Bosnia and Herzegovina	1
Canada	1
China	5
Colombia	3
Egypt	2
France	1
Germany	5
India	5
Iran	1
Ireland	2
Israel	3
Japan	2
Korea	2
Macedonia	1
Malaysia	1
Mexico	1
Morocco	1
Pakistan	1
Saudi Arabia	1
Singapore	2
Spain	2
Switzerland	3
Thailand	1
Tunisia	1
Turkey	1
United States	13
Total	71

Table 2. Participant statistics based on the region

Region	Number of Groups
North America	14
South America	4
Europe	15
Africa	3
Asia	29
Australia	6
Total	71

4 Features Used by Participants

To solve this multiclass classification problem of grouping digital x-ray image into four clusters, participants have taken different approaches. For feature extraction they utilized: Intensity Histogram (IH), Gradient Magnitude Histogram (GM), Shape Descriptor Histogram (SD), Curvature Descriptor Histogram (CD), Histogram of Oriented Gradient (HOG), Local Binary Pattern (LBP), Color Layout Descriptor (CLD), Edge Histogram Descriptor (EHD) from MPEG-7 standard, Color and Edge Direction Descriptor (CEDD), Fuzzy Color and Texture Histogram (FCTH), Tamura texture descriptor, Gabor texture feature, primitive length texture features, edge frequency texture features, autocorrelation texture features, Bag of Visual Words (BoVW), Scale invariant feature transform (SIFT), Speeded up robust features (SURF), Binary robust independent elementary features Brief (BRIEF), Oriented fast and rotated BRIEF (ORB), Multi-scale LBP Histogram with Spatial Pyramid, Sparse Coding with Max-pooling and Spatial Pyramid, Fisher Kernel Feature Coding, Global mean of rows and columns, Local Mean of rows and columns, and Gray Level Co-occurrence Matrix (GLCM).

5 Classifiers used by Participants

Classification is performed using Backpropagation Neural Networks (BPNN), Logistic Regression (LR), K Nearest Neighbors (KNN), Deep Belief Network (DBN), Convolution Neural networks (CNN), Decision Tree, Support Vector Machine (RBF Kernel, Poly kernel, Normalized Ploy kernel and Puk kernel), Random Forest, Logistic Model Tree (LMT), Naive Bayesian, and Ensemble Neural Network.

6 Performance Measure & Results

Each x-ray image can be classified as member of either of the four major groups: head-neck, upper-limb, body, and lower-limb. However, it might happen that a single image is classified as more than one class or classified as none of the classes. So we decided that the output for an input x-ray image is a 4-bit, bit string. Table 3 shows some sample input and output for reference.

Because one input can belong to multiple classes, we have tested the performance based on three different methods. The most conventional one is the **hamming similarity** calculation. However, a stricter version of classification accuracy checking is also used, that we are calling **exact matching**, which basically checks, for a given input how many of its multiple possible class is correctly identified. We also checked the accuracies using another method we are calling it **any match.** For an input image if the predicted class matches with any of the actual class of that image then it is considered as correct classification. Best result for exact match was 0.752; for any match was 0.864; and for Hamming similarity was 0.896 all produced by a group from IBM Australia. Final score for all seven groups is provided in table 4 and table 5.

Table 3. Sample input and Output

Input	Output			
	body	Head-Neck	U-Limb	L-Limb
	1	1	1	0
	0	0	0	1
	1	0	0	1
	0	0	0	0

Table 4. Final Results of the Digital X-Ray Image clustering task

Group Name	Country	Exact Match	Any match	Hamming Similarity
IBM MMAFL [3]	Australia	**0.752**	**0.864**	0.863
SNUMedInfo [4]	Korea	0.709	0.856	**0.895**
AmrZEGY [5]	Egypt	0.646	0.780	0.868
NLM [6]	USA	0.613	0.740	0.849
CASMIP [7]	Israel	0.606	0.732	0.843
BMET [8]	Australia	0.497	0.596	0.816
db Lab	China	0.219	0.264	0.664

Table 5. Output score of all 29 runs submitted by 7 participant groups

Group Name	Exact Match	Any match	Hamming Similarity
IBM MMAFL	**0.752**	**0.864**	0.863
SNUMedInfo	0.709	0.856	0.895
SNUMedInfo	0.699	0.844	0.890
IBM MMAFL	0.695	0.840	0.889
IBM MMAFL	0.692	0.832	**0.896**
IBM MMAFL	0.692	0.732	0.755
IBM MMAFL	0.689	0.820	0.890
SNUMedInfo	0.679	0.820	0.879
IBM MMAFL	0.672	0.812	0.874
AmrZEGY	0.646	0.780	0.868
NLM	0.613	0.740	0.849
CASMIP	0.606	0.732	0.843
CASMIP	0.603	0.728	0.847
IBM MMAFL	0.603	0.708	0.778
CASMIP	0.599	0.724	0.847
IBM MMAFL	0.599	0.724	0.838
NLM	0.593	0.716	0.842
CASMIP	0.589	0.712	0.842
CASMIP	0.589	0.712	0.842
CASMIP	0.589	0.712	0.842
CASMIP	0.573	0.692	0.833
CASMIP	0.573	0.692	0.833
CASMIP	0.563	0.680	0.830
CASMIP	0.550	0.664	0.826
NLM	0.543	0.656	0.810
IBM MMAFL	0.510	0.616	0.849
BMET	0.497	0.596	0.816
IBM MMAFL	0.470	0.568	0.835
db Lab	0.219	0.264	0.664

7 Observations & Conclusion

It is very likely that participants will use similar feature extraction and classification techniques. It is accepted that some features will be used by most of the participants, those are the so-called state of the art techniques. However, for this problem of clustering x-ray images into 4 clusters 6 participants have employed 27 different image feature extraction techniques. Different characteristics of the feature extractors are revealed. One interesting observation is that while exploring the famous HoG features one group claims it has poor discriminating capacity [6] on the other hand another group [7] is providing an accuracy above 80% using HoG feature and different classifiers. Another interesting observation is that, even though, x-ray images are gray, color features like CEDD, FCTH shows quite good discriminating ability. Most interesting yet obvious observation is the use of Convolution Neural Network (CNN). Recently, CNN is made popular by GoogLeNet. Out of six, 5 groups [4 - 8] used or experimented with Neural Networks. It is good news for the neural network researchers. We believe people have already started (rather restarted) to explore enormous ability of CNN and other computational learners other than SVM's.

Acknowledgements

The medical clustering task at the 2015 ImageCLEF is supported by the Independent University Bangladesh and European Science Foundation's (ESF) Research Networking Programmes (RNPs).

References

1. Villegas, M., Muller, 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 Roldan Garcia, M.: General Overview of ImageCLEF at CLEF2015 Labs. Lecture Notes in Computer Science, Springer International Publishing (2015)
2. Faruque, M. S. S., Banik, S., Mohammed, M. K., Hasan, M., Amin, M. A.: Teaching & Learning System for Diagnostic Imaging; Phase I: X-Ray Image Analysis & Retrieval. In Proceedings of the 6th International Conference on Computer Supported Education (2015).
3. Sedai, S., Liang, X., Abedini, M., Chen, Q., Chakravorty, R., Garnavi, R.: IBM Research - Australia at Image CLEF 2015: Medical Clustering Task, CEUR Workshop Proceedings, CEURWS.org, Toulouse, France (September 8-11 2015).
4. Choi, S.: SNUMedinfo at ImageCLEF 2015 Medical clustering task, CEUR Workshop Proceedings, CEURWS.org, Toulouse, France (September 8-11 2015).
5. Zeiadan, I.E., Zamel, A.A., Al Zohairy, A.: Clustering of medical X-ray images by merging outputs of different classification techniques, CEUR Workshop Proceedings, CEURWS.org, Toulouse, France (September 8-11 2015).
6. Vajda, S., Antani, S., Thoma, G.: National Library of Medicine (NLM) at ImageCLEF2015: Medical Clustering Task, CEUR Workshop Proceedings, CEURWS.org, Toulouse, France (September 8-11 2015).
7. Aboud, M., Spanier, A. B., Joskowicz, L.: ImageCLEF-2015 Clustering of body part x-ray Report, CEUR Workshop Proceedings, CEURWS.org, Toulouse, France (September 8-11 2015).
8. Lyndon, D., Kumar, A., Kim, J., Leong, P. h. W., Feng, D.: Convolutional Neural Networks for Medical Clustering, CEUR Workshop Proceedings, CEURWS.org, Toulouse, France (September 8-11 2015).