IBM Research Report

Semantic Model Vector Training Data for ImageCLEF2013

Noel Codella, Michele Merler
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598
USA

IBM Research Report: Semantic Model Vector Training Data for ImageCLEF2013

Noel Codella[1] and Michele Merler[1]

IBM T.J. Watson Research Center,
1101 Kitchawan Rd., Yorktown Heights, NY

Abstract. In this research report, we describe the dataset used to train the semantic model vector employed for the ImageCLEF 2013 medical modality visual classification task.

1 Introduction

The medical semantic model vector used for the ImageCLEF 2013 medical modality recognition task [1, 2] is trained from about 26,000 images collected from multiple sources, organized into a hierarchical taxonomy structure of 120 categories of various radiological categories. The images used to train our semantic model vector were acquired from web crawled data, the Image Retrieval in Medical Applications (IRMA) 2009 dataset [3], The Cancer Imaging Archives (TCIA) [4, 5], Cornell University SIMBA CT dataset [6], and the Japanese Society of Radiological Technology (JSRT) [7]. Web crawl was performed using the Microsoft Bing search engine, with queries as the semantic concept label (abbreviations expanded). The top 500 results were saved, and cleaned by a human labeler.

Each dimension in the model vector refers to a subcategory over 7 modality domains of CT, X-Ray, ECG, MRI, PET, Ultrasound, and Slide Microscopy. The rest of the dimensions represent subcategories covering body regions (such as chest, arm, leg, head, brain, neck, etc.), views (such as anterio-posterior, coronal, saggital), and disease states (such as pneumothorax) under each modality.

2 Categories & Number of Examples

This section gives a full accounting of the name of each semantic classifier in the semantic model vector, along with the number of positive examples used for training (in parentheses).

CT (10055)
CT_BRAIN (131)
CT_CHEST (9924)
CT_CHEST_LUNGCANCER (92)
CT_CHEST_NORMAL (9690)
CT_CHEST_PNEUMOTHORAX (142)
DX (12995)
DX_APPENDAGE (4650)

DX_APPENDAGE_ARM (2422)
DX_APPENDAGE_ARM_ELBOW (371)
DX_APPENDAGE_ARM_ELBOW_AP (151)
DX_APPENDAGE_ARM_ELBOW_LAT (220)
DX_APPENDAGE_ARM_FOREARM (157)
DX_APPENDAGE_ARM_FOREARM_AP (75)
DX_APPENDAGE_ARM_FOREARM_LAT (82)
DX_APPENDAGE_ARM_HAND (1188)
DX_APPENDAGE_ARM_HAND_FINGER (158)
DX_APPENDAGE_ARM_HAND_WHOLE (798)
DX_APPENDAGE_ARM_HAND_WHOLE_AP (688)
DX_APPENDAGE_ARM_HAND_WHOLE_OBL (110)
DX_APPENDAGE_ARM_HAND_WRIST (232)
DX_APPENDAGE_ARM_HAND_WRIST_AP (114)
DX_APPENDAGE_ARM_HAND_WRIST_LAT (118)
DX_APPENDAGE_ARM_SHOULDER_32 (632)
DX_APPENDAGE_ARM_UPPER (74)
DX_APPENDAGE_LEG (2228)
DX_APPENDAGE_LEG_ANKLE (345)
DX_APPENDAGE_LEG_ANKLE_AP (177)
DX_APPENDAGE_LEG_ANKLE_LAT (168)
DX_APPENDAGE_LEG_FOOT (672)
DX_APPENDAGE_LEG_FOOT_AP (408)
DX_APPENDAGE_LEG_FOOT_LATERAL (122)
DX_APPENDAGE_LEG_FOOT_OBLIQUE (142)
DX_APPENDAGE_LEG_KNEE (829)
DX_APPENDAGE_LEG_KNEE_PATELLA (103)
DX_APPENDAGE_LEG_KNEE_WHOLE (726)
DX_APPENDAGE_LEG_KNEE_WHOLE_AP (442)
DX_APPENDAGE_LEG_KNEE_WHOLE_LAT (284)
DX_APPENDAGE_LEG_LOWER (191)
DX_APPENDAGE_LEG_LOWER_AP (125)
DX_APPENDAGE_LEG_LOWER_LAT (66)
DX_APPENDAGE_LEG_UPPER (191)
DX_APPENDAGE_LEG_UPPER_AP (133)
DX_APPENDAGE_LEG_UPPER_LAT (58)
DX_CRANIUM (1070)
DX_CRANIUM_NOSE (314)
DX_CRANIUM_NOSE_LAT (50)
DX_CRANIUM_NOSE_OCCIPITOFRONTAL (264)
DX_CRANIUM_WHOLE (756)
DX_CRANIUM_WHOLE_AP (358)
DX_CRANIUM_WHOLE_FRONTOCCIPITAL (38)
DX_CRANIUM_WHOLE_LAT (360)
DX_HIP_JOINT_AP_ARTIFICIAL (84)

DX_HIP_JOINT_AP_NATURAL (127)
DX_HIP_JOINT_LAT_ARTIFICIAL (21)
DX_HIP_JOINT_LAT_NATURAL (53)
DX_TORSO (7275)
DX_TORSO_ABDOMEN (379)
DX_TORSO_ABDOMEN_AP (204)
DX_TORSO_ABDOMEN_BARIUMSWALLOW (143)
DX_TORSO_ABDOMEN_UPPER (32)
DX_TORSO_BREAST (262)
DX_TORSO_BREAST_LEFT (128)
DX_TORSO_BREAST_LEFT_AXIAL (63)
DX_TORSO_BREAST_LEFT_OBL (65)
DX_TORSO_BREAST_RIGHT (134)
DX_TORSO_BREAST_RIGHT_AXIAL (65)
DX_TORSO_BREAST_RIGHT_OBL (69)
DX_TORSO_CHEST (4583)
DX_TORSO_CHEST_FULL (4583)
DX_TORSO_CHEST_FULL_AP (3728)
DX_TORSO_CHEST_FULL_AP_LUNGCANCER (183)
DX_TORSO_CHEST_FULL_AP_NORMAL (302)
DX_TORSO_CHEST_FULL_AP_PNEUMONIA (204)
DX_TORSO_CHEST_FULL_AP_PNEUMOTHORAX (164)
DX_TORSO_CHEST_FULL_AP_UNKNOWN (2875)
DX_TORSO_CHEST_FULL_LAT (855)
DX_TORSO_CHEST_FULL_LAT_LUNGCANCER (1)
DX_TORSO_CHEST_FULL_LAT_NORMAL (18)
DX_TORSO_CHEST_FULL_LAT_UNKNOWN (836)
DX_TORSO_HIP (643)
DX_TORSO_HIP_JOINT (285)
DX_TORSO_HIP_JOINT_AP (211)
DX_TORSO_HIP_JOINT_LAT (74)
DX_TORSO_HIP_PELVIS (358)
DX_TORSO_HIP_PELVIS_ARTIFICIAL (74)
DX_TORSO_HIP_PELVIS_NATURAL (284)
DX_TORSO_SPINE (1408)
DX_TORSO_SPINE_CERVICAL (635)
DX_TORSO_SPINE_CERVICAL_AP (230)
DX_TORSO_SPINE_CERVICAL_LAT (405)
DX_TORSO_SPINE_LUMBAR (536)
DX_TORSO_SPINE_LUMBAR_AP (288)
DX_TORSO_SPINE_LUMBAR_LAT (248)
DX_TORSO_SPINE_THORACIC (237)
DX_TORSO_SPINE_THORACIC_AP (117)
DX_TORSO_SPINE_THORACIC_LAT (120)
ECG (904)

ECG_FIBRILLATION (566)
ECG_NORMAL (338)
MR (1042)
MR_APPENDAGE (347)
MR_BRAIN (414)
MR_BRAIN_AXIAL (259)
MR_BRAIN_AXIAL_COLLAGE (103)
MR_BRAIN_AXIAL_SINGLE (156)
MR_BRAIN_SAGITTAL (155)
MR_HIP (100)
MR_KNEE (347)
MR_SPINE (181)
PET (245)
PET_BW (134)
PET_COLOR (111)
SM (334)
SM_BLOODSMEAR (334)
SM_BLOODSMEAR_NORMAL (288)
SM_BLOODSMEAR_SICKLECELL (46)
US (400)
US_CARDIAC (190)
US_FETUS (210)

3 Abbreviations

AP: Anterior-posterior view
BW: Grayscale color spectrum
CT: Computed Tomography
DX: Digital X-Ray
ECG: Electrocardiogram
LAT: Lateral view
MR: Magnetic Resonance
OBL: Oblique
PET: Positron Emmission Tomography
SM: Slide Microscopy
US: Ultrasound

References

1. Cao, L., Codella, N., Connell, J., Merler, M., Nguyen, Q., Pankanti, S., Smith, J., "IBM Multimedia Analytics @ ImageCLEF2013" ImageCLEF 2013 Medical Image Retrieval Workshop
2. Seco de Herrera, A.G., Kalpathy-Cramer, J., Fushman, D.D., Antani, S., Muller, H., "Overview of the ImageCLEF 2013 medical tasks", ImageCLEF 2013 Medical Image Retrieval Workshop,
3. Image Retrieval in Medical Applications http://ganymed.imib.rwth-aachen.de/irma/index_en.php
4. The Cancer Imaging Archive (TCIA) http://cancerimagingarchive.net/

5. Armato SG III, McLennan G, Bidaut L, McNitt-Gray MF, Meyer CR, Reeves AP, Zhao B, Aberle DR, Henschke CI, Hoffman EA, Kazerooni EA, MacMahon H, van Beek EJR, Yankelevitz D, et al. "The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A completed reference database of lung nodules on CT scans." Medical Physics, 38: 915–931, 2011.
6. Cornell University Vision and Analysis Group SIMBA Dataset http://www.via.cornell.edu/visionx/simba/
7. Japanese Society of Radiological Technology http://www.jsrt.or.jp/jsrt-db/eng.php