# Detecting Figure-panel Labels in Medical Journal Articles using MRF

Daekeun You[1], Sameer Antani[2], Dina Demner-Fushman[2], Venu Govindaraju[1], George R. Thoma[2]

[1]Center for Unified Biometrics and Sensors
Dept. of Computer Science and Engineering
SUNY at Buffalo, Buffalo, NY 14260, USA
{dyou, govind}@buffalo.edu

[2]National Library of Medicine/NIH
Bethesda, MD 20894, USA
{santani, ddemner, gthoma}@mail.nih.gov

***Abstract*—We present a method for figure-panel (subfigure) label detection and recognition in multi-panel figures extracted from biomedical articles. Figures in biomedical articles often comprise several subfigures that are identified by superimposed panel labels ('A', 'B', …) which are referenced in the figure caption and discussion in the article body. Splitting such multi-panel figures into individual subfigures is a necessary step for improved multimodal biomedical information retrieval. Prior to feature extraction for indexing and retrieval of biomedical figures it is necessary to classify image content in each subfigure by its modality (X-ray, MRI, CT, etc.) and other relevant criteria. Subfigure labels are valuable in associating individual panels with relevant text in captions and discussion. We propose a 4-step panel label detection method based on Markov Random Field (MRF). Experiments on 515 multi-panel figures and analysis of the results show promising results. We present the successes and identify critical challenges.**

***Keywords- image-text detection; Markov Random Field; belief propagation; OCR; Neural network; image binarization; CBIR; image classification***

## I. INTRODUCTION

Multi-panel figures that contain several subfigures each identified by a panel label ('A', 'B', etc.) are frequently found in biomedical articles. In order to extract features from relevant images to classify them according to modality (X-ray, CT, MRI, gels, etc.), associate image content with relevant biomedical concepts extracted from the figure caption text and relevant discussion in the article text, and implement multi-modal (text + image) CBIR techniques, it is necessary to first separate them. Our goal is to automatically detect panel labels and to provide the location and associated text of panels to assist in this process.

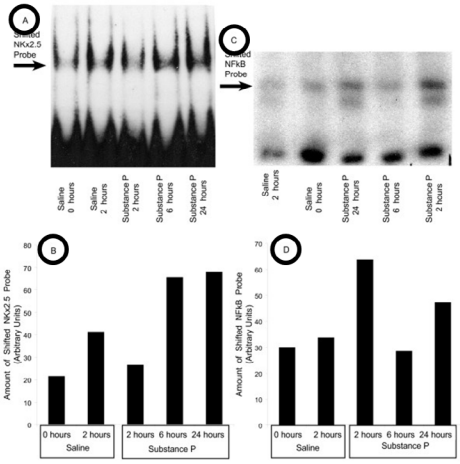



Figure 1. Example of multi-panel figure (panel labels shown in circles)

The need to separate subfigures is evident from the example shown in Figure 1, which is an illustration extracted from a biomedical article [1], where the subfigure panels A and C are images showing signal responses of a substance to saline for varying duration, while subfigure panels B and D show these responses as bar charts. Though the image pairs (A,C and B,D) look alike, they are clearly different, and image features extracted to represent their content must be computed separately. Further, to support efficient and accurate image retrieval, information about the subfigure content (response to saline, and "substance P") that is discussed in the figure caption and discussion in the original article text must be associated with relevant subfigure panels. As shown in the example, the author often places related images from different modalities as different subfigures, which are combined into a single image in the publication process. Meaningful multi-modal image retrieval is only possible if the visual content expressed in the image is unimodal (e.g., all CT, MRI, or X-ray images).

There has been prior effort to split multi-panel figures into subfigures based on structural image features extracted from panel boundaries using little textual information extracted from text captions [2]. Panel label detection could assist in improving subfigure splitting and image features used in panel splitting could result in increased accuracy in panel label detection. Combining the two approaches and utilizing textual features may be the best solution to the panel splitting problem.

In this paper, we propose an efficient panel label detection method based on character recognition (OCR) and Markov Random Field (MRF) theory. Candidate panel label characters are segmented from background and recognized. Then the labels are relabeled by MRF method. Section II explains our proposed methods, and experiment results and analysis of the results are presented in section III. Section IV discusses conclusions and future work.

## II. METHOD

Our panel label detection approach consists of four steps: i) preprocessing, ii) character recognition (OCR), iii) classification of OCR results into two classes (panel label or

1520-5363/11 $26.00 © 2011 IEEE
DOI 10.1109/ICDAR.2011.196

noise), and iv) post-processing. In this paper *panel label* denotes overlay characters identifying subpanels in multi-panel figures and *text label* denotes recognized text label of any extracted connected component (CC).

### A. Preprocessing

In the preprocessing step, candidate panel label characters are segmented from background. Several basic segmentation techniques such as image binarization or edge-based contour detection [3] could be applied to extract character CCs. In the multi-panel figures in our data set the color of panel labels is usually black or white and hence thresholding by one global threshold is sufficient to segment overlay characters from background. Two fixed threshold values, 50 and 200, are used to extract black and white characters, respectively, since the color of panel labels is not known in advance and certain images have both black and white panel labels within the same image.

### B. OCR

The candidate panel label CCs extracted in the preprocessing step are then recognized by an OCR engine. We implemented an alphanumeric OCR engine to recognize overlay characters including panel labels in biomedical images. Contour features described in [4] are used and a neural network (NN) classifier is trained on about 7,500 character images extracted from our image set. The average recognition rate is about 99% on a test set with 66,723 character samples extracted from the same image set.

### C. MRF panel label classification

The outputs obtained from the first two steps are CCs extracted from an input image and their text labels. The text labels contain true panel labels and noise as well. In order to detect panel labels from the text labels, several important factors such as structure (e.g., alignment and order) and characteristics (e.g., size) of panel labels need to be considered. Our approach to this problem is to label each text label as panel label or noise, based on features defined from a single text label and pair of neighboring text labels.

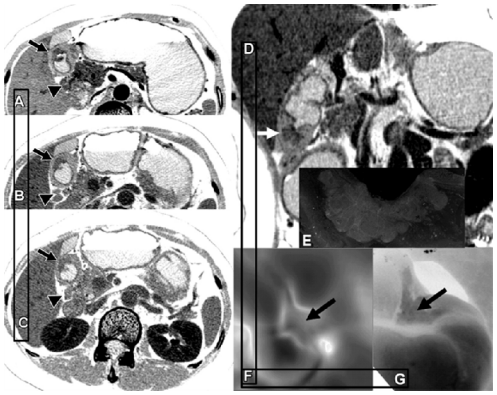


Figure 2. Aligned panel labels (in black rectangle zones).

Markov Random Field (MRF) is suitable for our labeling purpose since MRF defines and minimizes energy functions based on unary and binary relationships between neighbors within a neighborhood system [5-7].

#### 1) Neighborhood system

Unlike other applications [5] of MRF model that define a neighborhood system based on lattice structures (e.g., 4 or 8 neighbors) or *n* closest neighbors, the neighborhood system in our model is defined based on horizontal and vertical panel label location relationship among a text label *i* and other labels. One important assumption in our definition of a neighborhood system is based on our observation that

*Every panel label is aligned with at least one other panel label in the same horizontal or vertical zone.*

Fig. 2 illustrates the assumption. Panel labels 'A', 'B', and 'C' and panel labels 'D' and 'F' are aligned vertically as shown in black boxes. Panel labels 'F' and 'G' are aligned horizontally. Panel label 'E', however, has no other panel labels aligned in the same zone. We call images containing any unaligned labels like 'E' in the example *abnormal aligned* image.

Though true panel labels are generally aligned, noise text labels may appear anywhere within an image and hence a lattice structure neighborhood system is not suitable for our purpose. Distance based neighborhood system is not available either since distances among true labels differ from image to image and hence it is very difficult to choose an optimal distance to define a neighborhood system. We need to define a new neighborhood system that contains true panel labels and noise labels in the neighborhood $N_i$ of a text label *i*. Eq. 1 defines $N_i$ in our model.

$$N_i = \{i' \in S \mid (c_{i'x} \geq x_{il} \,\&\, c_{i'x} \leq x_{ir}) \, or \, (c_{i'y} \geq y_{it} \,\&\, c_{i'y} \leq y_{ib})\} \quad (1)$$

where $S$ denotes a set of text labels in an image, $c_{i'x}$ denotes $x$ coordinate of the centroid of CC of text label $i'$, and $x_{il}$ and $x_{ir}$ denote left and right bounding box of CC of text label $i$, respectively. The second term related to $y$ coordinates is defined in the same way. Fig. 3 illustrates our neighborhood system. The small rectangle boxes in (b) show bounding box of each CC extracted in the preprocessing step and the characters near the boxes are recognition results of each CC. The first and second terms in the $N_i$ form vertical and horizontal narrow zones as shown in Fig. 3(b) by dashed and solid lines, respectively. Only CCs satisfying Eq. 1 (i.e., locating in the zones) are considered as neighbors of text label *i*. Fig. 3(b) shows neighborhood of panel label 'D'. The neighbor set of panel label 'D' contains two true panel labels 'B' and 'C' and two noise labels 'I' and 'O', as pointed by black arrows.

#### 2) Compatibility functions and Belief propagation(BP)

Unary and binary compatibility functions are defined from single text label *i* and two text labels *i* and *i'*, respectively. Belief propagation (BP) is applied to

iteratively update and obtain optimal labels for each candidate text label. Eq. 2 shows the unary compatibility function.

$$r_i(f_i)=\alpha R_{pl}(i)+\beta C(i) \quad (2)$$

where $R_{pl}$ and $C$ are ratio of the number of actual and expected prior labels of text label $i$ and recognition confidence score provided by the OCR engine, respectively. $\alpha$ and $\beta$ are weights to control the influence of $R_{pl}$ and $C$, respectively. $r_i(f_i)$ represents the evidence of assigned label ($f_i \in \{0,1\}$) of text label $i$. We observe in most cases that panel labels are ordered from left to right or top to bottom starting from label 'A'. Hence, for example, the label 'D' in Fig. 3(b) is expected to have three text labels (i.e., 'A', 'B', and 'C') in the *prior label zone* marked in Fig. 3(c). Smaller $R_{pl}$ means that a text label is probably not a true panel label.

For the binary compatibility function, we define a function as shown in Eq. 3 based on the size of CC's bounding boxes and text labels between two candidates $i$ and $j$ ($\in N_i$)

$$r_{i,j}(f_i,f_j)=\gamma R_a(i,j)+\delta L(i,j) \quad (3)$$

where $R_a$ denotes ratio of areas of the two CCs, $L$ denotes text labels relationship, and $\gamma$ and $\delta$ are weights. Our two observations that *i) the size of two true panel label bounding boxes are almost similar and ii) a panel label that is located left or above of a panel label have smaller ASCII code than the other label* provide enough evidence of the definition. We compare two text labels and assign 1 to the $L(i,j)$ if the second observation is satisfied and 0, otherwise.

BP message update rule is defined as

$$m_{i,j}(f_j)=\max_{f_i}\{r_i(f_i)r_{i,j}(f_i,f_j)\max_{f_k} m_{k,i}(f_i)\} \quad (4)$$

where $m_{i,j}(f_j)$ is message from text label $i$ to $j$, $r_i(f_i)$ and $r_{i,j}(f_i, f_j)$ are compatibility functions defined above, and $k \in N_i - \{j\}$ are the neighbors of $i$ except $j$. Unlike general BP message update rule that multiplies all incoming messages to $i$ for label $f_i$, we picked the maximum message since text labels in $S$ do not have the same number of neighbors in their neighborhood and hence multiplying different number of messages can cause incorrect result. Messages are updated iteratively until there is no label flip. Final belief is computed by Eq. (5)

$$b_i(f_i)=r_i(f_i)\max_{f_{i'}} m_{i',i}(f_i) \quad (5)$$

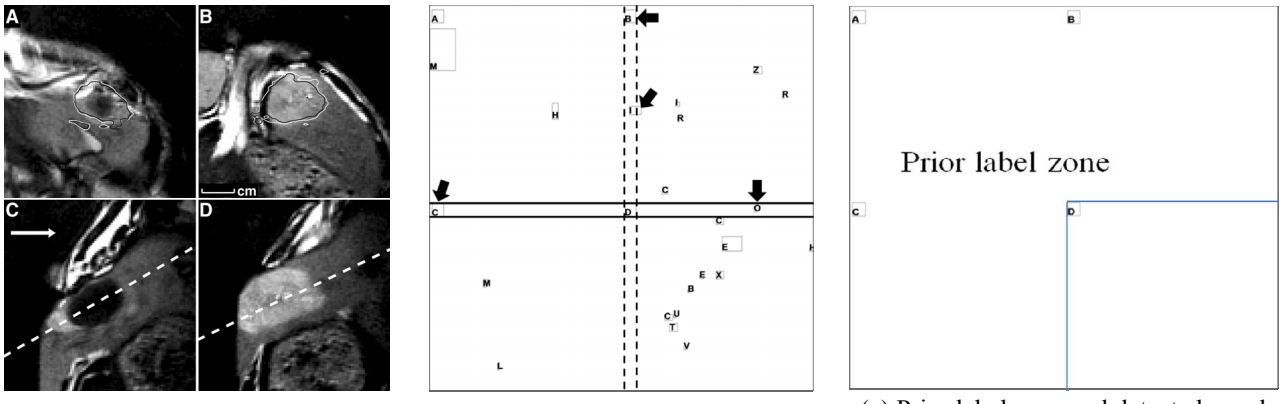
### D. Post-processing

True panel labels are easily identified by selecting text labels that have relatively large $b_i(1)$ (e.g., >0.8). Fig. 3(c) shows the detected panel labels that appear in Fig. 3(a). However, some noise labels may be detected as panel labels in some complicated cases containing many noise labels. Some post-processing can increase accuracy of panel label detection. Two post-processing methods can be considered to eliminate noise labels. One method uses textual information obtained from text caption analysis. Available information includes the number of panel labels and the text labels (e.g., 'A', 'B, 'C', 'D' for the Fig. 3 example). Another method is to check the belief $b_i(1)$ and find neighbors that propagate high messages for label '1' to text label $i$. Messages between text labels of true panel labels are likely to have high values and hence panel labels can be easily detected by tracing neighbors propagating high-valued messages starting from a certain label. We start tracing from 'A' since panel labels generally begins with 'A' and consider 'B' or 'C' in case 'A' is not detected for any reasons. The second method is applied in our detection approach since all necessary information for the method is available from the result of belief propagation.

## III. EXPERIMENT

### A. Data set

Our data set contains 515 multi-panel biomedical images that are selected from ImageCLEF [8] collection. The




(a) Input (b) Neighborhood system (c) Prior label zone and detected panel labels

Figure 3. Illustration of panel label detection

minimum and maximum numbers of panel labels in this set are 2 and 18, respectively.

### B. Test results

Our program performed the four proposed steps automatically on the images and the results were manually observed and analyzed. We judged a result as successful if all panel labels are detected with no noise labels added. Table 1 shows the test results and some important factors causing detection errors.

We judged a result as an error when a panel label was not detected or when a noise label was present. The main cause of errors, OCR error, is responsible for about 11% of the cases. We identified several factors in OCR errors.

- OCR errors due to narrow characters: The contour feature extraction from 3×3 subimages does not work well with some narrow characters such as I, l (lowercase 'L'), 1 (digit), j, and t. Some post-processing following the OCR engine is necessary to handle those characters (especially 'I') properly.
- Low recognition confidence: NN OCR engine provides floating point recognition confidence (< 1.0) and the score is used to compute the unary compatibility function. Low score results in low belief resulting in undetected panel labels.
- Errors in preprocessing step: Some extracted panel label CCs were severely distorted due to image characteristics or background interference. Any errors in preprocessing step may cause errors in later steps.

Fig. 4(a) and (b) show examples of panel label CCs touched by background and distorted due to binarization error, respectively. Preprocessing errors mainly occurred in low quality images or images with panel labels that may be binarized better with other thresholds than 50 or 200 used in our preprocessing step. Fig. 4(c) shows binarization results obtained with different thresholds. The result using a threshold of 170 (right most) is much better than the result using the default threshold 200 (middle). Overlay annotation segmentation method discussed in [9] may solve the binarization problem. Other minor error factors are listed in the table. Among them abnormal label alignment may be the most difficult problem since the neighborhood system in our model is defined based on the assumption of normal alignment, i.e., vertical, or horizontal. Observing more multi-panel figures and modifying the neighborhood system to accommodate for some variation in label alignment may solve the problem. Figure 5 shows additional sample panel label detection results.

(a) Touching background (b) Broken character (c) Binarization results w/ threshold 200 (middle) and 170 (right most)

Figure 4. Main causes of OCR errors

Table 1. Result of panel label detection

| | | # images (%) |
|---|---|---|
| **Success** | | **76.3** |
| Main causes of error | **OCR error** | **10.5** |
| | **Preprocessing error** | **5.0** |
| | Abnormal label alignment | 4.0 |
| | Post-processing error | 1.6 |
| | Abnormal label order | 1.4 |
| | Noise label added | 1.2 |
| Total | | 100% |

## IV. Conclusion

In this paper, a method to detect panel labels in multi-panel figures in biomedical articles is proposed. A MRF-based approach was applied to label candidate panel label connected components. An efficient neighborhood system and compatibility functions in belief propagation (BP) were defined based on our observation of various multi-panel figures in our data set. Error factors identified from our preliminary experiment are helpful to enhance performance. Future work includes i) finding a more robust character segmentation method to reduce OCR errors, ii) modifying the neighborhood system to handle abnormal label alignment, and iii) combining panel label detection results with textual information on panel labels extracted from figure captions and text mentions to increase accuracy of the detection results.

## Acknowledgment

This research is supported by the Intramural Research Program of the National Institutes of Health (NIH), National Library of Medicine (NLM), and Lister Hill National Center for Biomedical Communications (LHNCBC). We would like to thank the CLEF [8] organizers for making the database available for the experiments.

## References

[1] R. Saban, C. Simpson, R. Vadigepalli, S. Memet, I. Dozmorov, and M. Saban, "Bladder inflammatory transcriptome in response to tachykinins: Neurokinin 1 receptor-dependent genes and transcription regulatory elements," BMC Urol, 2007. 7: p. 7.

[2] S. Antani, D. Dember-Fushman, J. Li, B. Srinivasan, and G. R. Thoma, "Exploring use of images in clinical articles for decision support in evidence-based medicine," Proc. SPIE-IS&T Electronic Imaging. San Jose, CA, vol. 6815, January 2008.

[3] M. Sonka, V. Hlavac, and R. Boyle, Image processing, analysis, and machine vision. Thomson-Engineering, 2007.

[4] G. Kim and V. Govindaraju, "A lexicon driven approach to handwritten word recognition for real-time applications," IEEE Trans. Pattern Anal. Mach. Intell. vol. 19(4): 366-379, 1997.

[5] S. Z. Li, Markov Random Field modeling in image analysis. Springer-Verlag, 2009.

[6] D. You, S. Antani, D. Dember-Fushman, M. Rahman, V. Govindaraju, and G. R. Thoma, "Biomedical article retrieval using multimodal

features and image annotations in region-based CBIR," Proc. SPIE-IS&T Electronic Imaging. San Jose, CA, vol. 7534, January 2010.

[7] X. Peng, S. Setlur, V. Govindaraju, R. Sitaram, and Kiran Bhuvanagiri, "Markov Random Field based text identification from annotated machine printed documents," 10th Int. Conf. on Document Analysis and Recognition. pp. 431-435, 2009.

[8] H. Müller, J. Kalpathy-Cramer, I. Eggel, S. Bedrick, S. Radhouani, B. Bakke, C. E. Kahn Jr., and W. Hersh, "Overview of the CLEF 2009 medical image retrieval track," Working Notes of CLEF 2009 (Cross Language Evaluation Forum).

[9] D. You, S. Antani, D. Dember-Fushman, M. Rahman, V. Govindaraju, and G. R. Thoma, "Automatic identification of ROI in figure images toward improving hybrid (text and image) biomedical document retrieval," Proc. SPIE-IS&T Electronic Imaging. San Francisco, CA, vol. 7874, January 2011.

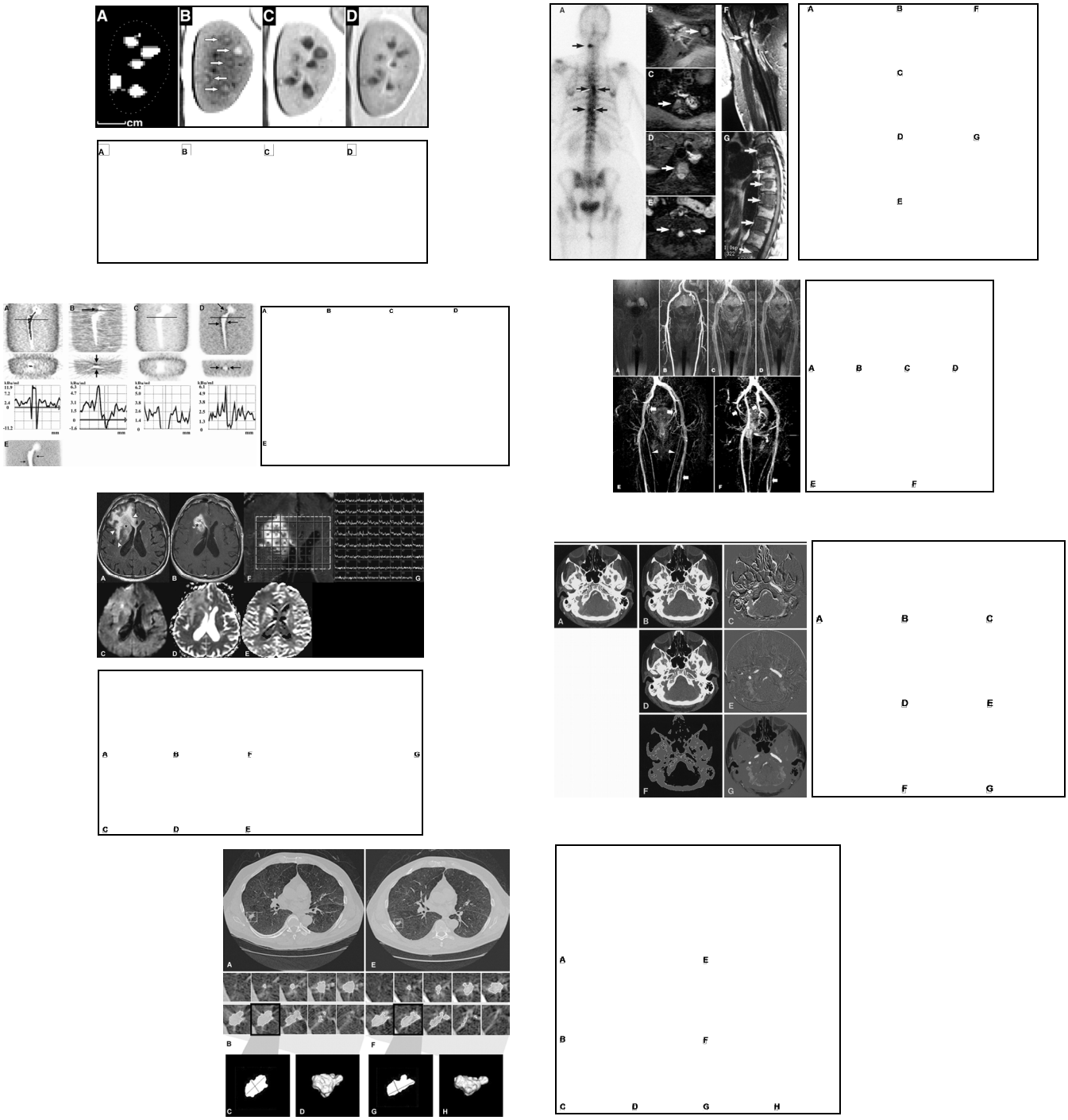


Figure 5. Sample multi-panel images in our test set and detection results