

Semantic Description of Medical Image Findings: Structured Learning Approach

Pavel Kisilev¹
pavelki@il.ibm.com

Eugene Walach¹
walach@il.ibm.com

Sharbell Hashoul^{1,2}
sharbelh@il.ibm.com

Ella Barkan¹
ella@il.ibm.com

Boaz Ophir¹
boazo@il.ibm.com

Sharon Alpert¹
asharon@il.ibm.com

¹ IBM Haifa Research Lab,
Haifa, Israel

² Carmel Medical Center,
Haifa, Israel

Abstract

Computer Aided Diagnosis (CADx) systems are designed to assist doctors in medical image interpretation. However, a CADx is often thought of as a "black box" whose diagnostic decision is not intelligible to a radiologist. Therefore, a system that uses semantic image interpretation, and mimics human image analysis, has clear benefits.

We propose a new method for automatic textual description of medical image findings, such as lesions in medical images. The method performs joint estimation of semantic features of lesions from image measurements. We formalize this problem as learning to map a set of diverse medical image measurements to a set of semantic descriptor values. We use a structured learning framework to model individual semantic descriptors and their relationships. The parameters of the model are efficiently learned using the Structured Support Vector Machine (SSVM).

The proposed approach generates radiological lexicon descriptors used to make a diagnosis of various diseases. This can help radiologists easily understand a diagnosis recommendation made by an automatic system, such as CADx. We apply the proposed method to publicly available and to proprietary breast and brain imaging datasets, and show that our method generates more accurate descriptions, as compared to other alternative approaches.

1 Introduction

One of the most important medical imaging tasks is lesion classification, which is used afterwards for disease diagnosis. For example, tumour lesions in various organs and modalities are classified by radiologists as malignant or benign, according to their *semantic*

features. Usually, the most important semantic features are lesion shape, boundary type, density characteristics, and others. Therefore, automatic classification of lesions requires explicit or implicit representation of the above semantic features by corresponding image measurements used during the classification process.

During the automatic diagnosis process, Computer Aided Diagnosis (CADx) systems use various *image features* derived from various image measurements. The features usually include histograms of intensity values, shape and texture descriptors, and others (see [1],[2] for an overview of such methods in breast imaging). Based on such features, CADx system makes a diagnosis decision.

Doctors often complain that CADx diagnostic decision process is not intelligible to them. Therefore, a system that has an ability to interpret medical images in semantic terms, offers several important advantages. In particular, automatically extracted semantic lesion descriptors can be used as a basis for a simplified radiological report. This, in turn, would reduce the load in a laborious reporting process, resulting in a faster and more coherent reporting system. Furthermore, providing a lucid presentation of the reasons for a given image interpretation, using the standard radiological lexicon would promote a physician's trust in the performance of the automated system. For example, a physician would know that the system classified a given tumour as benign due to its '*sharp edges*' and '*parallel to skin orientation*'.

The physician would be able to weigh these features relative to, for example, '*inhomogeneous tumour structure*'. This approach would allow setting the system parameters to match a physician's individual preferences. Today, the user can control only system trade-offs between sensitivity and specificity. In the semantic system, it would be possible to also control the relative impact of each feature.

Notice, that the *semantic step* in a CADx system is not a necessary step for making lesion classification. As mentioned above, typical CADx system uses various image features to make its final diagnosis. However, there are benefits in designing a system that can produce textual lesion description as an intermediate step.

In natural images, there is a growing number of papers dealing with the problems of automatic semantic tagging (see, for example, [4]), and of automatic description generation of images (see, for example, [5]-[7]). However, in medical imaging domain, it seems that this topic is yet to gain popularity. Adopting the aforementioned semantic approach to medical image description poses its own set of problems. Foremost is the choice of semantic descriptors. It is important that such a semantic system would be able to interact with a human reader using the standard radiological lexicon. The need for standardized terminology is not new. It is used in breast image reporting by human readers even without a CADx system being used. For that reason, the American College of Radiology developed the Breast Imaging Reporting and Data System (BI-RADS) [3] that standardizes the assessment and reporting of breast lesions. Although there are no analogues to BI-RADS standard in other radiological fields, doctors use similar semantic features, and describe lesions and write their reports in a similar manner.

In medical imaging, semantic descriptors have been primarily used in content-based information retrieval (CBIR) for the automatic retrieval of similar cases. In [8], for example, the authors presented a content-based mammogram retrieval system, which uses a query example to search for similar mammograms in the database. The mammographic lesions are interpreted based on their medical characteristics as specified in the BI-RADS lexicon. Then, the authors apply a hierarchical similarity measure based on a distance weighting function; each medical feature is considered *independently*.

Another use of semantic descriptors was proposed in [9], and aimed to extend the classification to specific disease classes, in contrast to usual benign-or-malignant binary classification traditionally used in CADx. The authors used semantic features that were extracted manually by radiologists. They introduced a Bayesian network that models the probabilistic relationships between breast diseases, mammographic findings, and patient risk factors to provide a disease specific classification. As in [8], the semantic terms are assumed to be conditionally independent.

In [10], the BI-RADS lexicon is used to extract texture characteristics associated with image regions obtained from a human reader's mammography reports. The method searches similar regions by computing the Mahalanobis distance of feature vectors that describe the shape and texture characteristics of the selected regions. Then, using the k-Nearest Neighbours (KNN) approach, the authors extract BI-RADS semantic descriptors in new images.

The above methods can be categorized into two major types. The first category of methods performs independent estimation of semantic descriptors. This assumption is used, for example, in [9]. The second category of methods is based on the KNN approach. This method is used, for example, in [10]. In the first type of the above methods, the independent estimation of each of the semantic descriptors is performed by a multiclass classification approach. The second type of methods applies unsupervised clustering approach. Our approach is genuinely different from these methods, since it learns *individual representations and dependencies* of the semantic features from data in a *discriminative* manner.

The main contributions of this work are as follows: 1) a new model for textual description of lesions, which captures semantic feature relationships; 2) a novel, discriminative method for automatic extraction of a basic radiological report which is medically sound and based on the standard radiological terminology; 3) the proposed method mitigates one of the major complaints of radiologists about CADx systems, namely, the lack of intelligibility of their decision process. Our method 'explains' to a radiologist why a particular diagnosis is made, using the standard radiological language.

2 Proposed approach

The main mode of operation of the proposed method is illustrated in Figure 1. Two examples of input images, of the detected lesion contours, and of the corresponding automatically generated textual descriptions of lesions using the proposed method are depicted in Figure 1a-1c, respectively.

In the following sections we describe the methodology of the proposed method in details. Here, we give a brief description of the main steps of the method. Given a medical image, the first step is to localize a lesion and to find its contour. In this work, we concentrate on the problem of automatic generation of semantic description of lesions. We, therefore, use semi-automatic lesion detection and contour extraction, instead of a fully automatic approach. We assume that the bounding box around a lesion is found or given by a radiologist. Then, an active contour type method is applied to find the lesion contour inside of the bounding box. Given the found lesion contour, we calculate various image measurements and, based on it, construct visual features. Finally, we use the learned in advance model of mapping from visual features to semantic values, and generate the semantic description of a new lesion.

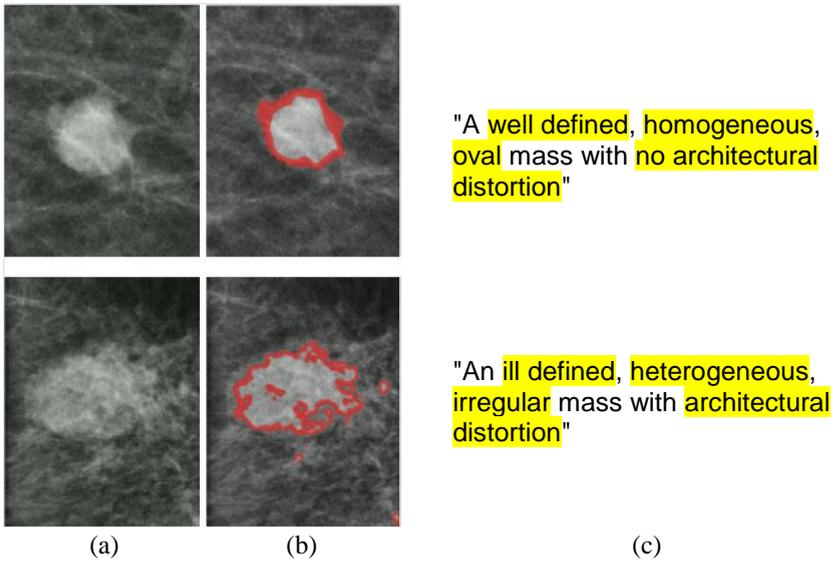


Figure 1: Examples of automatically generated textual description of lesions, using the proposed method: (a) region of interest (ROI) taken from a breast mammography image, (b) found lesion contour (c) textual description of the lesion (estimated semantic descriptors are marked in yellow colour).

2.1 Semantic descriptors and their corresponding image measurements

Having found a lesion, a radiologist examines its visual appearance characteristics to make a final diagnosis. In particular, in breast ultrasound (US), the radiological lexicon standardized by BI-RADS contains the following semantic descriptors: shape, margin, orientation, acoustic transmission (posterior enhancement/shadowing), boundary, and echo pattern. A graphical illustration of these semantic descriptors and corresponding semantic values is presented in Figure 2. Similarly, in mammography (MG), the semantic descriptors defined in BI-RADS are: shape, margin, and density. In contrast, in brain imaging, for example, there is no strict standard such as BI-RADS. Nevertheless, doctors describe brain lesions using similar terms. In particular, brain tumour semantic descriptors usually include: shape, margin, boundary type, contrast enhancement, localization, mass effect on surrounding tissues, edema and others.

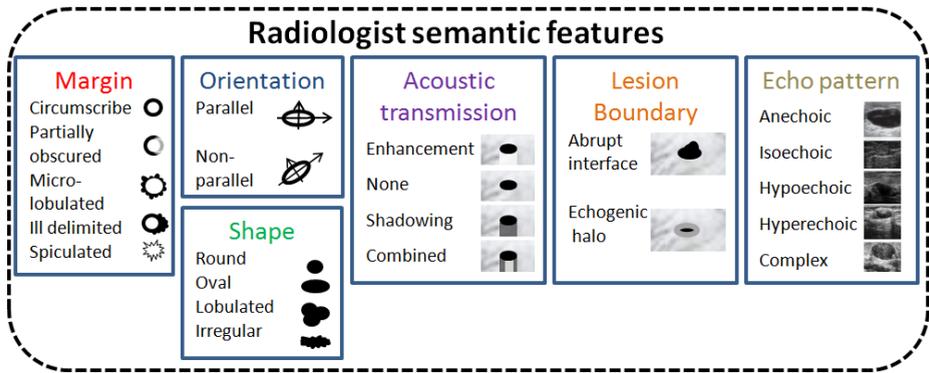


Figure 2: Example of semantic descriptors and corresponding semantic values in the ultrasound imaging modality (graphical illustration).

The estimation of semantic descriptor values requires explicit or implicit representation by a diverse set of image measurements that describe each one of the semantic descriptors quantitatively. We use these image measurements to calculate the informative features, such as histograms of pixel values, shape and texture descriptors and others. The following are brief explanations of semantic descriptors and their corresponding image measurements that form image features.

Shape and orientation. The shape of a lesion is the most important semantic characteristic. Malignant tumours tend to have more irregular and lobular shapes. We calculate different quantities such as the area of the lesion, its aspect ratio, and the curvature along the boundaries. Other shape features obtained by fitting an ellipse to the borders include: the ellipse orientation, the ratio between the minor and the major axes, and L1 norm and the maximum of distances between the contour and the ellipse. We also use a variant of the shape context descriptor [11].

Margin and boundary. Sharp margins may indicate a benign tumour while smooth margins may indicate a malignant one. To assess the sharpness of the boundaries we divide the mass into 8 sectors of 45 degrees, and calculate the sharpness of the boundary in each sector. The sharpness is calculated as the maximal slope of the boundary profile.

Intensity and texture characteristics. *Density* and *echo pattern* semantic descriptors are defined by the pixel intensity measurements. We compute two normalized intensity histograms of the inner and the outer (that is, next to the boundary), areas of the lesion. In addition, we characterise texture content of a lesion using several descriptors: entropy values of a local patch around each pixel computed from three dyadic scales of an image, and the Local Binary Pattern (LBP) descriptor [12]. We found these simple texture descriptors work well.

Other descriptors and measurements. More complex measurements, such as *acoustic transmission* and *echo pattern*, are used in ultrasound images. These measurements require additional detection capabilities. The posterior of the mass is an important characteristic when assessing the risk of malignancy. Strong enhancement and edge shadowing are common in benign masses such as cysts, while posterior shadowing is common in malignant tumours. To assess the level of the posterior enhancement or shadowing, we automatically detect the area below the mass, and calculate ratios of the median intensities and intensity variances inside its different segments.

Another important characteristic of masses examined by doctors is their *echogenicity* compared to the fat tissue. High values may indicate malignancy; the echogenicity and mass uniformity are useful for diagnosis of specific types of tumours. To quantify these features we use heuristics to recognize the fat tissue which is located on the upper side of the US images. We then compare the histogram of the lesion interior values to the one of the fat tissue.

An additional semantic descriptor which is used sometimes in radiological reporting is the *architectural distortion*. It is defined by the BI-RADS system as "an appearance that may include spiculations radiating from a point and focal retraction or distortion at the edges". The shape features described above serve as a good predictor of this descriptor value. (In contrast to other semantic descriptors that can take multiple possible values, the architectural distortion is a binary value descriptor.)

All the above image measurements represent feature channels, and are combined into the feature vector containing 236 values. These features are used to learn the model parameters and to make the inference as described in the next section.

2.2 Structured learning formulation for semantic description of a lesion

We pose the problem of semantic description of a lesion as learning to map a set of image based informative features to a set of semantic descriptor values. A lesion is described by a set of J semantic descriptors. Semantic description of the i -th lesion is an assignment $\mathcal{Y}_i = [y_{i,1}, \dots, y_{i,J}]$ where each j -th semantic descriptor $y_{i,j}$ can have one of the possible discrete values $Y_j \in \{1, \dots, V_j\}$ corresponding to the radiological lexicon. Following the standard practice in structured learning [14], the energy function of the above assignment is a sum of unary and pair-wise terms:

$$E(\mathbf{y}_i) = \sum_j \mathbf{u}_1^T \phi_1(y_{ij}, \mathbf{X}_i) + \sum_{j,k \in S} \mathbf{u}_2^T \phi_2(y_{ij}, y_{ik}, \mathbf{X}_i), \quad (1)$$

where \mathbf{X}_i are image measurements (visual features); ϕ_1 , and ϕ_2 are unary and pair-wise potentials, respectively, defined below; S is the set of all possible pairs of semantic descriptors; $\mathbf{u}_1, \mathbf{u}_2$ are the model parameters. The unary potentials capture the relationship between the image measurements and semantic descriptor values; the pairwise potentials capture joint relationships between the semantic descriptors, and reflect the likelihood of semantic descriptors to jointly have particular values. The unary potentials for j -th descriptor and feature channel c are defined as:

$$\phi_{1,j,c}(y_{ij} = Y_j, \mathbf{X}_i^c; \theta_{1,j}) \sim P_{Y_j} \doteq Pr(y_{ij} = Y_j | \mathbf{X}_i^c; \theta_{1,j,c}). \quad (2)$$

They are modelled using a softmax classifier whose parameters $\theta_{1,j,c}$ are learned from a training set of examples. ϕ_1 can be seen as a predictor of a semantic descriptor $y_{i,j}$ given a set of measurements \mathbf{X}_i . Similarly to the unary potentials, the pairwise potentials

$$\phi_{2,j,c}(y_{ij} = Y_j, y_{ik} = Y_k, \mathbf{X}_i^c; \theta_{2,j,c}) \sim P_{Y_j, Y_k} \doteq Pr(y_{ij} = Y_j, y_{ik} = Y_k | \mathbf{X}_i^c; \theta_{2,j,c}) \quad (3)$$

are modelled using a softmax classifier with parameters $\theta_{2,j,c}$.

2.3 Learning and inference

We learn the parameters $\mathbf{u} = [\mathbf{u}_1^T \ \mathbf{u}_2^T]^T$ of the model (1), using Structured SVM (SSVM) framework. In particular, given N training examples, the model parameters are learned by optimizing the regularized large-margin objective [13], [14]:

$$\begin{aligned} \hat{\mathbf{u}} = \min_{\mathbf{u}, \xi \geq 0} & \frac{1}{2} \|\mathbf{u}\|^2 + C\xi \\ \text{s.t.} & \frac{1}{N} \sum_{i=1}^N \max_{\bar{\mathbf{y}}_i \in \mathcal{Y}} \left[\Delta(\bar{\mathbf{y}}_i, \mathbf{y}_i^*) - \langle \mathbf{u}, \psi(I_i, \bar{\mathbf{y}}_i) \rangle + \langle \mathbf{u}, \psi(I_i, \mathbf{y}_i) \rangle \right] \leq \xi \end{aligned} \quad (4)$$

where the unary and the pairwise potentials are concatenated into a column vector:

$$\psi(\bar{\mathbf{y}}_i, \mathbf{X}_i) = \left[\left(\sum_j \phi_1(\bar{y}_{ij}, \mathbf{X}_{ij}) \right)^T \quad \left(\sum_{j,k \in S} \phi_2(\bar{y}_{ij}, \bar{y}_{ik}, \mathbf{X}_{ij}, \mathbf{X}_{ik}) \right)^T \right]^T. \quad (5)$$

We use the normalized weighted Hamming loss as the task loss:

$$\Delta(\bar{\mathbf{y}}_i, \mathbf{y}_i^*) = \sum_j w_j \mathbf{I}(\bar{y}_{ij} \neq y_{ij}^*) / \sum_j w_j, \quad (6)$$

where weights w_j are defined (or learned in advance) by the relative importance of semantic descriptors in a diagnosis process, and \mathbf{I} is the indicator function that equals 1 whenever the predicted value \bar{y}_{ij} is different from the actual value y_{ij}^* .

Once the model parameters are learned (as described above), the inference goal is, given a new a lesion, to find the best assignment whose semantic values result in the lowest energy (1). This is achieved by solving:

$$\hat{\mathbf{y}}_i = \arg \min_{\bar{\mathbf{y}}_i \in \mathcal{Y}} \langle \hat{\mathbf{u}}, \psi(\bar{\mathbf{y}}_i, \mathbf{X}_i) \rangle. \quad (7)$$

The solution of (7) can be found using the Sequence Alignment algorithms [13]. Notice, that the above formulation of the learning problem is similar to the Conditional Random Field (CRF) modelling used in object segmentation, natural language processing, and other fields (see for example, [15]).

After semantic descriptor values are estimated, they can be presented to a doctor in a form of a simplified radiology report. In addition, they can be used for making a lesion classification or a diagnosis, wherein they are used as features in a standard binary (malignant-or-benign) or a multiclass (specific disease) classification problem. This classification can be performed using, for example, SVM classifier. As mentioned above, the final diagnosis or lesion classification can be performed without the 'semantic step', directly from the image data. In fact, we compared both methods (with and without the intermediate 'semantic step'), and did not find any statistically meaningful difference in the

final diagnosis accuracy. However, the 'semantic step' provides readable explanation and insights, and, therefore, we believe, is beneficial for radiologists.

3 Experiments

We tested the accuracy of our method, and compared it with other approaches on breast and brain lesion imaging datasets. We used the widely known DDSM dataset [16], where we chose MG images with masses with annotated semantic descriptors of shape, margin, and density. After removing images with rare combinations (that make it difficult to train the system properly), the final set contains total of 1546 images (corresponding to 773 cases, or patients). Additionally, our proprietary dataset contains 408 US images (203 cases), and 270 digitally acquired MG images (160 cases). The values of semantic descriptors in our US and MG datasets were annotated by a trained radiologist according to BI-RADS: 6 descriptors for US, and 3 for MG. We used data from the Brain Tumor Segmentation (BRATS) MICCAI 2013 challenge. For BRATS sequences, our radiologist annotated each case using 4 descriptors: shape, margin, enhancement, and T1 homogeneity. The resulting brain dataset contained 1400 images (30 cases).

To extract accurate lesion boundaries, we used semi-automatic segmentation in all datasets. The radiologist chose a square region of interest (ROI) around a lesion. We then used the active contour type algorithm [17] to extract the contours. In DDSM, we used the original annotated contours to define the ROI, and applied the above active contour type algorithm to extract more accurate boundaries. We used these accurate contours to compute features described in Section 2.1.

To the best of our knowledge, there are no systems that attempt to model and describe medical findings using complex relationships between the semantic descriptors. We therefore compared our method with the approaches that can be considered as competing methods: 1) independent estimation of semantic descriptors (for example, [9]), and 2) the KNN based approach (for example, [10]).

One of the goals of our experiments is to compare the proposed SSVM based semantic descriptor estimation method with other possible approaches. In particular, we show that modelling the relationships between the semantic descriptors has a clear advantage over independent assumption based ones. Therefore, to exclude the factor of predictive ability of specific image features, we used the same image features in all experiments. (Obviously, using more sophisticated features is expected to give higher predictive ability). Notice also, that our goal is not the accuracy improvement of the final diagnosis, but rather introducing an intermediate 'semantic step' which helps to radiologists to understand a CADx system decision. We, therefore, do not present the final diagnosis accuracy. (As mentioned above, using or omitting the 'semantic step', had minor effect on the final diagnosis accuracy).

We compared the proposed SSVM-based method with KNN-based and with independent semantic descriptor estimation. In the KNN-based method, we implemented the approach proposed in [10]. For independent estimation, we used multiclass SVM classifier with RBF kernel.

In all the experiments, we used the following experimental methodology. The set of images was divided (with stratification) into three equal parts (denoted by segments A, B and C). Segment C was reserved as a testing set. Every algorithm was trained on segment A. The optimal values of parameters, evaluated on the segment B (the validation segment)

were picked, and the algorithm was retrained on both segment A and B. Then, the algorithm was tested on segment C. This process was repeated with reversed roles for segments A and B, namely with segment B as the training set and segment A as the validation. This procedure was repeated five times (5x2 cross validation), and concluded with 10 trials. For each one of the semantic descriptors, we calculated the means and the standard deviations (STD) of the following three performance metrics: 1) the accuracy, $ACC=(TP+TN)/M$, 2) the positive predictive value, $PPV=TP/(TP+FP)$, and 3) the true positive rate, $TPR=TP/(TP+FN)$. Here, M is the total number of testing examples, TP, TN, FP, and FN are the number of true positives, of true negatives, of false positives, and of false negatives, respectively.

In DDSM experiments, we used the following descriptors and their corresponding values: *shape* {*round; oval; lobulated; irregular; architectural distortion*}, *margin* {*circumscribed; ill-defined; spiculated; microlobulated; obscured*}, and four density values. The results of the means of the performance metrics are summarized in Table 1. The STD's of the metrics were all under 5% of the mean values.

The results of the means of the performance metrics for our proprietary breast US and MG datasets are summarized in Table 2 and Table 3 respectively. In this case, the STD's of the metrics were all under 4.7 and 7.6% of the mean values, respectively. Because of the relatively small number of examples, we used a reduced set of semantic values. In particular, in US experiments, we used 3 classes (possible values) for margin, shape, and echo, and 2 classes for the rest. In MG experiments, we used 3 classes for shape and margin, and 2 classes for density. We report only the accuracy for these experiments, since these numbers represent well the overall tendency. The results of the means of the performance metrics for the BRATS brain tumour dataset are presented in Table 4.

The proposed approach outperformed the two other methods in nearly all figures. We attribute this to the ability of our more sophisticated model to better capture hidden relationships between different semantic descriptors. It should be noticed that in some cases, such as density value prediction, there was very marginal gain in the performance, as compared to the independent estimation. This may indicate that this descriptor has weak relationship with the other descriptors.

Semantic descriptor Estimation method	Shape			Margin			Density		
	ACC	PPV	TPR	ACC	PPV	TPR	ACC	PPV	TPR
Independent estimation	0.64	0.64	0.66	0.62	0.63	0.63	0.71	0.72	0.71
k-NN based [10]	0.67	0.67	0.68	0.64	0.65	0.67	0.72	0.72	0.73
Proposed SSVM based	0.71	0.71	0.72	0.69	0.68	0.69	0.73	0.72	0.73

Table 1. DDSM dataset: semantic descriptor estimation; mean performance (bold indicates the best result). The STD's of the metrics are all under 5% of the mean values.

Semantic descriptor Estimation method	Shape ACC	Margin ACC	Density ACC
Independent estimation	0.73	0.72	0.81
k-NN based [10]	0.74	0.76	0.80
Proposed SSVM based	0.79	0.78	0.82

Table 2. Mammography dataset (proprietary): semantic descriptor estimation; mean performance (bold indicates the best result). The STD's of the metrics are all under 4.7% of the mean values.

Semantic descriptor Estimation method	Shape ACC	Orient. ACC	Margin ACC	Echo ACC	Transm. ACC	Boundary ACC
Independent estimation	0.62	0.98	0.6	0.75	0.79	0.74
k-NN based [10]	0.64	0.92	0.63	0.76	0.78	0.76
Proposed SSVM based	0.68	0.94	0.69	0.78	0.81	0.76

Table 3. Ultrasound dataset (proprietary): semantic descriptor estimation; mean performance (bold indicates the best result). The STD's of the metrics are all under 7.6% of the mean values.

Semantic descriptor Estimation method	Shape ACC	Margin ACC	Enhancement ACC	Homogeneity ACC
Independent estimation	0.72	0.84	0.8	0.79
k-NN based [10]	0.74	0.85	0.83	0.77
Proposed SSVM based	0.78	0.9	0.83	0.8

Table 4. Brain dataset (BRATS2013): semantic descriptor estimation results; mean performance (bold indicates the best result). The STD's of the metrics are all under 4.9% of the mean values.

4 Discussion

We presented a novel discriminative method for automatically generating semantic descriptions of lesions using a structured learning approach. The method addresses one of the major complaints of radiologists: the lack of intelligibility of the CADx decision process. Our system can 'explain' to a radiologist why a particular diagnosis is made, using the standard radiological lexicon. The proposed method improves the accuracy of semantic feature estimation by modelling their relationships. It outperforms the competing KNN-type method by up to 5% in accuracy. Similar gain in the accuracy is achieved on the brain MRI data (MICCAI BRATS 2013 challenge).

The main obstacle in making such a system completely automatic is the difficulty to design highly accurate fully automatic lesion localization and segmentation algorithms. Obviously, the accuracy of the semantic descriptor estimation is greatly dependent on the accuracy of the feature calculation, which, in turn, relies on the segmentation method. We plan to investigate dependable lesion segmentation methods, along with improvements in image feature measurements that will make our system less sensitive to the segmentation errors and provide more accurate results. Also, in order to use our system for free-text radiological report generation, natural language processing (NLP) methods should be engaged in the future.

References

- [1] A. Oliver, J. Freixenet, J. Martí, E. Pérez, J. Pont, E. R. Denton, and R. Zwiggelaar. A review of automatic mass detection and segmentation in mammographic images. *Medical Image Analysis*, 14(2):87–110, 2010.
- [2] T. Ayer, M. U. Ayvaci, Z. X. Liu, O. Alagoz, and E. S. Burnside. Computer-aided diagnostic models in breast cancer screening. *Imaging in medicine*, 2(3):313–323, 2010.

- [3] D'Orsi CJ, Mendelson EB, Ikeda DM, et al. Breast Imaging Reporting and Data System: ACR BI-RADS - Breast Imaging Atlas. Reston, VA: American College of Radiology; 2003.
- [4] M. Guillaumin, T. Mensink, J. J. Verbeek and C. Schmid. Tagprop: Discriminative metric learning in nearest neighbor models for image auto-annotation. In ICCV, pages 309–316, 2009.
- [5] Ali Farhadi, Seyyed Mohammad Mohsen Hejrati, Mohammad Amin Sadeghi, Peter Young, Cyrus Rashtchian, Julia Hockenmaier, and David A. Forsyth. Every picture tells a story: Generating sentences from images. In ECCV (4), pages 15–29, 2010.
- [6] V. Ordonez, G. Kulkarni and T. L. Berg. Im2Text: Describing Images Using 1 Million Captioned Photographs. In NIPS 2011, pages 1143-1151.
- [7] D. Elliott and F. Keller. Image Description using Visual Dependency Representations. EMNLP, pages 1292-1302, 2013.
- [8] C.-H. Wei, Y. Li, and P. J. Huang. Mammogram retrieval through machine learning within bi-rads standards. Journal of biomedical informatics, 44(4):607–614, 2011.
- [9] D. L. Rubin, E. S. Burnside, and R. Shachter. A bayesian network to assist mammography interpretation. In Operations Research and Health Care, pages 695–720. Springer, 2004.
- [10] F. Narvaez, G. Diaz, and E. Romero. Automatic bi-rads description of mammographic masses. In Digital Mammography, pages 673–681. Springer, 2010.
- [11] S. Belongie, J. Malik, and J. Puzicha. Shape Matching and Object Recognition Using Shape Contexts. IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (24): 509–521.
- [12] T. Ojala, M. Pietikäinen, and D. Harwood (1994), "Performance evaluation of texture measures with classification based on Kullback discrimination of distributions", Proc. 12th IAPR International Conference on Pattern Recognition (ICPR 1994), vol. 1, pp. 582 - 585.
- [13] I. Tsochantaridis, T. Joachims, T. Hofmann, and Y. Altun. Large Margin Methods for Structured and Interdependent Output Variables. Journal of Machine Learning Research (JMLR), 6(Sep):1453-1484, 2005.
- [14] Nowozin, Sebastian, and Christoph H. Lampert. "Structured learning and prediction in computer vision." Foundations and Trends in Computer Graphics and Vision 6.3–4 (2011): 185-365.
- [15] Lafferty, J., McCallum, A., Pereira, F. (2001). "Conditional random fields: Probabilistic models for segmenting and labeling sequence data". Proc. 18th International Conf. on Machine Learning. Morgan Kaufmann. pp. 282–289.
- [16] Michael Heath, Kevin Bowyer, Daniel Kopans, Richard Moore and W. Philip Kegelmeyer, "The Digital Database for Screening Mammography", in Proceedings of the Fifth International Workshop on Digital Mammography, M.J. Yaffe, ed., 212-218, Medical Physics Publishing, 2001. ISBN 1-930524-00-5.
- [17] Lankton, S.; Tannenbaum, A., "Localizing Region-Based Active Contours," Image Processing, IEEE Transactions on , vol.17, no.11, pp.2029,2039, Nov. 2008.