

# Using Non-lexical Features to Identify Effective Indexing Terms for Biomedical Illustrations

Matthew Simpson, Dina Demner-Fushman, Charles Sneiderman,  
Sameer K. Antani, George R. Thoma

Lister Hill National Center for Biomedical Communications  
National Library of Medicine, NIH, Bethesda, MD, USA  
{simpsonmatt, ddemner, csneiderman, santani, gthoma}@mail.nih.gov

## Abstract

Automatic image annotation is an attractive approach for enabling convenient access to images found in a variety of documents. Since image captions and relevant discussions found in the text can be useful for summarizing the content of images, it is also possible that this text can be used to generate salient indexing terms. Unfortunately, this problem is generally domain-specific because indexing terms that are useful in one domain can be ineffective in others. Thus, we present a supervised machine learning approach to image annotation utilizing *non-lexical* features<sup>1</sup> extracted from image-related text to select useful terms. We apply this approach to several subdomains of the biomedical sciences and show that we are able to reduce the number of ineffective indexing terms.

## 1 Introduction

Authors of biomedical publications often utilize images and other illustrations to convey information essential to the article and to support and reinforce textual content. These images are useful in support of clinical decisions, in rich document summaries, and for instructional purposes. The task of delivering these images, and the publications in which they are contained, to biomedical clinicians and researchers in an accessible way is an information retrieval problem.

Current research in the biomedical domain (e.g., Antani et al., 2008; Florea et al., 2007), has investigated hybrid approaches to image retrieval, combining elements of content-based image retrieval (CBIR) and annotation-based image retrieval (ABIR). ABIR, compared to the image-

<sup>1</sup>Non-lexical features describe attributes of image-related text but not the text *itself*, e.g., unlike a bag-of-words model.

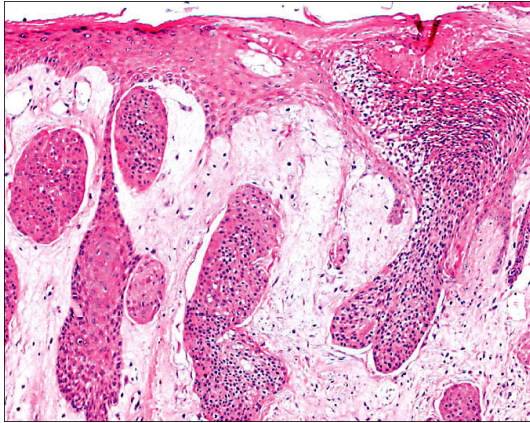
only approach of CBIR, offers a practical advantage in that queries can be more naturally specified by a human user (Inoue, 2004). However, manually annotating biomedical images is a laborious and subjective task that often leads to noisy results.

Automatic image annotation is a more robust approach to ABIR than manual annotation. Unfortunately, automatically selecting the most appropriate indexing terms is an especially challenging problem for biomedical images because of the domain-specific nature of these images and the many vocabularies used in the biomedical sciences. For example, the term “sweat gland adenocarcinoma” could be a useful indexing term for an image found in a dermatology publication, but it is less likely to have much relevance in describing an image from a cardiology publication. On the other hand, the term “mitral annular calcification” may be of great relevance for cardiology images, but of little relevance for dermatology ones.

Our problem may be summarized as follows: Given an image, its caption, its discussion in the article text (henceforth the image *mention*), and a list of potential indexing terms, select the terms that are most effective at describing the content of the image. For example, assume the image shown in Figure 1, obtained from the article “Metastatic Hidradenocarcinoma: Efficacy of Capecitabine” by Thomas et al. (2006) in *Archives of Dermatology*, has the following potential indexing terms,

- Histopathology finding
- Reviewed
- Confirmation
- Diagnosis aspect
- Diagnosis
- Eccrine
- Sweat gland adenocarcinoma
- Lesion

which have been extracted from the image mention. While most of these do not uniquely identify



**Caption:** Figure 1. On recurrence, histologic features of porocarcinoma with an intraepidermal spread of neoplastic clusters (hematoxylin-eosin, original magnification x100).

**Mention:** Histopathologic findings were reviewed and confirmed a diagnosis of eccrine hidradenocarcinoma for all lesions excised (Figure 1).

**Figure 1: Example Image.** We index an image with concepts generated from its caption and discussion in the document text (mention). This image is from “Metastatic Hidradenocarcinoma: Efficacy of Capecitabine” by Thomas et al. (2006) and is reprinted with permission from the authors.

the image, we would like to automatically select “sweat gland adenocarcinoma” and “eccrine” for indexing because they clearly describe the content and purpose of the image—supporting a diagnosis of hidradenocarcinoma, an invasive cancer of sweat glands. Note that effective indexing terms need not be exact lexical matches of the text. Even though “diagnosis” is an exact match, its meaning is too broad in this context to be a useful term.

In a machine learning approach to image annotation, training data based on lexical features alone is not sufficient for finding salient indexing terms. Indeed, we must classify terms that are not encountered while training. Therefore, we hypothesize that non-lexical features, which have been successfully used for speech and genre classification tasks, among others (see Section 5 for related work), may be useful in classifying text associated with images. While this approach is broad enough to apply to any retrieval task, given the goals of our ongoing research, we restrict ourselves to studying its feasibility in the biomedical domain.

In order to achieve this, we make use of the previously developed MetaMap (Aronson, 2001)

tool, which maps text to concepts contained in the Unified Medical Language System<sup>®</sup> (UMLS) Metathesaurus<sup>®</sup> (Lindberg et al., 1993). The UMLS is a compendium of several controlled vocabularies in the biomedical sciences that provides a semantic mapping relating concepts from the various vocabularies (Section 2). We then use a supervised machine learning approach, described in Section 3, to classify the UMLS concepts as useful indexing terms based on their non-lexical features, gleaned from the article text and MetaMap output.

Experimental results, presented in Section 4, indicate that ineffective indexing terms can be reduced using this classification technique. We conclude that ABIR approaches to biomedical image retrieval as well as hybrid CBIR/ABIR approaches, which rely on both image content and annotations, can benefit from an automatic annotation process utilizing non-lexical features to aid in the selection of useful indexing terms.

## 2 Image Retrieval: Recent Work

Automatic image annotation is a broad topic, and the automatic annotation of biomedical images, specifically, has been a frequent component of the ImageCLEF<sup>2</sup> cross-language image retrieval workshop. In this section, we describe previous work in biomedical image retrieval that forms the basis of our approach. Refer to Section 5 for work related to our method in general.

Demner-Fushman et al. (2007) developed a machine learning approach to identify images from biomedical publications that are relevant to clinical decision support. In this work, the authors utilized both image and textual features to classify images based on their usefulness in evidence-based medicine. In contrast, our work is focused on selecting useful biomedical image indexing terms; however, we utilize the methods developed in their work to extract images and their related captions and mentions.

Authors of biomedical publications often assemble multiple images into a single multi-panel figure. Antani et al. (2008) developed a unique two-phase approach for detecting and segmenting these figures. The authors rely on cues from captions to inform an image analysis algorithm that determines panel edge information. We make use of this approach to uniquely associate caption and mention text with a single image.

<sup>2</sup><http://imageclef.org/>

Our current work most directly stems from the results of a term extraction and image annotation evaluation performed by Demner-Fushman et al. (2008). In this study, the authors utilized MetaMap to extract potential indexing terms (UMLS concepts) from image captions and mentions. They then asked a group of five physicians and one medical imaging specialist (four of whom are trained in medical informatics) to manually classify each concept as being “useful for indexing” its associated images or ineffective for this purpose. The reviewers also had the opportunity to identify additional indexing terms that were not automatically extracted by MetaMap.

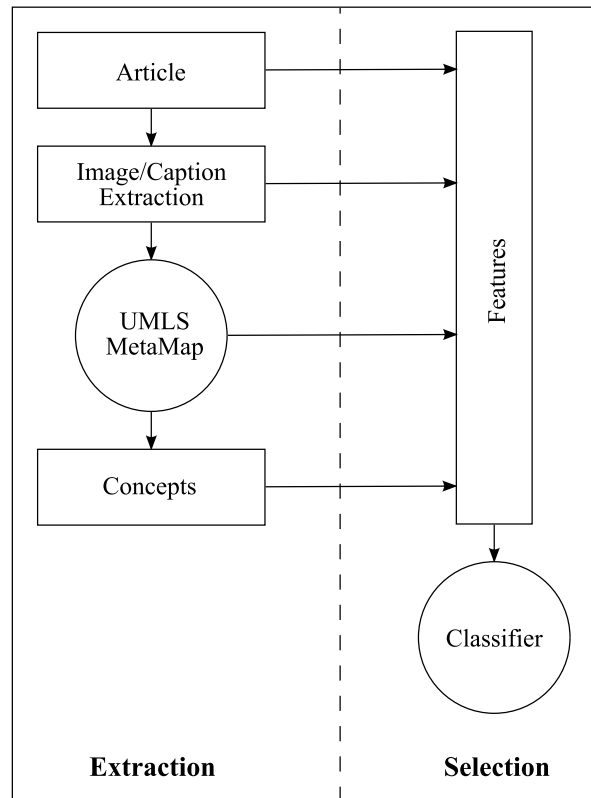
In total, the reviewers evaluated 4006 concepts (3,281 of which were unique), associated with 186 images from 109 different biomedical articles. Each reviewer was given 50 randomly chosen images from the 2006–2007 issues of *Archives of Facial Plastic Surgery*<sup>3</sup> and *Cardiovascular Ultrasound*<sup>4</sup>. Since MetaMap did not automatically extract all of the useful indexing terms, this selection process exhibited high recall averaging 0.64 but a low precision of 0.11. Indeed, assuming all the extracted terms were selected for indexing, this results in an average  $F_1$ -score of only 0.182 for the classification problem. Our work is aimed at improving this baseline classification by reducing the number of ineffective terms selected for indexing.

### 3 Term Selection Method

A pictorial representation of our term extraction and selection process is shown in Figure 2. We rely on the previously described methods to extract images and their corresponding captions and mentions, and the MetaMap tool to map this text to UMLS concepts. These concepts are potential indexing terms for the associated image.

We derive term features from various textual items, such as the preferred name of the UMLS concept, the MetaMap output for the concept, the text that generated the concept, the article containing the image, and the document collection containing the article. These are all described in more detail in Section 3.2. Once the feature vectors are built, we automatically classify the term as either being useful for indexing the image or not.

To select useful indexing terms, we trained a binary classifier, described in Section 3.3, in a



**Figure 2: Term Extraction and Selection.** We gather features for the extracted terms and use them to train a classifier that selects the terms that are useful for indexing the associated images.

supervised learning scenario with data obtained from the previous study by Demner-Fushman et al. (2008). We obtained our evaluation data from the 2006 *Archives of Dermatology*<sup>5</sup> journal. Note that our training and evaluation data represent distinct subdomains of the biomedical sciences.

In order to reduce noise in the classification of our evaluation data, we asked two of the reviewers who participated in the initial study to manually classify our extracted terms as they did for our training data. In doing so, they each evaluated an identical set of 1539 potential indexing terms relating to 50 randomly chosen images from 31 different articles. We measured the performance of our classifier in terms of how well it performed against this manual evaluation. These results, as well as a discussion pertaining to the inter-annotator agreement of the two reviewers, are presented in Section 4.

Since our general approach is not specific to the biomedical domain, it could equally be applied in

<sup>3</sup><http://archfaci.ama-assn.org/>

<sup>4</sup><http://www.cardiovascularultrasound.com/>

<sup>5</sup><http://archderm.ama-assn.org/>

any domain with an existing ontology. For example, the UMLS and MetaMap can be replaced by the Art and Architecture Thesaurus<sup>6</sup> and an equivalent mapping tool to annotate images related to art and art history (Klavans et al., 2008).

### 3.1 Terminology

To describe our features, we adopt the following terminology.

- A *collection* contains all the articles from a given publication for a specified number of years. For example, the 2006–2007 issues of *Cardiovascular Ultrasound* represent a single collection.
- A *document* is a specific biomedical article from a particular collection and contains images and their captions and mentions.
- A *phrase* is the portion of text that MetaMap maps to UMLS concepts. For example, from the caption in Figure 1, the noun phrase “histologic features” maps to four UMLS concepts: “Histologic,” “Characteristics,” “Protein Domain” and “Array Feature.”
- A *mapping* is an assignment of a phrase to a particular set of UMLS concepts. Each phrase can have more than one mapping.

### 3.2 Features

Using this terminology, we define the following features used to classify potential indexing terms. We refer to these as *non-lexical* features because they generally characterize UMLS concepts, going beyond the surface representation of words and lexemes appearing in the article text.

F.1 *CUI (nominal)*: The Concept Unique Identifier (CUI) assigned to the concept in the UMLS Metathesaurus. We choose the concept identifier as a feature because some frequently mapped concepts are consistently ineffective for indexing the images in our training and evaluation data. For example, the CUI for “Original,” another term mapped from the caption shown in Figure 1, is “C0205313.” Our results indicate that “C0205313,” which occurs 19 times in our evaluation data, *never* identifies a useful indexing term.

F.2 *Semantic Type (nominal)*: The concept’s semantic categorization. There are currently 132 different semantic types<sup>7</sup> in the UMLS Metathesaurus. For example, The semantic type of “Original” is “Idea or Concept.”

F.3 *Presence in Caption (nominal)*: **true** if the phrase that generated the concept is located in the image caption; **false** if the phrase is located in the image mention.

F.4 *MeSH Ratio (real)*: The ratio of words  $c_i$  in the concept  $c$  that are also contained in the Medical Subject Headings (MeSH terms)<sup>8</sup>  $\mathcal{M}$  assigned to the document to the total number of words in the concept.

$$R^{(m)} = \frac{|\{c_i : c_i \in \mathcal{M}\}|}{|c|} \quad (1)$$

MeSH is a controlled vocabulary created by the US National Library of Medicine (NLM) to index biomedical articles. For example, “Adenoma, Sweat” is one MeSH term assigned to “Metastatic Hidradenocarcinoma: Efficacy of Capecitabine” (Thomas et al., 2006), the article containing the image from Figure 1.

F.5 *Abstract Ratio (real)*: The ratio of words  $c_i$  in the concept  $c$  that are also in the document’s abstract  $\mathcal{A}$  to the total number of words in the concept.

$$R^{(a)} = \frac{|\{c_i : c_i \in \mathcal{A}\}|}{|c|} \quad (2)$$

F.6 *Title Ratio (real)*: The ratio of words  $c_i$  in the concept  $c$  that are also in the document’s title  $\mathcal{T}$  to the total number of words in the concept.

$$R^{(t)} = \frac{|\{c_i : c_i \in \mathcal{T}\}|}{|c|} \quad (3)$$

F.7 *Parts-of-Speech Ratio (real)*: The ratio of words  $p_i$  in the phrase  $p$  that have been tagged as having part of speech  $s$  to the total number of words in the phrase.

$$R^{(s)} = \frac{|\{p_i : \text{TAG}(p_i) = s\}|}{|p|} \quad (4)$$

This feature is computed for noun, verb, adjective and adverb part-of-speech tags. We

<sup>6</sup>[http://www.getty.edu/research/conducting\\_research/vocabularies/aat/](http://www.getty.edu/research/conducting_research/vocabularies/aat/)

<sup>7</sup>[http://www.nlm.nih.gov/research/umls/META3\\_current\\_semantic\\_types.html](http://www.nlm.nih.gov/research/umls/META3_current_semantic_types.html)

<sup>8</sup><http://www.nlm.nih.gov/mesh/>

obtain tagging information from the output of MetaMap.

- F.8 *Concept Ambiguity (real)*: The ratio of the number of mappings  $m_i$  of phrase  $p$  that contain concept  $c$  to the total number of mappings for the phrase:

$$A = \frac{|\{m_i^p : c \in m_i^p\}|}{|m^p|} \quad (5)$$

- F.9 *Tf-idf (real)*: The frequency of term  $t_i$  (i.e., the phrase that generated the concept) times its inverse document frequency:

$$\text{tfidf}_{i,j} = \text{tf}_{i,j} \times \text{idf}_i \quad (6)$$

The term frequency  $\text{tf}_{i,j}$  of term  $t_i$  in document  $d_j$  is given by

$$\text{tf}_{i,j} = \frac{n_{i,j}}{\sum_{k=1}^{|\mathcal{D}|} n_{k,j}} \quad (7)$$

where  $n_{i,j}$  is the number of occurrences of  $t_i$  in  $d_j$ , and the denominator is the number of occurrences of all terms in  $d_j$ . The inverse document frequency  $\text{idf}_i$  of  $t_i$  is given by

$$\text{idf}_i = \log \frac{|\mathcal{D}|}{|\{d_j : t_i \in d_j\}|} \quad (8)$$

where  $|\mathcal{D}|$  is the total number of documents in the collection, and the denominator is the total number of documents that contain  $t_i$  (see Salton and Buckley, 1988).

- F.10 *Document Location (real)*: The location in the document of the phrase that generated the concept. This feature is continuous on  $[0, 1]$  with 0 representing the beginning of the document and 1 representing the end.

- F.11 *Concept Length (real)*: The length of the concept, measured in number of characters.

For the purpose of computing F.9 and F.10, we indexed each collection with the Terrier<sup>9</sup> information retrieval platform. Terrier was configured to use a block indexing scheme with a Tf-idf weighting model. Computation of all other features is straightforward.

### 3.3 Classifier

We explored these feature vectors using various classification approaches available in the RapidMiner<sup>10</sup> tool. Unlike many similar text and image

<sup>9</sup><http://ir.dcs.gla.ac.uk/terrier/>

<sup>10</sup><http://rapid-i.com/>

classification problems, we were unable to achieve results with a Support Vector Machine (SVM) learner (libSVMlerner) using the Radial Base Function (RBF). Common cost and width parameters were used, yet the SVM classified all terms as ineffective. Identical results were observed using a Naïve Bayes (NB) learner.

For these reasons, we chose to use the Averaged One-Dependence Estimator (AODE) learner (Webb et al., 2005) available in RapidMiner. AODE is capable of achieving highly accurate classification results with the quick training time usually associated with NB. Because this learner does not handle continuous attributes, we pre-processed our features with equal frequency discretization. The AODE learner was trained in a ten-fold cross validation of our training data.

## 4 Results

Results relating to specific aspects of our work (annotation, features and classification) are presented below.

### 4.1 Inter-Annotator Agreement

Two independent reviewers manually classified the extracted terms from our evaluation data as useful for indexing their associated images or not. The inter-annotator agreement between reviewers A and B is shown in the first row of Table 1. Although both reviewers are physicians trained in medical informatics, their initial agreement is only moderate, with  $\kappa = 0.519$ . This illustrates the subjective nature of manual ABIR and, in general, the difficulty in reliably classifying potential indexing terms for biomedical images.

Annotator	Pr( $a$ )	Pr( $e$ )	$\kappa$
A/B	0.847	0.682	0.519
A/Standard	0.975	0.601	0.938
B/Standard	0.872	0.690	0.586

**Table 1: Inter-annotator Agreement.** The probability of agreement  $\text{Pr}(a)$ , expected probability of chance agreement  $\text{Pr}(e)$ , and the associated Cohen’s kappa coefficient  $\kappa$  are given for each reviewer combination.

After their initial classification, the two reviewers were instructed to collaboratively reevaluate the subset of extracted terms upon which they disagreed (roughly 15% of the terms) and create a

Feature	Gain	$\chi^2$
F.1 CUI	0.003	13.331
F.2 Semantic Type	0.015	68.232
F.3 Presence in Caption	0.008	35.303
F.4 MeSH Ratio	0.043	285.701
F.5 Abstract Ratio	0.023	114.373
F.6 Title Ratio	0.021	132.651
F.7 Noun Ratio	0.053	287.494
Verb Ratio	0.009	26.723
Adjective Ratio	0.021	96.572
Adverb Ratio	0.002	5.271
F.8 Concept Ambiguity	0.008	33.824
F.9 Tf-idf	0.004	21.489
F.10 Document Location	0.002	12.245
F.11 Phrase Length	0.021	102.759

**Table 2: Feature Comparison.** The information gain and chi-square statistic is shown for each feature. A higher score indicates greater influence on term effectiveness.

*gold standard* evaluation. The second and third rows of Table 1 suggest the resulting evaluation strongly favors reviewer A’s initial classification compared to that of reviewer B.

Since the reviewers of the *training* data each classified terms from different sets of randomly selected images, it is impossible to calculate their inter-annotator agreement.

#### 4.2 Effectiveness of Features

The effectiveness of individual features in describing the potential indexing terms is shown in Table 2. We used two measures, both of which indicate a similar trend, to calculate feature effectiveness: Information gain (Kullback-Leibler divergence) and the chi-square statistic.

Under both measures, the MeSH ratio (F.4) is one of the most effective features. This makes intuitive sense because MeSH terms are assigned to articles by specially trained NLM professionals. Given the large size of the MeSH vocabulary, it is not unreasonable to assume that an article’s MeSH terms could be descriptive, at a coarse granularity, of the images it contains. Also, the subjectivity of the reviewers’ initial data calls into question the usefulness of our training data. It may be that MeSH terms, consistently assigned to all documents in a particular collection, are a more reliable determiner of the usefulness of po-

tential indexing terms. Furthermore, the study by Demner-Fushman et al. (2008) found that, on average, roughly 25% of the additional (useful) terms the reviewers added to the set of extracted terms were also found in the MeSH terms assigned to the document containing the particular image.

The abstract and title ratios (F.6 and F.5) also had a significant effect on the classification outcome. Similar to the argument for MeSH terms, as these constructs are a coarse summary of the contents of an article, it is not unreasonable to assume they summarize the images contained therein.

Finally, the noun ratio (F.7) was a particularly effective feature, and the length of the UMLS concept (F.11) was moderately effective. Interestingly, tf-idf and document location (F.9 and F.10), both features computed using standard information retrieval techniques, are among the least effective features.

#### 4.3 Classification

While the AODE learner performed reasonably well for this task, the difficulty encountered when training the SVM learner may be explained as follows. The initial inter-annotator agreement of the evaluation data suggests that it is likely that our training data contained contradictory or mislabeled observations, preventing the construction of a maximal-margin hyperplane required by the SVM. An SVM implementation utilizing soft margins (Cortes and Vapnik, 1995) would likely achieve better results on our data, although at the expense of greater training time. The success of the AODE learner in this case is probably due to its resilience to mislabeled observations.

Annotator	Precision	Recall	$F_1$ -score
A	0.258	0.442	0.326
B	0.200	0.225	0.212
Combined	0.326	0.224	0.266
Standard	0.453	0.229	0.304
Standard <sup>a</sup>	0.492	0.231	0.314
Training	0.502	0.332	0.400

**Table 3: Classification Results.** The classifier’s precision and recall, as well as the corresponding  $F_1$ -score, are given for the responses of each reviewer.

<sup>a</sup>For comparison, the classifier was also trained using the subset of training data containing responses from reviewers A and B only.

Classification results are shown in Table 3. The precision and recall of the classification scheme is shown for the manual classification by reviewers A and B in the first and second rows. The third row contains the results obtained from combining the results of the two reviewers, and the fourth row shows the classification results compared to the gold standard obtained after discovering the initial inter-annotator agreement.

We hypothesized that the training data labels may have been highly sensitive to the subjectivity of the reviewers. Therefore, we retrained the learner with only those observations made by reviewers A and B (of the five total reviewers) and again compared the classification results with the gold standard. Not surprisingly, the  $F_1$ -score of this classification (shown in the fifth row) is somewhat improved compared to that obtained when utilizing the full training set.

The last row in Table 3 shows the results of classifying the training data. That is, it shows the results of classifying one tenth of the data after a ten-fold cross validation and can be considered an upper bound for the performance of this classifier on our evaluation data. Notice that the associated  $F_1$ -score for this experiment is only marginally better than that of the unseen data. This implies that it is possible to use training data from particular subdomains of the biomedical sciences (cardiology and plastic surgery) to classify potential indexing terms in other subdomains (dermatology).

Overall, the classifier performed best when verified with reviewer A, with an  $F_1$ -score of 0.326. Although this is relatively low for a classification task, these results improve upon the baseline classification scheme (all extracted terms are useful for indexing) with an  $F_1$ -score of 0.182 (Demner-Fushman et al., 2008). Thus, non-lexical features can be leveraged, albeit to a small degree with our current features and classifier, in automatically selecting useful image indexing terms. In future work, we intend to explore additional features and alternative tools for mapping text to the UMLS.

## 5 Related Work

Non-lexical features have been successful in many contexts, particularly in the areas of genre classification and text and speech summarization.

Genre classification, unlike text classification, discriminates between document *style* instead of *topic*. Dewdney et al. (2001) show that non-lexical

features, such as parts of speech and line-spacing, can be successfully used to classify genres, and Ferizis and Bailey (2006) demonstrate that accurate classification of Internet documents is possible even without the expensive part-of-speech tagging of similar methods. Recall that the noun ratio (F.7) was among the most effective of our features.

Finn and Kushmerick (2006) describe a study in which they classified documents from various domains as “subjective” or “objective.” They, too, found that part-of-speech statistics as well as general text statistics (e.g., average sentence length) are more effective than the traditional bag-of-words representation when classifying documents from multiple domains. This supports the notion that we can use non-lexical features to classify potential indexing terms in one biomedical subdomain using training data from another.

Maskey and Hirschberg (2005) found that prosodic features (see Ward, 2004) combined with structural features are sufficient to summarize spoken news broadcasts. Prosodic features relate to intonational variation and are associated with particularly important items, whereas structural features are associated with the organization of a typical broadcast: headlines, followed by a description of the stories, etc.

Finally, Schilder and Kondadadi (2008) describe non-lexical word-frequency features, similar to our *ratio* features (F.4–F.7), which are used with a regression SVM to efficiently generate query-based multi-document summaries.

## 6 Conclusion

Images convey essential information in biomedical publications. However, automatically extracting and selecting useful indexing terms from the article text is a difficult task given the domain-specific nature of biomedical images and vocabularies. In this work, we use the manual classification results of a previous study to train a binary classifier to automatically decide whether a potential indexing term is useful for this purpose or not. We use non-lexical features generated for each term with the most effective including whether the term appears in the MeSH terms assigned to the article and whether it is found in the article’s title and caption. While our specific retrieval task relates to the biomedical domain, our results indicate that ABIR approaches to image retrieval in *any* domain can benefit from an automatic annota-

tion process utilizing non-lexical features to aid in the selection of indexing terms or the reduction of ineffective terms from a set of potential ones.

## References

- Sameer Antani, Dina Demner-Fushman, Jiang Li, Balaji V. Srinivasan, and George R. Thoma. 2008. Exploring use of images in clinical articles for decision support in evidence-based medicine. In *Proc. of SPIE-IS&T Electronic Imaging*, pages 1–10.
- Alan R. Aronson. 2001. Effective mapping of biomedical text to the UMLS metathesaurus: The MetaMap program. In *Proc. of the Annual Symp. of the American Medical Informatics Association (AMIA)*, pages 17–21.
- Corinna Cortes and Vladimir Vapnik. 1995. Support-vector networks. *Machine Learning*, 20(3):273–297.
- Dina Demner-Fushman, Sameer Antani, Matthew Simpson, and George Thoma. 2008. Combining medical domain ontological knowledge and low-level image features for multimedia indexing. In *Proc. of the Language Resources for Content-Based Image Retrieval Workshop (OntoImage)*, pages 18–23.
- Dina Demner-Fushman, Sameer K. Antani, and George R. Thoma. 2007. Automatically finding images for clinical decision support. In *Proc. of the Intl. Workshop on Data Mining in Medicine (DM-Med)*, pages 139–144.
- Nigel Dewdney, Carol VanEss-Dykema, and Richard MacMillan. 2001. The form is the substance: Classification of genres in text. In *Proc. of the Workshop on Human Language Technology and Knowledge Management*, pages 1–8.
- George Ferizis and Peter Bailey. 2006. Towards practical genre classification of web documents. In *Proc. of the Intl. Conference on the World Wide Web (WWW)*, pages 1013–1014.
- Aidan Finn and Nicholas Kushmerick. 2006. Learning to classify documents according to genre. *Journal of the American Society for Information Science and Technology (JASIST)*, 57(11):1506–1518.
- F. Florea, V. Buzuloiu, A. Rogozan, A. Bensrhair, and S. Darmoni. 2007. Automatic image annotation: Combining the content and context of medical images. In *Intl. Symp. on Signals, Circuits and Systems (ISSCS)*, pages 1–4.
- Masashi Inoue. 2004. On the need for annotation-based image retrieval. In *Proc. of the Workshop on Information Retrieval in Context (IRiX)*, pages 44–46.
- Judith Klavans, Carolyn Sheffield, Eileen Abels, Joan Beaudoin, Laura Jenemann, Tom Lipincott, Jimmy Lin, Rebecca Passonneau, Tandeep Sidhu, Dagobert Soergel, and Tae Yano. 2008. Computational linguistics for metadata building: Aggregating text processing technologies for enhanced image access. In *Proc. of the Language Resources for Content-Based Image Retrieval Workshop (OntoImage)*, pages 42–47.
- D.A. Lindberg, B.L. Humphreys, and A.T. McCray. 1993. The unified medical language system. *Methods of Information in Medicine*, 32(4):281–291.
- Sameer Maskey and Julia Hirschberg. 2005. Comparing lexical, acoustic/prosodic, structural and discourse features for speech summarization. In *Proc. of the European Conference on Speech Communication and Technology (EUROSPEECH)*, pages 621–624.
- Gerard Salton and Christopher Buckley. 1988. Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, 24(5):513–523.
- Frank Schilder and Ravikumar Kondadadi. 2008. FastSum: Fast and accurate query-based multi-document summarization. In *Proc. of the Workshop on Human Language Technology and Knowledge Management*, pages 205–208.
- Jouary Thomas, Kaiafa Anastasia, Lipinski Philippe, Vergier Béatrice, Lepreux Sébastien, Delaunay Michèle, and TaïebAlain. 2006. Metastatic hidradenocarcinoma: Efficacy of capecitabine. *Archives of Dermatology*, 142(10):1366–1367.
- Nigel Ward. 2004. Pragmatic functions of prosodic features in non-lexical utterances. In *Proc. of the Intl. Conference on Speech Prosody*, pages 325–328.
- Geoffrey I. Webb, Janice R. Boughton, and Zhihai Wang. 2005. Not so naïve bayes: Aggregating one-dependence estimators. *Machine Learning*, 58(1):5–24.