A Controlled Vocabulary to Represent Sonographic Features of the Thyroid and its application in a Bayesian Network to Predict Thyroid Nodule Malignancy

Yueyi I. Liu, PhD1,2, Aya Kamaya, MD1, Terry S. Desser, MD1, Daniel L. Rubin, MD, MS1,2,
1Department of Radiology, Stanford University, Stanford, CA; 2Stanford Medical Informatics, Stanford University, Stanford, CA

Abstract
It is challenging to distinguish benign from malignant thyroid nodules on high resolution ultrasound. Many ultrasound features have been studied individually as predictors for thyroid malignancy, none with a high degree of accuracy, and there is no consistent vocabulary used to describe the features. Our hypothesis is that a standard vocabulary will advance accuracy. We performed a systemic literature review and identified all the sonographic features that have been well studied in thyroid cancers. We built a controlled vocabulary for describing sonographic features and to enable us to unify data in the literature on the predictive power of each feature. We used this terminology to build a Bayesian network to predict thyroid malignancy. Our Bayesian network performed similar to or slightly better than experienced radiologists. Controlled terminology for describing thyroid radiology findings could be useful to characterize thyroid nodules and could enable decision support applications.

Introduction
Thyroid cancer is the 7th most commonly diagnosed malignancy in females. The American Cancer Society estimates that 33,550 new cases of thyroid cancer will be diagnosed in 20071. Thyroid cancers most often present as thyroid nodules. It is crucial to distinguish malignant nodules from benign ones so that early intervention can be performed to reduce morbidity and mortality.

Many sonographic features have been described and studied as potential predictors of thyroid malignancy. These include size, multiplicity, echogenicity, presence of microcalcifications, margin, contour, shape, architecture, and vascularity (see 2,3 for review). In the prior studies of individual features, no consistent terminology was adopted, and different studies examining the same feature use different names. Consequently, it is difficult to combine results across studies for meta-analysis or to build multi-feature models.

Controlled vocabularies are lists of standard terms used to describe a domain. They have many benefits, including facilitating, communication, data mining, and data retrieval. An example is the Medical Subject Headings (MeSH), a controlled vocabulary created and maintained by the National Library of Medicine to index MedLine articles. Controlled terminology is appearing in radiology, such as breast imaging, where the BI-RADS controlled terminology is used to describe mammogram features. Recently, the radiology community built RadLex, a lexicon for uniform indexing and retrieval of radiology information resources4. RadLex has been translated into an ontology, which facilitates computational analysis and applications5. However, currently there is no controlled vocabulary to describe the sonographic features of the thyroid.

We believe that more accurate diagnosis of thyroid nodules can be achieved by creating a model incorporating all the features seen in thyroid ultrasound imaging. Bayesian classifiers are multi-features models that have been used in many areas of medicine. For example, Burnside et al. built a Bayesian classifier to predict breast cancer risk based on mammography findings6. Kline et al. created a classifier to identify a low-risk subset of patients suspected of having a venous thromboembolism using clinical data that are readily available7.

The goal of our work is to create a controlled vocabulary for sonographic features of the thyroid with the goal of providing a means of creating standard “imaging phenotype” descriptions of the images. We also hypothesize that such a controlled vocabulary can be used to drive decision support applications such as a Bayesian network for thyroid nodule classification.

Materials and Methods
We performed a systemic review and identified 16 articles2,8-22 that discussed either one or a number of sonographic features associated with either benign or
malignant thyroid nodules. While reviewing this literature, we unified features that were the same but named differently, resulting in a controlled terminology of features for thyroid ultrasound imaging. For each feature, we identified how they were defined in each article. We also obtained from literature the sensitivity, specificity, and positive predictive values of these features for malignancy. When these data were not directly available from the literature, they were computed from raw data in the article, using the total number of nodules, the number of benign/malignant nodules with the feature, if such data were provided in the articles.

We used the Netica development environment (http://www.norsys.com) to construct our Bayesian classifier and perform inference. We created a Bayesian network (BN) comprising a node for disease and nodes for the observed sonoanographic findings and patient demographics. We represented the pathology of thyroid nodules as a disease node with two states (benign vs. malignant). All sonoanographic features from our controlled terminology known to be predictors of malignancy were included in the BN. Given that age and gender also significantly influence the probability of a nodule being malignant, we included these demographic features in our model as well. The structure of our BN is shown in Figure 1. The pretest probabilities of thyroid malignancies by age and gender were derived from the SEER database (http://seer.cancer.gov/faststats/sites.php?site=Thyroid+Cancer).

To evaluate our classifier, we randomly selected 21 benign thyroid nodules and 20 malignant nodules from 37 patients who underwent ultrasound guided FNA in 2007 and early 2008. All final diagnoses were determined by pathology. The 20 malignant nodules included 18 papillary thyroid carcinomas, one lymphoma, and one poorly differentiated carcinoma. Follicular lesions were not included since FNA, the test used to establish the final diagnosis in our study, cannot distinguish benign follicular adenomas from malignant follicular carcinomas.

We compared the performance of our classifier to that of two radiologists specializing in thyroid ultrasound, one with five years of experience (radiologist 1), the other with 20 years of experience (radiologist 2). They each rated each nodule on a scale of 1 to 5 (1-benign, 2-probably benign, 3-not sure, 4-probably malignant, 5-malignant). Receiver operating characteristic (ROC) curves were generated using the ROCKIT 1.1B software (http://www-radiology.uchicago.edu/krl/KRL_ROC/software_indexx6.htm). One radiologist was aware of the case mix (number of benign versus malignant nodules), and the other was not.

**Results**

We performed a comprehensive literature review and identified 16 articles that discussed sonoanographic features that may help identify malignant thyroid nodules. These features are listed in Table 1.

The vocabulary used to describe lesions was highly varied. This was particularly true for the description of vascularity based on color Doppler patterns. Some articles divided vascularity into three broad categories: avascular, peripheral vascularity, or intrinsic vascularity. Frates et al. listed intrinsic hypovascularity as another class. Innuccilli et al studied only internal vascularity and ranked it for each nodule on a scale of 0-4 in order of increasing flow, with 0 corresponding to avascular, approximately 25% or less, 26% to 50%, 51% to 75%, and greater than 75% on color Doppler cross-sectional imaging. Chan et al. listed intrinsic vascularity and peripheral vascularity as another class. Innuccilli et al studied only internal vascularity and ranked it for each nodule on a scale of 0-4 in order of increasing flow, with 0 corresponding to avascular, approximately 25% or less, 26% to 50%, 51% to 75%, and greater than 75% on color Doppler cross-sectional imaging. Frates et al. also characterized the color flow of each nodule into 4 types, with 0 being avascular, 1 for minimal internal flow without a peripheral ring, 2 for a peripheral ring of flow (defined as >25% of the nodule’s circumference) but minimal or no internal flow, 3 for a peripheral ring of flow and a small to moderate amount of internal flow, and 4 for extensive internal flow with or without a peripheral ring.

To create a controlled vocabulary for the sonoanographic features of thyroid nodules, we selected only those descriptors that have been used in multiple studies (those highlighted in bold in table 1) (Fig. 1). We obtained the conditional probability table of the Bayesian network by two independent methods. In the first method, an aggregate sensitivity and specificity of each feature was calculated as the weighted average of these parameters from all the articles from which sensitivity and specificity are available, using our controlled terminology to unify features that were the same but named differently in each article. The weight used was the number of thyroid nodules studied in each article, so that the aggregate values were biased towards large studies. In the second method, we had an experienced radiologist supply conditional probabilities from her experience. These two sources of conditional probably agree remarkably (data not shown), except in the case of vascularity, which is most likely due to the varied definition of vascularity in the literature. As a result, we used the conditional probability table from the expert radiologist.

Our model is evaluated using 41 thyroid nodules from 37 patients. ROC curves of our classifier and
the radiologists’ predictions are shown in Figure 2. The area under the curve (Az) value of our model is 0.851 (95% confidence interval (CI): 0.745-0.939), which is similar to or slightly better than those of the radiologists (0.846 (CI: 0.678-0.943) for radiologist 1 and 0.719 (CI: 0.543-0.854) for radiologist 2). Using the classifier for decision support and choosing decision threshold of p=0.2 probability of malignancy (4 of 20 malignant nodules could be missed), only 5 of 21 (24%) benign nodules would undergo biopsy, implying PPV = 76%. At the p=1.0 decision threshold of malignancy (all malignant nodules would be biopsied), 16 out of 21 benign nodules (76%) would undergo biopsy, implying PPV = 56%.

Discussion
Ultrasound presents a rich set of features that are useful for diagnosing thyroid disease. Much of the literature has focused on evaluating one or a few features for the task of recognizing thyroid cancer. We wanted to use the full range of features to create a decision support model in evaluating thyroid nodules. However, creating this model was thwarted by the lack of controlled terminology for describing features.

We performed a comprehensive literature review of sonographic features implicated in thyroid malignancy to create a controlled vocabulary using the common features. Controlled vocabularies have been advocated to standardize radiology reporting, to facilitate communication, data retrieval, and data analysis. However, our results demonstrate the benefit of controlled terminology for creating a decision support application—in our case, a BN to help diagnose malignant thyroid nodules. In our initial evaluation of the BN using 41 nodules from 37 patients, our classifier performed similarly or slightly better than expert radiologists. One of the radiologists (radiologist 2) evaluated the ultrasound images completely blinded. The other radiologist (radiologist 1), though unaware of the final diagnosis of each nodule, was familiar with the cases by enumerating the sonographic features of each nodule. Hence, radiologist 1 was likely biased by awareness of the prior probability of a nodule being malignant in our test cases. We plan to undertake a more thorough evaluation in which only cases which have never been seen by either radiologist will be used to reduce potential bias.

One limitation of our controlled vocabulary is that it includes only features that are well studied. In the future, as the implication of other features become more evident, a more comprehensive vocabulary can be developed. Adopting controlled terminology for reporting imaging features could have benefit in other domains of imaging in terms of driving decision support, not only radiology but also potentially pathology.

References
1 Jemal, A. et al., Cancer statistics, 2007. CA Cancer J Clin 57 (1), 43-66 (2007).
2 Frates, M.C. et al., Management of thyroid nodules detected at US: Society of Radiologists in Ultrasound consensus conference statement. Radiology 237 (3), 794-800 (2005).
3 Desser, T.S. & Kamaya, A., Ultrasound of thyroid nodules. Neuroimaging Clin N Am 18 (3), 463-478 (2008).
4 Langlotz, C.P., RadLex: a new method for indexing online educational materials. Radiographics 26 (6), 1595-1597 (2006).
5 Rubin, D.L., Creating and Curating a Terminology for Radiology: Ontology Modeling and Analysis. J Digit Imaging (2007).
6 Burnside, E., Rubin, D., & Shachter, R., A Bayesian network for mammography. Proc AMIA Symp, 106-110 (2000).
7 Kline, J.A., Novobilski, A.J., Kabrhel, C., Richman, P.B., & Courtney, D.M., Derivation and validation of a Bayesian network to predict pretest probability of venous thromboembolism. Ann Emerg Med 45 (3), 282-290 (2005).
8 Alexander, E.K. et al., Natural history of benign solid and cystic thyroid nodules. Ann Intern Med 138 (4), 315-318 (2003).
9 Cappelli, C. et al., The predictive value of ultrasound findings in the management of thyroid nodules. QJM 100 (1), 29-35 (2007).
10 Chan, B.K., Desser, T.S., McDougall, I.R., Weigel, R.J., & Jeffrey, R.B., Jr., Common and uncommon sonographic features of papillary thyroid carcinoma. J Ultrasound Med 22 (10), 1083-1090 (2003).
11 Frates, M.C., Benson, C.B., Doubilet, P.M., Cibas, E.S., & Marqusee, E., Can color Doppler sonography aid in the prediction of malignancy of thyroid nodules? J Ultrasound Med 22 (2), 127-131; quiz 132-124 (2003).
12 Frates, M.C. et al., Prevalence and distribution of carcinoma in patients with solitary and multiple thyroid nodules on sonography. J Clin Endocrinol Metab 91 (9), 3411-3417 (2006).
Iannuccilli, J.D., Cronan, J.J., & Monchik, J.M., Risk for malignancy of thyroid nodules as assessed by sonographic criteria: the need for biopsy. J Ultrasound Med 23 (11), 1455-1464 (2004).

Kakkos, S.K. et al., Relative risk of cancer in sonographically detected thyroid nodules with calcifications. J Clin Ultrasound 28 (7), 347-352 (2000).

Khoo, M.L., Asa, S.L., Witterick, I.J., & Freeman, J.L., Thyroid calcification and its association with thyroid carcinoma. Head Neck 24 (7), 651-655 (2002).

Kim, E.K. et al., New sonographic criteria for recommending fine-needle aspiration biopsy of nonpalpable solid nodules of the thyroid. AJR Am J Roentgenol 178 (3), 687-691 (2002).

Kim, J.Y. et al., Radiologic and pathologic findings of nonpalpable thyroid carcinomas detected by ultrasonography in a medical screening center. J Ultrasound Med 27 (2), 215-223 (2008).

Koike, E. et al., Ultrasonographic characteristics of thyroid nodules: prediction of malignancy. Arch Surg 136 (3), 334-337 (2001).

Papini, E. et al., Risk of malignancy in nonpalpable thyroid nodules: predictive value of ultrasound and color-Doppler features. J Clin Endocrinol Metab 87 (5), 1941-1946 (2002).

Rago, T., Santini, F., Scutari, M., Pinchera, A., & Vitti, P., Elastography: new developments in ultrasound for predicting malignancy in thyroid nodules. J Clin Endocrinol Metab 92 (8), 2917-2922 (2007).

Wienke, J.R., Chong, W.K., Fielding, J.R., Zou, K.H., & Mittelstaedt, C.A., Sonographic features of benign thyroid nodules: interobserver reliability and overlap with malignancy. J Ultrasound Med 22 (10), 1027-1031 (2003).

Ahuja, A., Chick, W., King, W., & Metreweli, C., Clinical significance of the comet-tail artifact in thyroid ultrasound. J Clin Ultrasound 24 (3), 129-133 (1996).

Figure 1. Our Bayesian classifier for thyroid nodules. All possible values for each node are listed.

Figure 2. ROC curves for the Bayesian classifier and two radiologists.
Table 1. A systemic literature review revealed many sonographic features that may help distinguish malignant thyroid nodules from benign ones. Those in bold are terms selected for our controlled vocabulary.

| Category         | Feature                          | References                  |
|------------------|----------------------------------|-----------------------------|
| Calcification    | All calcifications               | 14,15                      |
|                  | Punctuate calcification          | 12                          |
|                  | Coarse calcification             | 10,12                      |
|                  | Rim-only calcification           | 12,17                      |
| Microcalcification|                                 | 2,9,10,16-20               |
| Margin           | Irregular margin                 | 10,16-18                   |
|                  | Blurred margin                   | 19                          |
| Smooth margin    |                                  |                             |
|                  | Microlobulated margin            | 21                          |
|                  | Macrolobulated margin            | 21                          |
| Ill-defined margin|                                | 12                          |
| Shape            | Taller than wide                 | 2,13,16,17                 |
|                  | Irregular shape                  | 10,17,18                   |
|                  | Round/oval in shape              | 10,17                      |
| Echogenicity     | Hypoechoic                       | 2,9,12,13,17,19-21         |
|                  | Marked hyperechoic               | 16,17,21                   |
|                  | Mixed hypoechoic/isoechoic       | 10                          |
|                  | Heterogenous internal echogenicity| 18                          |
| Architecture     | Solid                            | 2,8,10-13,17,19-21         |
|                  | Almost solid (<25% cystic)       | 8,12,21                    |
|                  | Mixed (25-75% cystic)            | 12                          |
|                  | Cystic                           | 8,10-13,21                 |
| Invasion         | Extracapsular invasion present   | 17-19                      |
|                  | Nodal metastasis                 | 17,19                      |
| Vascularity      | Intrinsic vascularity            | 9-11,13,21                 |
|                  | Perinodular vascularity          | 9,10,21                    |
|                  | Hypovascular                     | 10                          |
|                  | Avascular                        | 9,11,13,21                 |
| Halo             | Halo present                     | 2,10,12,18                 |
|                  | Halo present >=50%               | 12                          |
|                  | Halo present <50%                | 12                          |
| Ring down artifact| Ring down artifact present       | 22                          |