Estimating the probability of malignancy of thyroid nodules using adjusted naïve Bayes

Boas FE, Liu YI, Kamaya A, Desser TS, and Rubin DL
2012-03-04

Abstract

Purpose. Determining the malignant potential of thyroid nodules using ultrasound can be challenging. It was previously shown that the probability of malignancy can be estimated using a naïve Bayes model, based on a nodule's ultrasound features (microcalcifications, taller than wide, margin, capsular invasion, architecture, echogenicity, ring down artifact, and vascularity). However, probabilities generated by naïve Bayes are too confident, because these features contain redundant information. Our goal was to improve this model.

Materials and Methods. Using a dataset published by Liu et al (98 nodules evaluated by 2 radiologists), we developed an adjusted naïve Bayes model, in which the likelihood ratios from naïve Bayes are downweighted in order to improve the fit between the calculated probability of malignancy and the actual probability of malignancy. We also examined subsets of nodules in the dataset that have identical features. (In an ideal model, the probability of malignancy would be determined from other nodules with identical features.) Most of these subsets are either 100% malignant or 100% benign, but many of them contain a mixture of benign and malignant nodules. Based on this information, we calculated the theoretical maximum sensitivity and specificity for any model using this set of features.

Results. An adjusted naïve Bayes model in which the likelihood ratios are raised to the 0.50 power yields accurate predicted probabilities, without changing the ROC curve. This model has a sensitivity of 68% and specificity of 73% for detecting malignant thyroid nodules. In comparison, the theoretical maximum sensitivity and specificity for any model using this set of features is 77%. Finally, we developed a simple online calculator for thyroid nodule malignancy (http://www.stanford.edu/~boas/calc/thyroid.html).

Conclusion. Adjusted naïve Bayes is close to optimal for calculating the probability of malignancy of a thyroid nodule, given the current set of features. The concepts in this abstract can easily be applied to other classification problems.

Clinical Relevance (134 characters; max 200 characters). The probability of malignancy of a thyroid nodule can be estimated based on its ultrasound features, using a simple online calculator.

Introduction

Liu et al estimated the sensitivity and specificity of various ultrasound features of thyroid nodules for detecting papillary thyroid carcinoma, using a literature review and expert opinion. They examined the following features: microcalcifications, taller than wide, margin (smooth, irregular, or ill defined), capsular invasion, architecture (solid, almost solid, mixed, or cystic), echogenicity (hypoechoic, isoechoic, and hyperechoic), ring down artifact, and vascularity (intrinsic, perinodular, or avascular). They then developed a naïve Bayes model, where the pre-test probability of malignancy was based on the patient's age and gender, and the post-test probability of malignancy was determined using the sensitivity and specificity of each feature. Two radiologists each examined 99 nodules (98 of which have complete data), and they showed that the naïve Bayes model has a similar sensitivity and specificity (ROC curve) for detecting malignancy nodules as a radiologist.

Adjusted naïve Bayes

The naïve Bayes model described above assumes that each feature is independent. In reality, a nodule with one benign feature is more likely to have another benign feature, and a nodule with one malignant feature is more likely to have another malignant feature. Thus, naïve Bayes overcounts the evidence, generating probabilities that are a little too confident. For example, a calculated probability of malignancy of 2% corresponds to an actual probability of 11%, and a calculated probability of 99% corresponds to an actual probability of 89%.

This can be fixed by downweighting the likelihood ratios from naïve Bayes by a constant factor (see Appendix for details). For the Liu data set, the optimum value of this factor was 0.50 (95% confidence interval: 0.36 to 0.76). This means that on average, naïve Bayes is double counting evidence for benignity or malignancy. The adjusted naïve Bayes model is given in Table 1.

There are two ways to evaluate the adjusted naïve Bayes model. The ROC curve, which gives sensitivities and specificities, tells how well the model can distinguish benign from malignant nodules. The calibration curve shows how well calculated probabilities match the actual probabilities. For example, a model that always produces a probability of malignancy of 0.4 for benign nodules, and 0.6 for malignant nodules would have a perfect ROC curve, but inaccurate probabilities. On the other hand, a model that always gives the pre-test probability of malignancy for every nodule would have a terrible ROC curve, but accurate probabilities. Adjusted naïve Bayes improves the accuracy of calculated probabilities, but does not affect the ROC curve (Figure 1).

Feature		**Value**
Demographics	Age < 50, Male	-1.39
	Age < 50, Female	-1.99
	Age ≥ 50, Male	-0.85
	Age ≥ 50, Female	-1.39
Microcalcification	Yes	0.74
	No	-0.35
Taller-than-wide	Yes	0.72
	No	-0.09
Margin	Smooth	-0.44
	Irregular	0.20
	Ill-defined	0.08
Capsular invasion	Yes	3.09
	No	-0.14
Architecture	Solid	0.25
	Almost solid (< 25% cystic)	-0.27
	Mixed (25–75% cystic)	-0.61
	Cystic (> 75% cystic)	-0.84
Echogenicity	Hypoechoic	0.21
	Isoechoic	-0.24
	Hyperechoic	-0.93
Ring-down artifact	Yes	-3.70
	No	0.04
Vascularity	Intrinsic	0.16
	Perinodular	0.05
	Avascular	-1.50

Table 1. Adjusted naïve Bayes model for thyroid nodule malignancy. Add up the values for each feature to get the log odds of malignancy (x). The probability of malignancy then equals $e^x/(1+e^x)$.

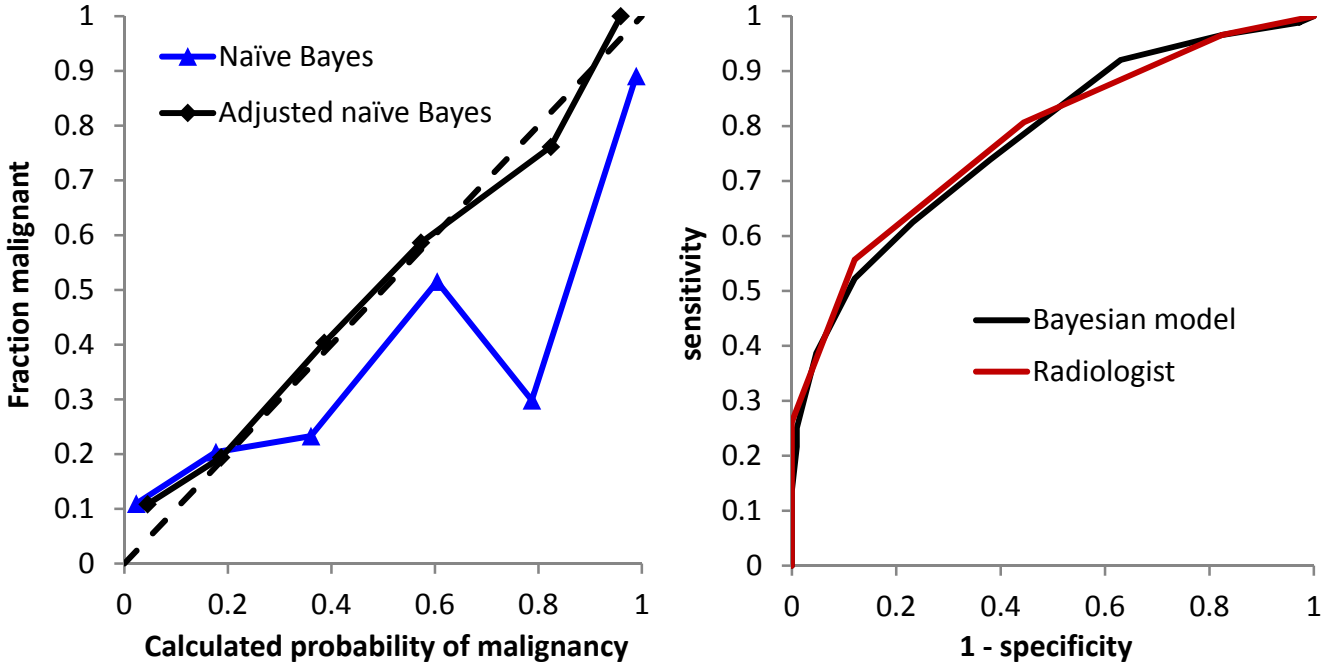

Figure 1. Left: Adjusted naïve Bayes improves the accuracy of calculated probabilities (p<0.05). Each point corresponds to a range of calculated probabilities: 0-10%, 10-30%, 30-50%, 50-70%, 70-90%, and 90-100%. Right: The adjustment factor does not change the relative ordering of probabilities, so the ROC curve is the same with naïve Bayes and adjusted naïve Bayes. The Bayesian model has a similar performance as a radiologist.

Ideal model

An ideal model would give the probability of malignancy based on other nodules with the exact same set of features. There are 2×2×3×2×4×3×2×3 = 1728 possible combinations of features (excluding demographics; see Table 1), so we do not have enough data to construct the ideal model using the current set of features. However, in the Liu data set, 64 of the 98 nodules with complete data (for one of the radiologists) have another nodule in the data set with exactly identical features (excluding demographics). This allows us to calculate the expected accuracy of the ideal model.

If each subset of nodules with identical features is either 100% benign or 100% malignant, then the accuracy, sensitivity, and specificity of the ideal model is 100%. If each subset is either x·100% benign or x·100% malignant, then the accuracy, sensitivity, and specificity of the model will be x·100%. The maximum likelihood estimate of x is 77%. (This was calculated by adjusting x to maximize the probability of seeing the observed distribution of benign versus malignant nodules in subsets of nodules with identical features.) By comparison, naïve Bayes or adjusted naïve Bayes can achieve a sensitivity of 68% and specificity of 73%, which is only slightly less than the ideal model. None of the subsets of nodules with identical features had a statistically significant difference between the fraction malignancy and the probability of malignancy calculated using adjusted naïve Bayes.

Thus, adjusted naïve Bayes is close to optimal for calculating the probability of malignancy of a thyroid nodule, given the current set of features. Further improvements to the accuracy of this model will require more features.

Online calculator

The adjusted naïve Bayes model is available at http://www.stanford.edu/~boas/calc/thyroid.html. As you select information about a thyroid nodule, it dynamically updates the probability of malignancy. When you only select a single feature, the calculator gives you the probability from Bayes theorem. When you have entered in all of the features, it gives you adjusted naive Bayes. With an intermediate number of features, it gives you both naive Bayes and adjusted naive Bayes (presumably the actual probability is between these two values).

Appendix

According to Bayes theorem:

$$\frac{P(A|B)}{P(\sim A|B)} = \frac{P(B|A)}{P(B|\sim A)} \cdot \frac{P(A)}{P(\sim A)}$$

Then, the odds of malignancy *A* given multiple independent features B_i is given by:

$$Odds(A|B_1 \cap B_2 \cap \ldots \cap B_n) = (\prod_{i=1}^{n} C(B_i)) \cdot Odds(A)$$

where

$$C(B_i) = \left(\frac{P(B_i|A)}{P(B_i|\sim A)}\right)^w$$

In naïve Bayes, w=1, and $C(B_i)$ is called the likelihood ratio. (In other words, you multiply the pre-test odds by the likelihood ratio to get the post-test odds.) In adjusted naïve Bayes, w is adjusted to account for redundancy in the features. In logistic regression, each coefficient $C(B_i)$ is optimized individually, which can account for simple correlations between features. Logistic regression does not significantly improve the ROC curve (data not shown).

For both adjusted naïve Bayes and logistic regression, the parameters of the model are adjusted to minimize the sum of squared differences between the calculated probability of malignancy, and the actual pathology (1=malignant, 0=benign). Confidence intervals are calculated by performing this optimization on multiple simulated data sets that are generated by sampling random nodules with replacement. We pooled data from both radiologists, and counted each case of malignancy as 0.354 cases, in order to match the pre-test probability of malignancy.

Reference

Liu YI, Kamaya A, Desser TS, and Rubin DL. (2011) "A Bayesian Network for Differentiating Benign From Malignant Thyroid Nodules Using Sonographic and Demographic Features." *AJR* 196(5): W598-W605.