

# Biomedical Applications of Uncertainty Modeling and Analysis with Fuzzy Receiver Operating Characteristic Methodology

James M. DeLeo and Gregory Campbell

National Institutes of Health  
Bethesda, Maryland USA

## Abstract

*Receiver Operating Characteristic (ROC) methodology is popular and well-established in medicine and other disciplines. It applies where a variable is used for two-state classifications. Fuzzy ROC methodology, introduced earlier by the authors as an extension of ROC methodology for fuzzy classifications, is further developed in this paper. Of particular interest here is the idea of modeling an individual's fuzzy membership with one or many, numerical or ordinal values, the use of the bootstrap to obtain nonparametric confidence intervals for all ROC plot parameters, the notion of using ROC-derived fuzzy functions to estimate degree of membership when a gold standard is unavailable, and the use of fuzzy ROC methodology to enhance artificial neural network performance. Medical examples with fever, macrosomic newborns, and adverse drug reactions as fuzzy membership classes are used to demonstrate basic concepts.*

## 1. Introduction

Receiver Operating Characteristic (ROC) methodology is popular and well-established in medicine [3, 13, 19]. It applies where a single variable,  $x$ , is used for classifying individual instances into one of two states ( $d=0$  or  $d=1$ ). Fuzzy ROC methodology, recently introduced as an extension of ROC methodology [5, 7-10], is further developed in this paper. Software used for the studies reported here is described in Appendix A.

## 2. What is fuzzy?

It has been made clear that modeling decision processes can be enhanced by using fuzzy membership [17, 18], especially in medicine [1] and other disciplines in which categories that are naturally fuzzy are often treated as dichotomous. For example, although "fever" is fuzzy and

individuals have fuzzy degrees of membership in "fever", an analyst could model "fever" as dichotomous and assign dichotomous membership values,  $d=0$  or  $d=1$ , to individuals. It will be demonstrated later that to do so is to discard information and thereby to sacrifice power. Fuzzy membership may be expressed as a single value or as a distribution and it may be numeric ( $0 \leq d \leq 1$ ) or ordinal. A distribution of fuzzy membership values occurs when more than one assessment is made, such as by a panel of experts. Thus, an individual might have a single 0.8 value in "fever" or a distribution of values centered about 0.8. Alternatively, ordinal classes such as "very low", "low", "medium", "high", and "very high" may be used as fuzzy membership values for "fever".

## 3. Fuzzy ROC methodology

ROC methodology was recently extended by allowing fuzzy membership values, the  $d$ 's, to be in the 0 to 1 interval. [5, 8, 10]. This was accomplished by expanding definitions of sensitivity and specificity to include both fuzzy and dichotomous membership values. It is as if each fuzzy observation is divided into two observations; one having full membership in the class with weight  $d$ , and the other having full non-membership with weight  $(1-d)$ . Then the cumulative distribution function using as weights the  $d$ 's is the fuzzy sensitivity and the cumulative distribution function with  $(1-d)$ -weights is the fuzzy specificity. As the threshold is varied over the range of  $x$ , pairs of sensitivity and 1-specificity values can be plotted to obtain the empirical fuzzy ROC plot. Each point on this plot then corresponds to an interval in  $x$ .

## 4. Fuzzy gold standards

ROC methodology requires gold standards for "true" event classification. Often gold standards are expressed as functional relationships between a reliable quantitative variable and class membership. For example, measurements of body temperature from an accurate digital

thermometer might be used to define a gold standard for "fever". A single threshold value could define a dichotomous gold standard and a monotonic function could define a fuzzy gold standard (see Figure 1A). Gold standards may also be expressed without mathematical functions such as when an expert directly assigns dichotomous or fuzzy membership values (see Figure 3A). Later it will be shown how to use ROC plots to develop useful x-based fuzzy functions that could substitute as gold standards.

## 5. Fever example

In this contrived example, body temperature measurements in degrees Fahrenheit were made with digital and mercury thermometers on 100 sequential emergency room arrivals. The objective was to use the digital thermometer as a gold standard for evaluating the mercury thermometer. A fuzzy gold standard was defined using 98.6 to 100.6 degrees as the range in which fuzzy membership increases linearly from 0 to 1, and a dichotomous gold standard was selected using a 99.6 degree threshold. Both gold standards are plotted in Figure 1A with the abscissa read as the digital thermometer scale. A fuzzy membership data set composed of 100 mercury thermometer readings and degree of membership values from the fuzzy gold standard was generated; it is plotted in Figure 1A. (From this data set, a dichotomous membership data set was constructed by converting fuzzy membership values below 0.5 to  $d = 0$ , and those at and above 0.5 to  $d = 1$ .) The ROC plot for the fuzzy gold standard data set is shown in Figure 1B. Also displayed in this figure is the "perfect" fuzzy ROC plot which is the one that has the largest possible area for the degree of membership values present in the data set. It is computed using a surrogate x-variable that is perfectly correlated with the fuzzy d-values.

## 6. Glucose example

Gestational diabetes indicated by high glucose levels in expectant mothers could signal macrosomic (big) newborns as well as pregnancy and birthing complications. In this example, a fuzzy gold standard for macrosomia degree of membership based on infant birth weight was defined using weights from 3.3 to 4.0 kilograms as the range in which fuzzy membership increases linearly from 0 to 1 as suggested by a Dubowitz scale of birth weights for term newborns [2, 14]. Plasma glucose values in milligrams per deciliter for expectant mothers,  $x$  and the corresponding weights of their full-term babies for 250 cases were extracted from the published data of Tallarigo et al [16]. The glucose and

associated macrosomia degree of membership values are plotted in Figure 2A, and the fuzzy ROC and "perfect" fuzzy ROC plots are shown in Figure 2B.

## 7. Adverse drug reaction example

This example concerns adverse drug reaction (ADR) detection and it is presented to illustrate ordinal fuzziness and fuzzy gold standards derived through expert panel assessments. An ADR is any untoward event related causally to an intended therapeutic agent [15]. ADRs can be thought of as discrete or fuzzy events. A methodology using panel consensus, Delphi polling, and artificial neural networks to establish ADR gold standards is being developed at NIH. In one study, ten clinical pharmacists independently assessed the presence of ADRs in 100 computer-simulated patient cases represented by 8 covariates. The pharmacists rated ADR presence with the following 5-point ordinal scale: A=most unlikely, B=unlikely, C=uncertain, D=possible, E=probable. A "Delphi" meeting was held at which the pharmacists discussed and modified divergent ADR ratings. For the present example, the ordinal fuzzy ADR ratings of nine pharmacists were used to compute gold standard ratings for each of the 100 cases. A numerical scale of A=0.05, B=0.20, C=0.50, D=0.80, and E=0.95 was used to convert the ADR ordinal ratings to numerical degrees of membership. Gold standard ratings for each case were computed as the average value for the panel of nine. These ratings were employed to evaluate the performance of the tenth pharmacist. The data are plotted in Figure 3A and the ROC plots in Figure 3B.

## 8. Area under the fuzzy ROC plot

The area under the fuzzy ROC plot can be used to quantify the global performance of the classification variable,  $x$ . For the fever example the area is 0.892 and for the glucose example it is 0.619. The "perfect" ROC areas,  $A_p$ , are 0.934 and 0.966 respectively. Whenever fuzzy membership values are present, maximum area is restricted to the area under the "perfect" ROC plot. One way to adjust for this restriction is to convert the fuzzy ROC area,  $A$ , to a normalized area,  $A_n$ , as follows:

$$A_n = 0.5 + (A - 0.5) / [2(A_p - 0.5)] , \quad (1)$$

which produces normalized areas of 0.952 and 0.628 respectively for the examples. Note that for fuzzy data, when  $A = A_p$ , then  $A_n = 1$ , and that for dichotomous data,  $A_p = 1$ , so that  $A_n = A$ .

## 9. Bootstrap confidence intervals

The bootstrap is an important statistical device used here for estimating fuzzy ROC parameters such as sensitivity and specificity with confidence intervals (see Appendix B). In the fever example, at 99.8 degree on the mercury thermometer, the sensitivity is 0.741 and the specificity is 0.877 (see Figure 1B). The 95% confidence intervals based on 2000 bootstrap samples are (0.647, 0.826) for sensitivity and (0.808, 0.937) for specificity. In the glucose example, at a glucose level of 77, the sensitivity is 0.931 and the specificity is 0.248 (see Figure 2B). The 95% confidence intervals based on 2000 bootstrap samples are (0.872, 0.966) for sensitivity, and (0.193, 0.317) for specificity.

The bootstrap can also be used to obtain confidence intervals for the area under the fuzzy ROC plot. For the fever example, the 95% confidence interval for the (non-normalized) fuzzy ROC plot area based on 2000 bootstrap samples is (0.852, 0.928), and for the glucose example it is (0.557, 0.676). The latter suggests that the area in the glucose example is different from 0.5 and hence glucose is somewhat effective in predicting macrosomia, a notion that could be formally tested. One way to do this would be to compute a z-score as follows:

$$z = (A - 0.5)/sd(A) , \quad (2)$$

in which the standard deviation of the area,  $sd(A)$ , is also estimated with the bootstrap. This is  $z = (0.619 - 0.500)/0.031 = 3.82$  for the glucose ROC plot. The bootstrap can also be used to compare fuzzy ROC plots based on different  $x$ 's using areas [7].

## 10. Dichotomous versus fuzzy

Sensitivity and specificity values corresponding to a particular  $x$ -value are not directly comparable across ROC plots based on different gold standards, nor are ROC areas. However, the z-score introduced above could be used for area comparisons. Whereas for the glucose example  $z = 3.82$  using the fuzzy gold standard, for the dichotomous one it is  $z = (0.618 - 0.500)/0.037 = 3.22$ . This suggests not only that glucose predicts macrosomia, but also that the fuzzy degree of membership gold standard seems to have more power than the dichotomous one.

## 11. ROC plot slopes

The slope of a line segment on the fuzzy ROC plot is equal to the ratio of  $d$ 's to  $(1-d)$ 's for the value or values of  $x$  that correspond to that line segment. In the ADR example, at the E=probable rating, the slope,  $s$ , is 10.2.

This can be converted to a degree of membership value,  $f$ , as follows:

$$f = s/(1+s) . \quad (3)$$

The associated degree of membership for the panel rating is 0.91. The bootstrap can be used to obtain confidence intervals for these values. The confidence interval for the slope is (8.43, 13.16) and for the fuzzy degree of membership is (0.894, 0.927) based on 2000 bootstrap samples.

## 12. Fuzzy prevalence and decision rules

Medical decisions based on dichotomous gold standards can also be based on fuzzy gold standards when they are known. In the dichotomous case for a known prevalence,  $p$ , and a known cost error tradeoff ratio,  $c$  (false positive cost divided by false negative cost), a line segment with slope,  $s$ , given by

$$s = c(1-p)/p , \quad (4)$$

is lowered to the ROC plot from above and intersects the ROC plot at a point that corresponds to a decision rule. In the fuzzy case, prevalence is the proportion of weighted  $d$ 's in the population. If the original data reflect this prevalence, as in the glucose example, then it can be estimated from the data. Here the sum of the  $d$ 's is 89.2 and the total is 250, so fuzzy prevalence is 0.357. For a non-fuzzy rule with  $c = 0.25$ , the slope is  $s = 0.25(0.643/0.357) = 0.450$  and the point on the fuzzy ROC plot corresponds to specificity of 0.248 and sensitivity of 0.931, circled in Figure 2B. This notion is now carried over to the generation of a degree of membership function based on the diagnostic variable and not the gold standard variable from which the original function was constructed. This point on the ROC plot corresponds to the glucose interval (75, 77] so if the glucose level of the mother is larger than 77 mg/dl, the decision is in favor of macrosomia and the appropriate action is taken. Note that on Figure 2A, there is a corresponding line segment on to this degree of membership function; it is 0.29 at  $x = 75$  and 0.51 at 77 and then linearly interpolated.

## 13. ROC-derived fuzzy functions

In general, each straight line with slope,  $s$ , ranging from 0 to infinity, when lowered to the empirical ROC plot will intersect a point on the convex hull [8] of the ROC plot. In general, this point corresponds to an interval of  $x$ -values. Slopes can be mapped into values on

the 0 to 1 interval with equation (3) above. These  $f$ -values can be thought of as fuzzy degree of membership values, and the function  $f$  versus  $x$  as a fuzzy function used to scale data input to classifiers, and in certain cases, to substitute for a gold standard.

## 14. Artificial neural networks

Artificial neural networks (ANNs) have been used effectively as discrete classifiers where each case is assigned to one class, and fuzzy classifiers where each case is assigned to all classes with fuzzy degrees of membership. As suggested above, ROC-derived fuzzy functions can be used to scale ANN input data. Also, fuzzy ROC methodology may be applied to ANN output classification variables to assess ANN performance and to enhance ANN decision making by accounting for prevalence and misclassification costs [9, 11]. The latter is particularly true when an ANN is trained as a fuzzy classifier using fuzzy class membership data, in which case the ANN behaves more like a function estimator than a discriminator.

## 15. Conclusions

Fuzzy ROC methodology has been further developed here. In particular, the idea of modeling fuzzy membership with single or multiple, numerical or ordinal values, the use of the bootstrap to obtain nonparametric confidence intervals for all ROC plot parameters, the notion of using ROC-derived fuzzy functions to estimate degree of membership if the gold standard is unavailable, and the use of fuzzy ROC methodology to enhance artificial neural network performance have been presented with medical examples about fever, macrosomic newborns, and adverse drug reactions to demonstrate practical utility.

## Appendix A: Software

The PC-based software used here to compute fuzzy ROC plot parameters and their standard deviations and nonparametric confidence intervals with the bootstrap is under continuing development by one of the authors (JMD). The software provides a choice of linear congruential random number generators and sampling methods including with and without replacement, and with and without class control.

## Appendix B: The bootstrap

The bootstrap is a recent computer intensive resampling statistical tool [12] that is used here for the

first time with fuzzy ROC methodology. The basic idea is to produce a bootstrap sample of size  $n$ , using sampling with replacement of the original sample of size  $n$ . For each of  $B$  such samples, the statistic of interest, such as the area under the ROC plot, is calculated. The standard deviation computed with these  $B$  values is an estimate of the standard deviation of the population statistic. These  $B$  values can also be used to generate a nonparametric percentile confidence interval; for example, a 95% confidence for ROC area is obtained as the 2.5 and the 97.5 percentiles of the distribution of the  $B$  area values. The bootstrap is effective for data sets that adequately represent the population, for then the bootstrap distribution of a statistic approximates its population distribution [12]. It can be utilized in ROC methodology to estimate standard deviations and nonparametric confidence intervals for sensitivity, specificity, the area under the ROC plot, and slopes on the ROC plot. For ROC analysis of dichotomous classification data, bootstrap confidence intervals have advantages over large sample statistical approximations [6]. For fuzzy membership data, the statistical distribution theory is not well-known and the bootstrap becomes an invaluable tool. It is important that the bootstrap procedure be used in a way that reflects how the data were collected.

## Acknowledgements

The authors wish to express their gratitude to Drs. Clifford Eng and Colleen Gilstad at the National Naval Medical Center for their help with the glucose example, to Drs. Frank Pucino, Karim Calis, and the other clinical pharmacy staff at the National Institutes of Health Clinical Center Pharmacy Department who participated in the successful collaboration that produced the data used in the adverse drug reaction example, and to Justin Sands for his help in preparing the figures.

## References

- [1] Adlassnig K-P. A Fuzzy Logical Model of Computer-Assisted Medical Diagnosis. *Methods of Information in Medicine* 1980;19:141-148.
- [2] Battaglia FC, Lubchenco LO. A Practical Classification of Newborn Infants by Weight and Gestational Age. *Journal of Pediatrics* 1967;71:159-163.
- [3] Beck JR, Shultz EK. The Use of Relative Operating Characteristic Curves in Test Performance Evaluation. *Archives of Pathology and Laboratory Medicine* 1986;110:13-20.
- [4] Campbell G. Advances in Statistical Methodology for

the Evaluation of Diagnostic and Laboratory Tests. *Statistics in Medicine* 1994;13:499-508.

[5] Campbell G, DeLeo JM. Fundamentals of Fuzzy Receiver Operating Characteristic (ROC) Functions. *Computing Science and Statistics: Proceedings of the Twenty-first Symposium on the Interface* (Malone L, Berk K, eds.) 1989;543-548.

[6] Campbell G, DeLeo JM. Bootstrapping ROC Plots. *Computing Science and Statistics: Proceedings of the Twenty-seventh Symposium of the Interface* (in press).

[7] Campbell G, Levy D, Bailey JJ. Bootstrap Comparison of Fuzzy R.O.C. Curves for ECG-LVH Algorithms Using Data from the Framingham Heart Study. *Computerized Interpretation of the Electrocardiogram XIV* (Bailey JJ, Kligfield P, eds.): *Journal of Electrocardiology* 1990;23:S132-137.

[8] DeLeo, JM. Receiver Operating Characteristic Laboratory (ROCLAB): Software for Developing Decision Strategies that Account for Uncertainty. *Proceedings of the Second International Symposium on Uncertainty, Modeling and Analysis* 1993 (Ayyub BM, ed.). IEEE Computer Society Press, 318-325.

[9] DeLeo, JM. The Receiver Operating Characteristic Function as a Tool for Uncertainty Management in Artificial Neural Network Decision-making Uncertainty. *Proceedings of the Second International Symposium on Uncertainty, Modeling and Analysis*, 1993 (Ayyub BM, ed.). IEEE Computer Society Press, 141-144.

[10] DeLeo, JM, Campbell G. The Fuzzy Receiver Operating Characteristic Function and Medical Decisions with Uncertainty. *Proceedings of the First International Symposium on Uncertainty Modeling and Analysis*, 1990 (Ayyub BM, ed.). IEEE Computer Society Press, 694-699.

[11] DeLeo JM, Campbell G. Receiver Operating Characteristic (ROC) Methodology in Artificial Neural Networks with Biomedical Applications. *Proceedings of the World Congress on Neural Networks* (in press).

[12] Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. New York: Chapman & Hall, 1993.

[13] Heffner JE, Brown LK, Barbieri C, DeLeo JM. Pleural Fluid Chemical Analysis In Parapneumonic Effusions: A Meta-Analysis. *American Journal of Respiratory and Critical Care Medicine* 1995;151:1700-1708.

[14] Lubchenco LO, Hansman C, Boyd E. Intrauterine Growth in Length and Head Circumference as Estimated from Live Births at Gestational Ages from 26 to 42 Weeks.

*Pediatrics* 1966;37:403-409.

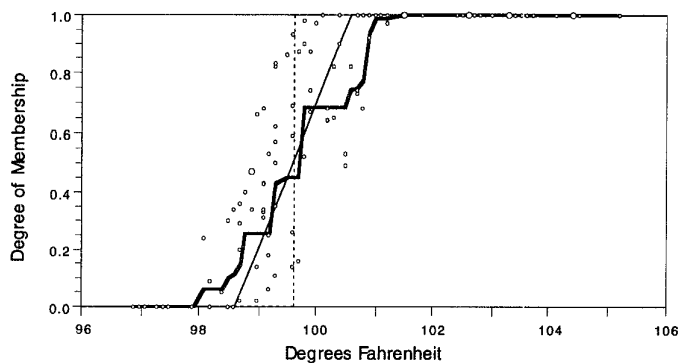
[15] Naranjo CA, Busto U, Sellers EM, Sandor P, Ruiz I, Roberts EA, Janecek E, Domecq C, and Greenblatt DJ. A Method for Estimating the Probability of Adverse Drug Reactions. *Clinical Pharmacology and Therapeutics* 1981;30:239-245.

[16] Tallarigo L, Giampietro O, Penno G, Miccoli R, Gregori G, Navalesi R. Relation of Glucose Tolerance to Complications of Pregnancy in Nondiabetic Women, *New England Journal of Medicine* 1986;315:989-92.

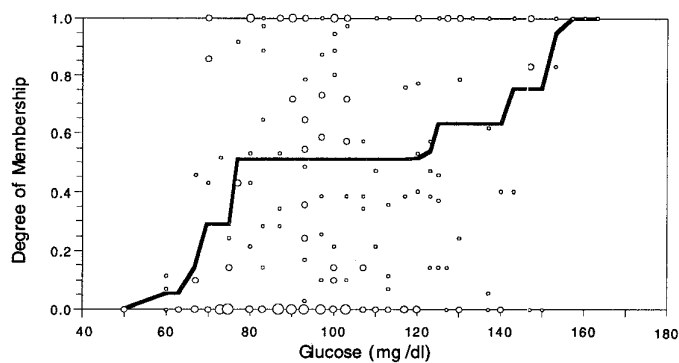
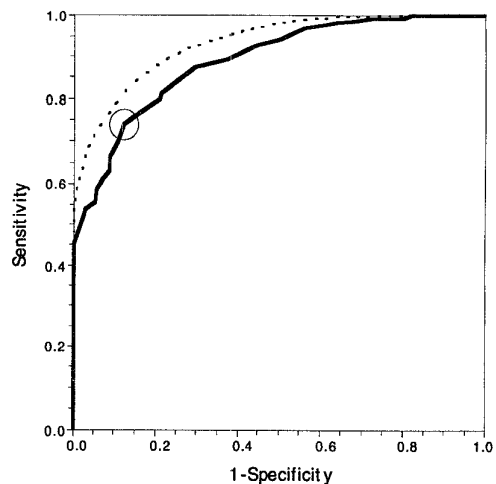
[17] Zadeh, L. Fuzzy Sets. *Information and Control* 1965;8:338-353.

[18] Zadeh LA, Fu K-S, Tanaka K, Shimura M. (eds). *Fuzzy Sets and Their Applications to Cognitive and Decision Processes*. New York: Academic Press, 1975.

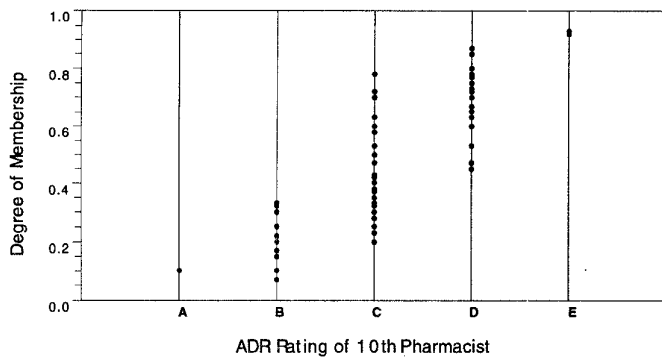
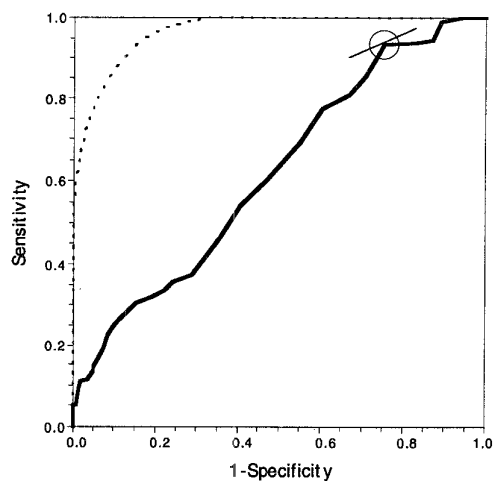
[19] Zweig MH, Campbell G. Receiver Operating Characteristic Plots: A Fundamental Evaluation Tool in Clinical Medicine. *Clinical Chemistry* 1993;39:561-577.



**Figure 1A** (above): Fever data (larger circles indicate multiple occurrences), fuzzy (thin line) and dichotomous (dotted line) gold standards, and ROC-derived fuzzy function (solid line). For gold standards, the abscissa is read as digital thermometer readings; for ROC-derived fuzzy function, as mercury thermometer readings. **Figure 1B** (right): Fever ROC plot (solid line) and "perfect" ROC plot (dotted line). The circled point at sensitivity = 0.741, specificity = 0.877 corresponds to 99.8 degrees on the mercury thermometer.



**Figure 2A** (above): Glucose data (larger circles indicate multiple occurrences) and ROC-derived fuzzy function. **Figure 2B** (right): Glucose example ROC plot (solid line) and "perfect" ROC plot (dotted line). A line with slope = 0.45 intersects the ROC plot at sensitivity = 0.248, specificity = 0.931, corresponding to the interval 75 to 77 on the glucose scale.



**Figure 3A** (above): Adverse drug reaction data. Rating scale: A=most unlikely, B=unlikely, C=uncertain, D=possible, E=probable. Degree of membership values are mean ratings of nine pharmacists based on numeric scale: A=0.05, B=0.20, C=0.50, D=0.80, E=0.95. **Figure 3B** (right): Adverse drug reaction ROC plot (solid line) and "perfect" ROC plot (dotted line). The slope of the lower left line is 10.2.

