Discriminant Functions and Decision Tree Induction Techniques for Antenatal Fetal Risk Assessment

Nilgün Güler[1], Olcay Taner Yıldız[2], Fikret Gürgen[2], Füsun Varol[3], Ethem Alpaydın[2]
[1]Math. Eng. Dept, Yıldız Technical University, Besiktas, Istanbul/TURKEY
[2]Comp. Eng. Dept, Bogazici University, Bebek, 80815, Istanbul/TURKEY
[3]Gyn. and Obs. Dept, Trakya University, Edirne/TURKEY
{guler@yildiz.edu.tr, yildizol@boun.edu.tr, gurgen@boun.edu.tr}

Abstract

This study concentrates on the comparison of the discriminant functions and the decision tree induction techniques in antepartum fetal evaluation. These classification techniques are applied to antenatal fetal risk assessment problem and the performances, the computational complexities and the importance of each technique in terms of diagnostic clues are observed. The task is to investigate the doppler ultrasound measurements of umbilical artery (UA) to relate the health conditions of fetuses using discriminant functions such as linear discriminant functions (LDF), multilayer perceptron (MLP), decision trees (C4.5, CART) and neural trees. We use the following UA blood flow velocity waveforms: pulsatility index (PI), resistance index (RI) and systolic/diastolic ratio (S/D) in terms of weeks (week index: WI as a normalized value) to decide if there is any hypoxia suspicion. It is observed that the performances of MLP and CART are better but C4.5 defines understandable diagnostic clues. On the other hand, the time complexity of LDF and C4.5 are becomes favorable. Experiments support that C4.5, MLP, CART and neural trees are favorable medical aids to phycisians during intensive surveilance of fetuses. With the limited number of indices, we obtain a specificity and sensitivity of 100% and 93% with these decision techniques.

1 Introduction

Several studies have already demonstrated the possibilities and limits of using umbilical doppler for the assessment of fetal growth [1]-[7]. The most commonly used indices are:

1. Systolic/diastolic (S/D) ratio
2. Resistance index (RI)
3. Pulsatility index (PI)

where $PI = \frac{S-D}{mean\ velocity}$ and $RI = \frac{S-D}{D}$. These indices are defined by using the blood flow velocity waveform and are independent of the angle between the ultrasound beam and the direction of blood flow.

In these studies during pregnancy follow up of IUGR fetuses, it has been shown that increased impedance to flow in the UA is associated with fetal hypoxemia and acidemia [1]-[4], [9]. Later, it has also been shown that indices of middle cerebral artery (MCA) and umbilical resistance index to cerebral resistance index ratio (URI/CRI) are important for fetal monitoring [1]-[7]. Even though, the effectiveness of the other indices such as middle cerebral artery (MCA) are known to be useful, we, in this study, only make use of UA indices for a comparison purpose of the discriminant functions and decision tree induction. The results can be extented for the other important indices of fetal monitoring.

On the other hand, in recent years, the family of methods suitable for classification problems has been extended to include a range of new techniques, such as discriminant functions and decision tree induction. Our purpose here is to make a comparison of these classification techniques in antepartum fetal evaluation. Briefly, we employ doppler ultrasound measurements of umbilical artery (UA), then, discuss the usage of discriminant functions and decision tree induction for the assessment of hypoxic conditions. Section 2 presents the proposed monitoring system for fetal health. Section 3 presents decision by discriminant functions such as linear discriminant analysis (LDA) and multilayer perceptron (MLP). Section 4 introduces decision tree induction techniques such as C4.5, CART and neural trees. Section 5 defines the relationship between the discrimination functions and decision trees. Section 6 introduces the experiments, the results and the diagnostic outcomes of these techniques. Finally, the dis-

0-7803-7044-9/01/$10.00 ©2001 IEEE

Table 1: PI, RI, S/D ratio for UA between 20 and 40 weeks[1]

Gestational age (Week)	PI	RI	S/D ratio
20	1.35	0.77	4.40
22	1.25	0.73	3.95
24	1.19	0.72	3.60
26	1.12	0.67	3.40
28	1.08	0.64	3.20
30	0.97	0.63	3.00
32	0.95	0.60	2.80
34	0.90	0.60	2.65
36	0.80	0.55	2.55
38	0.75	0.52	2.40
40	0.72	0.51	2.20

cussion and conclusions are made in the Section 7.

2 Fetal risk assessment system for antenatal care

The antenatal care system receives weekly doppler ultrasound examinations [1] as input and produces an output correlated with clinical conditions. Our study employs UA measurements and defines feature vectors by using weekly ultrasound values that are PI, RI, S/D and WI. The PI, RI and S/D are based on *pulsatility* which is defined as the difference between the peak systolic and end-diastolic components of the maximum frequency shift and the end-diastolic component itself [3]-[5]. The WI is the normalized gestational age in terms of weeks between 0 to 40 [8]. The UA doppler examinations of pregnant women in the range between 20 and 40 weeks are shown in Table 1.

The proposed system uses only the UA indices to have an assessment about the risk of fetal hypoxic conditions. We monitor the patterns of the data to see the effect of the specific measurements on the hypoxia. Then, we use decision techniques such as discriminant functions LDA and MLP and decision trees C4.5 and CART to classify the fetal conditions with these indices. As mentioned previously, the other doppler measurements such as MCA, URI/CRI generally improve the performance and the quality of the assessment decision.

3 Decision by discriminant functions

Discriminant function analysis has been an important decision tool for various pattern recognition applications [10], [11], [12], [13]. Discriminant functions generally transform input feature space on the directions of *maximum separability.* Discriminant analysis is a technique for identifying the "best" set of attributes or variables, known as the discriminator for an *optimal decision.* We can define the discriminant functions in various degrees of polynomials such as linear, quadratic.

3.1 Decision by Linear Discriminant functions (LDF)

The linear discriminant is the first order polynomial that is used for decision. The classifier can be obtained as a result of the application of Bayes rule to the problem of classification under the following assumptions:

- the data are normally distributed classes: $N_1(\mu_1, \sum_1)$, $N_2(\mu_2, \sum_2)$ where μ's are mean vectors and $\sum$'s are covariance matrices.
- the covariance matrices of every class are equal: $\sum_1 = \sum_2 = \sum$.

With these assumptions, a LDF is computed as:

$$y = (\mu_1 - \mu_2)' \sum^{-1} x \tag{1}$$

These assumptions impose restrictions to problems to which LDF are applied. But it is known that, despite these restrictions, the LDF still performs well on data which is only approximately normally distributed, and where the classes have different covariances.

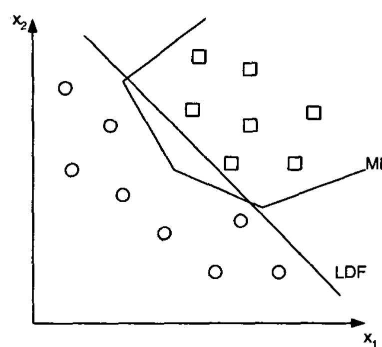

Figure 1: Discriminant functions LDF and MLP

3.2 Decision by nonlinear discriminant functions

We use a nonlinear discriminant function to draw decision boundaries between class regions. As an example case, we utilize *multilayer perceptrons* (MLP) with sigmoidal transfer function [14], [15], [16]. The feedforward network learns from the input data by the *supervision* of the output data creating single linear discriminant functions by each sigmoid hidden unit and then combines them. Thus, this piecewise linear discriminant function works as a nonlinear discriminator.

For training a MLP, the error back-propagation (BP) has become very popular [14]. The back-propagation is an optimization technique for implementing gradient descent in weight space for multilayer feedforward networks. The basic idea of the technique is to efficiently compute partial derivatives of an approximating function $F(w;x)$ realized by the network with respect to all the elements of the adjustable weight vector w for a given value of input vector x and output vector y. The weights are adjusted to fit linear piecewise discriminant functions to feature space for the best class separability. The difference between the network's output and the supervisor output is *minimized* according to a predefined *error function* (performance criterion) such as mean square error (MSE) (2) etc. The error function helps to place the discriminators to right location and positions:

$$MSE = \sum_q (y_q - F(w;x_q))^2 \qquad (2)$$

4 Decision by decision tree induction techniques

Decision tree construction algorithms are greedy in that at each step, we decide on a decision node that best splits the data for classification. Different decision tree learning methods differ in the way they make decisions at a node. In a *univariate decision tree*, at each node, only one feature is used. If the feature is continuous, the decision is of the form

$$x_l > c_k$$

making a binary split where x_i is the ith input feature and c_k is a suitably chosen threshold. If x_i is discrete valued with m values, then the node makes an m-ary split. A univariate test using feature x_i can only split a space with a boundary that is orthogonal to x_i axis. This can result in large trees and poor generalization.

In a *multivariate decision tree*, each decision node is based on more than one feature. The linear multivariate tree-constructing algorithms select not the best attribute but the best linear combination of the attributes. The decision at each node is of the form

$$\sum_{i=1}^{d} w_i x_i > w_0$$

where w_i are the weights, and w_0 is a threshold.

4.1 Decision by C4.5

Univariate decision tree construction algorithm C4.5 [17] learns attributes by constructing them top-down manner starting with selecting the best attribute to test at the root of the tree. To find the best attribute, each instance attribute is put into a statistical test to determine how well it alone classifies the training examples. The best feature is selected and used as a test node of the tree. A child of the root node is then created for each possible value of the attribute namely two children for ordered features as $x_i < a$ and $x_i > a$, and m children for unordered feature as

$$x_i = a_1, x_i = a_2 \ldots x_i = a_m$$

where m is the number of different possible values of the feature x_i. The entire process is then repeated recursively using the training examples associated with each child node to select the best attribute to test at that point in the tree. This form of C4.5 algorithm never backtracks and is a *greedy search algorithm.*

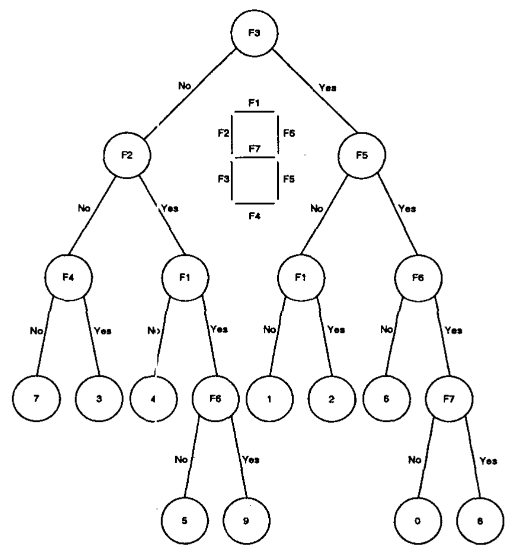

Figure 2: Sample Univariate Decision Tree for LED Problem

4.2 Decision by CART

The multivariate decision tree construction algorithm CART [18] selects not the *best attribute* but the *best linear combination of the attributes.* The linear combination consists of multiplication of weights w_i with each feature x_i. The main operation in CART is determining the weights w_i of those features.

CART algorithm for finding the coefficients of the available features is a step-wise procedure, where in each

step one cycles through the features $x_1, x_2 \ldots x_n$ doing a search for an improved linear combination split. At the beginning of the L'th cycle, let the current linear combination split be $v \leq c$. For fixed γ, CART searches for the best split of the form:
$v - d(x_1 + \gamma) \leq c$, such that

$$\delta \geq \frac{v-c}{x_1+\gamma}, where x_1 + \gamma \geq 0$$
$$\delta \leq \frac{v-c}{x_1+\gamma}, where x_1 + \gamma \leq 0 \qquad (3)$$

The search for δ is carried out for $\gamma = -0.25, 0.0, 0.25$. The resulting three splits are compared, using the chosen partition-merit criterion, and the δ and γ corresponding to the best are used to update v, producing:

$$v' = w_1' x_1 + w_2' x_2 \ldots, where$$
$$w_1' = w_1 - \delta$$
$$c' = c + \delta\gamma \qquad (4)$$

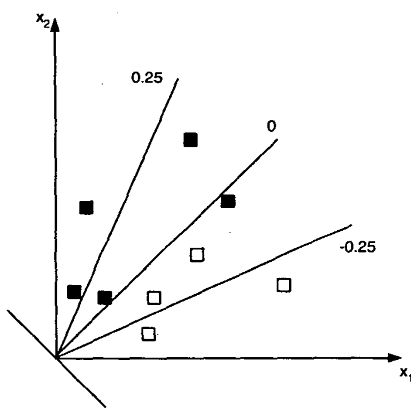

Figure 3: A Step in CART Algorithm

Figure 3 shows the first step of the CART algorithm for an example data set. The initial line is given as $x_1 + x_2 < 0$. The lines shown as -0.25,0 and 0.25 are the best splits found for γ = -0.25,0 and 0.25. Here only the coefficient of attribute x_1 is changed. The line with $\gamma = 0$ will be selected for further iteration.

This search is repeated for other features $x_2, x_3 \ldots x_n$ to obtain an updated split $v_l < c_l$. The final step of the cycle is to find the best c_l, and the system searches explicitly for the split that minimizes the impurity of the resulting partition. The cycles end when the reduction of impurity is below a constant say e[18].

4.3 Decision by Neural Trees

The decision at a binary multivariate node is a binary classification problem and thus any binary classifier can be used to implement it. Guo and Gelfand (1992) propose to use multilayer perceptrons in decision nodes which implement multivariate *nonlinear* decision trees. We replaced the multilayer perceptron with a single layer, linear perceptron. This type of decision trees, where at each decision node there is a linear perceptron is called a Neural Tree.

Table 2: Confidence Table for algorithms

	C4.5	CART	ID-LP	MLP	LDA
C4.5	-	38.62	85.95	88.06	27.67
CART	38.62	-	74.26	72.69	40.37
ID-LP	85.95	74.26	-	41.16	81.96
MLP	88.06	72.69	41.16	-	85.55
LDA	27.67	40.37	81.96	85.55	-

Table 3: Learning times for algorithms (in sec.)

	Learning Time
C4.5	2±0
CART	45±3
ID-LP	2±1
MLP	2±0
LDA	1±0

5 On the relationship of discrimination functions and decision trees

C4.5 decision tree algorithm constructs a stepwise discriminator by choosing the best ordered attribute each time (Figure 1,2). CART tree algorithm, Neural Trees and LDF draws a linear discriminator by the linear combination of the same attributes. Finally, MLP defines a nonlinear discriminator that have the ability to draw any discriminator if it is trained properly. In general, according to application faced, we can use one these techniques by considering data size, computation time, and the importance of the solution produced.

6 Experiments and Results

In this section, our purpose is to compare the algorithms with their confidence interval and time complexity. We test five algorithms using combined 5x2 cross validation (cv) F test [21]. We partition the data into two sets five times: training set and test set and we repeat experiments on these sets. We then compare the computation times of these algorithms. The set of data are used for the experiments: the PI, RI and S/D ratio values of UA [1] between 20 and 40 weeks.

Table 4: Prevalence Data from UA

	Sensitivity	Specifity	PPT	PNT
LDA	100%	76%	68%	100%
MLP	100%	93%	88%	100%
ID-LP	100%	93%	88%	100%
C4.5	100%	74%	68%	100%
CART	100%	93%	88%	100%

The result of the F test is shown in Table 2, and the learning time of the algorithms are given in Table 3. The results show that there is no significant difference between the algorithm(s) in terms of accuracy therefore we can choose the algorithms with the minimum learning time and maximum comprehensibility. Therefore we can choose C4.5 algorithm.

The sensitivity of the test is the probability that the test result will be abnormal when the disease is present. The specificity of the test is the probability that the test result will be normal when the disease is not present. The predictive value of an abnormal test (PNT) would be that fraction of fetus with an abnormal test result who have the abnormal condition, while the predictive value of a normal test (PPT) would be the fraction of fetus with a normal test result who are normal. Table 4 shows the Sensitivity, specifity, PPT and PNT results of the algorithms.

In this study, the highest specificity and sensitivity of the system is achieved by MLP and CART. The other techniques(expecially LDA and C4.5) are poorer to specify the predicting normal cases while they are very good at the predicting abnormal cases.

Both, the discriminant functions and decision trees of pattern recognition are verified to be valuable tools for revealing adverse conditions in antenatal fetal risk assessment. As the nonlinearity of the discriminator increases, the accuracy of the assessment also improves. A MLP, as a nonlinear discriminator, outperforms the performances of the LDA and the C4.5 induction, in the fetal risk assessment task. Also, in the case of CART, Linear combination of attributes (variables) are used for the decision induction. It is observed that there is a group of discriminators that is constructed with the techniques: the outcome of LDA is a linear discriminator, C4.5 draws stairs according to the best induced features. The CART induces a linear combination of the best features for an optimal decision. The MLP, as a nonlinear discriminator, draws piecewise linear combinations, in each of them a local optimality is searched by training.

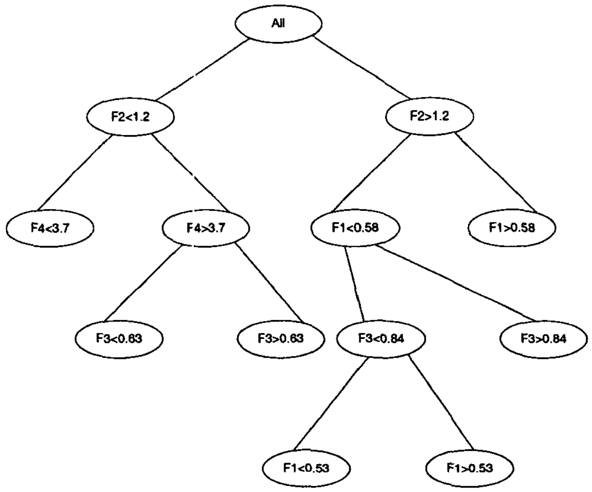

Figure 4: A sample result tree for C4.5 algorithm (WI= Featurel, PI= Feature2, RI=Feature3 ,S/D=Feature4)

7 Discussion and Conclusions

The study points the following facts:

1. The decision trees C4.5, CART and neural trees are found to be applicable to the selection of the best attributes and/or the combination of them to make the best decision for antepartum fetal evaluation.
2. The discriminant functions LDA and specifically MLP are also shown to be effective class discriminators for the same problem [19], [20].
3. The discriminant functions and decision tree induction techniques produce discriminators. The first group obtains an optimal decision by the combination of attributes in the linear or piecewise linear form. The second group obtains similar decision by employing a tree that give the result by selection of the best attribute or the linear combination of the best attributes.
4. The MLP, CART and neural trees are found to be helpful to support the doctor's decision.
5. It is proven that the risk assessment by using the doppler ultrasound based indices PI, RI, S/D ratio of UA according to WI can be done with our system. The experiments that were performed here confirm the observations of the doctors. But the special cases also need special attention by the doctors: For example, it must be pointed out

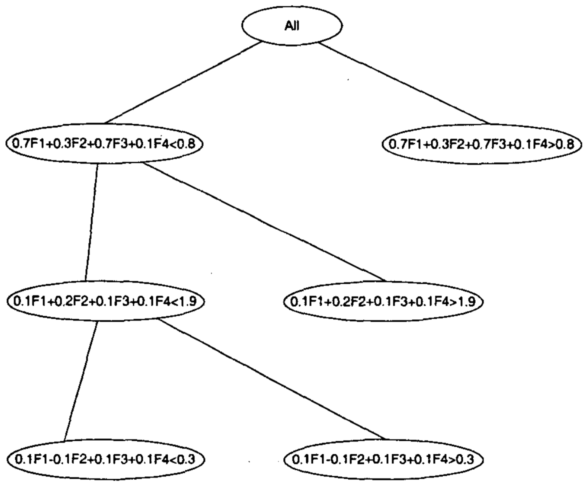

Figure 5: A sample result for CART algorithm

that the PI in the UA may even be raised several weeks or months before fetal hypoxia is clinically suspected.

6. This study points a fruitful line of enquiry for helping doctors in the risk assessment of antenatal fetal evaluation.

References

[1] Giancarlo, M., and J. A. Copel, *Doppler Ultrasound: Fetal Physiology and Clinical Application*, 1996.

[2] Creasy, R., *Maternal-Fetal Medisine*, 1999.

[3] Chervenak, F., and A. Kurcak, *The Fetus as a Patient*, 1996.

[4] De Haan, J., *Pathophysiologic Backgrounds of changes in doppler flow indices*, Elsevier, Amsterdam, 1994.

[5] D. Maulik, P. Yarlagadda, and G. Downing, "Doppler velocimetry in obstetrics", *Obstet Gynecol Clin North Am*, Vol 17, pp.163-86, 1990.

[6] M. S. Beksac, F. Basaran, S. Eskiizmirliler, A. M. Erkmen, and S. Yorukan, "A Computerized diagnostic system for the interpretation of umbilical artery blood follow velocity waveforms", *European Journal of Obstetrics & Gynecology and Reproductive Biology*, Vol 64, pp 37-42, 1996.

[7] J. W. Wladimiroff, J. A. G. W. V. D. Wijngaard, S. Degani, M. J. Noordam, J. V. Eyck, and H. M. Tonge, "Cerebral and Umbilical Arterial Blood Flow Velocity Waveforms in Normal and Growth Retarded Pregnancies", *European Journal of Obstetrics & Gynecology and Reproductive Biology*, Vol 69, pp 705-709, 1987.

[8] F. Gurgen, "Neural-Network-Based Decision Making in Diagnostic Applications", *IEEE Engineering in Medicine and Biology*, pp 89-93, July/August, 1999.

[9] F. Gurgen, E. Onal, and F. G. Varol, "IUGR Detection by Ultrasonographic Examinations Using Neural Networks", *IEEE Engineering in Medicine and Biology magazine*, Vol. 16, No. 3, pp. 55-58, 1997.

[10] Morrison, D. F., *Multivariate Statistical methods*, McGraw-Hill Publishing Company.

[11] Hair, J. F., jr. R. E. Anderson, R. L. Tatham, *Multivariate Data Analysis with Readings*, Prentice Hall, 1995.

[12] Fukunaga, K., *Introduction to Statistical Pattern Recognition*, Academic Press Inc., 1990.

[13] Subhash, S., *Applied Multivariate Techniques*, John Wiley & Sons Inc., 1996.

[14] Haykin, S., *Neural Networks*, Macmillan College Publishing Company Inc., 1994.

[15] Ripley, B.D., *Pattern Recognation and Neural Networks*, Cambridge Press, 1996.

[16] Smyth, P., and W. Buntine, *Probabilistic Machine Learning: Putting Theory in to practice: Tutorial Notes.*

[17] J. R. Quinlan, "Induction of Decision Trees", *Machine Learning*, Vol. 1, pp. 81-106. 1986.

[18] Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*, 1984.

[19] N. Guler, F. Gurgen, F. Varol, "Decision of Normal and Growth Retarded Pregnancies by Discriminant Functions Using Umbilical Arterial Blood Flows Velocity Waveforms", *Advance in Intelligent Systems and Computer Science*, pp. 186-190, 1999.

[20] N. Guler, F. Gurgen, F. Varol, "Ultrasound Blood Flow Waveform Discriminant Functions for Antenatal Fetal Risk Assessment", *IEEE Engineering in Medicine and Biology magazine*, 2000.

[21] E. Alpaydin, "Combined 5x2 cv F Test for Comparing Supervised Classification Learning Algorithms", *Neural Computation*, Vol. 11, pp. 1975-1982, 1999.

[22] H. Guo, and S. B. Gelfand, "Classification Trees with Neural Network Feature Extraction", *IEEE Transactions on Neural Networks*, Vol. 3, pp. 923-933, 1992.