Research Paper ■

Machine Learning for an Expert System to Predict Preterm Birth Risk

LINDA K. WOOLERY, PHD, RN, JERZY GRZYMALA-BUSSE, PHD

Abstract **Objective:** Develop a prototype expert system for preterm birth risk assessment of pregnant women. Normal gestation involves a term of 40 weeks, but because 8–12% of the newborns in the United States are delivered prior to 37 weeks' gestation, problems associated with prematurity continue to plague individuals, families, and the health care system.

Design: A knowledge-base development methodology used machine learning, statistical analysis, and validation techniques to analyze three large datasets (18,890 subjects and 214 variables). The dependent (i.e., decision) variable studied was weeks of gestation at delivery, with dichotomous coding of preterm delivery (prior to 37 weeks) and full-term delivery (37+ weeks).

Results: Machine learning with a program named Learning from Examples using Rough Sets (LERS) induced 520 usable rules that were entered into a prototype expert system. The prototype expert system was 53–88% accurate in predicting preterm delivery for 9,419 patients.

Conclusion: The prototype expert system was more accurate than traditional manual techniques in predicting preterm birth.

■ **J Am Med Informatics Assoc.** 1994;1:439–446.

Determining preterm birth risk and decision making related to interventions remain problematic in the clinical setting.[1–3] Accurate assessment of preterm risk will permit intervention with educational programs, bed rest, and early symptom management to prolong gestation or prevent preterm birth and will provide for improved perinatal survival and treatment outcomes. One problem related to preterm delivery risk assessment appears to be a poorly defined and complex knowledge base. The plethora of information about preterm risk remains disorganized, poorly validated through research, and of little guidance to patients and providers of prenatal care. Preterm risk information, including risk factors and outcomes, is increasing at a rate that confounds traditional techniques of information management and patient management. Previous approaches to studying the problems of preterm birth prediction have failed to validate linear models. Factors traditionally used to assess risk are not clearly or consistently associated with weeks of gestation at birth.[4–6]

Many risk scoring and screening instruments are available, but no conceptual or theoretical model of preterm risk has been reported, which may account for the poor reliability and validity of traditional manual screening techniques. McLean et al.,[2] on reviewing previous studies, found that manual risk assessment scoring tools were 17–38% accurate in their abilities to predict preterm delivery. This astonishingly poor predictive ability of accepted risk screening methods is a significant degradation of the accuracy that can be achieved by flipping a coin. The primary reason for this degradation is a psychometric problem where there is no underlying conceptual model[7] of preterm birth risk. Existing preterm birth risk screening tools include factors that are not valid predictors of preterm birth risk and fail to include reported factors that may be valid predictors of preterm labor. Although existing tools do not adequately predict preterm birth, current prenatal practice uses

Affiliations of the authors: School of Nursing, University of Missouri, Columbia, MO (LKW); Department of Electrical Engineering and Computer Science, University of Kansas, Lawrence, KS (JG-B).

Supported by National Institute for Nursing Research Grant No. 1 R43 NR02899-01A1 (IntelliDyne, Inc.), September 1992–February 1993.

Correspondence and reprints: Linda Woolery, PhD, RN, S329 School of Nursing, University of Missouri, Columbia, MO 65211. Internet: nurswool@mizzou1.missouri.edu

Received for publication: 10/28/93; accepted for publication: 6/20/94.

Table 1 ■

Example of a Decision Table and the Rules Discovered From It by the Learning from Examples using Rough Sets (LERS) Machine Learning Program*

	Attribute			Decision
Example	pregnancy_#	maternal_age	bleeding	delivery
patient_1	1	<20	yes	preterm
patient_2	3	30..39	no	fullterm
patient_3	2	20..29	no	preterm
patient_4	2	20..29	no	fullterm
patient_5	1	20..29	yes	fullterm

*This decision table and its rules are simplified to demonstrate the theoretical foundations. Actual decision tables and rules are much more complex.

these invalid, unreliable tools daily when dealing with pregnant women, resulting in an increasing trend to treat all pregnant women as though they are at "high risk" for preterm labor. Alternative solutions to the problem may be achieved using machine learning and expert system technology to support health care providers' assessments in this complex domain.

The purpose of this research was to improve clinical outcomes for childbearing families, first through the development of a knowledge base and then through the development of an expert system for improved preterm birth risk assessment of pregnant women. This was the second study in an ongoing program of informatics research in which artificial intelligence techniques, called *machine learning*, were used for knowledge acquisition to develop and describe a knowledge base for preterm birth risk assessment.

Machine Learning

Knowledge acquisition is the transfer of knowledge from the (expert) source to a knowledge base.[8] Traditional approaches to knowledge acquisition are tedious and frequently based on manual techniques, e.g., verbal protocol analysis.[9] The difficulty encountered in studying experts lies in the experts themselves and the processes by which they become experts. Experts have two kinds of knowledge—knowledge used to explain a task and knowledge that actually is used to perform a task.[10] Johnson[11] called this the "paradox of expertise" and argued that the knowledge we wish to capture is that which the expert is least able to discuss. The human experts developed preterm risk scoring and screening tools in the 1980s, but these tools remain only 17–38% accurate in predicting preterm birth risk.[2] There is no expert with a proven track record for accurately predicting preterm birth risk, and research findings are frequently contradictory. Traditional approaches to expert system development, where knowledge is acquired from human experts, simply have not worked in the domain of preterm birth risk assessment.

Newer approaches to knowledge acquisition using machine learning techniques were developed during the 1980s. Algorithms with different strengths and limitations have been developed to extract patterns from data for the creation of decision trees, production rules, and other representations for expert systems.[8] Four general paradigms of machine learning have evolved: the analytic paradigm, the genetic paradigm, the connectionist paradigm, and the inductive paradigm.[12] The analytic paradigm generally starts with a strong underlying theory and attempts to build and learn concepts from that theory. Because preterm birth risk lacks a strong underlying theory, this paradigm is not appropriate for this study. Another paradigm not well suited for preterm delivery risk assessment is the genetic paradigm, which generates new solutions to a problem and then tests the "fitness" of these solutions. The connectionist, or neural network, paradigm recognizes patterns in input data, and learns how to classify the input based on previous classifications with similar input patterns. This paradigm is appropriate for preterm birth risk assessment and will be used in future studies. The inductive paradigm generally works from sets of data where the classification of an example (patient) is known, and the system learns to discriminate between different classifications given the data values associated with the patient. The inductive paradigm of machine learning was used for this study.

The most successful inductive machine learning technique employed in this project involved the use of a computer program called Learning from Examples using Rough Sets (LERS), which was developed in the Computer Science Department of the University of Kansas by one of the authors (JG-B).[8,13–18,21] In clinical practice, where inconsistency is an integral part of caring for humans, the rough set approach seems a desirable theory to explore more fully in developing expert systems.

LERS works with a decision table that presents data about real-world phenomena that will be used for decision making. In the decision table, *objects* or *examples* are characterized by *attributes* and *decisions*. Examples are described by values of attributes, while decision values are often provided by experts. The primary goal is for LERS to extract information from various sets of data to discover rules from the decision table. A simplified description of this approach involves building a table with examples, attributes, decisions, attribute values, and decision values that are then read by the machine learning program for evaluation (Table 1).

In machine learning from examples, a *concept* is understood as the subset of the set of all examples having the same value of the decision. Let d denote a decision (e.g., *delivery* from Table 1) and let w denote a value of the decision (e.g., *preterm* from Table 1). Formally, a concept, denoted $[(d, w)]$, is a set of all examples that have value w for decision d. In our example, the concept [(*delivery*, *preterm*)] is the set {*patient_1*, *patient_3*}. Similarly, let q be an attribute and v its value. The block of an attribute–value pair (q, v), denoted $[(q, v)]$, is the set of all examples that have value v for attribute q. In our example, the blocks of all attribute–value pairs are:

[(*pregnancy_#*, 1)] = {*patient_1*, *patient_5*},
[(*pregnancy_#*, 2)] = {*patient_3*, *patient_4*},
[(*pregnancy_#*, 3)] = {*patient_2*},
[(*maternal_age*, <20)] = {*patient_1*},
[(*maternal_age*, 20..29)] = {*patient_3*, *patient_4*, *patient_5*},
[(*maternal_age*, 30..39)] = {*patient_2*},
[(*bleeding*, *yes*)] = {*patient_1*, *patient_5*},
[(*bleeding*, *no*)] = {*patient_2*, *patient_3*, *patient_4*}.

LERS is based on rough set theory, a method for managing uncertainty in knowledge acquisition.[13,19,20] Uncertainty may be caused by data errors, ambiguity of exact meanings of data in the table, or doubtful connections between conditions and a conclusion of the rule. A special case of uncertainty is an inconsistency when a decision table contains two examples having identical attribute values but different decision values. For example, patient_3 and patient_4 have identical attribute values—2 for *pregnancy_#*, 20..29 for *maternal_age*, and *no* for *bleeding*—but different decision values—*preterm* for patient_3 and *fullterm* for patient_4. A problem exists with other approaches to managing uncertainty where inconsistencies are removed from the table and ignored by the learning program. In rough set theory, inconsistencies are not removed.

In the rough set approach used in LERS, the basic concepts are *lower* and *upper approximations* of a concept. The lower approximation of the concept $[(d, w)]$ is the largest set of all examples that may be described as being *certainly* a part of the concept, taking into account all attributes. For example, for the concept [(*delivery*, *preterm*)], the only patient who may be classified as being *certainly* a part of the concept is *patient_1*. *Patient_3* is not so classified because even looking at values of all attributes we cannot distinguish her from *patient_4*, and *patient_4* does not belong to the concept.

On the other hand, the upper approximation of the concept $[(d, w)]$ is the smallest set of all examples that is described as *possibly* containing the concept. In our example, for the concept [(*delivery*, *preterm*)], not only *patient_1* but also *patient_3* and *patient_4* may possibly belong to the concept. Therefore, the upper approximation of the concept [(*delivery*, *preterm*)] is the set {*patient_1*, *patient_3*, *patient_4*}.

LERS computes both lower and upper approximations for concepts of the decision table. In the next step, LERS induces rules from these lower and upper approximations. The rules that are computed from lower approximations are called *certain*, while the rules that are computed from upper approximations are called *possible*. The definitions of these rules are similar to those mentioned above for lower and upper approximations of the concept. Certain rules are completely backed up by data, as long as we restrict our attention to available data. Certain rules describe regularities in the data, without any uncertainty. Possible rules are also supported by the same data, but it is possible that some data may support a rule and some other data may contradict the same rule. Thus it is only possible that a rule is true. Possible rules are further quantified by a special measure, called the *rough measure*. The rough measure of the rule describing concept $[(d, w)]$ is the ratio of the number of all examples from the concept $[(d, w)]$ correctly described by the rule to the number of all examples described by the rule. The rough measure may be interpreted as a conditional probability of the conclusion of the rule given all rule conditions. Obviously, the rough measure of a certain rule is equal to 1. The higher the rough measure for a possible rule, the more reliable the rule.

A basic algorithm that was used in the system LERS is called LEM2. In LEM2 an attribute–value pair is selected first by looking for attributes with the highest priorities. Attribute priorities should be allocated by the expert. In our project, the assumption was that all priorities are equal or that no priority is allocated, i.e., no bias was added. The next criterion for selection of an attribute–value pair is its relevance to a goal. Goal initially is a concept; later on, it is a concept with deleted examples that are already described by rules. The relevance of an attribute–value pair and the goal is evaluated as the cardinality of the common part of both sets. When a tie occurs, an attribute–value pair is selected on the basis of the maximum of conditional probability of a block of the attribute–value pair given the goal.

Table 1 describes two concepts. The first concept is characterized by value *preterm* for the decision *delivery* and is equal to the set {*patient_1*, *patient_3*}. The second concept, the set {*patient_2*, *patient_4*, *patient_5*}, describes all patients who have value *fullterm* for decision *delivery*.

The lower approximation of the concept {*patient_1*, *patient_3*} is the set {*patient_1*}. For the set {*patient_1*}, the set of all relevant attribute–value pairs is:

{(*pregnancy_#*, 1), (*maternal_age*, <20), (*bleeding*, *yes*)}.

Obviously, the attribute–value pair (*maternal_age*, <20) should be selected, since it describes only the set {*patient_1*}. Thus, the only *certain* rule is:

(*maternal_age*, <20) → (*delivery*, *preterm*).

On the other hand, possible rules are computed on the basis of the upper approximation of the concept, i.e, the set {*patient_1*, *patient_3*, *patient_4*}. The set of all relevant attribute–value pairs is:

{(*pregnancy_#*, 1), (*pregnancy_#*, 2), (*maternal_age*, <20), (*maternal_age*, 20..29), (*bleeding*, *yes*), (*bleeding*, *no*)}.

The best attribute–value pair is (*pregnancy_#*, 2) because [(*pregnancy_#*, 2)] = {*patient_3*, *patient_4*} is the most relevant set contained in the goal, the set {*patient_1*, *patient_3*, *patient_4*}. Thus, the first possible rule is:

(*pregnancy_#*, 2) → (*delivery*, *preterm*).

All certain rules, induced by LEM2, are:

(*maternal_age*, <20) → (*delivery*, *preterm*),
(*pregnancy_#*, 3) → (*delivery*, *fullterm*),
(*pregnancy_#*, 1) & (*maternal_age*, 20..29) → (*delivery*, *fullterm*).

Possible rules, induced by LEM2, are:

(*pregnancy_#2*) → (*delivery*, *preterm*),
(*maternal_age*, <20) → (*delivery*, *preterm*),
(*bleeding*, *no*) → (*delivery*, *fullterm*),
(*maternal*_age, 20..29) → (*delivery*, *fullterm*).

Sample, Setting, and Methodology

The methodology used in this study was refined from earlier knowledge base development methodology work[6,21] using simplified classification schemes, multiple large datasets (n = 18,890), and multiple machine learning programs (ID3,[22,23] LERS,[18] and CONCLUS[24]). An earlier study[6] used ID3 to generate 88 rules from a single database, but the classification criterion used was weeks of gestation at delivery and all 88 rules were deemed meaningless by a panel of experts. The conclusion reached was that the classification was too complex for ID3 to manage the large dataset. Based on this prior experience, we used a dichotomous decision classification where the machine learning programs analyzed preterm delivery or full-term delivery. The research procedure included the following steps:

1. Data from three large databases (n = 18,890) were loaded into appropriate computers and formats. The original intent was to merge databases, but their sizes and dissimilar variables made this unwieldy, so each database was processed separately. The clinical data represented a mixture of high-risk and low-risk pregnant women collected by a level III perinatal center in the midwest and two private companies providing home uterine-monitoring services for high-risk patients throughout the United States. The data in each database were split in half at this step. Half of the data were used for statistical analysis, machine learning, and rule generation for the prototype expert system. The other half of the data were set aside and used only to test the prototype expert system with real patient cases.
2. Exploratory factor analysis was conducted.
3. Multivariate regression analysis was conducted to determine predictors of preterm delivery risk.
4. Knowledge acquisition based on machine learning was conducted to induce rules directly from the data.
5. Rules were validated using content validity techniques and perinatal experts.
6. The prototype expert system was built and tested.

Results of Statistical Analysis (Methodology Steps 2 and 3)

All statistical analyses were conducted with datasets that were used for rule induction. Descriptive statistics were collected and exploratory factor analyses and multiple regression analyses were conducted for 9,419 subjects and 214 variables. Database 1 collected 52 variables, including patient demographic data, high-risk factors (such as multiple gestation, smoking, or drug use), medical complications (such as bleeding, diabetes, or hypertension), intervention data (such as medications and monitoring results), and outcome data (such as gestational age, birth weight, and American Pediatric Gross Assessment Record scores). Database 2 collected 77 variables, including items similar to those collected by the first database as well as numerous variables for biophysical markers such as height, weight, blood pressure, pulse, and uterine contractions. Database 3 collected 85 variables and included minimal demographic and high-risk data but detailed data with regard to *International Classification of Diseases*, 9th edition (ICD-9) diagnostic codes and Current Procedural Terminology procedure codes associated with patients who experienced preterm labor.

According to descriptive statistics, the average age of women in all three databases was in the late twenties. The number of adolescent subjects was relatively small, and these data may not reflect risk factors of adolescent pregnancy. Only three of the subjects analyzed had not received prenatal care, thus this study was unable to address preterm birth risk of women who do not seek prenatal care. Dichotomous coding and small numbers of subjects who had positive responses on numerous variables produced several problems for statistical analysis that will be managed in future studies using logistic regression techniques.

In general, conclusions drawn from descriptive data analysis were that the data were voluminous, sometimes erroneous, poorly organized, inconsistently recorded, and frequently dichotomous, and that data items needed were often not collected. The multiple regression statistics used in this study did find statistical significance for many of the variables, but the low correlations between most of the 213 predictor variables and the criterion variable (weeks of gestation at delivery) rendered statistical significance meaningless for assessment purposes in clinical practice. The multiple regression findings in this study may lend additional support to an earlier study[4] that found no statistically significant results for race, age, marital status, parity, or socioeconomic status and to a study[5] that found no statistically significant relationships between gestational age at delivery and maternal age, gravidity, parity, or race. More work is needed to replicate and analyze preterm birth risk factors in relation to age, race, and other items believed to predict or to be strongly associated with preterm birth risk. It is possible that preterm birth risk does not fit a linear model, and alternative analyses may be more appropriate in future studies.

The inability to predict preterm birth risk from the data was somewhat surprising, at first, but this finding can be clarified through several explanations. The low correlations between predictor variables and the criterion may be due, in part, to the possibility that health care providers continue to collect a great deal of data that have little to do with preterm birth risk. The data in the perinatal databases reflected risk factors that are consistent with traditional preterm birth risk screening instruments developed in the 1980s. However, review of the literature found that preterm risk scoring indices were not developed according to psychometric standards. It is possible that current clinical practice operates with assumptions about risk factors for preterm birth that are invalid. And it is important to remember that the exact cause of labor, whether full-term or preterm, remains unknown.

Table 2 ■

Accuracy Rates of the Prototype Expert System for Predicting Preterm or Full-term Delivery

	Database 1	Database 2	Database 3
Total no. of test cases	1,593	1,218	6,608
No. correctly classified	1,415 (88.8%)	722 (59.2%)	3,533 (53.4%)
No. misclassified	171 (10.7%)	456 (37.4%)	2,796 (42.3%)
No. unclassified	7 (0.4%)	40 (3.2%)	279 (4.2%)

Results of Machine Learning and Expert Validation (Methodology Steps 4 and 5)

Multiple approaches to machine learning were conducted using software programs named ID3, LERS, and CONCLUS. One of us (JG-B) previously indicated[8] that most successful research activity in the area of inductive machine learning worked with data that were free of errors and conflicts, or inconsistencies. The study tested the robustness of machine learning with data that contained both errors and inconsistencies. Examples with missing values and obvious errors, such as maternal 10-pound or 700-pound weights, systolic pressures of 14,000, and pulses of less than 40, were excluded from machine learning analysis, leaving 9,419 cases for further analysis. Of the programs tested, LERS produced the only usable output, inducing 1,655 rules directly from the data.

A content validity technique was used for rule validation, where two certified perinatal nurses who were experts were asked to verify rules using categories described by Fieschi[25] for tests of incompleteness and logical, structural, and semantic verification of rule output. Some of the verification process was accomplished through LERS programming enhancements that guaranteed that contradictory rules were not generated (logical verification), unattainable and circular rules (where rules iterate and never end) were not generated (structural verification), and rules with erroneous value limits were not generated (semantic verification). Programming enhancements to the LERS software were also able to determine which patient cases in the test data were unable to be classified by the rules, thus pointing out where rules were missing and where more data and rules were needed (tests of incompleteness).

LERS analysis of database 3 generated 1,133 rules, but there were multiple problems with the data. First, listwise deletion of cases with missing values created problems for database 3, where only 9 of 6,616 cases were without missing values. The first test of database 3 produced predictive accuracy rates of 98%, which were exciting until careful analysis revealed an analyzed attribute for preterm delivery, as well as a decision variable with the same value. This actually served to confirm the machine learning classification process, but was not clinically useful. Considering the confusion with duplicate attributes and problematic missing values, the experts recommended the prototype be built without the 1,133 rules from database 3. The remaining 520 rules from databases 1 and 2 were used for further expert verification.

Expert verification involved checking for redundant rules, irrelevant attributes, erroneous rules, and meaningless or nonsense rules. For example, a single rule the expert was asked to verify was:

(abortions, 0) & (gravida, 2) & (pregnancy complication, 2nd trimester bleeding) & (pregnancy complication, incompetent cervix) & (pregnancy complication, premature rupture of membranes) ===> (birth, preterm).

Experts found the above format difficult to analyze, and, at their requests, programs were written to make the rule output easier for the experts to analyze (see below); however, the process of verifying 520 complex rules remained tedious and difficult.

(abortions, 0) &
(gravida, 2) &
(pregnancy complication, 2nd trimester bleeding) &
(pregnancy complication, incompetent cervix) &
(pregnancy complication, premature rupture of membranes)
===> (birth, preterm)

Expert verification of the rules deemed all 520 rules usable, since there were no redundant rules, irrelevant attributes, erroneous rules, or meaningless or nonsense rules. In general, the experts indicated that each *individual* rule did not appear to provide enough information and that important data seemed to be missing. Considering the predictive accuracy of the prototype expert system, described in the next section, limitations of expert validation in complex and disorganized domains would benefit from further study.

Results of the Prototype Expert System (Methodology Step 6)

The prototype expert system used 520 rules in an object-oriented expert system shell named "Kappa"[26] that ran in a Windows (Microsoft)[27] environment. Forward chaining techniques and priority encoding of the rules were used to develop the prototype. It is important to remember that none of the testing data was used in building the prototype. A computer program was written to "feed" each of the 9,419 test subjects through the prototype expert system to analyze the system's ability to accurately predict preterm delivery. Accuracy was tested by having the expert system analyze each test case's data and predict either preterm or full-term delivery. The computer program then retrieved the actual preterm or full-term outcome from the database, and the expert system prediction was compared with the actual patient outcome. Where the predicted outcome and the actual outcome matched, there was 100% accuracy. Accuracy rates are reported in Table 2.

Considering the limitations with databases used, "noisy" data, and difficulties encountered with expert validation, the accuracy rates reflected in Table 2 were both surprising and encouraging. The results achieved with database 1 were 88.8% accurate in predicting preterm birth for both low-risk and high-risk pregnant women. Database 2 was 59.2% accurate in predicting preterm delivery in a population of high-risk pregnant women, most of whom were referred for home uterine monitoring because they were in preterm labor. The predictive accuracy of database 2 was less impressive, but it would be expected that predicting preterm delivery in a high-risk group being treated for preterm labor would be more difficult since medical interventions will, it is hoped, influence subsequent birth outcomes. And the seemingly poor results (53.4%) for database 3 were actually quite remarkable from two perspectives. First, there was no rule in the expert system that was derived from data in database 3, suggesting that the expert system may be tapping a construct of preterm birth risk independent of any particular database. Second, even the 53.4% accuracy rate was an improvement over existing manual screening tools that remain only 17–38% accurate.[2]

Discussion

While the ultimate goal of expert system development is to predict preterm *labor* risk, the definition of preterm labor and data needed to analyze preterm labor risk are less amenable to study presently. There were numerous confounding variables in the data that made prediction of preterm labor impossible, and it was determined that accuracy of predicting preterm delivery was more viable. Therefore, the purpose of this study was to determine the feasibility of using machine learning to generate expert system (knowledge-base) rules for prediction of preterm delivery. Each of the databases tested surpassed traditional manual accuracy rates in predicting preterm birth. Future studies are planned to determine the feasibility of using the expert system to predict preterm labor risk.

Future studies using prospective, carefully planned, and quality-controlled data collection methods are expected to improve rule induction and accuracy predictions to high levels in a fully implemented expert system, but this needs testing and validation. The statistical, machine learning, and prototype expert system findings from this study confirmed that preterm risk assessment is a complex and disorganized knowledge domain. But even with this complexity, the research methodology and machine learning techniques used in this study were able to extract rules directly from data and use these rules in a prototype expert system that was more accurate than traditional manual systems in predicting preterm delivery.

Thompson and Thompson[28] recommended adding attributes to improve machine learning classification and suggested that, when selecting attributes, it is better to err on the side of having too many. In other words, the 214 variables analyzed by LERS were inadequate to classify preterm birth for all subjects studied. The notion that additional attributes, or variables, are needed for preterm birth classification is consistent with findings indicating that data items that may be associated with preterm birth risk were missing. For example, the database did not include data about stress, sexual activity, substance abuse, nutritional status, or infections. The overall indication is that the variables needed to predict, or classify, preterm birth were not all available for LERS analysis. Future studies should find improved prediction accuracy as variables are added for machine learning classification.

The problems that prompted this study involve the difficulties encountered in acquiring and processing an overload of information for decision making related to preterm birth risk assessment. While this study provides a foundation for improved preterm birth prediction, the clinical problems associated with accurate assessment and treatment of women at risk for preterm labor need continued research. The knowledge base development methodology used in this study offers a mechanism to further develop linkages between technology and clinical problem solving in a variety of health care settings.

The authors thank Michele VanDyne (IntellDyne, Inc.) for her participation in this study and St. Luke's Regional Perinatal Center (Kansas City, MO), Healthdyne Perinatal Services (Marietta, GA), and Tokos Corporation (Santa Ana, CA) for providing data.

References ■

1. Alexander G, Weiss J, Hulsey T, Papiernik E. Preterm birth prevention: an evaluation of programs in the United States. Birth. 1991;18(3):160–9.
2. McLean M, Walters W, Smith R. Prediction and early diagnosis of preterm labor: a critical review. Obstet Gynecol Surv. 1993;48(4):209–25.
3. Rosen M, Merkatz I, Hill J. Caring for our future: a report of the expert panel on the content of prenatal care. Obstet Gynecol. 1991;77(5):782–7.
4. Lockwood C, Senyei A, Dische M, et al. Fetal fibronectin in cervical and vaginal secretions as a predictor of preterm delivery. N Engl J Med. 1991;315(10):669–74.
5. McGregor J, French J, Richter R, et al. Antenatal microbiologic and maternal risk factors associated with prematurity. Am J Obstet Gynecol. 1990;163(5):1465–73.
6. Woolery L. Knowledge Acquisition for Assessment of Preterm Labor Risk. Unpublished doctoral dissertation. Lawrence, KS: University of Kansas, 1992.
7. Nunnally J. Psychometric Theory. New York: McGraw-Hill, 1978.
8. Grzymala-Busse J. Managing Uncertainty in Expert Systems. Boston: Kluwer Academic Publishers, 1991.
9. Ericsson K, Simon H. Protocol Analysis: Verbal Reports as Data. Cambridge, MA: MIT Press, 1984.
10. Polanyi M. Personal Knowledge: Towards a Post-Critical Philosophy. Chicago: University of Chicago Press, 1958.
11. Johnson P. What kind of expert should a system be? J Med Philos. 1983;8:77–97.
12. Carbonell J. Machine Learning Paradigms and Methods. Cambridge, MA: MIT Press, 1989.
13. Grzymala-Busse J. Knowledge acquisition under uncertainty: a rough set approach. J Intell Robot Syst. 1988;1:3–16.
14. Dean JS, Grzymala-Busse J. An Overview of the Learning from Examples Module LEM1. Lawrence, KS: University of Kansas, 1988.
15. Grzymala-Busse J, Sikora D. LERS1: A System for Learning from Examples Based on Rough Sets. Lawrence, KS: University of Kansas, 1988.
16. Grzymala-Busse J. An Overview of the LERS1 Learning System. Proc Int Conf Ind Eng Applications Artif Intell Expert Syst. 1989:838–44.
17. Budihardjo A, Grzymala-Busse J, Woolery L. Program LERS_LB

2.5 as a Tool for Knowledge Acquisition in Nursing. Proc Int Conf Ind Eng Applications Artif Intell Expert Syst. 1991:735–40.
18. Grzymala-Busse J. LERS: a system for learning from examples based on rough sets. In: Slowinski R, ed. Intelligent Decision Support. Handbook of Applications and Advances of the Rough Sets Theory. Boston: Kluwer Academic Publishers, 1992: 3–18.
19. Pawlak Z. Rough sets. Int J Comput Inform Sci. 1982;11:341–56.
20. Pawlak Z. Rough Sets: Theoretical Aspects of Reasoning about Data. Boston: Kluwer Academic Publishers, 1991.
21. Woolery L, Grzymala-Busse J, Summers S, Budihardjo A. On the use of LERS-LB knowledge acquisition for expert system development in nursing. Comput Nurs. 1991;9(6):227–34.
22. Quinlan R. Learning efficient classification procedures and their application to chess games. In: Michalski RS, Carbonell JG, Mitchell TM, eds. Machine Learning: An Artificial Intelligence Approach. San Mateo, CA: Morgan Kaufman Publishers, 1983:463–82.
23. Quinlan R. C4.5: Programs for Machine Learning. San Mateo, CA: Morgan Kaufman Publishers, 1993.
24. Adalier M. Technical Report: CONCLUS. Unpublished technical report. Lawrence, KS: University of Kansas, 1991.
25. Fieschi M. Towards validation of expert systems as medical decision aids. Int J Biomed Comput. 1990;26(1–2):93–108.
26. Kappa PC. Version 1.2. IntelliCorp, 1991.
27. Windows. Version 3.1. Microsoft Corporation.
28. Thompson B, Thompson W. Finding rules in data. Byte. 1986;11:149–55.

■

Statement of Ownership, Management and Circulation
(Required by 39 U.S.C. 3685)

1A. Title of Publication: Journal of the American Medical Informatics Association
1B. Publication No.: 9 9 9 - 9 1 4
2. Date of Filing: 9/30/94
3. Frequency of Issue: bimonthly
3A. No. of Issues Published Annually: 6
3B. Annual Subscription Price: $92.00
4. Complete Mailing Address of Known Office of Publication (Street, City, County, State and ZIP+4 Code) (Not printers): American Medical Informatics Association, 4915 St. Elmo Street, Suite 302, Bethesda, MD 20814
5. Complete Mailing Address of the Headquarters of General Business Offices of the Publisher (Not printer): 210 South 13th St., Philadelphia, PA 19107
6. Full Names and Complete Mailing Address of Publisher, Editor, and Managing Editor (This item MUST NOT be blank)
Publisher (Name and Complete Mailing Address): Hanley & Belfus, Inc., 210 South 13th St., Philadelphia, PA 19107
Editor (Name and Complete Mailing Address): William W. Stead, MD, 2209 Garland Avenue South, Nashville, TN
Managing Editor (Name and Complete Mailing Address): Sandra Lovegrove, Hanley & Belfus, Inc., 210 S. 13th Street, Philadelphia, PA 19107
7. Owner (If owned by a corporation, its name and address must be stated and also immediately thereunder the names and addresses of stockholders owning or holding 1 percent or more of total amount of stock. If not owned by a corporation, the names and addresses of the individual owners must be given. If owned by a partnership or other unincorporated firm, its name and address, as well as that of each individual must be given. If the publication is published by a nonprofit organization, its name and address must be stated.) (Item must be completed.)

Full Name	Complete Mailing Address
American Medical Informatics Assn.	4915 St. Elmo St., Suite 302 Bethesda, MD 20814

8. Known Bondholders, Mortgagees, and Other Security Holders Owning or Holding 1 Percent or More of Total Amount of Bonds, Mortgages or Other Securities (If there are none, so state)

Full Name	Complete Mailing Address
NONE	

9. For Completion by Nonprofit Organizations Authorized To Mail at Special Rates (DMM Section 424.12 only)
The purpose, function, and nonprofit status of this organization and the exempt status for Federal income tax purposes (Check one)
(1) ☒ Has Not Changed During Preceding 12 Months
(2) ☐ Has Changed During Preceding 12 Months
(If changed, publisher must submit explanation of change with this statement.)

10. Extent and Nature of Circulation (See instructions on reverse side)	Average No. Copies Each Issue During Preceding 12 Months	Actual No. Copies of Single Issue Published Nearest to Filing Date
A. Total No. Copies (Net Press Run)	3,187	2,730
B. Paid and/or Requested Circulation 1. Sales through dealers and carriers, street vendors and counter sales	0	0
2. Mail Subscription (Paid and/or requested)	2,330	2,193
C. Total Paid and/or Requested Circulation (Sum of 10B1 and 10B2)	2,256	2,155
D. Free Distribution by Mail, Carrier or Other Means Samples, Complimentary, and Other Free Copies	74	38
E. Total Distribution (Sum of C and D)	2,330	2,193
F. Copies Not Distributed 1. Office use, left over, unaccounted, spoiled after printing	857	537
2. Return from News Agents	0	0
G. TOTAL (Sum of E, F1 and 2—should equal net press run shown in A)	3,187	2,730

11. I certify that the statements made by me above are correct and complete
Signature and Title of Editor, Publisher, Business Manager, or Owner: Sandra A. Lovegrove