Special Report

Development and Preliminary Testing of EPINFORM, an Expert System for Predicting Wheat Disease Epidemics

JAMES CARISTI, Visiting Scientist, ALBERT L. SCHAREN, Research Leader, U.S. Department of Agriculture, Agricultural Research Service, EUGENE L. SHARP, Professor, and DAVID C. SANDS, Associate Professor, Department of Plant Pathology, Montana State University, Bozeman 59717-0002

Estimates of yield reduction due to disease in cereal crops are used at the national, international, corporate, and local levels. At the national level, the best possible worldwide yield predictions are needed in order to make intelligent trade decisions. Agrichemical companies must be able to plan the manufacturing and marketing of appropriate pesticides. Individual farmers need accurate and early estimates of damage caused by disease in order to decide whether to take countermeasures. In this paper we report the development of two rule-based programs designed to improve the timeliness and quality of disease predictions.

The expert systems approach

Two major wheat diseases are stripe rust caused by *Puccinia striiformis* West. and Septoria nodorum blotch caused by *Leptosphaeria nodorum* Müller. Both fungi are parasitic on portions of the plants above the ground. Stripe rust and Septoria nodorum blotch are usually controlled by planting resistant wheat cultivars and, in certain circumstances, by using chemical fungicides. Although wheat cultivars planted in a region may have some resistance to a disease, extensive yield losses may still occur in times of severe epidemics or in the presence of a new race of the pathogen. Because fungicides are expensive and carry the threat of adverse environmental consequences, one strategy is to apply fungicides only when yield is threatened severely.

The difficulty in this strategy is in discriminating between major and minor threats to yield. Although disease symptoms may be visible on the plants, the effect on yield is not easily determined, since much depends on how many infection cycles have occurred at appropriate times in the various stages of plant growth (10,12).

A computer can help in accurately and tirelessly tracking infection cycles and calculating the expected yield reduction. If a computerized system is to make use of the best available knowledge in plant pathology; be easily adaptable to different regions, cultivars, and climatological conditions; and be useful at both national and local levels, it must be modifiable by people who are not programmers. This type of programming is available through expert systems.

Expert computer systems consist of programming and information that exhibit and utilize expert knowledge in a given field. The expert knowledge is kept separate from general programming logic by putting facts and rules derived from experts into a file called a "knowledge base." The program that then considers and applies those facts and rules is called an "inference engine." The separation of the expert knowledge base and inference engine makes it possible to modify the knowledge without extensive programming changes. It is also possible to obtain explanations from the system concerning the way in which a conclusion was reached. Numerous survey articles present the basic concepts of expert systems in more detail (e.g., 1,2,4,5,8).

Working expert systems have been developed in a variety of areas, including disease control on winter wheat (2), pest and drought control for apple orchard management (2), diagnosis of soybean diseases (9), and determination of irrigation and fertilization schedules for cotton crops (7). These systems differ not only in the expertise contained in their knowledge bases but also in the types of conclusions that can be generated. For example, most expert systems relating to the biological sciences are "diagnostic," in that they attempt to identify an organism or a therapeutic treatment. Some biologically related expert systems focus on the relationships among weather, soil composition, human intervention, and plant development (7).

The success of an expert systems project depends largely on the importance and tractability of the problem under consideration and on the availability of suitable expertise (1,4). The problem itself should be part of a narrow field of knowledge, whereas the value associated with its solution should be substantial. Experts must be willing to invest a considerable amount of time in communicating their expertise, and it would be helpful if they were accustomed to articulating their knowledge to neophytes. The problem-solving process for the human experts should take more than a few minutes but less than an hour. If a solution requires only a moment of the expert's time, methodologies simpler than an expert systems approach would probably suffice. If a solution requires more than an hour of the expert's cogitation, the thought processes probably could not be incorporated adequately in a computer program.

Design constraints

In this study, the expert system EPINFORM was developed for predicting the effects of stripe rust and Septoria nodorum blotch on wheat yields relatively early in a growing season. EPINFORM was designed for use on any IBM or compatible personal computer running DOS 2.1 or greater. This was done to enable individual farmers to use the system on a personal computer with as little as 64K bytes of main memory.

A more difficult design constraint involves the management of unknown data values by the expert system. Many of the conclusions in the knowledge bases depend on having values for critical input variables, such as the number of hours of dew during one night. Although it is possible that the user will know each day how many hours of dew there were the previous night, the system has to allow this value to be unknown at times. In this case, a knowledge base rule is used that estimates hours of dew in terms of other variables. In general, the expert system may determine a value for a variable from a rule in the knowledge base, or it may use a default value, or it may have to ask the user for critical information that cannot be estimated artificially. For example, default values are used for such things

Present address of first author: Department of Mathematics and Computer Science, Valparaiso University, Valparaiso, IN 46383.

Contribution 1895 from Montana Agricultural Experiment Station.

as heading and flowering dates, but the user must supply the amount of precipitation (a critical quantity in the model that cannot be derived in terms of other knowledge).

The knowledge bases of EPINFORM do not contain any knowledge of plant growth and development. Although there are many models of plant growth that depend on climatological data, such models do not always correspond exactly to what is occurring in the field. To avoid introducing damaging plant growth errors, the knowledge bases state conclusions as a percentage depreciation of the expected yield in the absence of disease. Of course, those using EPINFORM would find it possible to incorporate a plant growth model into a knowledge base by including additional rules.

Construction of EPINFORM

The expert system EPINFORM consists of several files and programs. The major component of the system is the inference engine, INFER (Fig. 1). This program issues requests to the user for daily data and then progresses through the knowledge base, recording conclusions both on the screen and in a file that maintains a log of important events, e.g., the predicted occurrence of an infection. Another program called EXPLAIN can provide justifications for derived conclusions, when desired, by retrieving conclusions and the conditions that generated them during a specified time interval. This is accomplished by examining the appropriate parts of the log and database files, then printing a compact report of selected variables.

The knowledge bases consist of "if-then" rules that can be understood and modified, even by users who are not experienced programmers. Figure 2 shows some of the rules contained in the stripe rust knowledge base. The complete stripe rust knowledge base currently has 54 rules, 18 of which estimate yield reduction either before the flowering date (e.g., rule 27) or at the flowering date (e.g., rule 28). The conclusions of the rules involve either printing and logging a deduction (rules 10, 27, and 28) or assigning a value to a variable (all the other rules). The conditions of the rules are expressions involving variables and any combination of elementary arithmetic and logical operators. Variables can be integer, real, string, or Boolean in type. The actual value of a variable may be unknown. Whenever such a variable is encountered by the inference engine as part of a condition, a rule is sought that concerns the variable as the conclusion. When no rule can determine the value of the variable sought, the system uses a default value or, in the case of a critical variable, requests that the user supply a value.

Each knowledge base models the progression of a disease. Both stripe rust and Septoria nodorum blotch occur in cycles that begin with the infection of leaves by spores and end with the production of new spores days later. The knowledge base rules maintain a cycle clock for tracking the infection cycles. The clock is started when the climatological conditions for an infection exist (Fig. 2, rule 11). These conditions are associated with the germination of a large majority of the available inoculum. The influence of any remaining spores and intermediate cycles they might initiate is assumed to be negligible. Thus, for our purposes, cycles are always distinct. When the clock reaches the cycle length, as determined by a knowledge base rule (rule 15 or 16), another rule records the effect of the cycle (rule 7 or 8), and the system awaits conditions necessary for reinfection (rules 9–11). Several infection cycles must occur before their cumulative influence would cause the rules to indicate any prospective yield reduction.

Both knowledge bases involve the assumption that some effective initial inoculum is present to begin an infection. These pathogens are not killed by extremely cold winters because of their ability to survive as pycnidiospores in fruiting bodies, as dormant mycelium in leaves (11), or on senescent leaves and plant debris (12). Furthermore, in regions where these diseases are often damaging to yields, there are so many spores constantly available that it seems reasonable to be unconcerned with the proliferation of the spores.

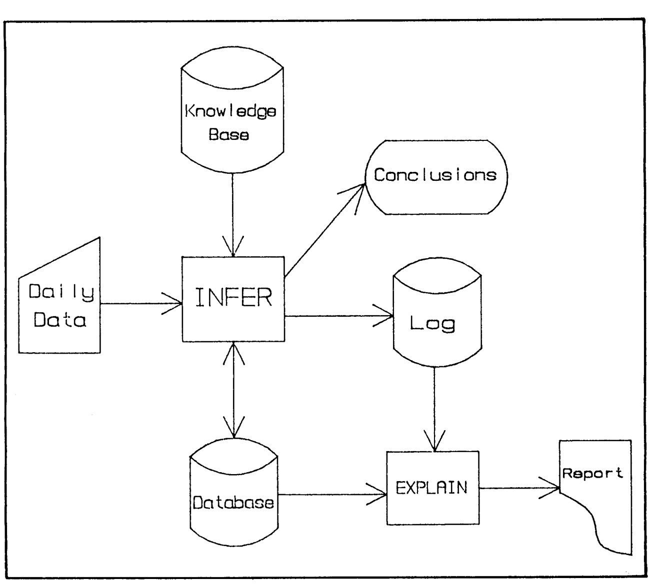

Fig. 1. EPINFORM system flowchart.

Accuracy of EPINFORM's predictions

Because expert systems involve the separation of expert knowledge from other programming, the rules contained in a knowledge base can be modified as a result of experience in running the system. Consequently, further development of the knowledge bases is expected as testing takes place. An expert system is presumed to require many years of refinement during its entire useful life (4).

Ideally, each knowledge base would be tested by comparing its prediction of yield reduction with the actual percentage of yield reduction caused by the disease. Obtaining documented cases of yield reduction attributable only to the disease in question is difficult, however. The yield testing performed thus far has involved yield data obtained through fungicide trials. Since a fungicide will usually affect more than one pathogen, we expect the percentage yield difference between treated and nontreated plants to be larger than that predicted by EPINFORM.

Data from the years 1975, 1976, and 1978 for Pullman and Walla Walla, Washington, all indicate larger losses from stripe rust than those predicted by EPINFORM (Table 1). Other diseases, such as leaf rust, Septoria tritici blotch, and powdery mildew, were present to varying degrees in the experimental plots examined. Because the diseases were affected to different extents by the fungicides applied and because EPINFORM was predicting losses caused only by stripe rust, we believe the predictions of losses attributed to stripe rust to be reasonably accurate. In 1981, when the principal losses were caused by

Table 1. Percentage yield differences observed in fungicide trials and predicted by EPINFORM for stripe rust

Year	Location	Observed[a]	Predicted
1975	Pullman, WA	21	10
1976	Pullman, WA	42	10
1978	Pullman, WA	19	10
	Walla Walla, WA	17	10
1981	Pullman, WA	68[b]	70
1982	Pullman, WA	13[c]	0

[a] Percentage yield difference: (treated − check)/treated for stripe rust-susceptible cultivars. Data from *Fungicide and Nematicide Tests*, vols. 31, 32, 34, 37, and 38.
[b] Leaf rust reported to have no influence on yield.
[c] Leaf rust reported to be predominant.

stripe rust alone, the prediction of EPINFORM was very close. And in 1982, when only leaf rust symptoms were evident, EPINFORM correctly predicted no yield losses due to stripe rust.

Another kind of testing is needed to determine whether EPINFORM's predictions of disease cycles are valid. For stripe rust, field measurements of disease severity have been published. Severity refers to the percentage of a leaf showing disease symptoms. Severity readings taken at various stages of plant growth can be compared with predictions made by EPINFORM.

EPINFORM was tested against all early-season stripe rust readings that we have been able to locate. We obtained daily high and low temperature and precipitation readings from climatological data provided by the National Oceanic and Atmospheric Administration. A 10-year mean was used to

```
 1. Humidity:95 IF AridArea and Precipitation > .01;
 2. Humidity:75 IF AridArea and
                   Precipitation > 0 and Precipitation <= .01;
 3. BootDate: HeadingDate - 5 ALWAYS;
 4. HoursDew:8 IF Humidity > 80;
 5. HoursDew:4 IF Humidity <= 80 and Humidity >= 70;
 6. CycleDay: CycleDay + 1 IF CycleDay>0;
 7. Severity: Severity*10 IF Severity>0 and
                          CycleDay>=CycleLength;
 8. Severity:1 IF Severity=0 and Prevalence>=100 and
                 CycleDay>=CycleLength;
 9. CycleDay:0 IF CycleDay>=CycleLength;
10. "INFECTION HAS OCCURRED" IF
                      HoursDew>4 and
                      (LowTemp>33 and LowTemp<56) and
                      (CycleDay=0 or CycleDay>CycleLength);
11. CycleDay:1 IF HoursDew>4 and
                 (LowTemp>33 and LowTemp<56) and
                 (CycleDay=0 or CycleDay>CycleLength);
12. TotalLows:0 IF CycleDay=1;
13. TotalLows: TotalLows + LowTemp IF CycleDay>0;
14. AverageLow: TotalLows/CycleDay IF CycleDay>0;
15. CycleLength:20 IF AverageLow<=41;
16. CycleLength:18 IF AverageLow>41 and AverageLow<49;
17. Prevalence:100 IF Prevalence<100 and CycleDay<5 and
                     HoursDew>8 and
                     (LowTemp>41 and LowTemp<49);
18. Prevalence: 11*Prevalence IF
           Prevalence<100 and CycleDay<5 and
           (( HoursDew>8 and (    (LowTemp>33 and LowTemp<43)
                               or (LowTemp>49 and LowTemp<56)))
             or ( HoursDew>4 and HoursDew<9 and
                 LowTemp>41 and LowTemp<49));
19. InfectionType:8 IF AverageLow<=59 and CycleDay=4;
20. InfectionType:2 IF AverageLow>=67 and CycleDay=4;
21. ResponseSymbol:0.1 IF InfectionType=2;
22. ResponseSymbol:1 IF InfectionType=8;
23. ResponseSymbol: ResponseSymbol/10 IF Resistance="R";
24. Coefficient: Severity*ResponseSymbol ALWAYS;
25. BootCoeff:Coefficient IF JulianDate=BootDate;
26. FlowerCoeff:Coefficient IF JulianDate=FlowerDate;
27. "YIELD LOSSES DUE TO RUST MAY BE DEVASTATING" IF
              CycleDay=4 and Coefficient>=50 and
              JulianDate<FlowerDate;
28. "YIELD REDUCTION DUE TO RUST WILL BE 50 TO 70%" IF
              JulianDate>=FlowerDate and
              FlowerCoeff>=50 and FlowerCoeff<90 and
              BootCoeff>=9 and BootCoeff<90;
```

Fig. 2. Some of the rules from the stripe rust knowledge base.

Table 2. Stripe rust severity, observed and predicted

Year	Location	Dates of measurements	Number of severity observations recorded in categories[a] <10%	10-50%	>50%	Severity predicted by EPINFORM (%)
1984	Pullman, WA	2 and 3 July	3	3	4	10–50
1983	Pullman, WA	14 June	0	3	0	10–50
		17 June	0	3	0	10–50
		28 June	0	2	1	>50[b]
1981	Pullman, WA	26 and 27 May	4	2	0	<10
		1–5 June	0	3	3	10–50
		15 June	0	1	5	>50
		30 June–1 July	0	1	5	>50
1977	Walla Walla, WA	30 May–2 June	0	3	0	10–50
1975	Pullman, WA	22 June–3 July	1	2	1	10–50
1974	Pullman, WA	9–15 June	6	3	0	<10

[a] Data from "Results from Cooperative Wheat Varietal Experiments in the Western Region of the USA." An observation refers to severity reading recorded for a particular cultivar at a particular farm. Categories correspond to theoretically distinguishable levels of infection.
[b] Severity predicted for 27 June was in 10–50% range.

estimate the date on which plants emerge from winter dormancy (3). The expert system was run from that point to produce its predictions. The results are given in Table 2. EPINFORM seems to identify well the times when observed severity measurements are uniformly high or low.

The Septoria knowledge base does not consider severity but, rather, the number of disease cycles occurring relative to important points in the plants' development. Thus, the testing of this knowledge base must rely on documented statements of yield reduction attributable to Septoria nodorum blotch. Since estimated yield reduction due to disease is usually stated qualitatively (high, moderate, or low) in wheat reports, thorough testing of the Septoria knowledge base is anticipated to take much longer than the stripe rust testing. We have verified that EPINFORM's predictions of yield reduction caused by Septoria nodorum blotch correspond to our human expert's predictions when provided with the same input data.

Obtaining adequate historical data for testing is difficult. In addition to daily high and low temperatures and precipitation readings, the expert system must also be supplied with data such as the date on which plants emerge from winter dormancy and the date on which most of the plants reach certain stages of growth. These dates vary for different cultivars and regions and from year to year, thus complicating the testing process. Reports of actual disease intensities must be found that include field location, wheat cultivars grown, and dates on which readings were taken. Only those readings taken at or before the milk stage of plant development are used, since predictions must be made relatively early to be of value.

Difficulties in obtaining documented disease intensities and yield consequences have limited testing based on published data. The initial results reported are extremely encouraging, however. Both knowledge bases have been provided to over a dozen researchers who work in different experimental environments for real-time testing. The generation of a sufficient amount of field data for use in testing the system's validity is expected to take two more years.

In contrast to a regional model for predicting stripe rust on winter wheat developed by Coakley et al (3), our expert system is more local and, more important, actually models the progress of the disease. The Coakley et al model is statistical and focuses on the key variables of degree days and disease intensity. Although it is subject to error on a local level because of particular climatological conditions relative to the disease, the Coakley et al model seems to be quite accurate for regional predictions. Conversely, our expert system appears to give good predictions at the level of an individual weather reporting station, but some method would have to be developed to integrate results from many reporting stations to give predictions for a larger region.

The expert systems could be expanded to include a cost-benefit analysis for fungicide use. If this were done, market expectations and fungicide costs would have to be included in the knowledge bases. Another area for possible enhancement of the system could involve making use of in-field weather sensors as described by Jones et al (6). We are grateful to W. L. Pedersen for pointing out that the rule-based nature of EPINFORM and its ability to generate explanations suggest it could be applied usefully in a teaching situation. A student of epidemiology could employ a variation of parameters technique with the expert system to determine, for example, the sensitivity of the disease to diverse environmental conditions.

EPINFORM is in the public domain and can be obtained at cost ($10) from the authors. An IBM formatted floppy disk contains all the programs and files as well as a user's manual. The source code also is available on request at no additional charge.

LITERATURE CITED

1. Bobrow, D. G., Mittal, S., and Stefik, M. J. 1986. Expert systems: Perils and promise. Commun. ACM 29:880-894.
2. Buchanan, B. G. 1986. Expert systems: Working systems and the research literature. Expert Syst. 3:32-51.
3. Coakley, S. M., Line, R. F., and Boyd, W. S. 1983. Regional models for predicting stripe rust on winter wheat in the Pacific Northwest. Phytopathology 73:1382-1385.
4. Davis, R. 1984. Amplifying expertise with expert systems. Pages 17-40 in: The AI Business: The Commercial Uses of Artificial Intelligence. P. H. Winston and K. A. Prendergast, eds. MIT Press, Cambridge, MA. 324 pp.
5. Hayes-Roth, F. 1984. The knowledge-based expert system: A tutorial. Computer 17:11-28.
6. Jones, A. L., Fisher, P. D., Seem, R. C., Kroon, J. C., and Van DeMotter, P. J. 1984. Development and commercialization of an in-field microcomputer delivery system for weather-driven predictive models. Plant Dis. 68:458-463.
7. Lemmon, H. 1986. COMAX: An expert system for cotton crop management. Science 233:29-33.
8. Michaelsen, R. H., Michie, D., and Boulanger, A. 1985. The technology of expert systems. Byte 10:303-312.
9. Michalski, R. S., Davis, J. H., Bisht, V. S., and Sinclair, J. B. 1983. A computer-based advisory system for diagnosing soybean diseases in Illinois. Plant Dis. 67:459-463.
10. Sharp, E. L. 1965. Prepenetration and postpenetration environment and development of *Puccinia striiformis* on wheat. Phytopathology 55:198-203.
11. Sharp, E. L., and Hehn, E. R. 1963. Overwintering of stripe rust in winter wheat in Montana. Phytopathology 53:1239-1240.
12. Shipton, W. A., Boyd, W. J. R., Rosielle, A. A., and Shearer, B. I. 1971. The common septoria diseases of wheat. Bot. Rev. 37:231-262.