

The Evaluation and Impact of NEPER Wheat Expert System

Ahmed Rafea
Computer Science Dept., AUC
Email: rafea@aucegypt.edu

Mostafa Mahmoud
Central Lab. for Agriculture Expert System, ARC
Email: mostafa@esic.claes.sci.eg

Abstract: This paper presents the laboratory and field evaluation results of NEPER Wheat expert system. The laboratory evaluation showed that NEPER performance is comparable with human experts. Field evaluation has revealed that NEPER has good economic and environmental impacts. The field testing results have also shown that NEPER is usable, applicable and needed. *Copyright* © 2001 IFAC

Keywords: Expert Systems, Diagnosis, Knowledge-based systems, Hierarchical structures, Classification, Intelligence.

1. INTRODUCTION

Bread, known as aish, or life, is a vital component of the Egyptian diet. In 1993, the country produced 4.5 million tons of wheat on 2.2 million feddans. Given the crucial role wheat plays in Egypt, CLAES cooperated with the Intelligent Systems laboratory (ISL) at Michigan State University in developing the Egyptian Regional Wheat Management System, funded by NARP a United States Agency for International Development (USAID) project in the period from 1992 to 1995. This project integrates an ES with a crop simulation model and aims at addressing all aspects of irrigated wheat management in Egypt. This integrated system is named NEPER (Kamel et al, 1995). In order to achieve this goal, NEPER is designed to perform the following functions:

- Select the appropriate variety for a specific field
- Advise the farmer on field preparation
- Design schedules for irrigation and fertilization
- Control pests and Weeds
- Manage harvests
- Diagnose malnutrition
- Diagnose disorders
- Suggest Treatments

In 1997, another project between CLAES and ISL was funded by the ATUT which is also a USAID project. One of the objectives of this project was conducting field-testing to measure the ES performance. The objective of conducting the field-testing was to evaluate the economical and environmental impacts and to measure the ES performance from three aspects: usability, applicability and need. The results of this testing were also used to enhance the user interface and extend the knowledge base of Neper .In this project,

there is a component for the evaluation of a new enhanced version of NEPER that considers the whole agricultural operations. This new version has been developed according to the results and recommendation of the field testing that is presented in this paper.

In this paper, the technical background of NEPER is presented in section 2. The laboratory evaluation is summarized in section 3. The experiments description is presented in sections 4. The economical and environmental impact of the ES are summarized in sections 5 & 6, respectively. The ES performance, section 7, has been measured using three aspects namely usability, applicability, and need of ES. ES enhancements as a result of those experiments are presented in section 8.

2. BACKGROUND

In developing Neper, the Generic Task Approach to ES development proposed by Chandrasekran (Chandrasekran, 1986) has been used. The idea behind the Generic Task approach, is that the way a problem is to be solved, depends largely on its type e.g., diagnosis, design, planning, etc. Consequently, problems of the same type could share some sort of a generic problem solver. So, according to the Generic Task methodology approaching a diagnosis problem will be inherently the same regardless of the domain in which such a problem is being addressed. The classical example of a problem solver that could be applied to a diagnosis problem is Hierarchical Classification (Gomez & Chandrasekran, 1981; Chandrasekran, 1983) and it is this problem solver that has been used in implementing the Wheat disorders ES, which is a component of NEPER.

This system component has been implemented using a Generic Task Tool developed at Michigan State University (MSU). In this tool, the knowledge base is created as a hierarchy of nodes. In each node, the knowledge is represented in a table, where each entry in this table represents either a database variable or a variable pointing to another table. Each database variable is associated with a question. A user will be presented with that question only if the database variable has never been assigned a value. The combination of possible inputs for each question denotes different rules and matching patterns. If a combination of inputs results in a match value greater than a given threshold, the node is said to be established. By asking the user a series of questions, the system is able to pursue or rule out paths in the classification in which the leaves represent disorders. Basically, if a path from a root to a leaf exists, then the disorder at the leaf is presented as the diagnosis.

3. Laboratory Evaluation

Laboratory evaluation is conducted before disseminating the ESs in the field. The Laboratory evaluation methodology consists of three main procedures namely Verification, Validation, and Evaluation. Verification is defined as the demonstration of consistency, completeness, and correctness of software (Adrion et al, 1982). O'Keefe et al. (1987, 1989, and 1990) have defined verification as "Building the system right", that is making sure that the implemented system is functionally matching the proposed design, and free of semantic and syntactic errors. Validation is the process whereby the system is tested to show that its performance matches the original requirements of the proposed system. It is defined as the determination of the correctness of the final program or software produced from a development project with respect to the user needs and requirements (Adrion et al, 19982). As noted by O'Keefe et al. (1987, 1989, and 1990) "Validation means building the right system". Evaluation is the process whereby we ensure the usability, quality, and utility of the ES (O'Keefe et al. 1987, 1989, and 1990). A complete testing cycle is performed in iterations through which, the ES is updated and refined.

Verification process evolves through two main stages during the development of the ES: the development stage and the examination stage. In the development stage, the developer practices different functions of the implemented systems, looking for potential errors that may exist. This is accomplished using two broad techniques: non case-based and case-based. Non case-based techniques include tracing, spying and other traditional debugging techniques. Case-

based verification techniques are applied by preparing "Typical Cases". These cases should be selected to serve requirements satisfaction as spelled out in the requirement specification. In the examination stage, the ES is tested to make sure that it is running properly, by testing all the functions of the system trying to examine the performance of the system in different situations. The output of this stage is the *verification report* that is a document of differences between system design and implementation. This report is used to update , the design document and implementation.

The validation step is done through conducting meeting with the doamin experts who provided the knowledge to check that the right system has been developed. This is done by going throug the generated test cases during the meeting with the domain experts. Their comments on the content and user interface are considered. Necessary updating of the design and implementation is done.

The evaluation step is to assess the quality, usability, and utility of the ES from the point of view of human experts other than the domain expert, who participate in the system development. Typical cases are created and distributed to three domain experts in the specialty of a specific sub system. If one sub system includes more than one specialty, cases are distributed to all experts in different specialties. For example in the remediation subsystem, we have three specialties: plant pathology, entomology, and nutrition. Therefore 9 experts have participated in the validation of this subsystem. For each specialty, an evaluator is selected to blindly assess the responses of the three human experts and the ES. After the evaluation, the domain expert participated in the development, the evaluator, and the domain experts participate in an evaluation meeting together with the knowledge engineer to discuss the evaluation results till they reach to a consensus.

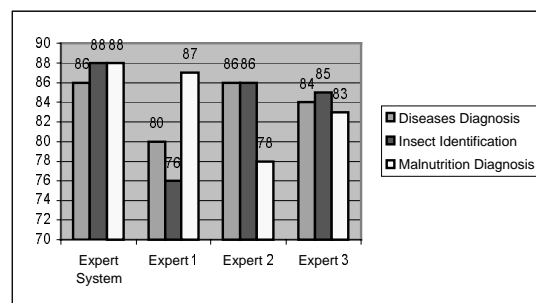


Figure 1 NEPER Diagnosis evaluation result

Applying this methodology on NEPER, verification and validation were done successfully. In this paragraph we will presnt the evaluation results of the diagnosis and treatment subsystems. Figure 1 shows the evalaution scores of NEPER diagnosis

subsystem. NEPER diagnosis subsystem over performs human expert in the insect and malnutrition specialties, and its score in the disease diagnosis results (86%) is equivalent to those of the best human expert.

The evaluation scores of NEPER treatment subsystem is shown in figure 2. NEPER treatment over performs human expert in disease treatment, and its score in the insecta and malnutrition treatment are 0.95 and 0.85 respectively of the best expert-group. After this experiment, NEPER has been trained to reach the scores of the best experts..

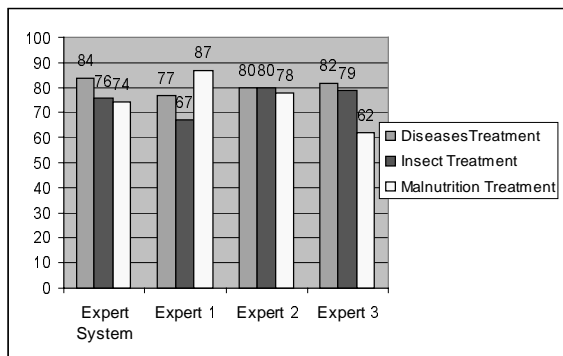


Figure 2 NEPER Treatment evaluation result

4. Field Test Experiment Description

Many experiments were conducted in the last few years for NEPER ESs. The objectives of those experiments were to validate the system in the field, and to measure the impact of using the system. The experiments were conducted in different locations by selecting two fields at same area and location: one is to be cultivated using NEPER Wheat recommendations without any interference from the agriculture engineer or any specialist, and the other one is to be cultivated as usual, this is a control field.

In order to get the best results from the experiment, the following issues and activities were considered and followed:

- ◆ Formal training on the usage of NEPER was conducted for the staff who are going to use the system.
- ◆ A computer engineers from CLAES were responsible for supporting the site staff on the usage of the system and handling trouble-shooting problems of hardware and software.
- ◆ A number of the wheat researchers from Field Crop Research Institute (FCRI) were assigned to supervise different fields, i.e., a researcher for each site.
- ◆ Periodical fields visits were conducted by researchers from CLAES and FCRI

Three experiments were conducted in three different seasons for NEPER ES. Two of those three experiments had the same number of fields, both of them consisted of a total number of 32 fields carefully selected for conducting the experiment. The third one consisted of a total number of 44. These fields were equally divided so that 16 fields in the first two experiments and 22 fields in the third one were assigned to utilize NEPER and managed by the ES, and the other fields were to be managed in the usual practice and acts as control. The selected fields were located in four different geographical areas, namely: Noubaria, Gemiza, Sharkia, and Decerns. In Noubaria two sites were selected to cover the different types of soil at that area. One of those sites located in Bostan and the other one located in Banger El-Sokar.

The first experiment covered only the diagnosis and treatment part of the NEPER Wheat ES including Weed Identification. The second one included the strategic part and tactic part. Strategic part includes six subsystems called: Variety Selection, Pre-cultivation Pest Control, Tillage, Planting, Irrigation & fertilization, and Harvest. The third one also included the strategic part and tactic part. Strategic part includes six subsystems called: Variety Selection, Planting, Land Preparation, Irrigation, fertilization, and Harvest. Tactical part includes two subsystems called diagnosis and weed identification, each of them includes the treatment function.

5. Economical Impact

In the first experiment (CLAES, 1996), the averages of treatment costs, yields, and straw per feddan was calculated for both NEPER and the control fields. By taking the averages of treatment cost, yield, and straw per Feddan, it was found that the average net income per feddan for ES fields is 2049.85 LE and for control fields is 1600.05 LE, consequently, the net production increase in Egyptian Pound was 449.8. This represents 26.78% increase in the production

In the second and third experiments (CLAES, 1999, CLAES, 2001), the complete system was tested. Tables (1) and (2) summarize the result of those experiments in the new reclaimed area and the Delta area. The following remarks were observed:

- In both the new reclaimed and Delta area, there was an increase in the production and net profit consistently in the two consecutive seasons.
- The percentage of increase in the net profit in the newly reclaimed is greater than the percentage of increase in the net profit in the Delta area.
- The production in the newly reclaimed area is less than the Delta area because the lack of expertise

Table 1: The result of the experiments in the new reclaimed area

Item	Season 97/98				Season 98/99			
	ES	Control	Differance	%	ES	Control	Differance	%
Average Production	1701	1506	195	13	1647	1431	216	15
Average cost	747	861	-114	-13	468	603	-135	-22
Average Net profite	954	645	309	48	1179	828	351	42

Table 2: The result of the experiments in the Delta area

Item	Season 97/98				Season 98/99			
	ES	Control	Differance	%	ES	Control	Differance	%
Average Production	2130	1830	300	16	2117	1759	358	20
Average cost	631	597	34	5	445	422	23	5
Average Net profite	1499	1233	265	22	1672	1337	335	25

in the reclaimed area. Hence expertise transfer in this area has led to a relatively high impact.

6. Environmental Impact

The conservation of natural resources has two aspects. The first is pertinent to the management of these resources on the macro level, such as controlling the expansion of urban development in order not to loose agricultural land. The second is concerned with the management of these resources on the micro level such as adding chemical fertilizers to the soil. In this paper, the focus will be on the status of the water and land resources because they are the two main resources related to our work on crop management ESs.

Water is the scarcest resource in Egypt, since its supply is nearly fixed and water demand for different sectors is continuously increasing. The water supply can be classified into three categories: surface water, ground water, and (isle) water reuse after treatment either from agriculture drainage or domestic usage. The decision makers concerned with water resource management in Egypt are challenged by how to balance the limited water supply with an increasing water demand for the future, since water is the major constraint for land expansion to satisfy food self-sufficiency. Another challenge is how to reduce the water pollution resulting from using chemical fertilizers and pesticides. After water, land is the major limiting factor for sustainable agricultural development (Rafea, 1996).

There are two problems facing decision-makers to conserve water resources namely: the efficient utilization of water resources, and the pollution resulting from the usage of chemical fertilizers and pesticides. Regarding soil conservation, there are two main problems namely: the urban expansion, and the soil degradation resulting from excessive use of fertilizers and other bad agricultural practices. Therefore, the main contribution of ESs for soil and water conservation is to transfer the agricultural practices according to certain strategy or a combination of strategies namely: environmental sustainability, economical sustainability and/or social

sustainability. In the ESs that have been built so far, we are concerned with economic sustainability taking into consideration the environmental sustainability in the second place. In other words, we are trying to acquire the recommendations that optimize the output relative to the agricultural inputs. As a consequence, environmental conservation is achieved, because no extra input is provided such as water, fertilizers and pesticides without a return in the yield.

The results of experiments conducted for the ES agree with the goals of environmental conservation. The fields managed by the ES have used fewer resources in terms of fertilizers and pesticides than the control fields and hence conserve environment. The cost is an indicator of the increase or decrease of using chemicals in general. Hence, we have used the cost as a factor in determining the quantity of used fertilizers and pesticides.

The average cost of pesticides used by NEPER Wheat fields in the first experiment was more than the control fields by 15.7 Egyptian pound/Fadden, but the production increased by 449.8 Egyptian pound/Fadden. Notice that the increase of cost in this experiment is negligible. In the second experiment the average cost of fertilizers and pesticides used by NEPER Wheat fields was less than the control fields by 5.57 Egyptian pound/Fadden and the production increased by 247.4 Egyptian pound/Fadden. In the third experiment the average cost of fertilizers and pesticides used by NEPER Wheat fields was less than the control fields by 1.2 Egyptian pound/Fadden and the production increased by 287.12 Egyptian pound/Fadden. In fact, this indicates that changes to the NEPER Wheat system have made it more compliant to the goals of resource management and environmental conservation.

In the second experiment of NEPER wheat ES, the average water quantity used to produce one Ardab of wheat in the ES fields was 112.74 M3 water, while in the control fields the farmers used 152.58 M3 water on average to produce the same quantity of wheat. This represents 35% decrease in the use of irrigation water.

7. Expert System performance

The expert system performance has been measured using three aspects namely usability, applicability, and need of ES.

7.1 Expert Usability

In order to measure the usability of the ES, the developers in CLAES have re-run the system on the cases reported in the forms of the fields managed by NEPER and compared the conclusions with the results represented in the field books by the researchers and extension agriculture engineers in different locations. In the first experiment (CLAES, 1996) was examining the comparison results, it was found that in 86% of the cases, the trained researchers have used the system correctly while this percentage has decreased to 38% for untrained researchers. This indicates the importance of training on the usage of the ES. It is worth noting that there is no great difference between the researchers and extension officers in using the system as the differences was only 4%, although the system was in English. This proves the importance of ES. It raised the performance of extension officers to the level of researchers, in the underlying domain of the NEPER.

In the second experiment there was discrepancy between the ES recommendation and the agriculture practices documented in the field books. When this discrepancy was discussed with the ES users, we found that this discrepancy was due to their rejection of the ES recommendation and not due to bad use of the system. Therefore, we concluded that the usability of the system in the second experiment was high.

7.2 Expert system applicability

The applicability can be measured by comparing the ES recommendation and to what extent the ES users have applied them. This discrepancy must not be due to bad usability of the system.

In the first experiment, the discrepancy between the ES results and the applied practices by the users were due to bad use of the system.

In the second experiment (CLAES, 1999), it was difficult to quantify the comparison result as it was found that sometimes the recommendations are applied partially. Hence qualitative measures were found more appropriate, especially in the strategic part. The applicability of the modules: Pre-Cultivation Pest Control, Planting, and Weed Control was found low because in Noubaria fields' users did not accept the ES recommendations of the pre-cultivation pest control and the planting modules. In the weed control module the actual practice is different from the ES advice. The applicability of the modules: Tillage and Fertilization are moderate, as the ES fields' users

did not accept the ES recommendation in about 50% of the cases. The applicability of the modules Diagnosis and Treatment were above moderate as the advice of the ES fields supervisors matches the advice generated by the ES in the range of 80 to 87.5% in diagnosis and 50% of the cases in treatment. The applicability of the modules: Variety Selection, Irrigation, and Harvest are high. In the Variety Selection, the ES recommendations are compatible with the actual varieties cultivated in the ES fields. In the Irrigation, the Delta area (there is only 10% difference). In the Harvest, the ES recommendations are compatible with the actual practice in the ES fields.

7.3 Need of Expert System

In order to measure the need of NEPER, a comparison has been done between the advice given by the researchers and extension workers supervising the control field in the experiment locations and the advice that would be generated if NEPER were used. In the first experiment (CLAES, 1996), examining the comparison results it was found that the ES performance is better in 76% of the cases, and hence there is a great need for having the ES.

In the second experiment (CLAES, 1999), it was found that there is a high need for the ES modules: Tillage, Irrigation, Fertilization, Diagnosis, and Treatment. In the Tillage module, it was found that the performance of the ES is better as all control fields supervisors did not apply laser and plowers, appropriately. In the Irrigation module, it was found that the ES recommends less water than what was recorded in the control fields books. In the fertilization module, it was found that the ES is better as ES recommends the adequate quantities of phosphorus and potassium fertilizers whereas some control fields did not apply these types of fertilizers at all. In diagnosis and treatment modules, the performance of the ES is better as the advice of the control fields supervisors match the advice generated by the ES in only 37.5% of the diagnosis cases, and 20% of the cases in treatment. The experiment showed that there is a need for such module especially if the treatment part is modified to be more applicable.

8. Expert System Enhancements

According to the results obtained from field-testing, the following enhancements were done:

- Arabic language support was introduced.
- The irrigation module was revised to be accepted by users.
- User interface become more flexible.
- Basic information about the field and the environment have been included in the reasoning

(i.e. drainage system, previous crops, water source, length and width of the field, etc.)

- The variety selection module has been enhanced to produce the most suitable variety for each field and produce justification for this selection.
- Basin recommendation has been revised completely.
- The harvest module has been enhanced to generate real advice about the suitable date of start harvest

The following enhancements were also suggested and the ES are going to include them:

- Most of the users were unable to understand what was meant by some operations so, more explanation like video clips should be provided.
- Some of the terms are difficult to understand, e.g., “spindly stem” and “leaf chlorosis”. Consequently, pictures are necessary for symptoms at different plant growth stage.
- Currently, the ES is capable of diagnosing severe nutrition deficiency. However, it is not equally capable of detecting early stages of nutrition deficiency. This should be rectified. A very good example of this is Nitrogen deficiency.
- Drought and Water Logging should be covered by the system specially that their symptoms coincide with the symptoms of Nitrogen def. and Potassium def.

9. CONCLUSION

The work done in this project has revealed and emphasized the effectiveness and importance of ES as a decision support tool for extension services. It was very clear that there is a difference in the advice quality and consistency given by the ES and the extension agriculture engineers.

In the mean time, field experiments showed that Usage of ES has an economic and environmental impact. Currently there are efforts to disseminate NEPER, nation wide, and to avail it on the Internet.

The field testing was found to be very useful as many aspects of the usability, applicability, and need were not possible to be identified without this field test. NEPER was found to be user friendly, and can be used by both researchers and extension workers. The recommendations generated by NEPER were applicable in most of the cases. The cases that were not accepted by the researchers and extension workers conducting the experiment, were discussed and the right recommendations were included in the successor version. Most of the NEPER modules are found to be needed. The modules which were found not needed, were examined. The result was that this was not needed by the researchers and extension

workers conducting the experiment but they are badly needed by the growers and extension workers in remote locations.

REFERENCES

- Adrion, W., Branstad, M., Cherniovsky, J. Validation (1982) "Verification and Testing of Computer Software" ACM Computing Surveys, Vol. 14, No. 2, 1982
- Chandrasekran, B. (1986). Generic Tasks in Knowledge-Based Reasoning: High-Level Building Blocks for expert system design. IEEE Expert, 1(3), 23-30.
- Chandrasekran, B. (1983). Towards a Taxonomy of Problem Solving Types. AI Magazine, 4(1), 9-17.
- CLAES (1996) "Validating NEPER Wheat Expert System and CERES Wheat Simulation Model", Technical report, No: TR/CLAES/ATUT(1)/3/96.12, 1996.
- CLAES (1999) "Validating NEPER Wheat Expert System - Field testing for season 97/98", Technical report, No: TR/CLAES/ATUT(w4)/5/99.2, 1999.
- CLAES (2001) "Validating NEPER Wheat Expert System - Field testing for season 98/99", Technical report, No: TR/CLAES/ATUT(w4)/10/2001.3, 2001.
- Nazareth, D (1989) Issues in the verification of knowledge In Rule- based System; International Journal of Man-Machine Studies, Vol.30, 1989, PP.255-271.
- O' Keefe R.M (1990) "Consultant Report" Report No-CR-88-024-08 the Expert Systems for Improved crop management project. Project No EGY/88/024, August 1990
- O' Keefe, R.M., O. Balci, and E. P. Smith, (1987) "Validating Expert System Performance" IEEE Expert, Vol. 2, No. 4, Winter 1987, PP 81-90.
- O' Leary, D., O'Keefe, R. (1989) "Verifying and Validating Expert Systems", Tutorial: MP4, IJCAI, 1989.
- Rafea, A. (1996) "Natural Resources Conservation and Crop Management Expert Systems", Workshop on Decision Support Systems for Sustainable Development, UNU/IIST, Macau. 26 February - 8 March, 1996.
- Gomez, F., & Chandrasekran, B. (1981). Knowledge Organization and Distribution for Medical Diagnosis. IEEE Transactions on Systems, Man, and Cybernetics, SMC-11(1), 34-42.
- Kamel, A., Schroeder, K., Sticklen, J., Rafea, A., Salah, A., Schulthess, U., Ward, R. and Ritchie, J. (1994). Integrated Wheat Crop Management System Based on Generic Task Knowledge Based Systems and CERES Numerical Simulation. AI Applications 9(1):17-27