Artificial Neural Networks
Can Distinguish Novice and
Expert Strategies during
Complex Problem Solving

RONALD H. STEVENS, PhD, ALINA C. LOPO, MD, PHD, PETER WANG, MS

Abstract

Objective: To determine whether expert problem-solving strategies can be identified within a large number of student performances of complex medical diagnostic simulations.

Methods: Self-organizing artificial neural networks were trained to categorize the performances of infectious disease subspecialists on six computer-based clinical diagnostic simulations that used the sequence of diagnostic tests requested as the input data. Six hundred seventy-six student solutions to these problems were presented to these trained neural networks to determine which, if any, of the student solutions represented those of the experts.

Results: For each simulation, the expert performances clustered around one dominant output neurode, indicating that there were common problem-specific features associated with the experts' problem-solving performances. When the performances of students who also made correct problem diagnoses were tested on these expert-trained neural networks, 17% were classified as representing expert strategies, indicating that expert performance was a somewhat rare and inconsistent occurrence among the students.

Conclusions: The ability to identify a small number of expert-like strategies within a large body of student performances may provide an opportunity to study the dynamics of complex learning at both individual and population levels as well as the emergence of medical diagnostic expertise.


Throughout the science-education community is the need for a theory of learning that illuminates the transitions from little to partial to full to expert knowledge within a domain. Although the general cognitive features of expertise have been well explored in many content domains, the manner in which expertise is developed is less clear. This area represents an important gap in our knowledge of learning, one that has implications for how instruction and assessment are to be performed in the future.

Artificial-neural-network technology is based on substantial theoretic foundations. Trained neural networks have had significant practical utility in solving classification problems for which categories are ill defined, for which the patterns are deeply hidden within the data, and for which models of behavior are ill defined. A previous study showed that supervised, backpropagation artificial neural networks trained with medical students' performances on clinical diagnostic problems in the area of immunology or in-
fectious disease correctly identified the problem-solving outcomes of other students more than 85% of the time. The same neural networks, however, poorly identified expert performances. Although our prior results indicated that detectable differences existed between novice and expert performances on our task, they also demonstrated the limitations of applying a supervised neural-network learning architecture for deducing information regarding the differential strategies of novices and experts. For instance, although low output weights for an expert's performance following presentation to a novice-trained network would indicate that the expert solution was not representative of, or generalizable from, that of novices, it was not clear where the expert performance actually resided on the neural-network problem space. The same limitation would occur in performing the converse comparison of student performances on a network trained with expert performances.

We have extended our previous studies by creating expert-trained neural networks with a self-organizing neural network architecture to avoid the above limitations, and to show that artificial neural networks may be useful for identifying strategic patterns that discriminate within the novice–expert performance continuum. It now seems possible to engage novices in complex problem solving and to have their performances evaluated as more "novice-like" or "expert-like" by suitably trained novice and expert neural networks.

**Methods**

**Construction of the Medical Diagnostic Simulations**

Our approach to investigating medical diagnostic performances is based on the cognitive principles of there being a starting condition (i.e., case history), a goal condition (i.e., diagnosis), and the access to the information needed to transit these conditions. Although this approach represents a basic form of problem solving, it is also powerful in that it applies to a broad range of disciplines across multiple levels of education.

Twelve computerized problems relating to medical diagnosis in infectious disease were constructed; six were used for student practice and six were used for student and expert testing. In these simulations, patient histories were presented, and the participants then accessed, in a noncued manner, additional information regarding the patient, laboratory tests, and so forth, needed to make a medical diagnosis. Analysis of the test items chosen has indicated that experts and students order between 4 and 25 test items while making a correct diagnosis, and that few participants

---

**Figure 1** Sample classifying characteristics for the infectious disease problem set. The results from the Structured Query Language (SQL) queries to the student performance database are exported to a spreadsheet in the form of sequential two-test identifiers. The total of all these classifying characteristics for the problem set constitutes the template for the input nodes of the unsupervised neural network.
make the correct diagnosis with the identical sequences of test selections.

Student and Expert Populations

The training set for the self-organizing neural networks consisted of 102 successful expert performances acquired from members of the teaching faculty of the University of California at Los Angeles and the University of Texas, Health Science Center at San Antonio. Successful student performances on these problems were collected from two different classes of second-year medical students (n = 310), who were required to solve these problems for a portion of their examination grades. The logistic and practical details of administering these problems as an examination, and the collection of performance data, have been previously reported. The problems were solved by 87% of the students and 81% of the experts; only data from solved problems were used for these studies.

Collection and Preprocessing of Performance Data

For the artificial-neural-network training, the input data of the training set were the test items placed in the sequence in which they were selected by participants while solving a problem, because the ordering of procedures or steps, given a particular set of conditions, provided useful information regarding knowledge of these conditions and procedures. Capturing and using these data had additional advantages, including those of providing a uniform test mode for both novices and experts and providing performance-based outcomes rather than recall-based outcomes.

Performance data were retrieved from the transaction database through standard Structured Query Language (SQL) queries based on problem, student identity, and additional criteria, such as whether a problem was solved and how many dollars had been spent (prorated on actual laboratory costs to provide a score) at the end of the problem. In general, a query to obtain data for training an artificial neural network consisted of all correct problem solutions, regardless of score, whereas during the testing phase, additional selection criteria may have been employed.

The data were exported from our analysis software to the input file for the artificial neural network in the form of two-test selections, or classifying characteristics, such as "Start TO Blood Culture" and "Blood Culture TO Echocardiogram." A problem-set template was first constructed from the total number of these two-test selections, which were returned from a query to the training set of performances (Fig. 1). There are usually between 500 and 1,000 of such classifying characteristics, which become the input nodes to the Kohonen self-organizing feature network. The Kohonen self-organizing network is a neural network architecture that associates to each input pattern a representative output pattern. This method of model building can be seen as performing vector quantization in that it seeks models that minimize the quantization error. Models are adjusted incrementally as new data are presented. During this adjustment, some ordering takes place: adjacent models in pattern space are near each other in model space. Such a neural network architecture appeared useful for our studies in that the distance of the student performances from those of the experts would provide an estimate of the similarity of the strategies. Such information cannot be as easily obtained by supervised backpropagation networks. The Kohonen-feature-map neural networks were constructed with locally developed software using libraries from Ward Systems Group (Rockville, MD).

Performance patterns in individual training or testing were generated by comparing the test selections made during the problem solving with the classifying-characteristics template and by entering a "1" at the location of a match. The network therefore would show a large number of 0s (classification characteristics on the template not in the test pattern) interspaced with
1s (4 to 25 of them, depending on the number of test items selected while solving a problem).

This pattern was sent to each neurode of the Kohonen feature layer through a set of adaptive weighted connections, as shown in Figure 2. The neural network for these studies had a 36-neurode output layer arranged in a $6 \times 6$ configuration. These neurodes in the feature layer competed with one another, through interneurode connections, for the privilege of producing the single output of 1 from this layer in response to the presented pattern; the remaining neurodes responded with a 0.

During training, patterns of problem-solving performances with similar features stimulated winning neurodes that were the same as or physically close to one another. Training also involved an attempt to maximize the distance between patterns that differed from one another. Both of these features resulted from the interneurode connections, in which each neurode typically had strong positively weighted connections to its immediate physical neighbors in the layer and somewhat negatively weighted connections to neurodes farther away (Fig. 3). This process generated a topologic map of the problem space, in which performances similar to one another became clusters on the 36-node grid.

Training Parameters

The expert neural network used for these studies had a 36-neurode output layer arranged in a $6 \times 6$ configuration, with an initial neighborhood of 6 and a learning rate of 0.2. These values were lowered to a neighborhood of 1 and a learning rate approaching 0 by the end of training. As shown in Figure 1, 423 input nodes consisted of sequences of test items arranged in the form of two-test classifying characteristics. The network was trained for 1,000 epochs, and the weights were updated after each epoch. On an x486 computer, the network training took 250 hours. The distribution of expert and student performances on the 36-neurode problem space was generated by presenting the total number of performance patterns to the trained neural network, selecting the winning neurode, and then summing the winning nodes across the entire performance set.

![STRENGTH OF INTERLAYER CONNECTIONS](image)

**Figure 3** Lateral inhibition connections within the Kohonen layer.

![DISTANCE FROM NEURODE](image)

**Figure 4** Topology of experts' problem space generated by self-organizing artificial neural networks. The distribution of expert and student performances on the trained problem space was generated by presenting the training patterns to the network and selecting the winning neurode. The students' and experts' performances are shown for the following problems. (A) Endocarditis, viridans-group Streptococcus: expert, $n = 18$; student, $n = 69$. (B) Meningoencephalitis, *Listeria monocytogenes*: expert, $n = 15$; student, $n = 100$. (C) Extrapulmonary *Mycobacterium tuberculosis*: expert, $n = 12$; student, $n = 34$. (D) Osteomyelitis, *Salmonella* sp: expert, $n = 18$; student, $n = 44$. (E) Disseminated *Mycobacterium avium-intracellulare*: expert, $n = 15$; student, $n = 57$. (F) Myocarditis, rheumatic fever: expert, $n = 24$; student, $n = 54$. 
Results

Distribution of Expert and Student Performances on the Expert-trained Neural Network

The distribution of the expert performances after training for the six different problems within the 36-neurode topologic map of the problem space is shown in Figure 4. For each problem, the expert performance clustered around one dominant neurode, indicating that there were common problem-specific features associated with the problem-solving performances of the experts. In addition, these dominant problem-specific neurodes were well separated from one another, indicating that the network was able to distinguish problem-specific features. For all problems, clusters of student performances were found at or near the problem-specific expert neurode, suggesting that the expert-trained neural network recognized performance features of some student solutions. The remaining student performances were clustered at neurodes often quite distant from the problem-specific expert neurode, and these performances represented many of the student performances overall, depending on the problem. Given the architecture of self-organizing artificial neural networks, the second group of student performances would not be representative of those of the experts.

Expertise Is Not a Common Occurrence among Students

Of 676 student solutions across the six infectious disease simulations, 117 were grouped at the expert-specific winning neurode. As shown in Figure 5, however, few students solved more than one simulation classified as expert by the expert-trained neural network. The finding that such a limited number of student performances were classified with those of experts indicates that the neural network was not performing simply as a problem classifier, but also as a classifier that was recognizing features of only a limited subset of student solutions.

What Performance Features Are the Neural Networks Using in Making the Novice–Expert Distinctions?

We next used search-path map analysis to try to understand how the expert-trained artificial neural network was generating these classifications, and to explore the significance of these classifications. The problem selected for this analysis was one in which the diagnosis was “Endocarditis, viridans-group Streptococcus.” This problem appeared the easiest of the problem set for both novices and experts, based on the high percentage (>90%) of correct diagnoses and the limited average number of test selections per solution (4–10 tests).

The group search-path maps for the 14 experts and for the 37 students who solved this problem are shown in Figure 6 and indicate the complexity of the strategies on even an easy problem. During group search-path mapping, each potential test selection available during problem solving is displayed on the computer screen as a rectangle, and the lines connecting these rectangles map from the upper left-hand corner of the “from” test to the lower center of the “to” test; the thickness of the lines is proportional to the number of persons making that sequential choice. For instance, many students went from “Echocardiogram TO Diagnosis,” whereas only a few persons went from “Catalase TO Diagnosis.”

Certain experts’ test selections were represented in the student pattern of performances (“Blood Bacterial Culture . . . TO Echocardiogram”), whereas others were missing (“Catalase TO Optochin Sensitivity”). From the student group search-path map alone, it would be difficult to identify which, if any, students were solving the simulations like experts.

This student population was then divided into two groups based on the expert–novice classification of the expert-trained artificial neural network (Fig. 7). These search-path maps indicate that the neural network classification of these student performances decomposed the search-path map shown in Figure 6B into two groups whose test selections differed significantly from one another. For instance, many of the students classified as nonexperts went from “Start TO Consult,” or made a diagnosis after choosing “Optochin Sensitivity”—selections that were poorly represented in the expert performances and absent in the student performances classified as expert. Similarly,
several common features seem to exist between the expert search-path map and that of the students classified as experts, such as "Start TO Blood Bacterial Culture..." and "Echocardiogram TO Diagnosis." However, the finding that none of the student performances classified as expert contained the sequence "Blood Bacterial Culture... TO Diagnosis" and that some of the novice-classified performances did contain this sequence suggests that the classifications being performed are not as simple as similar starting and ending test selections. A more detailed analysis of these differences across all six problems, with a focus on defining different novice-expert strategies, will be reported elsewhere.

**Discussion**

Artificial neural network technologies represent a promising, but unexplored, adjunct to medical education. In prior studies, we demonstrated that supervised neural networks trained with the performances of other student performances could classify subsequent student performances as successful or unsuccessful more than 85% of the time.9,10 These studies were useful in that they provided a peer-based evaluation of each student and also an opportunity to view the dynamics of the student's problem-solving strategies and exploration of alternative hypotheses. These networks, however, were inadequate for probing the differences between the strategies of novices and experts, although they did indicate that such differences existed.

The data in this paper suggest that self-organizing artificial neural networks, trained with expert performances on infectious disease diagnostic simulations, can detect problem-specific and expertise-specific problem-solving features and may therefore be useful for classifying the problem-solving performances of subsequent students into expert and nonexpert categories. The classifications being conducted are based on features of expert performances that the neural network has "discovered" during training and are independent of an arbitrary or formally constructed scoring algorithm. Because artificial neural networks model the probability-distribution functions of the training data, other mathematical procedures could

---

**Figure 6** Search-path mapping analyses of expert and student problem-solving performances. Search-path maps were generated from correct performances of 14 experts (top) and 37 students (bottom) who correctly solved the endocarditis infectious disease problem. The numbers of expert and student performances were chosen to provide similar numbers of test selections for the groups (42–44 test selections) for the subsequent analysis shown in Figure 7. The lines are drawn from the upper left-hand corner of the "from" test to the lower center of the "to" test, thereby providing an indication of the test-item sequencing. For each figure, only the test items selected are shown. The tests in the upper 25% of the window, and which are shaded light gray, represent informational resources rather than laboratory data.
Figure 7 Search-path analysis of student performances classified as novice (top) or expert (bottom) by the expert-trained unsupervised neural network. The student performances in the bottom part of Figure 6 were first classified by the expert-trained neural network into the expert and novice categories. Search-path maps were then generated for each group. Tests that are shaded light gray represent informational resources rather than laboratory data.

also have been used for these studies. The neural network architecture was chosen because direct-comparison studies have shown that artificial neural networks perform as well as or better than other procedures, such as multiple linear regression. Trained artificial neural networks are also easy to distribute.

A central issue regarding any results obtained with artificial neural networks is the adequacy of the training set, which can be enhanced with diversity. Diversity in the training set was fostered by using experts from two different institutions and across the expert continuum from residents to physicians to academic scientists. Obtaining sufficient performances to train a neural network can also be a limiting factor. However, given the distribution of expert performances on the topologic problem space generated after training, and the low proportion of student performances being classified as "expert-like," the training set appears adequate.

Given the difficulty of understanding the features from which artificial neural networks generalize during training, it is uncertain whether all students who are performing like experts are being detected. Multiple neural networks with different configurations, and with reduced and expanded training data sets, have yielded similar results, indicating that the neural network used for this study is not unique. The search-path analysis of the different performance classifications may also contribute information in this regard. First, such an analysis, when viewed by a domain expert, can provide a validation of the classifications being generated by the expert-trained artificial neural network. This feedback can be useful for enhancing and expanding the problem set and optimizing the architecture and training regimen of the artificial neural network. Second, working more from the perception of expert-trained networks that identify expert-like performances, search-path analyses may be useful for exploring domain-specific features of developing expertise.

What types of information regarding the development of expertise can be obtained with this approach? The important features of expert performance are believed to be not only the quantity of information in a do-
main, but also the structure and organization of this knowledge in memory. The transition from a novice's skills to competence to expertise in an area is accompanied by the acquisition of new information and concepts and by the subsequent reorganization of the appropriate preexisting mental structures relating to the domain. The classification of student performances by expert-trained artificial neural networks, followed by the decomposition of these classifications by search path analysis, is beginning to reveal some features of the early mental structures of novices. First, students seem to acquire and use information items that are seldom ordered by experts. For the problem shown in Figure 6, an example would be the use of the optochin and bile solubility tests. These procedures are emphasized in textbooks and laboratory sections for differentiating species of the genus Streptococcus but seem to be of little practical utility for experts. Second, the performances of novices reflect more reliance on rules than do the performances of the experts. As an example, on many of the student performances classified as novice across the six problems of the set, the first test selected is a blood culture, regardless of whether the pathogen is likely to be blood-borne. This selection reflects use of the rule of thumb “Always do a blood culture,” which is often emphasized to medical students. Finally, on some problems, the students classified as novices order the same tests as experts, or as students classified as experts, but they do so in the reverse order.

Each of the examples above can be practically addressed at the instruction level, and the possible effect of these interventions can begin to be explored in a rational way. Educational interventions can be designed and evaluated on their ability to increase the proportion of students who solve problems with expertise on real-world tasks. As such, artificial neural networks may provide a practical tool for developing alternative assessments in any situation in which the sequence of actions can be recorded and in which this sequence accurately reflects important features of the problem task. Once collected, the data on large numbers of students can be rapidly analyzed, and once suitable neural networks are trained, they can be distributed for local onsite assessments. This approach may be a broadly useful tool for investigating the emergence of expertise and for identifying learning deficiencies or data misconceptions that can be self-corrected by comparison with expert performances. An important feature for maximizing the utility of this approach, however, is the generation of the problem space or task. In addition to creating problem-solving spaces that provide opportunities for considerable strategic individuality, the focus should also be on the construction of problem spaces of varying levels of complexity and challenge, which could encompass a broad part of the continuum of novice and expert performance.

References