Using a Neural Network Expert System in Grading Dental Cavity Preparations

Tar C. Aw, D.D.S., M.S.,[1] W. Paul Lang, D.D.S., M.P.H.,[2] Peter Yaman, D.D.S., M.S.,[2] and Joseph B. Dennison, D.D.S., M.S.[2]

[1]School of Dentistry, University of Washington, Seattle, Washington.
[2]School of Dentistry, University of Michigan, Ann Arbor, Michigan.

INTRODUCTION

Grading of amalgam cavity preparations is based on recognizing patterns of scored evaluation criteria. The traditional approach uses rule-based summation of quantified evaluation criteria by humans. The purpose of this study was to evaluate the success of a neural network[1] expert system in assigning grades to amalgam cavity preparations, compared to a human expert. Formal, standardized criteria were used in the decision-making process, and a unique inference mechanism or reasoning method hitherto unused in dentistry was employed to answer this clinical problem.

METHODS

Amalgam Cavity Grader(ACG) was implemented using NeuroShell[2](Ward Systems Group Inc., Frederick, MD) on an IBM-compatible 66-MHz 486 DX2 computer with an Intel coprocessor. There were 4 Input Nodes, 12 Hidden Nodes and 5 Output Nodes. This configuration seemed to work the best, in terms of accuracy, and time to train.

A human expert was presented with cases of amalgam cavity preparations, each preparation with four evaluation criteria - Smoothness, Retention, Internal Outline, and External Outline. Each criterion is qualitatively identified by scores of Excellent, Good, Acceptable, Poor, and Unacceptable. Based on the combinations of scores of evaluation criteria, the human expert would assign a letter grade of A, B, C, D, or E. Training and testing sets were obtained by this method.

In the first experiment, a set of graded cases(N=214) was used to train the network. In the second experiment, another set of graded cases(N=85) was used to train the network. A testing set (N=98) of cases(separate from the training sets) was graded independently by the human expert and the two trained neural networks. The performances of the human expert and the neural networks were then analyzed and compared by statistical methods such as the ranked correlation, the paired t-test and the kappa statistic.

In a third experiment, the training sets were used as the testing sets for each trained network, and the accuracy compared with using a separate testing set, as above.

RESULTS

Experiment 1. Of the 98 cases, 80 were correctly assigned by the neural network, i.e. 81.63% correct. Using ranked correlation, the correlation coefficient, $r = 0.94$, indicating that there is a significant correlation between the two sets of grades, almost a linear relationship. Using paired t-test, the t-statistic = 0.47, indicating that the two graders do not differ, that there is no difference between the human expert and the neural network in grading cases of evaluation criteria scores. The kappa statistic, $k = 0.83$, indicating that there is a significant agreement between the human expert and the neural network, more than by chance alone.

Experiment 2. Of the 98 cases, 71 were correctly assigned by the neural network, i.e. 72.45% correct. The correlation coefficient, $r = 0.91$, the t-statistic = 0.61, and the kappa statistic, $k = 0.65$.

Experiment 3. Of the 214 cases, 210 were correctly assigned by that neural network, i.e. 98.13% correct. Of the 85 cases, 83 were correctly assigned by that neural network, i.e. 97.65% correct.

DISCUSSION

The neural network output a high percentage of results correctly on cases that it had not encountered previously. The neural network utilized the fact that the cavity evaluation criteria were associated in a certain manner(akin to that outlined by the human expert) to produce resultant grades. The human expert's technique of grading was more subtle and complex than perceived, as is most human cognitive processes. There seemed to be an interplay in the various combinations of individual criteria scores, more synergistic than a simple additive process.

This study demonstrated that the trained ACG could grade amalgam cavity preparations based on evaluation criteria, and performed at a level close to that of our human expert. By every measure, using a larger and more diverse training set produces improved neural network performance. Also, testing with cases that the network was trained with produces highly accurate results, as would be expected.

Our results are consistent with other previous studies comparing neural network performance to human experts. Computerized decision systems cannot replace practitioners but instead are designed to support the highest level of clinical competence by supplying access to current knowledge, and rapid computation of complicated decision analyses. As computers lack intuition, it can serve best as a second opinion, to bolster human decision-makers with its strengths, and the combination thereof will be superior than either alone. The promising results warrant further investigation.

REFERENCES

[1]. Nelson MM, Illingworth WT. A practical guide to neural nets. Reading, MA: Addison-Wesley Press, Feb. 1992.

[2]. NeuroShell User Guide Manual. Ward Systems Group Inc. Frederick, MD, 1995.

0195—4210/96/$5.00 © 1996 AMIA, Inc.