

## DISCUSSION

Both the ANN with the RBM and CSD data and the ANN with the additional inputs for TCM data were able to accurately distinguish activations from examples that were not activations. The ANNs misclassified very few of the unseen test examples. For the ANN with RBM and CSD only, the 200 unseen test examples were all classified correctly when staged training was used. The classification results for non-staged training were lower; 197 test examples were classified correctly with non-staged training. The ANN that incorporated information from RBM, CSD, and TCM showed a slightly higher error rate. This ANN correctly classified 196 of 200 test examples when trained with staged training and 195 of 200 test examples when non-staged training was used. Both ANNs identified examples that were not activations more accurately than examples of activations.

Some errors of classification by both ANNs may have been due to the test examples including only those examples with low magnitude derivatives. The examples with low magnitude derivatives had a lower signal-to-noise ratio than the examples with higher magnitude derivatives. Therefore, it would be expected that the low magnitude test examples would be more difficult for the ANN to distinguish and to classify correctly than a set of test examples that included higher magnitude derivatives. The low magnitude activations may also have more closely resembled waveforms of distant activity. This may

explain the ANNs misclassification of several test examples of activations as not activations. The ANN with RBM, CSD, and TCM data may have generated more errors in classification because of the large number of inputs to this ANN. Since the Sensitivity and +P of the ANNs were dependent upon classification results, the use of test examples with low magnitude derivatives and the number of inputs could also have affected these parameters.

The Sensitivity of the ANNs was high for both examples that were activations and examples that did not represent activations. When staged training was utilized, the Sensitivity of the ANNs for detecting activations was improved. For examples that were not activations, the type of training used did not affect the Sensitivity of the ANNs.

The +P of both the ANN with RBM and CSD data and the ANN with RBM, CSD, and TCM data was consistently greater than 95 %. For the activation class, the +P values of 100 % indicated that the ANNs did not incorrectly classify examples that were not activations as activations. The +P values for the class of examples that were not activations were lower due to the ANNs misclassifying activations as not activations. The +P of both ANNs was higher when staged training was used.

An interesting aspect of this study was that staged training produced better results than non-staged training. A possible explanation of the superior performance of the ANN when trained with staged training is that the staged training may represent a form of Bayesian learning. According to Bayesian

theory, there is a probability density function that represents the uncertainty in the values of parameters. A prior distribution, which describes the parameters, exists before any data are observed. The prior distribution is generally very broad since there is a high degree of uncertainty about parameter values before the data are examined. As data are observed, a posterior distribution is formed. The posterior distribution is based on knowledge gained from the observed data and tends to be narrower than the prior distribution [20].

In staged training, the training examples are presented to the ANN in specific sets or stages in a particular order. In each successive stage of training, more data are presented to the ANN. The first stage of training consisted of training the ANN with the set of examples that had activations with very high magnitude derivatives. Therefore, the ANN was initially trained to distinguish between the examples that were activations and the examples that were not activations on the data with the highest signal-to-noise ratio. These examples would be expected to have the clearest distinction between the two classes of examples, activations and not activations.

The initial stage of training with the examples of activations that had very high magnitude derivatives resulted in a change of the weights of the ANN. Due to the information provided by the data in the first training stage, the weight values appear to have been modified so that ANN learning of the next set of examples, which had activations with derivatives of lesser

magnitude, is facilitated.

The ANN is trained to classify the set of examples that had activations with high derivative magnitudes during the next successive stage of training. The high derivative magnitude examples should have a lower signal-to-noise ratio than the examples with very high magnitudes. The difference between the examples that are activations and the examples that do not represent activations is less than with the very high magnitude derivative examples. The information that the examples with high magnitude derivatives provided caused further changes in the values of the ANN weights. The weights seem to have been set so that the ANN was better able to learn to recognize examples with even smaller magnitude derivatives.

In the next training stage, the set of examples with activations that had medium magnitude derivative examples and an even lower signal-to-noise ratio are used to train the ANN. The values of the ANN weights are again modified as the ANN learns to differentiate between the examples in this training set that are activations and the examples that are not activations. The weights have been changed in order that the ANN can more easily recognize examples of activations with derivatives of decreased magnitude.

Finally, the ANN is trained with the example set that had activations with the lowest magnitude derivatives and the lowest signal-to-noise ratio. This final training stage causes further modifications in the values of the ANN weights. The ANN has learned to distinguish between the examples of

activations with the lowest magnitude derivatives and the examples that did not represent activations.

As the magnitude of the derivatives in the examples decreases, the signal-to-noise ratio also becomes lower. As the signal-to-noise ratio becomes diminished, the difference between the two classes of examples -- activations and not activations -- decreases. In addition, as the magnitude of the derivative decreases, the activation becomes more difficult to distinguish from distant activity.

Yet, perhaps the posterior distribution has narrowed with each successive training stage. The weights of the ANN have been gradually altered and refined with each stage of training so that the learning capabilities of the ANN in the next successive stage of training appear to have been facilitated. Therefore, the ANN that has been trained using staged training is better able to distinguish the low magnitude test examples that are activations than if non-staged training had been used.

The process of staged training needs to be studied further. Simple synthetic models are being used to examine various aspects of staged training. The preliminary experiments with synthetic models have shown that staged training improves the performance of the ANN. However, questions, such as how does staged training affect the structure of the ANN to produce superior results and in what cases is staged training likely to improve the performance of the ANN, need to be answered.