Artificial Neural Networks for Modelling of the Impact Toughness of Steel

D. DUNNE, H. TSUEI1 and Z. STERJOVSKI

School of Mechanical, Materials and Mechatronic Engineering, University of Wollongong, NSW 2522, Australia.
1) Department of Aviation Management, Republic of China Air Force Academy, Taiwan.

(Received on December 16, 2003; accepted in final form on June 10, 2004)

1. Artificial Neural Networks

Artificial neural networks are computational networks that attempt to simulate those processes occurring in the human brain and nervous system that enable pattern recognition, information filtering and functional control.1) They are part of a larger group of methods used for data mining i.e. the use of databases to extract models to classify or predict classes or trends. Decision trees, Bayesian belief networks, regression analysis, fuzzy logic and neural networks are all examples of data mining methods.2)

Biological neural networks transmit electrical charge through a massive network of discrete neurons (nerve cells) consisting of a cell body (or head) surrounded by dendrites and a long axon (or tail), see Fig. 1(a). Electrical charge is displaced along the axon and to the dendrites of other neurons via interconnections called synapses. The “weight” of the electrical message is controlled by the electrochemical properties of the synapse regions as well as the biochemistry of the dendrites.1) A symbolic representation of a neuron is shown in Fig. 1(b).

Artificial neural networks, based on the characteristics of biological neural networks, were first proposed in 1943 by McCulloch and Pitts,3,4 but detailed system designs did not appear until the late 1950s.3,4 One of the earliest and most important network structures was the Perceptron proposed by Rosenblatt5 which forms the basis of most ANNs since developed, including the Back Propagation network developed by Rumelhart and McClelland6 that is being used increasingly for the modelling of materials properties.6)

A schematic diagram of a Perceptron (an artificial neuron) is shown in Fig. 2(b). The output y is not simply the summation of the inputs but is modified by an activation function, \( f_N(z) \), that controls the amplitude of the output. If the amplitude exceeds a set threshold the neuron “fires”, sending its signal to other neurons.1) The continuously differentiable sigmoidal and hyperbolic tangent functions are commonly used as the activation function.

The application of artificial neural networks (ANNs) to the prediction of the Charpy impact toughness of quenched and tempered (QT) steels and ferrous weld metals is examined in detail. It is demonstrated that the Charpy impact toughness can be accurately predicted using the selected input variables and their ranges of values.

The capacity of ANNs to handle problems involving large sets of input variables is illustrated by a model developed to predict the impact energy of weld metal (WM) produced by flux cored arc welding (FCAW). The usefulness of ANNs for alloy design and process control is demonstrated through another model developed to predict the toughness of a QT structural steel as a function of composition and postweld heat treatment.

Although comparison of the two models indicates that the trends in toughness with changes in Mn and B concentrations are in opposite directions for weld metal and QT steel, it is shown that these trends can be reconciled with reported experimental results and theoretical interpretations.

KEY WORDS: artificial neural networks; impact toughness; modelling; flux cored arc weld; quenched and tempered steel; postweld heat treatment.

1. Artificial Neural Networks

Artificial neural networks are computational networks that attempt to simulate those processes occurring in the human brain and nervous system that enable pattern recognition, information filtering and functional control.1) They are part of a larger group of methods used for data mining i.e. the use of databases to extract models to classify or predict classes or trends. Decision trees, Bayesian belief networks, regression analysis, fuzzy logic and neural networks are all examples of data mining methods.2) Biological neural networks transmit electrical charge through a massive network of discrete neurons (nerve cells) consisting of a cell body (or head) surrounded by dendrites and a long axon (or tail), see Fig. 1(a). Electrical charge is displaced along the axon and to the dendrites of other neurons via interconnections called synapses. The “weight” of the electrical message is controlled by the electrochemical properties of the synapse regions as well as the biochemistry of the dendrites.1) A symbolic representation of a neuron is shown in Fig. 1(b).

Artificial neural networks, based on the characteristics of biological neural networks, were first proposed in 1943 by McCulloch and Pitts,3 but detailed system designs did not appear until the late 1950s.3,4 One of the earliest and most important network structures was the Perceptron proposed by Rosenblatt5 which forms the basis of most ANNs since developed, including the Back Propagation network developed by Rumelhart and McClelland6 that is being used increasingly for the modelling of materials properties.6)
An artificial neural network can be defined as a massively parallel interconnected network of simple, organised and adaptive processing elements (PEs) that is analogous to a biological nervous system. The processing elements (equivalent to neurons) are non-linear functions that transform multiple weighted input signals to a single numerical output signal. The interconnections (synapses) between the PEs are non-linear functions that transform the signals that pass through them and the network signal. The interconnections (synapses) between the PEs are non-linear functions that transform the signals that pass through them and the network signal.

A Back Propagation neural network learns iteratively by modifying its synaptic weights. Back Propagation networks are suited to solving problems of “hard learning”, as categorised by Minsky and Papert. Hard learning problems occur in nets that are required to map a well-defined set of inputs into a well-defined set of outputs.

The inputs are fed into a series of processing elements called the input layer and the weighted outputs of this layer are passed to the first hidden layer, which is regarded as an output layer. If the outputs of the hidden layer are fed to the final output layer the network is described as a two-layer neural network, Fig. 2(a). Multiple-layer neural networks incorporate more than one hidden layer, linked in series, before the final output layer. The number of hidden layers and the number of processing elements in each layer determine the “topology” of the neural network and affect the accuracy of the trained network. The optimum number of processing elements in each hidden layer is determined by trial and error and is related to the complexity of the problem.

A Back Propagation neural network learns iteratively by processing sets of data and comparing the output of the network with the known (measured) result. This process of network training involves modification of the weights to minimise the root mean square (RMS) error between the predicted and measured outputs. The changes in weights are made “backwards”, from the output layer through each hidden layer. This process normally results in convergence of the weights and completion of the learning process. The completed network is then tested on data that it has not seen before to assess the accuracy of the model.

Overtraining can occur if the learning capacity of the network is too large relative to the number of data sets used in its training. The network can start to learn specific features of the data at the expense of its capacity to generalize. The performance of the test data can be used to optimize the number of units in the hidden layer and effectively match the network complexity to the size and quality of the training and test data sets.

Instability of the weights is possible under certain conditions and strategies such as use of a momentum term and smoothing methods have been developed to reduce the risk of instability.

Just as the weights bias the effect of the input variables on the output, bias can also be applied to the PEs in the form of trainable weights. Interpreting a Back Propagation neural network model is not straightforward and extracting knowledge and meaning from the model is often difficult. In this respect ANNs are in stark contrast to physical models based on the fundamental mechanisms of the processes involved. ANNs are arbitrarily constructed, without any mechanistic or theoretical basis, by the set of linked differential equations and their coefficients that best predict the output. For scientific problems they serve a similar, but less transparent, function to multiple linear regression models for predicting the value of a particular property based on the key dependent variables. However, ANNs can handle much more complex, non-linear, multi-variable problems, particularly those where interactive effects are significant. Although ANNs have been described as “black box” models, it is probably more appropriate to refer to them as “grey boxes”, because they can be interrogated to establish causal relationships.
by Rumelhart and McClelland\(^5\)) and is a widely used standard. Delta Rule learning reduces the error between the output of a processing element and the desired output by modifying the incoming weights.

Random weighting of each processing element in the model is an arbitrary starting point and the weightings are progressively altered on exposure to numerous repetitions of training examples. Hence, the neural network is in the process of learning and the constant adjustment of weights between processing elements leads to a reduction in the RMS error of the overall training data.

The transfer function is the method of transforming the input sum (Fig. 2(b)) into a single output and is used for all layers in the network. The transfer function is usually a non-linear function that transforms the internally generated sum for each processing element (PE) into a potential output value. Available transfer functions for Back Propagation networks are sigmoid, hyperbolic tangent, linear and sine. The tanh transfer function is a smooth version of a \([-1, 1]\) step and was used for the model developed for weld metal toughness because it is recommended for hard learning problems.\(^5\)

The sigmoid function is a smooth version of a \([0, 1]\) step function and was selected for the model built for QT steel toughness because of its compatibility with the 0,1 values assigned to cooling conditions and sample test orientation (see Table 3). This transfer function normalises the input and output data ranges to values between 0 and 1.

The RMS error is a general index for the performance of the model created by the neural network. It is measured through all the PEs in the output layer by summing the squares of the errors between calculated outputs and measured outputs, dividing by the number of examples in the training (or testing) set, and then taking the square root.

The number of PEs in the hidden layer was adjusted by examining performance on the training/testing data. The RMS errors for different numbers of PEs in the hidden layer of the weld metal toughness model are given in Fig. 3. The result shows that nine PEs has a very good consistency for both the training data and testing data. Therefore, this number was selected for the hidden layer. Larger numbers of PEs in the hidden layer did not give significantly lower values of RMS error for the test data. Moreover, the complexity of the model increases with a large number of PEs.

For the QT steel toughness model, five processing ele-
ments in the hidden layer were considered to be the optimum to represent the variations in toughness as a function of the ten input variables. The architecture of this model is shown schematically in Fig. 4. The bias is effectively a processing element that sets the weightings of all other processing elements.

2.3. Model Sensitivity Analysis

Sensitivity analysis was applied to each of the trained models. This procedure determines the sensitivity of the output (toughness) to fluctuations in the value of any given input variable. A sensitivity analysis was also carried out to further validate the model and establish predicted trends in toughness with changes in the values of selected input variables. Sensitivity measures the contribution of each input variable on the desired output. The following equation was used for sensitivity, $S (%)$:

$$S(\%) = \frac{(O_{(n+1)} - O_{n}) \times I_{\text{range}}}{(I_{(n+1)} - I_{n}) \times O_{\text{range}}} \times 100 \ldots \ldots \ldots \ldots (1)$$

where $I_{\text{range}}$ is the range of the particular input variable, $O_{\text{range}}$ is the range of the entire output of the sensitivity data, $O_{n}$ is a data point (output) and $O_{(n+1)}$ is the ensuing data point, $I_{n}$ is a data point (input) and $I_{(n+1)}$ is the ensuing data point.

The average sensitivity ($S$) of each input variable can be calculated through all of the input data. The mean or standard values of the other variables can be selected for this calculation.

$$S = \frac{S_1 + S_2 + \cdots + S_n}{n} \ldots \ldots \ldots \ldots (2)$$

where $S$ is the average sensitivity for an individual input datum for a certain input variable and $S_1$ to $S_n$ are the sensitivities calculated for the variable for each of the 1 to $n$ training data sets.

3. Development and Testing of the Models

3.1. Input Variables—WM Toughness Model

Using experience of welding metallurgy, there are four major fields that are expected to influence the Charpy toughness: chemical composition, microstructure, non-metallic inclusions and welding conditions. These fields are also interactive. The model developed considered only the first three fields in order to reduce the complexity of the model and to concentrate on the nexus between structure and properties. The structural characteristics of the weld metal are, of course, derived from the welding process conditions.

Three different types of FCAW consumables were examined: rutile, rutile-boron and metal cored. Five different welding conditions were used for each consumable type, resulting in 15 deposited welds that were subjected to chemical analysis, mechanical testing and structural characterization. The structural variables that were selected for analysis are given in Table 1. There were 31 variables and 83 sets of data. The variables for each training sample included the weight percentages of C, Mn, Si, S, P, Ti, Al, B, Mo and O; the volume fractions of reaustenitised region, double-reheated region, acicular ferrite (AF), grain boundary ferrite (PF(G)), intragranular polygonal ferrite (PF(I)), ferrite with a second phase (FS), grain coarsened (GC) sub-zone, grain refined (GR) sub-zone, intercritical (IC) sub-zone, and non-metallic inclusions ($V_o$); the overall ferrite mean free path, $(L^oF)$; the ferrite mean free path of solidified weld metal $(L^wF)$, grain coarsened sub-zone $(L^GC)$, grain refined sub-zone $(L^GR)$, and intercritical sub-zone $(L^{IC})$; the mean three dimensional diameter of inclusions $(D_i)$, the number of inclusions per unit volume $(N_i)$, and the volume fraction of inclusions $\geq 1.5 \mu m (F \geq 1.5)$; the overall mean hardness $(H)$; and the impact test temperature $(T)$.

The subzones listed above are those normally used to characterize the heat affected zone (HAZ) of the parent plate or the weld metal. The GC sub-zone is that part of the HAZ adjacent to the fusion boundary; the GR subzone is the region of the HAZ which is fully austenitised but reaches a peak temperature less than about 1 100°C; and the IC subzone is bounded by peak temperatures between $A_{c1}$ and $A_{c3}$, for which partial austenitisation occurs.

The training data set contained 66 examples and the testing data set consisted of 17 examples. There was no overlap for the two data sets and both data sets had 31 input data fields and 1 desired output data field.

3.2. Input Variables—QT Steel Toughness Model

The ANN model was designed to predict the impact toughness of three QT 690 MPa minimum yield strength steels (see Table 2) for various test temperatures and notch orientations, after exposure to multiple PWHT cycles at 570°C followed by slow (250°C/h) or fast (still air) cooling. The main issue was the effect of repeated PWHT processes on the impact toughness of the steels (BIS 80 and BIS 80P), which are used in Australia for pressure vessel construction. There were 10 input variables that can be grouped into: heat treatment conditions (3 variables); a composition set (5 variables); and testing conditions (2 variables). The ranges of these variables are given in Table 3.
The data consisted of 83 separate sets involving the variables and the ranges shown in Table 3. About 85% of the data was used for training and 15% for testing the model. Sensitivity data were also used to further validate the model.

### 4. Discussion of Model Predictions

#### 4.1. Testing the Models

The predicted output of the models was compared with the measured output for both the training and testing data sets. Strong agreement between the predicted and measured toughness values for each model is demonstrated in Figs. 5
and 6. The close agreement indicates that there would be a high degree of confidence in the prediction of the impact toughness of a sample characterised by control variables lying within the ranges over which the model is based.

4.2. WM Toughness Model

Within the range of the experimental data used for training, predicted trends in impact energy with changes in the values of selected variables were calculated using the trained model. This procedure was used to establish the sensitivity of toughness to changes in the input value of a specific variable. The mean values of the other input variables (Table 1) were used in performing these calculations. In this way, trends in toughness with changes in the values of selected input variables—relationships that could require considerable cost and effort to establish experimentally—can be determined easily using the neural network model. An example is given in Fig. 7 for the effect of volume fraction of acicular ferrite. The curves are parametrised in terms of Charpy test temperature. The strongly beneficial effect of acicular ferrite, particularly at higher test temperatures, is evident and is consistent with the results of numerous experimental investigations.

The average sensitivity calculated for selected input variables for the WM toughness model is shown in Fig. 8. Since the model indicated that the direct effect of non-
metallic inclusions on impact toughness was small for the range of conditions examined, the associated variables (Table 1) are not included in Fig. 8. Instead, the figure treats variables that show relatively high sensitivities to impact toughness: chemical concentrations, microstructural features and impact test temperature.

For the composition ranges investigated, the elements: Mn, Si, B and O showed beneficial effects on toughness of the weld metal, whereas C, S, P, Ti, Al, Mo and N had detrimental effects. Although increased alloy content generally has an adverse effect on the impact toughness of steels, steel weld metals differ insofar as they rely on reactions between alloying elements and O and S to produce non-metallic inclusions that catalyse intragranular nucleation of ferrite within the coarse as-solidified grains. The structural refinement generated in this way can improve the toughness substantially.

The analysis indicates that increasing manganese over the range 0.94–1.5 wt% leads to increased toughness, consistent with the reported results. Silicon, and to a lesser extent manganese, are strong deoxidising elements that form inclusions in weld metals capable of nucleating acicular ferrite. In this way, increases in both silicon and manganese can contribute to increased impact toughness. In addition, manganese is known to lower the impact transition temperature of steels for C levels within the range of the weld metals examined (0.05–0.09 wt%).

4.3. QT Steel Toughness Model

The consistency of the predicted and measured impact energies for the test data set indicates that the model can be used as a design tool to investigate the effect of changes in control variables on the impact toughness. Moreover, the model has potential for use on-line during steelmaking or subsequent heat treatment to predict the toughness and, if necessary, to guide remedial action to bring the steel back into specification for impact toughness.

The average sensitivity of each of the ten input variables to impact toughness is shown in Fig. 10. Apart from the test variable and a variable selected as a parameter, the default values of the remaining variables are based on the data for 11 mm QT BIS80PV steel tested in the T-L orientation at −20°C with slow cooling after PWHT. These reference conditions correspond to the plate thickness and steel composition commonly used in the fabrication of transportable pressure vessels in Australia.

The model successfully predicts the form of the impact transition curve that is characteristic of ferritic steels (Fig. 11), boosting confidence in the validity and predictive power of the ANN model.

However, since this model was developed on the basis of only three plate steels with limited compositional ranges for the five elements considered, trends indicated by sensitivity analysis should be viewed within this context. Figure 12 shows the predicted sensitivity of toughness to Mn level and incorporates error bars on predicted toughness for the measured Mn levels that were inputs in the model. These bars represent ±5 J which is the RMS error reflecting the agreement between predicted and measured toughness values (Fig. 6). On this basis, the trend of toughness with increasing Mn content predicted by the model is not unrealistic.

Increasing the concentrations of all of the elements ex-
amined resulted in decreases in impact toughness, with the effect diminishing in the order: S, Mn, Cr, C, and B. Sulphur is widely acknowledged to adversely affect the impact toughness of wrought steels because of the formation of sulphide inclusions. Manganese also has a relatively strong negative effect over the composition range investigated (1.10–1.43 wt%), Fig. 12. This trend, and the predicted effects of Cr and C, are consistent with established relationships for QT steels of decreasing impact toughness with increased elemental concentration.\(^1\)\(^9\) Manganese and chromium in rolled steel plate are associated with compositional segregation in the form of banding that intensifies with increasing concentration.\(^2\)\(^0\) This banding can result in splitting on impact fracture, with a decrease in impact energy. Splitting was observed in the QT steels investigated.

Boron is predicted by the model to have a slightly negative effect on impact toughness, contrary to the prediction for weld metal. However, the composition ranges in the two cases do not overlap and the rationales for adding B are slightly different, as discussed below.

4.4. Comparison of Model Predictions

Both models include test temperature as an input variable and the sensitivity graphs (Figs. 8 and 10) both show that it is clearly the most sensitive variable, in accordance with the well known transition in impact fracture mode as a function of test temperature in ferritic steels.

The predicted trends with Mn and B are in opposite directions for the two models, but these apparent anomalies can be rationalised as follows. For the QT steels, B is added to enhance hardenability by inhibiting grain boundary nucleation of diffusional transformation products during quenching of fine grained austenite.\(^2\)\(^1\) This effect is considered to arise from grain boundary segregation of B. For this purpose, small concentrations (typically 5–20 ppm) are used in fully de-oxidised steels.\(^2\)\(^2\) However, subsequent tempering and PWHT are likely to result in formation of grain boundary borides (Fe\(_2\)B or M\(_{23}\)(BC)\(_6\)) that could adversely affect toughness.

Boron is added to weld metals for the similar purpose of inhibiting grain boundary nucleation of ferrite or bainitic ferrite. However, much higher levels of B are required because reaction with N and Ti reduces the effective solute B content. Figure 8 indicates that the ANN model predicts a significant improvement in toughness with increasing B level over the composition range investigated. The physical cause of this trend could be promotion of acicular ferrite. Moreover, the weld metal is not subjected to the prolonged tempering process required for QT steel that could generate grain boundary precipitation and a reduction in toughness.

Comparison of Figs. 9 and 12 shows that, for similar concentration ranges, manganese has opposite effects on toughness in the WM and QT steels. As discussed above, the decreasing trend for QT steels is consistent with reported results and also with the presence of Mn-dependent compositional banding. The increasing trend for WM also conforms to experimental observations and is probably related to an increase in amount of acicular ferrite through modification of the inclusion distribution.

5. Conclusions

(1) Artificial neural network models have been developed that accurately predict the impact toughness of QT structural steels and FCA weld metals.

(2) These models have been shown to be capable of predicting realistic trends with changes in the control variables, thereby providing guidelines for alloy design and/or thermal-mechanical treatment to obtain a specified toughness level. Thus the need to undertake expensive laboratory testing can be greatly reduced.

(3) The model for FCA weld metal identified that the compositional/microstructural variables that have the most favourable effect on impact toughness are volume fraction of acicular ferrite and the weight percentages of Mn, Si, B and O. The most adverse factors are volume fraction of grain coarsened weld metal HAZ and the weight percentages of P and S.

(4) The model developed for QT steel indicated that all of the compositional variables examined, as well as increasing PWHT time, have adverse effects on impact toughness.

(5) Although predicted trends with changing concentrations of Mn and B were in opposite directions for weld metal and QT steel plate, these trends are consistent with reported relationships and explanations.

REFERENCES

1) D. Graupe: Principles of Artificial Neural Networks—Advanced Series on Circuits and Systems, Vol. 3, World Scientific Publishing, Singapore, (1997), 4.
2) W. S. McCulloch and W. Pitts: Bull. Mathematical Biophys., 5 (1943), 115.
3) R. J. Lee: Proc. of 14th ACM National Meeting, ACM Press, New York, (1959).
4) F. Rosenblatt: Psychol. Rev., 65 (1958), 386.
5) D. E. Rumelhart and J. L. McClelland: Psychol. Rev., 89 (1986), 60.
6) H. K. D. H. Bhadeshia: ISIJ Int., 39 (1999), No. 10, 966.
7) T. Kohonen: Associative Memory: A System-Theoretical Approach, Springer Verlag, New York, (1987), 1.
8) R. K. Miller: Neural Networks, Fairmont Press Inc., Georgia, (1989), 11.
9) M. Minsky and S. Papert: Perceptrons, MIT Press, Cambridge, MA, (1969), 10.
10) J. Han and M. Kamber: Data Mining-Concepts and Techniques, Academic Press, CA, (2001), 303.
11) T. J. Sejnowski and C. R. Rosenberg: Complex Systems, 1 (1987), 145.
12) T. Cool, H. K. D. H. Bhadeshia and D. C. J. Mackay: Mathematical
13) L. X. Kong, P. D. Hodgson and D. C. Collinson: _ISIJ Int._, 38 (1998), No. 10, 1121.
14) H. K. D. H. Bhadeshia, D. C. J Mackay and L. E. Svensson: _Mater. Sci. Technol._, 11 (1995), No. 10, 1046.
15) Y. Y. Yang, D. A. Linkens, M. Mahfouf and A. J. Rose: _ISIJ Int._, 43 (2003), No. 7, 1040.
16) Z. Sterjovski, D. P. Dunne and S. Ambrose: _Aust. Weld. J., Weld. Res. Suppl._, 48, (2003), No. 2, 41.
17) H. Tsuei, D. P. Dunne and H. Li: _Sci. Technol. Weld. Joining_, 8 (2003), 205.
18) G. M. Evans: _Oerlikon-Schweissmitt_, 36 (1978), No. 82, s4; also IIW Doc. II-A-432-77.
19) Metals Handbook, Vol. 1, 10th Ed. ASM, Metals Park, Ohio, (1990), 735.
20) L. E. Samuels: Optical Microscopy of Carbon Steels, ASM, Metals Park, Ohio, (1980), 127.
21) F. B. Pickering: Physical Metallurgy and the Design of Steels, Materials Science Series, Applied Science Publishers, Essex, England, (1978), 103.
22) C. S. Siebert, D. V. Doane and D. H. Breen: The Hardenability of Steels, ASM, Metals Park, Ohio, (1977).