Contract No:

This document was prepared in conjunction with work accomplished under Contract No. 89303321CEM000080 with the U.S. Department of Energy (DOE) Office of Environmental Management (EM).

Disclaimer:

Proceedings of the ASME 2022
Pressure Vessels & Piping Conference
PVP2022
July 17-22, 2022, Las Vegas, Nevada, USA

Draft

PVP2022-84908

MACHINE LEARNING MODELS OF BURST STRENGTH FOR DEFECT-FREE PIPELINES

Xian-Kui Zhu, William R. Johnson, Robert Sindelar, Bruce Wiersma
Materials Technology
Savannah River National Laboratory
Aiken, SC 29808, USA

ABSTRACT

Burst strength of line pipes is essential to pipeline design and integrity management. The simple Barlow equation with the ultimate tensile strength (UTS) was often used to estimate burst strength of line pipes. To consider the plastic flow effect of ductile steels, Zhu and Leis (2006, IJPVP) developed an average shear stress yield criterion and obtained the Zhu-Leis solution of burst strength for defect-free pipelines in term of UTS and strain hardening exponent, n, of materials. The Zhu-Leis solution was validated by more than 100 burst tests for various pipeline steels.

The Zhu-Leis solution, when normalized by the Barlow strength, is a function of strain hardening rate, n, only, while the experimentally measured burst strength, when normalized by the Barlow strength, is a strong function of n and a weak function of UTS and pipe diameter to thickness ratio D/t. Due to difficulty of three-parameter regressions, this paper adopts the machine learning technology to develop alternative models of burst strength based on a large database of full-scale burst tests. In comparing to the regression, the machine learning method works well for both single and multiple parameters by introducing an artificial neural network (ANN), activation functions and learning algorithm for the network to learn and make predictions.

Three ANN models were developed for predicting the burst strength of defect-free pipelines. Model 1 has one input variable and one hidden layer with three neurons; Model 2 has three input variables and one hidden layer with five neurons; and Model 3 has three input variables and two hidden layers with three neurons for the first hidden layer and two neurons for the second hidden layer. Those ANN models were then validated by the full-scale test data and evaluated through comparison with the Zhu-Leis solution and the linear regression result. On this basis, the best ANN model is recommended.

KEYWORDS: Burst strength, pipeline, regression approach, machine learning, artificial neural network

1. INTRODUCTION

Pipeline systems are the most important nation's energy infrastructures for transporting large volumes of natural gas, crude oil, and other hydrocarbons over long distances to meet the increasing demands of energy consumptions. A transmission pipeline usually consists of many large-diameter thin-wall pipes made of carbon steels through welding. Pipelines are operated at high internal pressure. The burst strength of line pipes is required for material selection, pipeline design, safe operation, integrity management or engineering critical analysis.

There are extensive theoretical, numerical, and experimental investigations on burst strength of pressure vessels that have led to many predictive models for estimating the burst strength of pressure vessels. A comparative study of earlier predictive models was performed by Christopher et al. [1] for thick-wall cylindrical vessels and by Law et al. [2, 3] for thin-wall pipes. These authors found that none of the models was generally applicable to all steels because most earlier models were developed from either a simple analysis or an empirical data fit to limited experimental data for specific steels of interest. Some models focused on a conservative lower-bound prediction of burst strength, while other models reflected an upper bound prediction of burst strength.

Many models, such as Barlow strength, were obtained from the hoop stress of the pipes and the strength theory that requires only one material strength property, such as the yield stress (YS) or ultimate tensile stress (UTS), and thus did not consider the effect of plastic flow or strain hardening rate. Steward and Klever [4] first reported a set of experimental burst test data that showed a strong effect of plastic flow response on the burst pressure of end-capped vessels for various structural steels. Their burst test data have a large scatter within two bounds. The upper bound was defined by the von Mises solution of burst pressure [5, 6], while the lower bound was defined by the Tresca solution of

This work was produced by Battelle Savannah River Alliance, LLC under Contract No. 89303321CEM000080 with the U.S. Department of Energy. Publisher acknowledges the U.S. Government license to provide public access under the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

burst pressure [4]. The full-scale test data tended to fit the averaged result of those two bound solutions.

Motivated by these experimental observations, Zhu and Leis [7, 8] developed a new multi-axial plastic yield theory that was referred to as the average shear stress yield criterion, or simply as the Zhu-Leis criterion in public literature. This new theory uses a weighted average of the maximum and octahedral shear stresses and improves the representation of the flow response from the elastic domain into and through the elastic-plastic response [8]. For carbon steels, the Zhu-Leis solution predicts a burst pressure result that is close to the average of the von Mises and Tresca solutions, and thus matches well with the averaged test data [8]. Since the Zhu-Leis solution was published, many researchers [9-12] have used and validated it using different full-scale burst data for pipeline steels and concluded that the Zhu-Leis solution is the best prediction of burst pressure for thin-wall line pipes. This conclusion was further confirmed by Zhu and Leis [13] through comparison with test data from more than 100 full-scale burst tests for pipeline grades ranging from Grade B to X120, and by Zhou and Huang [14] through a model error assessment of existing burst models using another full-scale test database for line pipes.

The Zhu-Leis solution is a function of the strain hardening exponent only, when normalized by the Barlow strength. In contrast, the experimentally measured burst strength, when normalized by the Barlow strength, is a strong function of the strain hardening exponent and a weak function of the UTS and pipe diameter to thickness ratio D/t as well. Therefore, a complex three-parameter regression is needed if one wants to fit a better burst model from the test data.

Machine learning method using the artificial neural network (ANN) is a promising option for developing models for prediction of burst strength of PVs based on the full-scale burst test data while avoiding the difficulties of multiple parameter regression. In comparison to regression, ANNs work well for both single and multiple parameters. From detailed comparisons, Kumar [15] and Lee et al. [16] showed that ANNs have superior performance and higher accuracy than the traditional regression and that both methods are comparable if a linear relationship exists between dependent and independent variables. Recently, the deep learning or ANN approaches were successfully applied to predict the burst strength for pressure vessels or pipelines with or without defects [17-20] and for metal weld joints [21-22].

So encouraged, this paper developed alternative models of burst strength for line pipes using the ANN method. Three models were developed for predicting the burst strength of defect-free line pipes using test data from over one hundred burst tests. Model 1 has one input variable and one hidden layer with three neurons, Model 2 has three input variables and one hidden layer with five neurons, and Model 3 has three input variables and two hidden layers with three neurons for the first hidden layer and two neurons for the second hidden layer. These ANN models were validated using a validation dataset reserved from the full-scale test data in consideration. The ANN predictions are then evaluated through comparison with the Zhu-Leis solution and the linear regression results. After that, the best ANN model is finally determined.

2. MACHINE LEARNING-NEURAL NETWORK BASICS

2.1. Basic concepts

Machine learning is the ability of computers (machines) to learn without being programmed specifically [24]. In machine learning the computer learns from past experience (labeled input and output data) and makes future predictions. Machine learning is generally divided into two overarching categories: supervised and unsupervised learning.

In supervised learning the input and output variables are provided to the model, and a prescribed algorithm is used to determine the relationship and mapping function between the input and output data. The learning process of the algorithm will be supervised and modified until the level of performance is acceptable. For unsupervised learning only input data are provided to the model, and it analyzes the data to find patterns in the data sets. This paper uses the more often utilized supervised learning. In this research the machine learning and prediction are performed using an ANN model.

The concept of ANN was inspired by the biological neural networks. The basic unit of a biological network is called a neuron (or nerve cell). Most neurons have four basic components: a cell body (or soma), an axon, dendrites and synapses which receive input signals from other neurons or from the outside and combines them to perform a generally nonlinear operation on the received data [20]. Figure 1 shows a typical biological neuron. The human brain consists of numerous neurons that are connected by synapses to create a biological network to transfer and analyze data they receive.

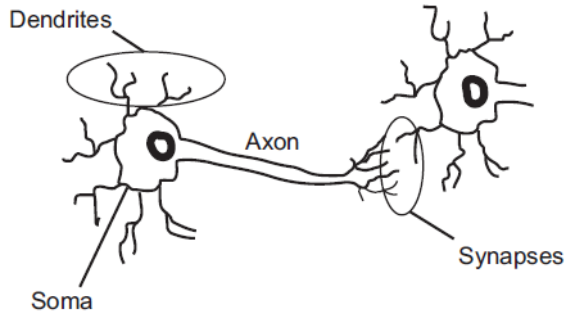

Figure 1. A typical biological neuron [20]

2.2. Architecture of ANN

An ANN works almost the same way as a biological neural network and acts like a function, which receives data as input and outputs the analyzed data after being trained. An ANN that consists of interconnected processing units can be thought as a highly simplified model of a biological neural network. Application has demonstrated that ANNs can be very powerful tools for solving complicated engineering problems [20]. Instead of complex rules and mathematical formulations, ANNs are

capable of learning key information patterns within a multi-information domain. A simple ANN consists typically of three layers: input, hidden and output layer, as shown in Fig. 2. The first layer is the input layer, which accepts independent variables. The last layer is the output layer, which provides output variables. The layers between the input and output layers are considered to reside in a "black box" and are named "hidden layers." In general, the number of both hidden layers and hidden neurons are unknown, and normally determined using the trial-and-error analysis.

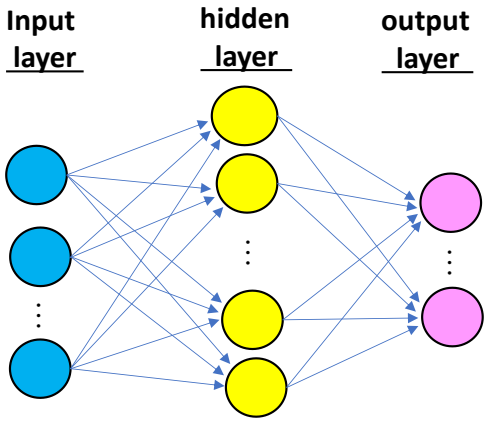

Figure 2. Illustration of a typical neural network model

2.3. Activation function

An ANN model should work if an algorithm is properly prescribed, in which each hidden neuron has a similar structure and works in the same way. Figure 3 illustrates a typical neuron in an ANN model, where each neuron unit has four components: a cell body, an axon, dendrites, and synapses. These components are similar to those for the biological neuron shown in Fig. 1.

Each neuron in the ANN model receives multiple input data, and then adds and processes the combined results with an activation function. The value processed by the activation function becomes an output value to other neurons in the next layer or final output data.

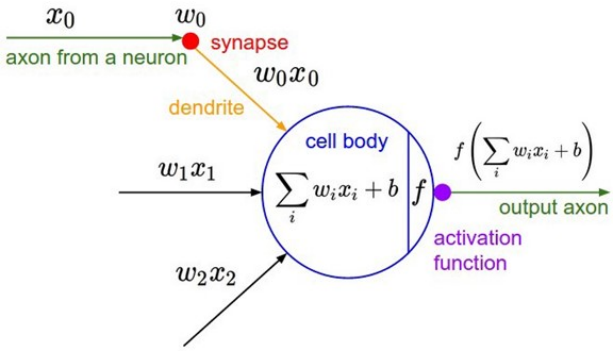

Figure 3. A typical neuron in an ANN model

A connection (dendrite) between layers is weighted to control the strength of the connection. The signal received at neuron *i* is described by the following combined linear function:

$$u_i = \sum_{j=1}^{n} w_{ij}x_j + b_i \quad (1)$$

and the output information is processed by an activation function:

$$y_i = f(u_i) \quad (2)$$

where u_i is the linear combined variable from the input data; x_1, x_2, …, x_N are the independent input variables; N is the total number of input variables; y_i is the dependent output variable of the neuron and can be the final output or the input to another hidden layer; w_{i1}, w_{i2}, …, w_{iN} are the weights of input variables at neuron *i*; *f* is an activation function; and b_i denotes a bias that has an effect of increasing or lowering the net input of the activation function and the flexibility of the model to fit the training data. An activation function is needed to introduce nonlinear real-world data to ANNs. If not, the output signal becomes a simple linear function, and then the ANN will act as a linear regression with limited learning power.

There are many activation functions used in an ANN modeling analysis. This paper adopts the most often used sigmoid (or logistic) function:

$$y(x) = \frac{1}{1+e^{-x}} \quad (3)$$

where e is the base of an exponential function.

Figure 4 shows the variation of the sigmoid function with its variable x. This function takes a value in the range of (0, 1). In a logistic regression the response variable describes the probability that the outcome is the positive case. If the response variable is equal to or exceeds a discrimination threshold, the positive class is predicted; otherwise, the negative class is predicted. The sigmoid function is used for the hidden units, and the linear function is usually used only for output units.

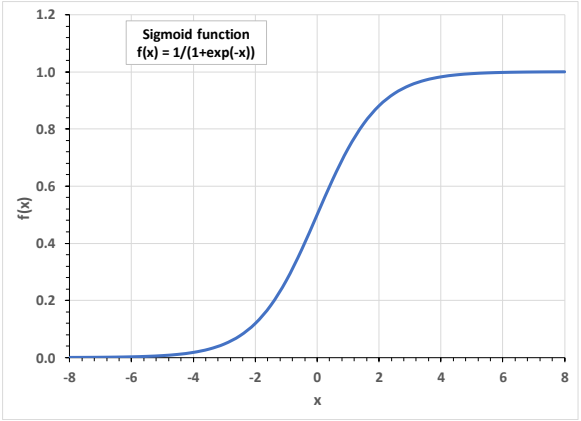

Figure 4. Sigmoid function

2.4. Cost function

For modeling burst strength, the goal of the ANN is to predict a real-value variable y from a given pattern of input variables x_i. The performance of the ANN needs to be measured to know how well the model predicts the output variables.

During the learning step an initial network structure is created with assumed numbers of hidden layers and hidden neurons and initial weights being applied to the connections. Observations enter the ANN, and the output $\hat{y}$ is obtained. There will be a difference or error between the ANN prediction $\hat{y}$ and the real observation y. Different indices were used to measure the error, such as the mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE). This paper adopts the most often used RMSE that is expressed by:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \quad (4)$$

To minimize the model error during data training, the initial network weights w_{ij} are updated. In the error calculations, the error when plotted against the weights w_{ij} is called the cost function J(w), because it determines the cost/penalty of the model. Figure 5 shows the cost function J(w) versus the weights w, where a real minimum cost J_{min} exists. As illustrated, minimizing the model error is also called minimizing the cost function. Usually, a gradient descent algorithm is used to determine the minimum cost function J_{min}. It starts with random model parameters, calculates the error for each learning iteration, and keeps updating the model parameters to move closer to the values that results in a minimum cost. During each learning iteration, the error is propagated backward to the model parameters on the previous layers, and thus the whole process is also referred to as backpropagation, which is one of the data training methods in the ANN modeling.

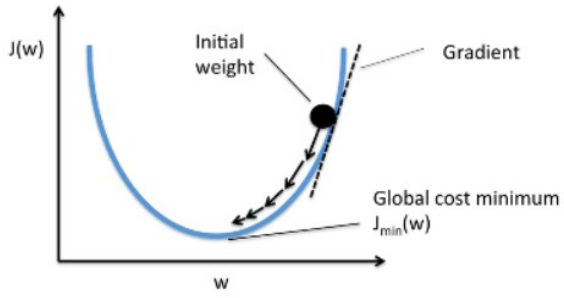

Figure 5. An illustration of the cost function

2.5. Dependent variable normalization

Because the sigmoid function has a value in the range (0, 1), all outputs from Eq. (2) take a value in the same range. As a result, all dependent variables are required to be transformed to the same scale. This is accomplished using normalization or scaling. In this study the min-max scaler is used to rescale all independent variables. This scaler would transform the dependent variables into values between 0 and 1 using following equation:

$$y_{norm} = \frac{y - y_{min}}{y_{max} - y_{min}} \quad (5)$$

where y is a raw data point before normalization from an independent variable, and y_{norm}, y_{min} and y_{max} are the normalized data and maximum and minimum value of the y dataset, respectively.

2.6. Machine learning procedures

With the fundamental knowledge introduced above, an ANN model can be established and then executed to predict the desired output results through the following steps.

Step 1. *Analyze database and select training dataset*. In general, a database used for a machine learning needs to be divided into a training dataset using about 80% of the database, and a validation dataset using the remaining 20% of the database.

Step 2. *Determine input and output variables*. Based on the database, the primary independent input variables and the desired output variable(s) should be determined.

Step 3. *Build ANN model architecture*. Based on the number of input and output variables, the number of hidden layers and the numbers of hidden neurons for each hidden layer should be estimated, and then an ANN model is built.

Step 4. *Initialize the model weights and biases*. Because all values at hidden neurons are unknown, the model parameters of weights and biases are unknown, and thus their initial values are randomly assigned.

Step 5. *Solve the cost function equation to minimize the model error*. Different computational tools can be used to minimize the model cost function, including MATLAB machine learning toolbox [20], Microsoft Excel Solver, and other proprietary computer software. The error calculations are iterated through training samples until the convergence is reached.

Special attention needs to be paid to avoid under-fitting when the model has fewer features and thus is not able to learn from the training data very well. Or over-fitting, when the model has complex functions and thus is able to fit the training data very well but is unable to generalize to predict new data also must be avoided.

3. THEORETICAL SOLUTION OF BURST STRENGTH OF LINE PIPES

As introduced previously, the Tresca and von Mises flow theories determine two bound solutions of burst pressure for pipes, while the average shear stress flow theory or Zhu and Leis theory [7-8] determines a more accurate flow solution of burst pressure. The three flow solutions of burst pressure were expressed in terms of the Tresca, Zhu-Leis and von Mises flow theories as:

$$P_T = \left(\frac{1}{2}\right)^{n+1} \frac{4t}{\bar{D}} \sigma_{uts} \quad (6)$$

$$P_A = \left(\frac{2+\sqrt{3}}{4\sqrt{3}}\right)^{n+1} \frac{4t}{\bar{D}} \sigma_{uts} \quad (7)$$

$$P_M = \left(\frac{1}{\sqrt{3}}\right)^{n+1} \frac{4t}{\bar{D}} \sigma_{uts} \quad (8)$$

where $\bar{D} = D - t$, D is the pipe outside diameter, t is the wall thickness, and n is the strain hardening exponent that can be measured from a simple tensile test or estimated from the YS and UTS. Zhu and Leis [25] obtained two simple equations for estimating the n from the YS defined at the 0.2% offset strain and the 0.5% total strain, respectively.

When the n approaches to zero, Equation (6) becomes:

$$P_0 = \frac{2t}{\bar{D}} \sigma_{uts} \quad (9)$$

This reference burst pressure is often called the Barlow strength for a defect-free pipe.

When normalized using the Barlow strength, the von Mises, Zhu-Leis and Tresca solutions are a function of strain hardening exponent only. As shown in Fig. 6, the von Mises solution is an upper bound prediction, the Tresca solution is a lower bound prediction, and the Zhu-Leis solution is an intermediate prediction of the two bounds. Figure 6 demonstrates that the normalized burst strength, P_b/P_0, is a function of strain hardening exponent for all three burst pressure solutions.

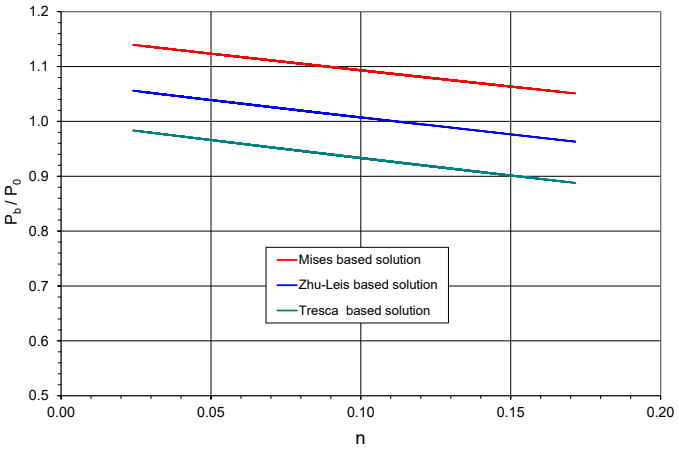

Figure 6. Comparison of Mises, Zhu-Leis and Tresca-based solutions of burst pressure of defect-free pipes

4. EXPERIMENTAL DATA AND REGRESSION

4.1. Experimental database of burst strength

A large database involving 109 full-scale hydrostatic burst tests has been assembled by Zhu and Leis [8, 13] from open literature for end-capped, defect-free pipes. The database covers a wide range of pipeline steel grades from Grade B to X120, with the yield to tensile strength (Y/T) ratios from 0.63 to 0.98, nominal diameters from 8 to 48 inches (203-1219 mm), and the diameter to thickness ratios from 28 to 114. In addition to pipeline steels, the database also includes a small portion of test data for other structural steels, like those employed in tubular goods. Because a wide range of steel grades is involved, the dependence of burst pressure on the strain hardening response should be appropriately assessed. Figure 7 shows that the experimental burst pressure is a function of n, with all outcomes normalized by the Barlow strength, P_0, in Eq. (9). The Barlow strength is represented by the horizontal line of $P_b/P_0 = 1$ in Fig. 7 and is independent of the strain-hardening exponent.

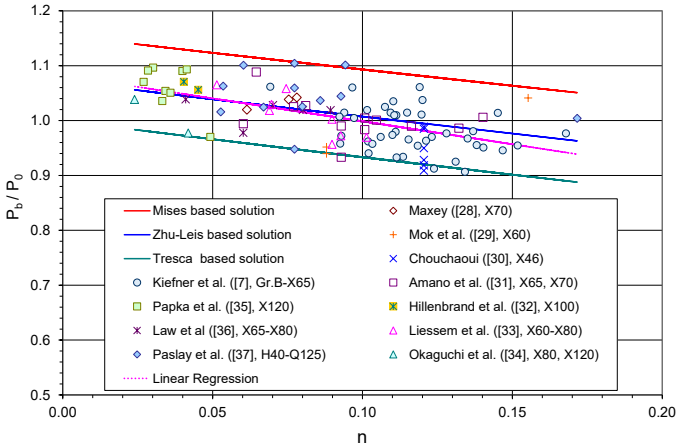

Figure 7. A collection of full-scale burst test data for various pipe sizes and different steels

Also included in Fig. 7 are the three burst solutions from Eqs (6), (7) and (8) for von Mises, Zhu-Leis and Tresca criteria.

The symbols in Fig. 7 represent the full-size burst test data with a focus on larger-diameter, thin-walled line pipes. This includes 43 tests by Kiefner et al. [7] for Grade B to X65 steels; 12 tests by Amano et al. [31] for X65 and X70 steels; three tests by Maxey [28] for X70 steel; two tests by Mok et al. [29] for X60 steels; six tests by Chouchaoui [30] for X46 steels; two tests by Hillenbrand et al. [32] for X100 steel; eight tests by Liessem et al. [33] for X60 to X80 steels; two tests by Osaguchi et al. [34] for X80 and X120 steels; nine tests by Papka et al. [35] for X120 steels; five tests by Lawet al. [35] for X65 to X80 steels; 12 tests by Paslay et al. [37] for H40 to Q125 well casing steels; one test by Netto et al. [38] for AISI 1020 mild steel; three tests by Knoop et al. [12] for X80 steel; and one test by Besel et al. [39] for X100 steel. It should be noted that all the reference numbers cited in this section are taken from Zhu and Leis [8] rather than those given in the present paper. These full-scale tests comprise a large comprehensive database for the pipe sizes and steel grades of interest and serve as a strong basis for the following data analysis and evaluation using the regression or machine learning. More details about the 109 full-scale tests can be found in Appendix 1 of Zhu and Leis [13].

The following observations can be drawn from the data comparisons in Fig. 7:

(1) The burst database shows a large scatter band with a clear dependence on the n. The scatter bandwidth is constant and independent of the n. The scatter shows an approximately linear relationship as the n decreases.

(2) The von Mises solution provides an upper bound prediction of burst pressure data for all n values, including n = 0. As a result, the von Mises solution overestimates the burst strength for all full-scale tests.
(3) The Tresca flow solution provides a lower bound prediction of burst pressure data and underestimates the burst strength generally for all full-scale tests.
(4) The Zhu-Leis solution tracks the central tendency of the database for all n values and provides the best predictions of experimental data on average.

4.2. Regression analysis

With the database discussed above, the regression analysis is performed to determine a curve-fit model and the major factors that affect the normalized burst strength measurement of P_b/P_0. This includes the relationships between the normalized burst strength P_b/P_0 and the strain hardening exponent n, D/t, or UTS in reference to the database, as discussed next.

4.2.1 Strain hardening exponent effect

Figure 8 replots the normalized strength data against n that were reported in Fig. 7 for the full-scale tests on burst pressure in consideration along with a linear best curve fit. The goodness-of-measure of this linear curve fit is R^2=0.336. It shows that the data scatter is rather large, but the linear curve fit captures the trend in the test dataset. The linear regression model of burst strength is expressed as

$$P_b = (-0.8363n + 1.0823)\frac{2t}{\bar{D}}\sigma_{uts} \quad (9)$$

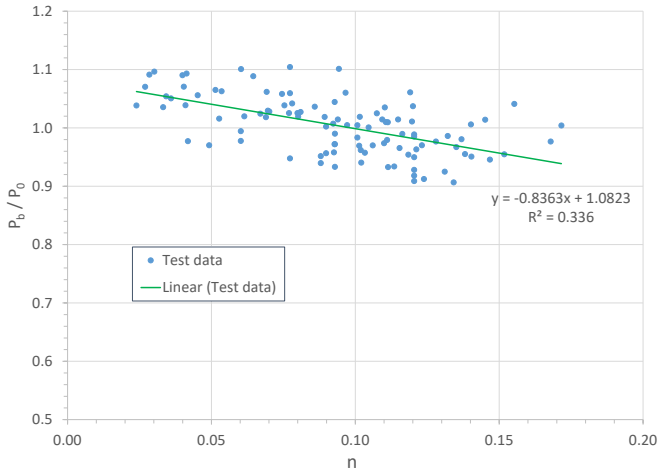

Figure 8. Curve-fit relationship between normalized burst strength with n

To evaluate the quality of the linear regression model, Figure 9 compares the normalized curve-fitted burst strength with the normalized measured data of burst strength. Again, this figure shows a large scatter of the test data. The linear regression model has the goodness-of-measure R^2=0.336, and the slope of the linear curve fit is much slower than the slope of the 1:1 line. When the measured strength $P_b < P_0$, the regression model tends to overpredict the burst strength. When the measured strength $P_b > P_0$, the regression model tends to underpredict the burst strength. Therefore, the Barlow strength P_0 is an overall averaged estimate of burst strength for all tests in consideration.

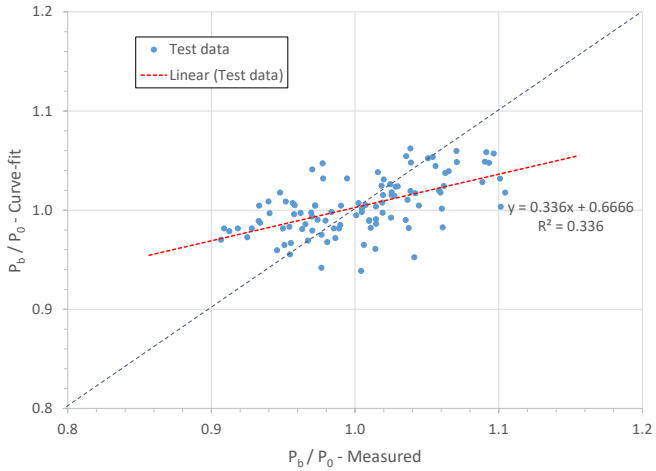

Figure 9. Comparison of normalized curve-fitted burst strength with normalized measure burst strength

4.2.2 Pipe size D/t effect

Figure 10 shows the effect of pipe size D/t on the normalized burst stregnth P_b/P_0. When D/t < 30, all values of P_b/P_0 > 1 and has a linear correlation. When D/t > 30, the test data randomly distribute around P_b/P_0 = 1 with a large scatter. The curve-fit line of P_b/P_0 vs D/t has a much smaller slope (m=0.0005 vs 0.8363) than the curve-fit line of P_b/P_0 vs n and a small goodness-of-measure R^2=0.0682. Thus, the linear relationship between P_b/P_0 and D/t is very weak. This inidicates that the D/t has a weak effect on P_b/P_0 generally, and the effect is centeral to D/t < 30.

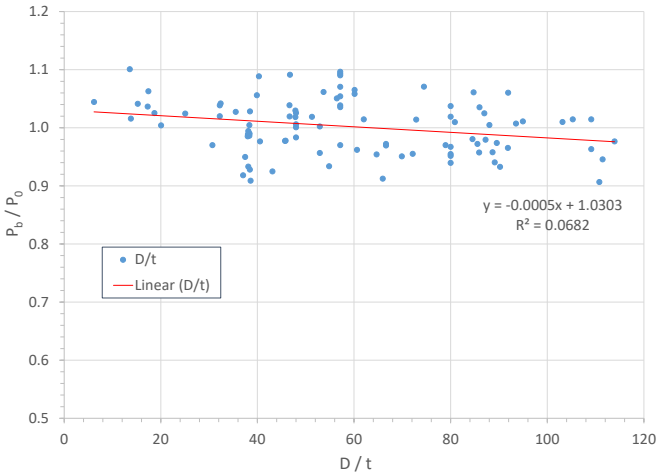

Figure 10. Curve-fit relationship between normalized burst pressure with D/t

4.2.3 Material property UTS effect

Figure 11 shows the effect of the UTS of pipeline steels used in the tests on the normalized data of burst strength P_b/P_0. As evident in this figure, the curve-fit line of P_b/P_0 vs UTS has a smaller slope (m=0.0011 vs 0.8363) than the curve-fit line of P_b/P_0 vs n but a larger slope (m=0.011 vs 0.0005) than the curve-fit line of P_b/P_0 vs D/t. In addition, the linear curve fit of P_b/P_0 vs D/t has an intermediate goodness-of-measure R^2=0.2685. As such, the material property UTS has a relatively larger effect on the P_b/P_0 than the pipe geometry D/t, but a weaker effect on the P_b/P_0 than n.

It is concluded that the strain hardening exponent n is the primary factor, and the D/t and UTS are the secondary factors that affect the normalized data of burst strength P_b/P_0.

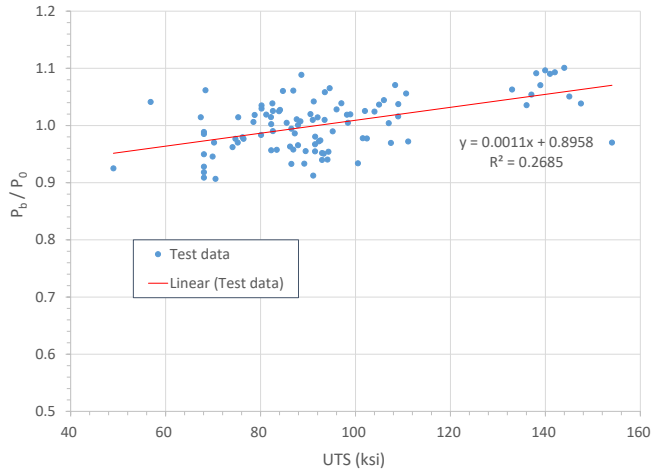

Figure 11. Curve-fit relationship between normalized burst pressure with UTS

5. ANN MODELING OF BURST STRENGTH

This section applies the machine learning to model and predict the burst strength from the database in Fig. 8. Following the procedures given in Section 2.6, the database is divided into two datasets: 80% used for training and 20% used for validation. Three ANN models will be built: the first one has one input variable (strain hardening exponent n) and the other two have three input variables (n, UTS, and D/t). All three models have one output variable (normalized burst strength P_b/P_0). One hidden layer is assumed for the first two models, and two hidden layers are assumed for the third model. In general, the number of hidden neurons is equal to or larger than the sum of the number of input and output variables.

5.1. ANN model 1 with one input variable

Figure 12 shows the architecture of the first ANN model, where the input variable is the strain hardening exponent n, and the output variable is the normalized burst strength (P_b/P_0) of line pipes. One hidden layer with three hidden neurons is assumed. Thus, this model is simply denoted as 1x3x1. The model contains six weights as marked in Fig. 12 and four biases, leading to 10 unknown parameters in total. These unknown model parameters will be determined by the training dataset. Each parameter is initialized to be a random value, and the Excel Solver is utilized to minimize the RMSE error of the cost function for this model. When a minimal RMSE is reached during iteration, the weights and biases are fixed from the training dataset and the model is completely created.

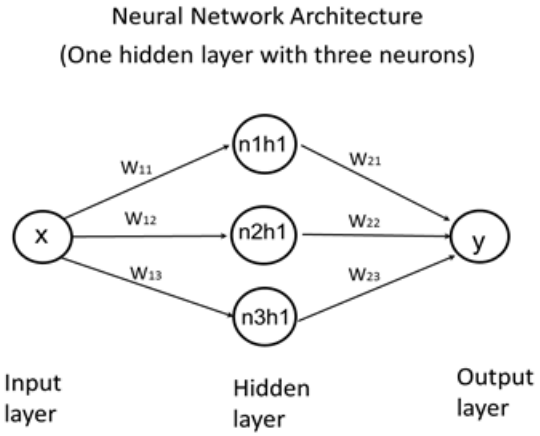

Figure 12. Architecture of ANN model 1

Figure 13 compares the ANN model 1 predictions (output data) with the scaled measured target data for each test. It shows that the ANN predictions match fairly well with most of target test data. Furthermore, Fig. 14 compares the ANN model 1 predictions with the measured data for both training and validation datasets. As evident from this figure, the model has the goodness-of-fit measure R^2 = 0.350 for the training dataset and R^2=0.384 for the validation dataset. This indicates that ANN model 1 has a comparable model error for the training and validation datasets. Therefore, ANN model 1 has obtained the best-fit parameters of weights and biases for its architecture.

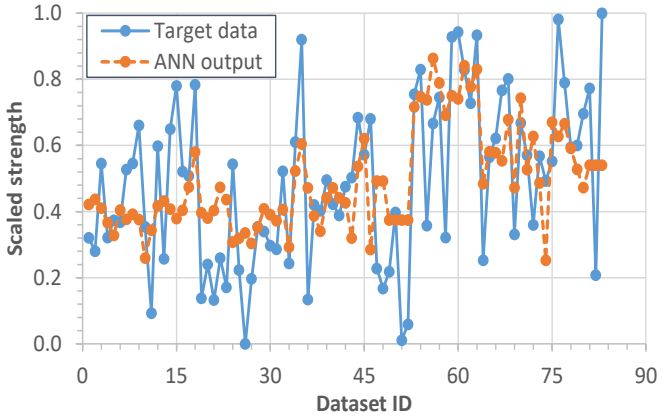

Figure 13. Comparison of ANN output data with target data of burst strength for ANN model 1

It should be noted that because the ANN model randomly initializes the weights and biases in each try of solution determination to minimize the model error, the initial weights and biases will be different. This results in different final weights and biases learned by the network, and the final model results will be slightly different. However the accuracy and performance will be approximately the same.

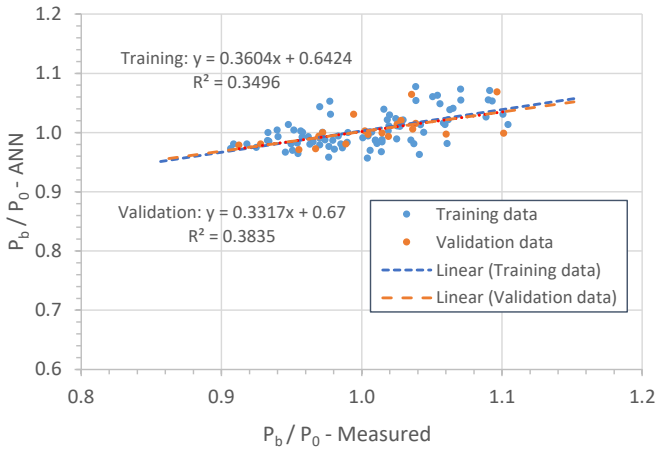

Figure 14. Comparison of ANN model 1 predictions with measured data for training and validation datasets

5.2. ANN Model 2 with three input variables

Figure 15 shows the architecture of the more complex ANN model 2, where three input variables are n, UTS and D/t, and one output variable is the normalized burst strength. One hidden layer with five hidden neurons is assumed. Accordingly, Model 2 is simply denoted as 3x5x1. This model contains twenty weights and six biases, leading to 26 unknown parameters in total. These unknown model parameters will be determined using the same training dataset as used for Model 1.

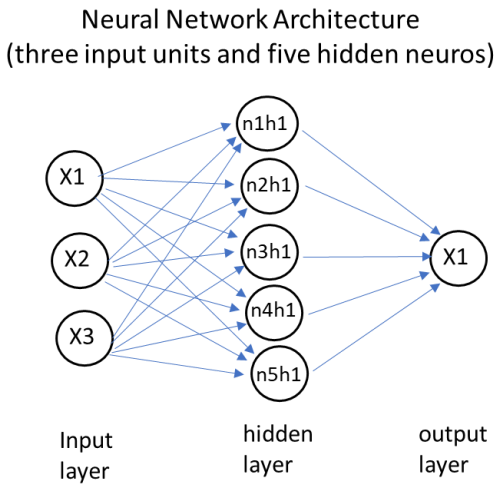

Figure 15. Architecture of ANN model 2

Again, each parameter is initialized to be a random value, and the Excel Solver is utilized to minimize the RMSE error of the cost function for this model. When a minimal RMSE is reached during iteration, the weights and biases are obtained from the training dataset and the model is fully built.

Figure 16 compares the ANN model 2 predictions (output data) with the scaled measured target data of burst strength for each test. It shows that the ANN predictions match fairly well with most of the target data.

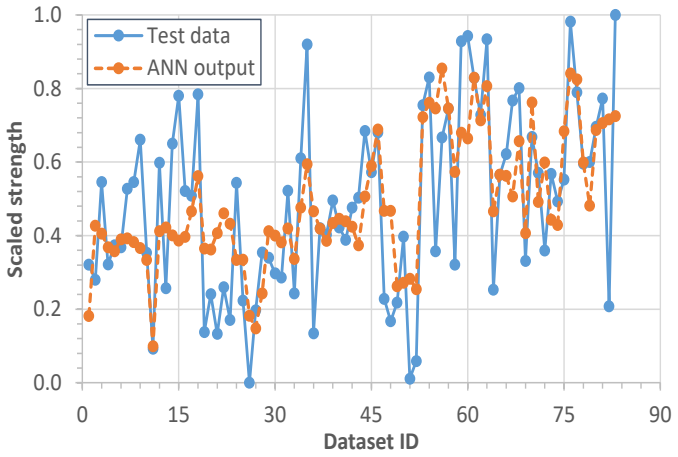

Figure 16. Comparison of ANN output data with target data of burst strength for ANN model 2

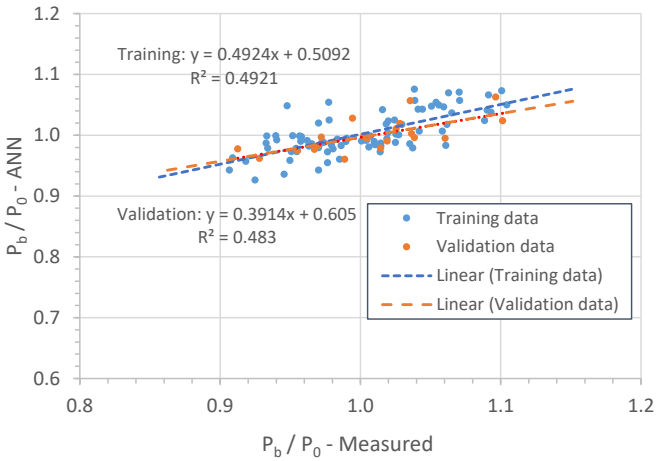

Figure 17. Comparison of ANN model 2 predictions with measured data for training dataset and validation dataset

Figure 17 compares the ANN model 2 predictions with the measured data for both training and validation datasets. As evident from this figure, this model has the goodness-of-fit measure $R^2 = 0.492$ for the training dataset and $R^2=0.483$ for the validation dataset. This indicates that ANN model 2 has a comparable model error for both training and validation datasets. Thus, ANN model 2 has obtained the best-fit parameters of weights and biases for its architecture.

5.3. ANN model 3 with three input variables

Figure 18 shows the architecture of model 3, the most complex ANN model 3 developed in this work. The three input variables are n, UTS and D/t, and the one output variable is the normalized burst strength. Two hidden layers are assumed, where the first hidden layer has three hidden neurons, and the second hidden layer has two hidden neurons. Accordingly, Model 3 is simply denoted as 3x3x2x1. This model contains 17 weights and six biases, leading to 23 unknown model parameters. These unknown model parameters are determined by the same training dataset used for Model 1. Again, each parameter is initialized to be a random value, and the Excel Solver is utilized to minimize the RMSE error of the cost function. When a minimal RMSE is reached during iteration, the weights and biases are obtained from the training dataset and the model is fully obtained.

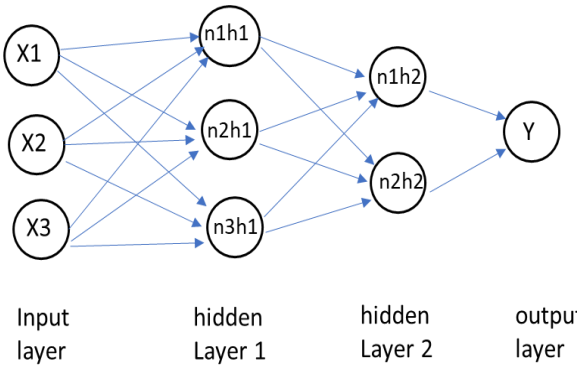

Figure 18. Architecture of ANN model 3

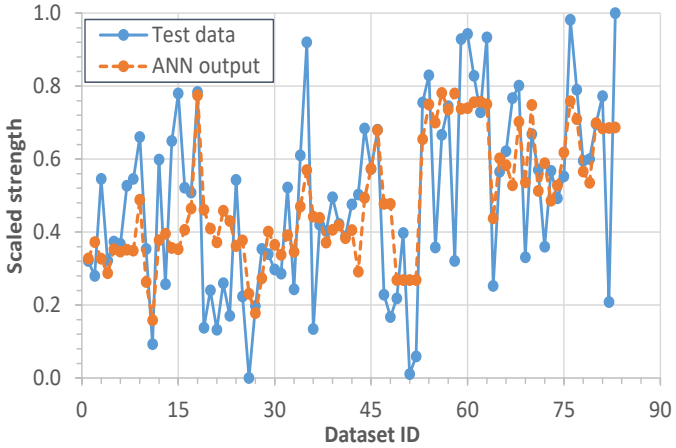

Figure 19. Comparison of ANN output data with target data of burst strength for ANN model 3

Figure 19 compares the ANN model 3 predictions (output data) with the scaled measured target data for each test. It shows that the ANN predictions match fairly closely with most of the target test data. Figure 20 compares the ANN model 3 predictions with measured data for both training and validation datasets. As evident from this figure, this model has the goodness-of-fit measure $R^2 = 0.465$ for the training dataset and $R^2=0.490$ for the validation dataset. This indicates that ANN model 3 has a comparable model error or similar prediction quality for both training and validation datasets. Therefore, ANN model 3 has obtained the best-fit hyperparameters in its architecture.

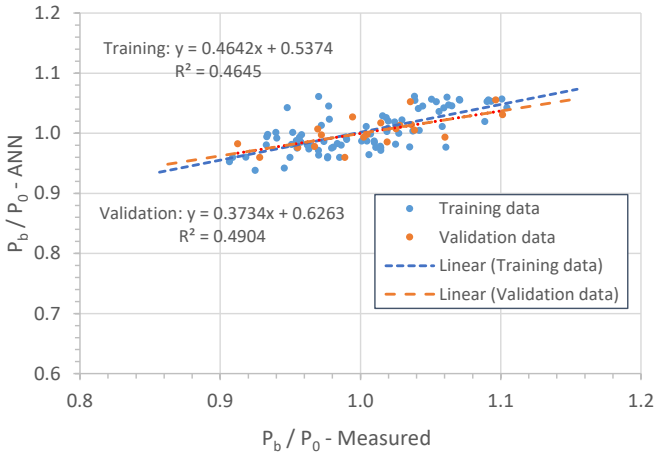

Figure 20. Comparison of ANN model 3 predictions with measured data for training dataset and validation dataset

6. COMPARISON OF ANN AND REGRESSION MODELS

Figure 21 compares the machine learning results with the linear regression results, the three theoretical burst solutions, and all measured data of burst strength for defect-free pipes, where the machine learning results include the predictions by three ANN models, and all predicted and measured burst pressure values are normalized by the Barlow strength. It is observed from this figure that (1) the ANN model 1, the linear regression and the Zhu-Leis solution predict comparable burst strengths with very small differences, and (2) ANN model 2 and model 3 predict comparable results with insignificant differences.

Figure 22 compares all model burst predictions that are normalized by the measured burst strength for all full-scale tests in consideration. This figure shows that all burst pressure predictions from three ANN models, linear regression and Zhu-Leis solution are comparable with insignificant differences.

Figure 23 compares the model predicted burst strength with the measured burst data for all full-scale tests in consideration, where the model predictions include those predicted by the three ANN models, the linear regression equation, and the Zhu-Leis solution. As evident in this figure, all model predictions are nearly identical to the measured burst data when the measured burst strength is less than 15 ksi (103.42 MPa), while some deviations are observable when the measured burst pressure is large than 15 ksi (103.42 MPa). For the latter case, the model predictions from ANN model 2 and model 3 are more accurate than other model predictions. Due to its simple architecture, ANN model 2 is recommended as the best machine learning model for predicting burst strength of defect-free thin-wall and thick-wall line pipes.

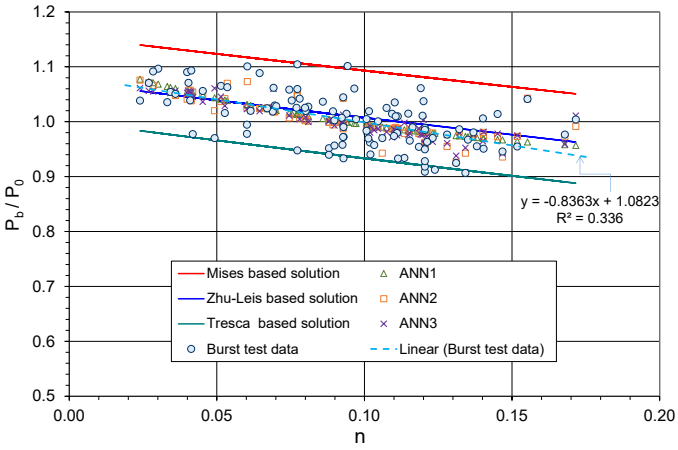

Figure 21. Comparison of machine learning results with regression results, burst solutions and measured test data

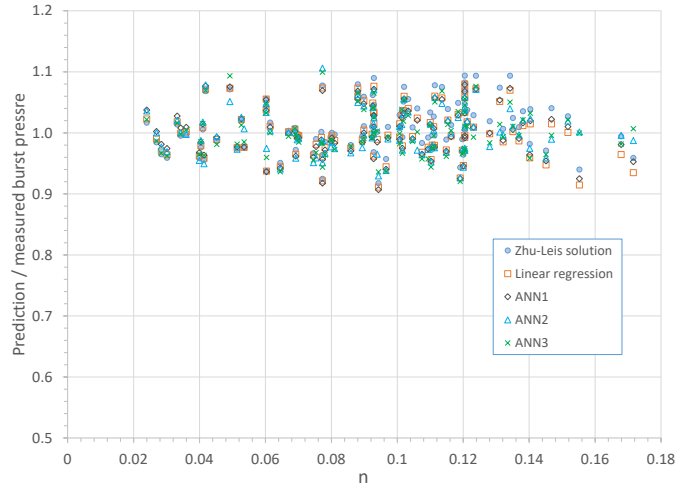

Figure 22. Comparison of model predictions normalized by the measure burst strength for all full-scale tests

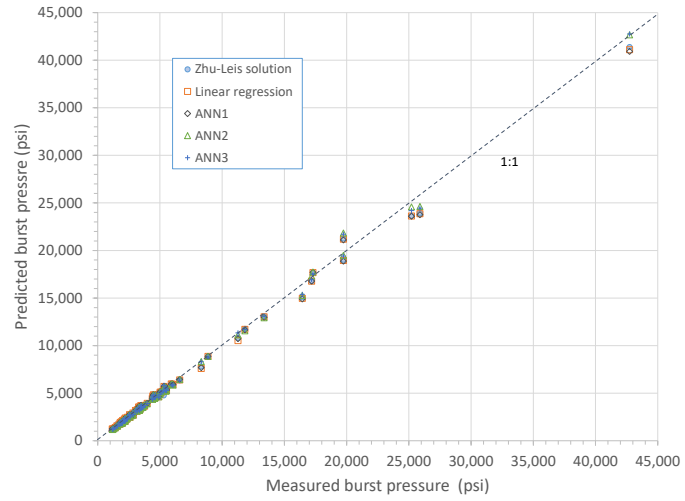

Figure 23. Variation of predicted burst strength with measured burst strength for all full-scale tests

7. CONCLUSIONS

This paper revisited a large burst database that contains more than 100 full-scale tests on burst strength of defect-free line pipes. The linear regression analysis showed that the measured burst strength, when normalized by the Barlow strength, is a strong function of strain hardening exponent n of the pipeline steels and a weak function of UTS and D/t as well. In contrast, three representative theoretical solutions of burst strength in terms of von Mises, Tresca and Zhu-Leis flow theories, when normalized by the Barlow strength, are a function of n only. Accordingly, this paper used ANNs and developed three alternative models of burst strength based on a large burst strength database for better characterization of the effect of material properties (UTS, n) and pipeline geometry (D/t) on the burst strength of line pipes.

Three ANN models were developed. Model 1 is a simple ANN model with one input variable and one hidden layer of three neurons. Model 2 is a common ANN model with three input variables and one hidden layer of five neurons. Model 3 is a more complex ANN model with three input variables and two hidden layers, where the first hidden layer has three neurons, and the second hidden layer has two neurons. The ANN model weights and biases were determined using 80% of the full-scale test data and validated using the remaining 20% of the full-scale data. The following conclusions follow from the results:

(1). The full-scale burst database has a large scatter band. The von Mises solution is the upper bound prediction, and the Tresca solution is the lower bound prediction. The Zhu-Leis solution agrees well with the averaged burst data and is comparable to the linear regression results.

(2). Since the output variable (P_b/P_0) and the independent input variable (n) have an approximately linear relation, ANN model 1 and the linear regression model are comparable. Both models predict similar averaged burst strength of test data.

(3). When the dependent output variable (P_b/P_0) is related in a complex function to three independent input variables (n, UTS, D/t), ANN model 2 and model 3 predict comparable results in a higher accuracy than ANN model 1 does. Because of the simple architecture, ANN model 2 is recommended as the best machine learning model developed in this work for more accurately predicting the burst strength of defect-free line pipes.

In summary, the Zhu-Leis solution is simple and accurate for predicting the burst strength of thin-wall pipes. Actually, all models, including the regression one, are very good. This paper successfully applies the machine learning to predict the burst strength for defect-free pipes and sets a stage for next step of using the machine learning to determine remaining strength for localized metal-loss or corroded line pipes.

ACKNOWLEDGEMENTS

The present authors are grateful to the financial support by the Department of Energy (DOE) and its Laboratory Directed Research and Development (LDRD) program through the LDRD Project 2022-00077 at Savannah River National Laboratory.

REFERENCES

[1] Christopher T, et al. "A Comparative Study on Failure Pressure Estimation of Unflawed Cylindrical Vessels," *International Journal of Pressure Vessels and Piping*, Vol. 79, 2002:53-66.

[2] Law M, Bowie G, Fletcher L, Piper J. "Burst Pressure and Failure Strain in Pipeline, Part 1: Comparison of Ring-Expansion and Tensile Testing in Gas Pipeline"; "Part 2: Comparison of Burst Pressure and Failure-Strain Formulas"; Part 3: "Failure Pressure Calculated by the Method of Plastic Instability," *Journal of Pipeline Integrity* Vol. 3, 2004: 95-113.

[3] Law M, Bowie G. "Prediction of Failure Strain and Burst Pressure in High Yield-to-Tensile Strength Ratio Linepipe," *International Journal of Pressure Vessels Piping*, Vol. 84, 2007: 487-492.

[4] Stewart G, Klever FJ. "An Analytical Model to Predict the Burst Capacity of Pipelines," *Proceedings of International Conference of Offshore Mechanics and Arctic Engineering*. Vol. V, Pipeline Technology, 1994: 177-188.

[5] Cooper WE. "The Significance of the Tensile Test to Pressure Vessel Design," *Welding Journal - Welding Research Supplement*; January 1957: 49s-56s.

[6] Svensson NL. "The Bursting Pressure of Cylindrical and Spherical Vessels," *Journal of Applied Mechanics,* Vol. 25, 1958: 89-96

[7] Zhu XK, Leis BN. "Accurate Prediction of Burst Pressure for Line pipes," *Journal of Pipeline Integrity*, Vol. 4, 2004: 195-206.

[8] Zhu XK, Leis BN. "Average Shear Stress Yield Criterion and Its Application to Plastic Collapse Analysis of Pipelines," *International Journal of Pressure Vessels and Piping*, Vol. 83, 2006: 663-671.

[9] Zimmermann S, Hohler S, Marewski U. "Modeling Ultimate Limit States on Burst Pressure and Yielding of Flawless Pipes," *Proceedings of the 16th Biennial Pipeline Research Joint Technical Meeting*, Canberra, Australia, April 16-19, 2007. Paper 13.

[10] Zimmermann S, Marewski U, Hohler S. "Burst Pressure of Flawless Pipes," *3R International*, Vol. 46, Special Edition, 2007: 28-33.

[11] Knoop FM, et al. "Mechanical Properties and Component Behavior of X80 Helical Seam Welded Large Diameter Pipes," *Proceedings of the 8th International Pipeline Conference*. Calgary, Alberta, Canada; September 27-October 1, 2010.

[12] Bony M, Alamilla JL, Vai R, Flores E. "Failure Pressure in Corroded Pipelines Based on Equivalent Solutions for Undamaged Pipe," *Journal of Pressure Vessel Technology*, Vol. 132, 2010: 051001.

[13] Zhu XK, Leis BN. "Evaluation of Burst Pressure Prediction Models for Line Pipes," *International Journal of Pressure Vessels and Piping*, Vol. 89, 2012: 85-97.

[14] Zhou W, Huang G. "Model Error Assessment of Burst Capacity Models for Defect-Free Pipes," *Proceedings of the 9th International Pipeline Conference*, Calgary, Canada, September 24-28, 2012.

[15] Kumar UA. "Comparison of Neural Networks and Regression Analysis: A New Insight," *Expert Systems with Applications*, Vol. 29, 2005: 424-430.

[16] Lee KY, et al. "Comparison and Analysis of Linear Regression and Artificial Neural Network," *International Journal of Applied Engineering Research*, Vol. 12 (20), 2017: 9820-9825.

[17] Zolfaghari A, Izadi M. "Burst Pressure Prediction of Cylindrical Vessels Using Artificial Neural Network," *Journal of Pressure Vessel Technology*, Vol. 142, 2020: 031303.

[18] Liu X., et al. "An ANN-Based Failure Pressure Prediction Method for Buried High-Strength Pipes with Stray Current Corrosion Defects," *Energy Science & Engineering*, Vol. 8, 2020: 248-259.

[19] Oh D, et al., "Burst Pressure Prediction of API 5LX-Grade Dented Pipelines Using Deep Neural Network," *Journal of Marine Science and Engineering*, Vol. 8, 2020: 8100766.

[20] Phan HC, Dhar AS, "Predicting Pipeline Burst Pressures with Machine Learning Models," *International Journal of Pressure Vessels and Piping*, Vol. 191, 2021: 104384.

[21] Arunchai T. et al. "Resistance Spot Welding Optimization Based on Artificial Neural Network," *International Journal of Manufacturing Engineering*, 2014: 154784.

[22] Tosun E, Calik A. "Failure Load Prediction of Single Lap Adhesive Joints Using Artificial Neural Networks," *Alexandria Engineering Journal*, Vol. 55, 2016: 1341-1346.

[23] Zhu JB, Zhu XK, Zhang W. "Machine Learning Modeling of Dynamic Strength of Resistance Spot Welds in High Strength Steels," *Proceedings of ASME Pressure Vessels and Piping Conference*, Virtual Online, July 12-16, 2021, PVP2021-63093.

[24] Nabi J. "*Machine Learning Fundamentals: Basic Theory Underlying the Field of Machine Learning*," 2018, online tutorial article at *https://towardsdatascience.com/machine-learning -basics-part-1-a36d38c7916*

[25] Zhu XK, Leis BN. "Influence of Yield-to-Tensile Strength Ratio on Failure Assessment of Corroded Pipelines," *Journal of Pressure Vessel Technology*, Vol. 127, 2005: 436-442.