Forecasting Incident Rates through Artificial Intelligence

Abdullah Al-Mutairi, M.S
HES Specialist
Chevron
Houston, TX

Joel M. Haight, Ph.D, P.E
Associate Professor
Pennsylvania State University
State College, PA

Introduction

Background

Safety can be an expensive aspect of industrial operations unless efforts are made to enhance and optimize health and safety programs to reduce the long-term cost associated with health and safety related incidents and damage. The objective of a health and safety program is to minimize or prevent loss to humans, the environment, property and profits due to incidents (OSHA 2006). These programs are implemented by applying human resource time to preventive intervention activities that are expected to prevent or minimize loss (OSHA 2006). The National Safety Council estimates the cost of workplace injury in the year 2004 to be $142.2 billion (NSC 2005). This cost is expected to rise (NSC 2005) due to increases in medical and legal fees unless optimization efforts and enhancements of health and safety programs that reduce the likelihood of incidents taking place. One step towards achieving this objective would be to quantify and analyze intervention activity and incidents for an existing health and safety program.

Using Neural Networks, which is a form of artificial intelligence, the researchers attempt to determine and identify a relationship between safety intervention activity and the incident rate. Once the relationship has been established it will then allow the analyst to use it as a forecasting tool to predict future incident rates given the level of safety intervention activities. In this study incidents recorded were comprised of physical injuries to workers as well as spills and equipment failure.

This research is a continuation of the previous work by Haight et al. (2001) and Iyer et al. (2004 & 2005) which focused on quantifying safety intervention activities with the incident rate. It is based on the relationship between four safety intervention factors which are considered inputs and the incident rate is the only output. Figure 1 is a graphical representation of the model established by Haight et al. (2001) that lays the foundation for quantifying safety intervention activities with the incident rate. Table 1 provides an example of a data sheet used during the data collection phase from the forestry division of a power company. This forestry division of the power company provided the setting and the context for this research.

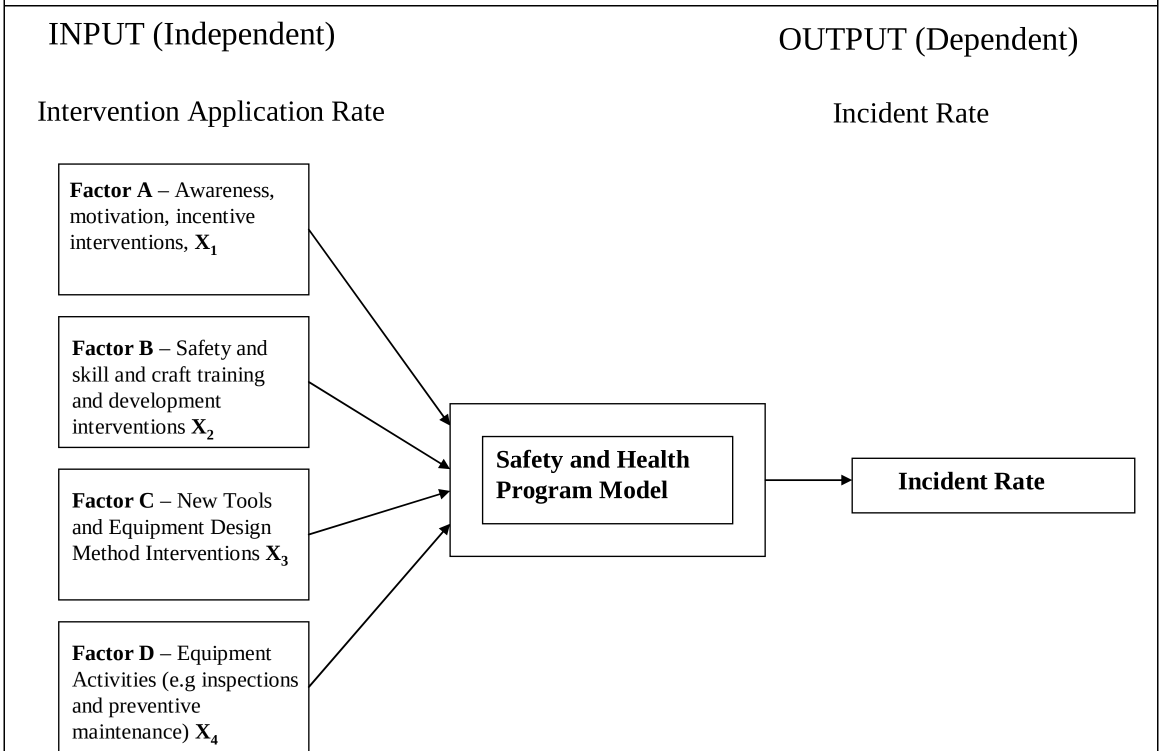

Figure 1: **Representation of the Safety and Health Program—Mathematical Model (adapted from Haight et al. 2001)**

	Week -	Date from:	**5-Jan-04**	Date to:	**9-Jan-04**	
Data Input Representative	No Issue	No Issue	No Issue	No Issue	No Issue	
Cost Center	917/7366	7367	7368	7752	7755	Totals
Safety Activities						
Factor A – Safety Awareness and Motivation Activities						392.00
1. Crew Inspections	2	4			2	43.00
2. Implementing Awards, Incentives, etc. Program						4.00
3. Reviewing and implementing Safety Programs						1.00
4. Implementing Joint Health and Safety Committee Activities and Programs						3.00
5. Developing and delivering safety related communications, bulletins, etc.						22.00
6. Providing Safety Related Feed Back to Employees						15.00
7. Job planning activities		24			24	150.00
8. Tail board conferences		24			37.5	154.00
9. Safety Supervision		78			70	613.00
Factor B - Skill Development and Training Activities:						297.00
1. Safety Training						29.00
2. Technical Training	150					230.00
3. Safety Meetings					17	38.00
4. Drills (emergency, safety, rescue practice and drills, etc.)						0.00
Factor C – New Tools and Equipment Design Methods and Activities:						26.00
1. New tool development activities						4.00
2. New methods and procedure development activities						21.00
3. Audits and Assessments						1.00
Factor D - Equipment Related Activities:						226.00
1. Equipment Inspections					10	111.00
2. Facilities Inspections						12.00
3. Personal Protective Equipment Inspections						27.00
4. Preventive Maintenance Activities						76.00
Total Hours For All Safety Activities/Week:	**152.00**	**130.00**	**0.00**	**0.00**	**160.50**	941.00

Table 1: An example of the data sheet used during data collection showing hourly data used in the study. Note, this is an example only (adapted from Haight et al. 2001, Iyer et al. 2004).

In Table **1** the safety intervention variables are located on the left side represented by Factors A, B, C, and D. The farthest column on the right *Totals* represents the sum of input levels of the various safety intervention factors. The input is the man hours allotted to each of the twenty safety intervention activities during a one week period. For further explanation of the data gathering process please refer to the methodology section.

Research Objectives

The goal of this research was to determine if incident rates could be accurately forecasted given a set of safety intervention inputs using artificial neural networks. In doing so, the researcher must ensure that the following objectives were adhered to during the study:

- Determine and develop the relationship between safety intervention activities and the incident rate.
- To develop a forecasting tool using Artificial Neural Networks that will allow an analyst to predict an incident rate based on the type and amount of safety intervention activities.

Hypothesis: Artificial Neural Networks is an accurate predictor of incident rates.

As this study involves a new approach in attempting to forecast incident rates, literature doesn't exist that defines what accuracy is. Therefore accuracy in this study is defined as:

- Absolute average percent error of less than 20%.
- Mean Absolute Deviation (MAD) of less than 1.0.
- Coefficient of Determination (R^2) greater than 0.50

Literature Review

The use of Neural Networks in a multi-faceted society is not a new concept but its use as a means of evaluating Health and Safety programs is a pioneering application. According to the literature, Artificial Intelligence has never been used in an attempt to correlate, analyze, or forecast safety intervention activity with the incident rate. In fact, except for Haight et al. (2001 & 2003) and Iyer et al. (2004 & 2005) research dealing with quantification of safety intervention activities and the incident rate and mathematical relationship modeling that exists today is minimal.

Guastello (1993) conducted research using regression analysis to relate the incident rates and intervention programs applied. He evaluated the programs as though the whole program was one intervention within each facility. So one input was compared to one output but the interactive effects between interventions were lost. He then realized that to determine the optimal level of interventions, it is imperative to know all the interventions that have an effect on the incident rate as well as the interactions amongst and between them.

Cleveland et al (1979) conducted comparative studies that distinguished successful from unsuccessful safety programs. These studies lay out specific practices of successful safety programs. This study does not quantify safety intervention hours with the incident rate nor does it address its correlation.

Behavior modification studies were performed similarly to the work of Cleveland and Guastello, as most of the safety programs were studied as a single intervention. A statistical analysis was performed by Frey and Ray (1999) that compared lost time injury and recordable rates with the mean behavioral safety index over a span of 30 months. Their single variable analysis did not show a lasting effect as their study did not distinguish what forms of safety intervention activities directly or indirectly affect the incident rate.

Geller performed behavior based safety research along with Kalsher et al. in (1989) that addressed the issue of incentives in reducing the incident rate and improving the program effectiveness, but the question of how effective was it still remains unanswered. DePasquale and Geller (1999) illustrated the factors in making a behavior based safety program successful, however they too did not quantify the input variables and output of a health and safety program.

Reinfort (1992) introduced a study that compared incident cost with safety intervention costs. He observed that the amount of money spent on safety was not the primary measure of the incident rate as opposed to the quality and type of safety intervention activities being adopted. His study opened the door to future research in addressing the issue of correlating safety intervention activities with the incident rate as he was the first to introduce the concept of a mix of activities.

The Haight et al. study in (2001) presented an analytical model that established a mathematical relationship between all intervention activities being implemented at the site and the incidents they were designed to prevent. The model provided a tool to develop a quantifiable design and to optimize a safety and health intervention program. The foundation of this study is based on the health and safety model shown previously in Figure 1 by Haight et al (2001) and Iyer et al. (2004).

Iyer et al. (2005 and 2004) developed a forecasting model and procedure to analyze, as well as, optimize a health and safety program by minimizing manpower input while concurrently minimizing incidents. He also produced a forecasting tool that would predict the incident rate given a set of safety intervention inputs. He determined that the carryover effect of an incident rate in a particular week had a statistically significant relationship with the safety intervention activity levels. Furthermore he developed forecasting models based on the results of his study using several statistical techniques such as transfer function modeling and regression analysis. Although the Iyer et al's. (2005) study is evidence that quantifying safety intervention activities with the incident rate is beneficial in terms of cost and reduction of losses, further research needs to be done to establish model reproducibility and its industry wide applicability.

Methodology

<u>Data Collection</u>

Data were collected on a weekly basis from September 2003 to February 2005. They were then entered in an Excel spreadsheet similar to the example displayed in Table 1.

In Table 1 the safety intervention variables adapted from Haight et al. (2001) are located on the left side represented by Factors A, B, C, and D. Factor A represents *Safety Awareness and Motivation Activities.* Factor B represents *Skill Development and Training Activities*. Factor C represents New *Tools and Equipment Design Methods and Activities* while Factor D represents *Equipment Related Activities*. The columns in the middle represent the hours spent on each intervention variable by the respective safety center within the company that participated in the research. The farthest column on the right labeled *Totals* represents the sum of inputs of the various safety intervention factors. The example in Table 1 shows only a fraction of cost centers reporting.

After gathering an adequate number of weeks' worth of data to proceed with a statistically significant study, the collection phase ended. The next step in the study involved organizing the data in a systematic way, which is explained in the next section so that they may be entered in an Artificial Neural Network (ANN).

An ANN is an information-processing prototype that mimics to an extent the way biological nervous systems, such as the brain, process information. According to the Defense Department's Advanced Research Projects Agency (DARPA) Neural Network Study (1988):

"... a neural network is a system composed of many simple processing elements operating in parallel whose function is determined by network structure, connection strengths, and the processing performed at computing elements or nodes. "

Data Organization For Use In ANN

In this phase, each of the twenty safety intervention inputs were summed on a weekly basis and placed in a separate Excel worksheet. The total number of weeks used in this study were 62 due to availability of data. Of those 62 weeks 37 were utilized to train the ANN and 25 weeks were used during the validation phase.

ANN does not have concrete rules regarding the number of weeks required for the validation and training phases, however guidelines do exist (Masters 1993). One of these guidelines suggest that the training set be representative of the entire population. Thus, the input data entered in the training set must encompass the range of incident rates displayed in the 62 weeks of data. Also, ANN does not have concrete rules defining network parameters, in fact only guidelines exist such as how many hidden layers to use and how many times to train the network. Furthermore ANN training capacity is partly based on the amount of patterns inputed, the fewer the data sets the less capable it becomes in formulating computational models based on the information given to it and vice versa. Therefore, there needed to be a correct mix of weeks inputted as the training phase so that there would still be enough weeks of data to produce a statistically significant measure of the ANN ability to help the analyst forecast incident rates.

That mix involved 25 weeks for validation and 37 weeks for training. ANN was trained with 10, 20 30, and 40 weeks prior to settling on 37. There was a tendency for the validation results to improve as the training set size increased. This was not always the case. So, it is at the discretion of the researcher to determine a suitable mix of training weeks to validation weeks, given the reasonableness of the results. The criteria used for this study to determine this "reason" were the Mean Square Error (MSE) and Mean Absolute Deviation (MAD). Note that, for every set of weeks not utilized for training, the remaining set of weeks were used for validation. This is evident as ANN requires that the training set be representative of the population so that when testing or validating takes place, outlier data should be nonexistant. An example would be having a trained ANN with incident rates ranging from 1-10 and then testing it with time spent on safety intervention activites that produced in real life incident rate within that range rather than having an incident rate of 15.

Throughout the 62 weeks of data gathered, not all of the work centers involved in this study were able to submit data on their safety intervention activities every week for various reasons. Therefore, a normalized set of data was determined and established.

It is important to note that an attempt was made to organize and input the data as a percent of available work hours similar to Iyer et al. (2004) but ANN was neither able to learn the pattern of these data nor forecast its outcome. This will be further discussed in the results and analysis section of this study.

Training and Forecasting

During the training phase of the ANN application, *supervised learning* took place. Whenever the term training is used throughout this study, it refers to the act of feeding ANN information and data and then running the program in order to enable it to learn and assimilate the information given to it. As for the term *supervised learning,* it refers to when the teacher or researcher gives ANN specific input patterns with the correct network output, in this case, the incident rate. So during the supervised learning phase, the researcher entered the safety intervention activity inputs with the corresponding output or incident rate. Once ANN was able to fully learn and assimilate the information, the researcher moved on to the validation phase, which is the forecasting stage of this research.

During the forecasting stage, the researcher performed validation. The term *validation learning* means that the network is not given any external indication as to what the correct responses should be nor whether the generated responses are right or wrong. It is simply projecting an output or forecasting the incident rate based on the safety intervention data given to it on a weekly basis i.e. the 40 inputs. The system, during validation looks back on the various input-output pairs that it learned during training and it learns by the environment, that is, by detecting regularities in the structure of input patterns. In this case, 25 weeks were used for validation.

ANN displays the results of the validation graphically and numerically by comparing the forecasted results to the actual results using the mean square error (MSE) formula. The MSE approach was chosen as it lies close to the center of normal distribution, thus, if errors are assumed to be normally distributed, minimizing the mean square error corresponds to other preferred optimizations. Furthermore the derivative of the mean square error can be easily computed relative to other performance measures. This signifies that when the optimization criterion is the mean square error, direct methods of optimizing performance can be achieved. To calculate the mean square error, sum the squared differences between the predicted output (ANN IR) versus the actual incident rate, then dividing by the number of components, in this case, weeks, that went into the sum. Equation 1 illustrates how the mean square error is calculated.

$MSE = \frac{1}{P}\sum_{p=1}^{P}\sum_{i=1}^{n}(d_{i,p} - a_{i,3})^2 \ldots$,	Eq. 1

Where $d_{i,\,p}$ equals desired output of output unit i for input pattern p and a_i equals observed output of output unit i. Also P equals total number of patterns in the data set, while n equals the number of output units.

Results may be improved by altering the architectural structure of the network by changing the amount of hidden layers, type of activation functions and the number of neurons utilized. This is an iterative process. Once enough iterations have taken place that lower the mean square error results without overtraining the network, the results are finalized and the forecasting stage of the research is concluded. Even though there is not an exact science to training the system, reducing the number of hidden neurons, helps the system avoid idiosyncrasies. Also, increasing the variety of the training set lessens the probability of overtraining the system. But it should not be forgotten that training usually starts with random initial weights and thus, there is no exact science of what constitutes adequate learning. Finally, statistical analysis is undertaken to support or refute the hypothesis of whether Artificial Neural Networks is an accurate predictor of incident rates.

Moving Average

This part of the research involved all the steps mentioned in the previous sections with one major difference, the inputs of one week were compared to the average incident rates for the following three weeks i.e. Week 1 inputs were compared to the incident rate for Week 1,2, and 3 since it is suggested that the effect from a health and safety program is neither instantaneous nor permanent. There were still 40 inputs entered into ANN per week with a corresponding output. The corresponding output was an average incident rate for 3 weeks, the week in which those 40 inputs originated from and the subsequent two weeks. The total number of weeks used in this part of the study was 58 weeks due to reduced availability of data. The training phase for this part of the research contained 35 weeks since the validation consisted of 23 weeks. The mix of training weeks to validation weeks used for this part of the study involved a similar approach to the week by week comparison detailed in the section *Data Organization for Use in ANN*. A 6-week moving average similar to Iyer et al. (2004) was not performed due to an inadequate availability of data, which would have meant loss of degrees of freedom, training strength as well as stastical significance of the results.

Results, Analysis and Discussion

Prelude to Results

In an attempt to reach optimal performance of ANN, modification to the network architecture needed to take place. Changing the architecture led to varying results and the network that produced the best results relative to other network runs was chosen. The network was chosen based on the lowest mean square error and the mean absolute deviation determined.

Figure 2 and Figure 3 display the architecture of the network with its corresponding output. The first number in brackets refers to the number of inputs in the input layer and the second number refers to the number of neurons associated with the activation function. Note the last number will always be one in this study as there is only one output function, in this case the incident rate. In Figure 5 **{**40,125,1} refer to 40 inputs in the input layer, 125 neurons in the hidden layer, and one output. The activation function in the hidden layer is logsig. ANN is Artificial Neural Networks forecasting capability and the Target or Actual is the incident rate for that particular week.

Figure 2: Example of an ANN validation output with its corresponding architecture [40,125,1], {'logsig','purelin'} MSE = 142.7 % R square = 0.05

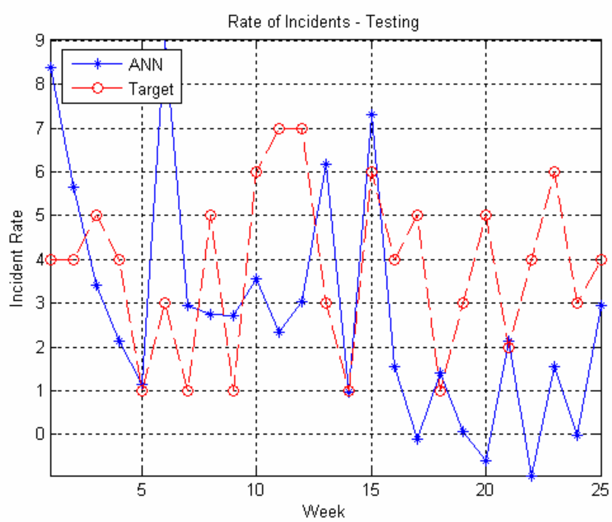

Figure 3: Example of an ANN validation output with its corresponding architecture [40, 125,15,1], {'logsig','tansig','purelin'} MSE = 76.8 % R square = 0.00

As mentioned earlier an effort was made to input the data as a percent of available work hours similar to Iyer et al. (2005) but as the following two graphs show ANNs were incapable of learning the data shown in Figure 4 nor of adequately predicting the results displayed in Figure 5.

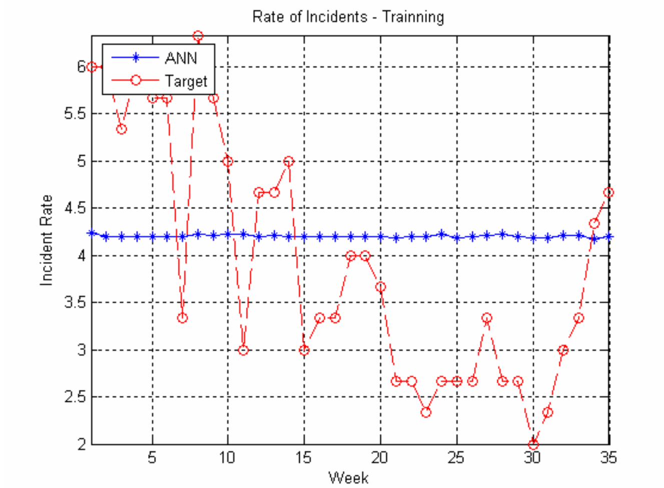

Figure 4: Training using percent of available man hours, MSE = 32.5 % R square = 0.00

As displayed by Figure 7 the network was not trained as ANN was unable to learn the specific input patterns and correlate it with the output given to it. This is illustrated by the points lying on a flat line rather than being remotely close to their output targets, in this case the incident rate.

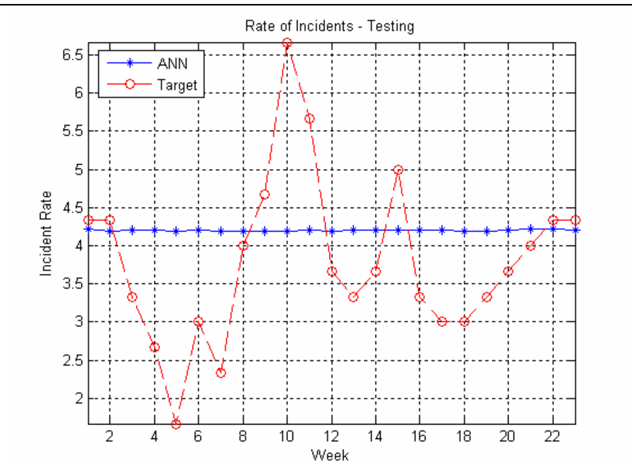

Figure 5: Validation using percent of available man hours MSE = 30.0 % R square 0.00

Furthermore lack of adequate training leads to poor validation results as illustrated in Figure 5. This phase of ANN is meaningless unless some form of training takes place. The output should attempt to correspond with what actually happened but in this case it is insignificant as the system was unable to be trained. After the attempt to use percent of available man hours failed, the data were inputted as total hours. The total hours represent the sum of hours for each cost center per safety intervention activity.

Forecasting Results and Analysis

After performing several runs and various ANN architectures the ANN system was trained and 25 weeks of safety intervention data were used for the validation phase. The following Figure 6 illustrates the finalized results.

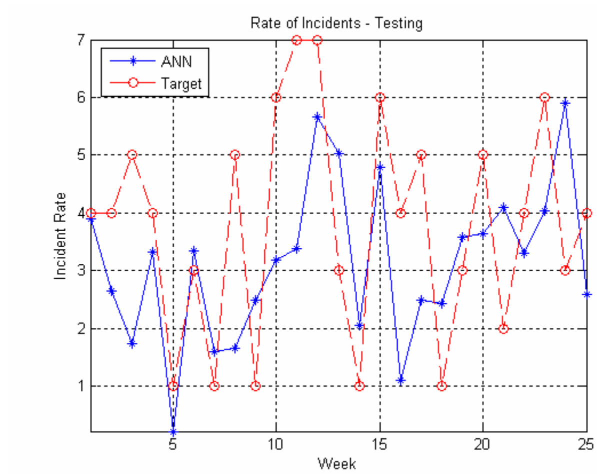

Figure 6: Forecasting Accuracy of Neural Networks MSE = 55.1% R square 0.13

The incident rates obtained from ANN were compared to the actual in a pair wise tabulation in Table 2**.** The pair wise comparison produced a residual result of a -0.63 indicating that on average the forecasted results tend to be lower than the actual incident rates. Also an average percent error of 55% indicates that the forecasted results were not close to the actual incident rates. Furthermore, the standard deviation revealed a relatively low statistical dispersion as the average standard deviation was 1.38.

Week	ANN	Actual	Residual	Absolute Percent Error
1	3.89	4.00	-0.11	2.70
2	2.64	4.00	-1.36	34.07
3	1.74	5.00	-3.26	65.17
4	3.32	4.00	-0.68	16.90
5	0.22	1.00	-0.78	78.34
6	3.35	3.00	0.35	11.62
7	1.60	1.00	0.60	60.30
8	1.65	5.00	-3.35	67.09
9	2.50	1.00	1.50	149.65
10	3.18	6.00	-2.82	46.94
11	3.39	7.00	-3.61	51.62
12	5.65	7.00	-1.35	19.22
13	5.03	3.00	2.03	67.79
14	2.05	1.00	1.05	104.96
15	4.80	5.00	-0.20	4.05
16	1.10	4.00	-2.90	72.58
17	2.49	5.00	-2.51	50.15
18	2.42	1.00	1.42	142.49
19	3.58	3.00	0.58	19.24
20	3.65	5.00	-1.35	27.07
21	4.09	2.00	2.09	104.48
22	3.31	4.00	-0.69	17.33
23	4.03	6.00	-1.97	32.88
24	5.90	3.00	2.90	96.66
25	2.59	4.00	-1.41	35.31
Average	**3.13**	**3.76**	**-0.63**	**55.14**

Table 2: Pair-Wise Comparison between ANN and Actual Incident Rates

To analyze the distribution of the data, a normality test was undertaken using the Anderson-Darling normality test. Both ANN and actual incident rates followed a normal distribution since both their respective P-values were greater than 0.05. The P value for both ANN and Actual were 0.867 and 0.096 respectively. As observed by Table 2 the average of ANN incident rates was 3.13 compared to the 3.76 which is the average of the Actual incident rate. That amounts to -16.9% error which indicates closeness among the means but further tests such as a Paired T-test need to be undertaken to verify this. Note that 55.14% is the average percent error of all 25 weeks while the 16.9% corresponds with the percent error of the means.

An F-test was performed to determine the ratio of two variances. If the two variances are not significantly different, their ratio will be close to 1. The resulting F-statistic was 0.551 and the associated P-value was 0.076. Since P was not less than 0.05, it can be concluded that there is no significant difference between the two standard deviations with a 95% confidence interval. This means that there is no significant variation between the population means of ANN and the actual incident rates.

After determining a lack of significant difference between the variances a Paired T-Test was performed using Minitab 14.0. A P-value of 0.103 indicates that there is not a statistically significant difference between the two means.

Also, a box plot of the analysis was performed using Minitab 14.0. Figure 7 illustrates the box plot of ANN and the actual incident rate. The box represents the middle 50% of the differences. The line through the box represents the median difference. The lines extending from the box represent the upper and lower 25% of the differences. The box plots of the data show the closeness in the means of the two data sets.

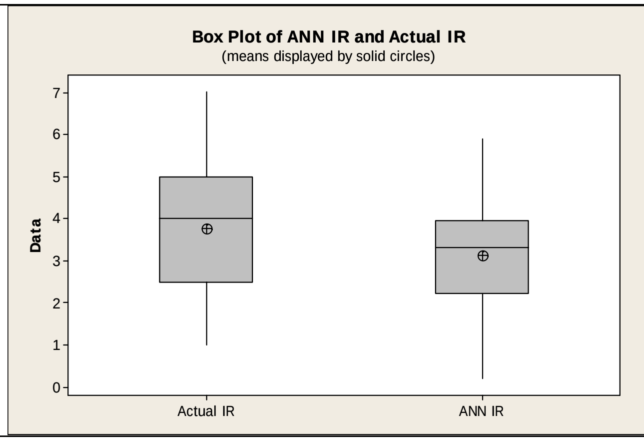

Figure 7: Box plot of ANN and Actual

Finally, to analyze the results obtained from the study, a Mean Absolute Deviation or MAD was determined as the measure of accuracy. Figure 8 displays the ANN forecasted incident rates versus actual incident rates. The plot indicates that ANN did not model accurately as the resultant R^2 was 0.13. Furthermore the points appear to be scattered rather than falling on a straight line. If they were to fall on a straight line, that would indicate that the ANN forecasted incident rates were linearly related. Also a Pearson correlation test was performed to see whether or not there is statistical significance in the R^2 value. The test produced a P-value of 0.075 which is greater than 0.05 which indicates zero statistical significance in its ability to correlate.

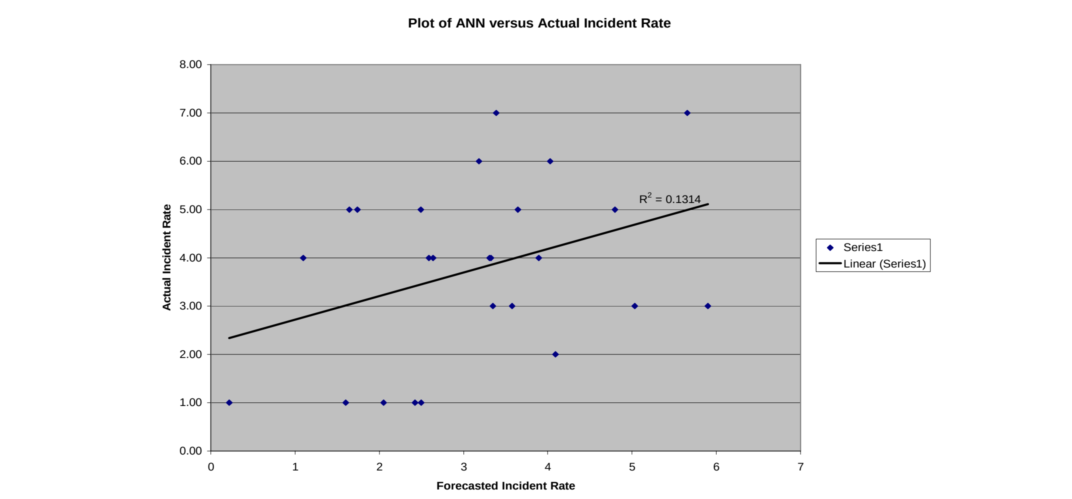

Figure 8: A plot comparing ANN with Actual incident rates

For the MAD, the closer the value is to 0 the more accurate one can claim that this prediction is. Eq. 2 displays the manner in which a Mean Absolute Deviation is obtained, where the sample size is N, the samples have values x_i, the mean is $\overline{x}$, and f_i is an absolute frequency. Furthermore, it shows the average deviation from the actual incident rates.

$$\text{M.A.D.} = \frac{1}{N}\sum_{i=1}^{N} f_i\,|x_i - \overline{x}|, \qquad \text{Eq. 2}$$

The MAD for the 25 weeks of forecasting was 1.63. The result of 1.63 incidents per week means that the predictions made by neural networks were on average within the range of +/- 1.63 incidents of the actual values.

Also, a normal average comparison was done to see whether or not simply taking the average incident rates of 25 weeks and projecting it every week produced better results than ANN. A summary of the results can be found in Table 3**.**

	Mean	**Average Percent Error**	**MAD**
ANN	3.13	55.14	1.63
Direct IR Average	3.76	75.86	1.49

Table 3: Normal Average Comparison

The results indicate that taking the average incident rate over the 25 weeks and comparing it to the incident rate of each of the 25 weeks yields a better mean absolute deviation of 1.49 as opposed to 1.63 for ANN. On the other hand, the absolute average percent error is far higher when using the normal average incident rate at 75.9% as opposed to ANN at 55.1%. The relative closeness between these results does not strengthen the hypothesis of ANN being an accurate forecasting tool.

A regression analysis was performed correlating the number of hours of safety intervention activities per week with the actual as well as the ANN forecasted incident rate. The resulting R square for ANN was 0.03 - poor correlation. But the regression analysis performed for actual incident rates with the number of hours of safety intervention activities per week also produced a poor R^2 value of 0.02. This indicates that the data itself has poor correlation with the incident rate which might suggest that further studies involving stronger correlation might yield better regression results when using ANN.

Moving Average

As stated earlier the moving average analysis involved comparing the inputs of one week to the average incident rates for the following three weeks. The methodology in attaining a desired ANN architecture for a moving average was similar to that of the safety intervention activities of one week versus the incident rate of that same week described in the previous sections. Figure 9 displays the optimized network results of the 23 weeks of validation involving a moving average.

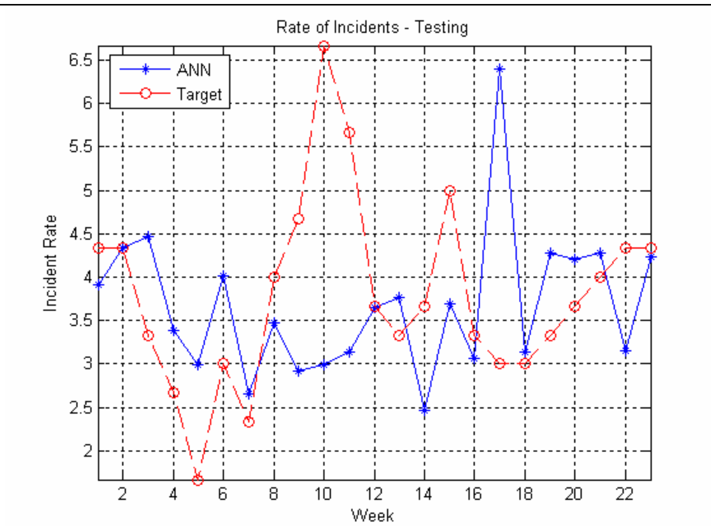

Figure 9: Forecasting Accuracy of ANN for Moving Average MSE = 27.2 % R square 0.01

A pair wise comparion was done to tabulate the results of the forecasting accuracy of ANN for the moving average. The pair wise comparison displayed in Table 4 produced a residual result of a -0.122 indicating that on average, the forecasted results tend to be lower than the actual incident rates. Also an average absolute percent error of 27.2% indicates that the forecasted results were "relatively" close to the actual incident rates. Furthermore, the standard was 0.83.

Week	ANN	Actual	Residual	Absolute Percent Error
1	3.905	4.333	-0.428	9.882
2	4.329	4.333	-0.004	0.095
3	4.467	3.333	1.134	34.010
4	3.385	2.667	0.718	26.919
5	2.994	1.667	1.327	79.616
6	4.014	3.000	1.014	33.803
7	2.647	2.333	0.313	13.426
8	3.471	4.000	-0.529	13.218
9	2.920	4.667	-1.747	37.426
10	2.992	6.667	-3.675	55.122
11	3.142	5.667	-2.525	44.555
12	3.644	3.667	-0.023	0.626
13	3.758	3.333	0.425	12.740
14	2.456	3.667	-1.210	33.013
15	3.688	5.000	-1.312	26.248
16	3.056	3.333	-0.277	8.314
17	6.392	3.000	3.392	113.067
18	3.139	3.000	0.139	4.637
19	4.273	3.333	0.939	28.184
20	4.200	3.667	0.533	14.537
21	4.279	4.000	0.279	6.983
22	3.148	4.333	-1.185	27.354
23	4.227	4.333	-0.106	2.454
Average	**3.675**	**3.797**	**-0.122**	**27.227**

Table 4: Pair-Wise Comparison between ANN & Actual Incident Rates for a Moving Average

Also, to find out the distribution of the data, a normality test was undertaken using the Anderson-Darling normality test. Both ANN and actual incident rates followed a normal distribution since both of their respective P-values were greater than 0.05.

The P value for both ANN and Actual were 0.082 and 0.367 respectively. Furthermore, as observed by Table 4 the average of ANN incident rates was 3.68 compared to the 3.80, which is the average of the actual incident rate. That amounts to a -3.2% error, which suggests closeness among the means but a Paired T-test needs to be performed to verify this.

Also just as previously done, an F-test was performed to determine the difference of two variances. If the two variances are not significantly different, their ratio will be close to 1. The resulting F-statistic was 0.590 and the associated P-value was 0.112. Since P was not less than 0.05, it can be concluded that there is no significant difference between the two standard deviations with a 95% confidence interval.

After determining a lack of significant difference between the variances, a Paired T-Test was performed using Minitab 14.0. A P-value of 0.687 indicates that there is not a statistically significant difference between the two means as it is far above the 0.05. Also, to go along with the Paired T-test a box plot of the analysis was performed using Minitab 14.0. Figure 10 illustrates the box plot for the moving average of ANN and the actual incident rate. Outliers are indicated by an asterisk.

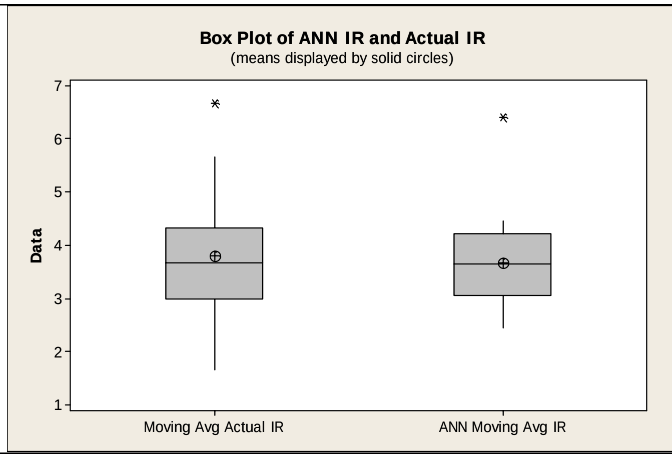

Figure 10: Box Plot of ANN and IR

Furthermore, a Mean Absolute Deviation or MAD was determined as the measure of accuracy. The MAD for the 23 weeks of forecasting using a moving average was 1.01 lower than 1.63 previously calculated for a nonmoving average. The result of 1.01 incidents per week means that the predictions made by neural networks were on average within the range of +/- 1.01 incidents of the actual values. A Pearson correlation test was performed to see whether there was statistical significance in the R^2 value. The test produced a P-value of 0.639 which is greater than 0.05. This indicates zero statistical significance in ANN's ability to correlate using a moving average.

A regression analysis was performed correlating the number of hours of safety intervention activities per week with the moving average incident rate. This was done for both actual as well as the ANN forecasted incident rate. The resulting R^2 square for ANN was 0.100 which is a bit higher than the previous regression analysis performed without using a moving average that yielded an R^2 of 0.03. However, a value of 0.10 indicates poor correlation. Also the moving average regression analysis performed for actual incident rates with the number of hours of safety intervention activities per week produced a poor R^2 value of 0.003.

Finally, a summary of the results comparing the performance of ANN using a moving average as opposed to not using one was tabulated. Table 5 displays these results.

	Residual	**Average Percent Error**	**MAD**	**R^2**
Direct	-0.63	55.14	1.63	0.03
Moving Average	-0.122	27.23	1.01	0.1

Table 5: Summary of ANN Results

The results indicate that moving average analysis performed better than a direct week-to-week comparison. This is indicated by a lower absolute average percent error of 27.23, a lower MAD of 1.01 and a higher R^2 of 0.1. This does not indicate however that ANN is an accurate forecasting tool it simply performs better with a moving average. However, the hypothesis stating ANN as an accurate predictor of incident rates must be rejected as none of the results obtained met the definition of accuracy for this study.

Conclusion and Future Work

After performing the analysis, the hypothesis that an artificial neural network is an accurate predictor of incident rates must be rejected. The low coefficient of determination of 0.10 and a relatively high average percent error indicates low statistical significance in accepting the hypothesis. It is important to note, that ANN performed better when utilizing a moving average to forecast the incident rate as opposed to a direct week-by-week comparison. This is evident by a lower moving average percent error of 27.2 %, a lower MAD of 1.01 and a higher correlation factor of 0.1.

Furthermore, even though this study provided us an example of artificial neural networks lack of statistically ability to correlate safety intervention measures with the incident rate, more research using a wider range of data over longer periods of time is needed to determine whether or not ANN can truly be used a forecasting tool of incident rates.

It is important to note that the results from this study are site specific and not industry wide applicable as the set of input variables and output used for training and validating originated from Hydro One. Some limitations exist with artificial neural networks such as the exclusion of outlier data and ANN inability to extrapolate the data. ANN effectiveness is as good as the data used to train the system. With that said, having an optimized set of input variables can lead to productive results.

In this study, the results of ANN illustrated the lack of significant statistical difference between the means and the variances as shown by a Paired t-test and F-test respectively. This does not mean that ANN has the potential to become an accurate predictor of incident rates but may prompt further studies and research. More research needs to be done by gathering more data and performing additional analysis such that it approaches normally distributed results and thus resulting in a lower mean absolute deviation and improving the results of regression analysis. Furthermore, different ways of optimizing the data or inputting in the ANN system might produce improved results.

As stated earlier, this study is site specific and not industry wide and thus gathering data from a different site may or may not produce similar results. One thing is for certain, if artificial neural networks can be determined to be an accurate forecasting tool, it will unlock doors that will enable companies, firms, and businesses to minimize incident rates and safety related costs by applying the appropriate mix of inputs. If artificial neural networks can show potential for this occurrence, incident driven lost time, medical impact and cost can be reduced.

Bibliography

1. Bailey, C. W. J. Am. Soc. Saf. Eng. 1997, 42(8), 33–5.
2. Bailey, C. W.; Petersen, D. J. Am. Soc. Saf. Eng. 1989, 34(2), 22–6.
3. Brigham, C. J. J. Am. Soc. Saf. Eng. 1981, 26(5), 18–24.
4. Chandler, B.; Huntebrinker, T. A. J. Am. Soc. Saf. Eng. 2003, 48(6), 33–40.
5. Cleveland, R.J.; Cohen, H.H.; Smith, M.J.; Cohen, A. Safety Program Practices in Record Holding Plants, National Institute for Occupational Safety and Health Report, 1979.
6. DARPA Neural Network Study (1988, AFCEA International Press, p. 60):
7. Demuth, H. Beale, M. Neural Network Toolbox 2000
8. DePasquale, J. P.; Geller, E. S. J. Saf. Res. 1999, 30(4), 237–49.
9. Fausett, L. (1994). *Fundamentals of Neural Networks.* New York: Prentice Hall.
10. Geller, E. S.; Rudd, J. R.; Kalsher, M. J.; Streff, F. M.; Lehman, G. R. J. Saf. Res. 1987, 18(1), 1–17.
11. Guastello, S. J. Saf. Sci. 1993, 16(3/4), 445–63.
12. Haight, J. M.; Thomas, R. E. Chem. Health Safe. 2003, 10(2), 21–5.
13. Haight, J. M.; Thomas, R. E.; Smith, L. A.; Bulfin, R. L.; Hopkins, B. L. J. Am. Soc. Saf. Eng. 2001a, 46(5), 38–44.
14. Haight, J. M.; Thomas, R. E.; Smith, L. A.; Bulfin, R. L.; Hopkins, B. L. J. Am. Soc. Saf. Eng. 2001b, 46(6), 33–7.
15. Haykin, S. (1994). *Neural Networks: A Comprehensive Foundation.* New York: Macmillan Publishing.
16. Ingalls Jr., T.S. J. Am. Soc. Saf. Eng. 1999, 44(12), 23–8.

17. Iyer, P. S., Haight, J. M., Del Castillo, Tink, B., & Hawkins, P. (2004). Intervention effectiveness research: Understanding and optimizing industrial safety programs. Chemical Health and Safety, 11(2), 9 –19.

18. Iyer, P. S., Haight, J. M., Del Castillo, Tink, B., & Hawkins, P. (2005). A research model—forecasting incident rates from optimized safety program intervention strategies. Journal of Safety Research, (36) 341-351.

19. James W. O'Sullivan, *Neural Nets: A Practical Primer*, AI In Finance, Spring, 1994.

20. Kalsher, M. J.; Geller, E. S.; Clark, S. W.; Lehman, G (1989).

21. National Safety Council website, www.nsc.org, 2006.

22. Petersen, D. J. Am. Soc. Saf. Eng. 1998, 43(10), 37–40.

23. Ray, P.S., Frey, A.J. Am. Soc. Saf. Eng 1999, 44(7) 25-8.

24. Rinefort, F. C. J. Am. Soc. Saf. Eng. 1992, 37(5), 42–5.

25. Rinefort, F. C. J. Am. Soc. Saf. Eng. 1977, 22(9), 36–48.

26. Smith, M. J.; Cohen, H. H.; Cohen, A.; Cleveland, R. J. J. Saf. Res. 1977, 10(1), 5–15.