INTERNATIONAL JOURNAL OF INTEGRATED ENGINEERING
ISSN: 2229-838X e-ISSN: 2600-7916

IJIE

Vol. 0 No. 0 (YEAR) 1-6
https://publisher.uthm.edu.my/ojs/index.php/ijie

Artificial Neural Network in Predicting Risk Exposure in Malaysian Shipyard Industry

David Chua Sing Ngie[1]*, Calvin Chin Yen Chih[1], Lim Soh Fong[2]

[1] *Department of Mechanical and Manufacturing Engineering, Faculty of Engineering, University Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, MALAYSIA*
[2] *Department of Chemical Engineering, Faculty of Engineering, University Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, MALAYSIA*

*Corresponding Author: csndavid@unimas.my
DOI: https://doi.org/10.30880/ijie.2024.00.00.000

Article Info

Received: Day Month Year
Accepted: Day Month Year
Available online: Day Month Year

Keywords

Risk exposure prediction, occupational accident, shipyard industry, risk management, supervised machine learning, artificial neural network

Abstract

Risk exposure prediction is an important task in risk management and control. The efficiency of occupational safety and health (OSH) risk prevention depends on the accuracy of predicting risk exposure. In this study, a multilayer perceptron training using the backpropagation algorithm neural network was developed and presented for risk exposure prediction in the Malaysian shipyard industry. The data was collected from industrial shipyards in Malaysia via related government agencies in order to train the model and evaluate its performance. The data was pre-processed to ensure homogeneity. The artificial neural network (ANN) model used 10 influencing factors as inputs for risk exposure prediction: gender, age, occupation, workplace factors, activities involved, nationality, working hours, educational level, years of employment, and working zone. Several network architectures were developed and the best model was selected for the risk exposure prediction of workers in the shipyard industry. Three evaluation metrics used for the selection of the best modal were mean square error (MSE), mean average percentage error (MAPE), and correlated of coefficient (R). The results showed that the ANN model, which has an accuracy performance of 90.2250% with a coefficient of correlation of 91.375%, can accurately estimate the risk exposure of workers in the shipyard industry. Sensitivity analysis also revealed that input factors, such as working hours and workplace factors, have significant effects on OSH risk prediction. Therefore, they should be taken seriously when dealing with the risk exposure in the Malaysian shipyard industry.

1. Introduction

Occupational safety and health (OSH) risk management is a common practice in many industries, such as construction, chemical production, electrical, and electronic production. In general, the OSH risk management framework that is practiced in the industries incorporates potential hazard identification, risk assessment, and risk control [1]. Despite the fact that many studies were carried out with regards to OSH risk management practices in the Malaysian shipyard industry in the past, there is still a lot of room for improvement in the area.

A previous survey about the Health, Safety, and Environmental Management System (HSEMS) of shipyard operations showed that most shipyards have HSEMS in place, but some are incomplete. The low priority assigned to Health, Safety, and Environment (HSE) resulted in nearly 10% of shipyards do not have a clear HSE policy or management system. Consequently, HSE management cannot be fully emphasized due to having less OSH awareness culture [2].

Furthermore, many activities in fabrication works are contracted out, and subcontractors are split into independent organizations that collaborate throughout the fabrication process at shipyards [3]. Therefore, there is no direct contact or link between shipyard management and subcontractors to promote OSH awareness and enforce basic risk management, including hazard identification, risk assessment, and risk control. This directly increases the OSH risk exposure of shipyard workers. Accidents are also bound to happen without the implementation of strict OSH management in shipyards. Among the examples of accidents that commonly occur at shipyards include workers suffering from lacerations, crushes, avulsion, fractures, amputations, and being caught in dangerous occurrences or between machinery [4].

Usually, OSH practitioners use their site audit investigation and experience to identify significant hazards from on-site incidents in order for improvements to be made accordingly. However, previous research demonstrated that OSH practitioners were having difficulties in risk assessment and prevention decision-making due to many factors contributing to the accidents [5]. In other words, the different judgments and experience levels of OSH practitioners would affect the accuracy of the risk assessment analysis and the priority of risk control decision-making.

Nevertheless, previous research also found that the risk control decision-making performed in the industry was not based on analytical findings from any collected data on the accidents. This highlights the analytical limitations in understanding the trend of the factors contributing to the accidents [6]. The data collected was not fully utilized to understand the limitations of the risk management framework currently practiced by shipyard companies. It also showed that most shipyard companies were sticking with their current standard operating procedure (SOP) without further revision. Their SOP documents were also found to have neither revision nor improvement as they were least enforced and partially practiced. This suggests that accidents are prone to happen in the shipyard industry due to poor management.

Therefore, this paper presents an innovative method to develop a risk prediction model to predict the risk exposure of workers in the Malaysian shipyard industry. So that OSH incidents can be prevented earlier and further improvements can be carried out by targeting the incidents that contributed to the hazards specifically. The method involves the use of a supervised data machine learning technique to train a model through optimization and classification via appropriate algorithm network architecture. The accuracy performance (AP) and sensitivity analyses are then applied to determine the risk prediction model's performance by analyzing the relationship between the input factors (independent variables) and the output OSH risk prediction (dependent variable) [7]. The proposed risk prediction model with artificial neural network (ANN), aims to assist OSH practitioners in estimating the OSH risk exposure of shipyard industry workers, ultimately contributing toward better risk management in the shipyard industry.

2. Literature Review

Predictive models for workplace accidents can generally be divided into two categories: machine learning (ML)-based and statistical learning-based. Supervised ML is one of the widely used intelligent techniques that are configured for a specific application, such as pattern recognition, function approximation, or data classification, through a learning process for prediction applications [8, 37].

In the medical field, a study [9] was conducted on the use and assessment of ANN model performance for the risk prediction of heart disease. It involved the use of an ANN architecture and Levenberg–Marquardt backpropagation algorithm for the network with training, validation, and testing from a dataset of 297 samples. Different training algorithms were examined in the research. Amongst them were scaled conjugate gradient backpropagation, BFGS quasi-Newton backpropagation, and resilient backpropagation. The data was pre-processed by eliminating some samples due to incomplete data. 13 dataset descriptions were converted into coding as an input to the ANN model. The output was a number that was linked to the risk of heart disease based on its ranking. A hyperbolic tangent sigmoid activation function and a log-sigmoid activation function were used for the hidden layer and the output layer respectively. They achieved a remarkable 100% accuracy performance with thirty hidden neuron layers for the prediction of heart disease.

Apart from the medical field, ANN has also been applied to construction project cost estimation. Researchers [10] also used the historical data of 169 completed projects to build an ANN model for predicting the total structural cost of building projects in the Gaza Strip. The data was randomly divided into 69% for training, 16% for validating the performance, and 15% for a completely independent test of network generalization. A total of 11 input factor parameters were used: type of project, area of a typical floor, number of floors, type of foundation, type of slab, number of elevators, type of external finishing, type of air-conditioning, type of tilling, type of electricity, and type of sanitary. Finally, the desired parameter (output) was specified, which was the total cost of the project. They used multilayer perceptron (MLP) network with architecture of one layer for each input, hidden and output layer. There were eleven neurons in the input layer and twenty-two in the hidden layer. One output layer with one output neuron. Hyperbolic tangent (tanh) was applied as a transfer function and momentum learning rate was used to perform the backpropagation algorithm. They achieved an accuracy performance (*AP*) of 94% with a mean absolute percentage error (MAPE) of 6% and a correlation coefficient (*R*) of 0.995.

Another related study was also carried out to explore how the ANN technique can be used to estimate productivity in construction project management [11]. The quality of construction management depends on the accuracy of estimations through construction labour productivity. In the study, an ANN model network was developed to estimate the labor productivity of marble finishing works for floors using the historical data of 150 data projects, encompassing 60% for training, 25% for validating the performance, and the remaining 15% for computation. The data used in the study was collected from residential, commercial, and educational projects in various locations across Iraq to train the model and evaluate its performance. The transfer function used was a hyperbolic tangent sigmoid function. 10 influencing factors for productivity forecasting were used in their study: age, experience, number of labour, height of floor, size of marble tiles, security conditions, health status of the work team, weather condition, site condition, and availability of construction materials. A model was then built to predict the productivity of marble finishing work for floors. Their model could predict the productivity for finishing works with a high degree of accuracy, with the values of *R*, average *AP*, and MAPE being 89.55%, 90.9%, and 9.1%, respectively. They proved that the backpropagation neural network (BPNN) model also performed exceptionally well in modeling input-output relationships.

Another study adopted the use of artificial intelligence (AI) techniques in cash flow forecasting for risk forecasting, ranging from statistical, mathematical, and simulation forecasting using the ANN method [12]. The study attempted to model the variation between the predicted and actual cost flows due to inherent risks in construction. The data was obtained through a questionnaire survey and empirical data collection. This dataset was collected from the actual cost flows through completed construction projects. A neural network was later employed to develop the cost flow risk assessment model using the backpropagation algorithm. The developed model was tested on 20 new projects with satisfactory predictions of variations between the predicted and actual cost flows at 30%, 50%, 70%, and 100% completion stages. On the other hand, 40 out of the 50 datasets collected from the questionnaire survey were used to develop a cost flow risk assessment model. The remaining 10 and another 10 datasets obtained from a construction company were used for the testing and validation of the model. The scores obtained from the questionnaire were associated with 11 identified significant risk factors which were used as input for the neural network. They attempted with 11 input neurons and 4 output neurons as a starting point for their network architecture. After a series of trials and errors, the network was found to be stabilized with 12 hidden nodes and a sigmoid transfer function. The results showed an error square (R^2) values of 0.626 (30%), 0.748 (50%), 0.653 (70%), and 0.767 (100%) with actual cost flow at 30%, 50%, 70%, and 100% completion stages, respectively. Based on the mean absolute deviation measured for 20 new projects, the model was able to predict the cash flow that was closest to the actual result.

Apart from the applications mentioned above, the application of ANN ML has also been proven in various fields, such as die cast's shrinkage prediction in engineering [38], analysis of stream flow trend and rain flow trend under environmental prospects [43], construction's capital forecasting [42], accident severity prediction [44], and it yielded very useful results [6, 13]. However, a thorough review of the literature showed that occupational accident risk analysis has only demonstrated the use of ANN in terms of its predictive and explanatory capacities, with limited details on the optimization of factors that cause workplace accidents [14, 15, 36, 41]. Hence, this paper investigates the use of the ANN technique with various network architectures to develop a good risk prediction model for estimating the OSH risk exposure of shipyard industry workers in Malaysian shipyard industry.

3. Methodology

3.1 Data Mining and Data Pre-processing

The data on 756 OSH accident cases within Malaysia was collected from several government agencies, including the Department of Occupational Safety and Health, Social Security Organization, and the Department of Manpower.

Data pre-processing was conducted to reduce the size of the dataset in order to obtain homogeneous subgroups from a complicated dataset consisting of 300 OSH accident cases recorded in the shipyard industry. This improved the dataset quality and prediction accuracy because a complex dataset would lead to serious queries while finding a meaningful relationship between elements of the dataset [16].

3.2 Data Input Attributes and Data Encoding

In this study, the general attributes of HSE contributing factors obtained from the data collection were categorized in textual terms and placed into five groups representing the main attributes: gender, age, occupation, workplace factors, and activities conducted [17]. The main attribute data was then elucidated in detail to show how the OSH accidents occur. The details (nationality, working hours, educational level, years of employment, and working zone) were also used to analyze the input-output paired HSE factors. These contributing input factors were identified as independent variables, whereas the severity condition was identified as the dependent variable (see Table 1).

Table 1 *Description of the Dependent Variable and Independent Variables*

Dependent Variable	Independent Variables
Severity of Accident	Gender Age Working Hours Years of Employment Occupation Nationality Education Level Workplace Factor Working Zone Activities Involved

Finally, 46 attributes from 10 categories of input factors were converted into coding for the computational progress, as presented in Table 2. However, an ordinal scale of 1 to 5 for the risk prediction was assigned to the targets to represent the output encoding.

Table 2 *Encoding Data for the Input and Output Attributes*

Attributes	Description	Encoding Data	Descriptive Statistical Data	Data Type
Gender	Gender of workers	Male: 1; Female: 2	Median: 1; Min: 1; Max: 2	Binary
Age	Age in years	<16 years: 1; 16–18 years: 2; >18 years: 3	Median: 3; Min: 1; Max: 3	Categorical
Working Hours	Working hours of the workers	8 hours or overtime working days: 1; Non-working hours: 2	Median: 1; Min: 1; Max: 2	Binary
Years of Employment	Working experience of the shipyard company	<3 years: 1; 3–5 years: 2; 6–8 years: 3; 9–10 years: 4; >10 years: 5	Median: 2; Min: 1; Max: 5	Categorical
Occupation	Position held in the shipyard company	General worker: 1; Operator: 2; Technician: 3; Contractor: 4; Visitor: 5	Median: 1; Min: 1; Max: 5	Categorical
Nationality	Nationality of workers	Malaysian workers: 1; Foreign workers: 2	Median: 2; Min: 1; Max: 2	Binary
Education Level	Educational level indicates the understanding of communication, safety procedures, self-risk assessment, and rational mind set for not conducting any miscommunication or unsafe act	Primary School: 1; Secondary School: 2; Certificate/Diploma: 3; Degree and above: 4; Non-educational background: 5	Median: 1; Min: 1; Max: 4	Categorical
Workplace	The working environment	Physical hazard: 1;	Median: 1;	Categorical

Factor	that leads to the incident.	Chemical hazard: 2; Biological hazard: 3; Ergonomic hazard: 4; Psychosocial hazard: 5; Environment hazard: 6	Min: 1; Max: 5	
Working Zone	Working zone at shipyards	Dry dockyard: 1; Wet dockyard: 2; Machinery workshop: 3; Inside building/others: 4; Welding/hot works: 1; Working at height: 2; Lifting & mobilization works: 3; Equipment installation works: 4; Use of hand tools/equipment: 5; Electrical installation works: 6; Sandblasting and painting works: 7; Repair/maintenance works: 8; Working in confined space: 9; Sea launching works: 10; Site supervision/visiting: 11; Other activities: 12	Median: 1; Min: 1; Max: 4	Categorical
Severity of Accident	OSH risk prediction	Negligible injury: 1; Non-permanent injury: 2; Permanent disability: 3; Fatality: 4; Catastrophic: 5	Min: 1; Max: 5	Categorical

3.3 Development of ANN Model for OSH Risk Prediction

The encoding data was entered into MATLAB 2017 to establish the network modeling (see Table 2). First, this research proposed that good modeling required at least 250 cases to develop a model based on a rule of thumb of multiplying all the neurons (10 × 25 × 1) [18, 19]. Second, this research proposed that the number of hidden layers should not exceed 2.5 times the input size as this would induce network instability [17, 20]. Therefore, the network model began training data with 19–31 neuron nodes with a single hidden layer using the nntool command [21]. Third, in terms of data extraction, this research extracted general data into seven modeling networks where extraction was performed based on the rule of thumb, namely 70% for training, 15% for validation, and 15% for testing. Finally, this research employed gradient descent with momentum weight and bias (learngdm) as the learning function, scaled conjugate gradient backpropagation (trainsg) as the transfer function, and Levenberg–Marquardt backpropagation (trainlm) [22] as the training function in the model development. Fig. 1 presents the proposed ANN modeling architecture for risk prediction.

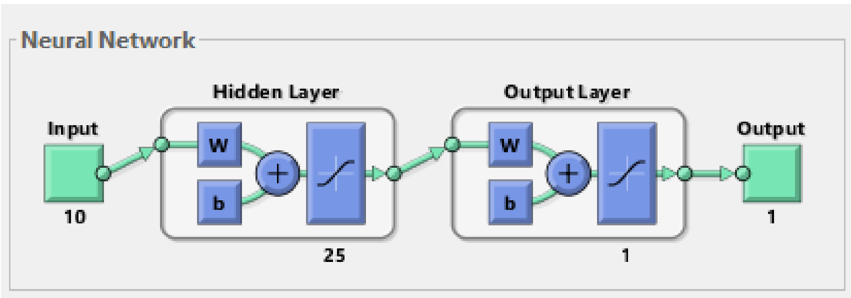

Fig. 1 *ANN Modeling Architecture Proposed for OSH Risk Prediction*

3.4 Selecting the Best Network Model

Next, various modeling networks were tested and trained, and the training would stop once the error in the testing set increased. The testing results were compared to meet the following requirements and criteria:

i. *R* computed in Equation 1 and R^2 should be greater than 0.5;
ii. A minimum of 50% of cases should be predicted with near-zero errors;

iii. The MSE calculated in Equation 2 should be less than 1.0.

The papers [23, 24, 25, 40] were referred to for the best modeling network requirements and criteria. After selecting the best modeling network, the MAPE values of each target were computed using Equation 3. The most acceptable MAPE applied in this research was less than 30% [17, 26].

Correlated Coefficient,

$$R = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{1}$$

where x_i denoted the input variables; $\bar{x}$ denoted the mean input; y_i denoted the actual target; and $\bar{y}$ denoted the mean target.

Mean Square Error,

$$MSE = \frac{1}{n}\sum_{i-1}^{n}(ti' - ti)^2 \tag{2}$$

Mean Average Percentage Error,

$$M_a = \frac{1}{n}\sum_{i=1}^{n}\frac{|ti' - ti|}{ti} \tag{3}$$

where t_i denoted the actual target; t_i' denoted the predicted one; and M_a denoted the incident target.

3.5 Sensitivity

It is well known that ANNs are powerful function approximators. However, there is a drawback to the method due to its inability to explain the obtained results, which is often called a black-box solution. With the best ANN modeling network chosen, sensitivity analysis was conducted in this study to evaluate the influence of each input parameter on the output variable. Journalist Gevrey advised on using the perturb and weights algorithms to perform the sensitivity analysis based on the relative importance of each input factor [27]. Both of these algorithms can classify the variables in the order of importance of their contribution to the output. However, the results observed using each method are not always the same. In this case, opinion from an ecologist is always helpful.

Sensitivity analysis is a method to analyse the impact of an independent variable (input) on a particular dependent variable (output) by percentage deviation. Neural network learning was disabled to maintain the network weights throughout the process of performing the perturb algorithm. The fundamental idea was to perturb the network's inputs and record the associated change in the output as a percentage deviation [28, 29, 30]. While the other inputs were fixed at their respective medians, the first input fluctuated between its median value (for binary or categorical variables) plus (or minus) upper and lower value limits. The absolute percentage change above and below the output variable's median was then calculated and recorded as the network output. Each input variable went through the same process until they were all done. [31].

4. Result

4.1 Effects on Neuron Nodes and Hidden Layers on the Performance of Neural Network

This paper examined the number of neuron nodes by increasing it by two and reinitializing the network weights. Then, the training process started until the optimum number of neuron nodes and hidden layers were reached corresponding to the observation of the training error, testing error, and regression square. Subsequently, the number of hidden layers for the optimum nodes was achieved. Fig. 2 and Fig. 3 demonstrate that the training and testing errors varied between 19 and 31 neuron nodes and a single hidden layer (7 networks). Network 4 showed the max training data with a correlation coefficient (R) of 0.91375. This can be justified by the fact that

70% of the total cases (210 cases) were used for training and the result indicated the minimum training error performance, 0.08625 (1–training R), compared to other networks. Furthermore, 15% of the cases (45 cases) were used for testing and network 4 showed the minimum testing error performance of 0.38855, which was the lowest error and almost equal to zero error compared to other networks. Therefore, 25 neuron nodes with a single layer were considered optimal with the least error found.

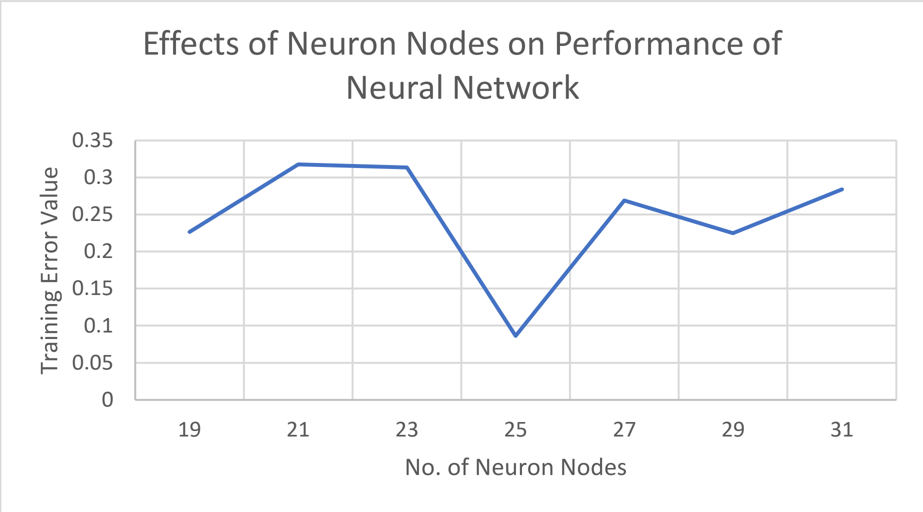

Fig. 2 *Training Error Performance Based on Different Number of Neuron Nodes*

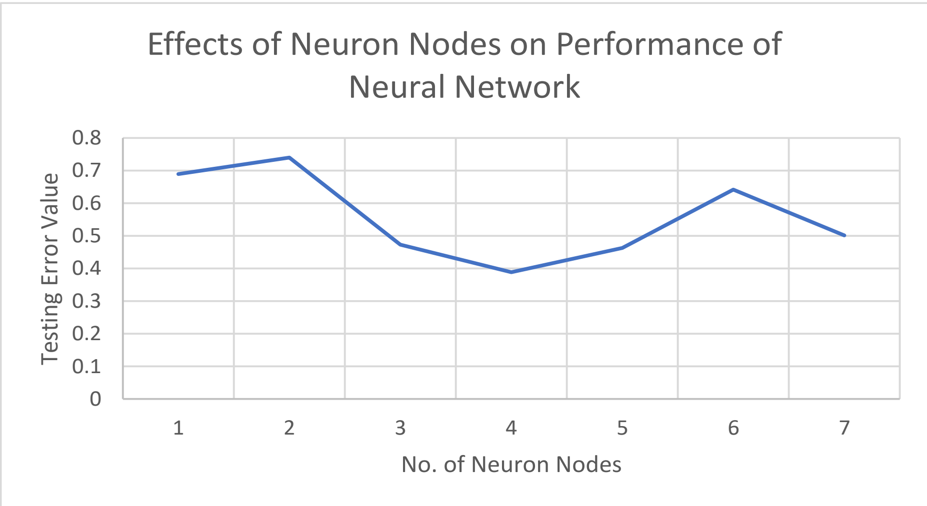

Fig. 3 *Testing Error Performance Based on Different Numbers of Neuron Nodes*

4.2 AP Performance

To select the best neural network, this study evaluated the performance of the ANN model in terms of *AP* and sensitivity analysis. From the seven networks developed (neuron nodes started from 19 to 31), this research discussed the least error calculated in the *AP* to select the best ANN modeling from different algorithms, architectures, number of neurons, types of training, and transfer functions [21].

Table 3 presents a comparison of the seven networks trained using the given criteria. This research found that regression analysis for network 4 can be conducted to determine the relationship between the input-output paired wise, as presented in Fig. 4. The linear regression result was illustrated graphically in the training data with R = 0.91375. The R^2 was 0.8349, which exceeded 0.5 and was higher than the R^2 of the other networks. This explains why during the training phase, network 4 showed a good linear correlation between the actual input factors and the output risk prediction. Furthermore, network 4 showed an MSE of 0.1387 for training, which was less than 1.0, indicating that it had the least MSE compared to the other networks. Subsequently, a minimum of 50% of cases were predicted for network 4 with the least errors, 0.1491, thus meeting the required criteria of almost-zero errors. This indicates that network 4 has the optimum architecture and algorithm. Therefore, network 4 was selected as the best modeling network to meet the prescribed criteria. Regarding the MAPE output in accidents, network 4 had an *AP* of 9.7750%, which was less than 30%. It can also be expressed in another form of *AP*, which is defined as (100−MAPE)% [32]. Therefore, network 4 can achieve an *AP* of 90.2250%. This explains that network 4 has high prediction accuracy. Table 4 summarizes the ANN modeling for OSH risk prediction.

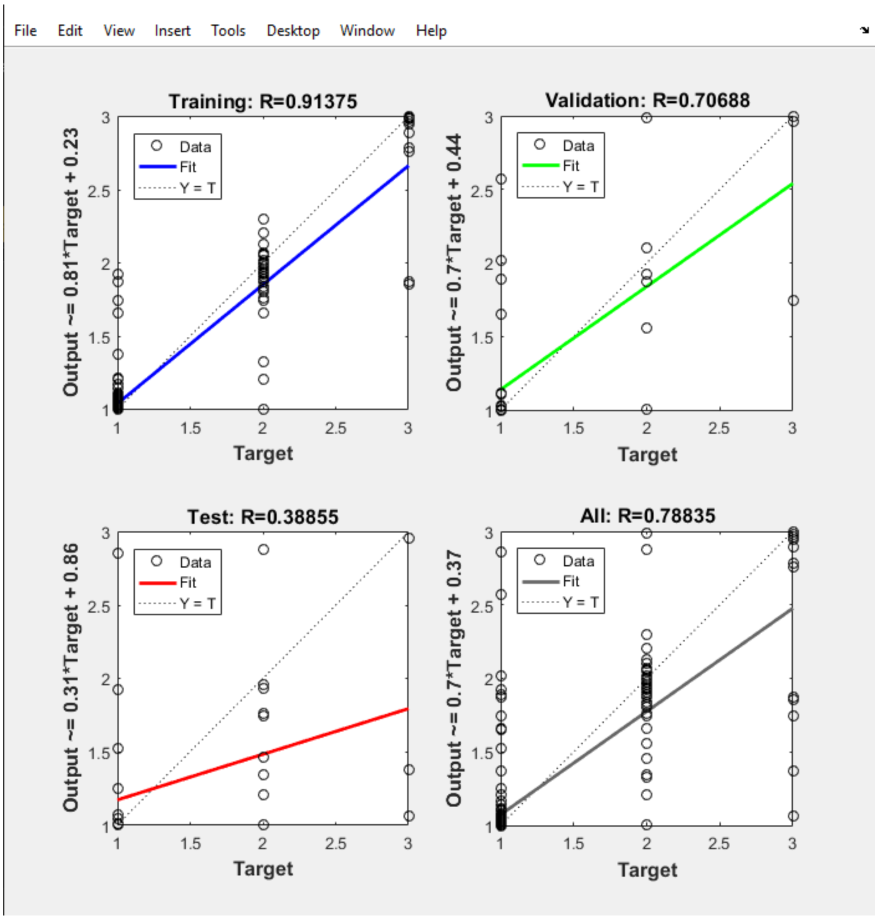

Fig. 4 *Result of the Training, Validation and Testing for Network Modeling No. 4*

Table 3 *Result of ANN Modeling for OSH Risk Prediction*

ID	Number of Neurons Used in the Training	Average 50% Sample Prediction Value Error	Error Square Training, R^2 (>0.5)	Best Mean Square Error, MSE (<1)	Mean Average Percentage Error, MAPE	Correlated of Coefficient, *R*
Network 1	19	0.2755	0.5983	0.1676	18.2435	0.7735
Network 2	21	0.2561	0.4656	0.1837	18.5641	0.6824
Network 3	23	0.2970	0.4711	0.2663	17.9784	0.6863
Network 4	25	0.1491	0.8349	0.1837	9.7750	0.9137
Network 5	27	0.2737	0.5344	0.1993	19.9539	0.7310
Network 6	29	0.2629	0.6007	0.1549	18.9556	0.7751
Network 7	31	0.3790	0.5126	0.2182	25.8134	0.7159

Table 4 *Summary of ANN Modeling for OSH Risk Prediction*

Description	ANN Model No.4
MAPE	9.7750%
AP%	90.2250%
R	91.375%
R^2	0.8349

To verify the *AP*, the testing performance of 15% (45 cases) was randomly gathered from the actual and predicted data risk exposure of workers obtained through network 4. The output also included an error percentage in prediction. From the comparison between the actual and predicted severities, any percentage of error that exceeded 9.7750% indicated that the prediction performance was poor. The remaining prediction performance was good. Fig. 5 presents the comparison of randomly picked data, demonstrating that the prediction value is almost the same as the actual severity. The randomly picked data for testing performance indicated a linear regression relationship between the actual data and predicted data, as presented in Fig. 6. Hence, the prediction result demonstrated a good estimation with a testing prediction performance of 90.2250%.

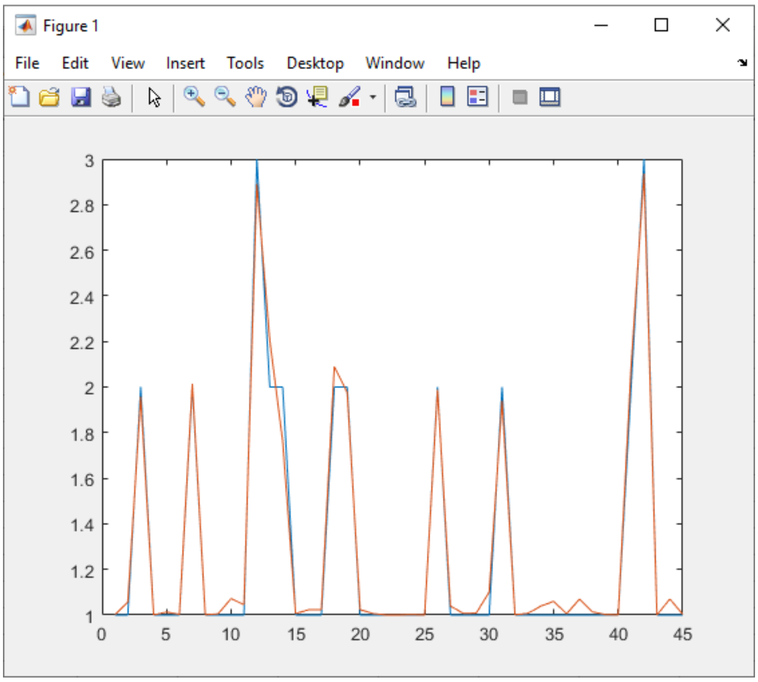

Fig. 5 *Comparison of Actual and Predicted Data*

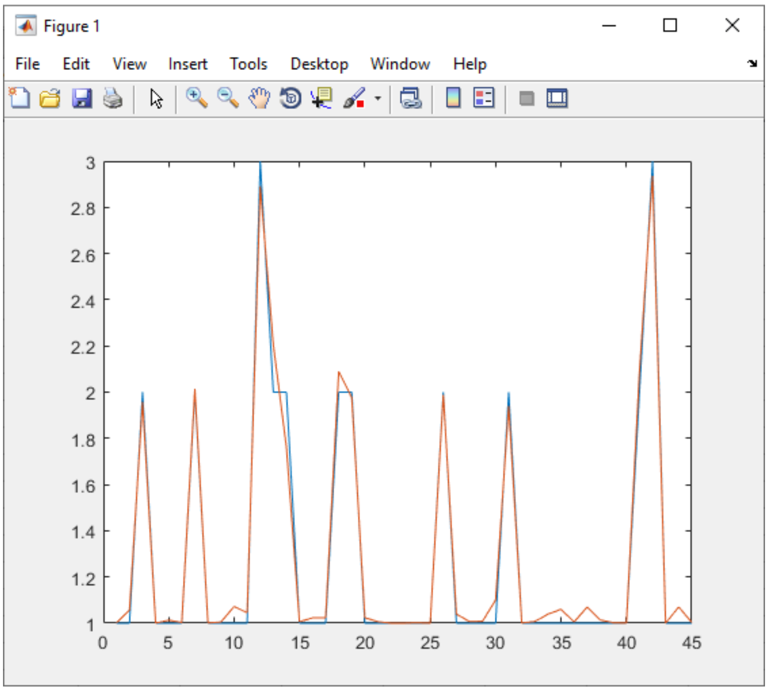

Fig. 6 *Linear Regression Analysis Between Actual and Predicted Data*

INTERNATIONAL JOURNAL OF INTEGRATED ENGINEERING
ISSN: 2229-838X e-ISSN: 2600-7916

IJIE

Vol. 0 No. 0 (YEAR) 1-6
https://publisher.uthm.edu.my/ojs/index.php/ijie

Table 5 *Variation of Input (Sensitivity) that Affects Variation of Output*

Output	Gender	Age	Working Hours	Years of Employment	Occupation	Nationality	Education Level	Workplace Factor	Working Zone	Activities Involved
Input Variables Sensitivity	2.0372	2.0372	2.0372	2.0372	2.0372	2.0372	2.0372	2.0372	2.0372	2.0372
Minimum	2.0372	1.5877	2.0372	2.4771	2.0372	2.0194	2.0372	2.0372	2.0372	1.5877
Median	2.0372	2.0372	2.0372	2.0372	2.0372	2.0372	2.0372	2.0372	2.0372	2.0372
Maximum	1.9189	2.0372	1.1224	2.2337	2.4798	2.0372	2.0689	1.3499	1.7692	1.5378

Table 6 *Summary of Output Variables Due to Input Variables*

Item	Gender	Age	Working Hours	Years of Employment	Occupation	Nationality	Education Level	Workplace Factor	Working Zone	Activities Involved
Output Variables Due to Input Variables	2.0372	1.5877	2.0372	2.4771	2.0372	2.0194	2.0372	2.0372	2.0372	1.5877
	2.0372	2.0372	2.0372	2.0372	2.0372	2.0372	2.0372	2.0372	2.0372	2.0372
	1.9189	2.0372	1.1224	2.2337	2.4798	2.0372	2.0689	1.3499	1.7692	1.5378

Table 7 *Output Variance in Percentage*

Items	Gender	Age	Working Hours	Years of Employment	Occupation	Nationality	Education Level	Workplace Factor	Working Zone	Activities Involved
Output Variance	4.85E-05	0.4494	4.85E-05	0.4399	4.85E-05	0.0177	4.85E-05	4.85E-05	4.85E-05	0.4494
	0	0	0	0	0	0	0	0	0	0
	0.1182	4.85E-05	0.9147	0.1965	0.4426	4.85E-05	0.0317	0.6872	0.2679	0.4993
Range	0.1181	0.4494	0.9146	0.2433	0.4426	0.0176	0.03169	0.6871	0.2678	0.0499
Variance (%)	3.6668	13.9462	28.3834	7.5523	13.7344	0.5490	0.9834	21.3229	8.3116	1.5495
Rank	7	3	1	6	4	10	9	2	5	8

INTERNATIONAL JOURNAL OF INTEGRATED ENGINEERING
ISSN: 2229-838X e-ISSN: 2600-7916

Vol. 0 No. 0 (YEAR) 1-6
https://publisher.uthm.edu.my/ojs/index.php/ijie

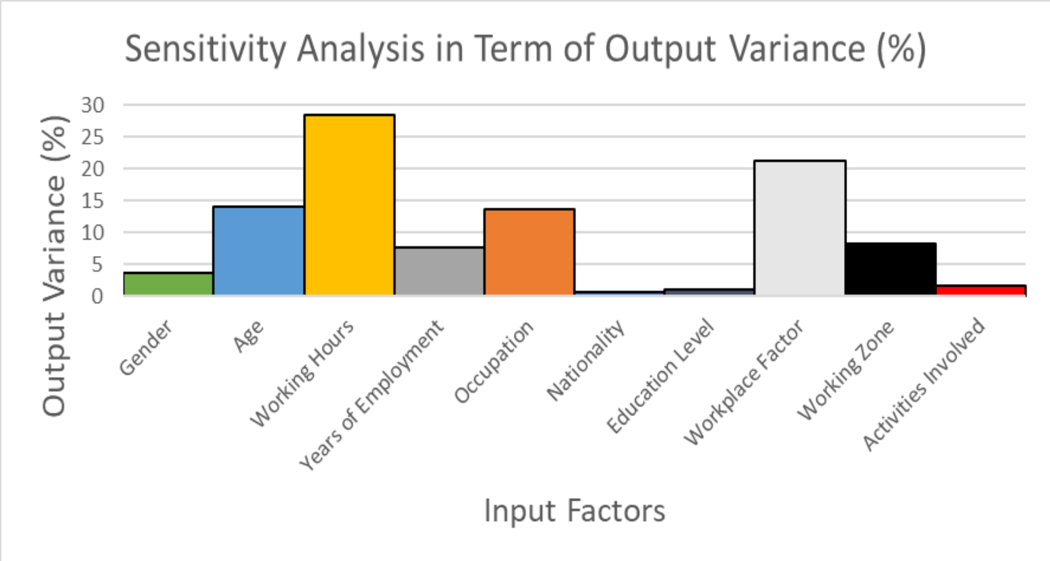

Fig. 7 *Sensitivity Analysis in Term of Output Variance in Percentage [31]*

4.3 Sensitivity Analysis

Finding the important variables affecting the perturbed levels of risk exposure severity in the shipyard industry was the aim of the sensitivity study. In order to determine the most significant contributors among the input variables at their median value, minimum limits, and maximum limits, a sensitivity analysis was carried out (see Table 5). Table 6 presents the output variables. Table 7 summarizes the variation of output based on the relative importance of each input factor, while Fig. 7 shows the relative importance of each input factor, from the most important to the least significant.

According to the sensitivity analysis result in Figure 7, nationality (0.5490%), education (0.9834%), activity involved (1.5495%), and gender (3.6668%) are the least significant predictors of increased levels of injury. It denotes that more serious injuries do not depend on whether the workers are local or foreign, educated or not, involved in any activities, or their gender. However, the finding should prompt a more detailed investigation into the influence of workers on risk exposure levels in shipyards. Likewise, researchers discovered that although men were more inclined to be in fatal accidents, women were more likely to sustain more serious injuries [33].

Another interesting observation is that input variables, such as years of employment (7.5523), working zone (8.3116), occupation (13.7344), and age (13.9462), can have an influence over risk prediction. Journalists Dissanayake and Lu obtained similar results when using a linear regression model to study age, location, and personal factors influencing the severity of injuries [34]. Their results demonstrated that location (working zone) with hazards has a higher probability of causing more severe injuries, as indicated by the positive coefficient. This indicates that older workers who are involved in accidents are less likely to have severe injuries; younger workers may have a higher probability of experiencing more severe injuries. However, age and years of employment are not influential factors in making a difference between fatal and incapacitating injuries.

Finally, the sensitivity analysis result indicated that workplace factors (21.3229) and working hours (28.3834) had a significant relationship with the level of injuries. Journalists Card et al. explained that workplace factors are crucial because they influence many other OSH accidents [5]. Furthermore, journalists Theofilatos et al. stated that the length of working hours positively correlates with the severity of the accidents [35]. This indicates that accidents that occur at night are more severe than those that occur during the day. It was also predicted that fewer accidents would occur during non-working hours as there would be less manpower involved.

In this study, the five main input variables (gender, age, occupation, workplace factors, and activities conducted) were enlarged into detail to illustrate the occurrence of OSH accidents. The variables of nationality, working hours, educational level, years of employment, and working zone were then used to analyze the input–output paired HSE factors. One might argue that such a phenomenon with these input variables may vary with the findings of other research. This is because our research separated the input data into categorical input at once. Therefore, more research should be done in which model sensitivity measures are computed for both single and multiple variable combinations as needed. The improved method would be more in line with the current industrial practice where a worker has numerous tasks and performs various jobs. This will affect the working zone, workplace factors, and activity involved as the variables that influence the risk exposure prediction.

5. Conclusion

Risk prediction has always been an inevitable challenge for OSH risk management, particularly during the investigation and prevention of accidents. Consequently, the necessity to perform more accurate risk exposure prediction stands as an active research area in the field of OSH risk management. Past literature denotes the recommendation of various models for their purpose based on the target output. However, there is a lack of risk prediction research in the shipyard industry. While ANN network modeling has been proven successful in various targets of prediction, this research is the first to use ANN modeling for risk prediction in the shipyard industry.

Our results advocate the efficacy of ANN machine learning with BPNN modeling to examine several variables and their interrelationships for risk prediction, which produced a successful linear modeling relationship. In the network architecture, 10 independent input variables and 1 single layer-dependent output variable (severity of risk prediction) were defined, followed by the development of 7 networks modeling between 19–25 neuron nodes. The accuracy performance of the developed networks was compared using several evaluation criteria to select the best model. It was found that the ANN model had a high *AP* of 90.2250% in predicting OSH risk exposure in the shipyard industry. This falls within the MAPE performance requirement, which is less than 30%.

In terms of sensitivity analysis, input factors, such as working hours and workplace factors, have significant effects on the output OSH risk prediction. Alternatively, educational level and nationality are the least influencing factors. Therefore, this paper proposes a room for improvement for future research by classifying more details in working hours and workplace factors to improve the accuracy performance of the ANN modeling [39].

Acknowledgement

Appendix

ANN	artificial neural network
AP	accuracy performance
BFGS	the Broyden-Fletcher-Goldfarb-Shanno algorithm
BPNN	backpropagation neural network
HSE	health, safety, and environmental
HSEMS	health, safety, and environmental management system
logsig	log-sigmoid activation function
M_a	mean average percentage error
MAPE	mean average percentage error
ML	machine learning
MLP	multilayer perceptron
MSE	mean square error
OSH	occupational safety and health
R	correlation coefficient
R^2	error square
SOP	standard operating procedure
tansig	hyperbolic tangent sigmoid activation function
trainbfg	quasi-Newton backpropagation
trainlm	Levenberg–Marquardt backpropagation
trainrp	resilient backpropagation
trainscg	scaled conjugate gradient backpropagation

References

[1] Anderson, W.E. “Risk Analysis Methodology Applied to Industrial Machine Development.” *IEEE Transactions on Industry Applications* 41, no. 1 (January 2005): 180–87. https://doi.org/10.1109/tia.2004.841006

[2] Othman, Idris, Mohd Fauzan Ibrahim, Nasir Shafiq, Hisham Mohamad, and Md Salim Kamil. “HSE Management System for Hotwork Operation at High Elevation in Shipbuilding Project.” *MATEC Web of Conferences* 203 (2018): 02005. https://doi.org/10.1051/matecconf/201820302005

[3] Etherton, John R. “Industrial Machine Systems Risk Assessment: A Critical Review of Concepts and Methods.” *Risk Analysis* 27, no. 1 (February 2007): 71–82. https://doi.org/10.1111/j.1539-6924.2006.00860.x

[4] Barlas, Baris. “Shipyard Fatalities in Turkey.” *Safety Science* 50, no. 5 (June 2012): 1247–52. https://doi.org/10.1016/j.ssci.2011.12.037

[5] Card, Alan J., James Ward, and P. John Clarkson. “Successful Risk Assessment May Not Always Lead to Successful Risk Control: A Systematic Literature Review of Risk Control after Root Cause Analysis.” *Journal of Healthcare Risk Management* 31, no. 3 (January 2012): 6–12. https://doi.org/10.1002/jhrm.20090

[6] Sarkar, Sobhan, Sammangi Vinay, Rahul Raj, J. Maiti, and Pabitra Mitra. “Application of Optimized Machine Learning Techniques for Prediction of Occupational Accidents.” *Computers & Operations Research* 106 (June 2019): 210–24. https://doi.org/10.1016/j.cor.2018.02.021

[7] Noori, Roohollah, Abdulreza Karbassi, and Mohammad Salman Sabahi. “Evaluation of PCA and Gamma Test Techniques on Ann Operation for Weekly Solid Waste Prediction.” *Journal of Environmental Management* 91, no. 3 (January 2010): 767–71. https://doi.org/10.1016/j.jenvman.2009.10.007

[8] Bouali, Hanen, and Jalel Akaichi. “Comparative Study of Different Classification Techniques: Heart Disease Use Case.” *2014 13th International Conference on Machine Learning and Applications*, December 2014. https://doi.org/10.1109/icmla.2014.84

[9] Yazdani, Armin, and Kannan Ramakrishnan. “Performance Evaluation of Artificial Neural Network Models for the Prediction of the Risk of Heart Disease.” *IFMBE Proceedings*, December 18, 2015, 179–82. https://doi.org/10.1007/978-981-10-0266-3_37

[10] El-Sawalhi, Nabil Ibrahim, and Omar Shehatto. “A Neural Network Model for Building Construction Projects Cost Estimating.” *Journal of Construction Engineering and Project Management* 4, no. 4 (December 1, 2014): 9–16. https://doi.org/10.6106/jcepm.2014.4.4.009

[11] “Factors Affecting Construction Labour Productivity for Floor Finishing Works.” *International Journal of Recent Trends in Engineering and Research* 3, no. 8 (September 2, 2017): 282–89. https://doi.org/10.23883/ijrter.2017.3415.cegvn

[12] Odeyinka, Henry A., John Lowe, and Ammar P. Kaka. “Artificial Neural Network Cost Flow Risk Assessment Model.” *Construction Management and Economics* 31, no. 5 (May 2013): 423–39. https://doi.org/10.1080/01446193.2013.802363

[13] Witten, Ian H., Eibe Frank, and Mark A. Hall. “Ensemble Learning.” *Data Mining: Practical Machine Learning Tools and Techniques*, 2011, 351–73. https://doi.org/10.1016/b978-0-12-374856-0.00008-0

[14] Matías, J. M., T. Rivas, J. E. Martín, and J. Taboada. “A Machine Learning Methodology for the Analysis of Workplace Accidents.” *International Journal of Computer Mathematics* 85, no. 3–4 (April 2008): 559–78. https://doi.org/10.1080/00207160701297346

[15] Martín, J.E., T. Rivas, J.M. Matías, J. Taboada, and A. Argüelles. “A Bayesian Network Analysis of Workplace Accidents Caused by Falls from a Height.” *Safety Science* 47, no. 2 (February 2009): 206–14. https://doi.org/10.1016/j.ssci.2008.03.004

[16] Hariharakrishnan, Jayaram, S. Mohanavalli, Srividya, and K. B. Sundhara Kumar. “Survey of Pre-Processing Techniques for Mining Big Data.” *2017 International Conference on Computer, Communication and Signal Processing (ICCCSP)*, January 2017. https://doi.org/10.1109/icccsp.2017.7944072

[17] Ayhan, Bilal Umut, and Onur Behzat Tokdemir. “Predicting the Outcome of Construction Incidents.” *Safety Science* 113 (March 2019): 91–104. https://doi.org/10.1016/j.ssci.2018.11.001

[18] Patel, D. A., and K. N. Jha. “Neural Network Model for the Prediction of Safe Work Behavior in Construction Projects.” *Journal of Construction Engineering and Management* 141, no. 1 (January 2015). https://doi.org/10.1061/(asce)co.1943-7862.0000922

[19] Alshihri, Marai M., Ahmed M. Azmy, and Mousa S. El-Bisy. “Neural Networks for Predicting Compressive Strength of Structural Light Weight Concrete.” *Construction and Building Materials* 23, no. 6 (June 2009): 2214–19. https://doi.org/10.1016/j.conbuildmat.2008.12.003

[20] Waziri, Baba Shehu, Kabir Bala, and Shehu Ahmadu Bustani. “Artificial Neural Networks in Construction Engineering and Management.” *International Journal of Architecture, Engineering and Construction* 6, no. 1 (March 1, 2017). https://doi.org/10.7492/ijaec.2017.006

[21] Abhishek, Kumar, Abhay Kumar, Rajeev Ranjan, and Sarthak Kumar. "A Rainfall Prediction Model Using Artificial Neural Network." *2012 IEEE Control and System Graduate Research Colloquium*, July 2012. https://doi.org/10.1109/icsgrc.2012.6287140

[22] Sapna, S. "Backpropagation Learning Algorithm Based on Levenberg Marquardt Algorithm." *Computer Science & Information Technology (CS & IT)*, October 31, 2012. https://doi.org/10.5121/csit.2012.2438

[23] Waziri, Baba Shehu, Kabir Bala, and Shehu Ahmadu Bustani. "Artificial Neural Networks in Construction Engineering and Management." *International Journal of Architecture, Engineering and Construction* 6, no. 1 (March 1, 2017). https://doi.org/10.7492/ijaec.2017.006

[24] Vakili, Masoud, Saeed-Reza Sabbagh-Yazdi, Koosha Kalhor, and Soheila Khosrojerdi. "Using Artificial Neural Networks for Prediction of Global Solar Radiation in Tehran Considering Particulate Matter Air Pollution." *Energy Procedia* 74 (August 2015): 1205–12. https://doi.org/10.1016/j.egypro.2015.07.764

[25] Khamis, M. F., Z. Baharudin, N. H. Hamid, M. F. Abdullah, and S. Solahuddin. "Electricity Forecasting for Small Scale Power System Using Artificial Neural Network." *2011 5th International Power Engineering and Optimization Conference*, June 2011. https://doi.org/10.1109/peoco.2011.5970423

[26] Azadeh, A., M. Rouzbahman, M. Saberi, and I. Mohammad Fam. "An Adaptive Neural Network Algorithm for Assessment and Improvement of Job Satisfaction with Respect to HSE and Ergonomics Program: The Case of a Gas Refinery." *Journal of Loss Prevention in the Process Industries* 24, no. 4 (July 2011): 361–70. https://doi.org/10.1016/j.jlp.2011.01.015

[27] Gevrey, Muriel, Ioannis Dimopoulos, and Sovan Lek. "Review and Comparison of Methods to Study the Contribution of Variables in Artificial Neural Network Models." *Ecological Modelling* 160, no. 3 (February 2003): 249–64. https://doi.org/10.1016/s0304-3800(02)00257-0

[28] Fish, Kelly E, and Jeffery G Blodgett. "A Visual Method for Determining Variable Importance in an Artificial Neural Network Model: An Empirical Benchmark Study." *Journal of Targeting, Measurement and Analysis for Marketing* 11, no. 3 (January 2003): 244–54. https://doi.org/10.1057/palgrave.jt.5740081

[29] Fish, Kelly E., and Richard S. Segall. "A Visual Analysis of Learning Rule Effects and Variable Importance for Neural Networks in Data Mining Operations." *Kybernetes* 33, no. 7 (August 1, 2004): 1127–42. https://doi.org/10.1108/03684920410534461

[30] Principe, Jose C. "Information Theoretic Learning." *Information Science and Statistics*, 2010. https://doi.org/10.1007/978-1-4419-1570-2

[31] Delen, Dursun, Ramesh Sharda, and Max Bessonov. "Identifying Significant Predictors of Injury Severity in Traffic Accidents Using a Series of Artificial Neural Networks." *Accident Analysis & Prevention* 38, no. 3 (May 2006): 434–44. https://doi.org/10.1016/j.aap.2005.06.024

[32] Willmott, CJ, and K Matsuura. "Advantages of the Mean Absolute Error (Mae) over the Root Mean Square Error (RMSE) in Assessing Average Model Performance." *Climate Research* 30 (2005): 79–82. https://doi.org/10.3354/cr030079

[33] Abdelwahab, Hassan T., and Mohamed A. Abdel-Aty. "Development of Artificial Neural Network Models to Predict Driver Injury Severity in Traffic Accidents at Signalized Intersections." *Transportation Research Record: Journal of the Transportation Research Board* 1746, no. 1 (January 2001): 6–13. https://doi.org/10.3141/1746-02

[34] Dissanayake, Sunanda, and Jian John Lu. "Factors Influential in Making an Injury Severity Difference to Older Drivers Involved in Fixed Object–Passenger Car Crashes." *Accident Analysis & Prevention* 34, no. 5 (September 2002): 609–18. https://doi.org/10.1016/s0001-4575(01)00060-4

[35] Theofilatos, Athanasios, Daniel Graham, and George Yannis. "Factors Affecting Accident Severity inside and Outside Urban Areas in Greece." *Traffic Injury Prevention* 13, no. 5 (September 2012): 458–67. https://doi.org/10.1080/15389588.2012.661110

[36] Erzin, Yusuf, and Yesim Tuskan. "The Use of Neural Networks for the Prediction of the Factor of Safety of the Soil against Liquefaction ." *Scientia Iranica* 0, no. 0 (May 6, 2018): 0–0. https://doi.org/10.24200/sci.2018.4455.0

[37] Dehghani, Farshad, Reza Eslamloueyan, and Mohammad Sarshar. "A Hybrid Model for Simulation of Lithium-Ion Batteries Using Artificial Neural Networks and Computational Fluid Dynamics." *Scientia Iranica* 29, no. 6 (May 11, 2022): 3208–3271. https://doi.org/10.24200/sci.2022.59292.6160

[38] Kumruoglu, Levent Cenk. "Prediction of Shrinkage Ratio of Za 27 Die Casting Alloy with Artificial Neural Network, Computer Aided Simulation and Comparison with Experimental Studies." *Scientia Iranica* 28, no. 5 (April 20, 2021): 2684–2700. https://doi.org/10.24200/sci.2021.54596.3824

[39] Büyükşahin, Ümit Çavuş, and Şeyda Ertekin. "Improving Forecasting Accuracy of Time Series Data Using a New Arima-Ann Hybrid Method and Empirical Mode Decomposition." *Neurocomputing* 361 (October 2019): 151–63. https://doi.org/10.1016/j.neucom.2019.05.099

[40] Gokler, Seda, and Semra Boran. “Determining Optimal Machine Part Replacement Time Using a Hybrid Ann-GA Model.” *Scientia Iranica* 29, no. (June 15, 2020): 721–782. https://doi.org/10.24200/sci.2020.52828.2902

[41] Niazkar, Majid, N. Talebbeydokhti, and S.-H. Afzali. “Bridge Backwater Estimation: A Comparison between Artificial Intelligence Models and Explicit Equations.” *Scientia Iranica* 28, no. 2 (February 12, 2020): 573–585. https://doi.org/10.24200/sci.2020.51432.2175

[42] Zeynalian, Mehran, and Iman Kalantari Dehaghi. “Choice of Optimum Combination of Construction Machinery Using Modified Advanced Programmatic Risk Analysis and Management Model.” *Scientia Iranica* 25, no. 3 (August 13, 2017): 1015–1024. https://doi.org/10.24200/sci.2017.4197

[43] Ceribasi, G., E. Dogan, U. Akkaya, and U.E. Kocamaz. “Application of Trend Analysis and Artificial Neural Networks Methods: The Case of Sakarya River.” *Scientia Iranica* 24, no. 3 (June 1, 2017): 993–99. https://doi.org/10.24200/sci.2017.4082

[44] Baykal, Tahsin, Fatih Ergezer, Ekinhan Eriskin, and Serdal Terzi. “Accident Severity Prediction in Big Data Using Auto-Machine Learning.” *Scientia Iranica* (February 22, 2023). https://doi.org/10.24200/sci.2023.60144.6626