ORIGINAL RESEARCH

Experimental study of neuro-fuzzy-genetic framework for oil spillage risk management

Oluwole Charles Akinyokun[1], Udoinyang Godwin Inyang[2]

1. Department of Computer Science, Federal University of Technology, Akure, Nigeria. 2. Department of Computer Science, University of Uyo, Uyo, Nigeria

Correspondence: Oluwole Charles Akinyokun. Address: Department of Computer Science, Federal University of Technology, Akure, Nigeria. Email: akinwole2003@yahoo.co.uk.

Received: January 3, 2013 **Accepted:** April 9, 2013 **Online Published:** June 5, 2013
DOI: 10.5430/air.v2n4p13 **URL:** http://dx.doi.org/10.5430/air.v2n4p13

Abstract

This paper reports the findings from the experimental study of an intelligent system driven by Neural Network (NN), Fuzzy Logic (FL) and Genetic Algorithm (GA) for knowledge discovery and oil spillage risk management. Application software was developed in an environment characterized by 11Ants Analytics, Matrix Laboratory (MatLab), Microsoft Excel, SPSS and GraphPadInstat as frontend engines; Microsoft Access Database Management System as backend engine and Microsoft Windows as platform. 11Ants Analytics served as a tool for oil spillage indicators rank analysis and predictive model building. Matlab served as a tool for the extraction of patterns from 11Ants Analytics Model of oil spillage. Microsoft Excel serves as an interface between 11Ants Analytics and Matlab. Microsoft Excel, SPSS and GraphPadInstat serve as tools for the generation of relevant statistics. Indicators of oil spillage risks serve as input to the NN. GA is used to provide optimal set of parameters for NN training while FL used for modelling imprecise knowledge and provision of membership functions for the GA and NN. Data on Oil Spill incidences associated with oil exploration activities in the Niger Delta Region of Nigeria were collected from National Oil Spill Detection and Response Agency (NOSDRA) and used to assess and evaluate the practical function of the intelligent system. Adaptive Neuro-Fuzzy Inference System (ANFIS) driven by Mamdani’s inference mechanism was used to predict and estimate oil spillage risks. The findings from the experimental study are presented.

Key words

Knowledge discovery, ANFIS, Risk analysis, Oil spillage management, Risk management

1 Introduction

Nations all over the world depend on oil and gas for fuelling of cars, generation of electricity and other domestic purposes. Oil exploitation and exploration with associated spillage have been increasing at alarming rates [1, 2]. During oil exploration, production, storage and transport activities, crude oil and products spill onto land and waterways. Oil spillage data are large, noisy and complex, and have some level of uncertainty associated with them. Statistical approaches, although offer precise methods for quantifying the inherent uncertainty that results from a particular sample or an overall population, they lack the ability to handle large, complex and noisy dataset and perform limited search during pattern extraction from databases. Conventional, database query methods produce limited and unreliable results desirable for effective decision-making.

Information Technology (IT) and advance modelling tools and techniques continue to help the society limit and manage disasters. The effectiveness of an oil spill response system and the robustness of a recovery plan are highly contingent upon an IT infrastructure that enhances information management, cooperation and coordination during disasters [3, 4]. Computer Systems, Decision Support Systems, KD and DM are some core tools that can assist in many aspects of oil spillage risk management. Where traditional analysis techniques fail to uncover hidden patterns from large and diverse datasets, knowledge discovery techniques succeed [5, 6]. KD and DM aims at extracting useful information and patterns from huge amount of data for prediction and modeling [7].

NN and DM tools offer ideal solutions to a variety of classification tasks such as speech, character and signal recognition as well as risk assessment and treatment. Although, gradient-based search techniques such as back-propagation are currently the most widely used optimization techniques for training NN, it has been shown that these gradient-based techniques are severely limited in their ability to find global solutions in a feasible computational time [8]. GA is a heuristic method used to find approximate solutions to complicated problems through application of the principles of evolutionary biology. The major strength of GA is that, bad proposals or noisy data do not affect the end solution negatively as they are simply discarded [9]. FL is a superset of Boolean Logic (BL), which handles the concept of partial truth [10, 11]. While FL performs inference mechanisms under cognitive uncertainty, NN use learning, adaptation, fault tolerance, parallelism and generalization to process data. Hence, to enable systems deal with cognitive uncertainty in a human like manner, one may incorporate the concept of FL into NN. Human operators can enhance NN by incorporating their knowledge with fuzzy membership functions, which are fine-tuned by a learning process. GA is a powerful tool for structure and weights optimization of NN. It is therefore useful to fuse NN, FL and GA techniques for offsetting the demerits of one technique by the merits of other techniques [12].

A combination of GA and a technique based on a localized Extended Kalman Filter (EKF) was used in the training of NN[13]. The GA provided a means of evolving a population of NN, to find optimal network architecture while EKF training algorithm was used in the network-learning phase. The hybrid model gave an accurate prediction of the future behavior of the currencies and in examining of the course of forecasting under a multi step horizon. A similar, but simpler, method was used successfully [14] for system structure identification using single layer NN. Liopa-Tsakalidi et al [15] demonstrates the effectiveness of GA in the estimation of unknown parameters of Richard’s function when used as a model to study the elongation of leaves and fruits of zucchini under the influence of two electrical conductivities (EC) of 2.2 dS/m and 4.4 dS/m in two nutrient solutions of hydroponics’ cultivation. The result showed that GA effectively simulates the plant growth and the effects of the two levels of EC in two different nutrient solutions than analytical/mathematical method. A study on the efficacy of extracts of Reynoutria Sachalinensis on cucumber growth using the Richards Function and Evolutionary Modelling is reported in [16]. The proposed technique of using GA effectively describes any unexpected or sudden changes that may occur in the plants’ elongation. A methodology for the development of Matrix Solid Phase Dispersion (MSPD) extraction for the determination of chlorinated compounds in fish using experimental design methods and NN is presented in [17]. The results show that the best possible performance of MSPD has been achieved using experimental design and NN modeling and demonstrate that the proposed soft computing strategy is very effective, efficient and achieves very satisfactory results. Due to the limitations of the ANN weights randomization techniques (slow training and sigmoid saturation problems), a GA model for ANN weight initialization and optimization was proposed and tested with its application in the classification and prediction of stroke disease [18]. A comparison of the output and the desired output of the NN and the hybrid system shows that the classification accuracy for all surfaces improved significantly with the Hybrid system. Furthermore, it was realized that the hybrid system provides solution to the shortcomings of the NN’s slow convergence and prevents it from stuck in the local minima.

The design of an expert system driven by NN, FL and GA for oil spillage risk management has been reported in[19]. The experimental study of that design is carried out in an environment characterized by Microsoft Windows as software platform, Microsoft Access Database Management System as backend, 11Ants Analytics Model Builder, Matrix Laboratory (Matlab), Microsoft Excel, SPSS and GraphPadInstat as frontend software is reported. Data collected from

Niger Delta Region of Nigerian National Oil Spill Detection and Response Agency (NOSDRA) on oil spill incidences associated with oil exploration activities was used in training and validating the system. The findings from the experimental study are reported in this paper. The primary objective is to demonstrate the practical function of that expert system.

2 Research method and materials

A review of existing literature on IT, NN, FL, GA, KD, DM, emergency risk management and oil spillage risks analysis was carried out. A multidimensional data model of emergency risk management using star, snowflake and facts constellation schemas was developed. The Fuzzy set (v) in V (universe of discourse) of oil spill attributes and its element denoted by x, is:

$$v= \{(x,\mu_v(x))| \ x \in V,\mu_v(x) \in[0,1]\} \tag{1}$$

where $\mu_v(x)$ is the membership function of x in v and μ_v is the degree of membership of x in v in the interval of [0,1]. It used the triangular membership function in Equation (2)

$$\mu_v(x) = \begin{cases} 1 & \text{if} \quad x < a \\ \dfrac{x-a}{b-a} & \text{if} \quad a \le x < b \\ \dfrac{c-x}{c-b} & \text{if} \quad b \le x < c \\ 0 & \text{if} \quad c = x \end{cases} \tag{2}$$

where a, b and c are the parameters of the membership function governing triangular shaped functions. Each of these attributes was described by linguistic terms of Very Low, Low, Medium, High and Very High}. A two layered feed forward neural networks was designed with the sigmoid neuron activation function given by Equation 3 and the output layer neuron given by Equation 4.

$$f(u) = [1+\exp(-u)]^{-1} \tag{3}$$

$$Z_k = \left[1+\exp(-\sum_{j=1}^{q}\omega_{ij}y_j - \theta_k)\right]^{-1} \tag{4}$$

The NN is trained with GA and the GA is created by an initial population of weights. Thereafter, the input variables, which were first encoded in binary, were converted to real value weights using the functions in Equations (5) and (6).

$$g_i = \begin{cases} 1 & \text{if} \quad b_1 = 1 \\ -1 & \text{if} \quad b_1 = 0 \end{cases} \tag{5}$$

$$R_i = \frac{g_i}{10}\sum_{t=2}^{m}\left(b_k \times 2^{m-k}\right) \tag{6}$$

where g_i is the sign bit of $gene_i$, R_i is the real value encoding for the ith gene, m represents the length of a gene, t=2, 3, …m. b_k is the kth bit in the gene.

GA operators were applied to adjust the weights of the neural network and the outcome of each adjustment was evaluated by the GA fitness function until 2n generations was reached and an optimal combination of weights chosen. The normalized fitness function is given as:

$$T_i = \frac{f_i}{\frac{1}{n}\sum_{j=1}^{n} f_i} \tag{7}$$

Where T_i is the normalized fitness of the ith chromosome, n is the total number of nodes in the neural network; f_i is the fitness of the ith chromosome; j=1,2,. . n.

A six layered neuro-fuzzy inference engine was developed. The first, second and fifth layers consist of adaptive nodes while the third, fourth and sixth layers are fixed nodes. The architecture implements the Mamdani's inference mechanism and can handle rules of the form:

IF (x is A) AND (y is B) AND (z is C) THEN (S is O)

where x, y, z are inputs variables, A, B, C are fuzzy sets of the input variables, S is the output and O is fuzzy set of the output variable within the fuzzy region specified by the rule. Based on the architecture, an expert system driven by FL, NN and GA was developed using 11Ants Analytics Model Builder, Matrix Laboratory (Matlab) programming languages, Microsoft Excel, SPSS and GraphPadInstat as software front end, Microsoft Access Database Management System as software back end and Microsoft Windows as software platform. The case study of the data on oil spill incidences of the National Oil Spill Detection and Response Agency (NOSDRA) of Nigeria was carried out. To determine the efficiency and effectiveness of the system, the results obtained were analyzed using precision, accuracy, sensitivity and Receivers Operating Characteristics (ROC) curve.

11Ants Model Builder and Matlab software packages which serve as frontend engine were used for knowledge [20]. 11Ants Model Builder has strengths in terms of easy data preparation tools and support for very large dataset. It features an impressive library of a wide range of machine learning algorithms including NN and GA. The proprietary HyperLearn technology effectively allows a computer to build accurate predictive models for a given dataset, even from a search space of billions [21]. It can efficiently process numeric and non-numeric values.

Matlab is a Microsoft Windows based application development tool. It is a numerical computing system and fourth-generation programming language developed by Mathworks Inc [22]. It provides exceptional features for deployment of patterns and relationships from models. In addition, Matlab is very efficient in matrix manipulation, has built-in functions for solving problems requiring data manipulation and analysis, signal processing, optimization and several other types of scientific computations. It also contains functions for 2D and 3D graphics and animation and has a rich Graphical User Interface Development Environment (GUIDE) that makes writing of codes very simple and easy. The availability of FL, adaptive Neuro Fuzzy Inference System and Database toolboxes is a major strength over other programming tools.

Matlab has spreadsheet Link EX tool that connects Microsoft Excel spreadsheet software with the Matlab workspace, enabling the access of Matlab environment from Microsoft Excel spreadsheet. With Spreadsheet Link EX software, exchange of data between Matlab and Excel is possible. Furthermore, Microsoft Excel is the environment for 11Ants Analytics suite. The conceptual diagram of the interconnection of 11 Ants Analytics, MatLab and Microsoft Excel is presented in Figure 1. Microsoft Excel, SPSS and GraphPadInstat were tools used for statistical analysis. Microsoft Access relational database management system is used in this research because its flexibility and effectiveness in the creation, storage, modification and retrieval of data. With Microsoft Access, it is also easier to build queries, forms and reports that are suitable for Matlab applications [23]. In addition, a combination of Microsoft Access and Matlab provides greater support for data abstraction, inheritance, polymorphism and encapsulation in Microsoft Windows platform than in any other platform [24]. 11Ants Analytics only works on Windows Operating System platform, therefore the choice of other platforms for the implementation of this system will be inappropriate.

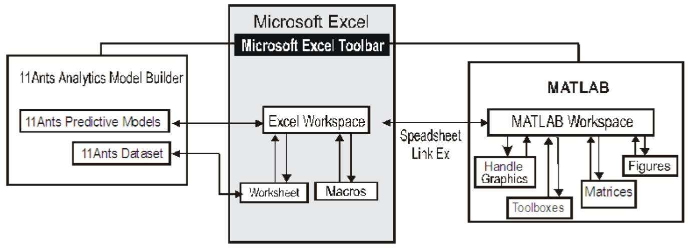


Figure 1. Interaction between 11Ants Analytics, Microsoft Excel and Matlab

Microsoft Access Relational Database Management system is used in this research because its flexibility and effectiveness in the creation, storage, modification and retrieval of data. With Microsoft Access, it is also easier to build queries, forms and reports that are suitable for MatLab applications [23]. In addition, a combination of Microsoft Access and MatLab provides greater support for data abstraction, inheritance, polymorphism and encapsulation in Microsoft Windows platform than in any other platform [24]. 11Ants Analytics only works on Windows Operating System platform, therefore the choice of other platforms for the implementation of this system will be inappropriate.

The architecture of Oil Spillage Risk Management (OSRM) is presented in Figure 2. The major components of the architecture are as follows:

a) Knowledge Base of OSRM.
b) Knowledge Mining Engine of OSRM.
c) Decision Support Engine of OSRM.

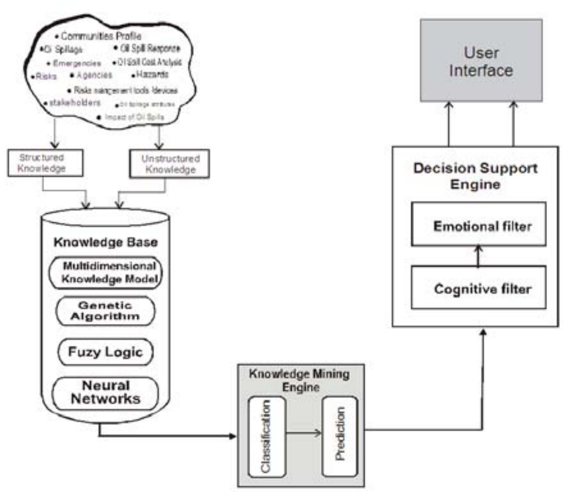


Figure 2. Architecture of OSRM

3 Data survey, collection and training

The Federal Government of Nigeria established NOSDRA as a parastatal under the Federal Ministry of Environment to implement the National Oil Spill Contingency Plan in line with International Conventions. The agency is also empowered to ensure timely, effective and appropriate response in terms of necessary equipment and resources to protect threatened environment and facilitate clean up of impacted sites to the best practical extent including remediation and restoration[25]. NOSDRA, therefore, regularly conducts visits to oil companies for the purpose of their facilities inspection, collects data and generate reports on oil spillage, and coordinates oil spill response activities throughout Nigeria.

The data used in the experimental study, which covered January 2005 through November 2011, were collected from NOSDRA. The details of the data have been presented in [19]. The knowledge mining which is performed with 11Ants Analytics advanced in the following stages as in [26]:

a) Selection and dataset pre-processing which involves the identification of the target variable in the dataset and selection of the input variables required for the knowledge discovery process. In this work, the target variable is the magnitude of spillage while the input variables, which are indicators of the NN are Day-of-Occurrence, Month-of-Occurrence, Year-of-Occurrence, Time-of-Occurrence, Location-of-Spill, Cause-of-Spill and Type-of-Spill.

b) Transformation of oil spillage indicators from gross form into atomic form. For example, the date component of oil spillage indicator is transformed into Day_Occurrence, Month_Occurrence and Year_Occurence. This was done to enable the extraction of knowledge from each category of day, month and year. The Time variable was transformed into two categories, namely; AM (12.00am -12.00noon) and PM (12.01pm – 11.59pm). The values of the variables are large, hence cannot be recorded in this paper but have been presented [19].

c) Knowledge mining was performed by loading Microsoft Excel 2007 application software, which automatically loads 11Ants Analytics, and Matlab 7.7.0 Software as Add-Ins. This permits the exchange of data between, Ms Excel, Matlab and 11Ants Analytics. The main menu list in the Ms Excel Ribbon with the 11Ants Analytics toolbar is composed of 'split data", "analyze data", "predict" and "manage". The indicators to the NN, day, month, year, time, location, cause and type of spill are in Columns A to G respectively, of the Ms Excel worksheet named 'Spill_Data'. Column H has Magnitude which is the target variable. The topmost row of the excel worksheet contains the name of the inputs. Upon selection of the columns A-G of the Excel worksheet and clicking any of the 11Ants menu item, the knowledge discovery begins with an activation request window, which is mandatory before using 11Ants Analytics.

After supplying the email address, activation key and clicking the 'Activate Now' button, the parameters are verified at the 11Ants Server (www.11Antsanalytics.com). Upon successful verification, the 11Ants Analytics menu is enabled on the Microsoft Excel Ribbon. In the Excel worksheet, the columns containing the variables (input and target) were selected and the "Split data" tab was clicked.

Upon supplying the appropriate percentages for the train and test datasets, and clicking the "Spilt Data" tab, the data was split into, training dataset and test dataset, each in separate worksheets, 857 (85%) and 151 (15%) of the data were used as train dataset and test dataset respectively. The training dataset allows the NN to get optimal parameters from the GA, both driven by 11Ants Analytics. The test dataset provides a true indication of performance of the NN model on the dataset. When the option "Analyze Data" is selected from the menu list, the Analyze Data Window for Oil Spillage is displayed on the screen.

By clicking "Start Analyzing data" on the menu list of the window, the 11Ants Model Building process window is displayed on the screen. The rank of the attributes based on their contribution to the target variable at 75% model quality is as depicted in Table 1. The results show the relative contributions of "Type of Spill Oil" as 36.3%, "Cause of Spill" as 29.6%, "Location of Spill" as 19.3%, "Day of Spill" as 8.21%, "Month of Spill" as 4.36%, "Time of Spill" as 2.35% and "Year of Spill" as 0.0%. Thus, "Year of Spill" has no contribution to the magnitude of spill and was eliminated from the list of indicators during training. The importance of indicators to magnitude of oil spill is presented in Table 1 and Figure 3. The built model, which was trained without year of spill, yielded 89% quality. Clicking the 'Build Model' tab from the

11Ants Model Builder Window, causes the display on the screen the Build Predictive Model Window. By clicking the "Start Building Predictive Model" of the window, the Oil Spill Hazard Model is built and ready for use.

Table 1. Relative Importance (weights) of Oil Spillage Indicators

Rank	Input Variable	Relative Importance	Relative importance (%)	Cumulative importance (%)
1	Type	0.363	36.2	36.2
2	Cause	0.296	29.6	65.8
3	Location	0.193	19.3	85.1
4	Day	0.0821	8.21	93.3
5	Month	0.0436	4.36	97.6
6	Time	0.0235	2.35	100.0
7	Year	0	0	100.0

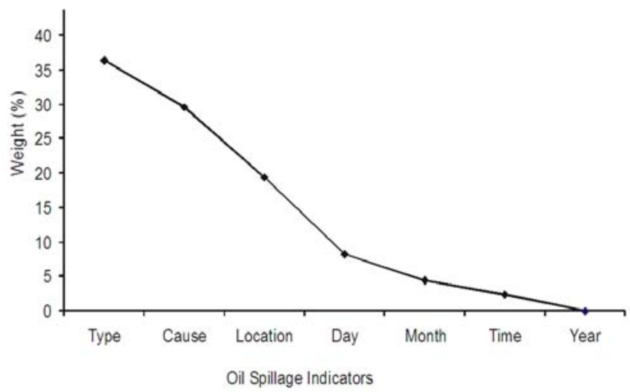


Figure 3. Graph of Relative Importance of Oil Spillage Indicators

4 Model evaluation and interpretation of results

When the columns of the Test Dataset worksheet of MS Excel containing the indicators are selected and "Predict" option of the 11Ants Menu bar is clicked, "Predict Using Model" window is displayed. When the "Predict Now' button is clicked, the predicted results, their analysis and summary are presented in Figure 4, Table 2 and Table 3 respectively.

11Ants Analytics
Data Mining Made Simple

Model Name: Oil Spill Hazard Model	
Statistic	**Predicted versus Known Values**
Count	151
Accuracy	97.35%
Default Accuracy (Always Predict Majority)	39.74%
Improvement over Default	57.62%

Figure 4. Analysis of Known and Predicted Values

Table 2. Predicted and Desired Values from the Test Dataset

S/N	Desired Values	Predicted Values	s/n	Desired Values	Predicted Values	s/n	Desired Values	Predicted Values
1	Medium	Medium	52	Medium	Medium	103	Very High	Very High
2	Medium	Medium	53	High	High	104	Very High	Very High
3	Very High	Very High	54	Very High	Very High	105	High	High
4	Very High	Very High	55	High	High	106	High	High
5	High	High	56	High	High	107	Very High	Very High
6	Medium	Medium	57	Medium	Medium	108	High	High
7	High	High	58	High	High	109	Very High	Very High
8	Medium	Medium	59	High	Medium	110	Very High	Very High
9	Very High	Very High	60	High	High	111	High	High
10	Low	Low	61	Medium	Medium	112	High	High
11	High	High	62	Medium	Medium	113	Very High	Very High
12	High	High	63	Very High	Very High	114	Medium	Medium
13	Medium	Medium	64	Very High	Very High	115	Medium	Medium
14	High	High	65	Medium	High	116	Very High	Very High
15	Very High	Very High	66	Very High	Very High	117	High	High
16	High	High	67	Very High	Very High	118	Very High	Very High
17	High	High	68	Very High	Very High	119	High	High
18	High	High	69	Very High	Very High	120	High	High
19	High	High	70	Very High	Very High	121	Medium	Medium
20	Very High	Very High	71	High	High	122	High	High
21	High	High	72	High	High	123	High	High
22	High	High	73	Medium	Medium	124	Very High	Very High
23	High	High	74	High	High	125	Medium	Medium
24	Medium	Medium	75	High	High	126	High	High
25	Very High	Very High	76	Very High	Very High	127	Very High	Very High
26	High	High	77	Very High	Very High	128	Low	Low
27	Very High	Very High	78	High	High	129	Medium	Medium
28	Medium	Medium	79	Medium	Medium	130	High	High
29	High	High	80	High	High	131	Very High	Very High
30	High	High	81	Very High	Very High	132	Very High	Very High
31	Very High	Very High	82	Low	Low	133	Very High	Very High
32	Very High	Very High	83	Very High	Very High	134	Very High	Very High
33	Very High	Very High	84	High	High	135	Very High	Very High
34	High	High	85	Medium	High	136	Low	Low
35	High	High	86	High	High	137	Very High	Very High
36	High	High	87	Medium	Medium	138	High	High
37	Very High	Very High	88	High	High	139	Very High	Very High
38	Medium	High	89	Low	Low	140	Low	Low
39	High	High	90	High	High	141	Medium	Medium
40	High	High	91	Very High	Very High	142	High	High
41	Very High	Very High	92	High	High	143	Medium	Medium
42	High	High	93	Very High	Very High	145	Medium	Medium
43	Low	Low	94	High	High	146	Medium	Medium
44	Very High	Very High	95	Very High	Very High	147	Medium	Medium
45	High	High	96	Very High	Very High	148	Very High	Very High
46	Very High	Very High	97	Medium	Medium	149	Medium	Medium
47	High	High	98	Low	Low	150	High	High
48	Very High	Very High	99	Very High	Very High	151	Very High	Very High
49	Very High	Very High	100	Medium	Medium			
50	High	High	101	High	High			
51	High	High	102	High	High			

Table 3. Summary of Predictions

	Known Label					
		High	Very High	Medium	Low	Total
Predicted Label	High	59	0	3	0	62
	Very High	0	53	0	0	53
	Medium	1	0	27	0	28
	Low	0	0	0	8	8
Total		60	53	30	8	151

Table 4. Model Confusion Matrix

Binary Breakdown per Label	**Positive Label**			
Statistic	**High**	**Very High**	**Medium**	**Low**
True Positives	59	53	27	8
True Negatives	88	98	120	143
False Positives (Type I Errors)	3	0	1	0
False Negatives (Type II Errors)	1	0	3	0
Accuracy	97.35%	100.00%	97.35%	100.00%
Precision	95.16%	100.00%	96.43%	100.00%
Recall / Sensitivity	98.33%	100.00%	90.00%	100.00%
Specificity	96.70%	100.00%	99.173%	100.00%
F-Measure	0.9672	1	0.9310	1

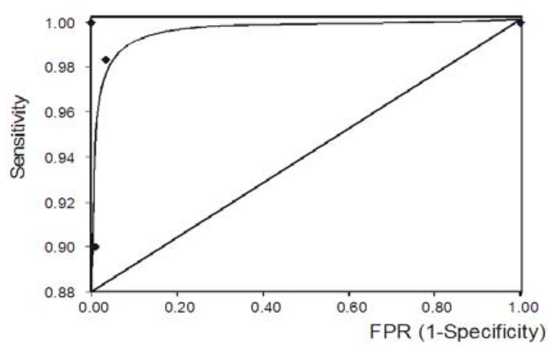


Figure 5. Model ROC Curve

From Table 3, patterns leading to, Very High and Low have 100% correct prediction. 3(10%) instances that belonged to Medium category were incorrectly predicted as high, while 1(1.67%) case that was high was incorrectly classified as medium. The breakdown of the result based on each fuzzy linguistic value is given in the Confusion Matrix depicted in Table 4.3 which shows that out of 60 cases that were known to be high, 1 case was not classified correctly (False Negative) while 3 cases that were not High were predicted as High (False Positive). In the medium category, out of 30 cases, 27 were correctly classified (True Positive) while 3 cases were excluded (False Negative). Only 1 case was incorrectly classified as medium (False Positive). True Positive Rate (TPR) for Very High and Low categories each is 100% (without type 1 and type II errors). The precision is 100% for Very High and Low categories, 95.16% for High category and 96.43% for medium category. The specificity for all categories is very high which implies insignificant False Positive Rate (FPR) and high TPR. The F-measure on the average is 0.974. The results of accuracy, precision, sensitivity, specificity and F-measure prove that the model is efficient and effective for classification and prediction of the magnitude of Oil Spill hazard. The Receivers Operating Characteristics (ROC) curve of the model is presented in Figure 5. Each point on the ROC curve is a coordinate (FPR, TPR) and represents an instance of a predicted fuzzy linguistic label of magnitude of spill. The ROC curve is closer to the perfect classification coordinate (0,1), which indicates an excellent performance of the model [27]. It also shows that any increase in sensitivity, is accompanied by a decrease in specificity, which is a major characteristic of effective and efficient classifiers. The model therefore is effective and efficient in the classification of magnitude of oil spill hazard.

4.1 Knowledge extraction and inference

Matlab was used to invoke the 11Ants oil spill model for knowledge extraction for associated risk evaluation. The major operations of the inference engine are as follows:

a) Extract patterns from 11Ants Analytic Model.

b) Build Adaptive Neuro Fuzzy Inference System (ANFIS) rules from the extracted patterns.

c) Analyze oil spillage risks.

d) Predict risk associated with the patterns from user specified indicators.

When the “Extract Patterns” option is selected from the main menu, the 11Ants Oil Spill Model is invoked and its arguments transferred to the ‘extract patterns’ module. The 11Ants Model training data is selected in the dataset box by clicking Browse. The browse button allows the user to explore the entire computer to specify the location of the required file. In the Knowledge Mining Criteria box, the minimum confidence and minimum support threshold was 10% and 15% respectively. Upon clicking the “Begin Extraction” option of the menu, a screen showing the extracted patterns (relationships and interdependencies between the input variables and the target variable) is displayed. Upon selecting ‘Build Rules’ option from the screen the “Rules Builder Screen” is displayed.

A total of 329 rules were built from the 11Ants Analytics model. 10% support and 15% confidence was the minimum threshold for rule pruning. The extracted rules have been presented in ref. 19. The graphical analysis of the rules gives the analysis of the number of rules, the support of rules and the Confidence of rules. When the ‘Number Rule Bar’ is clicked, a graph of the number of rules is displayed as shown in Figure 6. The graph shows that the 77 rules have single antecedent, while those with 2 indicators in the antecedent are the highest with 110 rules. Rules with the antecedent size of 3 are 95 while 40 rules have 4 indicators in the rule antecedent part and 7 rules are for antecedent size 5. When the “Rule Support” button is clicked, the graphical analysis of the rules based on their support depicted in Figure 7 is displayed. The graphical analysis shows that the support of rules decreases as the number of rules increases. When the “Rule Confidence” button is clicked, the visualization of the confidence of rules presented in Figure 8 is displayed.

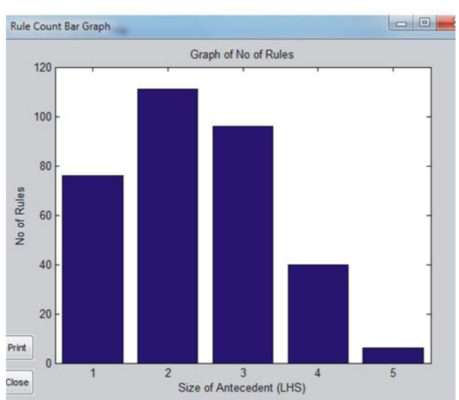


Figure 6. Graphical clustering of rules based on antecedent size

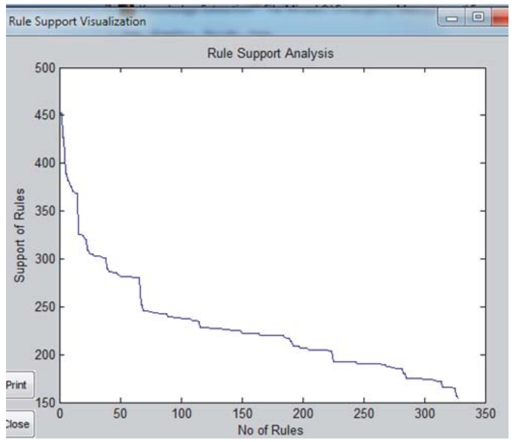


Figure 7. Graph Analysis of Rule Support

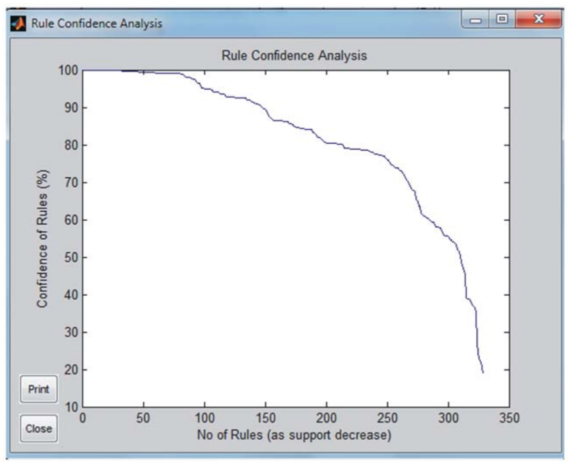


Figure 8. Graphical Analysis of Rule Confidence

4.2 ANFIS inference engine and rule pruning

When the “Oil Spillage Analysis” button is selected from the “Model Deployment Main Menu”, the “Analysis Window Menu” having options “Edit ANFIS”, “Summaries”, “Risk Analysis” and “Exit” is displayed on the screen. When the “Edit ANFIS” button is clicked, the ANFIS Editor showing that ANFIS inference engine has three inputs uses Mamdani’s inference mechanism with Min and Max for the OR operator and AND operator respectively. The Max and Centroid techniques were used for aggregation and defuzzification respectively. Five linguistics terms very low, low, medium, high and very high describes the input and output values. The membership function of each linguistic variable is in the range [0,1] and adopts the triangular membership function as presented in Figure 9.

The number of rules with the antecedent size of 1 and 2, though have high support, may not be interesting since they can occur by chance [28]. The rules are therefore pruned with antecedent size less than 3. In addition, the important values of oil spillage indicators in Table 1, serves as a measure, for filtering out insignificant or irrelevant input attributes from the training and test dataset [29, 30]. Three attributes of oil spillage (type, cause and location) have significant influence on the magnitude of spill. Rules with any of these significant attributes missing or including insignificant attributes are considered ‘noisy’ and were pruned, since they may not provide accurate analysis [31]. The ANFIS is therefore driven by:

a) Rules whose confidence and support are equal to or above the minimum threshold of 10% and 15% respectively.

b) Rules whose antecedent part is made up of all the significant indicators only.

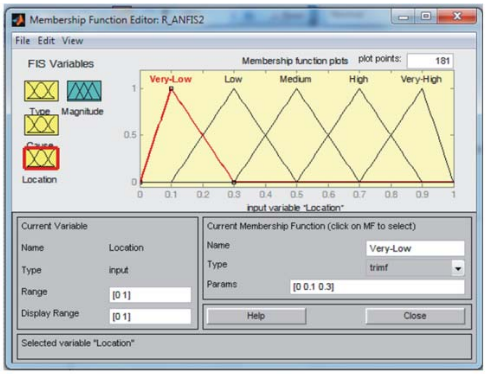


Figure 9. Membership Function of ANFIS Inference Engine

The ANFIS engine is therefore driven by 73 rules for the prediction of oil spillage risks. The system will interactively predict the risk from oil spill hazard indicators. Upon selection of the input indicators and clicking on 'predict' the class of risk associated with the fired rule is displayed. The interpretation of the risk is as follows:

a) Extreme: High probability that the consequence severity will cause deaths in the community.

b) High: Low probability of deaths in the community but high probability of acute injuries.

c) Moderate: Probability of deaths or acute injuries in the community is virtually zero but probability of minor injuries remains.

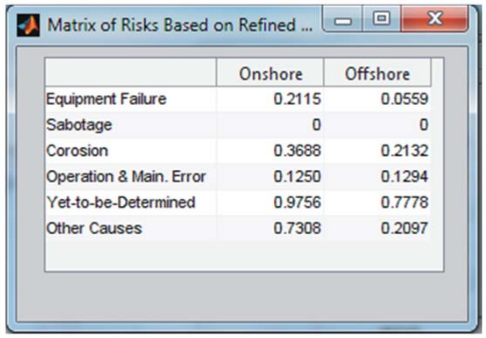
d) Low: There is no risk of deaths or any type or injuries to community assets.

e) Negligible: There is no risk to residents at all and consequences do not extend beyond the site boundary.

4.3 Results of oil spillage risk analysis and discussion

When any of the buttons on the 'Risk Analysis Menu', depicted in Figure 10 is clicked, the matrix of risks corresponding to that type of oil, for all the causes and in all locations of spillage is displayed.

Figure 10. Risk Analysis Menu



Matrix of Risks Based on Refined ...

	Onshore	Offshore
Equipment Failure	0.2115	0.0559
Sabotage	0	0
Corosion	0.3688	0.2132
Operation & Main. Error	0.1250	0.1294
Yet-to-be-Determined	0.9756	0.7778
Other Causes	0.7308	0.2097

Figure 11. Risk Based on Refined Product Spillage

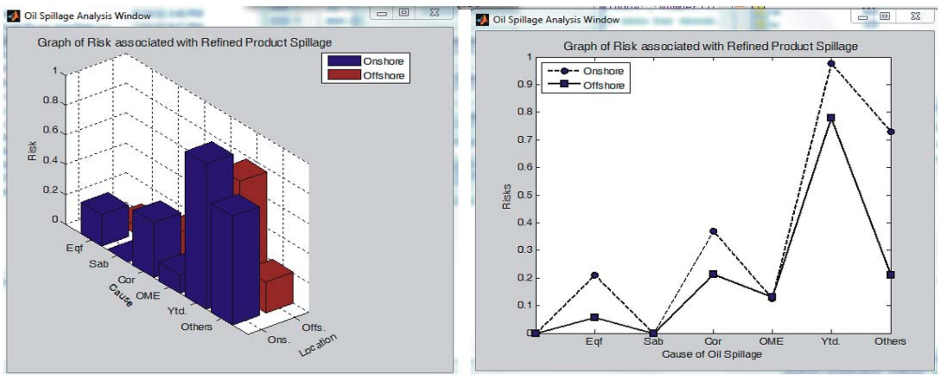
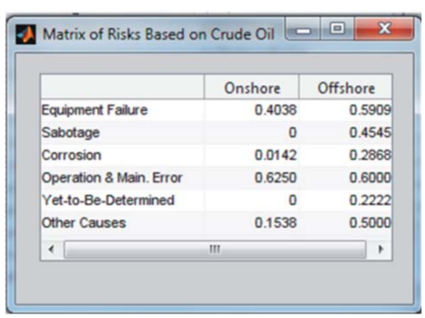


Figure 12. Graphs of Risk based on Refined Product Spillage

a. Risk Analysis by Refined Product Spillage: The matrix of risk associated with refined product spillage is presented in Figure 11 while the 3D plots and the line graph are presented in Figure 12. The results show that refined product spillage risk is relatively higher in onshore than offshore locations. In terms of cause, there is no refined product risks associated with sabotage while risk attributed to corrosion is higher than risk from equipment failure in both locations. Causes that are yet-to-be-determined (unknown) have a risk of 0.97 and 0.77 in onshore and offshore locations respectively. Therefore, risk of refined product spillage induced by corrosion, unknown and other causes are high and may pose significant harm on the environment. Risks from unknown sources are extreme; hence, a proper analysis to identify the source of the spill so that appropriate measures can be taken to prevent such spillage is recommended.



Matrix of Risks Based on Crude Oil

	Onshore	Offshore
Equipment Failure	0.4038	0.5909
Sabotage	0	0.4545
Corrosion	0.0142	0.2868
Operation & Main. Error	0.6250	0.6000
Yet-to-Be-Determined	0	0.2222
Other Causes	0.1538	0.5000

Figure 13. Matrix of Risk Based on Crude Spillage

b. Risk Analysis by Crude Spillage: The result of risk analysis based on crude spillage is presented in Figure 13. The graphical interpretation of the risks are shown in Figure 14. The results show that risks associated with crude spill are high and vary considerably by cause and location. This means that the impact of the crude spill on the environmental resources depends on the cause of the spill and location where the spill occurs. Apart from Operational and Maintenance Error (OME), risk associated with each cause of spillage is higher in offshore location than onshore location. Risk due to sabotage, corrosion and unknown causes are insignificant in onshore location. In general, risk posed by crude spillage from any source is high in offshore locations and relatively higher than the spill of any other oil type. Therefore, adequate measures should be put in place to promptly detect and respond to crude spillage.

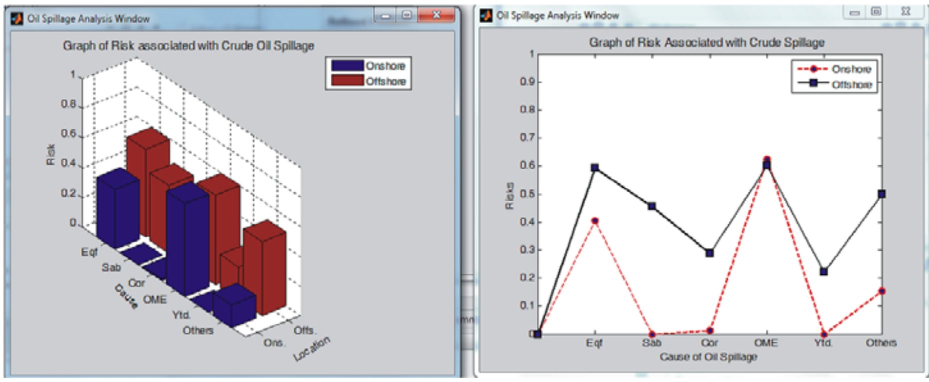


Figure 14. Graphs of Risk Based on Crude Spillage

c. Risk Analysis by Chemical Spillage: The matrix and graphically representation of risk due to Chemical spillage is depicted in Figure 15 and Figure 16 respectively. The results show that risks due to chemical spillage are almost negligible in onshore location. The most frequent cause of chemical spill in offshore location is Sabotage, followed by equipment failure while unknown causes are the least with 0.00 risk value. The result therefore shows that minimizing sabotage and equipment failure will help minimize the risk of chemical spillage in offshore locations.

Matrix of Risks Based on Chemical

	Onshore	Offshore
Equipment Failure	0.0385	0.1329
Sabotage	0	0.2727
Corrosion	0.0709	0.0956
Operation & Main. Error	0	0.0471
Yet-to-be-Determined	0	0
Other Causes	0.0385	0.0484

Figure 15. Matrix of Risk Based on Chemical Spillage

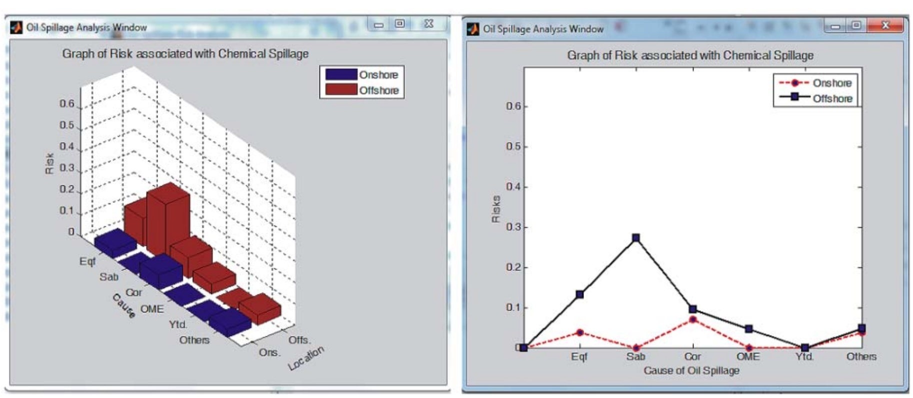


Figure 16. Graphs of Risk Based on Chemical Spillage

d. Risk Analysis by AGO Spillage: The result of oil spillage analysis based on AGO spillage for the different causes and locations are presented in Figure 17. The pictorial interpretation of the matrix is presented in Figure 18. The results show that risk associated with AGO spillage for the different causes and locations are below 2.0, therefore insignificant. However, there is absence of AGO spill risks from sabotage and unknown sources while OME and corrosion induced risk are 0.16 and 0.13 respectively for offshore location. In summary, the AGO risks induced by a specific cause do not vary significantly by location and are low irrespective of Cause and location.

Matrix of Risks Based on AGO

	Onshore	Offshore
Equipment Failure	0.1538	0.1189
Sabotage	0	0
Corrosion	0.1418	0.1324
Operation & Main. Error	0.1250	0.1647
Yet-to-be-Determined	0	0
Other Causes	0.0385	0.1452

Figure 17. Matrix of Risk Based on AGO Spillage

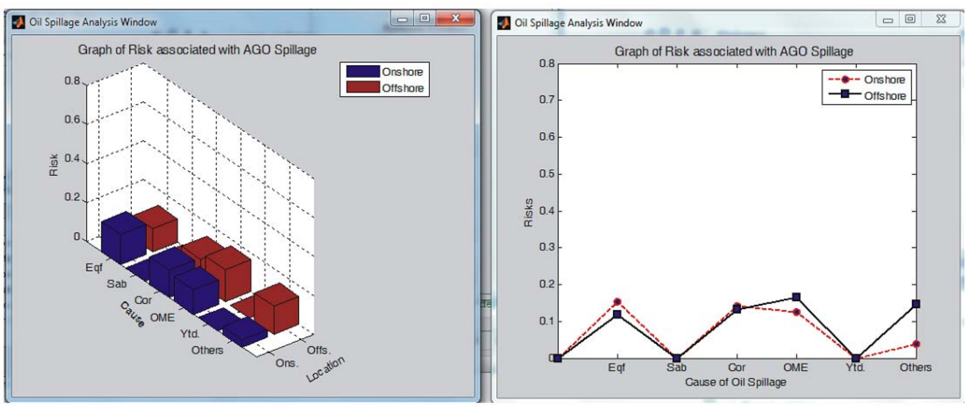


Figure 18. Graphs of Risk Based on AGO Spillage

e. Risk Analysis by Condensate Spillage: Condensate spillage risks are summarized as shown in Figure 19 and Figure 20. The results show that condensate risk caused by OME in onshore is greater than 0.1. Risks emanating from sabotage, Corrosion and other sources are almost zero in onshore locations while OME and other sources are almost zero in offshore locations. Therefore, condensate risks are negligible irrespective of cause and location and may not have significant effect on the environment.

Matrix of Risks Based on Condens...

	Onshore	Offshore
Equipment Failure	0.0192	0.0315
Sabotage	0	0.0909
Corrosion	0.0142	0.0441
Operation & Main. Error	0.1250	0
Yet-to-be-Determined	0.0244	0
Other Causes	0	0.0161

Figure 19. Matrix of Risk Based on Condensate Spillage

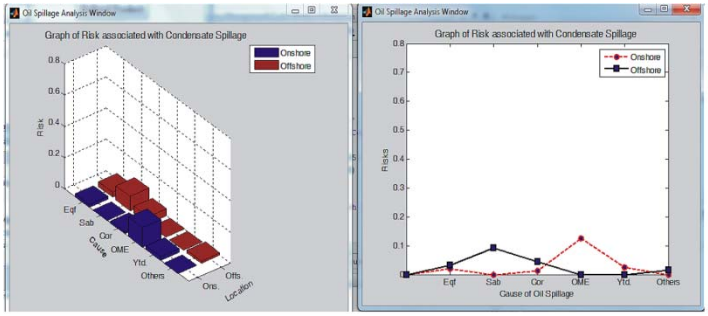


Figure 20. Graphs of Risk based on Condensate Spillage

Matrix of Risks Based on Other

	Onshore	Offshore
Equipment Failure	0.1731	0.0699
Sabotage	0	0.1818
Corrosion	0.3901	0.2279
Operation & Main. Error	0.1250	0.0588
Yet-to-be-Determined	0	0
Other Causes	0.0385	0.0806

Figure 21. Matrix of Risk Based on *Other* (Miscellaneous) Oil Spillage

f. Risk Analysis by “Other Oil type” Spillage: In addition to crude, chemical, AGO, condensate and refined product, other types of oil (sometimes unknown), which also poses risk to the environment, may be released. The matrix of risks associated with this category of oil with respect to its cause and location is presented in Figure 21. The pictorial representation of this is presented in Figure 22. The results show that corrosion in onshore location produces the highest risk, while sabotage and unknown causes induced no risk in onshore locations. However, in offshore location, risks associated with sabotage and corrosion are greater than 0.2. Other sources of spillage have risk less than 0.1 for both locations.

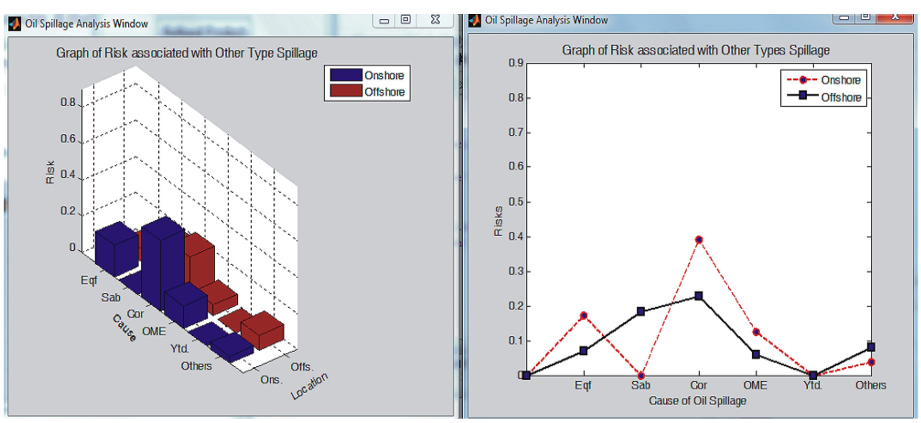


Figure 22. Graphs of Risk based on Spillage Other (miscellaneous) Oils

4.4 Statistical analysis of results and discussion

A 3-way Analysis of Variance (ANOVA) was carried out on the result obtained from oil spillage risk analysis, to determine the effects and interaction between type, cause and location on oil spillage risk. Statistical Package for Social Science (SPSS) version 17.0 and GraphPadInstat 5.1 are the statistical software tools used for the analysis. The factors (independent variables) are type, cause and location while the dependent variable is risk. The 3-way ANOVA output report is presented in Table 5.

Table 5. Result of 3-Way ANOVA

Tests of Between-Subjects Effects						
Dependent Variable: Risk						
Source	**Type III Sum of Squares**	**df**	**Mean Square**	**F**	**Sig.**	**Partial Eta Squared**
Corrected Model	2.968#	46	.065	9.699	.000	.947
Intercept	1.719	1	1.719	258.402	.000	.912
Cause	1.014	5	.203	30.485	.000	.859
Type	.074	5	.015	2.225	.083	.308
Location	.011	1	.011	1.598	.218	.060
Cause * Type	1.523	25	.061	9.156	.000	.902
Cause * Location	.273	5	.055	8.198	.000	.621
Type * Location	.074	5	.015	2.225	.083	.308
Error	.166	25	.007			
Total	4.853	72				
Corrected Total	3.134	71				

#. R Squared = .947 (Adjusted R Squared = .849).

As shown in Table 5, there is a significant main effect of cause of oil spillage, ($F(5,25) = 30.49$ and $p = 0.00$) on risk of spillage at 95% confidence level. This implies that, the mean risk from each cause of oil spillage vary significantly from one another. The risk severity rating for equipment failure (Mean = 0.32, S.D = 0.24) is significantly, the highest followed by sabotage (Mean = 0.32, S.D = 0.33) and unknown causes (Mean = 0.11, S.D = 0.12) with Corrosion (Mean = 0.03, S.D = 0.04) as the lowest. The mean risk level attributed to each cause of oil spillage is presented in Figure 23. Post hoc comparisons using the Least Significant Differences (LSD) procedure with an alpha value of 0.05 is presented in Table 6. The graph of the pairwise means difference in risks for the various causes is presented in Figure 24.

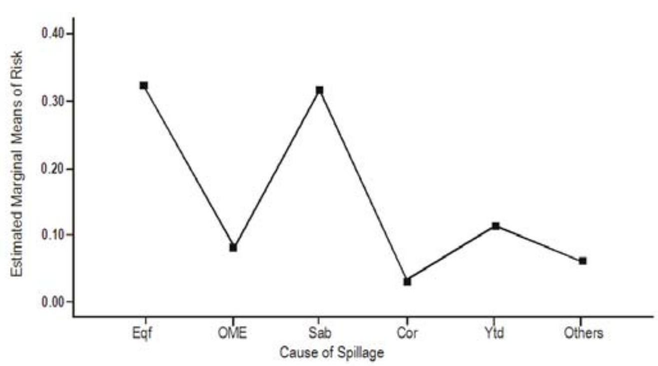


Figure 23. Estimated Marginal Means of Risk on Cause of Spillage

Table 6. Matrix of Pairwise Comparison for Causes of Spillage

	Eqf	OME	Sab	Cor	Ytd	others	Mean Difference
Eqf.	0	0.2359*	0.0045	0.2905*	0.2088*	0.2589*	0.2
OME	-0.2359*	0	-0.2315*	0.0546	-0.0271	0.023	-0.08
Sab	-0.0045	0.2315*	0	0.2860*	0.2043*	0.2544*	0.19
Cor	-0.2905*	-0.0546	-0.2860*	0	-0.0817*	-0.0316	-0.15
Ytd	-0.2088*	0.0271	-0.2043*	0.0817*	0	0.0501	-0.05
Others	-0.2589*	-0.023	-0.2544*	0.0316	-0.0501	0	-0.11

* The mean difference is significant at the .05 level.

The results show that there is a significant difference in the risk level among various causes of oil spill. The results also confirm that equipment failure, sabotage and unknown causes are major sources of risks variation. Corrosion has the least pairwise mean difference of risk. It is therefore necessary to attach high priority to cause control measures related to equipment failure and sabotage. Furthermore, investigations to reveal and isolate the various causes, which are yet-to-be-determined, is highly recommended so as to put adequate preventive measures in place for spills from such sources.

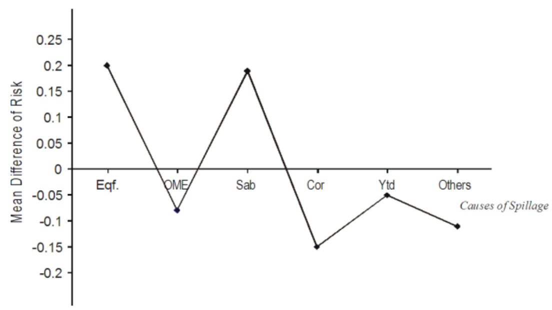


Figure 24. Graph of Pairwise Mean Difference of Risk for Cause

However, both Type (F(5,25) = 2.23), *p* = 0.83) and location ((F(1,25) = 1.59, *p* = 0.22) gave no significant effect on the level of risk at 95% level of confidence. This means that the severity level of oil spillage risk does not significantly vary by the type of spilled oil or by location in which the spill occurs. That is, the impact of oil spillage does not vary significantly by type of spilled oil or across the various locations in which the spill occurs. The marginal risk means for type and location of spillage are presented in Figure 25 and Figure 26 respectively. Post hoc based on LSD pairwise comparisons with an alpha value of 0.05 yielded the results in Table 7 where risks associated with AGO has the highest mean difference followed by crude, condensate and miscellaneous oils. Refined product has the least mean difference.

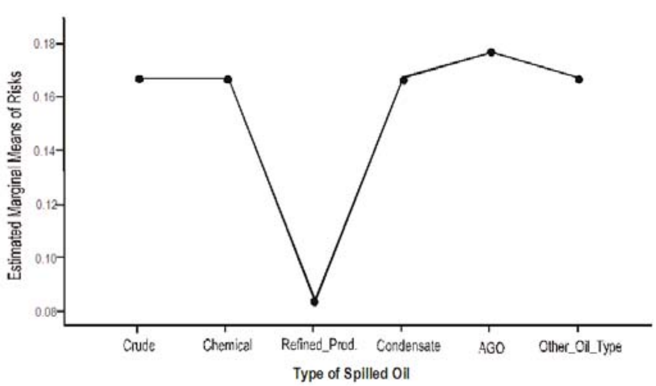


Figure 25. Graph of Estimated Marginal Means of Risk for Type of Oil

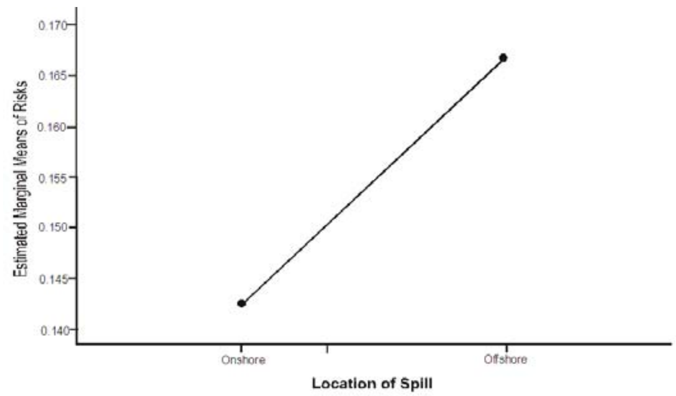


Figure 26. Graph of Estimated Marginal Means of Risk for Location

Table 7. Matrix of Pairwise Comparison for Type of Spill Oil

	Crude	Chemical	Refined_Prod	Condensate	AGO	Others	Mean Diff.
Crude	0	0	0.083*	0	-0.010	0	0.015
Chemical	0	0	0.083*	0	-0.010	0	0.015
Refined_Prod.	-0.083*	-0.083*	0	-0.083*	-0.094*	-0.083*	-0.085
Condensate	0	0	0.083*	0	-0.010	0	0.015
AGO	0.010	0.010	0.094*	0.010	0	0.0104	0.027
Others	0	0	0.083*	0	-0.010	0	0.015

* The mean difference is significant at the .05 level.

4.5 Interaction effect of oil attributes on risk

The interactions (interdependencies) of the various factors (cause, type, location) of oil spillage risk are also explained in Table 8. The interaction between cause and type (cause*type) yields F(25,25) = 9.16 and p = 0.00 which shows a significant cause by type combined effect at 95% level of significance. This shows that the severity of risks associated with a spill of a particular oil type significantly vary by cause of spill and vice versa. The rank of the influence of the association between cause and type on the risk severity is summarized in Table 8 and Figure 27. The results show that miscellaneous oil spill induced by sabotage (Mean = 0.88, S.D = 0.24), have the highest risk factor followed by crude spill from equipment failure (Mean= 0.61, S.D. = 0.22). AGO spill from sabotage is ranked third followed by chemical spill by equipment failure. The result reveals that the combined effect of corrosion and any cause of oil spill are significantly lower. Crude spill from OME is significantly less than that from unknown causes. Measures should be taken to prevent equipment failure, sabotage and unknown causes.

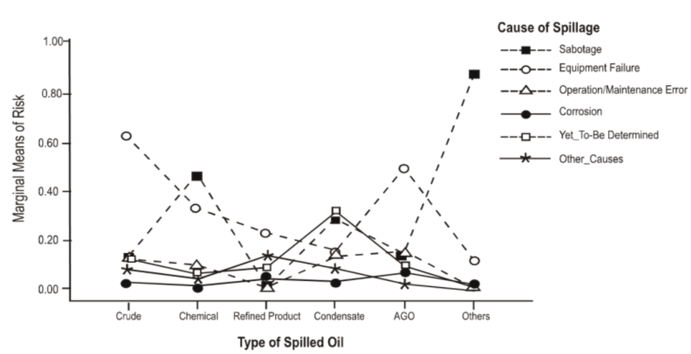


Figure 27. Measurement Risk by Cause of Spillage

Table 8. Mean Risk Ranking of Cause and Type Interaction Effect

Rank	Cause by Type Interaction	Mean Marginal Risks	Standard Deviation
1	Sabotage *Other_Oil_Type	0.88	0.24
2	Equipment_Failure * Crude	0.61	0.13
3	Equipment_Failure * AGO	0.50	0.22
4	Sabotage *Chemical	0.47	0.37
5	Equipment_Failure *Chemical	0.33	0.24
6	Yet_to_be_determined *Condensate	0.31	0.11
7	Sabotage *Condensate	0.29	0.11
8	Equipment_Failure * Refined_Product	0.23	0.32
9	Equipment_Failure *Condensate	0.15	0.19
9	Operation_Maintenance_ Error *AGO	0.15	0.03
10	Operation_Maintenance_ Error *Condensate	0.14	0.01
10	Operation_Maintenance_ Error *Crude	0.14	0.02
10	Other_causes*Refined_Product	0.14	0.19
11	Sabotage *Crude	0.13	0.11
11	Sabotage *AGO	0.13	0.00
12	Yet_to_be_determined *Crude	0.12	0.07
13	Equipment_Failure *Other_Oil_Type	0.11	0.16
14	Operation_Maintenance_ Error * Chemical	0.09	0.08
14	Yet_to_be_determined *AGO	0.09	0.05
14	Yet_to_be_determined *Refined_Product	0.09	0.13
14	Other Causes *Crude	0.09	0.07
15	Other Causes *Condensate	0.08	0.02
16	Corrosion *AGO	0.06	0.09
16	Yet_to_be_determined *Chemical	0.06	0.03
17	Corrosion *Refined_Product	0.05	0.06
18	Other_Causes *Chemical	0.04	0.01
19	Corrosion *Condensate	0.03	0.02
19	Corrosion *Crude	0.03	0.01
20	Other_Causes *AGO	0.02	0.03
21	Corrosion *Other_Oil_Type	0.01	0.02
21	Corrosion *Chemical	0.01	0.01
22	Operation_Maintenance_ Error *Other_Oil_Type	0.00	0.00
22	Sabotage *Refined_Product	0.00	0.00
22	Operation_Maintenance_ Error * Refined_Product	0.00	0.00
22	Other_Causes *Other_Oil_Type	0.00	0.00
22	Yet_to_be_determined *Other_Oil_Type	0.00	0.00

There was a significant difference in the level of risk of spillage attributed to the joint effect of cause of spillage and location of spillage ($F(5,25) = 8.198$, $p = 0.00$). The result of the Post hoc comparison with LSD procedure is presented in Table 9 and Figure 4.24. The results reveal that equipment failure (S.D.= 0.24) and sabotage (S.D = 0.17) significantly differs across locations, while a marginal significant difference is noticed with other causes (S.D = 0.073). The risk of corrosion, OME and unknown causes do not significantly vary by location. Moreover, the type/location interaction yielded no significant difference with $F(5,25) = 2.23$ and $p = 0.083$. This means that the impact of spillage of a particular oil type do not significantly differ in the various locations. The marginal means of risk based on combine effect of type/location and cause by location effect is presented in Figure 28 and Figure 29 respectively.

Table 9. Cause by Location Interaction Effect

Cause of Spill	Location of Spill	Mean	Deviation
Equipment_Failure	Onshore	0.199	0.243
	Offshore	0.442	
Operation_Maintenance_ Error	Onshore	0.077	0.017
	Offshore	0.094	
Sabotage	Onshore	0.402	0.171
	Offshore	0.231	
Corrosion	Onshore	0.030	0.0
	Offshore	0.030	
Yet_to_be_determined	Onshore	0.121	0.018
	Offshore	0.103	
Other_Causes	Onshore	0.025	0.074
	Offshore	0.099	

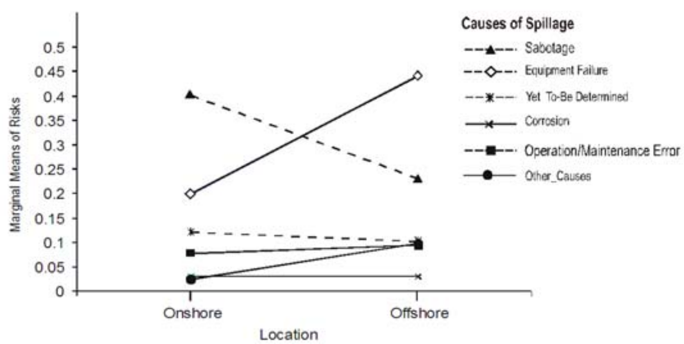


Figure 28. Graph of Cause by Location Marginal Means of Risk

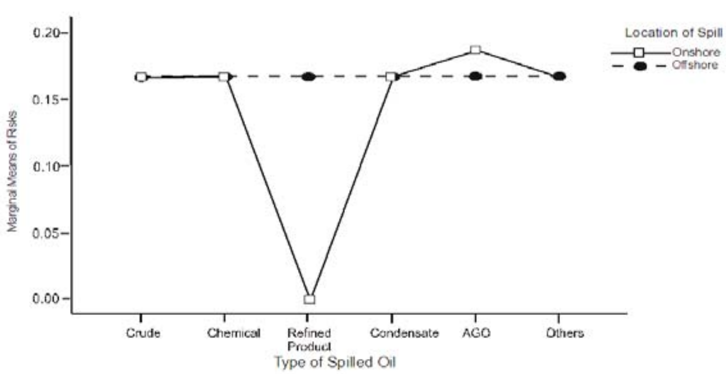


Figure 29. Graph of Marginal Means of Risks Based on Type*Location

In summary, the results of the statistical analysis show that spillage resulting from equipment failure, sabotage and unknown sources have the highest impact on community assets. In addition, miscellaneous oil spill induced by sabotage have the highest risk factor followed by crude spill from equipment failure. AGO spill from sabotage is ranked third followed by chemical spill by equipment failure. To reduce the impact of oil spillage, measures should be taken to prevent equipment failure, sabotage and unknown causes. Adequate response measures should be put in place to respond to crude and miscellaneous oil spill in both locations.

5 Conclusion

This paper presents the findings from the experimental study of a neuro-fuzzy-genetic hybrid model to predict the severity of risks associated with oil spillages. The system utilizes the advantages of NN, FL and GA. The NN component offers the advantages of adaptation, parallelism, fault tolerance and generalization. The GA component optimized the weights of the NN by providing optimal set of parameters for training the NN. The FL provides inference mechanism under cognitive uncertainty by modeling imprecise and vague knowledge, and provides evaluation and membership functions for the GA and NN. The hybrid platform consists of a 2-layered feed forward NN as the central component. The GA uses a hybrid encoding system (real and binary encoding) with the fitness proportionate selection strategy and a 1-point crossover to provide optimal parameters for the NN weights. Its inference engine is driven by ANFIS based on Mamdani's reasoning mechanism. The attributes of the community hazards are held in a multidimensional data model in the Knowledge Base.

Oil spillage data obtained from NOSDRA was used to assess the functionality of the system. The oil spillage indicators were fuzzified using the fuzzy Set {very high, high, medium, low, very low}. The NN and GA provided in 11Ants Analytics were trained with oil spill hazard data. The built model yielded an accuracy of 97.35% and average F-measure of 0.97 with the test dataset. The results also show ranking of the indicators to the NN based on their contributions (importance value) to the target variable. Matrix Laboratory (Matlab) software was used to invoke the parameters provided in the 11Ants Model to extract useful patterns and build ANFIS rules from the dataset. Microsoft Excel, SPSS and GraphPadInstat were used to generate the statistical data required for decision-making. A total of 940 patterns and 329 rules were extracted into knowledge base. The results show that the effectiveness of GA in the training of NN with reduced number of generations thereby drastically minimizing the cost of computation. Pruning of rules based on confidence, support, and number of indicators in the antecedent part of the rule resulted in 73 rules in the ANFIS and for oil spillage risk analysis. The results show that:

a) The severity level of oil spillage risks vary significantly by cause of spill while location and type of spilled oil had no significant effect on oil spillage risk.

b) Oil spillages resulting from equipment failure, sabotage and unknown sources have the highest risks on community assets.

c) AGO spillage has the highest mean risks followed by crude, condensate and miscellaneous oils spills

d) Miscellaneous oil spill induced by sabotage have the highest risk factor followed by crude spill from equipment failure. AGO spill from sabotage is ranked third followed by chemical spill by equipment failure.

e) Offshore locations are more exposed to oil spillage risks

References

[1] Egberongbe, F, Nwilo, P. C. and Badejo, O. T. Oil Spill Disaster Monitoring Along Nigerian Coastline: Marine and Coastal Zone Management – Environmental Planning Issues. Promoting Land Administration and Good Governance 5th FIG Regional Conference Accra, Ghana, March. 2006; 8-11: 1-26.

[2] Guevarra, J. Managing Oil Spill Risks of Transnational Onshore Pipelines. Society of Petroleum Engineers Oil and Gas Conference and Exhibition held in Mumbai, India, January. 2010; 20-22: 1-12.

[3] Foss, B. A. Knowledge Discovery – A Control Theory Perspective. In Kromorowski and Zykow J eds. Principles of Data Mining and Knowledge Discovery. Lecture Notes in Artificial Intelligence. Springer. 1997.
[4] Turrof, M., Chumber, M., Walle, B. V. D. and Yao, X. The Design of a Dynamic Emergency Response Management Information Systems (DERMIS). The Journal of Information Technology Theory and Applications (JITTA). 2004; 5(4): 1-35.
[5] Han, J., Cai, Y. and Cercone, N. Knowledge Discovery in Databases: An Attribute-Oriented Approach. In Proceedings of the 18th Conference of Very Large Databases, Vancouver, British Columbia, Canada August. 1992; 24-27: 547-559
[6] Miller, H.J. and Han, J. Geographic Data Mining and Knowledge Discovery-An Overview. Geographic Data Mining and Knowledge Discovery. In: Miller, H.J., Han, J.(eds),. Taylor and Francis. 2001: 3-32.
[7] Agrawal, R. and Srikant, R. Fast Algorithms for Mining Association Rules. In Proceedings of the 20th International Conference on Very Large Data Bases (VLDB 1994), September 12-15, 1994, Santiago de Chile, Chile. 1994: 87–499.
[8] Whitley, D., Starkweather, T., and Bogart, C. Genetic Algorithms and Neural Networks: Optimizing Connections and Connectivity. Parallel Computing. 1990; 14: 347-361. http://dx.doi.org/10.1016/0167-8191(90)90086-O
[9] Bodenhofer, U. and Herrera, F. Ten Lectures on Genetic Fuzzy Systems Slovak Technical. University, Bratislava. 1997: 1-69.
[10] Zadeh, L.A. Fuzzy Logic, IEEE Computer. 1988: 83-89.
[11] Akinyokun, O. C. Neuro-Fuzzy Expert Systems for Evaluation of Human Resources Performance. First Bank of Nigeria PLC Endowment Fund Lecture Series. Delivered at Federal University of Technology, Akure, December 10, 2002.
[12] Jain, L. and Martin, N. M. 1998. Fusion of Neural Networks, Fuzzy Systems and Genetic Algorithms: Industrial Applications. CRC Press, CRC Press LLC.
[13] Andreou, A.S., Georgopoulos, E. F., Likothanassis, S.D. Exchange-Rates Forecasting: A Hybrid Algorithm Based on Genetically Optimized Adaptive Neural Networks, Computational Economics. 2002; 20(3): 191-210. http://dx.doi.org/10.1023/A:1020989601082
[14] Adamopoulos, A., Georgopoulos E., Manioudakis, G. and Likothanassis, S. 1998. An Evolutionary Method for System Structure Identification Using Neural Networks. Proceedings of the Neural Computation'98, Vienna.
[15] Liopa-Tsakalidi, A., Konstantinidou-Doltsinis, S., Beligiannis, G.N. Studies on the efficacy of extracts of Reynoutriasachalinensis on Cucumber Growth Using the Richards Function and Evolutionary Modelling, Proceedings of the International Conference on Information Technology Interfaces, ITI , art. no. 5974081. 2011: 545-550
[16] Liopa-Tsakalidi, A., Savvas, D., Beligiannis, G.N. Modeling the Richards function using Evolutionary Algorithms on the effect of electrical conductivity of nutrient solution on zucchini growth in hydroponic culture, Simulation Modeling Practice and Theory. 2010; 18(9): 1266-1273. http://dx.doi.org/10.1016/j.simpat.2010.04.013
[17] Papadopoulos, V. D., Beligiannis, G.N., Hela, D.G. Combining Experimental Design and Artificial Neural Networks for the Determination of Chlorinated compounds in Fish Using Matrix Solid-Phase Dispersion, Applied Soft Computing Journal. 2011; 11(8): 5155-5164
[18] Shanthi, D; Sahoo, G. and N. Saravanan. Evolving Connection Weights of Artificial Neural Networks Using Genetic Algorithm with Application to the Prediction of Stroke Disease. International Journal of Soft Computing. 2009; 4(2): 95-102.
[19] Inyang, U. G. 2012. Development of A Knowledge Discovery System for Oil Spillage Risk Management. PhD Thesis, Federal University of Technology, Akure, Nigeria.
[20] Kinlay, J. and Rico, D. Can Machine Learning Techniques be used to Predict Market Direction? The 1,000,000 Model Test. Available from: http://jonathankinlay.com/?p=242. Accessed on September. 2011; 21.
[21] Fuyala, Tom. 11Ants Analytics: Data Mining Made Simple. Briefing paper for the data mining practitioner. Available from: www.11antsanalytics.com. Accessed on September 21, 2011.
[22] Mathworks (2007): Spreadsheet Link™ EX, User's Guide. Available from: www.mathworks.com. Accessed on July 23, 2012.
[23] Iwasokun, G., Akinyokun, O. C., Alese, B. and Olabode, O. Fingerprint Image Enhancement: Segmentation to Thining. International Journal of Advance Computer Science and Application (IJACSA). 2012; 3(1): 15-24.
[24] Raymond, T. Fingerprint Image Enhancement and Extraction. Unpublished Ph.D Thesis School of Computer Science and Sofware Engineering, University of West Australia. 2003.
[25] Ntukekpo, D.S. Spillage: Bane of Petroleum, Ultimate Water. Technology and Environment. 1996.
[26] Fayyad, U.M, Piatetsky-Shapiro, G., and Smyth, P. From Data Mining to Knowledge Discovery in Databases. AI Magazine. 1996; 17(3): 37-54.
[27] Fawcett, T. An Introduction to ROC Analysis. Pattern Recognition Letters. 2006; (27): 861–874. http://dx.doi.org/10.1016/j.patrec.2005.10.010
[28] Chandra, E. and Nandhini, K. Knowledge Mining from Student Data. European Journal of Scientific Research. 2010; 47(1): 156-163.

[29] Tsai, C. and Chen, M. Variable Selection by Association Rules for Customer Churn Prediction of Multimedia on Demand. Expert Systems with Applications. 2010; 37(3): 2006-2015. http://dx.doi.org/10.1016/j.eswa.2009.06.076

[30] Paris, I. H., Affendey, S. L. and Mustapha, N. Improving Academic Performance Prediction Using Voting Technique in Data Mining. World Academy of Science, Engineering and Technology. 2010; (62): 820-823.

[31] Inyang, U. G. and Umoren, M. U. Students' Dataset Mining for Academic Performance Risk Patterns Identification. World Journal of Applied Science and Technology. 2011; 3(2): 57-64.