Journal of Emerging Trends in Engineering and Applied Sciences (JETEAS) 4(2): 173-180
© Scholarlink Research Institute Journals, 2013 (ISSN: 2141-7016)
jeteas.scholarlinkresearch.org

Mining Significant Patterns for Oil Spillage Risk Analysis

[1]O. C. Akinyokun and [2]U. G. Inyang

[1]Department of Computer Science,
Federal University of Technology, Akure, Nigeria.
[2]Department of Computer Science,
Faculty of Science, University of Uyo, Uyo, Nigeria.
Corresponding Author: O. C. Akinyokun

Abstract
In an increasingly technological era, the dependence on oil and gas is on the increase. In order to meet this pressing need, oil exploration with associated spillage have been increasing at alarming rates, which in turn endangers the environment and generate huge oil spillage data repositories. Efforts to deal with oil spillages are inefficient because there is no clear knowledge and understanding of the relationships and inter-dependencies among oil spillage attributes. The purpose of this work is to discover patterns, rules and relationships among oil spillage attributes for oil spillage risks analysis. A hybrid methodology based on neural network, genetic algorithm and Fuzzy Logic for the discovery of patterns in terms of relationships, rules and inter-dependencies from oil spillage dataset is proposed. The system was implemented using 11Ants Analytics and Matlab using data on oil spillage incidences associated with oil exploration activities. A total of 33 comprehensible rules were discovered, the performance of the rules on the test data yielded an average accuracy and precision of 98.68% and 97.90% respectively. The rules are therefore adequate and suitable for the estimation of the oil spillage magnitude and can be deployed into the knowledge base of an expert system for oil spillage risk analysis. This knowledge will therefore guide human experts in the reduction of oil spillage incidences and associated risks.

Keywords: pattern discovery, data mining, oil spillage, risk analysis, rules.

INTRODUCTION

Information Technology (IT) and advance modeling tools and techniques continue to help the society to manage disasters. As nations all over the world, depend on oil and gas for fuelling of cars, generation of electricity and other domestic purposes the demand for oil and gas has also been increasing. In order to meet this pressing need, oil exploitation and exploration with associated spillage have been increasing at alarming rates (Egberongbe *et al.*, 2006, Guevarra, 2010). Failures of key resources, increased urbanization, critical infrastructure dependencies and interdependencies, supply shortages, market demands, changes in policies, terrorism and sabotage, and the heightened movement of people and goods around the world have increased the potential for the occurrence of oil spillage and other hazards (Sapateiro and Antunes, 2009, Nwilo and Badejo, 2010). Like any other type of emergencies, oil spillage is dynamic and changes continuously, thereby making it arduous for planners and responders to produce robust plans towards short term and long term management goals. Hence, the need for an understanding of the nature, sources, impact and responses required to prevent or control their occurrence.

In oil spillage databases, there is a lot more information in terms of hidden patterns, trends or relationships from what can be retrieved using traditional analysis and query methods. Although, statistical approaches quantify the inherent uncertainty that results when one tries to infer general patterns from a particular sample or an overall population, they lack the ability to handle large, complex, noisy and vague dataset. Furthermore, these approaches perform limited search during pattern extraction from databases, thereby producing incomplete and unreliable results, which cannot support effective decision-making. Where traditional analysis techniques fail to uncover hidden patterns from large and diverse datasets, knowledge discovery techniques succeed (Fayyad *et al.*, 1996; Han *et al.*, 1992; Miller and Han, 2001). Knowledge Discovery and Data Mining aims at extracting useful information and patterns from huge amount of data for prediction and modeling (Agrawal and Srikant, 1994).

Neural networks (NNs) are data mining tool used for classification and prediction, and are applied to an increasing number of real life problem of high complexity (Obot, 2007, Akinyokun *et al.*, 2009). They offer ideal solutions to a variety of classification tasks such as speech, character and signal recognition as well as risk assessment and treatment. Genetic Algorithm (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 (Bodenhofer and Herrera, 1997). Fuzzy Logic (FL) is a superset of the conventional Boolean logic that handles the concept of partial truth (Zadeh, 1988, Akinyokun, 2002). While FL performs inference mechanisms under cognitive uncertainty, NNs use learning, adaptation, fault tolerance, parallelism and generalization to process data. The fusion of FL with NN enables systems deal with cognitive uncertainty in a human like manner while GA is a powerful tool for structure and weights optimization of neural networks. It is therefore useful to fuse NNs, FL and GA techniques for offsetting the demerits of one technique by the merits of other techniques (Jain and Martin, 1998). These considerations are covered in this work.

DESIGN METHOD AND MATERIALS

The block diagram of the Knowledge Base (KB) of the system is presented in Figure 1. The design algorithms of hybrid platforms in (Shanthi *et al.*, 2009, Rivero *et al.*, 2009) are studied and modified to suit the design of the KB. As shown in Figure 1, the NN is the central component of the system as in (Jain and Martin, 1998). It receives fuzzified inputs and communicates risks levels associated with oil spillage to the environment as well. The GA component provides optimal set of parameters for training the neural network while the FL acts as a tool for modeling imprecise and vague knowledge, and for the provision of evaluation and membership functions for the GA and NN. Pieces of mined patterns (knowledge) are represented by fuzzy production rules and held in the Multidimensional Data Model (MDM).

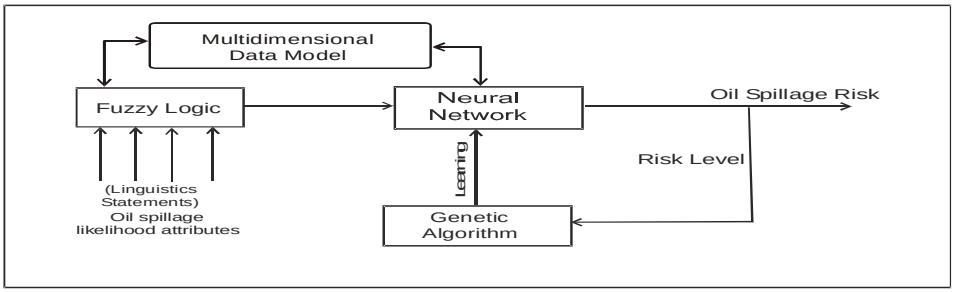


Figure 1: Block Diagram of Knowledge Base of Oil Spillage Risk Management System

The NN is a 2- layered feedforward architecture as in (Cao and Fan, 2010). It has multi-input units (Oil spillage attributes) and multi-output (severity of Oil Spillage Risks) and adopts sigmoid function for neuron activation. The hidden layer is the Oil spillage analysis. NN training by GA was implemented in four stages as in (Shanthi *et al.*, 2009; Cabestany *et al.*, 2009; Shahiar and Kianfar, 2009), namely, initial population generation, selection, crossover and mutation. A gene is represented by a connection weight between the *ith* input node and *jth* hidden node (ω_{ij}) or between the *jth* hidden node to *kth* hidden node (ω_{jk}). A chromosome is encoded as a string of genes { ω_{11}, ω_{12}, ω_{13}, . . ., ω_{mq}, ω_{11}, ω_{21}, . . ., ω_{q1} }, where *m* represents the number of input nodes, *q* represents the number of nodes in the hidden layer and *v* represent the number of output nodes. The set of weights therefore is represented as { ω_{11}, ω_{21}, ω_{31}, . . ., ω_{q1} }. The hybrid-encoding scheme, which begins with binary string encoding and then converted to real value code using Equations 1-2, is proposed.

$$g_i = \begin{cases} 1 \ if \ b_1 = 1 \\ -1 \ if \ b_1 = 0 \end{cases} \quad (1)$$

$$R_i = \frac{g_i}{10} \sum_{t=2}^{m} \left(b_t \times 2^{m-t} \right) \quad (2)$$

where R_i is the real value encoding of the *ith* gene, i=2, 3, …m. g_i is the sign bit of $gene_i$. . Fitness proportionate selection where the fitness function is evaluated for each individual, providing fitness values, which are then normalized. The normalized fitness value is given as:

$$T_i = \frac{f_i}{\frac{1}{N} \sum_{j=1}^{p} pf_i} \quad (3)$$

w*here j=1, 2,3 . . . p;* and f_i is the probability of the *ith* chromosome to be selected for crossover and mutation. The algorithm terminates when 2N iterations are completed. Individuals in the population with the largest fitness value are selected. This chromosome represents the optimal weights of the neural network.

In the fuzzification step, the fuzzy sets of each of the oil spillage indicators are expressed as functions while the elements of the set are mapped to their degree of membership. Triangular and trapezoidal membership functions have been used extensively due to their computational efficiency (Jang *et al.*, 1997). The works of (Dubois and Prade, 1996, Allenotor, 2010) guided the choice of the triangular membership function for this work and gives the general form as shown in Equation 4

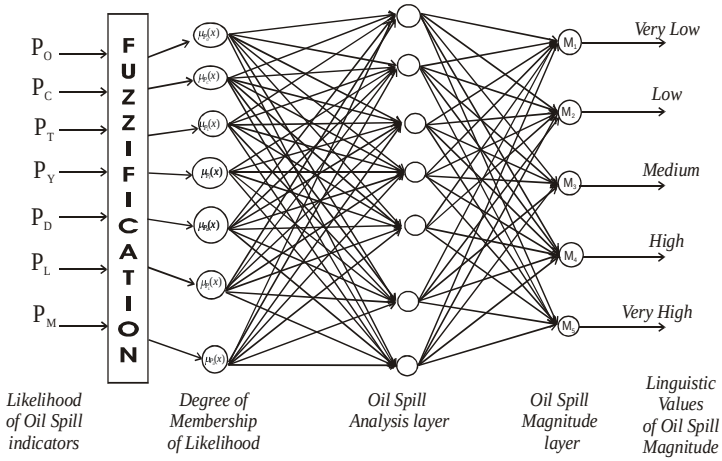
$$\mu(x)=\begin{cases}1, & if \quad x=b \\ \dfrac{x-a}{b-a} & if \ a \le x < b \\ \dfrac{c-x}{c-b} & if \ b \le x < c \\ 0 & if \ c = x \end{cases} \quad (4)$$

where a and c are the parameters governing triangular membership functions; *b* represents the value for which *μ(x)* = 1 and is defined as $b=\frac{(a+c)}{2}$. The actual membership functions of each element in the fuzzy set are derived by using the membership functions to determine the degree of membership (Jang et al., 1997; Akinyokun *et al.*, 2009). This is then mapped to the fuzzy linguistic value of *"very low", "low", "medium", "high"* or *"very high"* specified in the rules. Each linguistic term is a fuzzy set and has its own membership function. Equation 5 is an example of membership function for a linguistic term.

$$\text{Low(x)} = \begin{cases} 0 & if \ x < 0.1 \\ \dfrac{x - 0.1}{0.3} & if \ 0.1 \le x < 0.4 \\ 1 & if \ x > 0.4 \end{cases} \quad (5)$$

Neuro-fuzzy Inference Engine

The Neuro-fuzzy inference engine extracts and evaluates rules and produces fuzzy outputs. The architecture of the Neuro-fuzzy architecture for Oil Spillage Risk Management is presented in Figure 2. The likelihood of oil spill indicators is the input to the neuro-fuzzy inference system.



where P_O represents the probability that a particular type of oil has spilled, P_C represents the probability that an oil spill has a particular cause, P_L represents the probability that an oil spill occurred in a particular location, P_Y represents the probability that an oil spill occurred in a particular year, P_D represents the probability that an oil spill occurred in a particular day, P_M represents the probability that an oil spill occurred in a particular month, P_T represents the probability that an oil spill occurred in a particular time and M is the magnitude of spill.

For decision-making purposes, the output fuzzy sets must be defuzzified to crisp value in the real life domain. The deffuzzification method is the centroid method that returns the centre of area under the curve (Bodea and Dascalu, 2009). This is given as follows:

$$z = \frac{\sum_{i=1}^{n} \alpha_i y_i}{\sum_{i=1}^{n} \alpha_i} \quad (6)$$

where z is the crisp value and can be used for decision-making, α_i is the fuzzy implication (firing strength) of the *ith* rule, $\mu(\alpha_i)$ is the degree of membership of the likelihood of the *ith* indictor of oil spillage and y_i is the consequent of each rule. The crisp value for oil spillage analysis represents the magnitude of oil spillage. The conditional probability or influence of an attribute to the magnitude of spill is derived from the NN training by the GA component as follows:

$$P_i = \frac{\sum_{j=1}^{h} \omega_{ij}}{h} \quad (7)$$

where *h* is the number of hidden layer neurons, P_i is the conditional probability of the *ith* attribute. The combined probabilities (P_{ijk}) of a spillage are denoted in *v* x *w* matrix. There are *u* such matrices; each matrix represents a type of spilled oil. *v* is the number of possible causes of spillage while *w* is the various locations. The matrix of the combined probabilities of the *ith* type of spilled oil (T_i) is represented in Equation 8.

$$T_i = \begin{bmatrix} t_{11} & t_{12} & \cdots & t_{1w} \\ t_{21} & t_{22} & \cdots & t_{2w} \\ t_{11} & t_{12} & \cdots & t_{1w} \\ \cdots & & & \\ t_{v1} & t_{v2} & \cdots & t_{vw} \end{bmatrix} \quad \ldots \quad (8)$$

i= 1,2,3 … u; j = 1, 2, 3, . . . v; k = 1, 2, 3, . . . w;

KNOWLEDGE MINING

Spillage Dataset Pre-Processing and Dimension Reduction

The knowledge mining was performed using data collected from a National Oil Spill Detection and Response Agency (NOSDRA), Nigeria. It represents Oil Spill incidences associated with oil exploration activities from January, 2005 – November, 2011 in Akwa Ibom State, Nigeria. 1008 incidences were collected. Each tuple consists of *Date_of_Occurence, Location, Time, Type_of_Spilled_Oil (Type), Cause_of_Spillage (Cause)* as input indicators while Magnitude of spill is the target variable. Data items with any of these fields missing were filled-up while those with more than one fields missing were excluded from our records. Since each *Date_of_Occurence* are not likely to be frequent in the dataset, no useful knowledge, in terms of association and patterns may be derived from it. We therefore split *Date_of_occurrence* field into three separate components, *Day_of_Occurence (Day), Month_of_Occurence (Month)* and *Year_of_Occurence (Year)*. 11Ants Analytics was used for dataset splitting and input rank analysis, the dataset splitting was based on the holdout method where 85% representing 857 data items and 15% (151 data items) was the Train and Test dataset respectively while the input rank analysis was based on sensitivity analysis (Han and Kamber, 2001). The result is presented in Table 1.

Table 1: Relative Importance of Oil Spillage Indicators

Rank	Oil Spillage Indicators	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

The result depicted in Table 1 shows that *Type* has 0.362 as weight while Cause and Location 29.6% and 19.3% contribution to the Magnitude of Spillage respectively. The cumulative contribution of Day, Month, Time and Year is 14.9%, hence is insignificant and noisy in the estimation of oil spillage risks. We therefore eliminate Year from the training of the NN while the other insignificant indicators were basis for rule pruning. Table 3 gives a description of the possible values and the codes of each significant input indicator.

Table 2: Oil Spillage Indicators with their values

Input Indicators	Values	Codes
Location	Onshore	ON
	Offshore	OFF
Cause	Operational/Maintenance Error	OME
	Sabotage	Sab
	Equipment Failure	Eqf
	Corrosion	Cor
	Yet to be determined	Ytd
	Others	Oth
Type	Refined Product	Re
	Crude	Cr
	Chemical	Ch
	AGO	AGO
	Condensate	Con
	Others	Oth
Magnitude	Very Low	VL
	Low	LO
	Medium	ME
	High	HI
	Very High	VH

Pattern Discovery and Analysis

The system design is implemented in an environment characterized by MatLab 7.7.0 as the front-end tool for the implementation of the system; Microsoft Excel and Microsoft Access were the database management tools. Pattern discovery from the trained NN was performed in three stages using the modified Apriori Association rule mining algorithm. It involved the identification of frequent *k-antecedentset*, formulation of multidimensional rules and pruning less interesting rules. Figure 3 gives the lattice of indicators in a 3-dimensional data cube. The base of the cuboids contains the required antecedent size of three, which are used in producing interesting multidimensional rules in the form:

$$\text{Cause}(X, C_i) \wedge \text{Type}(X, `T_j') \wedge \text{Location}(X, `L_k') \Rightarrow \text{Magnitude}(X, `\mu(X))$$

where C_i is the *ith* Cause of Spillage, T_j is the *jth* Type of spilled oil, L_k is the *kth* Spillage location and $\mu(X)$ is the linguistic value specifying the degree of membership of the magnitude of spillage. 329 rules were extracted based on uniform minimum support and minimum confidence thresholds of 10% and 15 % respectively. The clustering of rules based on the size of antecedent of the rule is presented in Figure 4:

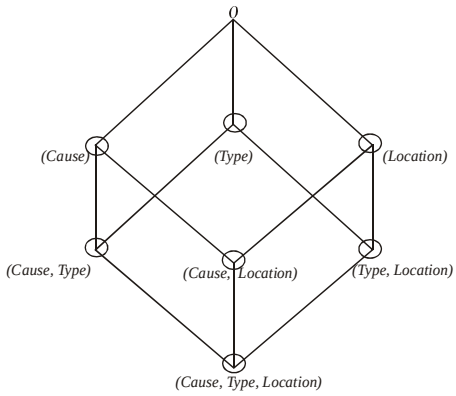


Figure 3: Lattice Cuboids of Oil Spillage Indicators

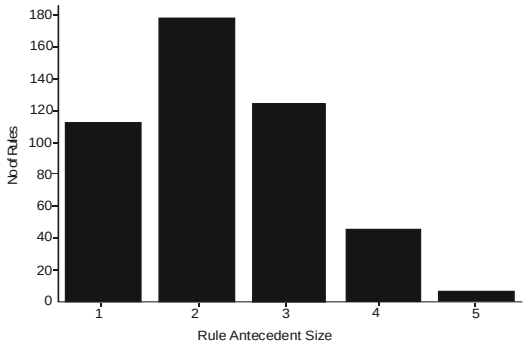


Figure 4: Antecedent Size Clusters of Rules

As shown in Figure 4, 77 rules have single antecedent, 110 rules are with antecedent size of 2. 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. The rule analysis based on support and confidence is presented in Figure 5 and Figure 6 respectively. In Figure 6, the support of rule decreases as the number of rules increases while in Figure 7, the confidence of rules also decreases as number of rules increases. The result shows that the confidence of a rule is higher than its support while an increase in the number of rules causes a decrease in the rule support and confidence.

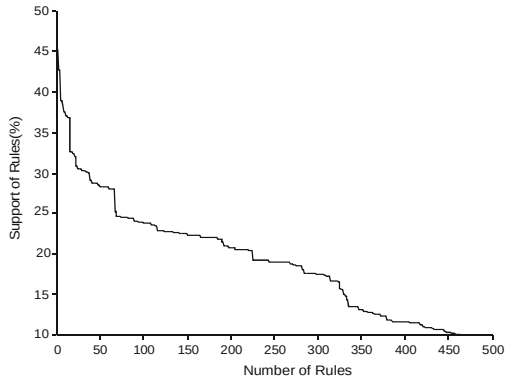


Figure 5: Graphical Analysis of Rule Support

Pruning of weak and uninteresting rules was performed in the following stages; firstly rules below the uniform minimum support and minimum confidence thresholds of 10% and 15% respectively were not extracted. Though rules with antecedent size 1 and 2 have high support and confidence, they may occur by chance and may be misleading (Chandra, 2010, Inyang and Umoren, 2011). We therefore delete rules whose antecedent size is less than 3 in the second stage. Thirdly, rules with antecedent part consisting of any insignificant oil spillage indicators

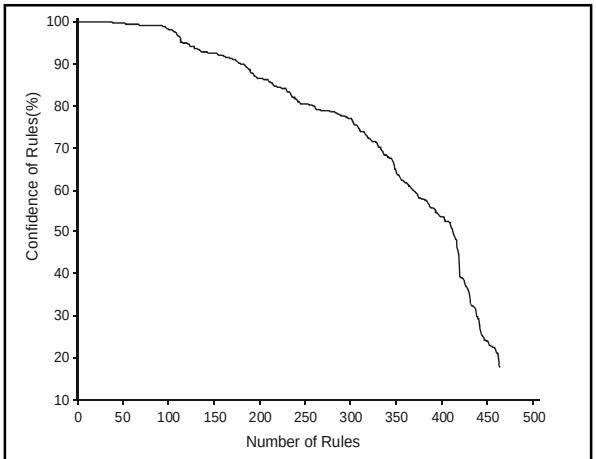


Figure 6: Graph of Confidence of Rules

were also eliminated. Finally, rules with all the attributes in the base of the cuboids depicted in Figure 3 were retained. The resultant ruleset contains 33 rules and the first 10 rules are presented in Table 3. An example of the interpretation of the rules is as follows:

Rule 5: IF (Cause=OME) *and* (Type=Con) *and* (Location=On) *Then* (Magnitude=Low)

Table 3: Ruleset for Oil Spillage Risk Analysis

S/N	Rule Antecedent			Rule Consequent	Rule Support (%)	Rule Confidence (%)
	Cause	Type	Location	Magnitude		
1	Cor	Re	On	ME	24.15%	99.04%
2	Eqf	Cr	Off	LO	23.86%	85.35%
3	Sab	Re	On	ME	22.40	100%
4	Eqf	Ch	Off	LO	19.43%	81.38%
5	OME	Con	On	LO	15.20%	69.23%
6	Oth	Re	Off	VH	20.97%	83.88%
7	Eqf	Oth	On	ME	22.22%	94.22%
8	Sab	Ch	On	VL	27.72%	63.76
9	Ytd	Cr	Off	HI	18.88%	85.11%
10	Ytd	Re	On	HI	22.97%	97.56%

Pattern Evaluation

The ruleset was subjected to the test dataset to assess its performance in the classification of oil spillage magnitude. The summary of the predications is presented in Table 4, the patterns leading to, Very High and Low have 100% correct prediction. 3(10%) instances that belonged to Medium category was 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 classification confusion matrix depicted in Table 5.

Table 4: Summary of All 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**

The classification confusion matrix in Table 5 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 for Very High and Low categories each is 100%. The precision is 100% for Very High and Low categories, 95% for High category and 96.43% respectively. The specificity for all categories is very high which implies an insignificant False Positive rate (FPR) and high True Positive Rate (TPR). Also, the F-measure on the average, is 0.974. The results of accuracy, precision, sensitivity, specificity and F-measure prove that the ruleset is efficient and effective for classification and prediction of the magnitude of Oil Spillage. The Receivers Operating Characteristics (ROC) curve of the ruleset is presented in Figure 7.

Table 5: Confusion Matrix of Rule set Evaluation

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

In Figure 6, each point is a coordinate (FPR, TPR) and represents an instance of a predicted fuzzy linguistic label of magnitude of spill. The ROC curve in Figure 7 is closer to the perfect classification coordinate (0,1), which indicates an excellent performance of the model (Fawcett, 2006). It also shows that any increase in sensitivity, will be 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.

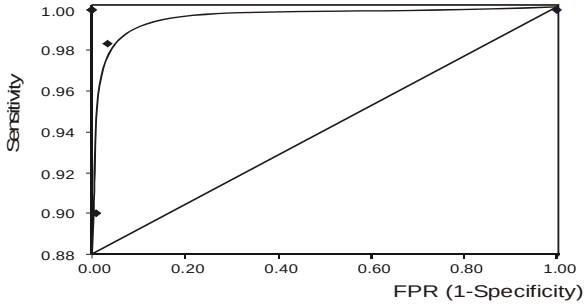


Figure 7: ROC Curve of Ruleset

CONCLUSION

The research reported in this paper proposes a hybrid methodology based on NN, GA and FL for the discovery of patterns in terms of relationships, rules and inter-dependencies from oil spillage dataset. The model integrates the advantages of NN, FL and GA thereby compensating for their disadvantages. The NN component offers the advantages of adaptation, parallelism, fault tolerance and generalization while the GA component is used to optimize the weights of the neural networks 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 providing evaluation and membership functions for the GA and NN. A total of 33 strong and comprehensible IF … THEN rules were extracted. The performance of the rules on the test dataset is very satisfactory. The work demonstrates the effectiveness of the hybridization of NN, GA and FL. It also provides clear knowledge in form of patterns and rules, and ranking of spillage attributes based on their relative importance to the magnitude of spillage. This knowledge will guide human expert in the reduction of oil spillage incidences and associated risks. As further research, these rules should be deployed in the knowledge base of an expert system driven by Adaptive Neuro Fuzzy Inference System for the analysis and assessment of oil spillage risks. In addition, GA could also be used in the determining the optimal structure of the NN model for improved performance.

REFERENCES

Agrawal, R. and Srikant, R., 1994. Fast Algorithms for Mining Association Rules. In Proceedings of the 20th International Conference on Very Large Data Bases (VLDB 1994), Santiago de Chile, Chile. September 12-15, 1994, pp: 487–499.

Akinyokun, O. C. 2002. 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.

Akinyokun, O. C., Obot, O. and Uzoka, FME. 2009. Application of Neuro-Fuzzy Technology in Medical Diagnosis: Case Study of Heart Failure, In Proceeding of World Congress on Medical Physics and Biomedical Engineering, Munich, Germany, pp: 301-304.

Allenotor, D. 2010. A Fuzzy Real Option Model for Pricing Grid Compute Resources. Unpublished Ph.D Thesis, Department of Computer Science, University of Manitoba, Winnipeg.

Bodea, C. N. and Dascalu, M. I.. 2009. Modeling Research Project Risks with fuzzy Maps. Journal of Applied and quantitative Methods. 4(1):17-30.

Bodenhofer, U. and Herrera, F., 1997. Ten Lectures on Genetic Fuzzy Systems Slovak Technical. University, Bratislava. pp:1–69

Cao, X and Fan, S. (2010): The Synthetic Assessment Modeling of Ships' Oil Spill Risk Based on Fuzzy Neural Network. In Proceedings of International Conference on Intelligent Systems and Applications (ISA). Wuhan, China, May 22-23, pp:368-371

Cabestany, J., Rojas, I. and Gonzalo, J. 2009. Advances in Computational Intelligence. In Proceedings of the 10th International Work-Conference on Artificial Neural Networks: Part I: Bio-Inspired Systems: Computational and Ambient Intelligence. Salamanca, Spain, June 10-12,. pp:433-448

Chandra, E. and Nandhini, K., 2010. Knowledge Mining from Student Data. European Journal of Scientific Research. 47(1): 156-163.

Dubois, D. and Prade, H. 1996. What are Fuzzy Rules and How to Use Them? Fuzzy Sets Systems, 84(2), pp:169–185.

Egberongbe, F, Nwilo, P. C. and Badejo, O. T. 2006. 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 8-11. pp:1-26

Fawcett, T. (2006): An Introduction to ROC Analysis. Pattern Recognition Letters. (27), pp 861–874.

Fayyad, U. M, Piatetsky-Shapiro, G. and Smyth, P.(1996): Knowledge Discovery and Data Mining: Towards a Unifying Framework. In Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96), Portland, Oregon, August 4-8. AAAI press, pp:1-30.

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

Han, J. and Kamber, M. 2001. Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers. San Francisco

Han, J., Cai, Y. and Cercone, N., 1992. Knowledge Discovery in Databases: An Attribute-Oriented Approach. In Proceedings of the 18th Conference of Very Large Databases, Vancouver, British Columbia, Canada August 24 - 27, pp:547-559

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

Jain, L. and Martin, N. M. 1998. Fusion of Neural Networks, Fuzzy Systems and Genetic Algorithms: Industrial Applications. CRC Press, CRC Press LLC

Jang, J. S., Sun, C.T. and Mitzutani, E.1997. Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence, Prentice Hall Inc., Upper Saddle River, USA

Miller, H.J. and Han, J., 2001. Geographic Data Mining and Knowledge Discovery-An Overview. Geographic Data Mining and Knowledge Discovery. In: Miller, H.J., Han, J.(eds),. Taylor and Francis, pp:3-32.

Nwilo, P. C. and Badejo, O.T., 2010. Sustainable Management of Oil Spill Incidents along the Nigerian Coastal Areas. Electronic Conference on Sustainable Development Information Systems, CEDARE2010. pp1-13

Obot, O. U., 2007. A Neuro-Fuzzy Expert System Model for the Diagnosis and Therapy of Cadiosvascular Diseases. (A Case Study of Heart Failure). Unpublished PhD Thesis, Submitted to the Department of Computer Science, Federal University of Technology, Akure ,Nigeria.

Rivero, D., Dorado, E., Efernandez-Banco, E. and Pazos, A., 2009. A Genetic Algorithm for ANN Design, Training and Simplification. In Proceedings of the 10th International Work-Conference on Artificial Neural Networks: Part I: Bio-Inspired Systems: Computational and Ambient Intelligence, pp:391-398.

Sapateiro, C and Antunes, P., 2009. An Emergency Response Model towards Situational Awareness Improvement. In Proceedings of the 6th International Conference of Information Systems for Crisis Response and Management ISCRAM2009 Conference, Gothenburg, Sweden, May 10-13, 2009. (eds. Landgren, J. and Juls., S.), pp:67-79

Shahriar, A. and Kianfar, J. 2009. A Hybrid Neuro-Genetic Approach to Short Term Traffic Volume Prediction. International Journal of Civil Engineering. 7(1): 41-48.

Shanthi, D; Sahoo, G. and Saravanan N., 2009. Evolving Connection Weights of Artificial Neural Networks Using Genetic Algorithm with Application to the Prediction of Stroke Disease. International Journal of Soft Computing 4(2): 95-102.

Zadeh, L.A. 1988. Fuzzy Logic, IEEE Computer, 83-89.