# Detecting Fraud Transactions Using Radial Basis Function-Artificial Neural Network

**Alabi I. O.1,\* , R. G. Jimoh 2**

1 Department of Information and Media Technology,

Federal University of Technology, Minna, Nigeria.

2 Department of Computer Science, University of Ilorin, Ilorin, Nigeria.

[Isiaq.alabi@futminna.edu.ng](mailto:Isiaq.alabi@futminna.edu.ng) (+234)8023130292, [jimoh\_rasheed@unilorin.edu.ng](mailto:jimoh_rasheed@unilorin.edu.ng) (+234)8168421369

------------------------------------------------------------------------------------------------------------------------------------------

**ABSTRACT**

The ubiquitous cases of abnormal transactions with intent to defraud are worrisome and are assuming a heightening proportion. Fraud detection and prevention mechanisms are concurrent processes in combating fraud malaise. The hitherto traditional methods of fraud detection are not enough to deal with the present level of sophistry with which financial fraudulent acts are perpetrated. In this work, an architecture that enhances fraud detection using an ensemble radial basis function and artificial neural networks was designed. This research provides a dynamic red flags of previously susceptible transactions that were properly classified to distinguish new cases. This approach is rather proactive than a reactive measures to fraud detection and would found relevance among corporate business professionals and government agencies, thereby minimizing the time and cost of fraud detection.

**Keyword:**Financial fraud detection,Basis radial function network, Artificial neural network, BRF-ANN model framework, Detecting fraud transactions.

1. **INTRODUCTION**

Fraud is innately criminal, it is an undesirable act of obtaining something of value that belongs to someone else under false pretence; recently, fraudulent activities are assuming frightening and sophisticated dimensions, whereby individuals, governments, agencies and businesses are deprived of substantial material benefits, yet detecting and preventing fraud is not a simple task. Much research efforts and other resources have been deployed but the syndrome persists. Then, a recurring question is how can this behaviour be eliminated or nipped before it materializes?

Gupta and Gill (2013) posit that fraud is as old as humanity and could assume an unlimited variety of forms. Mass media is replete with news of financial scam that is adversely affecting economies of nations and multinationals worldwide of which many people have been deprived of substantial amount of valuables. However, the development of new technologies has also provided further ways in which criminals may commit fraud (Bolton & Hand, 2002). Nonetheless, this undesirable trend must be curb to a barest minimum (cite).

Fraud detection and prevention are concurrent processes in combating fraud malaise; while fraud detection is the **s**potting of false claim, act or data; fraud prevention is the bursting of the crime before it materializes, by raising alarms thus preventing it from occurring. Because fraud is an adaptive crime, it requires special methods of intelligence gathering and data analysis in order to detect and prevent frauds (Nigrini, 2011).

Several methods and techniques had been proposed to detect and expunge fraudulent transactions, yet some problems persist. For example, most of the existing fraud detection systems do not timely alert when the fraud is committed, until some later time when it was almost too late to track offenders, perhaps due to their computational complexities or other deficiencies. In some situations where a fraud detection system alerts, it might be too rigid to keep pace with the current fraud trends, whereas fraud detection models must be dynamic to encompass emerging and future fraud options (Brause, Langsdorf, & Hepp, 1999).

This research work formulated a hybrid Radial Basis Function – Artificial Neural Network (RBF-ANN) model, a classification approach to explore large data in detecting fraud activities (such as money laundering, electronic commerce scam, dubious insurance, mortgage and health claims, etc.). Though, *RBF*s (a variant of ANN) are classical approximation functions for (non-) linear models, or a neural (single or multi-layer) network, due to their powerful convergence properties (Buhmann, 2004). Appositely, each RBF hidden unit executes a radial activated function and the output unit executes a weighted sum of hidden unit outputs.

In this work, we use the ensemble inductive models (for the same domain), to obtain better prediction quality, thereby reinforcing strengths and compensating weaknesses.

1. **RELATED WORKS**

There is a limited published work in fraud detection using BRF-ANN, yet various data mining techniques have been applied in financial fraud detection, such as neural networks (Cerullo & Cerullo, 1999), logistic regression models (Bermúdez, et al, 2008), the naïve Bayes method (Viaene, 2004) and decision trees (Kirkos, et al 2007). Broadly speaking, the techniques used for fraud detection fall into two major classes: statistical techniques and artificial intelligence.

Literatures revealed that statistical modelling techniques such as logistic regression, linear and quadratic discriminant analysis are widely used for modelling and prediction purposes, but their predetermined functional form and restrictive (often unfounded) model assumptions limit their usefulness. Statistical method such as Linear Discriminant Analysis (LDA) is the oldest and most common statistical tool in handling classification problems (Lee et al., 1999) that has been employed in fraud detection and credit scoring (Desai et al., 1996; Daskalaki et al., 2003; Lee et al., 2006).

On the other hand, supervised neural networks such as fuzzy neural nets, and combinations of neural nets and rules, have been extensively explored and used for detecting fraud in mobile phone networks and financial statement fraud (Green & Choi, 1997; Estevez & Perez, 2006). Also, the role of neural networks was to provide general and efficiently scalable parameterized nonlinear mappings between a set of input variables and a set of output variables (Bishop, 1995). Neural networks have shown to be very promising alternatives for modelling complex nonlinear relationships (Desai et al., 1996; Lacher et al., 1995; Lee et al., 1996; Mobley et al., 2000; Piramuthu, 1999; Salchenberger et al., 1997; Sharda & Wilson, 1996).

Maranzato, Pereira, Naubert, and Lago (2010), and Wilson (2009) used the logistic regression method as a tool to discriminate fraudulent actions from legitimate actions for insurance companies and e-commerce. Field and Hobson (1997) presented a neural network based fraud management technique based on profiling techniques. Fawcett and Provost (1997) presented a rule-based tool for fraud detection using a series of machine learning methods.

Fraud detection with Bayesian networks was presented in Ezawa (1996), where the author used a Bayesian network as a normative expert system. The author focused on the ratio of fraud cases to legitimate cases with different misclassification costs in determining correct classification.

Bishop (1995) and Hippert, et al. (2005) stated that some practical issues persist when implementing neural networks, such as the impact of the initial weight choice, how to set the weight decay parameter, and how to fit the noise in the training data. Other defects include long learning time, over-fitting error and black box characteristics (i.e. lack of explanatory power) Nonetheless, neural network role is crucial in providing general and efficiently scalable parameterized nonlinear mappings between a set of input and output variables (Bishop, 1995).

Neural networks have shown to be very promising alternatives for modelling complex nonlinear relationships (Desai et al., 1996; Lacher et al., 1995; Lee et al., 1996; Mobley et al., 2000; Piramuthu, 1999; Salchenberger et al., 1997; Sharda & Wilson, 1996).

Further, Chen et al. (2006) used Support vector machine (SVM) and Neural networks to show that when the data records are small, SVM can offer a better performance than neural network does. But, SVM tries to find a linear optimal hyper-plane to separate positive and negative cases by solving a quadratic optimization problem, but the data are not often linearly separable. In order to enhance the feasibility of the linear separation, the input space is transformed via a non-linear mapping into a higher dimensional feature space by using a kernel function (Steinwart & Christmann, 2008).

While some literatures (Daskalaki et al., 2003; Hung et al., 2006; Coussement & Poel, 2008; Hilas & Mastorocostas, 2008) favoured single classifiers like Neural net, Decision Tree, SVM, and other learning methods There are also some good empirical evaluations that strongly support ensemble methods (Kim et al., 2003; Caruana & Niculescu-Mizil, 2006). Because of the diversity among base classifiers, Kim, (2009) and Bauer and Kohavi (1999) have noted that ensemble methods generalizes better by combining base classifiers.

Researchers that have worked on fraud risk indicators include: Apostolou, Hassell, Webber & Sumners (2001); Hackenbrack (1993); Loebbecke , Einning & Willingham (1989); Majid, Gul, & Tsui ( 2001); Mock & Turner (2005); Moyes (2007), and Smith , Omar, Idris & Baharuddin (2005), their results indicated that fraud risk indicators are the most important factor in fraud detection. These indicators, also called red flags raise auditors’ sensitivity to the possibility of fraud (Krambia-Kapardis et al., 2010).

Research has shown that by combining two types of models, we can improve the overall detection rate of the system without compromising the benefits of either detection method. Based on these premises, the hybrid Radial Basis Function and Artificial Neural Network (RBF-ANN) was proposed in this work with higher predictive capability and accuracy.

The proposed BRF-ANN model interacts with online or operational transactions rather than mere historical warehoused data to analyse transactions in order to detect fraud and trigger timely alerts as necessary. A feed-forward radial basis function neural network with three-layers was introduced by Ghosh and Reilly (1994), the results showed more accuracy with shorter training time, but are slower on the application of new instances.

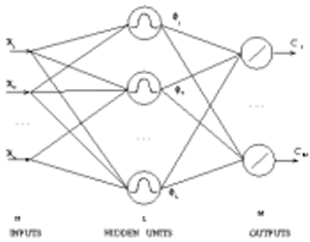
In this research, RBF-ANN is to be extended, optimized and applied to monitor and detect fraud risk indicators in a less complex, reliable and faster computations. An intelligent fraud monitoring and detection system based on this model over Transmission Control Protocol/Internet Protocol (TCP/IP). The results would indicate that the fraud detection system based on mining of the operational data in this manner is realizable, resilient and robust.

1. **RADIAL BASIS FUNCTION-ARTIFICIAL NEURAL NETWORK (BRF-ANN) LEARNING MODEL**

Early application of *Rbf* was in fire detection (Buhmann (2003)), where several measured parameters (the flame color, spectrum, intensity, direction, etc) were used to model a fire detector device. Also, in nuclear physics, *Rbf* was used to interpolate the data that come from the raster of a screen of a robot’s eye (Eckhorn, 1999; and Kremper, Schanze & Eckhorn, 2002). Other applications include *Rbf* as Interpolation (Broomhead & Lowe, 1988; and Matej & Lewitt, 1996), time-series modelling, Classification and Control engineering (Sanner & Slotine, 1996).

In this effort, *Rbf* was used as a learning model that fits a variety of previously unknown data objects and to make the model general as much as possible. Advantages of *Rbf* over other approximating algorithms (e.g. partial differential equations, PDEs) is its convergence power and its easy formulation of interpolants, remarkably resilient against irregular data distributions, besides, *Rbf*s have a variation-diminishing property to guarantee the approximations smoothness. (Buhmann, 2003).

Besides, RBF is a variant of ANN, where each hidden unit executes a radial activated function and the output unit executes a weighted sum of hidden unit outputs.

 Figure 3.4: A BRF Learning (Redraw the figure with weights)

Early application of *Rbf* was in fire detection, where several measured parameters (the flame color, spectrum, intensity, direction, etc) were used to model a fire detector device. In nuclear physics, *Rbf* has been used to interpolate the data that come from the raster of a screen of a robot’s eye (Eckhorn, 1999; and Kremper, Schanze & Eckhorn, 2002). Other applications include RBF as Interpolation (Broomhead & Lowe, 1988; and Matej & Lewitt, 1996), time-series modelling, classification and control engineering (Sanner & Slotine, 1996).

Advantages of *Rbf* over other approximating algorithms (e.g. partial differential equations, PDEs) is its convergence power and its easy formulation of interpolants, remarkably resilient against irregular data distributions, above all RBFs have a variation-diminishing property to guarantee the approximations smoothness. (Buhmann, 2003).

Typically, *Rbf* neural network is otherwise called ***n-h-m* neural network** because ithas *n*-inputs, *h* hidden nodes and *m* outputs, (See Fig. xx) The input vector, **x**= (*x1, x2,..,xn*)ᴛ ∈ R*n*, in which *ω* ∈ R*n×m* is the output weight matrix, the network output is:

Where, φ***i*(*x*)** is the activation function of hidden node *i*. The RBF network uses a variety of activation functions for hidden nodes (e.g. Thin-plate-spline (*x*2log(*x*)); Gaussian (e*-x*); multi-quadric (1+ *x*)½; inverse multi-quadric (1+ *x*)-½; and Cauchy (1+ *x*)-1), of which Gaussian function is frequently used of which this research also adhered to, i.e.

Where, **c***i* = (*x1, x2,..,xn*)ᴛ is the centre of the hidden node *i*; **δ*i*** is the constant extension of hidden node *i*. Another beauty of *Rbf* is that its response decreases (or increases) monotonically with distance from the centre and its ability to fit different functions, derived from the freedom to choose different values for the weights.

Suppose *h* initial cluster centres are created from the samples, and the first *h* of them are selected by default. *ci* (a scalar value) is the centre of cluster *i*; its corresponding mean square error is σi. The distance norms from all the sample inputs to the initial cluster centres are defined as:

Classify sample input **x** (vector) according to the minimum distance principle, then recalculate the new cluster centre of each cluster. When the first Di(**x**) = min Di(**x**) appears,

When the next Di(**x**) = min Di(**x**) appears,

For the rest,

Here, *v* is the learning rate of the winner cluster centre, *ξ* is the ratio of the penalty rate of the cluster centre *ρ* to *v*. then the *mse 0r MeR* of the winner cluster is:

Here, *μ* is a constant close to 1, but smaller than 1, usually taken to be 0.999. Learning rate of a further cluster centre is:

If this equation converges, the iteration ends. If it does not converge, the distance between the samples and the cluster centres has to be circulated. Order *k* = *k* +1, then re-cluster and calculate a new cluster centre. At the end of the iteration, we obtain the optimal cluster centre by removing empty ones and remove a cluster centre, if it is out of the data collection range.

After determining the cluster centre and extension constant of each hidden node, we can obtain the output weight vector ***ω*** = (*ω1, ω2,…, ωn*)**ᴛ** using the least squares method. However, it was noted that it would be erroneous to divide the sum-squared-training error by the number of patterns in order to estimate the noise variance, since some degrees of freedom have been used in fitting the model. The output of hidden node *i* is:

Then the output matrix of the invisible layer is:

Given the final teacher signal **y**; assuming the approximation error is:

Using the least square method,

In this algorithm, when the centre cluster begins, the penalty rate for the competition centre is fairly high. This way, redundant competition centres will move away from the dataset and become empty nodes that will not be learned anymore. As the computation progresses, the competition rate among cluster centres diminish too, so should the penalty rate for the competitions. It is obvious that the maximum number of iterations of the algorithm will not be larger than the number of the training samples.

1. **MODEL SELECTION**

Neural networks are renowned for its flexibility in modelling complex and non-linear models (Stern, 1996; Anderson & Rosenfeld, 1998), its downside is the impact of the initial weight selection/determination, how to set the weight decay parameter, and how to fit the noise in the training data. Other defects include long learning time, over-fitting error and black box characteristics (i.e. lack of explanatory power) (Bishop, 1995; Hippert et al., 2005).

In practice, methods for input selection must be carefully handled, as adding inputs (even relevant ones) beyond a certain point can actually degrade the performance of a predictive model (Viaene et al, 2005).

**Base Model Creation**

One can employ any of the following methods to ensure a variety of base models creation:

1. Different training sets
2. Different deterministic algorithms
3. Different parameter setups
4. Randomize or non-deterministic algorithm

We opted for varying instance weights, since BRF-ANN model is considered a stable algorithm as it does not react adversely when its parameters are perturbed unlike algorithms like Decision trees and Regression. Beside, BRF-ANN is weight-sensitive, therefore varying its weight vector is enough and sufficient to obtain different base models.

Hence, Base models *m1, m2,… , mh* could be generated using the same training set, but different vectors of per-instance weights *ω*(*1*), *ω*(*2*), … , *ω*(*h*) with each vector ω(*i*) containing weight *ω*(*i*)**x** for each instance **x** ∈ T. However, instead of generating the weights randomly, we used a refined weight adjustment scheme of AdaBoost (i.e. Adaptive boosting) algorithm, particularly suited for a 2-class classification tasks. The R implementation of AdaBoost and its variants found in R-Project: <http://cran.r-project.org/web/packages/> was adopted in this research.

**Models aggregation and selection**

This is the combination of base models *m1,m*2*,* … *,mn* into a better model *m*∗ using the BRF-ANN scheme, that makes it possible to compute *m*∗(*x*) based on *m1*(*x*)*, m*2(*x*)*,* … *, mn*(*x*) for arbitrary *xi* ∈ *X*. The combined model, *m*∗ is the representative of its base models. Essentially, the prediction error is estimated as the criterion for model selection, i.e. the estimation of how well the trained model will perform on future unseen datasets, by selecting the model whose estimated prediction error is least. The estimation of the prediction error in the final selection of base models, this is where BRF-ANN comes in handy.

Typically, the model aggregation technique for classification tasks is *class label* voting, usually, the combined prediction is obtained by:

(*Eq*. 3.14)

where the ∏condition notation is used to denote an indicator function that yields 1 when the

*condition* is satisfied and 0 otherwise, i.e.,

Finally, the aggregation function is:

This is the models’ representative.

1. **CLASSIFIER PERFORMANCE AND ISSUES**

Since RBF-ANN is a weight sensitive algorithm we used a weighted-miscalculation error for data attributes classification. Suppose a weight *wx* is assigned to each *x* ∈ *S*, the weighted misclassification error of model *m* with respect to the class *c* (normal or illegal transaction),on dataset *S* can be obtained as:

*Eq.* 3.16

The algorithm accepts the supplied weight vector to determine the weighted miscalculation error.

**Cost Matrix**

Generally, misclassification costs can be specified as a |C| × |C| matrix *ρ* (a square matrix), where *ρ*[*d1*, *d2*] is the misclassification cost of predicting class *d1* for an instance of a true class d2. The matrix is usually assumed to contain 1s on the main diagonal (i.e., *ρ*[*d*, *d*] = 1 for all *d* ∈ C). Typically, positive integer numbers are used for the remaining entries, with 1 corresponding to the least expensive misclassification, See figure 3.5.

|  |  |
| --- | --- |
| **True Class** | |
| Legitimate (Normal) | Fraudulent (Abnormal) |
| **Predicted Class** | Legitimate (Normal) | 1  (TP) | 7  (FP) |
| Fraudulent (Abnormal) | 7  (FN) | 1  (TN) |

**Figure 3.5:** A 2-class misclassification cost matrix for fraud transactions detection.

(TP= True Positive; FP=False Positive; TN=True negative; and FN= True Negative)

Since we are more concerned with a generic fraud classification model, a better approach to incorporating misclassification costs that is not algorithm dependent, is based on *instance weighting*, as discussed earlier, in which case BRF-ANN will convert its weight-sensitivity to cost-sensitivity.

Generally the cost matrix, *ρ* that assigns misclassification costs is defined as:

In fact, this represents the mean misclassification cost of model *m* as regards concept *c*, since minimizing the weighted training set error is equivalent to minimizing the mean training set misclassification cost.

The BRF-ANN minimum cost selects the class with the lowest expected cost, then the expected cost is:

= = c = c(1 – πd) (Eq. 3.2)

1. **METHODOLOGY**

This is a data mining application described in terms of three-level application Architecture on top of a data repository. The overall framework designed to support this data mining-based fraud detection system is describe in this section.

The proposed architecture (figure 3.2), is such that the, data sources are spread in multiple locations (Location A, B, C, …, N) usually local and remote data sources, which are subsequently aggregated in a single location (a Data Warehouse) via transactions Monitors, using a client-server configuration through TCP/IP protocol (Fig. 3.3 step1). The data are then pre-processed. The selected data is, cleaned and transformed as necessary under the guidance and knowledge of a domain expert. This process could be time-consuming, but could be mitigated in part if the data is already in a structured database, or a data warehouse, (Fig. 3.3 step 2).

The substantive data mining software, query applications and fraud detection algorithm are housed on an application server while the database and its associated drivers are on an independent server.

The model was simulated with the AH Bank dataset (in other climes, synthetic data could be applied as most fraud transactions are classified) and the performance evaluation of the model was derived from *k*-fold cross-validation metrics, which would be compared with some other methods in order to show how prompt the model is in fraud detection.

In order to minimize the network topology, a certain number of training cases were applied, the noise level was noted and a minimum hidden nodes with weight decay observed for k-fold cross validation stopping strategy.



**Figure 3.2**: The flow of BRF-ANN Fraud detection framework.

## 6.2.1 The System Monitors

The monitoring system based on RBF neural network consists of a set of PCs for mining operational transactions. Seventy per cent (95%) of the transactions are to be constructed as training samples set for the neural network. The remaining 5 per-cent of the dataset to be used as test transactions., to test the predictive ability and the generalization of the derived model.

**6.2.2 Data Attributes Transformation**

Next, we perform attribute selection based on the area of interest, prior to model creation thereby making good models easier to find. Specifically, the dataset attributes are split into two: namely the *main* parameters, while the unselected parameters of a transaction are called *subordinate* parameters (see figure 3.3).

**6.2.3 Data Access, Analysis and inductive training**

The next is Figure 3.2, step 3, where a generic cluster technique is applied on the *main* parameters. The data mining engine accesses the data through a model developer to learn, develop and subsequently predict anomalous (or normal) transactions accordingly.

Expectedly, the *AH* Bank datasets preliminary results indicate that most of the transactions are normal, while very few are suspiciously fraudulent. We then merge the segregated subordinate attributes to the main attributes. These combined data were fed into the BRF-ANN network for training and the output is stored in the fraud knowledge repository or Detector.

The function of the BRF-ANN model developer is to dynamically generate and share new fraud detection models as shown in the Figure 3.2. In this framework, the first instance of a detected fraud may have its exemplary data processed by the model developer, which is subsequently used to detect new frauds and shares it with the detector(s).

**6.2.4 Transactions Query Portal**

This is where users can verify the status of doubtful transaction(s), set or confirm red flags. In situations where a user can not readily classify or determine the fraud status of a transaction, (fig. 3.2, step 7), the dataset is passed on to the trained BRF-ANN fraud-knowledge database to determine its suspicious level (fig. 3.2, step 9), the level of the transaction suspiciousness would inform the fraud investigator whether it is enough proof of fraud or to probe further.

1. **CONCLUSION**

The quest for quick and accurate fraud detection is extremely important in order to minimize losses, strengthen value system and integrity in any given system (Cheng et al, 2014). Fraud prevention and detection is of utmost importance in any system, hence it is incumbent on system managers to put in place mechanisms to check and report fraudulent practices. This research provides a proactive rather than reactive solution of fraud detection and have found relevance among business professionals, policy makers in both public and private sectors, thereby minimizing the time and cost of fraud detection by adopting computationally efficient model and framework rather than mere relying on the primitive notion of manual long investigations and/or windy auditors findings.

The proposed BRF-ANN model interacts with OLAP transactions rather than mere historical warehoused data so that fraud is detected and alerted more timely from when it occurred. This research provides a dynamic framework and special methods of intelligence gathering and data analysis that constantly monitor real-time transactions as well as off-line transactions to come up with fraud detection methods to deal with emerging and future fraud trends.

**REFERENCES**

Anderson, J.A. & Rosenfeld, E., 1998. Neurocomputing: Foundations of Research. MIT Press, Cambridge.

Apostolou B., Hassell J., Webber S., & Sumners G., (2001). The relative importance of management fraud risk factors. *Behavioral Research in Accounting 13*, 1–24.

Bauer, E., & Kohavi, R., (1999). An empirical comparison of voting classification algorithms: bagging, boosting, and variants. *Machine Learning, 36* (1–2), 105–139.

Bishop, C.M.(1995). Neural networks for Pattern Recognition. Oxford University Press, Oxford, U.K.

Bermúdez, L., Pérez, J.M., Ayuso, M., Gómez, E. & Vázquez, F.J.,(2008). A Bayesian dichotomous, model with asymmetric link for fraud in insurance. *Insurance: Mathematics and Economics, 42* (2), 779–786.

Bolton R. & Hand D. (2002). Statistical Fraud Detection: A Review (With Discussion). *Statistical Science*, *17*(3), 235–255.

Brause, R., Langsdorf , T., & Hepp, M. (1999). Neural data mining for credit card fraud detection. *11th IEEE International Conference proceeding*, 103 -106.

Caruana, R., & Niculescu-Mizil, A., (2006). An empirical comparison of supervised learning algorithms. In: Proceedings of the 23rd International Conference on Machine Learning, Pittsburgh, USA.

Cerullo, M.J. & Cerullo, V., (1999). Using neural networks to predict financial reporting fraud. Computer Fraud & Security, May/June (1999) 14–17.

Chang, W. & Chang, J. (2012), An effective early fraud detection method for online auctions. *Electronic Commerce Research and Applications 11*, 346–360.

Chen, R.C., Shu-Ting, L., & Shiue-Shiun, L., ( 2006). Detecting credit card fraud by using support vector machines and neural networks. *International Journal of Soft Computing 1* (1), 30–35.

Coussement, K., & Poel, D.V.d, (2008). Churn prediction in subscription services: an application of support vector machines while comparing two parameter- selection techniques*. Expert Systems with Applications 34* (1), 313–327.

Daskalaki, S., Kopanas, I., Goudara, M., & Avouris, N., (2003). Data mining for decision support on customer insolvency in telecommunications business. *European Journal of Operational Research 145*, 239–255.

Desai, V.S., Crook, J.N., & Overstreet, J., (1996). A comparison of neural networks and linear scoring models in the credit union environment. *European Journal of Operational Research 95,* 24–37.

Estevez, P., C. Held, & C. Perez (2006). Subscription fraud prevention in tele-communications using fuzzy rules and neural networks. *Expert Systems with Applications 31*, 337–344.

Ezawa, K. J., & Norton, S. W., (1996). Constructing bayesian networks to predict uncollectible telecommunications accounts. IEEE Expert, 11, 45–51.

Field, S. & Hobson, P., (1997). Techniques for telecommunications fraud management. In Proceedings of International Workshop on Applications of Neural Networks to Telecommunications, 3, 107–115.

Ghosh, **S.** & Reilly, D. L., (1994). Credit card fraud detection with a neural-network. *Proceedings of the 27th Annual Hawaii International Conference on System Science. 3,* 621 430

Green, B. & Choi, J. (1997). Assessing the Risk of Management Fraud through Neural *Network Technology. Auditing, 16*(1): 14–28.

Gupta, R. & Gill, N. S. (2013), Prevention and Detention of Financial Statement Fraud – An implementation of Data Mining Framework*. International Journal of Advanced Computer Science and Applications, 3(*8), 65-76.

Hackenbrack, K. ( 1993). The effect of experience with different sized clients on auditor evaluations of fraudulent financial reporting indicators. *Auditing: A Journal of Practice and Theory, 1*, 99–110.

Hardy, R.L., 1990, Theory and applications of the multiquadric-biharmonic method, 20 years of Discovery, 1968 1988, Computational mathematics Application, 1( 8/9), 163 208.

Hippert, H.S., Bunn, D.W., & Souza, R.C., (2005). Large neural networks for electricity load forecasting: Are they overfitted. *International Journal of Forecasting, 21*, 425–434.

Hilas, C.S., & Mastorocostas, P.A., (2008). An application of supervised and unsuper- vised learning approaches to telecommunications fraud detection. *Knowledge- Based Systems 21* (7), 721–726.

Hung, S.Y., Yen, D.C., & Wang, H.Y., ( 2006). Applying data mining to telecom churn management. *Expert Systems with Applications 31* (3), 515–524.

Kim, J., Ong, A., & Overill, R., 2003. Design of an Artificial Immune System as a Novel Anomaly Detector for Combating Financial Fraud in Retail Sector. Congress on Evolutionary Computation, Canberra, Australia.

Kirkos E., Spathis C., & Manolopoulos Y., (2007). Data mining techniques for the detection of fraudulent financial statements. *Expert Systems with Applications, 32* (4), 995–1003.

Krambia-Kapardis, M., Christodoulou, C., Agathocleous, M. (2010). Neural net-works: the panacea in fraud detection*.* *Managerial Auditing Journal, 7*, 59–78.

Lacher, R. C., Coats, P. K., Shanker, S. C., & Fant, L. F. (1995). A neural network for classifying the financial health of a firm. *European Journal of Operational Research, 85*(1), 53–65.

Lee, K. C., Han, I., & Kwon, Y. (1996). Hybrid neural network models for bankrupt-cy predictions. *Decision Support Systems, 18*(1), 63–72.

Loebbecke J.K., Einning M.M..& Willingham J.J., (1989).Auditors’ experience with material irregularities: frequency, nature and detectability. *Auditing: A Journal of Practice & Theor,;1*,1–28.

Majid, A., Gul, F.A., & Tsui, J. ( 2001). An analysis of Hong Kong auditors’ perceptions of the importance of selected red flag factors in risk assessment. *Journal of Business Ethics, 3,* 263–74.

Maranzato, R., Pereira, A., Naubert, M., & Lago, A. P., (2010). Fraud detection in reputation systems in e-markets using logistic regression and stepwise optimization. In ACM SIGAPP Applied Computing Review.

Mock, T.J. & Turner, J.L. (2005). Auditor identification of fraud risk factors and their impact on audit programs. *International Journal of Auditing, 9*, 59–77.

Mobley, B. A., Schechter, E., Moore, W. E., McKee, P. A., & Eichner, J. E. (2000). Predictions of coronary artery stenosis by artificial neural network. Artificial Intelligence in Medicine, 18(3), 187–203.

Moyes, G. L.(2007). The differences in perceived level of fraud-detecting effectiveness of SAS No. 99 red flags between external and internal auditors. Journal of Business & Economics Research, 6, 9–25.

Ngai, E.W.T., Hu, Y., Wong Y.H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems, 50*, 559-569.

Nigrini, M. (2011). [Forensic Analytics: Methods and Techniques for Forensic Accounting Investigations](http://www.wiley.com/WileyCDA/WileyTitle/productCd-0470890460.html). Hoboken, NJ: John Wiley & Sons Inc. [ISBN](http://en.wikipedia.org/wiki/International_Standard_Book_Number) [978-0-470-89046-2](http://en.wikipedia.org/wiki/Special:BookSources/978-0-470-89046-2).

Orr, M. J., (1999). Introduction to Radial Basis function Networks, Recent advances in radial basis function networks, Edinburgh EH8 9LW, Scotland.

Palshikar, G. K.,(2002). The Hidden Truth – Frauds and Their Control: A Critical Application for Business Intelligence, Intelligent Enterprise, (5), 9, 28 May 2002, 46–51.

Phua, C., Lee, V., Smith, K., & Gayler, R. (2005). A comprehensive survey of data mining-based fraud detection research. *Artificial Intelligence Review*, 1–14.

Piramuthu, S. (1999). Financial credit-risk evaluation with neural and neurofuzzy systems. European Journal of Operational Research, 112(2), 310–321.

Salchenberger, L. M., Venta, E. R., & Venta, L. A. (1997). Using neural networks to aid the diagnosis of breast implant rupture. Computers and Operations Research, 24(5), 435–444.

Sharda, R., & Wilson, R. (1996). Neural network experiments in business failures prediction: A review of predictive performance issues. International Journal of Computational Intelligence and Organizations, 1(2), 107–117.

Sherly, K.K., Nedunchezhian, R., (2010). Boat adaptive credit card fraud detection system. IEEE (2010).

Smith, M., Omar, N.H., Idris, S., Baharuddin, I. (2005). Auditors’ perception of fraud risk indicators, Malaysian evidence. *Managerial Auditing Journal, 1,* 73–85.

SPSS (2000), Using Data Mining to Detect Fraud, White Paper – Technical Report.

Steinwart, I., & Christmann, A., (2008). Support VectorMachines. Springer, New York, USA.

Stern, H.S., (1996). Neural networks in applied statistics. *Technometrics 38* (3), 205–216.

Viaene, S., Derrig, R.A., & Dedene, G. (2004). A case study of applying boosting naive Bayes to claim fraud diagnosis*. IEEE Transactions on Knowledge and Data Engineering, 16(*5), 612–620.