Circulation in Computer Science
Vol.3, No.1, pp: (1-9), January 2018
https://doi.org/10.22632/ccs-2017-252-70

Discriminating Input Variables for Fraud Detection using Radial Basis Function Network

I. O. Alabi
Department of Information & Media Technology
Federal University of Technology
Minna, Nigeria

R. G. Jimoh
Department of Computer Science
University of Ilorin,
Ilorin, Nigeria

ABSTRACT
Fraud is an adaptive crime; special methods of data gathering and analysis are required to combat fraud issues as criminals often quest for dubious techniques to evade detection. Radial basis function (RBF) network, was used to build base models that identifies and detect the risk of fraud in transactions. At first, it is imperative to isolate the basic factors that are predictive of fraud occurrences so as to determine the Information gain of each attribute. The input variables' importance was ascertained to indicate how some of the input variables were distinguished as strong indicators or weak indicators of fraud. Hence, the relevant attributes were selected prior to examining the model's performance. This study has found relevance among corporate business professionals and government agencies, to minimizing the time and cost of fraud detection. The researcher recommended that fraud mining processes be regularly updated at fixed time intervals to checkmate criminals.

Keywords
Artificial neural network, attributes discrimination, detecting fraud transactions, fraud detection, radial basis function, and data mining.

1. INTRODUCTION
Fraud is a criminal act of depriving others of their valuables. Financial scam is pervasive and it is adversely affecting economies worldwide of which many people have been deprived of substantial amount of valuables. Fraud is a global scourge that harms corporate reputations, costs millions and ruins lives [21]. New technologies have provided further ways in which criminals may commit fraud [3]. In its 2015-2016 global fraud report, [23] observed that the number of businesses suffering a financial loss as a result of fraud is on the increase, specifically from 64% in the previous survey period to 69% during the year under review. [23] report posits that globalization of businesses have also contributed to increases in fraud risk, for instance, in situations where many cross-border businesses have thousands of companies in their supply chain, risks become more difficult to identify and keep under control. Some key fraud prone elements are shown in figure 1.

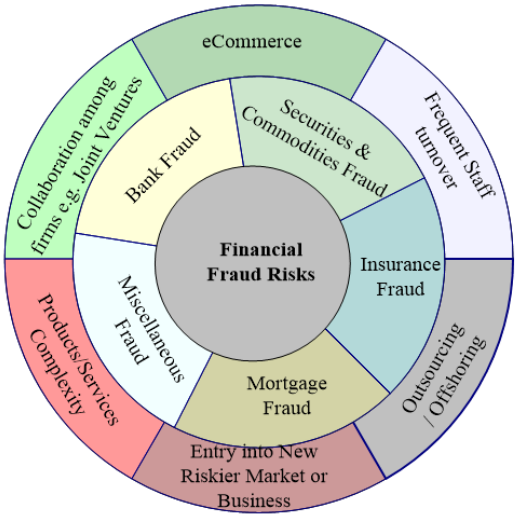

Figure 1: Financial Fraud Risks and Factors

Figure 1 depicts classical financial fraud types across different sectors of global economies such as bank, insurance mortgage, securities and commodities, etc., as well as some of the factors or elements of the fraud causes ranging from eCommerce activities, frequent staff turnover, collaboration among businesses, outsourcing or offshoring of employees, entrance into a new riskier market or new businesses outright. These are some elements, among others on which fraud threats could emanate.

Noteworthy are firms that were dissuaded from operating in some regions of the globe as a result of higher fraud risk exposure, especially in South America and Africa (See Table 1). Aside, other factors such as internal threats (employees at different cadre), cyber threats, opportunities, incentives and intent rationalization to defraud have all contributed to higher fraud risk exposure.

Table 1: Exposure to Frauds by Regions

Region	Fraud Rate (%)
North America	25
South America	10
Africa & Middle-East	12
Europe	29
Asia Pacific	**24**

Source: [23]

1.1 Fraud Detection and Prevention

Fraud detection and prevention are concurrent processes in combating fraud malaise; while fraud detection is the spotting 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. In some applications, fraud detection and prevention processes are in tandem [3], whereby fraud detection comes in once fraud prevention has failed thereby making fraud detection a continuous process. Nonetheless, fraud is an adaptive crime, so it needs special methods of intelligence gathering and data analysis in order to detect and prevent frauds [29].

Propensity towards fraud can never be eliminated; the onus is on management to create the most effective system possible to prevent it [35]. A corporate fraud policy formally sets out what an employee is expected to do when he or she spots suspicious transactions. Fraud detection procedure, employees' training and awareness must be in place.

Financial fraud detection (FFD) is vital to prevent fraud by distinguishing fraudulent financial data from authentic data, thereby alerting fraudulent behaviour or activities thus enabling decision makers to develop appropriate strategies to decrease the impact of fraud [28].

1.2 Fraud Data Mining

Data mining methods are evolving with supervised or unsupervised training data (of past cases) to build models to identify and detect risks of fraud [16]. Hence, researchers and business professionals have acknowledged data mining techniques for playing a key role in fraud detection due to its capability to extract knowledge from huge data heaps, and have been assisting auditors and crime investigators to track fraudulent practices [40], yet only 3% of firms surveyed by [21] used data analytics to detect fraud, indicating that technology is not well utilized. As stated earlier, notwithstanding greater and more innovative efforts organizations are making to combat fraud, it persists as a serious threat to businesses while its adverse impacts cannot be underestimated [19].

Aside due diligence, staff compliance training and other internal controls which could be useful, and which must be adopted with proactive data analytics to identify anomalous transactions or behavior. Researchers have proposed a number of fraud detection techniques, such as logistic regression, linear or quadratic discriminant analysis, and neural networks using various data mining as well as other computational techniques to model and predict fraudulent practices.

It is remarkable that 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 trends [4].

1.3 Input Features Selection

In this study, the researchers sought to discriminate some fraud attributes that are more significant in fraud indication using a radial basis function (RBF) network. Radial basis functions are classical functions that can be employed to approximate models (linear or non-linear) of a neural (single or multi-layer) network, especially useful to approximate multivariate functions with its remarkable convergence properties [5]. RBF has been successfully applied in many areas, such as fire detection, to measure several parameters (the flame color, spectrum, intensity, direction, etc.) to model a fire detector device. In robotics, RBF was used to interpolate the data that come from the raster of a screen of a robot's eye [11]; and [22].

Accordingly, RBF was used to interpolate the data that come from financial transactions to predict fraud occurrence or suspicious risk. RBF's resilience and convergence power over other interpolants such as regression and partial differential equations makes it outstanding [5]. First, we want to demonstrate the fact that some of the observed input attributes are significant to the RBF's network response, while others are not.

2. RELATED WORKS

Researchers have adopted many techniques for solving fraud issues, for instance, Statistical methods (parametric and non-parametric, Linear discriminant analysis (LDA), Linear Regression (LR)), machine learning and supervised neural networks such as fuzzy neural nets, and combinations of neural nets and rules, have been extensively explored and used for detecting fraud as ([15] and [12])'s works revealed.

[20] proposed a combination of data mining and natural computing techniques. Quite often, hybrid models that combines multiple inductive models for the same domain are used in order to obtain better prediction quality, thereby reinforcing strengths and compensating weaknesses [9].

[26] and [38] used the logistic regression method as a tool to discriminate fraudulent actions from legitimate actions for insurance companies and e-commerce. Though simple to interpret, the result of classification is not categorical (YES/NO), instead it was an estimated probability of each observation belonging to a given class [13].

[7] used decision trees (C4.5) and the instance-based learning algorithm to detect fraudsters. This approach is unsuitable where multiple attributes are being considered. Similarly, [34] compared logistic regression, neural networks and regression trees. They observed that neural networks and logistic regression approaches outperform decision tree in solving the fraud problem.

[37] noted that Machine learning and artificial intelligence solutions are increasingly explored for fraud detection and prediction, especially in insurance milieu. [20] also asserted that most financial fraud solutions are premised on sets of predefined rules and thresholds, perhaps based on statistical means and standard deviations, though these are hardly enough to trap recent sophisticated fraud means.

Recently, neural network has been widely used due to its ability to model complex and non-linear models, however not having any strict limitations and rigorous assumption for the type of input data ([36]; [1]). Its downsides include long learning time, over-fitting error, and black box characteristics ([2]; [17]).

Another concern is whether a fraud detection model is accurate enough to provide correct classification of a case into fraudulent or legitimate, since fraud detection tools with largest predictive capability are always required in practice.

3. KEY ATTRIBUTES DISCRIMINATION

If one could isolate the factors that indicate a fraud risk or a high probability of fraudulent practices, then develop rules (or controls) and use them to flag only those claims or requests susceptible to be fraudulent of which data mining techniques are replete with various techniques of such. By this, fraud investigators can identify the symptoms of fraud before large losses occur. Continual routines that monitor key symptoms and track risk trends can also be a major deterrent, thereby preventing or identifying fraud almost as soon as it occurs.

Usually, financial transactions are characterized by a set of features, say *m* (often numerous in dimensionality), some of which might not have correlation with fraud detection e.g. telephone number, while some of these transactions features are highly correlated to fraud detection e.g. credit history as would be discussed later. It is these significant, highly correlated features that were selected to construct the desired model.

[37] was emphasizing the importance of input relevance, arguing that it was not uncommon for domain experts to ask which inputs are relatively more important or contribute most to fraud detection. As such, methods for input selection are not only capable of improving the human understanding of the problem, but also allow for more efficient and lower-cost solutions. They concluded that adding inputs (even relevant ones) beyond a certain point can actually lead to a reduction in the performance of a predictive model.

As stated earlier, researchers have explored many models for fraud detection; the widely used is the Neural networks — 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 [2]. Neural networks have shown to be very promising alternatives for modeling complex nonlinear relationships ([10]; [24]; [25]; [27]; [31]; [32]; [33]).

As much as the modeling flexibility of neural networks is very attractive for modeling complex and non-linear models ([36]; [1]), yet some practical issues persist when implementing neural networks, such as the impact of the initial weight choice, setting the weight decay, and adjusting the training data noise. Other defects include long learning time, over-fitting error and black box characteristics (i.e. lack of explanatory power) [2]; [17]).

4. METHODOLOGY

An evolutionary algorithm as BRF network can be used to improve the deficiencies stated above. Radial basis function network is an approximation of a true model, which represents better the issue at hand. The score functions were slightly relaxed so that the model's parameters and predictions do not vary drastically.

4.1 The RBF Network Learning

RBF approximation and interpolation function of large, say *n*, of radial basis functions, each with different centres x_i and weights w_i. The weights can be approximated with linear least squares using linear algebra, which makes analysis easier and computations faster [30], See Figure 2.

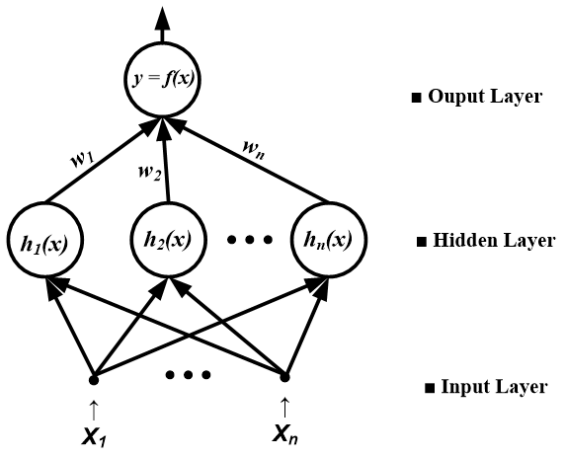

Figure 2: A basic RBF neural network architecture.

Typically figure 2 indicates how a conventional RBF network feeds-forward the input vector **x** to n basis functions whose outputs are linearly combined with weights into the network output.

The network is often used for its global approximation properties and freedom of local minimums. Then, we cluster the data and normalize the data sampling bias with k-mean clustering algorithm in combination of the least square method. Weight adjustment in network training is done by successive layer weight optimization.

Simply, we will adhere to a single hidden layer network as the one shown in Fig. 2. The network function,

$$\boldsymbol{y} = f(x) = \boldsymbol{W}(\boldsymbol{x}) = \sum_{i=1}^{h} \boldsymbol{\omega_i \varphi_i}(\boldsymbol{x}) \qquad (1)$$

Usually, RBF neural network has *n*-inputs, *h* hidden nodes and *m* outputs, otherwise called ***n-h-m* neural network** such that each of the vector **x** feeds forward to *m*-basis functions whose output are linearly combined with weights, ω_i into the network output *f(x)*, a weighted sum of hidden units (see equation 1). The equation (1) depicts the input vector, $\mathbf{x} = (x_1, x_2,..,x_n)^T \in R^n$, the weight matrix, $\boldsymbol{\omega} \in R^{n \times m}$, and the network output, where, $\varphi_i(\mathbf{x})$ is the activation function of hidden node *i*, ω_i is the weight of node *i* and *h* is the number of hidden nodes.

4.2 Model Creation

The supervised learning procedure adopted in this work to construct a BRF-ANN network whereby the network is trained with feature inputs $x_i = (x_{i1},..,x_{ip})$ and the corresponding outputs $y_i \in \{0,1\}$. The sole objective of the training algorithm was to ensure that a set of input features would yield the anticipated set of outputs using the BRF network framework, such that the developed final model could subsequently classify previously unseen data features into their respective true classes.

The RBF network could assume a variety of activation functions for hidden nodes of which Gaussian function is used in this study, i.e.,

$$\boldsymbol{\varphi_i}(\mathbf{x}) = \boldsymbol{e}^{\frac{\|\mathbf{x}\,\mathbf{c_i}\|}{\delta_i^2}}, \quad (for\ i\ 1, 2, \ldots, h) \qquad (2)$$

Where, $\mathbf{c}_i = (x_1, x_2,..,x_n)^T$ is the centre of the hidden node *i*; $\boldsymbol{\delta}_i$ is the constant extension of hidden node *i*. The obvious advantages of the Gaussian activation function are its flexibility and its ability to fit different weight values.

4.3 Model Selection

The *k*-mean algorithm, an indirect clustering approach based on inter-sample similarity measurement used to select *k* samples out of *n* samples as the initial cluster centre and assigns the other objects to clusters represented by cluster centres that are most similar to them according to the distances to the initial cluster centres. The inter-cluster distance, d_i is the distance from cluster centre *i* to other cluster centres, i.e. $d_i = \min_i \| c_j - c_i(k) \|$ in which *k* is the overlap coefficient. Then it calculates the cluster centres of new clusters, one after the other, until the metric function, usually, *mean square error* (*mse*) begins to converge.

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

$$\boldsymbol{D_i}(\mathbf{x}) = \boldsymbol{\sigma_i} \| \boldsymbol{x} \ \ \mathbf{c_i} \|^2 \quad , \boldsymbol{i} = \mathbf{1}, \mathbf{2}, \dots, \boldsymbol{h} \tag{3}$$

If this equation converges, i.e., the first $D_i(\mathbf{x}) = \min D_i(\mathbf{x})$, the iteration ends. If it does not converge, the distance between the samples and the cluster centres has to be re-calculated.

As stated earlier, the input attributes feature classification, the hidden layers of the network are actively interacted with one another such that the output of the hidden layer *j*-1 is the input of the hidden layer *j*. and that the radial basis function output is restricted to the interval (0,1) by the function:

$$Y_j(X) = \frac{1}{1+e^{-\varphi_j(X)}}, \ \ for\ j=1,\dots,k \tag{4}$$

Hence, the *j*th layer is the predicted response class, i.e.

$$\hat{y} = \boldsymbol{W}(\boldsymbol{X}) = \begin{cases} 1, & if\ transaction\ is\ fraudulent \\ 0, & if\ transaction\ is\ normal \end{cases} \tag{5}$$

This study used multi-class encoding to achieve this (see Table3). These transformations are required to adjust to the needs of our classification algorithm.

Hence, the BRF network is constructed once a weight is obtained by training the network with feature inputs $x_i = (x_{i1},..,x_{ip})$ and the corresponding outputs $y_i \in \{0,1\}$ yield. The sole objective of the training algorithm was to ensure that a set of input features would yield the anticipated set of outputs using the BRF network framework, such that the developed final model could subsequently classify previously unseen data features into their respective true classes.

4.4 Simulation

The model is simulated with the German Bank credit data (since most fraud transactions are classified, and are not available in public domain). 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.

In order to determine suitably minimal attributes worthy of selection that are sufficient to detect suspicious transactions that could yield utmost performance of the model. See Table 2.

4.5 Experimental Dataset Description

Due to dearth of fraud data, a German bank credit data, a true representative of the problem, available online was used to construct a classification model. The dataset used consists of real German Bank Credit data from the UCI Repository of Machine Learning Databases [18]. However, this dataset is of interest because it consists of a good mix of continuous and nominal attributes, a few missing values and no special knowledge is required to understand it. All attributes and values of real identities have been concealed due to the confidentiality of the data.

This dataset variable description is as shown in Table 2. The dataset has been edited to include several indicator variables to make it suitable for the algorithm under consideration in order to conform with the response categorical variables as defined in equation (5).

The dataset consists of 1000 records of bank customers, with 30% fraud cases consisting of 20 features attributes (factors), having a good mix of continuous and nominal attributes were taken. These attributes consist of several indicator variables that are suitable for the model development. The sample data was split into two using the algorithm adopted by [39], thus n_T (training) and n_Q (test) in the ratio 9:1 respectively. Finally, the response variable is a binary valued variable, coded as 0 (for *normal transaction*) and 1(for *abnormal or fraudulent transaction*). The data attributes used are:

Table 2: Table of Variable Description

S. No	Variables	Description
1.	A1	Status of current Account
2.	A2	Duration
3.	A3	Credit history
4.	A4	Credit purpose
5.	A5	Credit amount
6.	A6	Status of Savings Account
7.	A7	Present Employment Status
8.	A8	Installment rate in percentage of disposable incomes
9.	A9	Gender and Marital
10.	A10	Guarantor(s)
11.	A11	Present Address
12.	A12	Property Ownership
13.	A13	Age (in years)
14.	A14	Other Installment Plans
15.	A15	Housing
16.	A16	Number of existing credits in this bank
17.	A17	Employment Type
18.	A18	Number of Dependents
19.	A19	Telephone
20.	A20	Foreign Worker

4.6 Itemsets Runs and Classification

To test the predictive ability and the generalization of the derived model, we shall apply the remaining 10 per-cent of the test transactions.

The prediction performance of each base model at different hidden layers can be evaluated with the average *misclassification error rate* (MER) or the prediction error rate thus:

The prediction performance of each classification model at different hidden layers, H was performed using the average *misclassification error rate* (MER), $\hat{\vartheta}_H$ given as:

$$\hat{\vartheta}_H = \frac{1}{S \times n_T} \sum_{r=1}^{s} \sum_{i=1}^{n_T} I(y_T \neq \hat{y}_T) \qquad (6)$$

where, s is the cross-validation runs
$I(\bullet) \in \{0,1\}$ is an indicator function
H is the number of hidden nodes

Averaged over the number of cross-validation runs s, where $I(\bullet)$ is an indicator function whose value is 1 if the predicted class label $\hat{y}_{iT}$ of the ith sample at the rth cross-validation run does not equal the true class label y_{iT} of the sample transactions and 0 if otherwise. Hence, $\hat{\vartheta}_H$ is the prediction error rate of the model with H number of hidden nodes.

Misclassification costs are elusive in practice, difficult to quantify and are often left at readers' discretion, however, regardless of the assigned misclassification cost, the classification model with $\hat{\vartheta}_n = \min(\hat{\vartheta}_1, \hat{\vartheta}_2, \ldots, \hat{\vartheta}_j)$ is chosen as the best model that fits the dataset and the number of hidden nodes that yielded this best model $n \in H$ is the optimal hidden nodes number for this classification model.

Specifically, the *German* Bank datasets were fed into the BRF-ANN network for training and the output is stored in the fraud knowledge repository or Detector. A model developer is to dynamically generate and share new fraud detection models. 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).

Since fraud matters are typically binary classification issues, the hidden layers of the network actively interacted with one another such that the output of the hidden layer j-1 is the input of the hidden layer j (See Figure 3) and that the radial basis function output is restricted to the interval (0,1) by the function:

$$Y_j(X) = \frac{1}{1+e^{-\varphi_j(X)}}, \quad for\ j=1,\ldots,k \qquad (7)$$

Hence, the jth layer is the predicted response class, i.e

$$\hat{y} = \boldsymbol{W}(\boldsymbol{X}) = \begin{cases} 1, & if\ transaction\ is\ fraudulent \\ 0, & if\ transaction\ is\ normal \end{cases} \qquad (8)$$

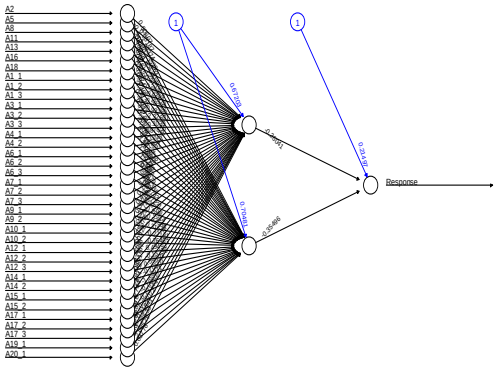

Figure. 3: A 1-layer Radial basis function network

Of importance is the attribute transformation, a pre-processing task of sort for inductive learning. Attribute transformation, though functionally dependent on the original data for easy analysis, in this study, we used multi-class encoding to achieve this (see Table 3). These transformations are required to adjust to the needs of our classification algorithm.

5. RESULTS

After experimenting with the equation (1) on the data set, it is necessary to measure the contribution of each independent (input) variable and the best performing variables on the response variable. The Garson Algorithm [14], (relative important of these independents variables) is a better technique to adopt. The 2-layer BRF network was obtained, i.e., a typical 20 : 1 : 1 network layer (See Figure 3).

The instance weights were varied, since BRF-ANN model is considered a stable algorithm as it does not react adversely when its parameters are perturbed (changes due to minor data variation), unlike some algorithms such as Decision trees and Regression, giving that BRF network is weight-sensitive, therefore varying its weight vector is enough and sufficient to obtain different base models.

5.1 Results Discussion

Table 3 shows the input variables performance chart, i.e., the input variables that positively and negatively affects the network's response. For instance, Credit history (Variable 3), Employment type (Variable 17), and provision of loan Guarantor are strong indicators of positive response, that is, if any of these variables are not provided in any instance, the network issues a red flag (an indicator for fraud transaction).

On the other hand, Account status (variable A1), employment status (variable A7) and instalment rate (variable A8) were not strong enough to pre-empt fraud, see Table 3.

Figure 4 is a graphical view of the input variables importance rank, while Table 4 shows the key variables importance values. The result indicated that variables A10_1 (none provision of Guarantors); variable A3_2 (credit payback duly); variable A3_1 (Credit history of whether or not credit had been taken in the past); variable A17_3 (whether or not a transaction Job status is employed or not); and variable A_20 (the resident status of a transaction, i.e. foreign or local) are all exerting positively on the network output, i.e. are all contributing significantly to the occurrence of fraud.

On the other hand, the variables A6_2 (Savings account holders having balances in the range 100 to 100 Dutch Mark); variable A1_3 (the account type: checking or savings, etc.); variable A1_2 (the balance status: zero or higher than 200 Dutch Mark); variable A10_2 (having co=applicant as a Guarantor); variable A7_3 (being employed within less than 3 years or higher than 7 years); variable 6_3 (Not having a saving account) are all having negative effect on the network variable, i.e. they are not strong indicators of fraud.

Table 3: Variables Importance using Garson Algorithm [14]

S/N	Independent Variable		Relative importance	Remarks on the Transaction status
1	Duration in month (A2)		0.014	Very Weak positive relationship
2	Amount of Credit (A5)		0.0000	No relationship
3	Installment Rate (%) (A8)		0.0917	Very Weak positive relationship
4	Present Residence in years (A11)		0.0294	Very weak positive relationship
5	Age in Years (A13)		-0.0003	Very weak Negative relationship
6	Number of Existing Credit (A16)		-0.0488	Very weak Negative relationship
7	# People liable for maintenance (A18)		0.0894	Very Weak positive relationship
8	Status of existing Account	$< 0DM$ (Ref., A1_1)	-0.1900	Weak Negative relationship
		$0 \leq \cdots \geq 200DM$ (A1_2)	**-0.6001***	**Intermediate Negative relationship**
		No checking Acc. (A1_3)	**-0.7571***	**Strong Negative relationship**
9	Credit History	**No credit taken (A3_1)**	**0.3966***	**Substantial positive relationship**
		All credit payback duly (A3_2)	**0.8276***	**Strong Positive relationship**
		Delay in paying back (Ref.,A3_3)	**0.3321**	**Substantial positive relationship**
10	Purse of the credit	Tangible asset (A4_1)	0.0330	Very Weak positive relationship
		Non-tangible (Ref., A4_2)	-0.0349	Very weak Negative relationship
11	Saving Acc./Bonds	$< 100DM$ (Ref., A6_1)	-0.2314	Very weak Negative relationship
		$100 \leq \cdots \geq 1000DM$ (A6_2)	**-1.000**	**Perfect Negative relationship**
		No saving Acc. (A6_3)	**-0.4297**	**Substantial Negative relationship**
12	Present Employment	Unemployed (Ref., A7_1)	-0.1354	Weak Negative relationship
		$1 \leq \cdots < 4\ years$ (A7_2)	-0.048	Very Weak Negative relationship
		$4 \leq \cdots \geq 7\ years$ (A7_3)	-0.2446	Weak Negative relationship
13	Personal Status and sex	Sex (M=1, F=0) (A9_1)	0.0767	Very weak positive relationship
		Marital status (M/D/S/W=1, S=0) (A9_2)	**0.3024**	**Substantial positive relationship**
14	Guarantors	**None (Ref., A10_1)**	**0.3842**	**Substantial positive relationship**
		Co-applicant/guarantors (A10_2)	**-0.8577**	**Strong Negative relationship**
15	Property	**Real Estate (A12_1)**	**-0.3387**	**Substantial Negative relationship**
		Car & others (A12_2)	**-0.2895**	**Substantial Negative relationship**
		No property (Ref., A12_3)	**-0.2708**	**Substantial Negative relationship**
16	Installment Plans	Bank/Store (A14_1)	0.1083	Weak Positive relationship
		None (Ref., A14_2)	**0.2322**	**Substantial Positive relationship**
17	Housing	Rent (Ref., A15_1)	-0.1995	Weak Negative relationship
		Own/free (A15_2)	**-0.3006**	**Substantial negative relationship**
18	Job	Unskilled (Ref., A17_1)	-0.1162	Weak Negative relationship
		Skilled (A17_2)	-0.0627	Very Weak Negative relationship
		Self-employed (A17_3)	-0.0509	Very Weak Negative relationship
19	Telephone	(None=0, Yes=1, A19_1)	-0.1607	Weak Negative relationship
20	Foreign Worker	**(Yes=1, No=0, A20_1)**	**0.4900***	**Substantial Positive relationship**

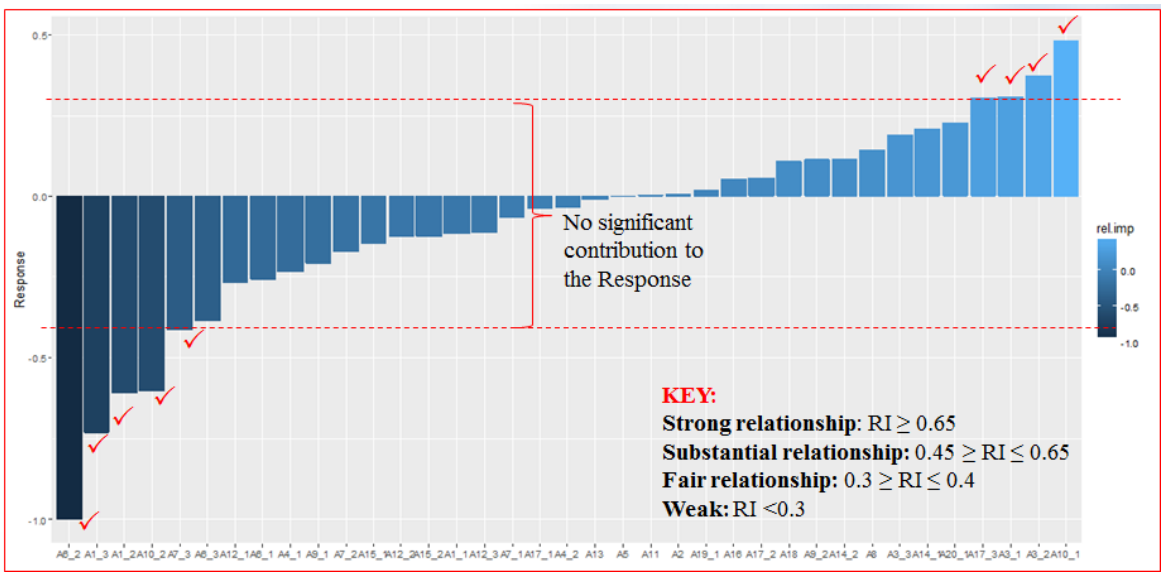

Figure 4: Overall attributes performance.

Table 4: Table of key variables based on variable importance values.

S. No.	Variables	Description	Variable importance value
1.	A10_1	Guarantors: None	0.3842
2.	A3_2	Credit History: All credit payback duly	0.8276
3.	A3_1	Credit History: No credit taken	0.3966
4.	A17_3	Job: Self-employed	-0.0509
5.	A20_1	Foreign Worker	0.4900
6.	A6_2	Saving Account/Bonds (btw 100 and 1000DM)	-1.0000
7.	A1_3	No checking Account	-0.7571
8.	A1_2	Status of existing Account : $0 \leq \cdots \geq 200DM$	-0.6001
9.	A10_2	Guarantors : Co-applicant/guarantors	-0.8577
10.	A7_3	Present Employment: $4 \leq \cdots \geq 7\ years$	-0.2446
11.	A6_3	No saving Account	-0.4297

6. CONCLUSION

As stated in section 1, financial frauds are abnormal activities hence are generic with similar characteristics but distinct parameters. Notwithstanding the fact that a German bank credit dataset was used for this experiment, it is believed that the model would exhibit similar behaviour with other localized datasets.

The future work of this study should be able to make some comparisons of the level of frauds in other sub-sectors of the financial industry, such as Microfinance banks, Commercial banks and some specialized banks like Export/Import banks, Industrial and Agro-allied development banks where fraud propensity is assumed low.

Meanwhile, this study, adhered to some data mining processes to harness data, pre-process it, then trained a BRF network model using a dataset from an online German bank credit data. Base models were created, with their *R* implementations, and in turn used in aggregation to build the required radial basis network model.

The input variables' importance was ascertained and the data summary shown. Also, it was shown how some of the input variables were distinguished as strong indicators or weak indicators of fraud. The dataset was split into 90:10 per-cent Training and Test data ratio the derived model's prediction shall be evaluated for prediction accuracy or misclassification error in the subsequent reports. The researcher is of the opinion that some more variables when explored further, could still be identified to positively/negatively affect the response of the network.

Quite often, fraud detection techniques or measures are not enumerated in great detail in the public sphere, as this gives perverse criminals the information that they require to evade detection, also since fraud perpetrators adapt their methods on an ongoing basis; their persistence and stealth is especially evident in the creative ways digital networks are constantly being attacked. Hence, models can be updated at fixed time intervals to checkmate criminals, set up protocols or filters for urgent or confidential transactions or to block fraudulent transactions outright.

7. REFERENCES

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

[2] Bishop, C.M., (1995). Neural Networks for Pattern Recognition. Oxford University Press, Oxford, UK.

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

[4] Brause, R., Langsdorf T., & Hepp M., (1999), Neural data mining for credit card fraud detection, 11th IEEE International Conference proceeding, pp103 -106.

[5] Buhmann, M. D., (2004). Radial basis functions: Theory and Implementations. Cambridge university press. ISBN 0-521-63338-9.

[6] 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.

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

[8] Cheng C., Wang, D. & Shi, W., (2014). *The Second Monitor Centre of China Earthquake Administration, Xi'an, Shaanxi, China.* Engineering Technology and Applications. Taylor & Francis Group, London, ISBN 987-1-138-02705-3.

[9] Cichosz, P., (2015). Data mining algorithms: Explained using R. John Wiley & sons Ltd., ISBN 978-1-118-33258-0

[10] 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.

[11] Eckhorn, R. (1999) 'Neural mechanisms of scene segmentation: recordings from the visual cortex suggest basic circuits for linking field models', IEEE Trans. Neural Net. 10, 1–16.

[12] Estevez, P., C. Held, & C. Perez (2006). Subscription fraud prevention in telecommunications using fuzzy rules and neural networks. Expert Systems with Applications 31, 337–344.

[13] Farvaresh, H. & Sepehri, M. M., (2011). A data mining framework for detecting subscription fraud in telecommunication. Engineering Applications of Artificial Intelligence 24, 182–194.

[14] Garson, G.D. 1991. Interpreting neural network connection weights. Artificial Intelligence Expert. 6(4), 46-51.

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

[16] 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.

[17] 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.

[18] Hoffman, H (2000). UCI Machine Learning Repository [http://archive. .ics.uci.edu/ml]. Institut f"ur Statistik und "Okonometrie Universit"at Hamburg.

[19] Interpol (2016). Suspicious activity reports. Retrieved from: (https://www.interpol.int/Crime-areas/Financial-crime/Social-engineering-fraud/Types-of-social-engineering-fraud) on January 13, 2017.

[20] Khac, N. A. L., & Kechadi, M., (2010). Application of data mining for antimony laundering detection: a case study, in: Data Mining Workshops (ICDMW), 2010 IEEE International Conference on, 2010, pp. 577–584.

[21] KPMG (2016). Global profiles of the fraudster: technology enables and weak controls fuel the fraud. Retrieved from: https://home.kpmg.com/xx/en/insights/2016/05/global-profiles-of-the-fraudster.html on January 13, 2017.

[22] Kremper, A., Schanze, T. & Eckhorn, R., (2002). 'Classification of cortical signals with a generalized correlation classifier based on radial basis functions', J. Neurosci. Meth. 116, 179–187.

[23] Kroll (2016). Global Fraud Report 2015-16: vulnerabilities on the rise. Retrieved from: http://www.kroll.com on January 13, 2017.

[24] 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.

[25] Lee, K. C., Han, I., & Kwon, Y., (1996). Hybrid neural network models for bankruptcy predictions. Decision Support Systems, 18(1), 63–72.

[26] 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.

[27] 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.

[28] Ngai, E.W.T., Hu, Y., Wong Y.H., Yijun Chen, & 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.

[29] Nigrini, M., (2011). Forensic Analytics: Methods and Techniques for Forensic Accounting Investigations. Hoboken, NJ: John Wiley & Sons Inc. ISBN 978-0-470-89046-2.

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

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

[32] 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.

[33] 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.

[34] Shen, A., Tong, R., & Deng, Y., (2007). Application of classification models on credit card fraud detection. In Proceedings of the 10th International Conference on Service Systems and Service Management, 1–4.

[35] Silverstone, H. & Davia, H. R., (2005). Fraud 101: Techniques and Strategies for detection, 2ed. John Wiley & Sons, Inc., Hoboken, New Jersey.

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

[37] Viaene, S., Dedene, G. Dedene, & Derring R. A., (2005). Auto claim Fraud detection using Bayesian learning neural networks. Expert systems with applications 29 (2005) 653-666.

[38] Wilson, J. H., (2009). An analytical approach to detecting insurance fraud using logistic regression. Journal of Finance and Accountancy, 1.

[39] Yahya, W. B., Oladiipo, M. O. & Jolayemi, E. T. (2012). A fast algorithm to construct neural networks classification models with high-dimensional genomic data. *Anaa Seria informatica,*10(1), 39-56.

[40] Yue, X. Wu, Y. Wang, Y. Li, & Chu C., (2007). A review of data mining-based financial fraud detection research. International conference on wireless communications Sep, Networking and Mobile Computing. 5519–5522.