Journal of Health, Medicine and Nursing
ISSN 2422-8419 An International Peer-reviewed Journal
Vol.56, 2018
www.iiste.org

Detecting and Combating Fraudulent Health Insurance Claims Using ANN

Ebenezer Larnyo*
Department of Health Policy and Management, Jiangsu University, School of Management, 301 Xuefu Road, Zhenjiang, Jiangsu Province, China

Baozhen Dai (PhD)
Department of Health Policy and Management, Jiangsu University, School of Management, 301 Xuefu Road, Zhenjiang, Jiangsu Province, China

Thomas Bilaliib Udimal
School of Management, Jiangsu University, 301 Xuefu Road, Zhenjiang, Jiangsu Province, China

Wu Chen (Prof.)
School of Computer Science and Engineering, Jiangsu University of Science and Technology, No.2 Mengxi Road, Zhenjiang, Jiangsu, 212003, China

This work was funded by the National Nature Science Foundation of China (71774069), 2014 "Six Talent Peaks" Project of Jiangsu Province (2014- JY-004)

Abstract
While governments and private sector stakeholders are taking steps to improve the access and quality of health care service to its citizenry, a lot of resources are lost every year due to fraudulent health insurance claims. The aim of this paper is to explore a more robust and accurate ways of predicting fraudulent health insurance claims by the use of artificial neural network (ANN). Using the fraud diamond theory (FDT)'s fraud elements as fraud indicators, a fraud prediction model was created to determine whether a claim presented by a subscriber (individual) is fraudulent or non-fraudulent by varying severally the number of epoch, hidden layer number and threshold of the artificial neural network on a 14 input data to obtain an optimal parameter for the model.The model was able to predict accurately 98.98% with an MSE of 0.0086, which outperformed other artificial neural network (ANN) methods used to predict fraudulent health care claims. The incorporation of the capacity indicator of the fraud diamond theory (FDT) makes this model a tool not only for prediction but also pre-empting the occurrence of fraud. This study is the first to adopt the fraud diamond theory's fraud elements as fraud indicators together with artificial neural network (ANN) in predicting fraudulent health insurance claims.
Keywords: health insurance claim, ANN, fraud prediction model, fraud diamond theory

1. Introduction

The health insurance industry has seen significant growth over the years as several non- insurance policy holders are now enrolled onto the national health insurance schemes in order to obtain cheaper, quicker and reliable health care services. However, this significant gain has been bedeviled with the attempts of some individuals and healthcare service providers to outsmart the system by presenting fraudulent claims. In 2015, $3.2 trillion was spent on health care in the USA according to the National Health Care Anti-Fraud Association (NHCAA)'s white paper issued in September, 2017 representing 17.8% of the nation's gross domestic product (GDP) (NHCAA, 2017). Out of this overwhelming expenditure, the institute of medicine of national academies estimated health care fraud to be ranging at about $75 billion (Olsen, Saunders, & Yong, 2010) to an astounding $640 billion (De Rugy & Jason J. Fitchner, 2015) a year.

According to the Association of Certified Fraud Examiners, health care fraud is a crime that involves misrepresenting information, concealing information, or deceiving a person or entity in order to receive benefits, or to make a financial profit (ACFE, 2018). These fraudulent claims ranges from the ones committed by individuals most commonly consists of using insurance that belongs to someone else, adding a person to an insurance policy that is not eligible for insurance coverage, providing false information, failing to remove someone who is no longer eligible from a policy, and many others (ACFE, 2018), while the ones committed by medical providers or practitioners include billing for services and procedures that were not actually provided to the patient, duplicating submission of a claim for the same service when it was only performed once, billing for a different more costly services, falsifying a patient's diagnosis to justify surgeries, etcetera (ACFE, 2018).

In order to curb this canker that threatens the sustainability of health insurance schemes and affordable health care provision, systems for processing electronic claims have been implemented to automatically perform audits and reviews of claims data. These electronic processing systems have been designed to identify areas of claims requiring special attention such as erroneous or incomplete data input, duplicate claims and medically

non-covered services amongst other (Li, Huang, Jin, & Shi, 2008). Though these tremendous breakthrough has been achieved, most of these processing system are usually limited in their capability to detect certain types of fraud due to the fact that most of these systems have to rely solely on predefined simple rules specified by domain experts. It is worth noting that data mining has contributed greatly to eliminating these bottlenecks by incorporating highly sophisticated data search capabilities and statistical algorithms that can be used to predict fraud.

It is refreshing to note that several insurance fraud detection and prediction methods have been researched and implemented to help claims officers and auditors detect and predict falsified or fraudulent claims in the health care sector.

Some of these methods incorporate supervised learning techniques where an algorithm is used to learn the mapping function from the input to the output of a training data. The main goal of the supervised learning is to approximate the mapping function or generalize the input so well that when there is a new input data, there can be a prediction of the output variable for that data hence making the supervised learning methods usually fast and accurate (Donalek, 2011).

1.1 Use of supervised learning techniques in fraud detection

Bayesian network (BN), a supervised learning technique whose weights are determined by how predictive each indicator is for specific types of fraud was used by (Ormerod, Morley, Ball, Langley, & Spenser, 2003) for early detection of insurance fraud. The system used automated knowledge updating to keep pace with dynamically changing fraud by adding new indicators that emerge from patterns of repeated anomalies. Also a K-nearest neighbor (KNN) algorithm with an optimized distance metric in combination with a genetic algorithm were applied to a medical fraud detection problem by (He, Hawkins, Graco, & Yao, 2000) to detect inappropriate practice of service providers and "doctor-shoppers". The genetic algorithm was used to determine the optimal weighting of the features used while the weights were used in the KNN algorithm to identify the nearest neighbor practice profiles after which two rules (the majority rule and the Bayesian rule) were applied to determine the classifications of the practice profiles. The classification methodology achieved good generalization in classifying general practitioners' practice profiles in a test dataset.

Researchers have also adopted another supervised learning algorithm known as the decision tree in the detection of fraud in the healthcare sector. A new multistage methodology was proposed by (Cooper, 2003; Johnson & Nagarur, 2016) for insurance companies to detect fraud committed by providers and patients. The first three stages of the methodology was aimed at detecting abnormalities among providers, services, and claim amounts while the fourth stage integrates the information obtained in the previous three stages into an overall risk measure after which a decision tree based method is used to compute risk threshold values. The decision as to whether a claim is fraudulent or not is made by comparing the risk value obtained in stage four with the risk threshold value resulting in a better performance on real-world insurance data. These real-world insurance data could however become very difficult to interpret due the challenge of overly complex decision trees with the generation of thousands of rules (He, Wang, Graco, & Hawkins, 1997).

In order to address this bottleneck, (He et al., 1997) proposed a three-step so-called "divide and conquer" procedure by first using a clustering algorithm to divide all insurance subscribers' profiles (40,000 insurance subscribers) into groups then a decision tree is built for each group and then converted into a set of rules which is evaluated by establishing a mapping from the rule to a measure of its significance using simple summary statistics such as the average number of claims made in that cluster and the average size of the claims; then extremes (either defined by experts or compared with the overall average) are signaled for further investigation. This procedure reduces significantly the number of rules to 280 summed over ten clusters. An improvement in the supervised learning algorithm was researched by (Ortega, Figueroa, & Ruz, 2006). They proposed a fraud detection system based on the use of one committee of multilayer perceptron neural networks (MLP) for each one of the entities they used in the fraud/abuse problem: medical claims, affiliates, medical professionals and employers. Results of the fraud detection system showed a detection rate of approximately 75 fraudulent and abusive cases per month, making the detection 6.6 months earlier than without the system. Despite these tremendous breakthroughs, there are still issues relating to the robustness of these techniques regarding their ability to accurately classify fraudulent claims and their inability to pre-empt the occurrence of health insurance fraud, largely due to their non-usage of theoretically tested fraud indicators. Thu, this research seeks to create a fraud prediction model by using MLP-NN with the help of the four key fraud indicating elements (Wolfe & Hermanson, 2004) to detect and predict fraudulent health insurance claims focusing on fraud committed by insurance policy holders. The decision to use the multilayer perceptron neural network (MLP-NN) is because of its wide usage in prediction model and its ability to learn from examples only once and after the learning process is done, they are able to catch hidden and strongly non-linear dependencies, even when there is a significant noise in the training set.

2. Methodology

2.1 Variables for estimating models of fraudulent health insurance claims.

This study applies the fraud diamond theory as a theoretical framework to predict fraudulent health insurance claims submitted by subscribers. The choice of the fraud diamond theory (FDT) over the very popular fraud triangle theory (FTT) (Cressey, 1953) is as a result of the enhancement it (FDT) introduces into the detection and prevention of fraud. In FDT, capability is introduced as a fourth element to the three elements proposed in the fraud triangle theory (FTT) by Cressey, which considers an individual's abilities or personal traits that drives him or her to seek an opportunity and exploit it. Capability goes into depth of what it takes mentally to commit fraud hence enabling managers, owners of businesses and employees to better understand fraudsters and their traits thus helping prevent the committal of these frauds.

According to (Wolfe & Hermanson, 2004) many frauds especially some of the multibillion-dollar ones, would not occur without the right person with the right capabilities in place. Whiles opportunity opens the doorway to fraud, pressure and rationalization draws the person towards it. However, the person must have the capability to recognize the open doorway as an opportunity and to take advantage of it by walking through, not just once, but repeatedly.

The variables selected for this study were classified using the four fraud diamond theory (FDT)'s elements (Wolfe & Hermanson, 2004) and the Association of Certified Fraud Examiners' red flags of insurance fraud as contained in the 2018 insurance fraud handbook (ACFE, 2018).

Incentive (Pressure): Incentive or pressure are the indicators that motivates a person to commit fraud in the first place. Pressures arises as a result of an individual having financial problem that he or she is unable to solve through legitimate means hence results to committing illegal acts such as stealing cash, falsifying financial statements or falsifying health insurance claims and injuries to acquire money to solve their financial problems. Some key indicators of financial distress are unemployment (UET), divorce (DIV), failing business (FAB), looming foreclosure (LBF), high debts (HID). Hence, the use of financial distress as a fraud risk indicator representing pressure: UET, DIV, FAB, LBF and HID (Wolfe & Hermanson, 2004).

Opportunity: In order for fraud to be committed, the perpetrator should have perceived some level of weakness in the system that he or she could exploit to his or her advantage. These weaknesses are what is referred to as opportunity. Some of the indicators for perceived opportunity for the committal of fraud are prior claims history (PCH), new claims policy (NCP), hand delivery of documents (HDD), familiarity with claims process (FCP) and weak internal control (WIC) (Wolfe & Hermanson, 2004).

Rationalization: The stage of rationalization is a cognitive stage and requires the fraudster to be able to justify the crime in a way that is acceptable to his or her internal moral sensibilities. Most fraudsters are first-time criminals and do not see themselves as criminals, but rather a victim of circumstance or failure in the system. Rationalizations are often based on external factors, such as a need to take care of a family, or a dishonest employer which is seen to minimize or mitigate the harm done by the crime. Some of the common rationalizations statements used by fraudsters includes: no one will notice more actions (NWN), "I deserve more after all these years" (IDT), "I pay and never use my premium" (IPIU), "Everyone does it" (EDI) and social bond theory (SBT).

Capability: The capability element focuses on the necessary traits or skills and abilities a person has that can influence their ability to commit fraud. It is where the fraudster recognized the particular fraud opportunity and ability to turn it into reality. Position and function of fraudster (PAF), confidence and ego (CAE), coercion and intelligence (CAI), deceit and stress (DAS), relationship to claims or insurance agent (RCIA), CEO's readiness to bend the rules for profit (CRICP) are the supporting elements of capability (Wolfe & Hermanson, 2004).

2.2 Data collection and Processing.

The dataset used to form the basis of this model was obtained from the State of Connecticut's local data catalog website (DATA.GOV) of the department of insurance as a secondary data. The dataset is a listing of consumer complaints filed against insurance companies licensed in Connecticut by individual insurance subscribers for a review of decisions taken against them by these insurance companies. The dataset included about 16,669 complaints on over 64 insurance coverage ranging from auto insurance to medical health insurance. The dataset included a categorization of the various forms or coverage of insurance, the insurance company name, date the case was opened, date the case was closed, reason forming the basis for the rejection or acceptance of claims by the insurance company, the disposition, conclusion, recovery and other categories. However, for the purpose of this research, the concentration for the coverage is on the individual health insurance claims and decisions taken with respect to them. Thus, the total samples we propose to use for creating the prediction model is 199 samples together with the variables derived by the use of the FDT elements constituting a total dataset of 2786 captured daily but not successively for five year (2013-2017). In contrast to model-based techniques, ANNs which we propose to use to create our prediction model are data-driven self-adaptive (through learning) systems, as they do not need any prior assumptions with regards to the models of the scenarios being investigated, or if they do, they

are minimal (Krambia-Kapardis, Christodoulou, & Agathocleous, 2010). ANNs can usually generalize pretty well after they are trained with a sample of the data, which could even be noisy (Haykin, 2009) enabling it to analyze new data in practice.

The data collection and processing is divided into two main parts, the first part is the selection of the target data which was based on the conclusions that was arrived at by the department of insurance of the state of State of Connecticut on the individual cases and the second part is the selection and extraction of the input variables by performing a principal component analysis using IBM SPSS statistics tool based on 20 fraud indicators.

2.3 Data Preparation for constructing the neural network.

The data preparation process consists of pre-processing which includes randomization where the non-fraudulent health insurance claims data set is mixed with the fraudulent data set to avoid bias when training the data set for the network. After the data has been randomized, it is then normalized to standardize the data between 0 (non-fraudulent health insurance claims) and 1 (fraudulent health insurance claims). Once both normalization and randomization of the data is done, the dataset is divided into two sub-data (training and validation data) for the construction of the network. The essence of pre-processing the sample data is to train the multilayer networks to generalize so as to ensure that the input is within the range for which it was trained. The division of the processed sample dataset is made in a proportion of 80 percent (80%) of training data and 20 percent (20%) validation data, this is to prevent validation error.

After the original training, we would use the optimal model to test the validation data set. The result which will consist of the original output from the validation data set and the prediction model will be used to ascertain the accuracy of classification of the result thus if the result obtained is less than 0.5 (>0.5), it will be considered to be fraudulent and when its greater than 0.5 (<0.5), it will be considered non-fraudulent.

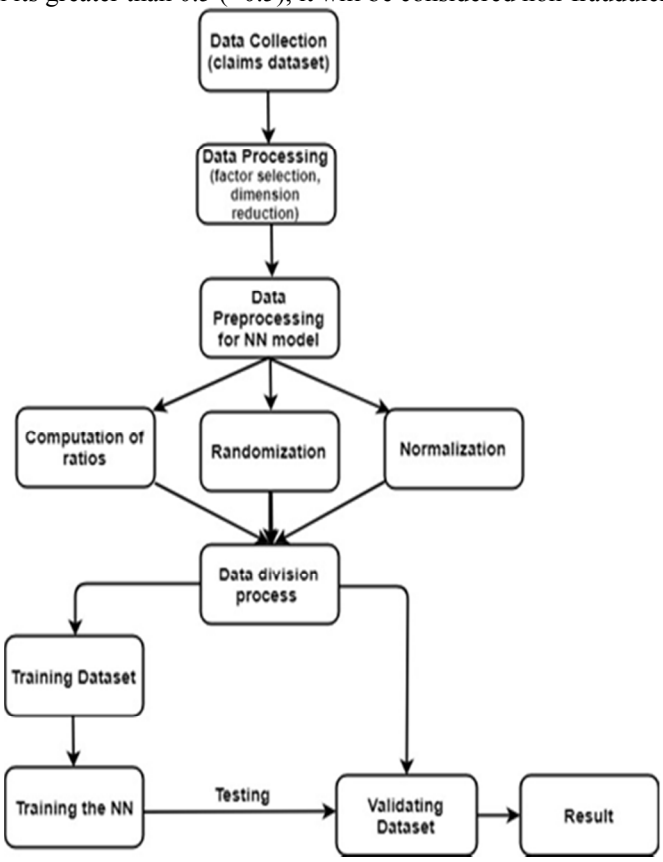

Fig. 1. Flow diagram for data preparation

The Fig. 1 shows the various processes involved in the data preparation through to when the results were obtained.

2.3 Proposed network model for analysis.

This study adopts MLP-NN because it can be used for prediction problems, function fitting and pattern recognition. Multilayer perceptron feed forward neural is one of the supervised networks made up of an interconnection of perceptron in which data and calculations flow in a single direction, from the input data to the outputs. The number of layers in a neural network is the number of layers of perceptron. The advantage of the

usage of neural networks for this prediction model is its ability to learn from examples only and once the learning process is done, they are able to catch hidden and strongly non-linear dependencies, even when there is a significant noise in the training set. Non-linearity is a very important property, particularly, if the relationship between input and output is inherently non-linear. An example picked from the training set is presented to the network, and the weight coefficients are modified to minimize the difference between the desired output and the actual response of the network. The training of the network is repeated for many examples in the training set until the network reaches the stable state. Thus, the network learns from the examples by constructing an input-output mapping for the problem (Omar, Johari, & Smith, 2017).

To achieve an optimized prediction model, several parameters based on the study by (Basheer & Hajmeer, 2000): transfer function, training function, learning function, number of hidden layers, epoch number (EN), Threshold (TH) and the mean squared error (*MSE*) will be adopted.

Training function will also be used to train the neural network for pattern recognition. It is a monotonically increasing, continuous, differential function, applied to the weighted input of a neuron to produce the final output. The values of the weights and biases of the network are tuned to optimize the performance of the network. The first training function used is the gradient descent back propagation (traingd), which is the batch's steepest descent training functions. The weights and biases are updated in the direction of the negative gradient of the performance function. The second training function used is the gradient descent with adaptive learning rule back propagation (traingdm) training function, this function allows the network to respond not only to the local gradient, but also to recent trends in the error surface. Acting like a low-pass filter, the momentum allows the network to ignore small features in the error surface. Without the momentum, a network can get stuck in a shallow local minimum and with the momentum, the network can slide through such a minimum.

One of the fastest back-propagation algorithm which is highly recommended as the first-choice for supervised algorithm called the Levenberg-Marquardt back-propagation (trainlm) would be used as a learning function, although it does require more memory than other algorithms. Feed forward networks often have one or more hidden layers of sigmoid neurons followed by an output layer of linear neurons. Multiple layers of neurons with non-linear transfer functions allow the network to learn non-linear relationships between input and output vectors. The linear output layer is most often used for function fitting (or non-linear regression) problems (Omar et al., 2017).

In our study, the outputs of the network are dichotomous that is 0 for fraudulent claims and 1 for non-fraudulent claims hence log- sigmoid transfer function (logsig) and tan-sigmoid transfer function (tansig) would be used to calculate the output layer. Eight different epoch numbers were selected: 200, 400, 600, 800, 1000, 1200, 1400, 1600, and the values would also be tested based on trial and error, and the epoch number that gives the optimal result was used for the creation of the prediction model. The epoch numbers measure the number of times all of the training vectors are used once to update the weights. For batch training, all of the training samples pass through the learning algorithm simultaneously in one epoch before weights are updated. Table 1 below shows the summary of the proposed parameters.

Table 1. Proposed parameters for this research

Factors	Parameters	Syntax
Training function	Gradient descent back propagation	traingd
	Gradient descent with adaptive learning rule back propagation	traingdm
	Levenberg-Marquardt back propagation	trainlm
Transfer function	Log-sigmoid transfer function	logsig
	Tan-sigmoid transfer function	tansig
Epoch Number	200, 400, 600, 800, 1000, 1200, 1400, 1600	
Threshold	0.72,0.74,0.76,0.78,0.80,0.82,0.84,0.86	
Number of hidden layer	10, 15, 20, 25, 30, 35, 40, 45	

Based on these parameters above, we created a neural network which was used for our prediction as shown in the figure 2.

3. Study Results

3.1 Results of data extraction and selection.

In selecting the best factors to use for our prediction model, the target data was grouped into two main categories; fraudulent and non-fraudulent health insurance claims so as to have a better comparison between these two major variables. The selection and extraction of the input variables was done by performing a principal component analysis using IBM SPSS statistics tool based on 20 fraud indicators.

Table 2. Total Variance Explained

	Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	10.150	53.422	53.422	5.703	30.016	30.016
2	3.557	18.721	72.143	4.556	23.978	53.994
3	1.445	7.604	79.747	3.037	15.985	69.979
4	1.041	5.478	85.225	2.897	15.246	85.225

This helped reduce the 20 variables to a smaller set in order to avoid information overlap while still maintaining the significance of the information in the variables. The Principal Component Analysis (PAC) extracted for the four (4) fraud diamond theory indicators showed a total variance in the scale 53.422%, 18.721%, 7.604%, and 5.478 % with a fixed number of factors to extract as 4 as shown in table 2.

The factor loadings on each of the rotated components that resulted from Varimax rotation are shown in table 3. Since the factor loading for the rotated components were high (mostly above 0.60), for the purpose of this study, we decided to use 14 variables with loadings above 0.722 thus variables X1, X2, X3, X5, X7, X8, X10, X11, X13, X14, X15, X16, X18, X20).

Table 3. Rotated Component Matrix

	Component			
	1	2	3	4
X16	.908			
X1	.908			
X15	.901			
X2	.874			
X14	.850			
X6	.663			
X8		.882		
X10		.878		
X20		.828		
X7		.805		
X3		.753		
X13			.809	
X18			.786	
X19			.722	
X4			.701	
X5				.856
X11				.848
X9				.697
X12				.688

3.2 Results of Prediction Model.

Table 4. Optimal values

Factors	Optimal Values
Transfer Function	logsig
Training Function	trainlm
Learning Function	learngd
Hidden Layer Number	10
Epoch Number	400
Threshold	0.83
Mean Squared Error	0.0086
R^2	0.9889

Table 4 shows optimal values of the prediction model. The transfer function: logsig, training function: trainlm, learning function: learngd were used to obtain the optimal values. Additionally, an epoch number of 400, hidden layer number of 10, threshold of 0.83 resulted in a mean squared error of 0.0086 and a regression of 0.9889.

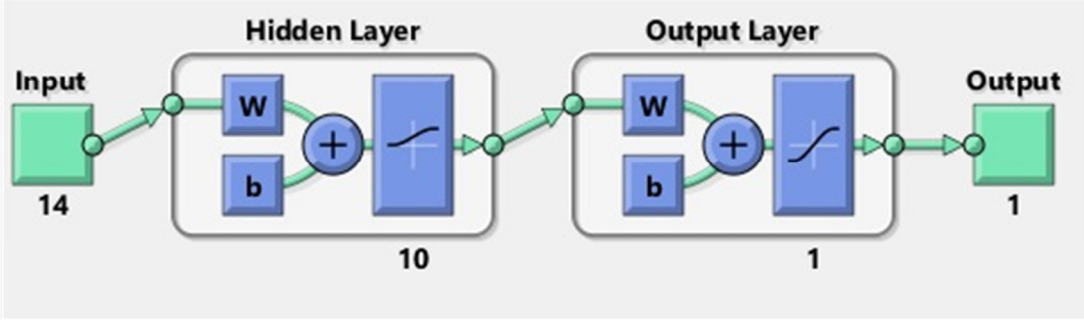

Figure 2 Created MLP feed forward Neural Network using MATLAB R2017a

Figure 2 above shows the network diagram created with the optimal parameters using MATLAB R2017a.

4. Discussion

This study applied a multilayer perceptron feed-forward neural network to create a model for the prediction of fraudulent health insurance claims. The model adjusted the weights and biases based on gradient-descent back-propagation to train the network. The mean square error was chosen as the statistical criteria for measuring of the model's performance. After several trial and error process using all the parameters as shown in Table 1, we obtained *MSE* of .0086. The mean square error helps to measure the average of the squares of the errors or deviations which is the difference between the network's output and the target. The regression analysis, R^2 after several trial and error process while using all the parameters as shown in Table 1, produced an R^2 of .9889. The R^2 helps us to compute the sum of squares of the residuals and the total sum of the squares which is proportional to the variance of the data. In using both logsig and tansig, this study obtained similar result of *MSE* which is at .0086 implying that there is no statistical difference in either using logsig or tansig to differentiate the transfer function of the neural network model. Thus, logsig was chosen as it helped to obtain our optimal result due to its ability to conform to minimum or normal runoffs (Dorofki, Elshafie et al. 2012). The results of the training function also showed no significant difference in *MSE* between each of training function factor values where all three factors used showed *MSE* of .0086. This an indication that any one of the three factors could be used to develop the prediction model. Also, this study used Levenberg Marquardt back propagation (trainlm) for the prediction model as it is able to minimize the mean square error function (Sapna, Tamilarasi, & Kumar, 2012) and performs better on function fitting.

After several trial and error processes, the results of hidden layer neural network that were obtained showed no statistical difference in *MSE* values as all eight hidden layer numbers generated *MSE* of .0086. This means that any one of the eight hidden layer numbers could be used to develop the prediction model. Hence, ten (10) was chosen as the number of hidden layer that is suitable for the prediction model. This choice was as a result of the 10 layer ability to generalize the data and could properly fit the input data. Also, the fewer the number of the hidden neuron of the network, the less incapable it is for the network to differentiate between complex patterns, leading to only a linear estimate of the actual trend. Conversely, if the network has too many hidden neurons, it will follow the noise in the data to over parameterize, leading to poor generalization for untrained data thereby making training to become excessively time-consuming. Results of epoch number (EN) show that there was no difference in *MSE* between each EN. All eight epoch numbers showed *MSE* of .0086. Epoch number helps in the training iterations and update of the weights. The reason for the use of the different numbers of epochs is to enable the network to stop the training process once the number of iterations exceeds epochs. Since the result of *MSE* for each epoch number is statistically the same, the study uses an epoch number of 400 to develop a prediction model as it could generalize the performance of the prediction model. Threshold values 0.72, 0.74, 0.76, 0.78, 0.80, 0.82, 0.84 and 0.86 were also used to compare the output value. Threshold function is regarded as an active parameter that continues to vary in accordance with input. Results of the threshold show that there is no difference in *MSE* between each Threshold value used as all show *MSE* of .0086. This study uses 0.83 to create the fraud prediction model.

The original dataset representing the target set of the validation data were compared with the actual result of the simulation to test the predictive ability of the model. The ability of the result of the ANN prediction model (*MSE* of .0086 and R^2 of .9889 or 98.89 per cent) to predict fraudulent and non-fraudulent insurance claims shows that the model is able to predict all 199 dataset that were separated for validation and testing. Also, the results are indicative of the ability of the prediction model to accurately classify fraudulent and non-fraudulent claims. For the dichotomous dependent variable, it was coded that when the result of the ANN is above 0.5 it is considered as non-fraudulent insurance claims. If the ANN result shows lower than 0.5 it is considered as fraudulent insurance claims. This is how the ANN results could be used to predict if the insurance claims is fraudulent or whether it contains any falsifications.

The findings of this study demonstrate that the fraudulent insurance claims prediction model is able to predict fraudulent insurance claims correctly at 98.89% (R^2 = .9889). The *MSE* that classify fraudulent insurance claims is at .0086 (*MSE* =.0086). This shows that the model could provide higher predictive level compared to previous studies that use ANN to predict fraudulent insurance claims. The use of fraud diamond theory's elements that represent fraud risk indicators for this study demonstrates the model's robust nature.

Table 5. Summary of prediction result compared with other previous studies.

Author	Techniques	Prediction Result (%)
He et al., 1997	MLP-NN	80.93
Liou, Fen-May et. al 2008	Neural Network	96.00
Liou, Fen-May et. al 2008	Logistic regression model	92.00
Current study	MLP-NN with FDT	98.98 (MSE= 0.0086)

In order to provide benchmarking, the result of this study was compared with other model results were either neural network or other supervised learning methods were used to predict fraud as shown in table 5. The comparison showed that this study's predictive percentage of 98.99 and MSE of 0.0086 outperformed other techniques such as Liou et. al. 2008 who applied data mining techniques to detect fraudulent or abusive reporting by healthcare providers using their invoices for diabetic outpatient services. Also our model's false prediction is only 1.11% compared to (Ortega, Figueroa et al. 2006)'s 6%, which goes to colloborate about 98.98% of the decisions that was taken by the department of insurance. One of the difficulties usually faced in the fight against preventing fraudulent claims in the health care sector is the classification and the determination of what actually constitute fraud, however in our model we were able to incorporated the four key fraud elements: pressure, opportunity, rationalization, and capacity (Wolfe and Hermanson 2004) which are normally used in predicting and detecting fraud thus making our model very robust.

5. Implications of study

While several researchers have relied on the ability of ANN to classify and predict fraud in several sectors, only few studies have done in health care sector to detect fraudulent health insurance claims, particularly using fraud theories (either fraud triangle theory or fraud diamond theory) indicators together with ANN to predict fraud. To the best of our knowledge, no research has attempted to use these key fraud indication element of the fraud diamond theory to predict fraud, hence this study contributes substantially to the literature deficit in the area of health insurance claims fraud detection. The predictive ability of this model suggests that, it is able to classify future studies having dichotomous variables of fraud and non-fraud and is able to predict more accurately whether an individual or company dealing with health insurance claims are committing fraud through their claims processing.

6. Conclusions

It is evident that the amount of resources (money) lost through fraudulent health insurance claims, is significant enough to have a negative effect on the GDP of a country, hence derailing the effort of governments and stakeholders in the health care sector to deliver quality, affordable and quicker health care service to its citizenry. Thus, to equip auditors, claims officers and other industry players in the area of health insurance claims processing with the necessary tool to predict and combat fraudulent health insurance claims, we created a prediction model with MLP-NN and fraud diamond theory (FDT). The reason for the usage of neural networks other than other statistical models for this prediction model is its ability to learn from examples only and once the learning process is done, they are able to catch hidden and strongly non-linear dependencies, even when there is a significant noise in the training set hence giving players the advantage in the dealing with the different characteristic of fraud and fraudsters. The results of our model indicates a prediction accuracy of 98.98% meaning the false positive prediction of our model is only 1.11% which is better than other previous studies, also our *MSE* was 0.0086 indicating the capability of our model to predict fraudulent health insurance claims with minimal error thus justifying about 98.98 % of the parameters that were used in taking final decisions on the claims complaints that were submitted to the department of insurance as were in the original dataset.

References

ACFE, A. O. C. F. E. (2018). Insurance Fraud Handbook. The Gregor Building, 716 West Avenue, Austin, Texas 78701 USA, Association of Certified Fraud Examiners, Inc.

Basheer, I. A. and M. Hajmeer (2000). "Artificial neural networks: fundamentals, computing, design, and application." Journal of microbiological methods 43(1): 3-31.

Cooper, C. (2003). Turning information into action. Computer Associates: The Software That Manages eBusiness, Report.

Cressey, D. R. (1953). "Other people's money; a study of the social psychology of embezzlement."

DATA.GOV. "Insurance Company Complaints, Resolutions, Status, and Recoveries." Retrieved 2nd January,

2018, from https://data.ct.gov/Business/Insurance-Company-Complaints-Resolutions-Status-an/t64r-mt64.
De Rugy, V. and Jason J. Fitchner. (2015, January 13, 2015). "“Is Federal Spending Too Big to be overseen?” “Retrieved January, 11, 2018, from http: //mercatus.org/publication/federal-spending-too-big-be-overseen.
Dorofki, M., A. H. Elshafie, O. Jaafar, O. A. Karim and S. Mastura (2012). "Comparison of artificial neural network transfer functions abilities to simulate extreme runoff data." International Proceedings of Chemical, Biological and Environmental Engineering 33: 39-44.
Haykin, S. S. (2009). Neural networks and learning machines, Pearson Upper Saddle River, NJ, USA:.
He, H., S. Hawkins, W. Graco and X. Yao (2000). "Application of Genetic Algorithm and K-Nearest Neighbour Method in Real World Medical Fraud Detection Problem." JACIII 4(2): 130-137.
He, H., J. Wang, W. Graco and S. Hawkins (1997). "Application of neural networks to detection of medical fraud." Expert Systems with Applications 13(4): 329-336.
Johnson, M. E. and N. Nagarur (2016). "Multi-stage methodology to detect health insurance claim fraud." Health care management science 19(3): 249-260.
Krambia-Kapardis, M., C. Christodoulou and M. Agathocleous (2010). "Neural networks: the panacea in fraud detection?" Managerial Auditing Journal 25(7): 659-678.
Li, J., K.-Y. Huang, J. Jin and J. Shi (2008). "A survey on statistical methods for health care fraud detection." Health care management science 11(3): 275-287.
Liou, F.-M., Y.-C. Tang and J.-Y. Chen (2008). "Detecting hospital fraud and claim abuse through diabetic outpatient services." Health care management science 11(4): 353-358.
NHCAA, N. H. C. A.-F. A. (2017). The U.S. Health Care System and the Challenges of Fraud, National Health Care Anti-fraud Association.
Olsen, L., R. S. Saunders and P. L. Yong (2010). The healthcare imperative: lowering costs and improving outcomes: workshop series summary, National Academies Press.
Omar, N., Z. A. Johari and M. Smith (2017). "Predicting fraudulent financial reporting using artificial neural network." Journal of Financial Crime 24(2): 362-387.
Ormerod, T., N. Morley, L. Ball, C. Langley and C. Spenser (2003). Using ethnography to design a Mass Detection Tool (MDT) for the early discovery of insurance fraud. CHI'03 Extended Abstracts on Human Factors in Computing Systems, ACM.
Ortega, P. A., C. J. Figueroa and G. A. Ruz (2006). "A Medical Claim Fraud/Abuse Detection System based on Data Mining: A Case Study in Chile." DMIN 6: 26-29.
Wolfe, D. T. and D. R. Hermanson (2004). "The fraud diamond: Considering the four elements of fraud." The CPA Journal 74(12): 38.