

# Use of Artificial Neural Network for Forecasting Health Insurance Entitlements

Sam Goundar, The University of South Pacific, Fiji

 <https://orcid.org/0000-0001-6465-1097>

Akashdeep Bhardwaj, University of Petroleum and Energy Studies, India\*

 <https://orcid.org/0000-0001-7361-0465>

Suneet Sonal Prakash, The University of South Pacific, Fiji

Pranil Sadal, The University of South Pacific, Fiji

## ABSTRACT

A number of numerical practices exist that actuaries use to predict annual medical claims expense in an insurance company. This amount needs to be included in the yearly financial budgets. Inappropriate estimating generally has negative effects on the overall performance of the business. This paper presents the development of artificial neural network model that is appropriate for predicting the anticipated annual medical claims. Once the implementation of the neural network models were finished, the focus was to decrease the mean absolute percentage error by adjusting the parameters such as epoch, learning rate, and neuron in different layers. Both feed forward and recurrent neural networks were implemented to forecast the yearly claims amount. In conclusion, the artificial neural network model that was implemented proved to be an effective tool for forecasting the anticipated annual medical claims. Recurrent neural network outperformed feed forward neural network in terms of accuracy and computation power required to carry out the forecasting.

## KEYWORDS

Artificial Neural Network, Back Propagation, Claims. Data Normalization, Feed Forward, Health Insurance, Sigmoid, Time Series

## 1. INTRODUCTION

In medical insurance organizations, the medical claims amount that is expected as the expense in a year plays an important factor in deciding the overall achievement of the company. BSP Life (Fiji) Ltd. provides both Health and Life Insurance in Fiji. Medical claims refers to all the claims that the company pays to the insured's, whether it be doctors' consultation, prescribed medicines or overseas treatment costs. Claims received in a year are usually large which needs to be accurately considered when preparing annual financial budgets. These claim amounts are usually high in millions of dollars every year. An increase in medical claims will directly increase the total expenditure of the company

DOI: 10.4018/JITR.299372

\*Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

thus affects the profit margin. Currently utilizing existing or traditional methods of forecasting with variance. This research study targets the development and application of an Artificial Neural Network model as proposed by Chapko et al. (2011) and El-said et al. (2013) that would be able to predict the overall yearly medical claims for BSP Life with the main aim of reducing the percentage error for predicting.

According to Rizal et al. (2016), Neural Network is very similar to biological neural networks. Neural networks can be distinguished into distinct types based on the architecture. Two main types of Neural Networks are namely feed forward Neural Network and Recurrent Neural Network (RNN). Artificial neural networks (ANN) have proven to be very useful in helping many organizations with business decision making. Example, Sangwan et al. (2020) proposed Artificial Neural Network is commonly utilized by organizations for forecasting bankruptcy, customer churning, stock price forecasting and in many other applications and areas. This research focusses on the implementation of Multi-Layer Feed Forward Neural Network with back propagation algorithm based on gradient descent method. The network was trained using immediate past 12 years of medical yearly claims data. Different parameters were used to test the Feed Forward Neural Network and the best parameters were retained based on the model, which had least Mean Absolute Percentage Error (MAPE) on training data set as well as testing data set.

In the insurance business, two things are considered when analysing losses: frequency of loss and severity of loss. Previous research investigated the use of artificial Neural Networks (NNs) to develop models as aids to the insurance underwriter when determining acceptability and price on insurance policies. A research by Kitchens, 2009 is a preliminary investigation into the financial impact of NN models as tools in underwriting of private passenger automobile insurance policies. Results indicate that an artificial NN underwriting model outperformed a linear model and a logistic model. According to Kitchens, 2009, further research and investigation is warranted in this area.

In the past, research by Mahmoud et al. (2013) and Majhi S (2018) on Recurrent Neural Networks or RNN have also demonstrated that it is an improved forecasting model for time series. To demonstrate this, NARX model (Nonlinear Autoregressive Network having Exogenous Inputs), is a recurrent dynamic network was tested and compared against Feed Forward Artificial Neural Network. Abhigna et al. (2017) state that Artificial Neural Network (ANN) has been constructed on the human brain structure with very useful and effective pattern classification capabilities. ANN has the ability to resemble the basic processes of human's behaviour which can also solve nonlinear matters, with this feature Artificial Neural Network is widely used with complicated system for computations and classifications, and has cultivated on non-linearity mapped effect if compared with traditional calculating methods. According to Zhang et al. (2016), ANN has the proficiency to learn and generalize from their experience. The authors Motlagh et al. (2016) emphasize that the idea behind forecasting is previous know and observed information together with model outputs will be very useful in predicting future values. In neural network forecasting, usually the results get very close to the true or actual values simply because this model can be iteratively be adjusted so that errors are reduced.

Viaene et al. (2005) explored the explicative capabilities of neural network classifiers with automatic relevance determination weight regularization. Reports the findings from applying these networks for personal injury protection automobile insurance claim fraud detection. The automatic relevance objective function scheme provides us with a way to determine which inputs are most informative to the trained neural network model. An implementation of MacKay's, (1992) evidence framework approach to Bayesian learning is proposed as a practical way of training such networks. The empirical evaluation was based on a data set of closed claims from accidents that occurred in Massachusetts, USA during 1993.

The first type of Artificial Neural Network is a Feed Forward Neural Network. The basic layout of multi-layer feed forward neural network has three layers, which are input layer, middle layer is known as hidden layer and one output layer. Input layer, which is independent, receives the data as input.

Hidden layer transforms the input into objects (Abdel et al., 2013) that the output layer can utilize. Output layer converts the hidden layer activations into the scale that has been specified by the user. According to Abhishek et al. (2012) Multi-layer networks utilize a number of learning techniques as proposed by Salama et al., (2014), with the most common being back-propagation. The output values are usually compared with the actual value to get the error rate. The calculated value is then put through the network as the input. This way the network adjusts the weight of each connector so that the error could be reduced. After several iterations, the values will eventually come to a state where the error is very small. Some other factors, which needs to be thoroughly considered in the process of designing the neural network, are Epoch (iterations), weight initialization, Neurons in the hidden layer, learning rate and the momentum constant or bias value. Training process usually involves adjusting these parameters for maximum accuracy.

Ajibola et al. (2012) looks at using artificial neural network's primary role of providing financial protection for other industries by the insurance industry. In spite of the harsh economic environment in Nigeria, the insurance industry has been crucial to the consummation of business plans and wealth creation. However, the continued downturn experienced by many countries, in the last decade, seems to have affected negatively on the financial health of the industry, thereby rendering many insurance companies inherently distressed. Although there is a regulator to monitor the insurance companies in order to prevent insolvency and protect the right of consumers this oversight function has been made difficult because the regulators appeared to lack the necessary tools that would adequately equip them to perform their oversight functions. One such critical tool is a decision making model that provides early warning signal of distressed firms. Their research constructs an insolvency prediction model based on artificial neural network approach, which could be used to evaluate the financial capability of insurance companies.

The second type of Artificial Neural Network is known as Recurrent Neural Network (RNN). The authors Zhang et al. (2016) state that unlike Feed Forward Neural Networks, Recurrent Neural Networks (RNN) are bi-directional where the output is fed back to the previous units. The feedback enables the neural network models to be trained and process sequences dynamically, which would have better prediction. According to Khan et al (2013) recurrent neural networks have inbuilt dynamic behaviour, which is why it is, referred as special neural networks. The main difference between the feed forward neural networks is the feedback path (s) where the output is fed back into the system as an input. The inter-relationship between the inputs and the internal state is processed for creating the output after the training. Learning process is supervised, where the target values are always the second source of information.

This research has been organized as follows. Section 2 presents literature survey of previously published research on the same domain and topic. Section 3 illustrates initial investigation and raw data collection process. Section 4 presents the Research Methods followed which includes performing an auto-correlation, data normalization, algorithm training of the test data and then checking the design for network accuracy and the stopping criteria. Section 5 is based on actual implementation and compares two models Feed Forward Neural Network and Recurrent Models. Section 6 compares the neural network models and the Human Prediction.

## 2. Literature Survey

Dong et al. (2019) presented the computational framework for Reservoir Computing using Recurrent Neural Network (RNN) having fixed weights. The authors proposed several physical implementations to improve the energy efficiency and speed using a scattering model.

Dumas et al. (2019) performed a review for a set of neural network architectures for intra image predictions. The authors proved that fully connected neural networks tend to provide better performance for small block sizes, while convolutional neural networks had better predictions for large complex textures blocks.

Liu et al. (2019) proposed Deep learning based neural network prediction model for image compression. The authors initially articulated prediction of multi-class classification problem and a framework that transformed the multi-class classification into binary classification problem solved by just one binary classifier. Then the authors constructed the deep learning based binary classifier to predict if an image is lossy with another or not. The authors also proposed a sliding window search strategy. This helped predict results for the lossless predictor. Experimental results show that the mean accuracy of the perceptually lossless predictor. The research displayed superior results for the proposed neural model as compared to the conventional models.

Random rough subspace based neural network ensemble method was proposed by Xu, Wang, Zhang, & Yang (2011) for insurance fraud detection. In this method, rough set reduction is firstly employed to generate a set of reductions, which can keep the consistency of data information. Secondly, the reductions are randomly selected to construct a subset of reductions. Thirdly, each of the selected reductions is used to train a neural network classifier based on the insurance data. Finally, the trained neural network classifiers are combined using ensemble strategies. For validation, a real automobile insurance case is used to test the effectiveness and efficiency of our proposed method with two popular evaluation criteria including the percentage correctly classified (PCC) and the receive operating characteristic (ROC) curve. The experimental results show that our proposed model outperforms single classifier and other models used in comparison. The findings of this study reveal that the random rough subspace based neural network ensemble method can provide a faster and more accurate way to find suspicious insurance claims, and it is a promising tool for insurance fraud detection.

Dong et al. (2019) proposed a novel neural network model for health related prediction model for whole population and individuals for blood glucose fluctuations. The authors integrated pre-training and fine-tuned processes to overcome the problem of insufficient data for individual patient and making full use of the population and individual differences and fluctuations. When compared with other machine learning and neural network approaches, the numerical results suggest that the proposed approach gains significant improvements on prediction accuracy.

Dong et al. (2019) proposed a new neural prediction approach based on recurrent neural networks (RNN). The authors incorporated pre-processing of clustering into the classical RNNs. Numerical results suggested that the proposed approach utilized more than one cluster for both type I and type II datasets and has gained improvements compared with support vector regression (SVR) and other RNN methods in terms of prediction accuracy.

Epilepsy is a neurological disorder associated with abnormal electrical activity in the brain, which causes seizures. The occurrence of seizure is not predictable; the duration between seizures, as well as the symptoms, varies from patient to another. Nashaat et al. (2019) designed and implemented a monitoring system for epileptic patients; the system should continuously check some vital signs, analyse the measurements, and decide whether the patient is nearly to have a seizure or not. Whenever a seizure is predicted, the system initiates an alarm. In addition, a notification should be sent to the health care responsible, as well as one preferred contact. By implementing the monitoring system, people who suffer from epilepsy will have more chance to work and live a normal life. Thus, this research study presents the concept of the overall system and shows results of the implemented systems: EEG, ECG and Fall Detection system. Results have shown that the fall detection accuracy reached 99.89% whereas the accuracy of the prediction using the ANN was about 97.34%.

Liou, Tang, & Chen (2008) looked at the usage of artificial neural networks in hospitals and health care providers. Because of the exaggerated and fraudulent medical claims initiated by national insurance schemes, artificial neural networks can be applied for detection. Their study applied data mining techniques to detect fraudulent or abusive reporting by healthcare providers using their invoices for diabetic outpatient services. This research was pursued in the context of Taiwan's National Health Insurance system. We compare the identification accuracy of three algorithms: logistic regression, neural network, and classification trees. While all three are quite accurate, the classification tree model

performs the best with an overall correct identification rate of 99%. The neural network (96%) and the logistic regression model (92%) follow it.

Predicting indoor air quality becomes a global public health issue. Commercial organizations have developed a smart connected object, which is able to measure different physical parameters including concentration of pollutants (volatile organic compounds, carbon dioxide and fine particles). This smart object must embed prediction capacities in order to avoid the exceedance of an air quality threshold. This task is actually performed by neural network models. However, when some events occur (change of people's behaviours, change of place of the smart connected object as example), the embedded neural models become less accurate. So a relearning step is needed in order to refit the models. The smart connected object must perform this relearning, and therefore, it must use the less computing time as possible. Thomas et al. (2019) proposed combining a control chart in order to limit the frequency of relearning, and to compare three learning algorithms (backpropagation, Levenberg-Marquardt, neural network with random weights) in order to choose the more adapted to this situation.

Jun et al. (2019) proposed a general framework that incorporates effective missing data imputation using VAE and multivariate time series prediction. We utilize the uncertainty obtained from the generative network of the VAE and employ uncertainty-aware attention in imputing the missing values. We evaluated the performance of our architecture on real-world clinical dataset (MIMIC-III) for in-hospital mortality prediction task. Our results showed higher performance than other competing methods in mortality prediction task.

Mahdi et al. (2019) presented Deep Learning technique called Long Short Term Memory (LSTM) recurrent neural networks to find sessions that are prone to code failure in applications that rely on telemetry data for system health monitoring. The authors used LSTM networks to extract telemetry patterns that lead to a specific code failure. For code failure prediction, treating the telemetry events, sequence of telemetry events and the outcome of each sequence as words, sentence and sentiment in the context of sentiment analysis, respectively. The proposed method is able to process a large set of data and can automatically handle edge cases in code failure prediction. The authors took advantage of Bayesian optimization technique to find the optimal hyperactive parameters as well as the type of LSTM cells that leads to the best prediction performance. The authors introduced the Contributors and Blockers concepts. In this research, contributors are the set of events that cause a code failure, while blockers are the set of events that each of them individually prevents a code failure from happening, even in presence of one or multiple contributor(s). Once the proposed LSTM model is trained, we use a greedy approach to find the contributors and blockers. To develop and test our proposed method, we use synthetic (simulated) data in the first step. The synthetic data is generated using a number of rules for code failures, as well as a number of rules for preventing a code failure from happening. The trained LSTM model shows over 99% accuracy for detecting code failures in the synthetic data. The results from the proposed method outperform the classical learning models such as Decision Tree and Random Forest. Using the proposed greedy method, we are able to find the contributors and blockers in the synthetic data in more than 90% of the cases, with a performance better than sequential rule and pattern mining algorithms.

According to Yunos, et al. (2016), expected claim frequency and the expected claim severity were used in predictive modelling for motor insurance claims. There are two category of claims were considered, namely, third party property damage (TPPD) and own damage (OD). Data sets from the year 2001 to 2003 are used to develop the predictive model (Mnasser et al., 2014). The main issues in modelling the motor insurance claims are related to the nature of insurance data, such as huge information, uncertainty, imprecise and incomplete information; and classical statistical technique (Azar et al., 2014) which cannot handle the extreme value in the insurance data. Their research proposes the back propagation neural network (BPNN) model as a tool to model the problem. A detailed explanation of how the BPNN model solves the issues is provided.

Zhang et al. (2019) proposed a hybrid neural network model, which combines Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN). The proposed model extracted local features and captured the degradation process. In order to show the effectiveness of the proposed approach, tests on the NASA Commercial Modular Aero-Propulsion System Simulation (C-MAPSS) dataset of turbofan engine. The experimental results show that the proposed CNN-RNN hybrid model achieves better score values than the Multilayer Perceptron (MLP), Support Vector Regression (SVR) and (CNN) on FDOOI, FD003 and FD004 data sets.

Kurtah et al. (2019) presented their research to provide a system that displays in real time the disease status. The system also predicts the propagation of diseases allowing the concerned health ministry to better plan remedial actions. The system consists of three mobile crowdsourcing applications that allow the public, doctors and pharmacies to report diseases and drugs sales in real time. Data regarding diseases for the year 2017 were retrieved and the corresponding daily weather information namely as temperature, humidity and wind for that year was then extracted and added to this dataset. An Artificial Neural Network (ANN) was then trained with this dataset and then used to predict the propagation of the diseases, which can be monitored, by the Ministry of Health and Quality of Life through another application. The prediction was performed based on the number of reported diseases on the current day along with weather forecasts for the forthcoming days and the results were promising. The model has been evaluated resulting in an accuracy of 90%. Finally, we believe that such a system can be very beneficial to the ministry, which can then take informed decisions to counteract the possible propagation of diseases.

Chuanchai et al. (2019) proposed a framework for spatially predicting the particulate matter concentration in the area without monitoring station. The proposed framework consisted of two components. One is a particulate matter monitoring station deployed in a reference location, and other is a spatial prediction model to apply spatial interpolation technique and machine learning technique to provide the particulate matter concentration value in the area without monitoring station. This study also explores the results from the variety of components in the model. Two spatial interpolation techniques, namely IDW: Inverse Distance Weigh and Kriging are compared. The evaluation results show that the model can spatially predict particulate matter concentration value with the average 10.16% error by using the Kriging technique with seven inputs for machine learning.

Decaro et al. (2019) illustrated application of machine learning techniques to predict hematic parameters using blood visible spectra during ex-vivo treatments. A spectroscopic setup was prepared for acquisition of blood absorbance spectrum and tested in an operational environment. This setup is

Figure 1 Medical Claims Amount Paid for each year

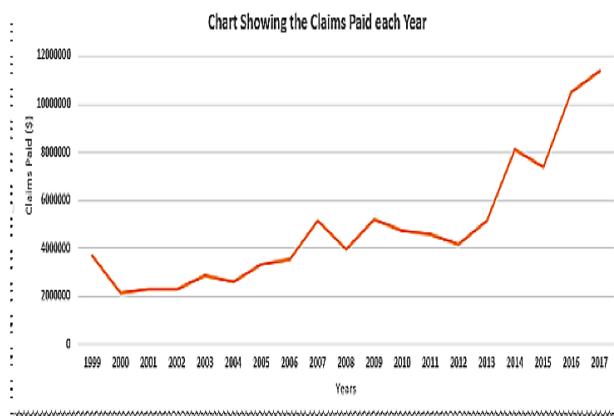


Figure 2 Medical Claims Paid for each Quarter

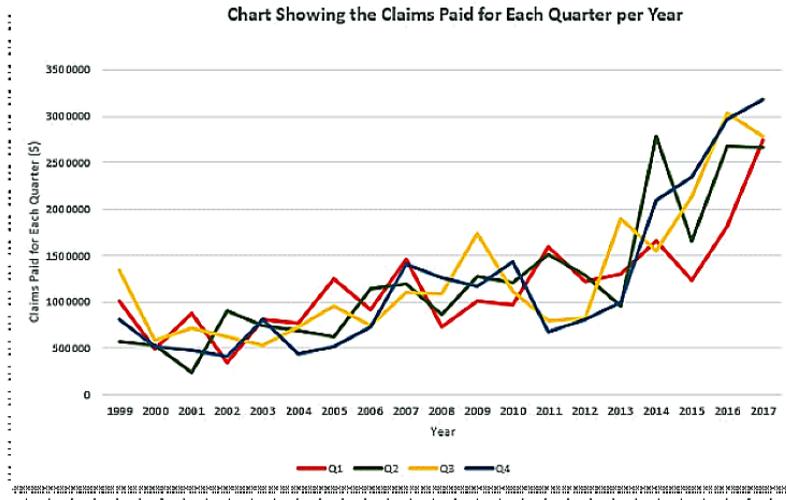


Table 1. Yearly Claims Figure and Feature Summary

Year	Claims Paid	Data	Average Age	Acute	Chronic
2001	3724016.038	Training	37.86	1677712.204	2046303.834
2002	2137491.717	Training	38.73	372907.245	1764584.472
2003	2318647.2	Training	41.48	128182.3	2190464.9
2004	2289772.26	Training	40.9	263764.24	2026008.02
2005	2892552.57	Training	39.85	125446.68	2767105.89
2006	2630915.71	Training	40.41	235286.64	2395629.07
2007	3337894.81	Training	40.78	243860.186	3094034.626
2008	3528283.74	Training	38.44	556824.34	2971459.4
2009	5161877.31	Training	40.73	921796.24	4240081.07
2010	3936563.29	Training	38.79	1217478.56	2719084.73
2011	5181380.24	Training	39.58	1014753.33	4166626.91
2012	4731157.85	Training	40.37	1159020.06	3572137.79
2013	4581545.17	Training	39.81	896958.86	3684586.31
2014	4140205.86	Test	36.35	1370092.91	2770112.95
2015	5159114.67	Test	31.48	1868632.78	3290481.89
2016	8094810.18	Test	31.89	3028274.76	5066535.42
2017	7378419.19	Test	31.35	1681138.83	5697280.36
2018	10483842.3	Test	32.72	3766492.5	6717349.8
2019	1362799.91	Test	32.16	4859328.41	6503471.5

non-invasive and can be applied during dialysis sessions. A support vector machine and an artificial neural network, trained with a dataset of spectra, have been implemented for the prediction of haematocrit and oxygen saturation. Results of different machine learning algorithms are compared, showing that support vector machine is the best technique for the prediction of haematocrit and oxygen saturation.

### 3. INITIAL INVESTIGATION

Before the actual implementation of the models, there was a need to carry out some analysis on the raw data and have formal research methods. BSP LIFE (Fiji) Ltd provided the raw claims data, relationships between the data sets needed to be identified. The raw data was summarized yearly to explore the trends in the yearly claimed amount paid. To understand and observe the overall trend between the annual medical claims figure, the aggregated sum of medical claims amount paid for each year was calculated. Figure 1 highlights the yearly trend of medical claims paid out.

The yearly medical claims amount was divided into the 4 quarters as the next step of analysis. After dividing and summarizing the results into quarters, observations concluded that quarterly series had noise and could not be used as the input for forecasting. The accuracy of the prediction model that would be designed later would also be affected negatively if quarterly amount were to be used. Figure 2 shows the trends of quarterly medical claims series.

A number of features are extracted from the raw data, which encompassed of average age of the insured, acute disease amount and chronic disease amount for each year. Table 1 shows the summary of features extracted.

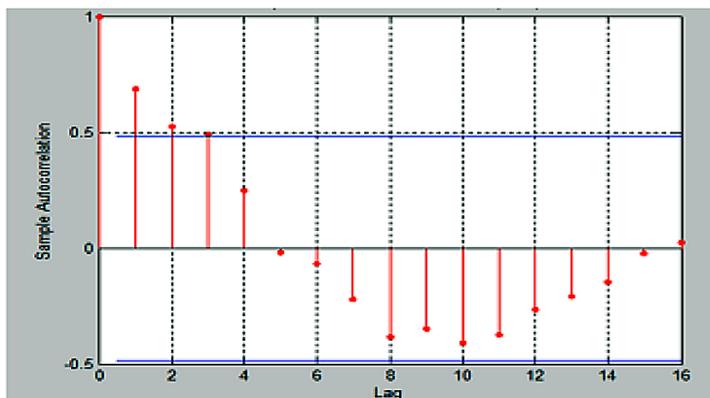
### 4. RESEARCH METHODS

The research methods initiated by taking the raw data, performing an auto-correlation and then initiating the data normalization followed by training the test data. Finally network design was checked for network accuracy and the stopping criteria. The below section illustrates the implementation process in form of figures and graphs.

#### 4.1 Auto- Correlation

Auto- Correlation is a depiction of the extent of similarity between a time series and a lagged version of itself over consecutive time intervals. In this research, Correlation was used to determine the window size of the input to the network model that would be developed. MatLab software was used to plot the auto-correlation of the series, which highlighted the positive correlation in 1, 2 and 3 lag

Figure 3 Auto-correlation plot of the yearly claims series



variables. The Lags were used to identify the number of input to the network i.e. the number of years of claims amount needed as the input. Figure 3 shows the plot from Mat-Lab which was achieved by using the correlation function.

### 4.2 Data Normalization

The process of scaling numeric data into a new range of values which are usually between [-1, 1] or [0, 1] is known as Data normalization. According to the authors Abhigna, Jerritta, Srinivasan and Rajendran (2017), a number of normalization methods exists which can be used for data normalization such as Min- Max Normalization, Median normalization, Stastical or Z-Score Normalization, Statistical Column Normalization and Sigmoid Normalization. In this research, Min-Max Normalization technique was utilized so that the claims value is scaled between the range of [0, 1]. Since sigmoid activation function was used at the hidden and output layer where the inputs were required to be normalized between 0 to 1, the authors used data normalization. The formula that was used to carry out the normalization was:

$$X_{norm} = (X - X_{min}) / (X_{max} - X_{min})$$

$X$  in the formula denotes the actual value to be normalised,  $X_{min}$  refers to the minimum value in the data set and  $X_{max}$  refers to the maximum value. Min-Max normalization had the advantage of retaining exactly all relationships in the data after the scaling which is why this is used. After the pre- processing of medical raw claims data were successfully completed, the next step required the actual model to be built.

### 4.3 Training and Test Data

The claim records provided had past 20 years of raw data, which was then, divided into training and test data. The split ratio between the training and test data was approximately 70% and 30%. For training, claim records from 1999 to 2011 were used and for testing, records from 2012 to 2018 were used.

### 4.4 Network Design

Training algorithm together with activation function needed to be determined before the actual implementation of the networks. Back propagation algorithm based on gradient descent method was chosen. The hidden layer and the output layer had activation function as sigmoid or logistic function, which was used at each neurons in these layers with the formula as:

$$\{\text{Sigmoid Activation}\} = 1 / (1 + e^{-x})$$

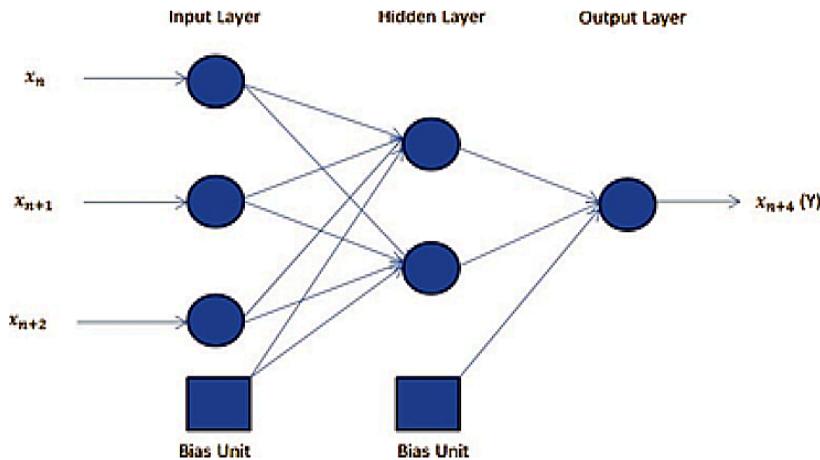
Table 2. Input structure to and from Network

Input Values to Network	Estimated Output Value from Network
$X_1, X_2, X_3$	$X_4$
$X_2, X_3, X_4$	$X_5$
$X_3, X_4, X_5$	$X_6$
.....	.....
$X_n, X_{n+1}, X_{n+2}$	$X_{n+3}$
$X_{n+1}, X_{n+2}, X_{n+3}$	$X_{n+4}$

There were three instance of the initial network model implemented with having 1, 2 and 3 input. There were three instances built with 3 different input simply because there was positive correlation with 1,2 and 3 lag variables. Separate bias units were used which connected to every neuron in the hidden and output layer. Table 2 shows the structure of the input to the model and the desired output for training.

An example of the illustration of inputs from Table 2, consider the first case of training, claims figure for 1999,2000 and 2001 which will be used to predict 2002 yearly claims amount. Figure 4 shows the basic layout and structure of the feed forward neural network implementation. The neurons in the hidden layer were adjusted during the training process.

Figure 4 Basic Structure of Network implemented



#### 4.5 Network Parameters

The parameters such as the learning rate, number neurons in the hidden layer, bias unit value and Epochs were adjusted during the training process. There was no confined method of identifying the parameters that would yield the best result in terms of actual vs predicted. The only way of adjusting these parameters were on a trial and error basis and observing the accuracy of the network. The method that worked well was adjusting these parameters on a trial and error basis, observing which parameters achieved lower error in the training, and testing process.

#### 4.6 Network Accuracy

The accuracy of the models was measured through Mean Absolute Percentage Error (MAPE). Root Mean Squared Error (RMSE) was also used at some instances just too see the accuracy of the network. The following was the formula used for calculating MAPE:

$$MAPE = \frac{100}{n} \sum_{t=1}^n \left( \frac{|Y_n - Y_t|}{|Y_t|} \right) \text{ Where } Y_n \text{ is the actual value and } Y_t \text{ is the predicted value.}$$

**4.7 Stopping Criteria**

Epochs and MAPE was used to stop the training and testing process. The neural network models were tested with a number of different Epochs where MAPE is observed. For each Epoch size, MAPE was observed and eventually reached a point where the lowest MAPE was observed. Root Mean Squared Error on the training data was also used as the stopping criteria and worked well.

**5. IMPLEMENTATION PROCESS**

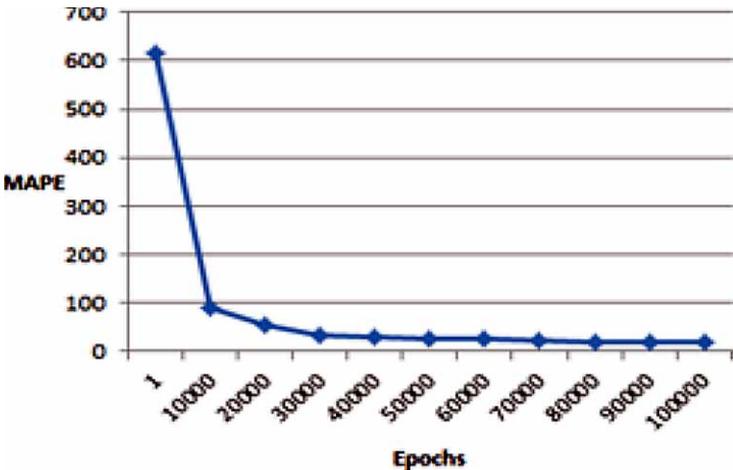
**5.1 Model 1 – Feed Forward Neural Network**

To find out which input size was best, the initial feed forward neural network had one input, 1 hidden and 1 output layer. The network was implemented in C++ to have better picture of how the actual algorithms worked and to observe how the weights were updated while training to reach a lower MAPE. During training, a number of neurons were chosen to observe the impact on MAPE of the network. During the initial runs, the model generated random weights, which were modified while training. Inputs fed to the neural network consisted of 1, 2 and 3 years of medical annual claims paid amount. From the observations, it was concluded that 3 years of data or 3 inputs had much accurate results compared to 1 and 2 inputs. Table 3 shows how the RMS error was reduced when the number of inputs were increased.

Table 3. Summary of RMS Errors with different Network Inputs

Epoc Count	Learning Rate	Hidden Neurons	RMS Error (1 Input)	RMS Error3 (2 Input)	RMS Error (3 Input)
2000	0.15	5	0.08965	0.0704	0.0208
2000	0.15	6	0.0982	0.0707	0.0332
2000	0.15	7	0.09043	0.0711	0.0415
2000	0.15	8	0.09065	0.717	0.0448
2000	0.15	9	0.0909	0.0722	0.0541

Figure 5 MAPE against Different Epochs on Training Set



Now that the input to the network was confirmed, another feed forward neural network model was implemented which consisted of all features i.e. chronic claimed amount, acute claimed amount, average age and annual claimed amount. Including all features as inputs did had positive impact on the MAPE on the training data set but failed on the test data set. The initial model (3-5-1) with three inputs, 5 neurons in the hidden layer and 1 output layer had the training data accuracy of 93% and test accuracy of 87.9%. Learning rate and the bias value were also adjusted during training. The neural network model that came to this accuracy had the learning rate as 0.15 and bias value of 1. The most appropriate Epoch size was found to be 50000. Figures 5 and 6 show how MAPE was reduced during the training process.

Figure 6. Actual vs Predicted on Training Set and on Test Data Set.

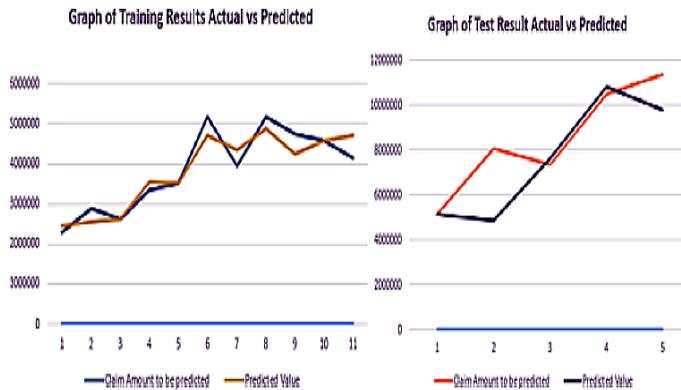
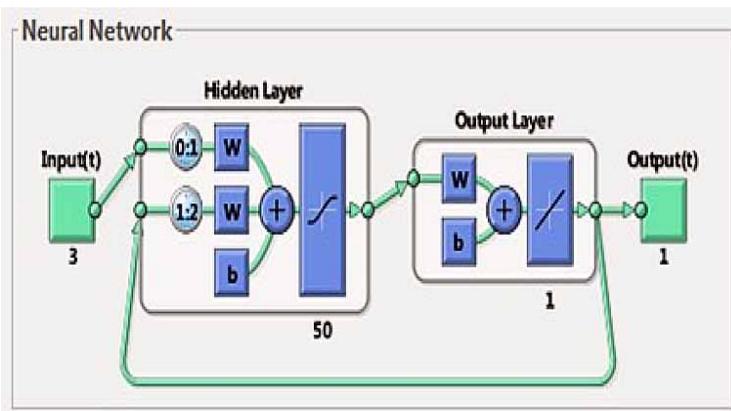


Figure 7 Recurrent Neural Network layout

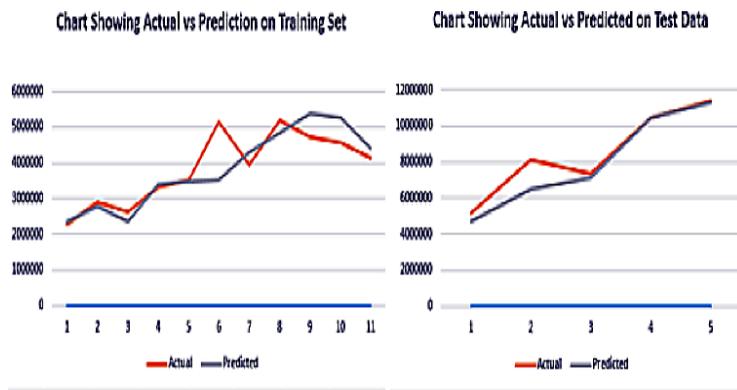


## 5.2 Model 2 – Recurrent Neural Network

After the successful implementation of Feed Forward Neural Network, it was time to test the network using Recurrent Neural Network. NARX (Non-Linear Autoregressive Network) is implemented as part of this research. NARX is a recurrent dynamic network. The number of inputs to the network remained the same as three. The accuracy of the network was MAPE. This network was implemented using MATLAB software.

The parameters such as the learning rate, number of neurons in the hidden layer were adjusted to obtain the least MAPE on both training and test data sets. The network managed to achieve 90.38% as the accuracy on the training set and 93.58% on test set. The number of neurons in the hidden layer was 50 and the learning rate was 2.5, which was used to achieve this. Figures 7 and 8 illustrate the layout of the recurrent network.

Figure 8 Actuals vs Predicted on Training Set and Test data Set



## 6. MODEL COMPARISON

### 6.1 Neural Network Models

After the actual implementation of both models and observing the accuracy and performance, a quick comparison was carried out. In terms of performance of the networks, Recurrent was much faster and required very low computation resources to complete the runs. Feed Forward on the other hand was very resource intensive and took a while to run, sometimes took 15 minutes to complete one run. Recurrent neural network model also outperformed Feed forward neural network in terms of the accuracy. Table 4 shows the summary of accuracy between the two models.

Table 4. Summary of Accuracy Between the two Neural Networks

Neural Network	MAPE on Training Set	Accuracy on Training Set	MAPE on Test Set	Accuracy on Test Set
Feed Forward	7.00	93.00	12.15	87.85
Recurrent	9.62	90.38	6.42	93.58

## 6.2 Comparing with Human Prediction

One of the main aim of the research carried out was to compare and contrast the level of accuracy between human prediction and Artificial Neural Network Prediction. BSP Life (Fiji) Ltd currently has a variance of about 18.10% between the actuals and forecasted value. After implementing the two models and observing the accuracy, there was a reduction of about 11.5% in the overall Mean Absolute Percentage Error (MAPE) when compared to human prediction. The recurrent neural network model had an MAPE of 6.42%. Table 5 shows the comparison of MAPE between BSP Life and the ANN models.

Table 5. Comparison of MAPE between BSP Life and the ANN models

Year	BSP Life Forecast % Error	ANN % Error	Difference %
2015	17.5	7.99	9.51
2016	8.5	19.47	-10.97
2017	4.5	3.9	0.6
2018	22.5	0.5	22
2019	37.5	0.22	37.28
<b>MAPE</b>	<b>18.1</b>	<b>6.42</b>	<b>11.68</b>

## CONCLUSION

This research focused on forecasting yearly claims amount for BSP Life (Fiji) Ltd. Neural network model has been trained using the data from 1999 to 2017. After observing the results from the implementation it can be concluded that forecasting medical claims figure is possible and provided strong results of forecasting using Artificial Neural Network Models. The ANN model outperformed the human prediction that is used now. The ANN model reduced the error rate by about 11.5%. NARX model outperformed the overall accuracy of the Feed Forward ANN together with the computational resources that is required for forecasting. Longer period were required to train the Feed Forward Neural Network when compared to NARX recurrent model. Training the network with larger Epoch size resulted in the network being over trained where the results in the training set was very pleasing but did not perform well on the test data set.

## FUNDING AGENCY

The publisher has waived the Open Access Processing fee for this article.

## REFERENCES

- Abdel Alim, O., Shoukry, A., Elboughdady, N., & Abouelseoud, G. (2013). A Probabilistic Neural Network-Based Module for Recognition of Objects from their 3-D Images. *International Journal of System Dynamics Applications*, 2(2), 66–79. Advance online publication. doi:10.4018/ijdsda.2013040105
- Abhigna, P., Jerritta, S., Srinivasan, R., & Rajendran, V. (2017). Analysis of feed forward and recurrent neural networks in predicting the significant wave height at the moored buoys in Bay of Bengal. *International Conference on Communication and Signal Processing (ICCSPP)*. doi:10.1109/ICCSPP.2017.8286717
- Abhishek, K., Kumar, A., Ranjan, R., & Kumar, S. (2012). Rainfall prediction of a maritime state (Kerala), India using SLFN and ELM techniques. *International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICT)*. doi:10.1109/ICICT1.2017.8342829
- Abroyan, N. (2017). Convolutional and recurrent neural networks for real-time data classification. *Seventh International Conference on Innovative Computing Technology (INTECH)*. doi:10.1109/INTECH.2017.8102422
- Azar, A., & Balas, V. (2012). Statistical Methods and Artificial Neural Networks Techniques in Electromyography. *International Journal of System Dynamics Applications*, 1(1), 39–47. Advance online publication. doi:10.4018/ijdsda.2012010103
- Banerjee, S., Chaudhuri, S. S., & Roy, S. (2017). Fuzzy logic based vision enhancement using sigmoid function. *IEEE Calcutta Conference (CALCON)*. doi:10.1109/CALCON.2017.8280692
- Chapko, A., Emrich, A., Flake, S., Golasowski, F., Gräßle, M., Kohlos, A., Laum, N., Lerche, C., Rust, C., Tacke, J., Werth, D., & Zoth, C. (2011). uRun: A Framework for User-Generated Mobile Services in the Health and Fitness Domain. *International Journal of Service Science, Management, Engineering, and Technology*, 2(4), 79–97. Advance online publication. doi:10.4018/ijssmet.201100108
- Chuanchai, P., Champrasert, P., & Rattanadoung, K. (2019). Particulate Matter Concentration Spatial Prediction using Interpolation Techniques with Machine Learning. *7th IEEE International Conference on Information and Communication Technology (ICoICT)*. doi:10.1109/ICoICT.2019.8835214
- Decaro, C., Montanari, G., Molinariz, R., Gilberti, A., & Bagno, A. (2019). Machine learning approach for prediction of Hematic parameters in Hemodialysis patients. *IEEE Journal of Translational Engineering in Health and Medicine*, 7, 1–8. Advance online publication. doi:10.1109/JTEHM.2019.2938951 PMID:32309060
- Din, G. M. U., & Marnerides, A. K. (2017). Short Term power load forecasting using Deep Neural Networks. *International Conference on Computing, Networking and Communications (ICNC)*. doi:10.1109/ICNC.2017.7876196
- Dong, J., Rafayelyan, M., Krzakala, F., & Gigan, S. (2019). Optical Reservoir Computing Using Multiple Light Scattering for Chaotic Systems Prediction. *IEEE Journal of Selected Topics in Quantum Electronics*, 26(1), 1–12. Advance online publication. doi:10.1109/JSTQE.2019.2936281
- Dong, Y., Wen, R., Li, Z., Zhang, K., & Zhang, L. (2019). A New RNN Based Approach to Diabetic Blood Glucose Prediction. *7th IEEE International Conference on Bioinformatics and Computational Biology (ICBCB)*. doi:10.1109/ICBCB.2019.8854670
- Dong, Y., Wen, R., Zhang, K., & Zhang, L. (2019). A Novel RNN-Based Blood Glucose Prediction Approach Using Population and Individual Characteristics. *7th IEEE International Conference on Bioinformatics and Computational Biology (ICBCB)*. doi:10.1109/ICBCB.2019.8854657
- Dumas, T., Roumy, A., & Guillemot, C. (2019). Context-Adaptive Neural Network-Based Prediction for Image Compression. *IEEE Transactions on Image Processing*, 29, 679–693. doi:10.1109/TIP.2019.2934565 PMID:31425077
- El-said, S. (2013). Reliable Face Recognition Using Artificial Neural Network. *International Journal of System Dynamics Applications*, 2(2), 14–42. Advance online publication. doi:10.4018/ijdsda.2013040102
- Hajiaghayi, M., & Vahedi, E. (2019). *Code Failure Prediction and Pattern Extraction Using LSTM Networks*. doi:10.1109/BigDataService.2019.00014

Ibiwoye, A., Ajibola, O. O., & Sogunro, A. B. (2012). *Artificial Neural Network Model for Predicting Insurance Insolvency*. Academic Press.

Jun, E., Mulyadi, A., & Suk, H. (2019). Stochastic Imputation and Uncertainty-Aware Attention to EHR for Mortality Prediction. *IEEE International Joint Conference on Neural Networks (IJCNN)*. doi:10.1109/IJCNN.2019.8852132

Khan, M., Khattak, A., Zafari, F., & Mahmud, S. A. (2013). Electrical load forecasting using fast learning recurrent neural networks. *International Joint Conference on Neural Networks (IJCNN)*. doi:10.1109/IJCNN.2013.6706998

Khatri, S. K., Singhal, H., & Johri, P. (2014). Sentiment analysis to predict Bombay stock exchange using artificial neural network. *3rd International Conference on Reliability, Infocom Technologies and Optimization*. doi:10.1109/ICRITO.2014.7014714

Kitchens, F. L. (2009). Financial implications of artificial Neural Networks in automobile insurance underwriting. *International Journal of Electronic Finance*, 3(3), 311–319. doi:10.1504/IJEF.2009.027853

Kurtah, P., Takun, Y., & Nagowah, L. (2019). *7th IEEE International Conference on Information and Communication Technology (ICoICT)*. doi:10.1109/ICoICT.2019.8835381

Liou, F.-M., Tang, Y.-C., & Chen, J.-Y. (2008). Detecting hospital fraud and claim abuse through diabetic outpatient services. *Health Care Management Science*, 11(4), 353–358. doi:10.1007/s10729-008-9054-y PMID:18998594

Liu, H., Zhang, Y., Zhang, H., Fan, C., Kwong, S., Kuo, C., & Fan, X. (2019). Deep Learning-Based Picture-Wise Just Noticeable Distortion Prediction Model for Image Compression. *IEEE Transactions on Image Processing*, 29, 641–656. doi:10.1109/TIP.2019.2933743 PMID:31425033

Mahmoud, M., & Sunni, F. (2012). Stability of Discrete Recurrent Neural Networks with Interval Delays: Global Results. *International Journal of System Dynamics Applications*, 1(2), 1–14. Advance online publication. doi:10.4018/ijdsda.2012040101

Majhi, S. (2018). An Efficient Feed Forward Network Model with Sine Cosine Algorithm for Breast Cancer. *International Journal of System Dynamics Applications*, 7(2), 1–14. Advance online publication. doi:10.4018/IJSDA.2018040101

Mnasser, A., Bouani, F., & Ksouri, M. (2014). Neural Networks Predictive Controller Using an Adaptive Control Rate. *International Journal of System Dynamics Applications*, 3(3), 127–147. Advance online publication. doi:10.4018/ijdsda.2014070106

Motlagh, T., & Khaloozadeh, H. (2016). A new architecture for modelling and prediction of dynamic systems using neural networks: Application in Tehran stock exchange. *4th International Conference on Control, Instrumentation, and Automation (ICCIA)*. doi:10.1109/ICCIAutom.2016.7483160

Pei, S., Tong, L., Li, X., Jiang, J., & Huang, J. (2017). Feed-forward network for cancer detection. *13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*. doi:10.1109/FSKD.2017.8393356

Petrescu, L., Cazacu, E., & Petrescu, C. (2015). Sigmoid functions used in hysteresis phenomenon modelling. *9th International Symposium on Advanced Topics in Electrical Engineering (ATEE)*. doi:10.1109/ATEE.2015.7133863

Pun, T. B., & Shahi, T. B. (2018). Nepal Stock Exchange Prediction Using Support Vector Regression and Neural Networks. *2018 Second International Conference on Advances in Electronics, Computers and Communications (ICAEECC)*. doi:10.1109/ICAEECC.2018.8479456

Rizal, A., & Hartati, S. (2016). Recurrent neural network with Extended Kalman Filter for prediction of the number of tourist arrival in Lombok. *International Conference on Informatics and Computing (ICIC)*. doi:10.1109/IAC.2016.7905712

Safarinejadian, B., Tajeddini, A., & Ramezani, A. (2013). Predict time series using extended, unscented, and cubature Kalman filters based on feed-forward neural network algorithm. *3rd International Conference on Control, Instrumentation, and Automation*. doi:10.1109/ICCIAutom.2013.6912827

- Salama, M., & Hassaniien, A. (2014). Fuzzification of Euclidean Space Approach in Machine Learning Techniques. *International Journal of Service Science, Management, Engineering, and Technology*, 5(4), 29–43. Advance online publication. doi:10.4018/ijssmet.2014100103
- Samarawickrama, A. J. P., & Fernando, T. G. I. (2017). A recurrent neural network approach in predicting daily stock prices an application to the Sri Lankan stock market. *IEEE International Conference on Industrial and Information Systems (ICIIS)*. doi:10.1109/ICIINFS.2017.8300345
- Sangwan, N., & Bhatnagar, V. (2020). Comprehensive Contemplation of Probabilistic Aspects in Intelligent Analytics. *International Journal of Service Science, Management, Engineering, and Technology*, 11(1), 116–141. Advance online publication. doi:10.4018/IJSSMET.2020010108
- Seo, J., Lee, J., & Kim, K. (2018). Decoding of Polar Code by Using Deep Feed-Forward Neural Networks. *International Conference on Computing, Networking and Communications (ICNC)*. doi:10.1109/ICCNC.2018.8390279
- Solovyeva, A. B. (2017). Types of recurrent neural networks for non-linear dynamic system modelling. *XX IEEE International Conference on Soft Computing and Measurements (SCM)*. doi:10.1109/SCM.2017.7970552
- Tang, T. A., Mhamdi, L., McLernon, D., Zaidi, S. A. R., & Ghogho, M. (2018). Deep Recurrent Neural Network for Intrusion Detection in SDN-based Networks. *4th IEEE Conference on Network Softwarization and Workshops (NetSoft)*. doi:10.1109/NETSOFT.2018.8460090
- Thomas, P., Suhner, M., & Derigent, W. (2019). Relearning procedure to adapt pollutant prediction neural model: Choice of relearning algorithm. *IEEE International Joint Conference on Neural Networks (IJCNN)*. doi:10.1109/IJCNN.2019.8852193
- Venkadesan, A., Bhavana, G., Haneesha, D., & Sedhuraman, K. (2017). Comparison of feed forward and cascade neural network for harmonic current estimation in power electronic converter. *International Conference on Innovative Research in Electrical Sciences (IICIRES)*. doi:10.1109/IICIRES.2017.8078295
- Viaene, S., Dedene, G., & Derrig, R. A. (2005). Auto claim fraud detection using Bayesian learning neural networks. *Expert Systems with Applications*, 29(3), 653–666. doi:10.1016/j.eswa.2005.04.030
- Wang, S., Ning, C., & Cui, W. (2015). Time series prediction of bank cash flow based on grey neural network algorithm. *International Conference on Estimation, Detection and Information Fusion (ICEDIF)*. doi:10.1109/ICEDIF.2015.7280205
- Wu, B., Li, K., Yang, M., & Lee, C. H. (2016). A study on target feature activation and normalization and their impacts on the performance of DNN based speech dereverberation systems. *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*. doi:10.1109/APSIPA.2016.7820875
- Xiao, D., Li, X., Lin, X., & Shi, C. (2015). A time Series Prediction method based on self-adaptive RBF neural network. *4th International Conference on Computer Science and Network Technology (ICCSNT)*. doi:10.1109/ICCSNT.2015.7490837
- Xu, W., Wang, S., Zhang, D., & Yang, B. (2011, April). Random rough subspace based neural network ensemble for insurance fraud detection. In *2011 Fourth International Joint Conference on Computational Sciences and Optimization* (pp. 1276-1280). IEEE. doi:10.1109/CSO.2011.213
- Yunos, Z. M., Ali, A., Shamsyuddin, S. M., Ismail, N., & Sallehuddin, R. S. (2016). Predictive Modelling for Motor Insurance Claims Using Artificial Neural Networks. *International Journal of Advances in Soft Computing and its Applications*, 8(3), 160-172.
- Zhang, W., Li, B., Zhao, D., Gong, F., & Lu, Q. (2016). Workload Prediction for Cloud Cluster Using a Recurrent Neural Network. *International Conference on Identification, Information and Knowledge in the Internet of Things (IIKI)*. doi:10.1109/IIKI.2016.39
- Zhang, X., Dong, Y., Wen, L., Lu, F., & Li, W. (2019). Remaining Useful Life Estimation Based on a New Convolutional and Recurrent Neural Network. *15th IEEE International Conference on Automation Science and Engineering (CASE)*. doi:10.1109/COASE.2019.8843078

*Sam Goundar is an Editor-in-Chief of the International Journal of Blockchains and Cryptocurrencies (IJFC) – Inderscience Publishers, Editor-in-Chief of the International Journal of Fog Computing (IJFC) – IGI Publishers, Section Editor of the Journal of Education and Information Technologies (EAIT) – Springer and Editor-in-Chief (Emeritus) of the International Journal of Cloud Applications and Computing (IJCAC) – IGI Publishers. He is also on the Editorial Review Board of more than 20 high impact factor journals. As a researcher, apart from Blockchains, Cryptocurrencies, Fog Computing, Mobile Cloud Computing and Cloud Computing, Dr. Sam Goundar also researches in Educational Technology, MOOCs, Artificial Intelligence, ICT in Climate Change, ICT Devices in the Classroom, Using Mobile Devices in Education, e-Government, and Disaster Management. He has published on all these topics. He was a Research Fellow with the United Nations University. He is a Senior Lecturer in IS at The University of the South Pacific, Adjunct Lecturer in IS at Victoria University of Wellington and an Affiliate Professor of Information Technology at Pontificia Universidad Catolica Del Peru.*

*Akashdeep Bhardwaj achieved his PhD from University of Petroleum & Energy Studies (UPES), Post Graduate Diploma in Management (PGDM), Engineering graduate in Computer Science. He has worked as Head of Cyber Security Operations and currently is a professor in a leading university in India. He has over 24 years of experience working as an Enterprise Risk and Resilience and Information Security and Technology professional for various global multinationals.*