ADVANCING INDUSTRIAL ENGINEERING IN NIGERIA THROUGH TEACHING, RESEARCH AND INNOVATION

A BOOK OF READING

Edited By
Ayodeji E. Oluleye
Victor O. Oladokun
Olusegun G. Akanbi

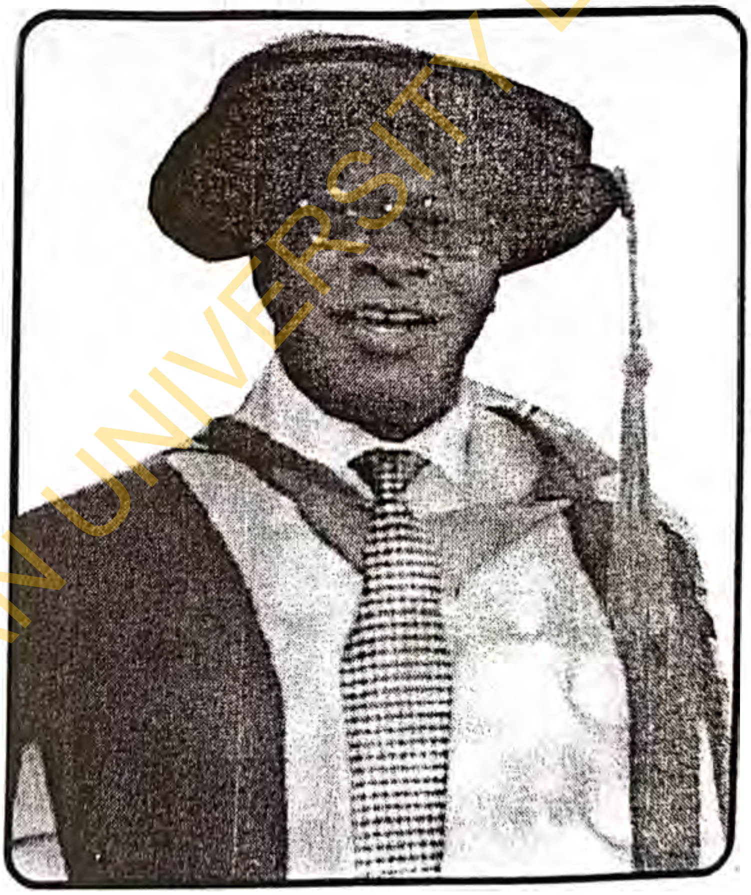
ADVANCING
INDUSTRIAL ENGINEERING IN NIGERIA
THROUGH
TEACHING, RESEARCH AND INNOVATION

Edited By

Ayodeji E. Oluleye
Victor O. Oladokun
Olusegun G. Akanbi

ADVANCING
INDUSTRIAL ENGINEERING IN NIGERIA
THROUGH
TEACHING, RESEARCH AND INNOVATION

(A Festchrift in honour of Professor O. E Charles-Owaba)



Professor O. E. Charles-Owaba

Advancing Industrial Engineering in Nigeria through Teaching, Research and Innovation.

ISBN : 978-078-515-9

Published by

Department of Industrial and Production
Engineering
University of Ibadan.

Printed by:
Leading Edge Printers & Publisher
Ibadan

FOREWORD

It gives me great pleasure writing the foreword to this book. The book was written in recognition of the immense contributions of one of Nigeria's foremost industrial engineers, respected teacher, mentor, and lover of youth – Professor Oliver Charles-Owaba.

His commitment to the teaching and learning process, passionate pursuit of research and demonstration of excellence has prompted his colleagues and mentees to write this book titled – Advancing Industrial Engineering in Nigeria through Teaching, Research and Innovation (A Festschrift in honour of Professor O. E Charles-Owaba) as a mark of honour, respect and recognition for his personality and achievements.

Professor Charles-Owaba has written scores of articles and books while also consulting for a medley of organisations. He has served as external examiner to various programmes in the tertiary educational system. The topics presented in the book cover the areas of Production/Manufacturing Engineering, Ergonomics/Human Factors Engineering, Systems Engineering, Engineering Management, Operations Research and Policy. They present the review of the literature, extension of theories and real-life applications. These should find good use in the drive for national development.

Based on the above, and the collection of expertise in the various fields, the book is a fitting contribution to the corpus of knowledge in industrial engineering. It is indeed a befitting gift in honour of erudite Professor Charles-Owaba.

I strongly recommend this book to everyone who is interested in how work systems can be made more productive and profitable. It represents a resourceful compilation to honour a man who has spent the last forty years building up several generations of industrial engineers who are part of the process to put Nigeria in the rightful seat in the comity of nations. Congratulations to Professor Charles-Owaba, his colleagues and mentees for this festschrift.

Professor Godwin Ovuworie
Department of Production Engineering
University of Benin

TABLE OF CONTENTS

CHAPTER 3

Development Of An Artificial Neural Network-Fuzzy-Markov Model For Industrial Accidents Forecasting

*[1] I. E. Edem and [2]O. A. Adebimpe

[1,2]Department of Industrial and Production Engineering, University of Ibadan, Ibadan, Nigeria.
*Corresponding Author. E-Mail: [1]ie.edem@mail1.edu.ng, [2]oa.adebimpe@mail1.edu.ng

Abstract

Industrial accidents possess the potential of causing physical, psychological and even fatal consequences when they occur. In that regard, Industrial accidents forecasting aid stakeholders in properly managing and improving workplace safety by anticipating accident occurrences to prevent or minimise their consequences. Due to the high variation, random and fluctuating characteristics of industrial accident occurrences, machine learning and Markov based models methods have become increasingly popular as a tool for understanding their occurrence patterns and detecting their vibrational directions. However, little investigation has been made towards combining the positive characteristics of these methods for industrial accidents forecasting. This study is concerned with the development of a neuro-fuzzy-Markov model for the prediction and forecasting of industrial accident occurrences.

The methodology employed essentially involves the implementation of the Artificial Neural Network (ANN) model through the development of structured control methods to enhance improved forecast candidate generation potential. Further, an analysis of the model's residual was undertaken to obtain an ANN forecast correction factor by using a

combination of fuzzy and Markov techniques. Based on this, investigations were then carried out to determine the direction of vibration of the ANN predictor model. Subsequently, results were generated from the prediction mechanism. The model was validated by comparing its one window ahead (OWA) forecast potential with those of the ANN model using secondary industrial accident data based on the mean absolute percentage error (MAPE) and the Root Mean Square Error (RMSE). Also, an evaluation was done by comparing the forecast performances of the model with those of two traditional (Autoregressive moving average [ARIMA] and exponential smoothing [EXPSM]) models and two non-traditional (Mao and Sun grey-Markov (MSGM) and Grey-Fuzzy-Markov Pattern Recognition GFMAPR) models.

The forecast performance results obtained on the model's application showed that it possessed the capability to correct and improve ANN forecasts. The MAPE and RMSE results obtained for the ANN-fuzzy-Markov model were 15.39 and 26.39, while those produced by the ANN were 20.13 and 30.57 respectively. The model produced more superior forecast when compared with ARIMA and EXPSM, and compared well with the MSGM and GFMAPR.

The obtained results indicate that the model possesses the capability of carrying the predictions of industrial fire accident to an acceptable degree of accuracy.

1.0 Introduction

Accident occurrences are an important component of industrial activities. Every occupational activity is inherent with potential accidents which if manifested can lead to accidents. Industrial accidents are undesired occurrences that can occur in forms such as burns, cuts, lacerations, radiation and amputations and even fatalities. Industrial accident occurrences can negatively impact the physical and psycho-social health conditions for workers and their families on the one hand, and potentially affect the fortunes and continuity of businesses on the other. As a result, understanding the dynamics of their occurrences with a view to anticipating and effectively managing them becomes imperative.

Forecasting accidents ahead of their occurrences is a popular practice adopted by organisations as part of their management processes (Xiaoping & Liu, 2009; Mock, Nugent, Kobusingye, & Smith, 2017). Accident forecasting is a guess at the safety state of a system (Xiaoping & Liu, 2009). In most cases, this guess is made by the mathematical or statistical characterisation of a prediction mechanism based on available information and observations (Shmueli, 2011). Generally, forecasting involves the deployment of single or multiple input dependent models (Bontempi, 2008; Manaloto & Balahadia, 2017; Makridakis, Spiliotis, & Assimakopoulos, 2018; Mapuwei, Bodhlyera, & Mwambi, 2020) to produce outputs based on the mathematical or algorithmic rules governing the deployed model.

Single input forecasting models are used particularly in situations where the accident occurrence causal variables are sparse, unavailable, or difficult to extract as is the case with accident occurrences (Aidoo & Eshun, 2012; Ratnayaka, 2017), or when a simple model in terms of input data is desired (Aksoy & Dahamsheh, 2009). In this regard, traditional time series (TS) models such as Auto-Regressive and Moving Average (ARIMA), exponential smoothing and regression models find relevance and are still prevalently deployed (Ghedira, Kammoun, & Saad, 2018; Svs, 2018; Urrutia *et al.*, 2018; Malysa, 2020).

The non-traditional forms of TS models such as Artificial Neural Networks (ANN), Markov models and Fuzzy Time Series Models (FTSM) based models have become more popular in the last decade as a result of their ability to handle imprecise, ambiguous and incomplete data (Zhang, Qu, Wang, & Zhao, 2020). ANNs are computing systems that learn without formal statistical training, and possess the ability to detect complex nonlinear relationships between dependent and independent variables (Tu, 1996; Xiaoping & Liu, 2009), Fuzzy logic is used for the detection, description and representation of uncertainties (Pabuçcu, 2017) while the Markov methods possess the characteristic of detecting fluctuations and the vibrational direction of TS occurrences (Wang, Mehrabi, & Kannatey-Asibu, 2002). In their pure forms, these models have been found to perform satisfactorily from their successful application in various industrial areas including weather forecasting (Voyant, Muselli, Paoli, & Nivet, 2011), health emergency preparedness (Mapuwei *et al.*, 2020), crude oil production

(Boaisha & Amaitik, 2010), stock price forecasting (Javedani Sadaei & Lee, 2014) and tool wear monitoring rolling bearings (Liu, Youmin, wu, Wang, & Xie, 2017).

Furthermore, the hybrid forms of these models have also been frequently applied. For example, Ali, Yohanna, Ijasini, and Garkida (2018) applied a Neuro-Fuzzy (NF) model in electricity load forecasting and the results of the study indicated a forecast accuracy greater than 98 %. Ganesan, Annamalai, and Deivanayagampillai (2019) used the fuzzy cognitive maps Markov chain model in the prediction of stock market behaviour. Das, Naik, and Behera (2020) developed a NF and fuzzy reduction model and successfully applied it to solving problems of data classification in data mining. In addition, a fuzzy neural network-Markov model was developed by (Shi, Hu, Yu, & Hu, 2016). The results obtained from the studies mentioned show that the hybrid forms of the non-traditional, non-linear models can be implemented in TS forecasting to produce outputs of high accuracies.

Some studies have been carried out concerning the use of single and hybrid forms of the ANN, Markov and FTSM models for industrial accident prediction and prevention. Lan and Zhou (2014) applied a variant of the grey-Markov model in carrying prediction and out a single step ahead forecast of annual coal mining deaths in China. A forecast of the number of occupational accidents, death and permanent incapacitation was made by (Ceylan, 2014) using models developed by ANN models. Results obtained showed that the aim for which the models were developed was achieved. Also, Edem, Oke, and Adebiyi (2018) developed a high variation tolerant grey-fuzzy-Markov model for a one-window ahead (OWA) forecast of industrial accidents. Sarkar, Vinay, Raj, Maiti, and Mitra (2019) observe that although some efforts have been made concerning carrying out studies in this area, the literature is sparse with the domain still slowly developing. As such, more research is required in this area for improved understanding.

Two major problems of utilizing the ANN for TS analysis is that of the choice of ANN architecture (R. Adhikari & Agrawal, 2013; Sánchez-Sánchez, García-González, & Coronell, 2020) and overtraining minimization (Faraway & Chatfield, 1998). Both of these problems have reduced the popularity of ANN for TS prediction. Two common attempts at overcoming this problem involve the use of model comparison criteria and cross-validation experiments (Ratnadip Adhikari, 2015). The use of the Markov technique as a tool for detecting TS vibrational directions and correcting model prediction anomalies is popular and has been successfully applied in various industrial areas (Yarmohammadi & Safaei, 2012; Liu *et al.*, 2017; Wilinski, 2019). However, little effort has been made at applying this technique in ANN TS prediction to correct and redirect forecasts that have been produced from inadequate architectures.

This study is aimed at the development of an ANN-Fuzzy-Markov model for OWA forecasting of industrial accidents. The objectives include the building of the model and its subsequent evaluation with other existing models currently utilised for the same purpose. The study seeks to demonstrate the effectiveness of the Markov technique in producing accurate ANN TS forecast without consideration given to the processes of optimal network architecture development. It is hoped that the model developed will serve as an approach for accident forecasting as well as present stakeholders with an effective accident management tool.

The rest of the work is presented as follows: Section 2 contains a brief presentation of the theoretical concepts of the model, a detailed description of the methodological steps undertaken in the development of the model is presented in section 3. Section 4 contains information regarding the application of the developed model and the discussion of the results obtained from its analysis and evaluation. The conclusions made from the study are presented in section 5.

2.0 Theoretical Background for the Model Development

The ANN-fuzzy-Markov industrial accidents (AFMIA) forecasting model is essentially founded on four concepts namely, ANN prediction procedure, fuzzy classification scheme, Markov transitions and the fuzzy-Markov probabilities determination. The ANN prediction procedure describes the method of setting up the ANN structure for industrial accidents TS prediction. The fuzzy classification scheme is concerned with capturing the imprecision detected from the prediction anomalies of the ANN while the Markov transitions are concerned with the re-direction of the prediction anomalies. The concept of the fuzzy-Markov probabilities determination is usually required in analysing situations where systems characterized by Markov properties also exhibit imprecise properties. These four concepts are subsequently discussed in this section.

2.1 General concept of Artificial Neural Network

ANN forecasting is a deep learning method employed for TS analysis. Unlike linear TS models, ANNs readily learn different types of relationships between variables in the absence of do not require complex mapping function assumption functions (Brownlee, 2018). Although different variants of ANN models such as the feed-forward network (FNN) or Multi-Layer Perceptron (MLP) network, as well as the back-propagation (recurrent Neural networks [RNN]) such as Convolutional Neural Networks (CNN) and the Long-Short Term Memory (LSTM) network, exist, the general architecture for all ANN models is essentially the same.

A typical ANN model is primarily made up of layers of nodes or neurons. The simplest ANN model must possess at least three neuron layers namely, not more than one input layer (responsible for receiving and processing input information), at least one hidden layer (concerned with the further processing of information received from the input layer) and not more than not output layer (which conveys the results of the information processed by the previously mentioned layers) [see Figure 1]. The key to the workings of the ANN is that within a layer, information processed, weighted and distributed to the next layer is made possible by the firing of biological system emulated signals arising from mathematically derived activation functions (MacLeod, 2013). Thus at the end of the process, the output layer makes a decision or produces an inference based on the values computed in the hidden layers.

The major difference between the FNN and the RNN variants is that in the FNN, information once weighted, cannot be modified as information flow is unidirectional while for the RNN, there is bi-directional information flow as the results received by the output layer can be communicated back to the hidden layer for weight readjustment towards improved learning.

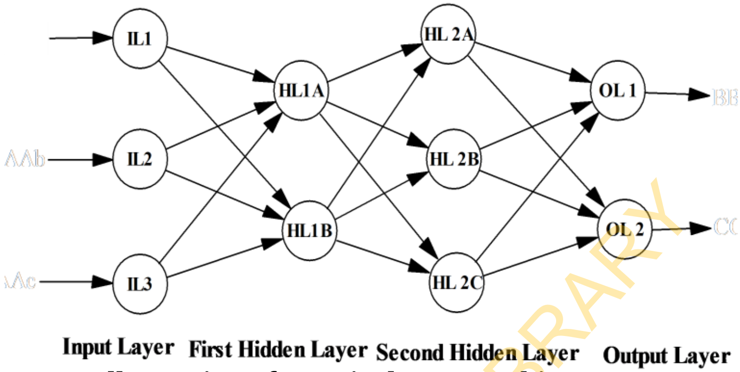


Figure 1: Illustration of a typical ANN architecture

2.2 Neural network model structure for forecasting The procedure for the deployment of ANN for forecasting involves firstly the preparation of a set of historical data X $\{X\colon x_i(i\colon 1,2,3,\dots,x_E)\}$ into the input vector form M and output vector form Y (Equations 1 and 2). M is a trajectory matrix made up of L lagged row vectors $m_i(i\colon 1,2,3,\dots,L)$ each of width N made up of a sequence of occurrences following each other. N $(1 \leq N \leq x_{E-1})$ is referred to as the forecast window length. Y is a column vector made up L rows for which each row value $y_i(i\colon 1,2,3,\dots,L)$ is mapped to m_i for learning and prediction.

It is worth noting that x_E does not constitute the entire historical data, but the portion of data set aside for building the model. M and Y are subsequently fed into the built network and a supervised learning process involving setting the process to the desired number of iterations (epochs) and learning rates (Kriesel, 2010) is executed to obtain forecast. Due to the time and computational complexities involved in the manual development and implementation of the process, the use of soft computational solutions such as Keras TensorFlow (in R and python programming) have become quite popular (Brownlee, 2018; Lewis, 2018)

(1) $$M = \left(X_{ij}\right)_{i=1,j=1}^{L,N} = \begin{array}{c} \begin{pmatrix} x_1 & x_2 & x_3 \cdots & x_{N-1} & x_N \\ x_2 & x_3 & x_4 \cdots & x_N & x_{N+1} \\ x_3 & x_4 & x_5 \cdots & x_{N+1} & x_{N+2} \\ . & . & . & . & . \\ . & . & . & . & . \end{pmatrix} \end{array} \begin{array}{c} m_1 \\ m_2 \\ . \\ . \\ . \\ m_{-1} \\ L \end{array}$$

(2) $$Y = \begin{array}{c} \begin{array}{cccccc} y_1 & y_2 & y_3 \cdots & y_{L-1} & y_L \end{array} \\ \left(x_{N+1} \; x_{N+2} \; x_{N+3} \cdots x_{E-1} \; x_E\right)^T \end{array}$$

2.3 Fuzzy set classification

A fuzzy set $\mathcal{F}$ is defined by a set (Equation 3). $\mu_{\mathcal{M}}(x)$ is called a membership function of the set and represents a discrete or continuous function that specifies the extent to which the element s, a member of crisp set M, belongs to $\mathcal{M}$. s is associated with a fuzzy value that exists between 0 and 1 describing the extent in which it belongs to or satisfies some property attributed to M. Fuzzy values can be described using triangular, quadratic or exponential numbers (Bojadziev & Bojadziev, 2007).

$$\mathcal{F} = \{(s, \mu_{\mathcal{M}}(s) | s \in \mathcal{F}, \mu_{\mathcal{F}}(s) \in [0,1])\} \qquad (3)$$

2.4 Markov chain probabilities

Let S be a discrete sample space $S = \{s_i : i = 1,2,3, \ldots, N\}$. A Markov chain can be defined as a random variable sequence X_t $(t: 1,2,3, \ldots, T)$ that takes values in X such that the probability of an event $P_{s_i \acute{s}_i}$ occurring is dependent on the probability the event $\acute{s}_i$ occurs at time $t+1$ conditioned on the probability that event s_i occurred at time t (Dobrow, 2016). With consideration given to a Markov chain of finite-state (Equation 4), the probability of event $\acute{s}_i$ occurring for n number of step transitions is described in equations 5 and 6.

$$
(4) \qquad P_{s_i\acute{s}_i} = \begin{array}{r} States(s_i) \\ 1 \\ 2 \\ 3 \\ . \\ . \\ . \\ N-1 \\ N \end{array}
\begin{array}{c} \begin{array}{cccccc} 1 & 2 & 3 & \ldots & N-1 & N \end{array} \\
\begin{pmatrix}
P_{11} & P_{12} & P_{13} & \ldots & P_{1N-1} & P_{1N} \\
P_{21} & P_{22} & P_{23} & \ldots & P_{2N-1} & P_{2N} \\
P_{31} & P_{32} & P_{33} & \ldots & P_{3N-1} & P_{3N} \\
. & . & . & \ldots & . & . \\
. & . & . & \ldots & . & . \\
. & . & . & \ldots & . & . \\
P_{n-11} & P_{n-12} & P_{n-13} & \ldots & P_{n-1N-1} & P_{n-1N} \\
P_{n-11} & P_{n-12} & P_{n-13} & \ldots & P_{n-1N-1} & P_{n-1N}
\end{pmatrix} \end{array}
$$

$$P^t_{s_i\acute{s}_i} = P(X_{t+1}) = \acute{s}_i | P(X_t) = s_i \quad \{s_i, \acute{s}_i \in S, t \geq 1\} \qquad (5)$$

2.5 Fuzzy Markov probabilities

Fuzzy Markov probabilities (FMB) are the probabilities of events that exhibit Markov properties. However, the precision of the state properties cannot be accurately estimated. FMB of events is determined from the combination of the concepts earlier presented in sections 2.3 and 2.4. Here, the FMB determination procedure of (Pardo & de la Fuente, 2010) is presented.

Suppose a set of fuzzy states $\mathcal{F}_k\{k = 1,2,3, \ldots, z\}$ is defined such that each $\mathcal{F}_i$ represents an event in the initial Markov chain (Equation 4), then the probability of a single step transition from a fuzzy initial state $\mathcal{F}_v$ to a fuzzy final state $\mathcal{F}_w$ is expressed in equation 6.

$$\tilde{P} = P(\mathcal{F}_w|\mathcal{F}_v) = P\left(\tilde{X}_1 = w \middle| \tilde{X}_0 = v\right) = \sum_{i=1}^{N} \bar{P}\left(\frac{P_{s_i}\mu_{\mathcal{F}_v(s_i)}}{P(\mathcal{F}_v)}\right) \qquad (6)$$

Where, $\mu_{\mathcal{F}_v(i)}$: Fuzzy membership value for initial Markov event i, P_{s_i}: Initial state probabilities; $P(\mathcal{F}_v)$: Fuzzy initial state probabilities; $\bar{P}$: The single-step transition probability to the final fuzzy state $\mathcal{F}_w$ given the initial state event.

$P(\mathcal{F}_v)$ and $\bar{P}$ are further described in equations 7 and 8.

$$P(\mathcal{F}_v) = \sum_{i=1}^{N} P\{X_0 = \acute{s}_i\}\mu_{\mathcal{F}_v}(\acute{s}_i) \qquad (7)$$

$$\bar{P} = P\left(\tilde{X}_1 = w \middle| X_0 = s_i\right) = \sum_{i=1}^{N} P_{s_i\acute{s}_i}\mu_{\mathcal{F}_w}(\acute{s}_i) \qquad (8)$$

Equation 6 can be conveniently solved to obtain $\tilde{P}$ by adopting the following stepwise procedure (Pardo & de la Fuente, 2010)

Step 1: Create the initial Markov chain (crisp set) transition matrix $P = P_{s_i \acute{s}_i}$ of the events (Equation 4)
Step 2: Create the matrix of the initial Markov chain states versus their corresponding transition fuzzy state components Q as well as that of the initial Markov chain states versus their corresponding initial fuzzy state components H
Step 3: Obtain $\bar{P} = PQ$ via matrix multiplication
Step 4: From H, determine $P_{s_i}\mu_{\mathcal{F}_v(s_i)}$ for $\forall i$, $P(\mathcal{F}_v)$ for $\forall k$, then obtain $(H^*)^T$ containing the values of $\frac{P_{s_i}\mu_{\mathcal{F}_v(s_i)}}{P(\mathcal{F}_v)}$.
Step 5:The fuzzy transition probability matrix $\tilde{P}$ for the event is finally obtained using equation 9.

$$\tilde{P} = (H^*)^T \bar{P} \tag{9}$$

3.0 Methodology

In this section, the AFMIA forecasting model is presented. Section 3.1 gives an overview model development process. The notations and definitions employed are presented in Section 3.2. Section 3.3 contains a detailed description of the methodological steps deployed in developing AFMIA.

3.1 Overview of the model development process

The development of the AFMIA involves the collection and preparation of industrial accidents, then deploying the prepared data in the development of the ANN prediction model. The model is used to generate various predictions and forecast which are subsequently screened based on certain performance criteria. Predictions that satisfy the screening conditions are deployed in carrying out a direction of vibration analysis to correct the ANN forecast producing a modified value in the process. If the desired ANN forecast population level and corresponding corrections are met, then a procedure for determining a single value forecast is implemented if not met, then the model proceeds to regenerate more forecast candidates. Figure 2 presents a summary of the key activities involved in the AFMIA development process while Table 1 lists the symbols and notations employed in the work.

3.2 Description of the AFMIA development

In this section, each phase of the forecast model development is described in detail in the following subsections.

3.2.1 Development and implementation of ANN prediction model

Given a set of industrial accident historical data D ($D = \{D_t : t = 1,2,3, \ldots, K\}$), the OWA forecast process aims to forecast D_{K+1}. The first step taken towards achieving this step was to convert D into the input and output vector forms described in equations 1 and 2 with the input vector window width N set. To guide the decision on the choice of N, a partial auto-correlation analysis was carried out on D to determine the extent of the lagged relationship that exists between the data points N steps apart. Subsequently, the CNN neural network model was deployed. The CNN was built using Keras python 3.7 model with TensorFlow backend. Table 2 shows the parameters that were defined in building CNN and their corresponding values. The model was subsequently implemented by training D at different epoch values ε to obtain the vector of predictions D^{ε} $\{D_t^{\varepsilon} : t = N+1, N+2, N+3, \ldots, K-1, K\}$ and corresponding forecast D_{K+1}^{ε}.

Table 1: Symbols and notations

C_K: ANN Forecast residual	CNN: Convolutional Neural Networks
D_t^{ε}: ANN prediction made at time t	N: Swing window width
C_t: Cumulative residual at time t	CE: cumulative error
$\tilde{F}_c$: Fuzzy correction span vector	TS: Time series
μ: Fuzzy membership	OWA: One window ahead
D: Vector of Industrial accident historical data	
D^{ε}: Vector of AFMIA predictions	Forecast residual direction of vibration range (VBR)
MAPE: Mean absolute percentage error	AS1, AS2: Accidents datasets 1 and 2.
β : Residual swing magnitude	MMFCR : Method of the most frequent cumulative residual
σ: Residual quality	RELU: Rectified Linear Activation Function
$\widehat{\beta}, \hat{\sigma}$: Residual swing control parameters	AdaM: Adaptive moment estimator optimiser
$\phi_{t,t+1}$: Residual span value between t and $t+1$	
S_i: Residual state i	
$\rho^{+}\{r_K^{\varepsilon}\}$: Residual r_K^{ε} of positive polarity	
T^R T^I: regular and irregular step transitions	
$\rho\left(r_j^{*\varepsilon}\right)$: Polarity of ANN residual of point j	
P: crisp Markov transition matrix	

3.2.2 Screening of ANN predictions

One of the major problems that impede effective ANN time series prediction is the existence of a plethora of local optima in the model search space. This feature creates a high probability of predictions being produced from unsuitable local optima during the learning process.

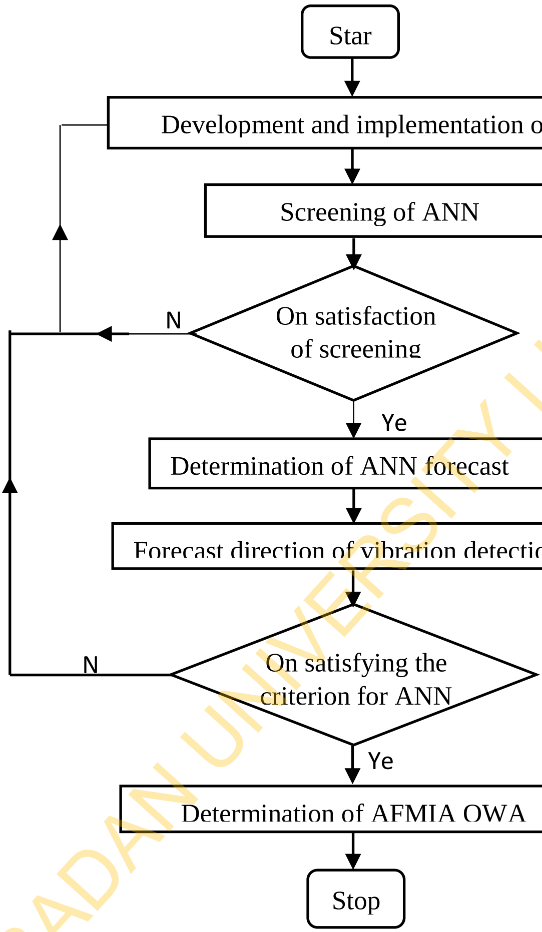


Figure 2: Flowchart of the AFMIA forecasting model

Table 2: CNN architecture parameters and corresponding values as used in the study

CNN architecture Parameter	Convolution dimension	Kernel size	Activation function	Input shape (batch size, samples, height, timestep, features)
	1	64	RELU	(1, dynamic ,4,1)
CNN architecture Parameter	**Pool size**	**Layers (dense/output)**	**Optimiser**	**Loss function**
	2	(5/1)	AdaM	MSE

To surmount this, D^{ε} was determined to exist in the range of acceptable predictions by evaluating some properties of their residuals r^{ε} and ensuring that they exist within a range which we define in this study as acceptable. The properties investigated were the residual swing magnitude β and residual quality σ.

β was obtained from the *MAPE* of D^{ε} and tries to capture the general swing magnitude prevalent in the residuals. σ, which estimates the were independent and identical distribution characteristic of the residuals was estimated using equations 10 and 11. The equation seeks to estimate the extent of dispersion of positive and negative peaks and troughs within the residuals.

The control property value $\hat{\beta}$ was set based on the maximum *MAPE* prediction fitness values produced by the grey-Markov $\{GM(1,1)\}$ and the double-declining exponential smoothing models on the preliminary application of the industrial accident data. $\hat{\sigma}$ was set to the value of 0.5. In both cases, the control property works by making D^{ε} and D_{K+1}^{ε} accepted prediction vector $D^{*\varepsilon}$ and forecast $D_{K+1}^{*\varepsilon}$ respectively when $\beta(D^{\varepsilon})$ and

$\sigma(D^{\varepsilon})$ do not exceed their respective control values and rejecting D^{ε} otherwise.

$$\sigma(D^{\varepsilon}) = \frac{\sum_{j=1}^{K-2} s_j}{2*(K-2)} \{\rho(r_{j-1}^{*\varepsilon}) \neq \rho(r_j^{*\varepsilon}), j = 1,2, \dots K-2\} \quad (10)$$

$$r_j^* = \begin{cases} r_{t+1}^{\varepsilon} & \{t = K\} \\ r_t^{\varepsilon}(\rho\{r_t^{\varepsilon}\} \neq \rho\{r_{t+1}^{\varepsilon}\}) & \{otherwise\} \end{cases} \quad (11)$$

$$r_t^{\varepsilon} = \frac{100(D_t - D_t^{\varepsilon})}{D_t} \quad (12)$$

$$t: N+1 \dots, K-1, K$$

3.3 Forecast correction

This is the most important stage of the neuro fuzzy-Markov model development. We assume that the D^{ε} is not accurate due to issues relating to local and plateau optima as well as model under-learning and over-learning. As such, the fuzzy Markov model was applied in the analysis of the ANN output towards obtaining improved outcomes.

The process which essentially involves the analysis to obtain the cumulative and span values of residuals, the fuzzy classification of r^{ε}, determination of a correction factor and subsequent modification of $D_{K+1}^{*\varepsilon}$ is discussed in this section.

3.3.1 Cumulative residual error and span residual error analysis

Based on the inferences provided by r^{ε} on the behaviour of the ANN model's attempt to accurately predict the historical data, we theorised that if a forecast residual r_{K+1}^{ε} was determined, then the forecast accuracy of the neural network model could be improved. To this effect, the concept of cumulative error (*CE*) tracking was introduced. *CE* tracking seeks to determine the sum of errors produced by the ANN at point t, dependent on the magnitude and polar direction of r_t^{ε}. Thus within the residual error space, the *CE* at t (C_t) is defined by equations 13. Also, to obtain C_t, the swing spans of r_t^{ε} ($d_{t,t+1}$) were determined (equation 14)

$$C_t = \begin{cases} C_{t-1} - |r_t^{\varepsilon}| & \{(\rho\{r_t^{\varepsilon}\} = \rho\{r_{t+1}^{\varepsilon}\}) \wedge (|r_t^{\varepsilon}| > |r_{t+1}^{\varepsilon}|)\} \\ C_{t-1} + |r_t^{\varepsilon}| & \{(\rho\{r_t^{\varepsilon}\} = \rho\{r_{t+1}^{\varepsilon}\}) \wedge (|r_t^{\varepsilon}| \leq |r_{t+1}^{\varepsilon}|)\} \\ |r_t^{\varepsilon}| & \{(\rho\{r_t^{\varepsilon}\} \neq \rho\{r_{t+1}^{\varepsilon}\})\} \end{cases} \quad (13)$$

$$\phi_{t,t+1} = \begin{cases} |r_{t+1}^{\varepsilon} - r_t^{\varepsilon}| & \{\rho\{r_t^{\varepsilon}\} = \rho\{r_{t+1}^{\varepsilon}\}\} \\ C_{t-1} + |r_t^{\varepsilon}| & \{otherwise\} \end{cases} \quad (14)$$

3.3.2 Fuzzy classification of span residual errors

The residual error spans ($\phi_{t,t+1}$), were subsequently classified on basis of their span magnitudes (*SM*). Three lingustic crisp *SM* classes namely, Low $\{L: (0, 0.33\theta)\}$, Medium $\{M: (0.33\theta, 0.67\theta)\}$ and Large $\{H: (0.67\theta, \theta)\}$ were defined. The fuzzy membership of $\phi_{t,t+1}$ in the defined classes were determined using triangular membership functions (equations 15-17).

$$\mu_L = \mu_{\phi_{t,t+1}}\{L\} = \begin{cases} 1 & \{\phi_{t,t+1} \leq 0.5\bar{L}\} \\ \frac{2\phi_{t,t+1} - \bar{M}}{\bar{L} - \bar{M}} & \{0.5\bar{L} < \phi_{t,t+1} \leq 0.5\bar{M}\} \\ 0 & \{otherwise\} \end{cases} \tag{15}$$

$$\mu_M = \mu_{\phi_{t,t+1}}\{M\} = \begin{cases} \frac{\bar{L} - 2\phi_{t,t+1}}{\bar{L} - \bar{M}} & \{0.5\bar{L} < \phi_{t,t+1} \leq 0.5\bar{M}\} \\ \frac{\bar{H} - 2\phi_{t,t+1}}{\bar{H} - \bar{M}} & \{0.5\bar{M} < \phi_{t,t+1} \leq 0.5\bar{H}\} \\ 0 & \{otherwise\} \end{cases} \tag{16}$$

$$\mu_H = \mu_{\phi_{t,t+1}}\{H\} = \begin{cases} \frac{2\phi_{t,t+1} - \bar{M}}{\bar{H} - \bar{M}} & \{0.5\bar{M} < \phi_{t,t+1} \leq 0.5\bar{H}\} \\ 1 & \{\phi_{t,t+1} \geq 0.5\bar{H}\} \\ 0 & \{otherwise\} \end{cases} \tag{17}$$

Where $\bar{L}, \bar{M}$ and $\bar{H}$ are the mid point values of L, M, H respectively.

$$f\{SM\}_{ij} = \begin{array}{c c} & \begin{array}{ccc} L & M & H \end{array} \\ \begin{array}{c} L \\ M \\ H \end{array} & \begin{pmatrix} f_{LM} & f_{LH} & \\ f_{ML} & f_{MM} & f_{MH} \\ f_{HL} & f_{HM} & f_{HH} \end{pmatrix} \end{array} \tag{18}$$

$$A_{SM}\{SM\}_{ij} = \begin{array}{c c} & \begin{array}{ccc} \mu_L & \mu_M & \mu_H \end{array} \\ \begin{array}{c} L \\ M \\ H \end{array} & \begin{pmatrix} A_{LL} & A_{ML} & A_{HL} \\ A_{LM} & A_{MM} & A_{HM} \\ A_{LH} & A_{MH} & A_{HH} \end{pmatrix} \end{array} \tag{19}$$

3.3.3 Determination of correction span for ANN forecast

The frequencies of $\mu_{\phi_{t,t+1}}\{SM\}$ as they occur in the different SM classes were determined, recorded and used to develop the crisp Markov transition matrix (P) as discussed in section XX (equations 18 and 20). Also, the fuzzy partition which shows the corresponding fuzzy states of the system $\left(Q = \mu_{SM}\{SM\}_{ij}\right)$ [equations 19 and 21] was also determined.

$$P\{SM\}_{ij} = \frac{f\{SM\}_{ij}}{\sum_{i=1}^{3} f\{SM\}_{ij}} \qquad \left\{\mu_{SM}\{SM\}_{ij} > 0\right\} \tag{20}$$

$$\mu_{SM}\{SM\}_{ij} = \max\left(A_{SM}\{SM\}_{ij}\right) \tag{21}$$

It is worth noting that $A_{SM}\{SM\}_{ij}$ is the set containing $\mu_{SM}\{SM\}_{ij}$ for all $\phi_{t,t+1}$ whose fuzzy membership falls within that set.

Based on P and Q and their derivatives, fuzzy-Markov transition probability matrix $\tilde{P}$ was then derived using equation 9. Let R be the vector containing the upper bounds of the SM classes, then the correction span vector was obtained as,

$$\tilde{F}_c = \sum_{j=1}^{3} \tilde{P}_{ij} R_j \qquad \left\{\mu_{\phi_{K-1,K}}\{SM\} > 0; R_j = \{0.33\theta, 0.67\theta, \theta\}\right\} \tag{22}$$

Finally, the ANN forecast correction span ($\phi_{K,K+1}$) was obtained as,

$$\phi_{K,K+1} = \max\left(\tilde{F}_c\right) \tag{23}$$

3.4 Forecast direction of vibration detection analysis

Observe from section 3.3 that the approach used in obtaining $\phi_{K,K+1}$ involved the use of values in their absolute forms. As such, the expected forecast cumulative residual is expected to be of the form described in equation (24). However, the direction of a TS event is usually exclusive to a single direction per instance of occurrence. This is what is referred to as the TS direction of vibration. This section aims to determine this direction thereby determining a single C_{K+1} value in the process.

$$C_{K+1} = 0.5(C_K \pm \phi_{K,K+1}) \tag{24}$$

The procedure essentially involves the creation of residual states, determination and classification of step transitions determined for r^{ε}, a procedure to obtain the forecast step transition and analysis to establish the forecast residual direction of vibration range (DVR). It is assumed here that r^{ε} satisfies the conditions for normally distributed residuals.

3.4.1 Creation of states for ANN prediction residuals

The second step involved the determination of the number of residual step transitions that occur from one ANN prediction point to another. In this regard, we initially apply the principle of grey-Markov forecasting of fluctuating time series (Zhan-li & Sun, 2011) which requires that if $\rho\{r_K^{\varepsilon}\}$ is of a particular polarity, then r^{ε} be reversed from the most polar residual form of $\rho\{r_K^{\varepsilon}\}$ to the least polar form. For example, if $\rho^{+}\{r_K^{\varepsilon}\}$, that is if r_K^{ε} is of positive polarity, then we reverse the set of residuals to obtain a new set $\hat{r}^{\varepsilon}$ arranged from the most positive r_t^{ε} to the least positive.

In the next step, we a series of residual states $S_i\{S_i = (r_{LB}, r_{UB}); i = 1,2,3, \dots W\}$ for $\hat{r}_K^{\varepsilon}$ using a fixed residual width w. In this work, the value of w was fixed using equation (25).

$$w = 0.5 * (MAPE^{*}\{D^{\varepsilon}\} - 1)$$

Where $MAPE^{*}\{D^{\varepsilon}\}$ is the integer form of $MAPE\{D^{\varepsilon}\}$.

As an example, suppose the vector V containing five residual values and $V = r^{\varepsilon} = \{-13.23, +4.71, +8.11, -10.24, +4.37\}$, then

$$r^{\varepsilon} = +8.11, +4.71, +4.37, -10.24, -13.23 \qquad (25)$$

$$S = \{-14.00, -10.50, -7.00, -3.50, 0.00, 3.50, 7.00, 10.50, +14.00\} \qquad (26)$$

$$S_i = \{i = 1: (-14.00, -10.50), i = 2: (-10.50, -7.00), \dots, i = 8: (10.50, 14.00)\}$$

A few observations can be made from S_i from equation 26.

i. The states are created from point 0.00 and progressively extended to the positive and negative polar coordinates using the mean of r^{ε} as w, but numbered in order of $\hat{r}^{\varepsilon}$.

ii. S_i may be filled (occupies r_t^{ε}) or empty.

iii. Within each polar coordinate, S_i generation should be terminated as soon as $\max(r_t^{\varepsilon})$ or $\min(r_t^{\varepsilon})$ has been accommodated by an already generated state except in cases of observation iv.

iv. An extra state is created to account for the imprecision in the residual location in situations where $\Delta S_a < \Delta S_b$.

$$\Delta S_a = \begin{cases} |(|min(S_i^{LB})| - |r_t^{\varepsilon}\{min(S_i)\}|)| \\ |(|max(S_i^{UB})| - |r_t^{\varepsilon}\{max(S_i)\}|)| \end{cases} \quad (27)$$

$$\Delta S_b = \begin{cases} |(|min(S_i^{UB})| - |r_t^{\varepsilon}\{min(S_i)\}|)| \\ |(|max(S_i^{LB})| - |r_t^{\varepsilon}\{max(S_i)\}|)| \end{cases} \quad (28)$$

3.4.2 Residual step transition determination and classification

This stage involved the determination of the step transitions of the residuals. The location of each r_t^{ε} in S_i and the state number $S_i\{r_t^{\varepsilon}\}$. The step transitions of r_t^{ε} from point t to $t+1$ were determined (equation 29)

$$T_{t,t+1} = S_i\{r_{t+1}^{\varepsilon}\} - S_i\{r_t^{\varepsilon}\} \quad (29)$$

The set of step transitions T was then grouped into two classes namely regular transitions T^R and irregular transitions T^I. T^R are those step transitions that occur frequently while T^I are those that may be termed abnormal or outlier swings. Based on preliminary observations on twenty normally distributed residuals, the swing classes are defined as

$$T_{t,t+1} \in \begin{cases} T_j^R \ \{|T_{t,t+1}| \leq 2\} \\ T_j^I \ \ \{otherwise\} \end{cases} \quad (30)$$

Each transition class was further split into two classes in terms of the direction of the step transitions. As in the case of regular transitions, $T_j^R = \{T_j^{R+}, T_j^{R-}\}$ where T_j^{R+}, T_j^{R-} are the regular step transitions in the forward and backward direction respectively for the transition start position S_K. T_j^I was similarly defined.

3.4.3 Forecast residual transition step determination

Deploying the defined classes, the forecast transition step was then determined. Due to the imprecise nature of the system, forecast transitions can either in the forward or reverse direction. As such both directions had to be investigated towards obtaining the desired forecast direction. The forward and backward transition step candidates were first derived (equation 31). Their corresponding candidates for the desired forecast residual states S_k^+ and S_k^- repectively were located in S_i (equation 32). The respective forecast residual state boundary values $r_{LB}^{\varepsilon}\{S_k^q\}$ and $r_{UB}^{\varepsilon}\{S_k^q\}$ and for all transition directions q were then located.

$$T_{k,k+1} = \begin{cases} \bar{T}_j^{*Rq} & \{T_j^{Rq} \neq \{\emptyset\}; \forall q\} \\ T_j^{Rq} & \{otherwise; \forall q\} \end{cases} \quad (31)$$

$$S_k^q = \begin{cases} S_K + T_{k,k+1} & \{T_{k,k+1} - Int(T_{k,k+1}) \leq 0; \forall q\} \\ S_K + T_{k,k+1} + 1 & \{otherwise; \forall q\} \end{cases} \quad (32)$$

3.4.4 Range of forecast residual direction and ANN forecast and prediction correction

This phase of the analysis concerns the determination of the range of forecast residual direction (V_{LB}, V_{UB}). This involved capturing potential scenarios that could result from the transition process. The aim is to guide the decision on the choice of the members of $r_{LB}^{\varepsilon}\{S_K^q\}$ and $r_{UB}^{\varepsilon}\{S_K^q\}$ that constitute (V_{LB}, V_{UB}). The scenarios and the recommended forecast residual direction values are presented in equations (32-35). Note that the computation of $C_{K+1}^a[C_{K+1}^a = 0.5(C_K + \phi_{K,K+1})]$ and $C_{K+1}^b[C_{K+1}^b = 0.5(C_K - \phi_{K,K+1})]$ using equation (24) was undertaken before proceeding.

If $|T_{K-1,K}| \in T_j^I$, then compute S_K^a,

$$S_K^a = S_k + min[T_{t,t+1}] \quad (33)$$

$$(V_{LB}, V_{UB}) = \begin{cases} [min(r_h^{\varepsilon}\{S_K^a\}), max(r_h^{\varepsilon}\{S_K\})] & \{\rho^-\{T_j^I\}\} \\ [max(r_h^{\varepsilon}\{S_K\}), min(r_h^{\varepsilon}\{S_K^a\})] & \{\rho^+\{T_j^I\}\} \end{cases}$$

Else, compute

$$R^{\varepsilon}\{S_K^q\} = r_{LB}^{\varepsilon}\{S_K^q\} \cup r_{UB}^{\varepsilon}\{S_K^q\} \quad (34)$$

If $\rho\{C_{K+1}^a\} = \rho\{C_{K+1}^b\}$, then

Obtain $R^{*\varepsilon}\{S_K^q\} \subseteq R^{\varepsilon}\{S_K^q\}$ such that, $[(\rho\{R_i^{*\varepsilon}\} = \rho\{C_{K+1}^q\}) \vee (R_i^{*\varepsilon} = 0),] \ \{\forall i\}$

where $R_i^{*\varepsilon}$ are elements of $R^{*\varepsilon}$

$$(V_{LB}, V_{UB}) = \begin{cases} \left[min\left(R_i^{*\varepsilon}\{S_K^q\}\right), max\left(R_i^{*\varepsilon}\{S_K^q\}\right)\right] & \left\{\rho^-\{C_{K+1}^q\}\right\} \\ \left[max\left(R_i^{*\varepsilon}\{S_K^q\}\right), min\left(R_i^{*\varepsilon}\{S_K^q\}\right)\right] & \left\{\rho^+\{C_{K+1}^a\}\right\} \end{cases}$$

(35)

Otherwise,

$$(V_{LB}, V_{UB}) = \left\{\left[min\left(R_i^{\varepsilon}\{S_K^q\}\right), max\left(R_i^{\varepsilon}\{S_K^q\}\right)\right]\right.$$

(36)

The scenario captured by equation 33 is that in which an irregular residual swing occurs at the period just preceding the forecast period $\left(T_{j=K}^I\right)$. In such a situation, the forecast DVR was fixed as that existing between the S_K and the maximum state length of existing step transitions in the reverse direction of $T_{j=K}^I$. If the first scenario is not observed, then we check for the scenario in the forecast cumulative residual candidates are of a similar polarity. If this exists, C_{K+1}^q must be evaluated using the DVR with the same polar origin as C_{K+1}^q (equation 35). If all $R^{\varepsilon}\{S_K^q\}$values are of non-similar poles to C_{K+1}^q, then $R_i^{\varepsilon}\{S_K^q\}$ of magnitude closest to C_{K+1}^q are deployed to create the DVR. However, if non of the first two scenarios play out, then the DVR is fixed by the most negative and least negative values of the forecast residual state boundaries (equation 36).

Once the $(V_{LB},\ V_{UB})$ has been determined, the C_{K+1}^q that meets the conditions of the DVR (C_{K+1}) is subsequently determined as the C_{K+1}^q having the least proximity value from the midpoint of the DVR (equation 37).

$$C_{K+1} = \begin{cases} C_{K+1}^a & \left\{\left|C_{K+1}^a - \bar{V}\right| < \left|C_{K+1}^b - \bar{V}\right|\right\} \\ C_{K+1}^b & \left\{\left|C_{K+1}^b - \bar{V}\right| < \left|C_{K+1}^a - \bar{V}\right|\right\} \\ 0.5(C_{K+1}^a C_{K+1}^b) & \{Otherwise\} \end{cases}$$

(37)

C_{K+1} serves as the correction factor to the ANN forecast D_{K+1}^{ε}. Thus, the corrected ANN forecast $\widehat{D}_{K+1}^{\varepsilon}$was subsequently obtained (equation 38).

$$\widehat{D}_{K+1}^{\varepsilon} = \frac{D_{K+1}^{\varepsilon}}{(1 - 0.01C_{K+1})} \qquad (38)$$

3.5 AFMIA forecast and prediction mechanism

The ANN aspect of the model requires that D_{K+1}^{ε} candidates be generated to ensure a statistically acceptable forecast. However, the iterative process involved in machine learning takes time and can affect model efficiency. In a bid to create a balance between the model's effectiveness and efficiency, each run of the AFMIA was designed to iterate and obtain 5 D_{K+1}^{ε} predictions within three epochs based on a specified start epoch (ε_s). Given that the AFMIA model is a OWA model, the ANN forecast generation and the fuzzy-Markov process was repeated for each $\widehat{D}_{K+1}^{\varepsilon}$ output generated.

Once the prediction and forecast process is completed and the set of $\widehat{D}_{K+1}^{\varepsilon}$ obtained $\widehat{D}$, the single value $\widehat{G}_{K+1}^{\varepsilon}$ was obtained by computing the mean of $\widehat{D}$ (MMean) or the method of the most frequent cumulative residual C_{K+1}^{q} (MMFCR) identified and used to forecast $\widehat{D}_{K+1}^{q\varepsilon}$ (equations 39 - 41).

$$\widehat{G}_{K+1}^{\varepsilon} = \begin{cases} \frac{\sum \widehat{D}^{a}}{\#\widehat{D}^{a}} & \{\#\widehat{D}^{a} > \#\widehat{D}^{b}\} \\ \frac{\sum \widehat{D}^{b}}{\#\widehat{D}^{b}} & \{\#\widehat{D}^{b} > \#\widehat{D}^{a}\} \end{cases} \qquad (39)$$

Where,

$$\widehat{D}^{a} = \left\{\widehat{D}_{(K+1)i}^{\varepsilon} \left\{C_{(K+1)i} = C_{(K+1)i}^{a}\right\}\right\} \qquad (40)$$

$$\widehat{D}^{b} = \left\{\widehat{D}_{(K+1)i}^{\varepsilon} \left\{C_{(K+1)i} = C_{(K+1)i}^{b}\right\}\right\} \qquad (41)$$

The ANN predictions D_t^{ε} $\{t{:}\,1 \ldots, K\}$ were also corrected using a modified form of equation 38. The correction factors for the predictions were taken as the mean of their respective $S_i\{r_t^{\varepsilon}\}$ (equation 42)

$$\widehat{D}_t^{\varepsilon} = \frac{D_t^{\varepsilon}}{\left(1-0.01\bar{S}_i\{r_t^{\varepsilon}\}\right)} \qquad (42)$$

4. Model implementation, validation and evaluation

Given the computational complexity of the model, its process routines were implemented via a computer program developed using the Keras package with the Tensorflow backend of Python 3.7 environment. The model's prediction and forecasting capability was investigated by applying it to two secondary industrial accident data obtained from the literature (Bureau of Safety and Environmental Enforcement, 2018; Gov.UK, 2020). In both cases, the datasets were split into a training set and a testing set using a 70-30 ratio format.

The model's relative performance was investigated by comparing its outputs with those of two traditional models (exponential Smoothing [EXPSM] and ARIMA). The EXPSM and auto-ARIMA feature of the SPSS software was used for the traditional models' analysis. Also, AFMIA's performance was also compared with two non-traditional models Markov based models (Grey-Markov model (GMM) and the grey–fuzzy–Markov and pattern recognition model (GFMAPR). Computer programs were developed for the GMM and GFMAPR based on the theoretical principles provided by Zhan-li and Sun (2011) and (Edem *et al.*, 2018).

4.1 Prediction and forecast accuracy

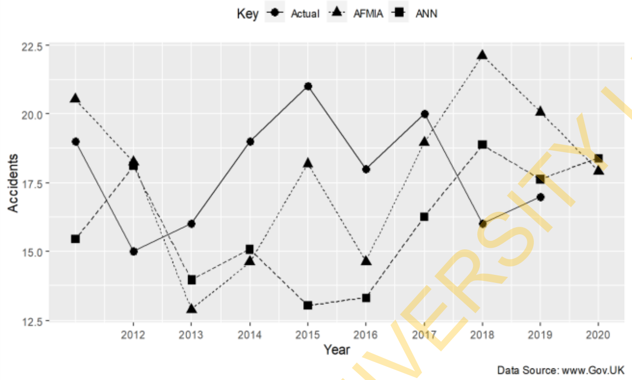
In a bid to ascertain the predictive capability of AFMIA, the industrial accident data (referred here as Accident data sets AS1 and AS2 respectively) were characterized to determine their suitability for TS prediction (Table 3). It was observed in both industrial accident data sets that. This observation helped to reaffirm the observation that although dependent on organisation's safety philosophy, industrial accidents occurrences are random, exhibiting fluctuating tendencies and non-seasonality characteristics. These characteristics fit the data profile for which the AFMIA was developed.

Table 3: Characteristics of fire accidents data to which the AFMIA was applied

Industrial Accidents Data	Characteristics of data		
	Variation	Fluctuation	Randomness
AS1	0.2462	89%	114.61%
AS2	0.2162	78%	76.81

4.2 Expected performance of the model

Given that the model was developed to obtain improved forecasts from the neural network predictor, the results obtained on the application of the model showed forecast superiority over the CNN-ANN base. It can be observed from Figures 3 and 4 that AFMIA exhibited good tracking and anticipation capability notwithstanding the level of variation observed within the two data sets. A summary of the MAPE and RMSE performances of the base and corrector models (Table 4) were observed to be 18.44 and 3.48 compared to 19.68 and 4.08 of the ANN. Similarly for AS2, 15.39 and 26.39 were the respective method of evaluation outcomes obtained from AFMIA application, while the ANN forecast evaluation results were 20.13 and 30.57. From the results and based on the data used, it is clear that the AFMIA is capable of correcting ANN thus improving the potential of improved occurrence monitoring.



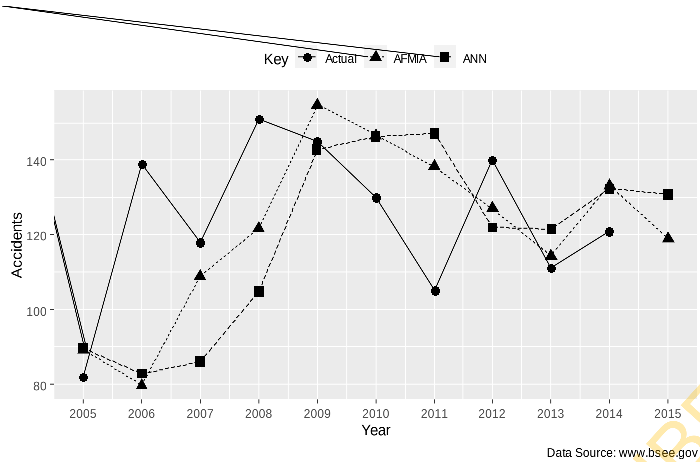


Figure 4: ANN and AFMIA OWA Forecasts for accident data AS2

Table 4: MAPE and RMSE of industrial accidents prediction by the ANN and AFMIA model

	AS1				AS2			
Industrial Accidents Data Type	**In-sample Fit**		**Out-of-sample**		**In-sample Fit**		**Out-of-sample**	
Evaluation Results	**AFMIA**	**ANN**	**AFMIA**	**ANN**	**AFMIA**	**ANN**	**AFMIA**	**ANN**
MAPE	19.50	15.35	18.439	19.679	8.052	7.171	16.828	21.012
RMSE	26.54	7.12	3.483	4.080	9.557	7.683	26.803	31.389

However, regarding the fitted or trained model results, Table 4 and Figures 5 and 6, show that although both models performed reasonably well, the ANN predictor results were more superior. This infers a greater degree of deep learning over the AFMIA. This tendency to overlearn has been one of the problems that have hampered the progress of ANN as a TS predictor and as such a more accurate fitting may lead to an inaccurate forecast of data that was not used in the model development. Further, the fitted data is only an end to a means in itself and is only recognized in situations of little or no information regarding future occurrence accidents. As such it is not considered relevant in this study as the OWA technique used in the validation process provides a more reliable approach at determining the validity of future occurrences.

4.3 AFMIA relative performances with the traditional and non-traditional models

The evaluation of the performance of the model with some of the existing models was also done. MAPE and RMSE results obtained from the application of the non-traditional model forms on AS1 were 49.94 and 10.20 for ARIMA and 20.20 and 3.92 for ESM. Given the degree of variation in AS1, the less accurate results produced by ARIMA is expected due to its inability to handle strong non-linearity situations. However, the ESM produced results that were comparable in accuracy to the AFMIA (Table 5). However, for AS2, both models produced less accurate forecasts in comparison to the AFMIA (Table 5).

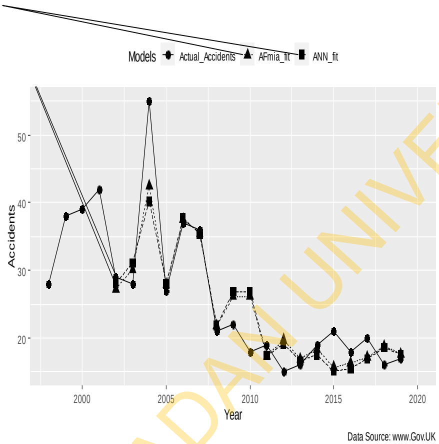


Figure 5: ANN fitted predictions for accident data AS1 and corresponding AFMIA correction

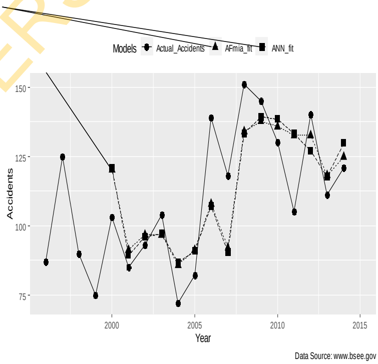


Figure 6: ANN fitted predictions for accident data AS2 and corresponding AFMIA correction

This result indicated that the non-traditional forecasting model types may be more accurate in industrial accident prediction. This view was strengthened by the results obtained from the non-traditional models MSGM and GFMAPR (Table 4).

Table 5: MAPE and RMSE of industrial accidents prediction by evaluated traditional an traditional forecasting models

Industrial Accidents Data Type	AS1				AS2			
	Out-of-sample				Out-of-sample			
Evaluation Results	ARIMA	ESM	MSGM	GFMAPR	ARIMA	ESM	MSGM	GF
MAPE	49.936	20.204	14.818	20.208	20.129	21.051	21.138	1
RMSE	10.20	3.917	2.943	4.216	31.794	30.800	32.789	2

It was observed that for AS1, the MSGM forecasts with MAPE of 14.82 and RMSE of 2.94 were more accurate than those of the AFMIA (MAPE of 18.44 and RMSE of 3.48) although slightly so. Similarly with respect to AS2, the performance evaluation results of the GFMAPR (MAPE = 16.76; RMSE = 27.63) generally matched those of the AFMIA (MAPE = 16.83; RMSE = 26.80). This could indicate the capability of the fuzzy and Markov component of the AFMIA, MSGM and GFMAPR to capture more accurately the imprecision that exists in industrial accident data. Generally, the AFMIA performed well as an industrial-accidents forecasting tool when compared with existing forecasting models.

5. CONCLUSIONS

This study developed a hybrid model (AFMIA) for industrial accident forecasting based on the use of neuro-fuzzy-Markov methods. The model was tested and validated using fitted and out-of-sample data from secondary sources. Also, the model's capability and accuracy were evaluated against existing traditional and non-traditional forecasting models.

Results obtained showed that the developed model produced more accurate forecasts than the ANN model as well as all the traditional models considered. However, AFMIA produced less accurate predictions in some cases when evaluated against the non-traditional models.

AFMIA relative prediction superiority ranges evaluated in terms of the MAPE were 6 – 25%, 9-170%, and 20 – 26% over the ANN, traditional models and non-traditional models respectively. The developed model's relative prediction inferiority range was 0 – 12. Thus based on the limit of

the quantitative evaluation done, the model is can be used in carrying predictions without the need for complex ANN architectural development processes. In addition, the obtained results make the model effective and adequate for the prediction, anticipation and management of industrial accidents.
This work aims to present a perspective on the potential of combining machine learning techniques and related soft computing methods in the development of models with new and better capabilities through the development of the novel ANN-fuzzy-Markov industrial accident forecasting model. Further, the study increases the scope of industrial accidents forecasting as it opens up the window of the potential of combining machine learning techniques and Markov based techniques in industrial accidents investigation and management.

The field of machine learning in forecasting is still unfolding. There is still a lot of potential areas of research that can be carried out to usher in improvement. One such is related to the understanding of the multiple local optima problem that impedes satisfactory machine learning based forecasting. Another area worth investigating relates to the determination of the number and unit size of swing residuals that can produce the most precise neuro-fuzzy-Markov predictions.

It is hoped that this study will motivate more involvement from researchers and stakeholders towards improved safety management through the use of these forms of soft computing methods.

REFERENCES

Adhikari, R. (2015). A neural network based linear ensemble framework for time series forecasting. *Neurocomputing, 157*, 231-242. doi: https://doi. org/10. 1016/j. neucom. 2015 .01.012

Adhikari, R., & Agrawal, R. (2013). An Introductory Study on Time Series Modeling and Forecasting. *ArXiv, abs/1302.6613*.

Aidoo, S. J., & Eshun, P. (2012). Time Series Model of Occupational Injuries Analysis in Ghanaian Mines-A Case Study. *Research Journal of Environmental and Earth Sciences, 4*, 162-165.

Aksoy, H., & Dahamsheh, A. (2009). Artificial neural network models for forecasting monthly precipitation in Jordan. *Stochastic Environmental Research and Risk Assessment, 23*, 917-931. doi:10.1007/s00477-008-0267-x

Ali, D., Yohanna, M., Ijasini, P. M., & Garkida, M. B. (2018). Application of fuzzy – Neuro to model weather parameter variability impacts on electrical load based on long-term forecasting. *Alexandria Engineering Journal, 57*(1), 223-233. doi:https://doi.org/10.1016/j.aej.2016.12.008

Boaisha, S., & Amaitik, S. (2010). *FORECASTING MODEL BASED ON FUZZY TIME SERIES APPROACH*.

Bojadziev, G., & Bojadziev, M. (2007). *Fuzzy Logic for Business, Finance, and Management* (2nd ed.). Singapore: World Scientific Publishing Co. Pte. Ltd.

Bontempi, G. (2008). Long term time series prediction with multi-input multi-output local learning. *Proceedings of the 2nd European Symposium on Time Series Prediction (TSP), ESTSP08*.

Brownlee, J. (2018). *Deep Learning for Time Series Forecasting* S. Martin (Ed.) *Predict the Future with MLPs, CNNs and LSTMs in Python* (pp. 557). Retrieved 10 September, 2020 from https://machinelearningmastery.com/deep-learning-for-time-series-forecasting/

Bureau of Safety and Environmental Enforcement. (2018). Fires Explosions. Available from Bureau of Safety and Environmental Enforcement Retrieved 24th September, 2020 from https://www.bsee.gov/stats-facts/offshore-incident-statistics/fires-explosions

Ceylan, H. (2014). An Artificial Neural Networks Approach to Estimate Occupational Accident: A National Perspective for Turkey. *Mathematical Problems in Engineering, 2014*, 756326. doi:10.1155/2014/756326

Das, H., Naik, B., & Behera, H. S. (2020). A Hybrid Neuro-Fuzzy and Feature Reduction Model for Classification. *Advances in Fuzzy Systems, 2020*, 4152049. doi:10.1155/2020/4152049

Dobrow, R. P. (2016). *Introduction to Stochastic processes with R* (2nd ed.). USA.: John Wiley & Sons, Inc.

Edem, I. E., Oke, S. A., & Adebiyi, K. A. (2018). A novel grey–fuzzy–Markov and pattern recognition model for industrial accident forecasting. *Journal of Industrial Engineering International, 14*(3), 455-489. doi:10.1007/s40092-017-0236-4

Faraway, J., & Chatfield, C. (1998). Time series forecasting with neural networks: a comparative study using the airline data. *Applied Statistics 47*, 231-250.

Ganesan, K., Annamalai, U., & Deivanayagampillai, N. (2019). An integrated new threshold FCMs Markov chain based forecasting model for analyzing the power of stock trading trend. *Financial Innovation, 5*(1), 35. doi:10.1186/s40854-019-0150-4

Ghedira, A., Kammoun, K., & Saad, C. (2018). Temporal Analysis of Road Accidents by ARIMA Model : Case of Tunisia. *International Journal of Innovation and Applied Studies, 24*, 1544-1553.

Gov.UK. (2020). Statistical data set Fire statistics data tables. Available from Gov.UK Retrieved 24th September, 2020 from https://www.gov.uk/government/statistical-data-sets/fire-statistics-data-tables

Javedani Sadaei, H., & Lee, M. H. (2014). Multilayer Stock Forecasting Model Using Fuzzy Time Series. *The Scientific World Journal, 2014*, 610594. doi:10.1155/2014/610594

Kriesel, D. (2010). *A brief introduction to neural networks* Retrieved 14 August, 2020 from http://www.dkriesel.com/en/science/neural_networks

Lan, J.-y., & Zhou, Y. (2014). Application of Gray Markov SCGM1,1c Model to Prediction of Accidents Deaths in Coal Mining. *International Scholarly Research Notices, 2014*. doi:10.1155/2014/632804

Lewis, N. D. (2018). *Deep Time Series Forecasting With Python Predict the Future with MLPs, CNNs and LSTMs in Python* (pp. 557). Retrieved 10 September, 2020 from https://machinelearningmastery.com/deep-learning-for-time-series-forecasting/

Liu, J., Youmin, H., wu, b., Wang, Y., & Xie, F. (2017). A Hybrid Generalized Hidden Markov Model-Based Condition Monitoring

Approach for Rolling Bearings. *Sensors, 17*. doi:10.3390/s17051143

MacLeod, C. (2013). *An introduction to practical neural networks and genetic algorithms for engineers and scientists*. Jerhi Wahyu Fernanda, 157.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PLOS ONE, 13*(3), e0194889. doi:10.1371/journal.pone.0194889

Malysa, T. (2020). Use of Selected Adaptive Models to Forecast Accidents During the Repair, Maintenance, Installation of Machines and Equipment. *Multidisciplinary Aspects of Production Engineering, 3*(1), 141-149. doi:doi:10.2478/mape-2020-0012

Manaloto, R., & Balahadia, F. (2017). *Time series forecasting using multiplicative model: A predictive model for fire risk in the city of Manila*. Paper presented at the IEEE 15th Student Conference on Research and Development (SCOReD), Universiti Teknologi PETRONAS (UTP), Malaysia. 115-120.

Mapuwei, T. W., Bodhlyera, O., & Mwambi, H. (2020). Univariate Time Series Analysis of Short-Term Forecasting Horizons Using Artificial Neural Networks: The Case of Public Ambulance Emergency Preparedness. *Journal of Applied Mathematics, 2020*, 1-11. doi:10.1155/2020/2408698

Mock, C. N., Nugent, R., Kobusingye, O., & Smith, K. R. (2017). *Injury prevention and environmental health* (Vol. 7). Washington, USA: World bank Group.

Pabuçcu, H. (2017). Time series forecasting with neural network and fuzzy logic (pp. 195-204).

Pardo, M. J., & de la Fuente, D. (2010). Fuzzy Markovian decision processes: Application to queueing systems. *Computers & Mathematics with Applications, 60*(9), 2526-2535. doi:https://doi.org/10.1016/j.camwa.2010.08.004

Ratnayaka, R. M. (2017). *Factors Affecting Industrial Accidents: Empirical Evidence from Manufacturing Setting*. Paper presented at the Confronting Development Challenges through Promoting Innovative Management Thinking"

Proceedings of the 6th International Conference on Management and Economics. University of Ruhuna, Sri Lanka. , University of Ruhuna, Sri Lanka. 299-318.

Sánchez-Sánchez, P., García-González, J., & Coronell, L. (2020). Encountered Problems of Time Series with Neural Networks: Models and Architectures.

Sarkar, S., Vinay, S., Raj, R., Maiti, J., & Mitra, P. (2019). Application of optimized machine learning techniques for prediction of occupational accidents. *Computers & Operations Research, 106*, 210-224. doi:https://doi.org/10.1016/j.cor.2018.02.021

Shi, B., Hu, C. H., Yu, X. H., & Hu, X. X. (2016). New fuzzy neural network–Markov model and application in mid- to long-term runoff forecast. *Hydrological Sciences Journal, 61*(6), 1157-1169. doi:10.1080/02626667.2014.986486

Shmueli, G. (2011). To Explain or to Predict? *Statistical Science, 25*(3), 289–310. doi:10.1214/10-STS330

Svs, R. (2018). Prediction of fatal accidents in Indian factories based on ARIMA. *Production Engineering Archives, 18*, 24-30. doi:10.30657/pea.2018.18.04

Tu, J. V. (1996). Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *Journal of Clinical Epidemiology, 49*(11), 1225-1231. doi:https://doi.org/10.1016/S0895-4356(96)00002-9

Urrutia, J., Villaverde, S., Algario, N., Malvar, R., Oliquino, A., & Gano, L. (2018). Forecasting the Number of Fire Accidents in the Philippines through Multiple Linear Regression. *Indian Journal of Science and Technology, 11*, 1-7. doi:10.17485/ijst/2018/v11i20/123341

Voyant, C., Muselli, M., Paoli, C., & Nivet, M.-L. (2011). Optimization of an artificial neural network dedicated to the multivariate forecasting of daily global radiation. *Energy, 36*(1), 348-359. doi:https://doi.org/10.1016/j.energy.2010.10.032

Wang, L., Mehrabi, M., & Kannatey-Asibu, E. (2002). Hidden Markov Model-based Tool Wear Monitoring in Turning. *Journal of Manufacturing Science and Engineering-transactions of The Asme - J MANUF SCI ENG, 124*. doi:10.1115/1.1475320

Wilinski, A. (2019). Time series modeling and forecasting based on a Markov chain with changing transition matrices. *Expert Systems with Applications, 133*, 163-172. doi:https://doi.org/10.1016/j.eswa.2019.04.067

Xiaoping, Z., & Liu, M. (2009). An overview of accident forecasting methodologies. *Journal of Loss Prevention in The Process Industries - J LOSS PREVENT PROC IND, 22*, 484-491. doi:10.1016/j.jlp.2009.03.005

Yarmohammadi, M., & Safaei, M. (2012). Markov Switching Models for Time Series Data with Dramatic Jumps.

Zhan-li, M. A. O., & Sun, J. (2011). Application of Grey-Markov Model in Forecasting Fire Accidents. *Procedia Engineering, 11*, 314-318. doi:10.1016/j.proeng.2011.04.663

Zhang, Y., Qu, H., Wang, W., & Zhao, J. (2020). A Novel Fuzzy Time Series Forecasting Model Based on Multiple Linear Regression and Time Series Clustering. *Mathematical Problems in Engineering, 2020*, 9546792. doi:10.1155/2020/9546792