Weather forecasting is an ever-challenging area of investigation for scientists. It is the application of science and technology in order to predict the state of the atmosphere for a given time and location. Rainfall is one of the weather parameters whose accurate forecasting has significant implications for agriculture and water resource management. In Zambia, agriculture plays a key role in terms of employment and food security. Rainfall forecasting is one of the most complicated and demanding operational responsibilities carried out by meteorological services all over the world. Long-term rainfall prediction is even more a challenging task. It is mainly done by experts who have gained sufficient experience in the use of appropriate forecasting techniques like modelling. It is mainly done by experts who have gained sufficient experience in the use of appropriate forecasting techniques like modelling. In this paper, a rainfall forecasting model using Artificial Neural Network is proposed as a model that can be 'trained' to mimic the knowledge of rainfall forecasting experts.

This makes it possible for researchers to adapt different techniques for different stages in the forecasting process. We begin by noting the five main stages in the seasonal rainfall forecasting process. We then apply artificial neural networks at each step. Initial results show that the artificial neural networks can successfully replace the currently used processes together with the expert knowledge. We further propose the use of these neural networks for teaching such forecasting processes, as they make documentation of the forecasting process easier and hence making the educational process of teaching to forecast seasonal rainfall easier as well. Artificial Neural Networks are reliable, handle more data at one time by virtual of being computer based, are less tedious and less dependent on user experience.

Keywords—Artificial Neural Networks; seasonal rainfall forecasting; soft computing; Artificial Intelligence; machine learning;

I. INTRODUCTION

Weather is the state of the atmospheric conditions at a place duration given time. It is defined in terms of weather parameters like temperature, pressure, wind (speed and direction) and rainfall. Weather forecasting is the application of science and technology to predict the state of the atmosphere for a given time and location [1].

Rainfall is a natural phenomenon resulting from atmospheric, oceanic circulation systems and complex physical processes that cause an amount of rain to fall at a place during a particular period. Rainfall is one of the weather parameters whose accurate forecasting has significant implications for agriculture and water resource management [2] [3].

Amongst all weather parameters, rainfall is the one that mostly affects human life and livelihood in developing countries and least developed countries like Zambia where majority of the population depends on rain fed agriculture [3] [4]. Rainfall also affects many sectors including but not limited to water resources management, energy, tourism, health, disaster risk reduction (DRR) and infrastructure development. Thus, accurate seasonal rainfall forecasting is essential for planning and management of many sectors [5] [6].

One of the most complicated and demanding operational responsibilities carried out by meteorological services all over the world is rainfall forecasting [3] [7]. Forecasting rainfall is complex due to the various dynamic environmental factors, both spatial and temporal random variations [8].

Seasonal rainfall forecast, which is the prediction of the expected rainfall performance for a given rainy season, is usually generated in August and issued in September in the Southern African Development Community (SADC) region. Current seasonal rainfall forecasting methods used in Zambia...
have been proved to be less accurate [9]. These techniques do not take into account all factors that may influence rainfall. The statistical models based on regression analysis and eyeball inspection are used.

This research proposes to use Artificial Neural Networks (ANNs) in order to increase the accuracy of seasonal rainfall forecast in Zambia, since these statistical models have some inherent limitations over long range rainfall forecasts [10] [11].

The current seasonal rainfall forecasting method assumes a direct correlation between the Pacific Sea Surface Temperatures (SSTs) and station rainfall observations. Atmospheric systems are not governed by only these two factors, but this assumption ignores the availability of other factors in influencing rainfall [12]. Other parameters that may have influence on rainfall include temperature, relative humidity and wind speed [3]. A common weakness of statistical rainfall forecasting models is that while the correlations are assumed to remain constant for the duration of the forecast, they usually change with time and slowly lose their significance [13]. Long range weather forecasts like seasonal rainfall forecast become less accurate as the difference in time between the present moment and the time for which the forecast is given increases.

A limitation of high spatial variability of station point rainfall observations increases the inaccuracy and uncertainty that reduce the skill (accuracy) of the seasonal rainfall forecasts. Moreover, changing climate has introduced further uncertainties that need to be considered in this assumption of a direct linear correlation between observed rainfall data and SSTs [12] [14]. Seasonal rainfall forecasts in Zambia are therefore, currently not of high efficacy.

Furthermore, some of the stages in the current seasonal rainfall forecasting process require expert knowledge through eyeball inspection which is not easy to pass on through an educational process.

This research targets to generate accurate seasonal rainfall forecasts which will help the country to plan and prepare for the farming season because the majority of the population rely on rain fed agriculture in Zambia. Seasonal rainfall forecasts provides information that help the government and stakeholders to prepare for and reduce the potential negative impacts of climate change, hence accurate and precise forecast will play an integral part in the preparedness [15].

This study aims to improve accuracy of forecasting seasonal rainfall in Zambia through the use of ANNs, which consequently make it easier to capture expert knowledge used in forecasting seasonal rainfall. We discuss the work done thus far in this paper. In the rest of this paper we outline how we plan to achieve this aim; in section II we discuss artificial neural networks (ANNs). In section III we discuss related technologies. These related technologies will be the bases for selecting a suitable method for generating seasonal rainfall forecasts in Zambia. In section IV we proceed to discuss a proposed proof of concept model for forecasting seasonal rainfall based on the ANNs discussed in section III. Finally, in section IV, we discuss the currently used procedure and also propose our new approach and how we will evaluate the proposed proof of concept model using the currently used forecasting method and real data in Zambia. We cover initial experimentation and results in section V.

II. ARTIFICIAL NEURAL NETWORKS

Soft computing deals with approximate models to provide imprecise but usable solutions to complex computation problems giving results in approximation. The three basic components of soft computing are Artificial Neural Network (ANN), Fuzzy Logic (FL) and Genetic Algorithm (GA) [10]. Artificial neural network which is a branch of Artificial Intelligence (AI) have a long history of interactions with Robotics. AI implies the use of computers to model intelligent behavior with minimal human intervention [16]. Fig. 1 shows a simple three layer artificial neural network.

Soft computing systems are flexible enough to adapt or cope with changes encountered. They are robust to be tolerant when confronted with imprecise information and able to react within a reasonable time in response to events. The flexibility, robustness, nonlinearity and quick response nature of soft computing made it attractive technology for weather forecasting [10].

Weather forecasting require some intelligent computing, which can read nonlinear data and generate some rules and patterns to study and train from the observation data to predict future weather [17].

![Fig 1: A simple three-layer neural network](image-url)

Artificial neural networks are a beautiful biologically-inspired information processing technique that draws inspiration from the way the human brain works into a programming paradigm which enables a computer to learn from observational data [18]. Using these neural networks, it is possible to train software tools that can maintain expertise on tasks using observational data. A neural network acquires knowledge through a learning (training) process and inter-neuron connection strengths are used to store the knowledge. The learning rules enables a network to gain knowledge from
available data and apply that knowledge to train the network [19]. A trained neural network can be thought of as an expert in the category of information it has been given to analyse [20]. Neural networks are currently used as a machine-learning technique for solving a variety of tasks, including language translation, image classification, image generation and weather forecasting [21].

ANN models are based on prediction by smartly analyzing trend from an already existing huge historical set of data [7]. Artificial neural network provides users a model free, which can generate input-output mapping for any set of data as complex pattern recognition can be attempted without making any initial assumptions [22][8].

Machine learning applies artificial intelligence that provides ability for systems to automatically learn and improve from experience without being explicitly programmed. It is relatively robust and does not require complete understanding of the physical processes that govern the atmosphere. One of the core objectives of machine learning is to instruct computers to use past experience to solve a given problem [23]. These features make machine learning a viable alternative for weather forecasting for longer periods.

Deep learning is a subset of machine learning in Artificial Intelligence that have capable networks of unsupervised learning from unstructured data. It is a powerful set of techniques for learning in neural networks that has also drawn the attention of researchers. Deep Neural Networks are capable of learning techniques sequentially hence able of overcoming catastrophic forgetting. This gives deep learning neural networks the ability of maintaining expertise even in tasks which they have not encountered for a long time [24]. Such capability is important especially when it comes to weather forecasting where certain phenomenon can go unobserved for long periods of time due to the non-linearity in weather data [4].

Researchers have used all three basic soft computing techniques to forecast rainfall. Nevertheless some important features of artificial neural networks, make many researchers prefer to use Artificial Neural Networks because:-

a) Artificial neural network has ability to implicitly detect complex nonlinear relationships between dependent and independent variables.

b) Artificial neural networks are data-driven, self-adaptive methods that learn from examples and don’t need restrictive assumptions about the form of the basic model. ANN discover the relationships among data which may be too complex to define.

c) Artificial neural networks are a family of massively parallel architectures capable of learning and generalizing from examples and experience to produce meaningful solutions to problems, although input data contain errors and are incomplete. This makes ANN a powerful tool for solving problems like forecasting. Neural network is efficient at training large-size data samples due to its parallel processing capability.

d) Artificial neural network can also predict patterns which are not provided during training [25][26].

The neural networks can be used to replace human expertise in situations like image analysis where eyeball inspection is the norm. They are also useful in situation where models are complex and not easy to derive.

III. RELATED TECHNOLOGIES

Soft computing techniques like artificial neural networks have been widely used in many different applications of weather forecasting. The results have proved to be powerful methods which excel at function approximation and pattern recognition than conventional approaches. Since the atmosphere is chaotic and weather data is nonlinear following a very irregular trend, soft computing techniques are considered to be better a approach for developing effective and reliable nonlinear predictive models for weather analysis [27].

Developments in Artificial Intelligence (AI) techniques, in particular ANNs provide superior rainfall forecasting methods. For example, the performance of the rainfall forecasts produced using ANNs in India has better accuracy than statistical and mathematical models [6]. In Bangkok, Thailand, forecasts by artificial neural network model were compared with simple persistent method and ANN forecasts were superior. Artificial neural networks are an aggressive method of forecasting rainfall over the linear regression method because of its ability to be trained and adapt [28]. Artificial neural network has been successfully used to forecast rainfall for longer periods, like one year in advance in the West mountainous region of Iran and the results have been said to be very accurate [29]. Research has been carried out for a long time in forecasting rainfall for longer periods using numerical and statistical models, and successes of these models are rarely visible [12].

Artificial neural networks rainfall forecasting methods provide aggressive models over the existing rainfall forecast methods used in Zambia because ANN has the ability to be trained using historical data and to adapt [9]. A. Chaturvedi, applied ANN methods using back propagation method for rainfall prediction in Delhi - India, a region that highly depends on monsoon and seasonal rainfall for agriculture activities and results showed a minimal Mean Square Error (MSE) [6]. A survey conducted in 2013 show that ANN techniques are more suitable than traditional statistical and numerical methods in forecasting rainfall [10]. Artificial neural network minimizes the error in weather forecasting using various algorithms and gives a predicted value which is nearly equal to the actual value [30].

G. Shrivastava et al argues that pattern recognition and prediction in a deterministic approach through ANN techniques based on back-propagation algorithm has been proved to be the most efficient way for long term rainfall prediction [12]. Neural network is the best tool for pattern recognition through training with past data [31].
Using ANN algorithms in forecasting rainfall has become an attractive approach because of its flexibility, nonlinearity and ability in data driven learning through building models without any prior knowledge. Artificial neural networks have been used as a suitable technique for the long-term climate variability like seasonal rainfall forecast due to the fact that learning is accomplished through training [10]. ANN forecasting models are based on prediction by smartly analysing the trends from an already existing voluminous historical set of data. Mathematical or statistical weather models have been found to be very accurate in calculation, but not in prediction as they cannot adapt to the irregularly varying patterns of atmospheric data which can neither be written in form of a function nor deduced from a formula [3].

Using artificial neural networks to forecast rainfall gives more details in terms of forecasting for a specific location [9]. In the two regions of Chhattisgarh – India, different datasets for a period of ten years were used as inputs of artificial neural network in rainfall prediction, and the output accuracy percentage was between 80 – 90% [23]. The ability of ANNs to cope with nonlinearities, speed of computation, learning capacity and prediction accuracy makes it a superior model of forecasting rainfall [32]. ANN multilayer perceptron has the ability to be trained with error correction learning. Most applications in rainfall forecasting utilize a feed-forward neural network that incorporates the standard static multilayer perceptron (MLP) trained with back-propagation algorithm [2].

IV. METHODOLOGY

In this section we discuss our proposed approach. We begin by describing the currently used process and then discuss our new proposed approach which is based on the currently used approach.

A. Currently used approach

We base our study and method for forecasting on the currently used procedure for forecasting seasonal rainfall. The main focus of the initial stages of our study is to replace the parts where human expertise especially through eyeball inspection is the norm for doing analysis with artificial neural networks. The currently used procedure is illustrated in Fig. 2. The first step is the identification of homogeneous rainfall zones. After this, SST data is downloaded and July SST data is used to identify correlation with rainfall regions for specified months i.e. January February March (JFM). This correlation is then used to identify basins that have influence on the rainfall patterns for the correlated months and respective region. The input data (rainfall and SST) obtained has to be normalized because they are of different units. The next stage is a regression based analysis which is used to find the relationship between rainfall amounts and the correlated basins. Empirical statistical forecasting model is developed using Simple Linear Regression model (SLRM) to describe a linear relationship between two variables: X as independent which is Sea Surface Temperature basins and Y as dependent is rainfall.

To find the forecasted rainfall figure for any region, for the given three months i.e. JFM the following formula is used:

\[ Y = mx + c \]

where,

- \( Y \) is forecasted rainfall figure,
- \( M \) is coefficient for the chosen basin,
- \( X \) is the SST for month of July and
- \( C \) is the constant.
Rainfall data from 1981 to latest year is arranged according to years and sorted by rainfall amount column using excel in ascending order. The number of records of sorted data is then divided into terciles. Any remaining number of records are assigned to the middle group. Calculated (forecasted) rainfall figure is placed to the figure it is closest to. Eyeball inspection is used to determine the category of forecasted rainfall as either Below Normal rainfall, Normal rainfall or Above Normal rainfall. Algorithm 1 shows how classification of seasonal rainfall is done.

Algorithm 1 Seasonal Rainfall Classification
1: Arrange Rainfall Amount from 1981 to Latest Year
2: Sort by Rainfall Column in Ascending order
3: Group Sorted Number of Records into 3 Groups
4: The 3 Groups to Contain Equal Number of Records
5: Assign any Remaining Records to the Middle Group
6: Place Calculated Rainfall Figure next to Number it is Close to
7: Return Either Above Normal Rainfall; Normal Rainfall; or Below Normal Rainfall

Algorithm 2 Rainfall Groups
1: Rainfall Figures are sorted in Ascending order
2: Divide the Number of Records by 3 to Create 3 Groups
3: Assign any Remaining Number of Records to the Middle Group
4: Group with Lower values is Below Normal Rainfall
5: Group in the Middle is Normal Rainfall
6: Group with higher values is Above Normal Rainfall
7: Calculated Rainfall Figure is placed Next to Number it is Close to
8: Return Group Category the Rainfall Figure is placed in

B. Proposed approach

We base our procedure on the currently used approach. However, we incorporate artificial neural networks to replace the steps that involve eyeball inspection by human experts.

1) Identification of Homogeneous Rainfall Zones

The first step in the seasonal rainfall forecasting process is demarcating the country into homogenous rainfall zones. These zones exhibit coherency in rainfall variations. Zambia is demarcated into three rainfall homogeneous zones. Region I covers the southern parts of the country and receives annual rainfall amount of less than 800mm; while Region II covers the central part of the country and receives annual rainfall amount of between 800mm to 1000mm; and Region III covers the northern parts of the country and receives annual rainfall amount of above 1000mm. We also use the already identified zones which are used in the current process. Fig. 5 shows how Zambia is demarcated into homogenous rainfall zones.

Fig. 5. Agro-ecological zones of Zambia and meteorological stations for each region

2) Downloading Sea surface temperatures (SSTs)

After identification of the homogeneous rainfall zones, the second stage is downloading of SSTs. We downloaded the sea surface temperatures from the International Research Institute for Climate and Society (IRI) website.


3) Correlation of rainfall with SSTs

After downloading of sea surface temperatures, we proceed by finding a correlation between station rainfall for a region and the SSTs for given period in months i.e. January February March (JFM). Fig. 6 is a map showing correlation between rainfall for the period of JFM, of region 3 and SST. From the correlation map, basins with a more than ±0.3 correlation are selected using an artificial neural network using pattern recognition to estimate the coordinates for the basin areas. In the old process this was done using eye ball inspection and then zooming to each basin, one at a time to determine best area with a more than ± 0.3 correlation as shown in Fig. 7. After this, we take note of the coordinates for the selected basins and give each basin an identity. After this, all best correlated basins coordinates are noted and recorded.
The coordinates for the selected basin in Fig. 7 are estimated using artificial neural network as lat1 = -01.2, lat2 = 01.3, long1= -13.8 and long2= -10.2, and given name as CWPAC.

4) Regression analysis

For all selected basins, July Sea Surface Temperature for the zones is collected together with rainfall seasonal averages for the same zone. We then apply an artificial neural network to learn the expected rainfall variable Y. In the currently used approach, an empirical statistical forecasting model is developed using Simple Linear Regression model (SLRM) to describe a linear relationship between two variables X as independent(s) (SST basins) and Y as dependent (rainfall). Standardized data for selected basins and rainfall is used.

5) Forecasting

We then use artificial neural network to predict the expected amount of rainfall. To determine the category of the season rainfall, we begin by first sorting the rainfall by amount and then classifying into 3 groups as either below normal, normal or above normal using ANN. We then use this data to train our ANN. For example, we pick rainfall amounts from say 1961 to 2017 and sort them in ascending order, divided into 3 groups and pick a group to which a forecasted rainfall figure belongs to using Neural Network.

In the current process the rainfall amounts from say 1981 to date are sorted in ascending order using excel. The data is then grouped sorted into 3 segments using eyeball inspection. Then the forecasted rainfall figure is compared with the sorted data using eyeball inspection by placing the rainfall figure close to a figure it is close to, thus determining category of the forecasted rainfall as either below normal, normal or above normal rainfall.

V. INITIAL EXPERIMENTATION AND RESULTS

A. Data set

To do our initial experimentation we downloaded data for SSTs from the IRI site. Then, we get data for rainfall statistics for Zambia from 1961 to 2017 see Table I. Apart from the data for rainfall amounts we also collect data from 1961 to 2017 which classifies each season as either above normal, normal or below normal. This data is based on the rainfall amounts recorded in each year and the demarcation into three groupings for the collected data.

<table>
<thead>
<tr>
<th>Location</th>
<th>From</th>
<th>To</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zambia</td>
<td>June 1961</td>
<td>June 2017</td>
</tr>
</tbody>
</table>

B. Setup

We then applied the procedure outlined earlier on in section IV (B). We however only used neural networks for the forecasting stage instead to replace the eyeball inspection and also in place of the regression analysis. To do the implementation of the neural networks, we use Matlab’s neural networking toolkit. All computations were done on a machine running on an i7 processor and windows 10 operating system. We divided the data into sections for training and also for retrospectively testing the predicting ability of our procedure.

C. Discussion and Results

To train our neural network, we used monthly rainfall data from 1961 to 2010 and tested using monthly rainfall data from 2011 to 2017. We were able to pick a group to which a forecasted rainfall figure belongs to using neural network the...
same way it is placed to the group using eyeball inspection in the current process, see Table II.

### TABLE II. PICKING GROUP FOR FORECASTED RAINFALL

<table>
<thead>
<tr>
<th>Year</th>
<th>Current Approach</th>
<th>Proposed Approach (ANN)</th>
<th>Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011</td>
<td>Below Normal</td>
<td>Below Normal</td>
<td>Below Normal</td>
</tr>
<tr>
<td>2012</td>
<td>Below Normal</td>
<td>Below Normal</td>
<td>Below Normal</td>
</tr>
<tr>
<td>2013</td>
<td>Above Normal</td>
<td>Above Normal</td>
<td>Above Normal</td>
</tr>
<tr>
<td>2014</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
</tr>
<tr>
<td>2015</td>
<td>Below Normal</td>
<td>Below Normal</td>
<td>Below Normal</td>
</tr>
<tr>
<td>2016</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
</tr>
<tr>
<td>2017</td>
<td>Below Normal</td>
<td>Below Normal</td>
<td>Below Normal</td>
</tr>
</tbody>
</table>

NNTOOL (Open network/Data manager) which allows to import, create, use and export neural network and data was used. In the performance graph, epochs is plotted against Mean Square Error (MSE). Train, test and validation parameters are plotted against the best case. The graph clearly shows the best validation check is at 44 epochs. Performance graph is plotted as shown in Fig. 8.

Errors reduce after more epochs of training, but could increase on validation data set as network starts over fitting training data. Training stops after consecutive increases in validation error. Best performance is taken from epoch with the lowest validation error.

Regression plots show results using NNTOOL. It is used to validate the network performance. The regression plots in Fig. 9, Fig. 10, Fig. 11 and Fig. 12 display network outputs with respect to targets for training, validation, testing and all. For a perfect fit, all the data should fall along a 45-degree line, where network outputs are equal to targets. In this instance, the fit is very good for all data sets, with regression values in all plots are around 0.99 or above. Thus observed Mean Square Error was under very tolerable level. It has high accuracy and minimized error.
VI. CONCLUSION

The main goal of this research is to enhance processes that are used to generate accurate seasonal rainfall forecasts. The skills required to do seasonal rainfall forecasting are not easy to pass on. In this paper we proposed the use of neural networks to replace the aspect of human expertise and thus make the teaching of the forecasting process easier. From the above discussions and results, it can be concluded that artificial neural network is a reliable long range rainfall forecasting tool and that we can replace human skills which require eyeball inspection with neural networks. It has also shown that this technique can be replicated, without depending on expert experience.

The accuracy level will also be enhanced over the current method as they will be no human errors. It is evident that generating seasonal rainfall forecasts with artificial neural network, will improve the planning and decision making of users that use the forecasts. Such forecasts are fundamental as they will have a positive result on agriculture performance and crop yield in particular. This is of great importance because majority of the population rely on rain fed agriculture in Zambia.

In addition, the seasonal rainfall forecasting provides information that helps the government and stakeholders to prepare for and reduce the potential negative impacts of climate change. This makes accurate forecasting an integral part in the preparedness.

For future works, we plan to replace all the processes that require eyeball inspection with neural networks. We also plan to carry out user testing to evaluate which method is easier to teach that is a comparison between use of neural networks and eyeball inspection.

We further plan to replace the model for selecting the basins with a deep neural network so as to incorporate many other parameters which influence seasonal rainfall forecasting but are not yet incorporated in the currently used model. The deep neural network will also help reduce the risk of forgetting the expertise that is not used over time.

REFERENCES


In our procedure, we started by reading the excel file containing the records into Matlab, count the number of records, and sort the data according to the desired column (rainfall in this case). The number of records were divided into three groups and assigned any remaining records to the middle group. Our procedure was able to predict whether there will be below normal rainfall, normal rainfall or above normal rainfall.

Thus, we successfully replaced the eyeball inspection done in the forecasting stage of the whole procedure as well the regression model with artificial neural networks.

Fig. 11. Testing Regression graph using NNTOOL

Fig. 12. All Regression graph using NNTOOL