Integration of IK, Satellite Imagery Data, Weather Data and Time Series Models in Season Behaviour Predictions. Case of Swayimane, KZN, South Africa

John Nyetanyane(✉)

Central University of Technology, Bloemfontein, South Africa
jnyetanyane@cut.ac.za

Abstract. Impacts of climate change continue to cripple the livelihood of many South Africans by sabotaging the rainfed agricultural systems that they rely heavily on. Despite the value of the indigenous knowledge (IK) in tackling the impacts of climate change, it continues to lose value and precision in season behaviour predictions. In this paper, the IK is integrated with scientific knowledge to enhance the season predictions by small scale farmers. This is achieved by quantifying the collection and processing of IK indicators. Secondly, collect and process the regional weather data (rain and average temperature) from the weather station close to the region of interest and satellite data(vegetation cover, waterbodies cover and soil moisture cover) collected within the region of interest. Thirdly, train the following time series models: Long Short Term Memory (LSTM) network, Seasonal Auto Regressive Moving Average (SARIMA) and Holt Winter's model on the historical weather and satellite data and perform rainy season predictions for scientific perspective. Fourthly, categorize the historical data into warm and cold rainy season periods when normal, below normal and above normal rains were experienced. Fifthly, develop the mobile application that will use the categorized historical data to complement the observation of IK indicators. Lastly, integrate the predictions made via the IK's perspective with the ones performed via the scientific perspective to come up with more robust season predictions.

Keywords: Indigenous knowledge · certainty level · LSTM · SARIMA · TES · Holt Winter's model

1 Introduction

Impacts of climate change continue to demoralize the livelihood of the planet earth's habitats. This is primarily observed in an agricultural sector which is the key to food, water, shelter and many other resources to support the human kind. Rainfed agricultures run by small holder farmers are in fragile state especially in South Africa that remains technologically developing and financially critical [4]. Despite the value that brought by the indigenous knowledge in season behaviour predictions, its accuracy continue to

M. Masinde et al. (Eds.): AFRICATEK 2023, LNICST 520, pp. 165–184, 2024.
https://doi.org/10.1007/978-3-031-63999-9_10

shrink as years progress due to among others climate vagaries, deforestation, pollution and many more. The seasonal climate forecasts on the other side are failing the farmers due to accessibility, interpretation and most importantly they are not locally scaled due to lack and sparse distribution of weather stations and rain gauges [1].

This paper aims at integrating the IK with scientific knowledge to come up with robust season predictions. This is achieved by first quantifying the collection and processing of the IK indicators and their observations. Collect and prepare the historical weather data and satellite data. Train the three time series models: Seasonal Autoregressive Integrated Moving Average (SARIMA), Long Short Term Memory (LSTM) network and Holt Winter's model on the historical data to come up with seasonal predictions for the scientific perspective. Categorize the historical data into warm and cold rainy season periods when normal, below normal and above normal rains were experienced. Further, develop the mobile based application that will use the categorized historical data to complement the predictions that will be made by the farmers using their IK system. The IK's perspective prediction will be integrated with the forecasts made for scientific perspective to come up with more robust predictions.

From the predictions that will be generated, farmers will be able to conclude on the possible season outcome by observing changes in the time series of rain, temperature, waterbodies, soil moisture and vegetation greenness cover.

2 Literature Review

The impacts of climate change continue to engulf the agricultural sector globally and worse in Sub-Saharan Africa which is primarily constituted by the small-holder farmers who are mostly in marginal locations with low level of technology and lack of essential farming resources [2, 21]. Majority of these farmers have claimed to have experienced excessive weather patterns that have resulted in prolonged wet, hot, dry weather conditions that result in crop damage, death of livestock, soil erosion, bush fires, poor plant germination, pests, lower incomes, and deterioration of infrastructure [2].

The government in association with the meteorologists are continuing to provide farmers with seasonal climate forecasts as an adaptation strategy to climate change. These forecasts have been doing well in predictions of rainfall and temperature anomalies to many places around the world. These areas have extensively benefited to food security, drought mitigation and adjustments of cropping decisions by many farmers [3].Despite the value of these forecasts, they are not doing justice to the localities that are remote from the weather stations. This is due to the limited number and sparse distribution of these stations in Africa [1, 22]. The seasonal climate forecasts can be complemented with local based knowledge and the local based land cover features and how they behave throughout the season. Tracking changes of the vegetation, water surfaces and soil moisture covers will help in explaining the precipitation and temperature given that these features are good proxy of the weather parameters. According to [5] increased sensitivity of these landcover features to precipitation and the use of remote sensing technologies can help to bridge the gap between the forecasts local farmers expect in their areas and what is provided by the weather stations.

The computation of the remote sensed land cover features is possible through ratioing which is a multi-satellite image manipulation enhancement technique that mathematically manipulates different spectral bands of the satellite image to generate new information that is not included in any single band. Ratioing gave birth to numerous indices that are now used to remotely explain the behaviour of vegetation, water and land resources to combat the impacts of climate change. For instance, The Normalized Difference Vegetation Index (NDVI) is a vegetation index that is used primarily to distinguish between healthy and stressed vegetation cover using red and near-infrared(NIR) bands [6]. The Macfeers' Normalized Difference Water Index (NDWI) and Normalized Difference Moisture Index (NDMI) on the other hand are primarily used to map water surfaces and moisture content on leaves and soil respectively [7].

There are number of projects conducted that are similar to the current study. The first project conducted by [8] was about the development of the IK mobile based application that quantifies the farmers' season predictions with the help of scientific data. The focus of the research was more on how the IK season predictions can be systematically explained with the help of scientific data. This is also part of the current study with respect to the IK's perspective in season behaviour prediction. The second study carried out by [9]was based on foreseeing of the season transition from one state to the other using the IK with vegetation data and temperature data. Few researches are presented where both IK and scientific knowledge are integrated: [10] proposed a way in which farmers' knowledge and scientific knowledge on tree species and soil fertility assessment can be combined to improve farmers' cropping decisions. [11] integrated both science and IK to diagnose types of hazards in diverse environmental and cultural settings. [12] presented the integration of IK, artificial intelligence and Information Technology to forecast an early onset of droughts in Western Africa to enhance adaptive capacities of the communities. [13] proposed a middleware that integrates the heterogeneous data sources with IK based on unified ontology for an accurate Internet of Things(IoT) Based Drought Early Warning System(DEWS). [14] proposed the participatory geographic information systems as an organizational platform for the integration of traditional and scientific knowledge in contemporary fire and fuel management. [15] emphasized on the integration of local knowledge and science to address the economic consequences of driftwood harvest in a changing climate. [16] proposed the Intelligent Agro-climate Decision Support tool for small scale farmers where they investigate the effectiveness of the integration of IK interpreted with fuzzy inference systems, mobile phone technology and smart sensors to enhance farmers decisions support systems. [17] proposed the use of Fuzzy Inference Rules System that integrates the local farmers' knowledge of soil and its fertility. Finally, [18] proposed an artificial intelligence technique called fuzzy cognitive maps to scientifically validate the farmers' indigenous knowledge on weather lore and the results have stimulated the opportunity for integrating consistent weather lore with modern systems of weather prediction and in enhancing applications offering decision support relying on weather effects.

3 Methodology

The action-based research methodology was adopted with the aim to identify the problem and come up with the solution to mitigate or lessen the effects of the problem. In this context, the researcher's aim was to evaluate the effectiveness of the IK in season behaviour predictions by small-scale farmers in Swayimane region in uMgungundlovu district KZN, South Africa. The 100 questionnaires were disseminated to the farmers and 88 of them were returned. From the analysis that was made, farmers rely heavily on the IK. Further, despite how important this knowledge system is to them, it continues to lose accuracy in predictions as most farmers claim failing to anticipate onset, magnitude, distribution and cessation of rains and temperature throughout the season. This is attributed to the change in climate that has caused shift of season patterns and erratic rains behaviour.

The framework presented below was followed to integrate these knowledge systems (IK and scientific knowledge) and present their integration via the mobile-based application (Figs. 1 and 2).

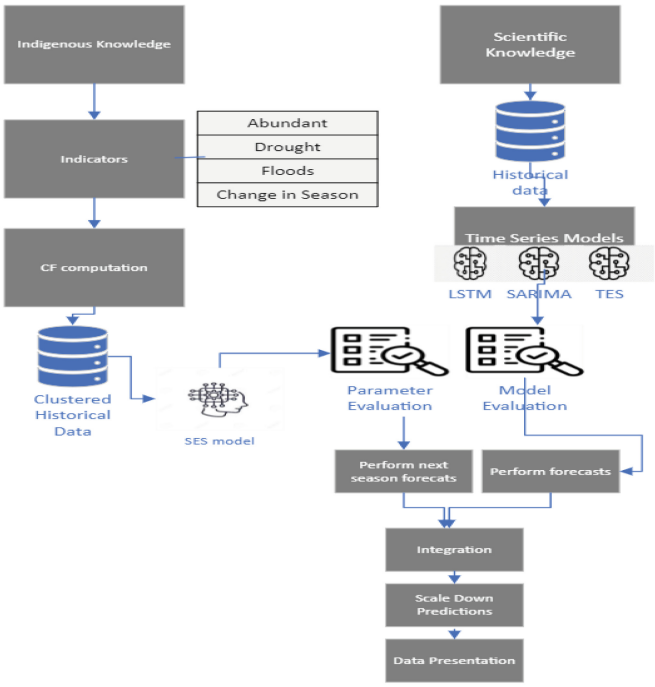

Fig. 1. Framework of the integrated Indigenous and Scientific Knowledge Systems

The indigenous and scientific approaches will be discussed separately and their integration later on.

3.1 Scientific Approach

3.1.1 Satellite Imagery and Weather Data Processing

An approximate of 400 free weekly Sentinel 2 satellite images between 2013 and 2021 were collected from USGS and Scihub Copernicus website. The multispectral bands of these images were ranging between band 1 (Coastal aerosol) and band 12 (short wave infrared (SWIR)). The temporal resolution of the satellite images was roughly 7 days with spatial resolution of 10 m in the visible part of the spectral bands and the NIR band. The Quantum Geographic Information System (QGIS) was used to process the satellite images. The Semi-Automatic Classification (SCP) and Base map plugins were installed into QGIS to aid in pre-processing of the images. Among tasks that were performed using these tools is band clipping, which is the process where set of pixels that represent the region of interest are extracted from the image with coarse spatial resolution. The clipped images were further calibrated by reducing the atmospheric effects. The images with high cloudy pixels were calibrated through cloud masking and mosaic where cloudy pixels in one image are replaced by noncloudy pixels in another image that is having a close temporal resolution as the one that is currently being calibrated.

Below is the satellite image representing the surface reflectance after applying the atmospheric corrections.

Fig. 2. Swayimane's satellite image

For every image, the raster calculator was used to compute the healthy vegetation cover represented by NDVI values of 0.3 and above using Red and NIR bands. The water surfaces cover and soil moisture cover were also computed using Macfeeter's NDWI and NDMI respectively. The Macfeeter's NDWI has the reflectance values ranging between –1 and + 1 where positive values denote waterbodies and 0 or negative values denote non-water bodies features such as soil and vegetation. The NDMI on the other side has values ranging between –1 to + 1, where the lowest values indicate low vegetation or soil water content, and the highest ones correspond to high water content. These pixel cover values were recorded together with the images' acquisition dates. The daily rain and average temperature data ranging between 1993 to 2021 were extracted from the weather station close to the region of interest. Given that there were less number of missed data, this was simply corrected by imputing the blank cells with the average data of a given column.

3.1.2 Time Series Analysis and Forecasting

The time series data (rain, average temp, waterbodies, vegetation greenness and soil moisture cover values) were loaded from the computer's directory and were manipulated using Anaconda IDE.

Three well-known time series models namely Holt Winter's model, SARIMA and LSTM were trained and evaluated using the historical data.

The time series data were split into training and validation sets as shown below (Table 1).

Table 1. Training and validation data table

Data	Training set	Test
Weather data	1993–2018	2019–2021
Satellite imagery data	2013–2018	2019–2021

The training set was used to train the model while the testing set was used to evaluate the performance of the model in predicting future values.

3.1.2.1. Background and Implementation of Holt Winter's Model

This model is also known as Tripple Exponential Smoothing (TES) is an extension of Double Exponential Smoothing and Single Exponential Smoothing models. It was assigned the name TES given its capacity to handle series with level, trend and seasonality components. Before forecasting using this model, the behaviour of the trend and seasonality must first be identified. Both trend and seasonality can either be additive or multiplicative in nature. An additive trend is observed when rate of change in the data is constant over time and can be represented by a linear graph. The multiplicative trend on the other side is observed when the rate of change in data is either increasing or decreasing exponentially over time and it can be represented by a non-linear graph. Below is the Holt Winter's additive forecast that is composed of an additive trend and multiplicative seasonality.

$$F_t = L_t + T_t + S_t = \left[\alpha \frac{y_t}{s_{t-m}} + (1-\alpha)L_{t-1} + T_{t-1}\right] + \left[\beta(L_t - L_{t-1}+) + (1-\beta)T_{t-1}\right] + \left[\alpha \frac{y_t}{L_t} + (1-\gamma)S_{t-m}\right] \quad (1)$$

The model was imported from the Stats Model library. To fit the model, the training set, the trend and seasonal type which can both either be additive or multiplicative were passed as parameters to the model. The trend and seasonality types were applied interchangeably for each time series data and the Akaike Information Criterion(AIC) metric was computed.

For each training set, the best model was selected based on the minimum AIC value. The best models were used to forecast the number of units or rows equivalent to the ones of the test sets. Both the models' forecast and test sets are plotted for each time series

data where the orange signal represents the predictions and the blue represents the actual data (Figs. 3, 4, 5, 6, and 7).

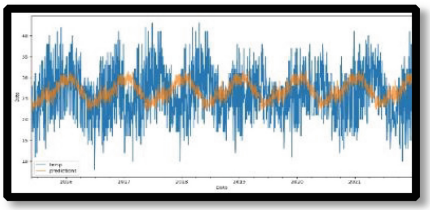

Fig. 3. Temperature data and model's predictions

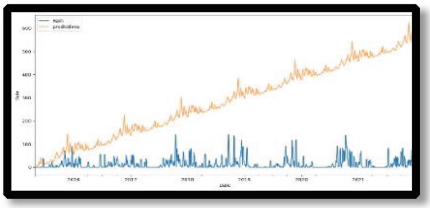

Fig. 4. Rain data and model's predictions

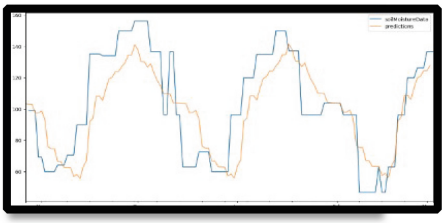

Fig. 5. Soil data and model's predictions

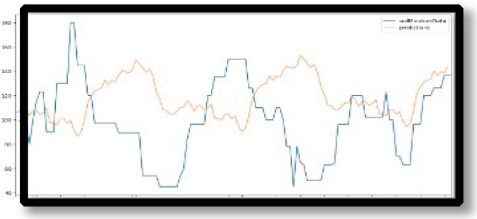

Fig. 6. Vegetation data and model's predictions

The Root Mean Squared Error (RMSE) was computed for each time series data.

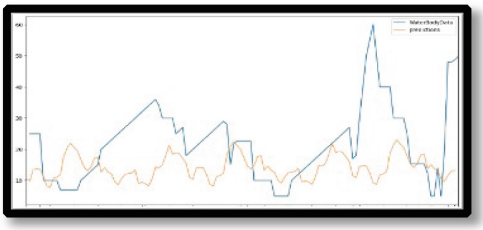

Fig. 7. Waterbodies data and model's predictions

3.1.2.2. Background and Implementation of SARIMA Model

SARIMA is an Auto Regressive Integrate Moving Average (ARIMA) model with seasonality. An ARIMA model is composed of two distinct models Auto Regressive (AR) and Moving Average (MA) that explain the behaviour of the series from two different angles. The integration part denoted as I, reveals how many times the differencing was performed to stationarize the series. The stationary series is the one with no increasing or decreasing trend, has constant mean and variance over time. The SARIMA model is presented with 7 parameters as shown below:

$$ARIMA(\mathrm{p},\ \mathrm{d},\ \mathrm{q})(\mathrm{P},\ \mathrm{D},\ \mathrm{Q})_m$$

where (m) is the seasonal factor (number of periods it takes for seasonality to repeat).

P-number of seasonal AR terms

D -number of seasonal differences

Q-number of seasonal MA terms

The general equation of the model is presented as

$$ARIMA(\mathrm{p,\ d,\ q})(\mathrm{P,\ D,\ Q})_m = \mu + \sum_{i=1}^{p} y_i f_{t-i} + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} + \sum_{i=1}^{P} \phi_j f_{t-si} + \sum_{i}^{q} = 1\mu_j \epsilon_{t-sj} + \epsilon_t \qquad (2)$$

To perform forecasts using the model, the SARIMA model identified as SARIMAX was imported together with Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF) models from Stats Model library. The daily rain and average temperature were aggregated into weeks using the Mean function. This was performed to minimize the model fitting process which is considered resource heavy computation.

The first step was to evaluate the stationarity of the time series data to determine if transformation has to be done or not. The Augmented Dickey Fuller (ADF) test was performed to test the stationarity of each time series data. To conclude that the series is stationery, the null hypothesis must be rejected, and alternative hypothesis must be accepted. The null hypothesis states that the series in question is not stationery by nature and the alternative hypothesis states that the series is stationery by nature. The P-statistic value is the one that is primarily used to evaluate the stationarity of the series. If it is less than 5%, the series is considered stationery, otherwise it is considered not stationery and as a result, the transformation techniques will be implemented to stationeries the series. The data were evaluated and were concluded to be stationary. The next step was to determine the order movement of the SARIMA model for each time series.

The autocorrelation plots (ACF and PACF) were created for all the time series data. However, the accurate estimation of the lags for SARIMA through visual observation

was complicated due to massive number of parameters that need to be estimated. To get around the selection of optimal model parameters, the function that uses nested loops to test wide range of candidate values representing model parameters was used. For each iteration, model fitting is performed, and the AIC matrix is computed.

The model for each time series that showed the minimum error rate was selected. Each model was used to forecast the number of units or weeks equivalent to test sets as illustrated below where the blue line represents the actual data in test sample and the orange line represents the model predictions (Figs. 8, 9, 10, 11 and 12).

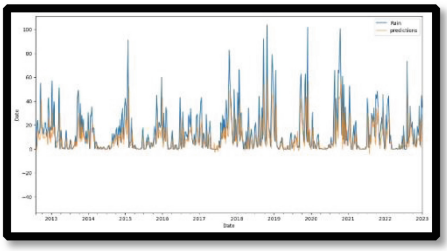

Fig. 8. Rain data and model's predictions

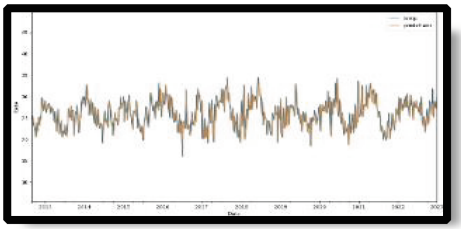

Fig. 9. Temperature data and the model's predictions

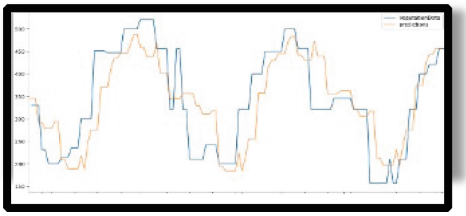

Fig. 10. Vegetation data and model's predictions

The RMSE was computed for each time series data.

3.1.2.3. Background and Implementation of LSTM RNN Model

The traditional Recurrent Neural Network (RNN) has been widely used for processing sequential data. These networks are fundamentally different from "vanilla" Neural Networks (NN) also known as feedforward NN that go for input and output, in one direction

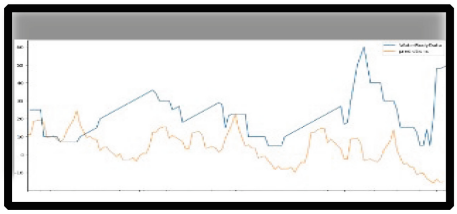

Fig. 11. Waterbodies data and model's predictions

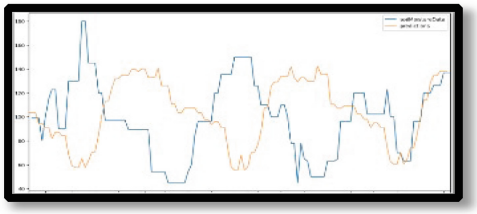

Fig. 12. Soil Moisture data and model's predictions

and are unable to maintain information about previous events. RNN are sequence-based models which are able to establish temporal correlation between previous information and current information and as such the decision the RNN made at time step t–1 can influence the decision it will reach at the current time step(t) [23]. These RNN architectures have loops in them that allow information to persist over time. The network computes the internal state update denoted as h_t from one-time step to the next as illustrated below.

$$h_t = f_w\left(W_{hh}^T h_{t-1} + W_{xh}^T X_t + b_t\right) \tag{3}$$

Where W_{hh}^T and W_{xh}^T are weights, h_{t-1} is the previous state, X_t is the current input and b_t is the bias, f_w is an activation function that in most cases can be hyperbolic tangent(tanh) or logistic sigmoid function (σ) presented as:

$$hyperbolic : f(x) = \tanh(x) = \frac{2}{1 + e^{-2x}} - 1 \tag{4}$$

$$tangent : f(x) = \sigma(x) = \frac{1}{1 + e^{-x}} \tag{5}$$

The hyperbolic tangent squashes the results (x) into –1 to 1, while the sigmoid function squashes the results(x) into 0 and 1.

The output y_t is the product of the current internal state and some weight vector.

$$y_t = W_{hy}^T h_t \tag{6}$$

The RNN uses backpropagation through time (BPTT) to train itself, this is where the model's output at time t, is compared to the ground output at time t and weights

are adjusted based on the magnitude of the error to minimize it to its optimal. This is illustrated below:

$$e_t = (y_t - \hat{y}_t) \tag{7}$$

$$\Delta w_t = n\frac{de_t}{dw_t} \tag{8}$$

$$w_t = w_t + \Delta w_t \tag{9}$$

During the process of BPTT, the model suffers from gradient vanish and exploding problems due to number of multiplications that occurs resulting into model's inability to learn long term dependencies [20]. This is where LSTM RNN come into place by eliminating the traditional RNN problems. The LSTM consists of internal memory cell state and gates that solves vanishing gradient and exploding problems. The previous hidden state (h_{t-1}) and the current input X_t are fed to the sigmoid activation function of the forget gate. The gate will decide what information to keep or to throw away. Values close to 0 will be thrown while those close to 1 will be kept.

$$f_t = \sigma\left(W_{fx}X_t + W_{fh}h_{t-1} + b_f\right) \tag{10}$$

The input gate receives the previous hidden state and the current input and uses the sigmoid function to decide what information to update

$$i_t = \sigma(W_{ix}X_t + W_{ih}h_{t-1} + b_i) \tag{11}$$

To compute the candidate cell state, the previous hidden state and the current input are fed to the tangent function which will squash the values between –1 and 1.

$$g_t = tanh\left(W_{gx}X_t + W_{gh}h_{t-1} + b_g\right) \tag{12}$$

The current cell state will be computed as the (product of candidate cell state and the input gate's output) added to the (product of the previous cell state and the forget gate's output) as shown below:

$$c_t = g_t \cdot i_t + c_{t-1} \cdot f_t \tag{13}$$

The output gate is computed using the previous hidden state and the current input

$$o_t = \sigma(W_{ox}X_t + W_{oh}h_{t-1} + b_o) \tag{14}$$

The current hidden state is computed as the product of the output from the output gate and the cell state that is first squashed to the tangent function.

$$h_t = tanh(c_t) \cdot o_t \tag{15}$$

The current hidden state and cell state are then carried to the next time step.

The key reason for using LSTM is that the cell state sum activities over time which can overcome vanishing gradient and better capture long term dependencies of time series [20].

To perform forecasting, the model was loaded from the Stats Model library. Given that the LSTM is sensitive to the scale of the data, the MinMax scaler function was applied to transform the time series data into a range of 0 and 1. Smallest values were close to 0, while largest values were close to 1. To prepare data for model fitting, train and test sets were further processed by splitting each sample into dependent and independent features based on timestamp that will be provided as an argument to the function. To predict y(t) considered as dependent variable, the model will first learn from the following regressed independent features y(t–1), y(t–2), y(t–3)……y(t–n). The number of independent features depends on the time stamp provided. The LSTM model will be expected to learn the independent features and predict the dependent feature for both train and test samples of every time series.

The stacked LSTM model with three hidden layers and output layer was created.

The model was compiled with RMSE loss function and 'Adam' optimizer. The model fit was performed for every time series' generated samples. The model was used to predict the dependent features in both train and test sets of every time series. To evaluate the model's performance, the data were rescaled back to original form. Below is the model's predictions plots for each time series where the blue signal represents the actual data, the orange signal represents the predictions made on training set and the green signal represents the predictions made on the test sets (Figs. 13, 14, 15, 16 and 17).

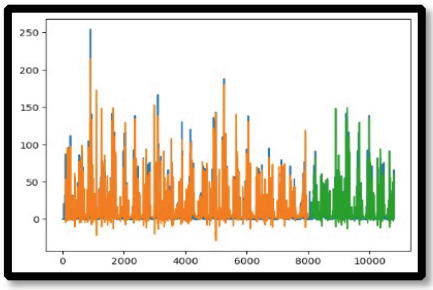

Fig. 13. Rain data predictions using LSTM

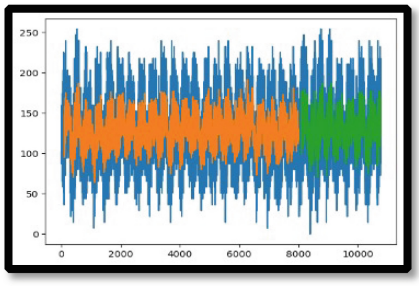

Fig. 14. Temperature data predictions using LSTM

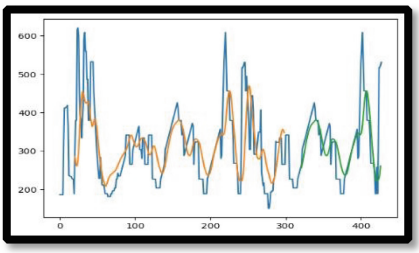

Fig. 15. Soil moisture predictions using LSTM

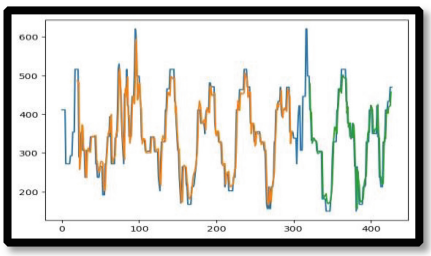

Fig. 16. Water surface data predictions using LSTM

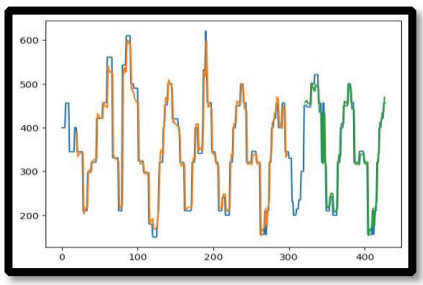

Fig. 17. Vegetation data predictions using LSTM

3.1.2.4. Conclusion on Models that Were Selected

The SARIMA (1,0,1) (1,0,2) model was identified as the best model to forecast weather data with average accuracy of 88.2%. The LSTM model was identified as the best model for satellite imagery data with average accuracy of 83.6%. These models were used to forecast the next 48 weeks. The data was stored in cloud database for further manipulation and integration.

3.2 Manipulation of the Historical Data to Complement the IK Predictions

The historic weather and satellite imagery data of only warm rainy season(Sept.–Jan.) and cold rainy season (March to May) were extracted from the main historical database. The

data were analyzed and were grouped based on how they behave throughout the season. For instance, some records were clustered into abundant rains, less rains and excessive rains periods by observing and analyzing the change in their series values. Further, the data were structured such that set of records can be extracted and manipulated based on the current week the mobile application is used. For instance, if the system is used in week 12, the historic data from the next week (week 13) will be extracted and further manipulated.

The historical data were stored in cloud database for later manipulation by the system.

3.3 IK System Development

To quantify the collection and processing of IK indicators' observations, the following class diagram was created (Fig. 18).

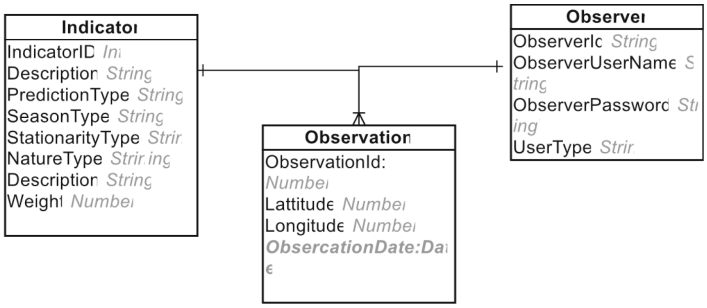

Fig. 18. Class diagram showing relationship between indicator and observer

From the diagram presented above, three entities were created, where the observation entity acts as an intermediary or a bridge between the observer and the indicator. This is due to the fact that an observer can observe one or more indicators and one indicator can be observed by one or more observers. The IK indicator object is comprised of the following properties:

Indicatorid: Uniquely identify an id,

IndicatorDesc: Describes the indicator

NatureType: describes the nature of an indicator which can either be meteorological, astronomical and environmental.

StationaryOrMotionType: Some indicators are in-motion or variable while others are fixed. For instance, birds or animal indicators will be different to trees indicators with respect to stationarity. The one that are in motion do not guarantee that their next observation will be on the same location at different time stamps in contrast to the ones that are fixed. This type of annotation will help in computation of time-stamp and location distance between two or more observations of the same indicator.

PredictionType: Indicators can predict either abundant rains, drought, floods or season changes.

SeasonType: Indicators are further clustered into either warm or cold season indicators.

Weight: The weight field measure the impact/accuracy the indicator has towards its prediction type.

The algorithm was developed to evaluate the registration of the IK indicators' observations. The aim of the algorithm was to ensure that indicators are registered strictly in the area of interest. If the indicator observation is registered for the first time, it will be accepted. However, if there are other observations already made of that indicator, the time stamp and location distance will be computed with the aim to avoid duplicate entries. The time stamp and location distance thresholds can be adjusted by an administrator of the mobile application. The role of an administrator is also to manage the indicators and their observations by performing the CRUD (create, read, update and delete) roles. The computation is also with respect to the nature and stationarity of the indicators. For instance, the location distance will apply between the two observations of a non-fixed environmental indicator. The timestamp distance will be sufficient for both the meteorological and astronomical indicators' observation given that they can be witnessed by everyone in the area. The Haversine equation presented below was used to compute the shortest distance between two points on an earth surface using latitude and longitude parameters.

$$d = 2r\ \sin^{-1}\left(\sqrt{\sin^2\left(\frac{x_2 - x_1}{2}\right)} + \cos(x_1)\cos(x_2)\sin^2\left(\frac{y_2 - y_1}{2}\right)\right) \tag{16}$$

where r represents the radius $x1$ represents latitude of point A

$x2$ represents latitude of point B

$y1$ represents longitude of point A

$y2$ represents longitude of point B

The time stamp distance is computed as a days' difference between two dates.

3.3.1 Certainty Level Computation

The certainty level (CL) is about the sureness of the predicted season behaviour with respect to the observation of the indicators. It is computed using the qualifying indicators' observations taking into account number of times they were observed and the number of the observers. For instance, the indicator observation will qualify if it was observed two or more times by two or more observers. The threshold for the frequency values will be set by an administrator of the application.

The weight of the qualifying indicators will be summed to determine the season prediction type (abundant, floods or drought) that has a maximum value. The CL will be computed with respect to the winner prediction type. The CL is computed as a total weight of observed indicators divided by the sum of weights of all indicators that belong to the same prediction type and season type. If the CL is above a given threshold set by an administrator, IK will be integrated with the scientific knowledge. If it will be integrated, the categorized historical data will be extracted with respect to the prediction type and timestamp of when the prediction was made. For instance, if the predictions of abundant rains were made early in the warm rainy season, the historical data of the periods when the abundant rains were observed early during the warm rainy season will be extracted by the system and these data will be fed to the Single Exponential Smoothing (SES)

model that will predict the next season data values. The selection of this model is due to the fact that it was easy to implement and its order movement parameter can be easily optimized by the system. Unlike other sophisticated models such as Double Exponential smoothing or Tripple Exponential Smoothing, this model is light in computational time given that only one parameter (its learning rate(alpha)) has to be optimized. The SES model is presented below

$$f_{t+1} = \alpha F_t + \sum_{i=1}^{n} a(1-a)^i F_{t-i} \tag{17}$$

The predictions that will be made for the IK forecasts will be integrated with the predictions already generated for the scientific perspective.

3.3.2 Downscaling of the Season Predictions

At this stage the predictions are still in pixel area cover values for remote sensing data, in millimeters (mm) for rain data and in degree Celsius for temperature data. Hence, they are not presentable to the farmers. To scale down the predictions, number of algorithms were developed to extract the farmers' historic data and perform mathematical computations to scale the predictions to the range of 1 to 6, where 1 represents no change, 2- very less change, 3- less change, 4- normal change, 5-above normal change and 6 represents extreme change. For instance, to scale down the satellite imagery data, the average pixel area cover values of periods when farmers have experienced floods, abundant rains and drought will first be computed. The researcher's expectation is that during excessive rains periods, water surfaces and soil moisture cover average pixel values will be high than during abundant and drought seasons, however, that is what the algorithms will figure out.

To scale the temperature values, the average of temperature during warm rainy seasons and cold rainy seasons will be computed.

Below is the screenshot on how rainy season prediction will be presented to the user by showing movement of temperature(red series), water series (navy blue series) that represents the integration of rain, waterbodies and soil moisture series, and vegetation greenness (green series) data (Fig. 19).

3.3.3 System Demonstration

The screenshot below shows the menu screen that the farmer will be displayed with (Figs. 20 and 21).

The indicator observation will be registered as shown below where the user will select the season type which can either be warm or cold, the indicator's prediction type and the indicator that they have observed.

When the save button is clicked, the camera option will be opened for the environmental type of the indicators as illustrated below on Fig. 22.

The administrator will be able to conclude on the validity of the indicator observation with respect to the image and the indicator's description as illustrated on Fig. 23. Lastly, the system will use the observed indicators to compute the certainty level of the predicted season behaviour.

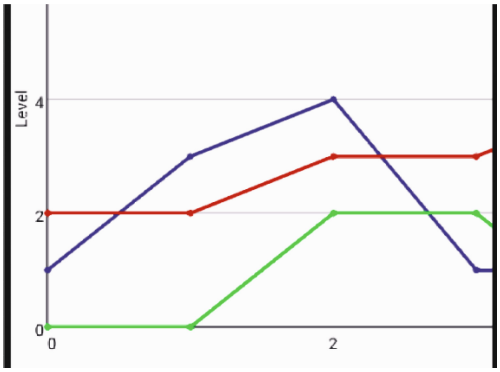

Fig. 19. Screenshot of the season prediction

Fig. 20. Screenshot of the menu screen

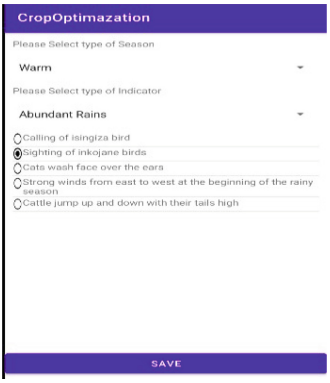

Fig. 21. Screenshot showing the registration of the indicator observation

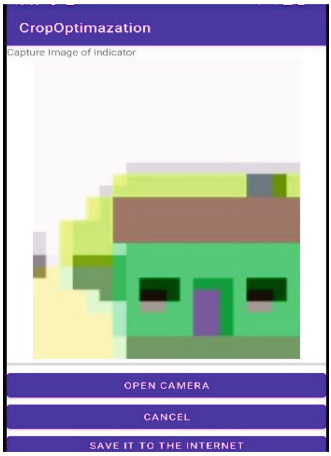

Fig. 22. Image capturing of environmental indicator

Fig. 23. Screenshot showing list of observed environmental indicators' images

4 Conclusion

This paper presented the integration of IK and SF approaches towards season behavior predictions. The development of this tool was motivated by the impacts of climate change and marginalization of the IK system, despite its contribution towards tackling the impacts of climate change. Despite the value brought by IK system, it is being directly and indirectly sabotaged for instance, through pollution and other human related activities where birds, animals, trees and insects that farmers use to build on the indigenous

knowledge system are being sabotaged. Further, the indigenous knowledge continues to be overlooked by an upcoming generations who tend to adopt to the western life at an enormous rate while sideling their cultural life [19].

This paper has emphasized that the IK in season behavior predictions can be systematically collected and processed and also be integrated with the scientific forecasts to come up with more robust season predictions. Further, this paper has also emphasize the importance of the satellite imagery and ML technologies in mapping and predictions of the changes in land cover features to bridge the gap between the lack and sparse distribution of weather stations and what is expected by the farmers in local areas remote from the weather stations. The next approach will be to submit the prototype to the farmers in Swayimane region for evaluation. This process will help in identifying the strengths and weaknesses of the proposed application. Further, finding ways to improve it.

References

1. Masinde, M., Bagula, A., Muthama, N.J.: The role of ICTs in downscaling and up-scaling integrated weather forecasts for farmers in sub-Saharan Africa. In: Proceedings of the Fifth International Conference on Information and Communication Technologies and Development, pp. 122–129, March 2012
2. Mutekwa, V.T.: Climate change impacts and adaptation in the agricultural sector: the case of smallholder farmers in Zimbabwe. J. Sustain. Dev. Africa **11**(2), 237–256 (2009)
3. Meza, F.J., Hansen, J.W., Osgood, D.: Economic value of seasonal climate forecasts for agriculture: review of ex-ante assessments and recommendations for future research. J. Appl. Meteorol. Climatol. **47**(5), 1269–1286 (2008)
4. Fanadzo, M., Ncube, B.: Challenges and opportunities for revitalising smallholder irrigation schemes in South Africa. Water SA **44**(3), 436–447 (2018)
5. Richard, Y., Poccard, I.J.I.J.O.R.S.: A statistical study of NDVI sensitivity to seasonal and interannual rainfall variations in Southern Africa. Int. J. Remote Sens. **19**(15), 2907–2920 (1998)
6. Yengoh, G.T., Dent, D., Olsson, L., Tengberg, A.E., Tucker III, C.J.: Use of the Normalized Difference Vegetation Index (NDVI) to Assess Land Degradation at Multiple Scales: Current Status, Future Trends, and Practical Considerations. Springer (2015)
7. Al-Hakeem, R., Al-Kubaisi, Q.Y.: Detection of physical and chemical parameters using water indices (NDWI, MNDWI, NDMI, WRI, and AWEI) for Al-Abbasia River in Al-Najaf Al-Ashraf governorate using remote sensing and Geographic Information System (GIS) techniques. IRAQI J. Phys. **20**(4), 10–17 (2022)
8. Nyetanyane, J.: Indigenous knowledge mobile based application that quantifies farmers' season predictions with the help of scientific knowledge. In: Masinde, M., Bagula, A. (eds.) Emerging Technologies for Developing Countries. AFRICATEK 2022. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol. 503. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-35883-8_13
9. Nyetanyane, J., Masinde, M.: Foresee transition to agricultural season by integrating indigenous knowledge, satellite imagery, weather data and ARIMA family models to enable good crop establishment by small-scale farmers in Swayamani Region, KwaZulu-Natal, South Africa. In: 2020 2nd International Multidisciplinary Information Technology and Engineering Conference (IMITEC), pp. 1–7. IEEE, November 2020
10. Bukomeko, H., Jassogne, L., Tumwebaze, S.B., Eilu, G., Vaast, P.: Integrating local knowledge with tree diversity analyses to optimize on-farm tree species composition for ecosystem service delivery in coffee agroforestry systems of Uganda. Agrofor. Syst. **93**, 755–770 (2019)

11. Pareek, A., Trivedi, P.C.: Cultural values and indigenous knowledge of climate change and disaster prediction in Rajasthan, India (2011)
12. Masinde, E.M.: Bridge between African Indigenous knowledge and modern science on drought prediction (Doctoral dissertation, UNIVERSITY OF CAPE TOWN) 2012
13. Akanbi, A.K., Masinde, M.: Towards semantic integration of heterogeneous sensor data with indigenous knowledge for drought forecasting. In: Proceedings of the Doctoral Symposium of the 16th International Middleware Conference, p. 2. ACM, December 2015
14. McBride, B.B., et al.: Participatory geographic information systems as an organizational platform for the integration of traditional and scientific knowledge in contemporary fire and fuels management. J. Forest. **115**(1), 43–50 (2016)
15. Jones, C., Kielland, K., Hinzman, L., Schneider, W.: Integrating local knowledge and science: economic consequences of driftwood harvest in a changing climate. Ecol. Soc. **20**(1) (2015)
16. Thothela, N.P., Markus, E.D., Masinde, M., Abu-Mahfouz, A.M.: A survey of intelligent agro-climate decision support tool for small-scale farmers: an integration of indigenous knowledge, mobile phone technology and smart sensors. In: Fong, S., Dey, N., Joshi, A. (eds.) ICT Analysis and Applications. Lecture Notes in Networks and Systems, vol. 154. Springer, Singapore (2021). https://doi.org/10.1007/978-981-15-8354-4_71
17. Wenisch, S.M., Uma, G.V., Ramachandran, A.: Fuzzy inference system for an integrated knowledge management system. Int. J. Comput. Appl. **975**, 8887 (2010)
18. Mwagha, S.M., Masinde, M.: Scientific verification of weather lore for drought forecasting–the role of fuzzy cognitive mapping. In: Proceedings of the IST-Africa 2015 Conference, Lilongwe, Malawi, pp. 6–8, May 2015
19. Nyetanyane, J.: Indigenous knowledge crop health recommendation expert system for the upcoming generations. In: 2023 IST-Africa Conference (IST-Africa), pp. 1–8. IEEE, May 2023
20. Lindemann, B., Müller, T., Vietz, H., Jazdi, N., Weyrich, M.: A survey on long short-term memory networks for time series prediction. Proc. CIRP **99**, 650–655 (2021)
21. Ayanlade, A., Radeny, M., Morton, J.F.: Comparing smallholder farmers' perception of climate change with meteorological data: a case study from southwestern Nigeria. Weather Clim. Extrem. **15**, 24–33 (2017)
22. Dinku, T., Hailemariam, K., Maidment, R., Tarnavsky, E., Connor, S.: Combined use of satellite estimates and rain gauge observations to generate high-quality historical rainfall time series over Ethiopia. Int. J. Climatol. **34**(7), 2489–2504 (2014)
23. Kong, W., Dong, Z.Y., Jia, Y., Hill, D.J., Xu, Y., Zhang, Y.: Short-term residential load forecasting based on LSTM recurrent neural network. IEEE Trans. Smart Grid **10**(1), 841–851 (2017)