



Integration of Indigenous Knowledge, Climate Data, Satellite Imagery and Machine Learning to Optimize Cropping Decisions by Small-Scale Farmers. a Case Study of uMgungundlovu District Municipality, South Africa

John Nyetanyane^(✉) and Muthoni Masinde

Department of Information Technology, Central University of Technology, Free State,
Bloemfontein, South Africa

{jnyetanyane, emasinde}@cut.ac.za

Abstract. Despite being the backbone of food security in most African countries, small-scale farmers are overlooked during the implementation of many of developmental projects. These farmers are financially incapable of equipping themselves with irrigation systems and other agricultural technologies that can assist in improving their farm yields. One of the challenges facing them is that they rely heavily on rain-fed agriculture which makes them extremely vulnerable in the face of climate change. They have continued to consult their indigenous knowledge systems to predict the onset of rains and in making critical decisions such as when to prepare land for crop cultivation. Evidence shows that this knowledge is no longer as precise as it used to be – among other reasons, this is due to the effects of climate change and deforestation. The second problem is that the only sources of weather information these farmers have (e.g. the media) are general and not scaled down to the specific locations where they reside. On the other hand, most of the small-scale farmers are educationally and technologically semi-literate and are financially crawling when it comes to adoption of the likes of sensors and other technologies that could help in predicting and monitoring crop health. There is however opportunity in that most of these farmers are now using android phones. In this research, a model that utilizes the indigenous knowledge, climate data and vegetation index to foresee the onset of the favourable weather season for crop cultivation, crop monitoring and crop health prediction is proposed.

Keywords: Indigenous knowledge · Satellite imagery · Simple moving average · Weighted moving average · Linear exponential smoothing · Double exponential smoothing · Average ndvi

1 Introduction

Africa is considered to have less of crop lands in comparison with other continents such as America and Asia [11]. Increase in number as well as severity of droughts, coupled

with population growth, cause serious implication on vegetation lands in Africa [21]. Remote sensing technology such as satellite imagery to assess and monitor the vegetation health in Africa is not yet widely adopted. African farmers, majority of whom are small scale farmers, continue to rely on indigenous knowledge as opposed to conventional science and technology [23]. Factors that demotivate the use of technology such as satellite images, comes at high cost and need for massive processing before being readily available. Moreover, the requisition of complex and sophisticated software applications to manipulate the satellite images are adding on the complications to adopt to technology. Yes, we do have the satellite images that can be accessed at any time; however, these free satellite images have less spatial resolution of up to 100 to 1000 m. Among the well-known satellites, the National Aromatics and Space Administration (NASA)'s satellite called MODIS satellite provides images with the minimum spatial resolution of 250 m. Therefore, if one pixel of an image is covering a distance of 250 m by 250 m, it will be difficult to detect small objects from the ground. Even if Africa's small-scale farmers were to perceive the information from the satellite images, it will serve minimum or no interests on them because of the formats and language used to represent them. Moreover, in most cases it takes time to receive the updated versions of free-satellite images, making it difficult to track the changes in near-real-time.

Illiteracy and poverty level amongst most African farmers make it difficult for agricultural information dissemination and requisition of technological resources to aid on their farming growth [3]. However, farmers do need technology that can assist in optimising their cropping decisions, for instance, to monitor their crop growth. Even though, indigenous knowledge has been effective in the yesteryears, its usefulness has slowly been eroded, besides, it takes a human labour to physically and periodically monitor a huge portion of land. There are several attributes that can be detected using the machine as opposed to IK. These include Normalized Difference Vegetation Index (NDVI), level of gases in the atmosphere, precipitation, among others. Early detection of crop problems can assist in reducing financial losses, increase food security and reaching targeted yielding and profit margins [9]. In this paper, we argue that integrating the farmer's knowledge with mainstream scientific approach can improve the accuracy of crop yield prediction.

2 Literature Review

Different crops are cultivated in different seasons mainly during summer (Nov-Dec-Jan) and winter (May-June-July). The summer crops are harvested in autumn while the winter crops are harvested in spring. Summer crops need more rain and warm temperature during their life span while the winter crops need minimum of rain and cold temperature. The well-known summer crops are maize, sorghum, beans, peanuts, sunflower seed and so forth. Whereas wheat, peas, spinach, onions, carrots, lentils and so forth are planted during winter.

The dispersal of vegetation in African land is gradually becoming extinct [13]. According to [21] most studies have proved that Africa has experienced more frequent, intense and widespread of droughts from the last 50 years. South Africa depend highly on the availability of rain as it stands out as one of the most water-scarce countries [13].

According to [7], statistics showed that there was a decrease of agricultural households in South Africa from 19.9% in 2011 to 13.8% in 2016 due to severe drought in 2014 to 2015. Among other factors that aggravate the limitation of vegetation is escalating population and degradation of land and environment due to development of industries, mining sectors, houses and other premises [21]. Therefore, land is a limited resource and hence the present efforts should mainly rely upon increasing the productivity of the existing land [6].

Vegetation is the source of food, it must be well planned and monitored for sustainability. Contribution of land, vegetation and agriculture is the back bone of every country [28]. [36, 37] explained that crop production and yielding must be planned and accurately predicted through the use of technology. They further accentuate that being able to optimize cropping decisions and predict crop yields accurately will allow the government to plan production, distribution and food consumption more effectively, combat food insecurity, prepare for shortages and supply shocks well in advance. Additionally, [43] explain that efficient and effective vegetation monitoring can help preventing famine and support humanitarian efforts in light of climate change and droughts. According to [29], being able to monitor the crop health and predict the crop yield at a growing season, marketed prices can be forecasted, import and export can be planned, the socio-economic impact of crop loss can be minimized, and humanitarian food assistance can be planned.

Although there are some interesting aspects with vegetation monitoring, prediction of crop yield before harvest is a challenging aspect in the field of machine learning and artificial intelligence [44]. Developed countries such as USA and Japan are equipped with approaches such as comprehensive surveys of weather parameters and land cover types that are publicly available and greatly facilitate the crop yielding task [43]. However, the aforementioned technologies are not available in developing countries such as South Africa and yet reliable yield predictions are most needed [8]. As a solution, [8] introduce the use of remote sensing data such as free satellite images that is globally available, of good quality and relatively inexpensive. The use of remote sensing combined with other form of technologies can be quite helpful to the small-scale farmers especially those that do not enjoy the comfortability of having the irrigation systems and rely only on rain [17]. Most of the local farmers are making living from the rain-fed agriculture and at the onset of rainy season, farmers make crucial decisions about what, when and where to plant [13, 17, 34].

2.1 Remote Sensing Technology

Remote sensing is a technology of acquiring information above the earth surface without coming into contact with it [20]. Remote sensing technology has been used as an effective tool to track and monitor the crop vigour, density, health, development and growth [5]. Further, this technology can be used in computational sustainability such as species distribution modelling, poverty mapping, climate modelling and natural disaster prevention [43].

One of the main examples of remote sensing technology is satellite imagery system. It is a powerful tool equipped with powerful sensors and cameras used to collect images

of the world surface [16]. These sensors can record the electromagnetic radiation emitted from the sun or any form of a solar and reflected from the ground and or objects on the ground [15].

2.2 Satellite Image Segmentation or Classification

Collected satellite images are also known as multispectral satellite images contains abundance of information visible and non-visible to the human eye that could be correlated with economic features, vegetation growth and agricultural outcomes [16, 43]. Because satellite images have high spatial graphics, data is high-dimensional and unstructured, deep learning and machine learning techniques come into operational when having to validate the image in to certain places of interest [16, 43]. Image segmentation also known as image classification is a process of classifying an image in to different groups or regions [9, 16, 43]. [40] define image classification as a part of computer vision that will help to classify pixels into finite set of individual classes based on their data values. It is a technique that will help to extract only useful data from a given image. Image classification is considered as a vital part of the remote sensing, image analysis and pattern recognition [38]. Object recognition and classification to extract meaningful insights from the large-scale image is considered as a challenging task by machines [16].

2.3 Vegetation Indices

When the features of interests such as crop lands have been extracted from the given satellite image using the machine vision algorithms, different types of indices specifically vegetation indices can be calculated. Among the existing indices, NDVI is a well-known index used to measure and assess the vegetation greenness, vegetation growth and vegetation health. According to [31], NDVI is a ratio based on the spectral reflectance from vegetation, fourthly it is from combined operations between red and NIR band regions of the multispectral light which is expressed mathematically as:

$$ndvi = \frac{I_{nir} - I_{red}}{I_{nir} + I_{red}} \quad (1)$$

NDVI values range between the scale of -1 and $+1$, healthy vegetation is presented by high NDVI values between 0.5 and <1 . Non-vegetation surfaces such as soil and water bodies yield negative values considered as low NDVI values [2]. The healthiness of plants also depends on the health of the soil. A well-known index used to measure the soil health is Soil Adjusted Vegetation Index (SAVI). The formula is presented below whereby L represents the soil brightness correction that is traditionally 0 in the highly vegetation areas and 1 in a low vegetation areas

$$savi = \frac{I_{nir} - I_{red}}{(I_{nir} + I_{red} + L)} * (1 + L) \quad (2)$$

The historic and current vegetation data collected from the remote sensing technology can be fed to a time series models to predict the crop yield or future vegetation so to enlighten the local farmers, business or economical entities, agricultural, political developers and so forth with the future [14, 39, 42].

2.4 Machine Learning (Time Series Analysis)

Forecasting is the process of feeding the machine a historical data and allow the machine to learn the pattern or the trend from the data so that it can be able to predict new or future values. Most of the time series prediction approaches are based on the analysis of historical data and assume that past and current data patterns can be used to forecast future data points. Different types of models that are used are listed below. The researcher will dwell only on few models relevant to the paper.

2.4.1 Linear Exponential Smoothing (LES) Model

This model also known as exponential smoothing or single exponential smoothing (SES). From [33], exponential smoothing is one of the simplest and popular forecasting techniques that is included in popular forecasting software products. The algorithm automates the weight assignment process and it doesn't utilize the order movement and its forecasting mechanism doesn't ignore the past data. To automate the weight assignment, it utilises the summation of exponents and constant known as alpha that is assigned a value of $0 > \alpha < 1$.

As presented below, the formula for linear exponential smoothing whereby f_{t+1} represents the next forecasted value and alpha (α) is constant value ranges between 0 and 1.

$$f_{t+1} = \alpha Y_t + (1 - \alpha)f_t \quad (3)$$

The formula can be expanded up until the n^{th} period.

$$f_{t+1} = \alpha Y_t + \alpha(1 - \alpha)Y_{t-1} + \alpha(1 - \alpha)^2 Y_{t-2} + \dots + \alpha(1 - \alpha)^{n-1} Y_1 + (1 - \alpha)^n f_t \quad (4)$$

2.4.2 Double Exponential Smoothing (DES) Model

This model known as Trend Adjusted Exponential Smoothing (TAES) and "Holt-Winter's trend and seasonal method". It outperforms single exponential smoothing, especially when there is a non-stationery seasonal effect [30]. This model improves its accuracy level by adjusting the forecasted values based on a trend. The seasonal forecast is calculated as the previous forecast with trend added to the product of alpha and the difference of the previous actual value and the previous forecast with trend. The trend is calculated as the previous trend added to a product of a beta (β) and the difference between the current seasonal forecasted value and the previous forecast with trend.

$$\begin{aligned} f_t &= fit_{t-1} + \alpha(A_{t-1} - fit_{t-1}) \\ T_t &= T_{t-1} + \beta(f_t - fit_{t-1}) \\ fit_t &= f_t + T_t \end{aligned} \quad (5)$$

2.4.3 Simple Moving Average (SMA) Model

SMA model is the simplest method that perform forecasting by calculating an average from the past data towards the recent data. The model has to be parameterized with the number of steps it should take towards forecasting of the future values. The algorithm will start to calculate from the past data towards the recent data. [30] said the objective of SMA is to smooth the past data to estimate the trend cycle component by computing the moving average by dropping the oldest observations and including the next observation. The weakness of this model is that its forecasting mechanism focuses only on recent data ignoring the past data. This model assigns equal weights on the regressed subset data to predict for the next observation. The formula is presented below.

$$f_t = \frac{1}{k} \sum_{j=1}^k Y_{t-j} \quad (6)$$

Where k represents the order or the step of the moving average. The summation will start from the recent observation (Y_{t-1}) of the selected subset and move towards the old observation of Y_{t-j} subset.

2.4.4 Weighted Moving Average (WMA) Model

This model is similar to SMA, it makes use of the order movement, and calculates the average from the selected subset of data. The predicted value is highly related to the subset of the previous values. However, each regressed subset is manually assigned a higher weight that is $0 > a < 1$, and it will decay towards the oldest observation in the subset. It is conventional that these weights are summed to 1. The formula is represented below.

$$f_t = \frac{1}{k} \sum_{j=1, r=1}^k Y_{t-j} W_r \quad (7)$$

Whereby W is the vector of user defined weights based on the order process of the WMA.

2.5 Indigenous Knowledge

Even though the use of science and technology to monitor vegetation is escalating in every direction. In Africa, the use of machines and remote sensing technology has not been widely adopted and fully utilized due to poverty, educational level and cultural norms. Unemployment rate has been a trending concern in South Africa; education level is also dissatisfying, according to the statistics 18.6% of the South African farmers never went to school, while 56.3% never reached Grade 12 education [7].

Most African farmers still retain and maintain the use of indigenous knowledge than technology. Africa is the continent that still preserves, manage and share IK from generation to generation as it is crucial for social and economic development especially in rural Africa [26]. Over the course of history and up to this day, traditional local communities have continued to rely heavily on their own indigenous knowledge systems in observing the environment and dealing with natural disasters [27]. It is unique to a given culture or society and it is the basis for local level decision making in agricultural

health care, food preparation, education and natural resource management [26]. It is a collection of a vast of body of knowledge on disaster prevention and mitigation, early warnings, preparedness and response acquired through observation and study, and is often based on cumulative experience handed down from generation to generation [27]. IK covers diverse areas of importance for society, spanning issues concerned with the quality of life from agriculture and water to health [41]. IK can also be defined as systematic body of knowledge acquired by local people through the accumulation of experiences and intimate understanding of the environment in a given culture [15]. It can also be explained as a formulation of expertise based on an observation of natural phenomena.

Although IK is considered as a valuable resource in sustainable development and peace building, it is not fully utilized by many domains such as government, donors, scientists and so forth [35]. According to [32] in western oriented academic circles and investigation, African voice is either side-lined or suppressed, and this is witnessed when IK and methods are ignored or not taken seriously. The notion of the static nature of indigenous knowledge has been disproved through numerous examples showing how IK can be changing continuously, especially as a culture or a community develops and grows, and is subject to changing environmental, cultural, physical and economical stressors [41].

Even though western science and technology are seen as a reliable resource, the use of scientific tools such as scientific weather forecasts in Africa is hampered by many technological challenges while their utilization is dismal [25].

As a way forward, [1] declare that IK and SK (scientific knowledge) can be merged together as long as IK is well collected and archived in a systematic manner. The integration of science, technology and indigenous knowledge as a way of performing information dissemination will help the farmers to respond to the forecast and cope with suboptimal climate impacts. In addition, [12] states that although the generation of knowledge has been confined by science alone, the emergence of a new contract between science and society has renewed knowledge management and production stakes. Evidence showed that local knowledge can and must be integrated with information and technology in efforts to improve the rural livelihoods ([10, 19, 24]). These researchers further enlighten that the integration of IK and science has proved its viability for the past decades and has expanded in the field of medicine, engineering, artificial intelligence, machine learning, social and economic sciences and other domains.

Some few researches are presented below: [27] explained the integration of both science and IK to diagnose types of hazards in diverse environmental and cultural settings. [22] proved the merge of IK and modern science as a bridge to forecast an early warning of drought, so to implement the strategies for preparations by the local communities. Further, [4] proposed a middleware that will integrates the heterogeneous data sources with IK based on a unified ontology for an accurate IoT-based drought early warning system (DEWS). [24] proposed the participatory geographic information systems as an organizational platform for the integration of traditional and scientific knowledge in contemporary fire and fuel management. Finally, [18] explained the integration of local knowledge and science to address the economic consequences of driftwood harvest in a changing climate.

3 Methodology

Mixed method approach was deemed most applicable because it enables elicitation of ideas and views of the phenomena in order to get descriptive and accurate findings. The sample of 50 farmers were interviewed while 100 surveys were distributed to the community of uMgungundlovu region in Pietermaritzburg South Africa. Among the selected farmers, 15 of them were growing sorghum and beans, 10 of them were growing wheat and peas and 25 were growing maize. The purposive sampling strategy was used, where by these farmers were selected on the basis that they are educationally and technologically illiterate, they are not equipped with irrigation systems and rely heavily on indigenous knowledge to perform cropping. During the interview, farmers were requested to state down the challenges that they are facing when using IK knowledge system alone to determine the onset of favourable weather season, crop monitoring and crop health and yield prediction. They accentuate on the impact of climate change on their crop health and yield and the impact of biotic (pests, crop diseases and other crop intruders) and abiotic (soil fertility) to their crop yield.

They were further requested to list the indigenous knowledge indicators that they use to foresee the onset of rains and winter season. Because IK indicators have different impacts towards the onset of favourable weather season, farmers were guided on scoring each indicator on a scale of 1 to 10.

The survey that was distributed to the local community was comprised of close and open-ended questions. The aim of the survey was to get the local insights about climate change, its impact on their vegetation and livestock, strategies that they use conserve water from rains and so forth.

The IK indicators and their associated weights together with local insights analysed from the surveys and interviews were documented for further processing.

4 Implementation

During the implementation phase, the system was developed based on the following domains.

- Determine the season onset
- Crop monitoring
- Crop health prediction

4.1 Determine the Season Onset

The indigenous knowledge, satellite imagery and machine learning were used to foresee the oncoming favourable weather season in the next three weeks. The summer season was identified with the successive rains and warm temperature while the winter season was identified with cold dry season (Fig. 1).

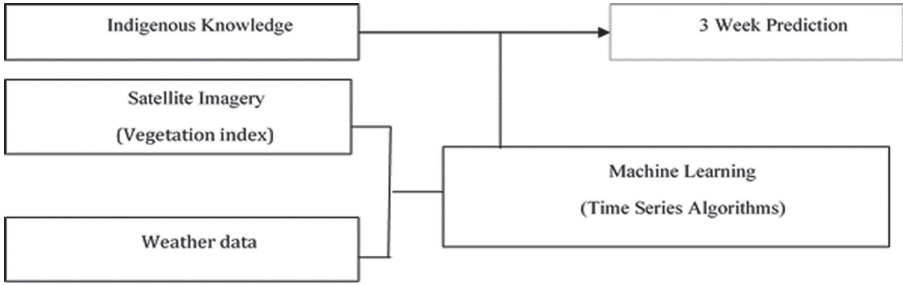


Fig. 1. Three-week prediction of season onset using three knowledge systems

4.1.1 Indigenous Knowledge

Data Preparation. Number of seasonal IK indicators associated with weights were obtained during the methodology phase. The indicators were clustered in to summer season and winter season. These indicators were stored in a file tagged with their appropriate weights. The weights were ranging between 0 and 1 where 0 represent no impact and 1 represents high impact.

Data Processing. The mobile app was developed whereby the user will first interact with the list of available IK indicators depending on the type of crop they want to cultivate and will be requested to select the indicators that they have observed from the environment in the last few days. The system will calculate the average percentage of the selected IK weights from the sum of all weights. The formula is presented below where $\sum_{b=1}^k L_b$ represents the selected weights and $\sum_{i=1}^n F_i$ represents the total weights.

$$AVG_IK = \left(\frac{1}{\sum_{i=1}^n F_i} * \sum_{b=1}^k L_b \right) * 100 \tag{8}$$

4.1.2 Satellite Imagery

Data Preparation. Previous years monthly Pietermaritzburg’s satellite images were requested from MODIS satellite. The edge detection algorithms were used to extract uMgungundlovu region from each satellite image. From the extracted region, the average NDVI was calculated using the formula below where N represents the number of either NIR or Blue pixel values.

$$AVG_NDVI = \left\{ \frac{1}{N} \sum_{i=1}^n \frac{NIR_i - B_i}{NIR_i + B_i} \right\} \tag{9}$$

The aim of performing the above operation was to get one value that will represent the level of vegetation greenness on each extracted image. The average NDVI value which ranges between .1 and <1 was further converted in to percentage weight based on the type of crop needed to be cultivated (summer or winter crop) using the algorithm presented below (Fig. 2).

```

if(crop_type = 'summer') then
  if(avg_ndvi > 0 AND avg_ndvi < 1) then
    percentage = ((avg_ndvi) / .99) * 100
  else then
    percentage = 0
  else then
    if(avg_ndvi > 0 AND avg_ndvi < 1) then
      percentage = (.1 / avg_ndvi) * 100
    else then
      percentage = 0
  
```

Fig. 2. An algorithm that assigns weight to average ndvi value

The calculated weights were sent to the time series models of machine learning. Four models were trained which are SMA, WMA, LES and DES. The RMSE (Root Mean Squared Error) formula was used to calculate the error rate of the models (Table 1).

Table 1. Representation of model evaluations

Times series models	RMSE
SMA	11.9
WMA	21.2
LES	15.6
DES	17.8

The SMA model outperformed other models, and it was used to perform four months’ average NDVI predictions. The selection of four months was to cover the life span of the summer crops. The monthly predictions were converted into weeks so it will be easier to extract value based on a certain week. The values were sent to the cloud for further processing.

Data Processing. Immediately after the average percentage of the selected IK indicators is calculated, the system will extract the next three-week average NDVI percentage and integrate them together.

The main reason average NDVI is utilized on the prediction of favourable weather season is because number of researches showed that vegetation greenness increases during the onset of rains and depreciates during the onset of winter season. Therefore, tracking the vegetation greenness can give insight of the onset of the favourable weather season.

4.1.3 Climate Data

Data Preparation. The previous years’ weekly rain and temperature data of Pietermaritzburg was obtained from South African climate data provider. These variables were analysed by the four-time series models which are WMA, SMA, LES and DES. The

WMA model outperformed other models with the error rate of 7.9. It was selected to predict the next four months 'weekly predictions to cover the life span of summer crops. The data was sent to the cloud for further processing.

Data Processing. Immediately after the average IK is calculated, the system will extract the next three-week rain and temperature predictions from the cloud and will be converted in to one weight. The following algorithm will be used to evaluate the type of crop needs to be cultivated (summer or winter crop), level of rains in millimeters and level of temperature in degree Celsius (Fig. 3).

```

if(crop_type = 'summer') then
  if(rain_level >= MIN_RAIN AND rain_level <= MAX_RAIN) then
    if(temp_level >= MIN_TEMP AND temp_level <= MAX_TEMP) then
      weight = ((temp_level/MAX_TEMP)*40) + (rain_level/MAX_RAIN)*60
    else if(temp_level < MIN_TEMP) then
      weight = ((temp_level/MAX_TEMP)*40) + (rain_level/MAX_RAIN)*60
    else then
      weight = √(((temp_level/MAX_TEMP)*40) + √(rain_level/MAX_RAIN)*60)
    else if (rain_level < MIN_RAIN) then
      weight = (rain_level/MIN_RAIN)*100
    else then
      weight = ((rain_level/MIN_RAIN)*100)/2
  
```

Fig. 3. An algorithm that assigns weight to climate data based on type of crop

In summary, the rain takes a high precedence when cultivating the summer crops while temperature takes a high precedence when cultivating the drought resistant crops.

4.1.4 Integration of Climate Data, Indigenous Knowledge and Satellite Imagery

The weights of all knowledge systems will be integrated together, and the average will be calculated. The user will receive the next three week forecast with a meaningful feedback.

$$FinalPrediction = \frac{IK\ weight(\%) + climate\ data\ weight(\%) + vegetation\ index\ weight(\%)}{3} \quad (10)$$

4.2 Crop Monitoring

The observation by an indigenous knowledge expert will be made on crops. The farmer will use the app to scale the health properties of the crops. These include crop greenness, leave size and stem size, soil health and fertility, level of crop pests and other crop intruders, crop diseases, and other properties. The system will calculate the average scaling done by the farmer. The results will be stored in a cloud. The farmer will monitor the crop health daily or at least 4 times per week. The scaling of each property ranges from 1 to 100, where 1 represent very weak health and 100 represents excellent health.

The formula below was used to obtain the average percentage of the monitored results where $\sum_{i=1}^n \frac{F_i}{100}$ represents the summation of weights scaled out of 100.

$$AVG_{Monitoring}(\%) = \frac{\sum_{i=1}^n \frac{F_i}{100}}{n * 100} \quad (11)$$

4.3 Crop Health Prediction

During this stage, the system will extract monitored results from the cloud and send them to machine learning algorithms to perform three-week crop health prediction. The system will automate the models' training, error rate calculation, best model selection and three-week crop health prediction. To add more weight to the predicted results, we integrated the 3-week forecast of climate data and average NDVI (Fig. 4).

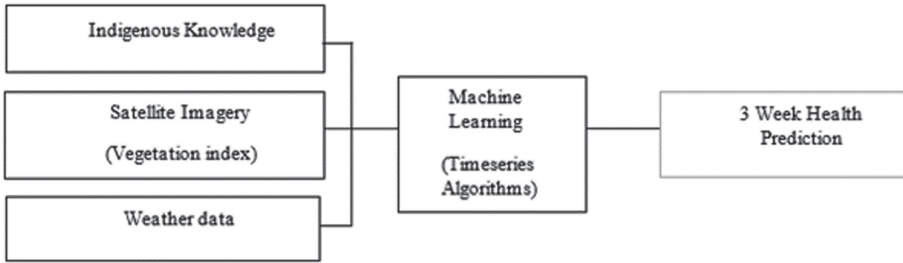


Fig. 4. Three knowledge systems used to predict next three-week crop health

$$FinalPrediction = \frac{Prediction\ of\ Monitored\ Results(\%) + next\ 3\ week\ climate\ data(\%) + next\ 3\ week\ avg}{3} \quad (12)$$

5 Model Evaluation and Conclusion

For this paper we evaluated the system performance to foresee the onset of rains. For the system to perform forecasting, the weights of indigenous knowledge indicators observed from the environment need to be computed. The researchers with the help of indigenous knowledge experts analysed the rainy season IK indicators that were stated during the methodology phase. We categorised the IK indicators to months that they frequently appear. We further calculate the average weights of those indicators as shown on the table below (Table 2):

Table 2. Category of IK indicators

Months	AVG_IK Indicators
January	60%
February	45%
March	20%
April	5%
May	7%
June	10%
July	25%
August	50%
September	55%
October	70%
November	80%
December	80%

The data presented above was converted into weeks to perform weekly forecasts and was used as an input to the system. The 2019 weekly climate data (rain and temperature calculated percentage weights) was extracted. The 2019 satellite imagery data (percentages of average NDVI) was extracted. The system forecasts were tested against 2019 climate data as displayed below: We validated the system starting from week 1 of January to week 48 of December. The system forecasted the next three-week observations which were compared to the next three week climate data (Fig. 5).

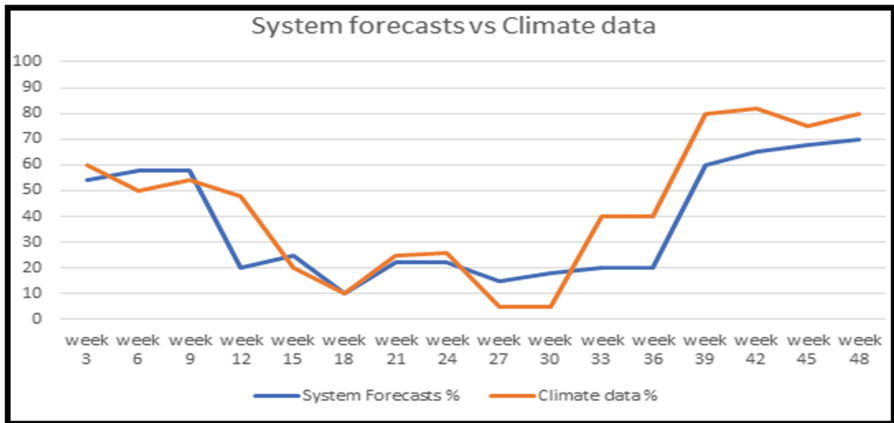


Fig. 5. System forecasts vs climate data chart

The mean percentage error(MPE) algorithm was used as modelled below.

$$mpe = \frac{100\%}{n} \sum_{j=1}^n \frac{a_j - f_j}{a_j}$$

The MPE of 15.35% was obtained. With the accuracy level of 84.65%. The system forecasts were again tested on 2019 satellite imagery data as displayed below (Fig. 6).

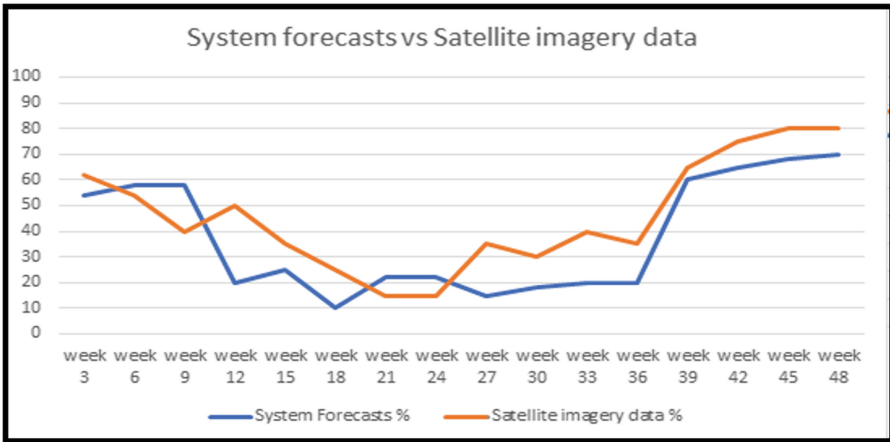


Fig. 6. System forecasts vs satellite imagery data chart

The MPE of 26.7% was obtained. With the accuracy level of 73.3%.

The system forecasts were tested against data from 2019 to 2017. The table shows the accuracy level of the prototype against climate data (Table 3).

Table 3. Accuracy measure against climate data

Years	Accuracy measure against Climate data
2019	85%
2018	58%
2017	62%
2016	70%
Average Accuracy	68.75%

The table below shows the accuracy level of the prototype’s forecasts against satellite imagery collected data (Table 4).

Table 4. Accuracy measure against satellite imagery data

Years	Accuracy measure against satellite imagery data
2019	73%
2018	61%
2017	63%
2016	69%
Average Accuracy	66.5%

The correlation was calculated between the 2019 climate data and 2019 satellite imagery data. The 87.6% of the results was obtained. The correlation calculation was extended from 2019 to 2016.

From correlation statistics we conclude that the vegetation health is highly correlated with the rainy season. Therefore, we can analyse the vegetation health to infer to the onset of rains and suitable temperature (Table 5).

Table 5. Comparison of satellite and climate data

Years	Satellite Imagery data vs Climate data (Correlation accuracy)
2019	88%
2018	81%
2017	70%
2016	83%
Average Correlation	80.5%

The average system accuracy of approximately 70% was reached. Some of the factors that hampered the accuracy of the system are climate conditions on the satellite images, ever-green vegetation that cannot be related to any season, and some flaws observed from NDVI such as failure rate on highly saturated areas and so forth.

The model is easier and cheaper to use in contrast with other agricultural technologies. The paper valued the indigenous knowledge and emphasizes that indigenous knowledge can be mathematically calculated and integrated with the sophisticated technologies such as satellite imagery and machine learning algorithms. Although the model is not perfect, it is useful to many local farmers that mostly rely on rains for life.

References

1. Agrawal, A.: Indigenous and scientific knowledge: some critical comments. *Antropologi Indonesia* (2014)

2. Ahmad, F.: Landsat ETM + and MODIS EVI/NDVI data products for climatic variation and agricultural measurements in Cholistan Desert. *Global J. Hum.-Soc. Sci. Res.* **12**(13), 1–11 (2013)
3. Aina, L.O.: Globalisation and small-scale farming in Africa: what role for information centres. In: *World Library and Information Congress: 73rd IFLA General Conference and Council*, pp. 1–8 (2007)
4. 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 (2015)
5. Al-Kindi, K.M., Kwan, P., Andrew, N.R., Welch, M.: Remote sensing and spatial statistical techniques for modelling *Ommatissus lybicus* (Hemiptera: Tropiduchidae) habitat and population densities. *PeerJ* **5**, e3752 (2017)
6. Bringezu, S.: Visions of a sustainable resource use. In: *Sustainable Resource Management*, pp. 155–215. Routledge (2017)
7. *Community Survey: Agricultural households*. Statistics South Africa (2016)
8. Davey, J.: *Sugar cane monitoring and analysis using remote sensing*. Wooding Geospatial Solutions (2018)
9. Dhanachandra, N., Manglem, K., Chanu, Y.J.: Image segmentation using K-means clustering algorithm and subtractive clustering algorithm. *Procedia Comput. Sci.* **54**, 764–771 (2015)
10. Gerke, S., Evers, H.D.: Globalizing local knowledge: social science research on Southeast Asia, 1970–2000. *J. Soc. Issues in Southeast Asia* **33**(S), S242–S263 (2018)
11. Gibbs, H.K., Ruesch, A.S., Achard, F., Clayton, M.K., Holmgren, P., Ramankutty, N., Foley, J.A.: Tropical forests were the primary sources of new agricultural land in the 1980s and 1990s. *Proc. Natl. Acad. Sci.* **107**(38), 16732–16737 (2010)
12. Girard, N.: Knowledge at the boundary between science and society: a review of the use of farmers' knowledge in agricultural development. *J. Knowl. Manag.* **19**(5), 949–967 (2015)
13. Goldblatt, A.: *Agriculture: Facts & Trends: South Africa*. Ceo Wwf-Sa (2010)
14. Gomez, C., White, J.C., Wulder, M.A.: Optical remotely sensed time series data for land cover classification: a review. *ISPRS J. Photogram. Remote Sens.* **116**, 55–72 (2016)
15. Gupta, A.D.: Way to study indigenous knowledge and indigenous knowledge system. *Cult. Anthropol.* **373**, 393 (2012)
16. Jean, N., Burke, M., Xie, M., Davis, W.M., Lobell, D.B., Ermon, S.: Combining satellite imagery and machine learning to predict poverty. *Science* **353**(6301), 790–794 (2016)
17. Jiri, O., Mafongoya, P.L.: Managing vulnerability to drought and enhancing smallholder farmers resilience to climate change risks in Zimbabwe. In: *Handbook of Climate Change Resilience*, pp. 1–17 (2018)
18. 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)
19. Kala, C.P., Farooquee, N.A., Dhar, U.: Prioritization of medicinal plants on the basis of available knowledge, existing practices and use value status in Uttaranchal. *India. Biodiversity & Conservation* **13**(2), 453–469 (2004)
20. Lei, L.I.U., Zhou, L.L., Bao, H.F.: Remote sensing image segmentation algorithm based on multiagent and fuzzy clustering. In: *DEStech Transactions on Environment, Energy and Earth Sciences, (ICESEE)* (2017)
21. Masih, I., Maskey, S., Mussá, F.E.F., Trambauer, P.: A review of droughts on the African continent: a geospatial and long-term perspective. *Hydrol. Earth Syst. Sci.* **18**(9), 3635 (2014)
22. Masinde, E.M.: *Bridge between African Indigenous knowledge and modern science on drought prediction* (Doctoral dissertation, UNIVERSITY OF CAPE TOWN) (2012)
23. Masinde, M., Mwagha, M., Tadesse, T.: Downscaling Africa's drought forecasts through integration of indigenous and scientific drought forecasts using fuzzy cognitive maps. *Geosciences* **8**(4), 135 (2018)

24. McBride, B.B., Sanchez-Trigueros, F., Carver, S.J., Watson, A.E., Stumpff, L.M., Matt, R., Borrie, W.T.: 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)
25. Mwachaga, S.M. and 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 (2015)
26. Owiny, S.A., Mehta, K., Maretzki, A.N.: The use of social media technologies to create, preserve, and disseminate indigenous knowledge and skills to communities in East Africa. *Int. J. Commun.* **8**, 14 (2014)
27. Pareek, A., Trivedi, P.C.: Cultural values and indigenous knowledge of climate change and disaster prediction in Rajasthan, India (2011)
28. Pavlovsky, G.: *AGRICULTURAL RUSSIA: On the Eve of the Revolution*. Routledge, London (2017)
29. Picchioni, F.: Monitoring the effects of changing food prices on food and nutrition security: the minimum calorie expenditure share (MCES) (Doctoral dissertation, SOAS University of London) (2017)
30. Prema, V., Rao, K.U.: Development of statistical time series models for solar power prediction. *Renew. Energy* **83**, 100–109 (2015)
31. Qiu, J., Wang, X., He, S., Liu, H., Lai, J., Wang, L.: The catastrophic landslide in Maoxian county, Sichuan, SW China, on June 24, 2017. *Nat. Hazards* **89**(3), 1485–1493 (2017)
32. Owusu-Ansah, F.E., Mji, G.: African indigenous knowledge and research. *African J. Disabil.* **2**(1), 30 (2013)
33. Ravinder, H.V.: Determining the optimal values of exponential smoothing constants—does solver really work? *Am. J. Bus. Educ.* **6**(3), 347–360 (2013)
34. Roncoli, C., Ingram, K., Kirshen, P.: Reading the rains: local knowledge and rainfall forecasting in Burkina Faso. *Soc. Nat. Resour.* **15**(5), 409–427 (2002)
35. Ross, A., Sherman, K.P., Snodgrass, J.G., Delcore, H.D., Sherman, R.: *Indigenous Peoples and the Collaborative Stewardship of Nature: Knowledge Binds and Institutional Conflicts*. Routledge, London (2016)
36. Sabini, M., Rusak, G., Ross, B.: Understanding Satellite-Imagery-Based Crop Yield Predictions. Technical Report. Stanford University. <http://cs231n.stanford.edu/reports/2017/pdfs/555.pdf>. Accessed 23 Oct 2017
37. Silver, D., et al.: Mastering the game of Go with deep neural networks and tree search. *Nature* **529**(7587), 484–489 (2016)
38. Sowmya, D.R., Rao, S.A., Shenoy, P.D., Venugopal, K.R.: Generation of digital elevation map for steep terrain region using Landsat-7 ETM + Imagery. In: 2018 International Conference on Data Science and Engineering (ICDSE) (pp. 1–5). IEEE (2018)
39. Swain, K.C., Zaman, Q.U., Schumann, A.W., Percival, D.C., Bochtis, D.D.: Computer vision system for wild blueberry fruit yield mapping. *Biosys. Eng.* **106**(4), 389–394 (2010)
40. Thakur, N., Maheshwari, D.: A review of image classification techniques. *Int. R. J. Eng. Technol. (IRJET)* **4**(11), 1588–1591 (2017)
41. Tharakan, J.: Indigenous knowledge systems for appropriate technology development. In: *Indigenous People*, p. 123 (2017)
42. Verbesselt, J., Hyndman, R., Newnham, G., Culvenor, D.: Detecting trend and seasonal changes in satellite image time series. *Remote Sens. Environ.* **114**(1), 106–115 (2010)
43. You, J., Li, X., Low, M., Lobell, D., Ermon, S.: Deep Gaussian process for crop yield prediction based on remote sensing data. In: *AAAI*, pp. 4559–4566 (2017)
44. Zhao, Y., et al.: Genome-based establishment of a high-yielding heterotic pattern for hybrid wheat breeding. *Proc. Natl. Acad. Sci.* **112**(51), 15624–15629 (2015)