Quest Journals
Journal of Research in Agriculture and Animal Science
Volume 7 ~ Issue 5 (2020) pp: 20-29
ISSN(Online) : 2321-9459
www.questjournals.org

Research Paper

"Locust Invasion Prediction for Future Food Security and Sustainable Agriculture in East Africa" Case Study SmartAfriHub Hackathon Challenge 4

ObotAkaninyene
Department of Agricultural Economics and Extension, NnamdiAzikiwe University, Nigeria, ORCID: 0000-0001-8775-4064
- *MaximilienHouel, SISTEMA GmbH*
- *Elias Cherenet, Haramaya University, ORCID: 0000-0001-5335-2120*
- *Winnie Kamou, Association of Freelance Journalists*
- *Lilian Ndugu, Regional Centre for Mapping of Resource for Development*
- *Paul Kasoma, Youths in Technology and Development Uganda (YITEDEV)*
- *KizitoOdhiambo, agriBORA UG*
- *ChristusCitoMiderho, Université Notre Dame du Tanganyika, ORCID: 0000-0001-8347-1374*

ABSTRACT
It is known that the locusts are depending on climatic factors (rain, temperature, soil moisture), by the combination of these data sets and some specific model as a "species distribution modelling", it can be possible to provide location based information by generating risk maps of potential locust invasion or at least shows an estimation of where the locust might appear in the next few days.
With the actual satellite open data available the best way to spot the hotspot of locusts is by detecting and measuring the impact. It is predicted that the locust swarm feed from the agriculture fields/green leaves, hence satellite imagery/earth observation data can be used as a proxy to measure the volume of the vegetation on the ground using the indices derived from it i.e Normalized Difference Vegetation Index (NDVI), therefore, any abrupt decrease in vegetation distribution/greenness on a given area can be considered a proxy of locust attack. Measuring the associated impact could permit us to identify the areas where urgent help is needed. Usually this approach is useful to estimate the degree of damage as a response to the locust swarm. This approach is also vital to the farmers as it prepares them for the worst cases, since the farmers have the exact location where the next locust invasion can happen and then will make them prepare as early as possible. For the objective 1, ensemble approach was used to analyze the East African countries most prone to locust invasion. The objective 2, SDM model was used to show the environmental variables favoring the occurrence of locust invasion in East Africa.

***R**eceived 28 November, 2020; **A**ccepted 14 December, 2020*
P**ublished with open access at **www.questjournals.org

I. INTRODUCTION

The Intergovernmental Panel on Climate Change (IPCC) has reported that the most hazardous manifestation of climate change is through increased temperature, wetter and drier climates, heat waves and prolonged droughts, which in turn causes severe fluctuations in crop production (Scherm et al., 2000; Landi and Giovanni, 2016). One of the serious impacts of climate change on crop productivity, especially in sub-Saharan Africa is the shift in the occurrence of pests and diseases. Despite losses caused by the pests and disease, the greatest challenge is the lack of knowledge regarding the probability of occurrence, distribution and direction of pests and diseases spread, as well as the severity of their effect on crops.

In Africa prior to the impact of COVID-19, it is estimated that about 73 million people in Africa are acutely food insecure. With 85% of the African population depending on food importation, COVID-19 has further caused a detrimental increase in the level of food shortage due to travel restrictions and other lockdown measures. Sub-Saharan African is a place of major focus in terms of food insecurity and it was estimated that one in five people is undernourished. Before the pandemic, these food crises have ensued due to factors like climate change, conflicts, lack of investment in agriculture, economic shocks, and unstable markets.

Unprecedented locust breaks have caused losses approximated at USD8.5 billion in crops and livestock, poor harvests, and have severely crippled food buoyancy in the continent (Dupe, 2020).Since the beginning of 2020, the world knows of an infestation of desert locust that goes from Uganda to Pakistan, The countries which are more affected by these locusts include; **Uganda**, **Kenya**, **Ethiopia**, **Somalia**, **Yemen**, **Iran** and **Pakistan**.

According to FAO 2020, "A typical swarm can be made up of 150 million locusts per square kilometer and is carried by the wind up to 150 km in one day, even a very small one square kilometer locust swarm can eat the same amount of food in one day which can be eaten by about 35 000 people."

So far, East African countries are spending a lot of money in controlling the spread of desert locusts. A lot of crops have been destroyed, revenues and export earnings have dropped as well as an increase in governments' expenditure in trying to contain the outbreak. The appearance of the locusts follows a period of extreme weather, including devastating floods that have further threatened the food supply.

The actual situation is alarming because it coincides with the growing season. According to the FAO and United Nations Office for the Coordination of Humanitarian Affairs (OCHA) 2020, "Desert locusts are considered the most destructive migratory pest in the world. Hundreds of thousands of hectares, including cropland and pasture, have already been affected."

The desert locust watch from the FAO is already monitoring the crisis over the African continent but also with an increase of the problem in the Middle East region. However with such a variety of threats the locusts impose onto harvests and yields, there is no silver bullet to protect against losses and damages. Rather, a cohesive approach is needed that incorporates all available tools in the toolbox, from better forecasting and monitoring technologies to other innovative means that preserve human life, crop life, animal life as well as soil biodiversity.

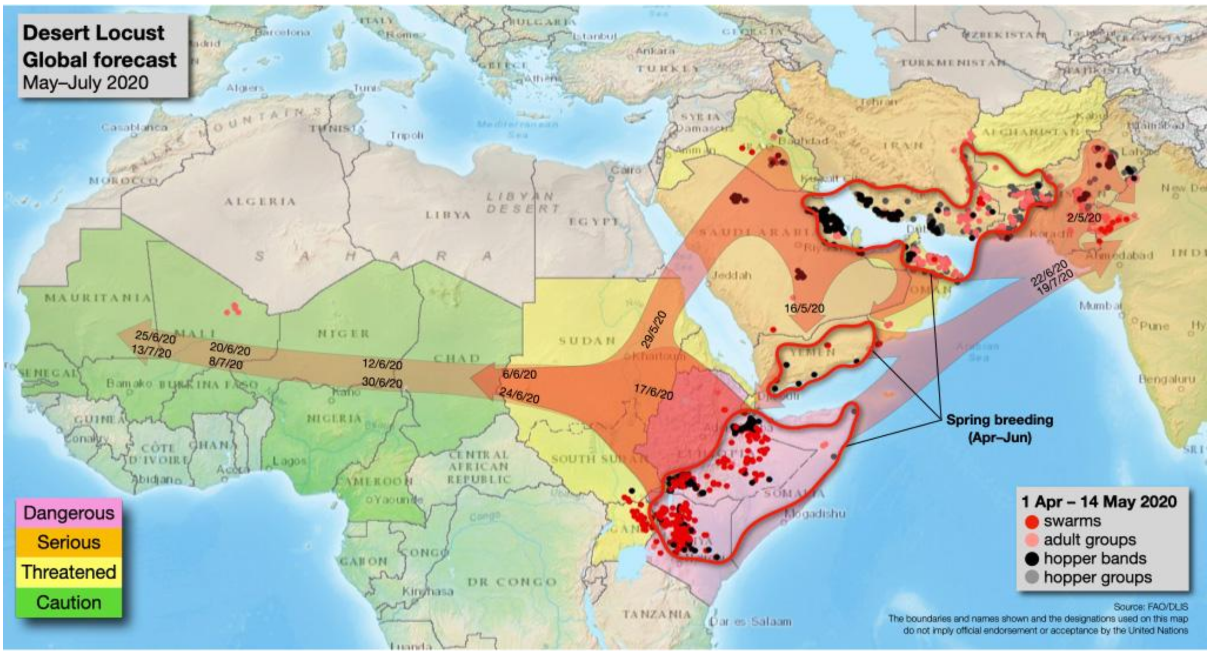

Figure 1: FAO desert locust watch situation and forecasting of the desert locust

The presence of the locusts is a real threat to food security and livelihoods, moreover the situation is showing an increase of swarms but especially spreading in neighboring countries.

Without timely or effective interventions, sporadic cases of desert locusts can easily turn into an upsurge and ultimately a plague.

Due to the actual situation, there is a willingness to find a solution to best respond to this crisis. To answer it, the use of technology is an option to answer quickly and accurately. Nowadays satellite imagery is covering the entire globe and provides a large amount of information concerning what is happening on the ground. It is possible to access data for temperature, soil moisture, precipitation, vegetation, etc. Moreover it is known that meteorological / climatic factors influence the movement and apparition of locusts. In this sense the main idea is to provide answers at two different stages, first to identify the impact of locusts and which area needs more help. Secondly, to prevent locusts by identifying more precisely where a new swarm can appear.

The two approaches are following the idea of developing a new pipeline for monitoring of the locusts with the combination of technological improvement by showing the capabilities of remote sensing data and increase the precision at the best.

Satellite imagery will allow governments and decision-makers to monitor crop conditions across the large expanse of land. This information aids preparation for any forthcoming disasters like droughts and floods, thereby, increasing productivity and profitability. This information when gathered for some time can also be used to forecast an entire region's agricultural output, including anticipating crop shortages. This critical information is necessary to adequately prepare for food shortages and avoid famine by positioning food supplies in areas likely to experience these challenges. With these tools, crises can be predicted instead of being in disarray or reacting after the crises have happened.

Therefore the main goal of this project was to develop a geospatial risk outbreak model for the timely location of desert locust invasion and mapping desert locust risk zones in Eastern Africa region. The identification of potential destruction of the locust and then its mapping to provide updated information on its damage status to the government/decision makers but also to farmers (in the format in which they best understand it).

II. OBJECTIVES

The main objective was to predict and map desert locust in Africa as

In order to achieve the main task the project was sub categorized into two specific objectives: i) **Estimate the region most prone to locust invasion**, and ii) **Identify the conditions favoring the locust invasion location.**

III. METHODOLOGY

Climate models have been used to map the potential distribution of pests and pathogens on basis of the fundamental niche concept (Baker etal., 2000). Thisapproach provides apragmatic tool to explore the potential of exotic pests and diseases spreading into new areas. However, climate data alone isnot exhaustive enough to map the distribution, and potentially vulnerable areas, since other factors come into play. Recent advances in optical remote sensing, which have facilitated the discrimination of health crops from the diseased ones provides an additional tool to mapping.

Vegetation index based on Sentinel-1 and Sentinel-2 satellite data was developed to detect the possible impact of locust invasion. Why S1? Because it has a better temporal resolution than Sentinel-2 and especially because it permits to avoid atmospheric perturbation, as locusts are influenced by rain, hence by using S1 potentially we can get better ground information. Second and foremost, S1 has a combination of the different signals (VV - VH - VV/VH) resulting in a better output close to what is obtained with NDVI derived from S2 optical /passive satellite imagery from S2.

Satellite data acquisition

Satellite products such as Sentinel-1 C-band synthetic aperture radar imaging and Sentinel-2 multispectral images were acquired from Copernicus Open Access Hub. Since Sentinel-1 data is in raw format, therefore, various image pre-processing and cleaning steps were applied to the downloaded imagery.

The image pre-processing methods such as thermal noise removal, image calibration, and speckle filtering and terrain correction were applied to all acquired sentinel-1 images. However, since Sentinel-2 satellite image was acquired in Level C1, only atmospheric correction was applied using the Sen2Cor function available in python to remove the atmospheric perturbation and cloud effects. All the data pre-processing was done entirely in python programming software.

Environmental data acquisition

Data on other biophysical features such as locust presence (GPS points), Temperature, precipitation, and vegetation were acquired from different open sources databases. The locust presences (GPS) points were obtained from agriBORA (194 points) and FAO database (1829 points). A total of 2023 presence points of adult locust data were acquired from both sources. The FAO database has rich records starting from 1985 up to date, therefore for the ease of harmonizing the two datasets, only the data collected in 2020 was considered for this study. All the GPS points are collected in geographic projection format (WGS 84). The dataset regarding temperature and precipitation were downloaded from Adam platform. The ADAM platform is a cross-domain application that supports various application fields and provides various environmental data.

Parametrizing the SDM model and generating risk map

Species Distribution Models (SDMs) are considered as the most powerful tools in many disciplines such as regional conservation planning, climate change impact assessment, and ecology (Naimi, 2015), phylogeography (Alvarado-Serrano & Knowles, 2014). It has robust capability in predicting probability of occurrence in geographical areas (i.e both in space and time) using the presence and absence of data, and can produce invasion risk mapping (Srivastava *et al.*, 2019).

The flowchart (figure 2) illustrates the various steps taken to set the SDM model. Multiple steps were applied to map the presence of locust invasion. Before generating the final binary map, correlation between various environmental predictors included in the model was calculated. Once after confirming the degree of correlation among the variables, different thresholds and ranking was used to select the most important variables to be included in the model, hence producing the binary maps. Finally, by using the selected variables in the previous step, a heat map was generated (i.e location and forecasting).

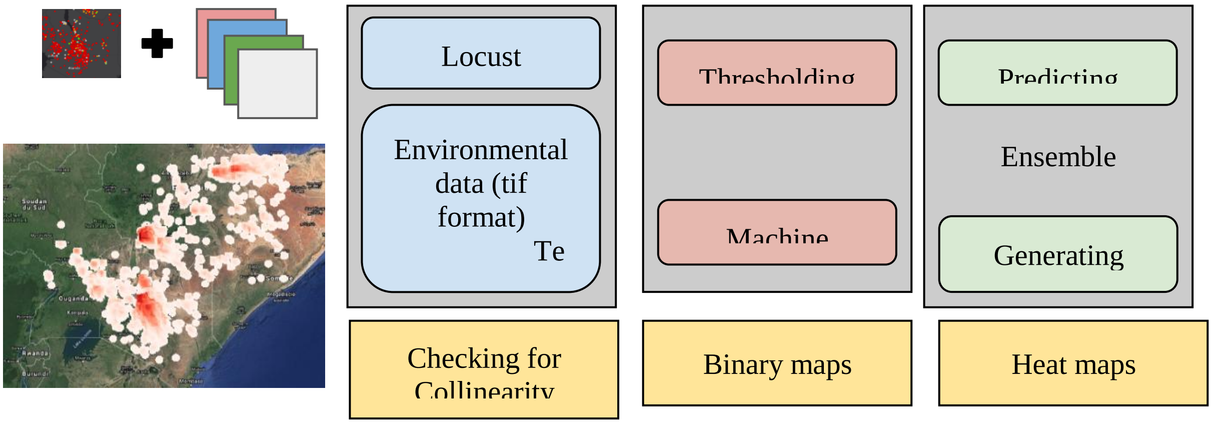

Figure 2: Locust invasion prediction workflow in SDM model

The proposed workflow is as said separated in two different parts: the first step of the work is corresponding to the mapping of the impact of the locust. As noticed below a locust swarm can devastate the agricultural fields. The evolution of the vegetation can thus be considered as a proxy of it. Many possibilities exist for analyzing the state of the vegetation, the most common metric corresponding to the NDVI (Normalized Difference Vegetation Index). This index can be computed with optical satellite imagery, and provide useful and accurate information on vegetation. However optical imagery can be limited in a way, open data used for this purpose is usually from Modis data, covering large areas but with a low spatial and temporal resolution of 250 meters and 16 days respectively. Landsat and Sentinel-2 collections can be a good option to also provide accurate data at 30 meters every 16 days and 10 meters of resolution with a revisit time of 5 days.

Unfortunately optical imagery is limited with meteorological effects, presence of clouds hides the information on the ground. Even if the temporal resolution is quite high, it leaves the possibility of missing a lot of information. An even bigger problem in the locust crisis is that the insect can move quickly in a limited time. To answer this problem we wanted to investigate the use of SAR imagery, with the open data from Sentinel-1, it reaches 10 meters of spatial resolution with a revisit time of 6 days. Moreover radar signals permit to avoid clouds as the signal is going through it. Even if the atmospheric perturbation still exists it permits more information from the ground.

One problem remains concerning Sentinel-1 data, compared to optical imagery it doesn't have a vegetation index as NDVI to analyze the evolution of vegetation. However Sentinel-1 data and Sentinel-2 NDVI are showing correlations that are encouraging the use of SAR data.

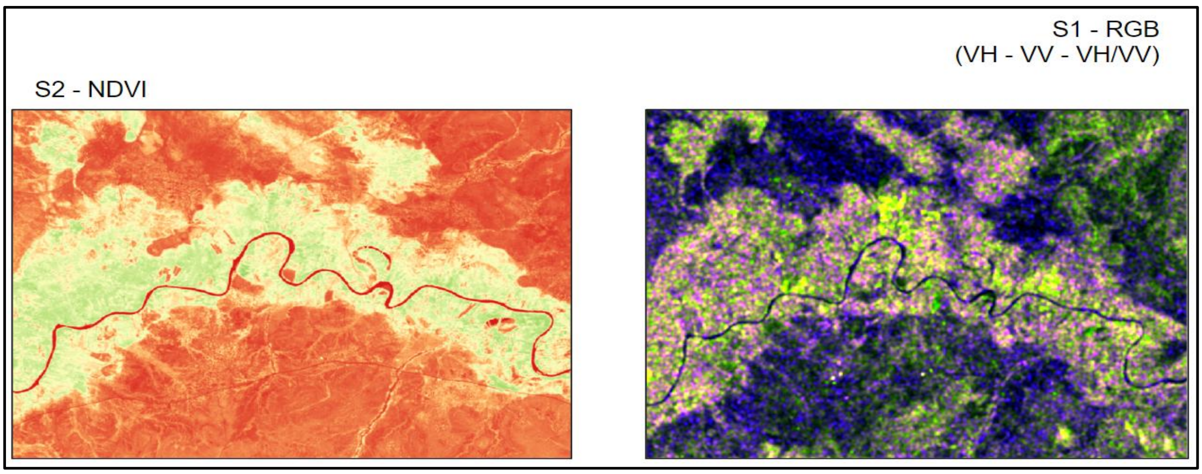

Figure 3: On the left, Sentinel-2 NDVI, high value in green are corresponding to higher volume of vegetation / on right, Sentinel-1 measurement, combination of polarization VH - VV - VH/VV.

For this purpose of analyzing the vegetation with Sentinel-1 we propose a workflow using machine learning methods, to train a model and then compute vegetation indices on Sentinel-1 data and analyze in time the evolution of vegetation over the areas impacted by the locusts.

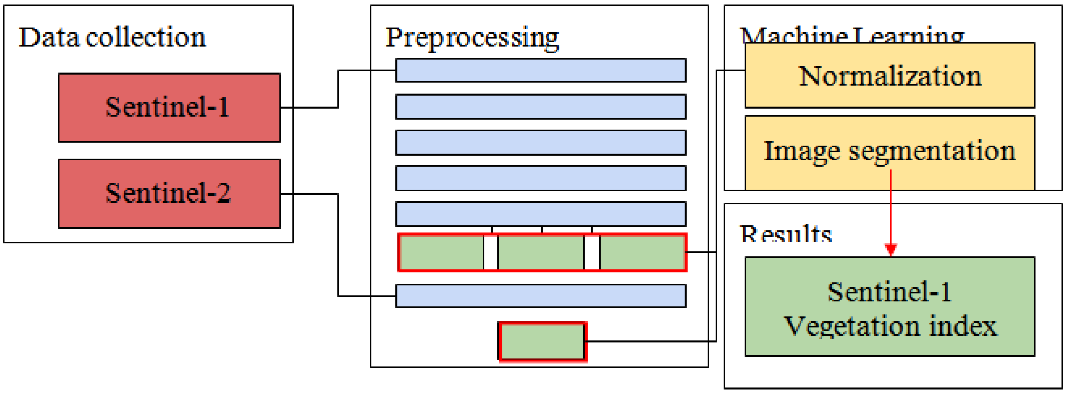

Figure 4: Computation of vegetation index workflow

The analysis of the impact of the desert locust is corresponding to the "answer" part to the crisis. **Knowing where the damages are** and identify the needs for future steps. The second approach corresponds to **a preventive part to the crisis**, the objective of this approach is to build a model similar to a species distribution model (SDM).

The link between environmental factors and the presence of locust is already proved now. In this way there is a possibility to imagine a model by combining climatic, meteorological, geological information to identify the habitat of the locusts and so to map their trajectory.

Locations of the swarm at a specific time are known and can be combined with satellite data for temperature, precipitation, soil moisture, etc. For a specific area and a specific time, the information can be compared. The workflow is separated into three main parts: an analysis of the link between the diverse information, a modelling based on this link and then a mapping of the potential presence of locusts.

IV. RESULT

For the first step of the work and the computation of a vegetation index based on Sentinel-1 measurement, some preprocessing was added to permit training of models. The NDVI has values between -1 and 1, to fit into an image segmentation work a re-class process was needed, then the result is between 0 and 100, and every class corresponds to 5% more than the previous class. As a baseline, we decided to use an Ensemble Method by using a Random Forest to create the NDVI and to see if the work can be then moved to a more complex model in deep learning as UNet.

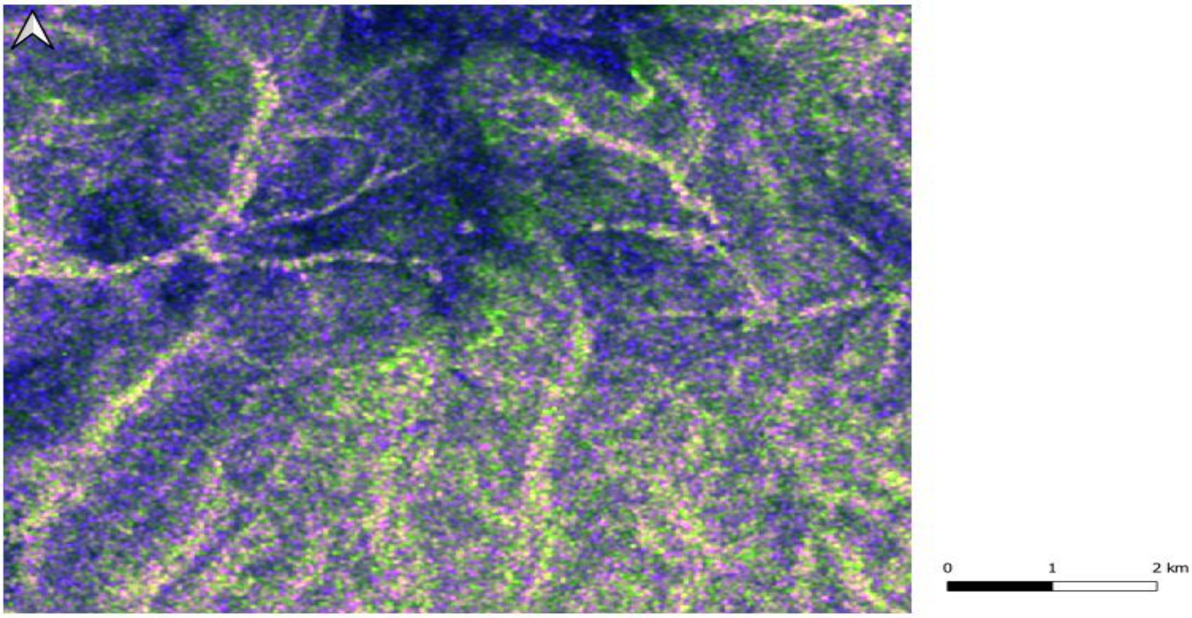

Figure 5: Sentinel-1 measurement (VV - VH - VH/VV)

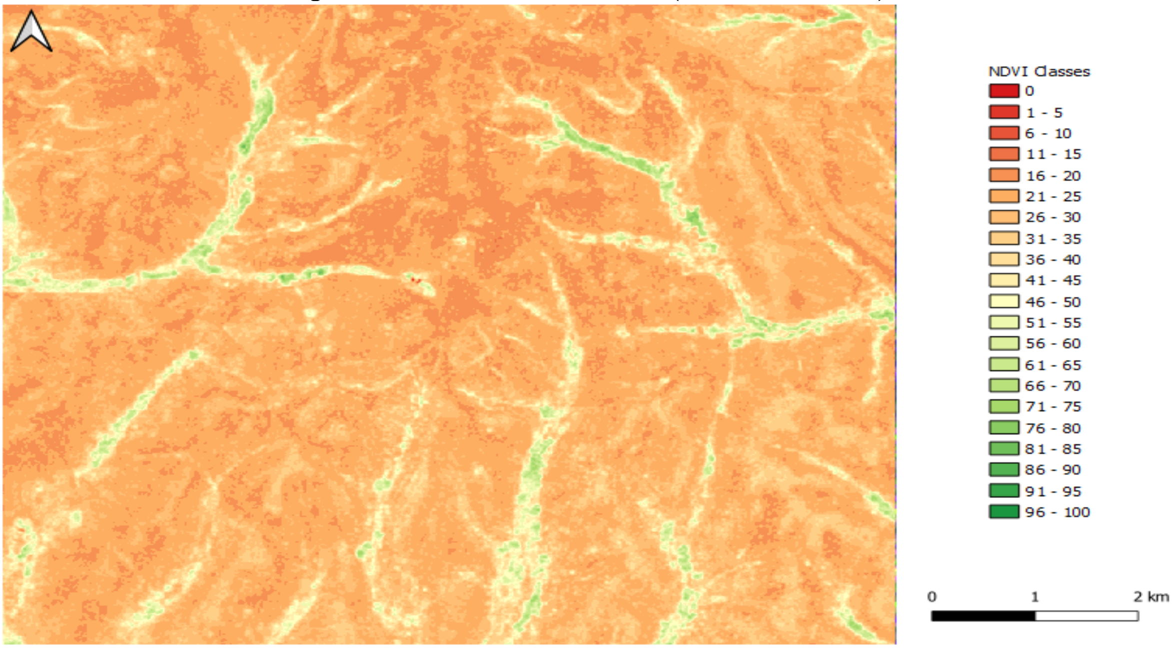

Figure 6: Sentinel-2 NDVI used as ground truth

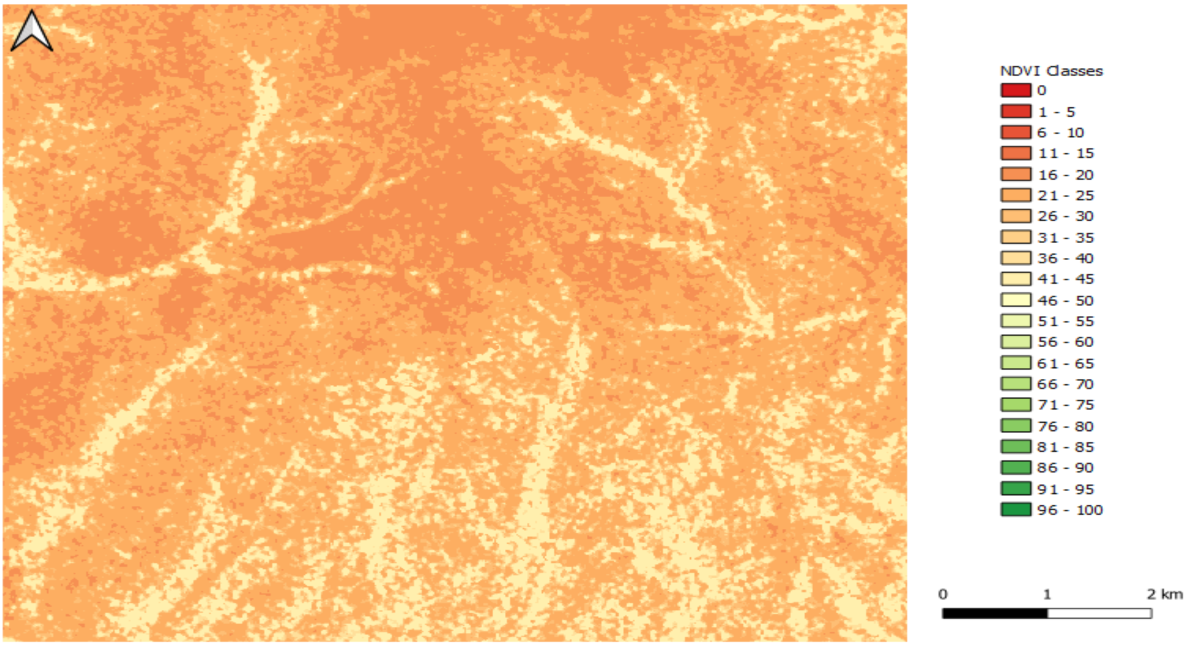

Figure 7: Result of the Random Forest model

Visually the first result obtained is encouraging, but still the result of the model is showing limitations as we can see especially in the south part of the example above. Due to the complexity of Sentinel-1 measurement the model is presenting a difficulty to identify precise areas that are supposed to be more vegetative. The model in itself is limited, of course the Random Forest is based on analysis of pixel values and is not looking in the structure of the image, it is in this way that moving to a more complex model as UNet could be a good option.

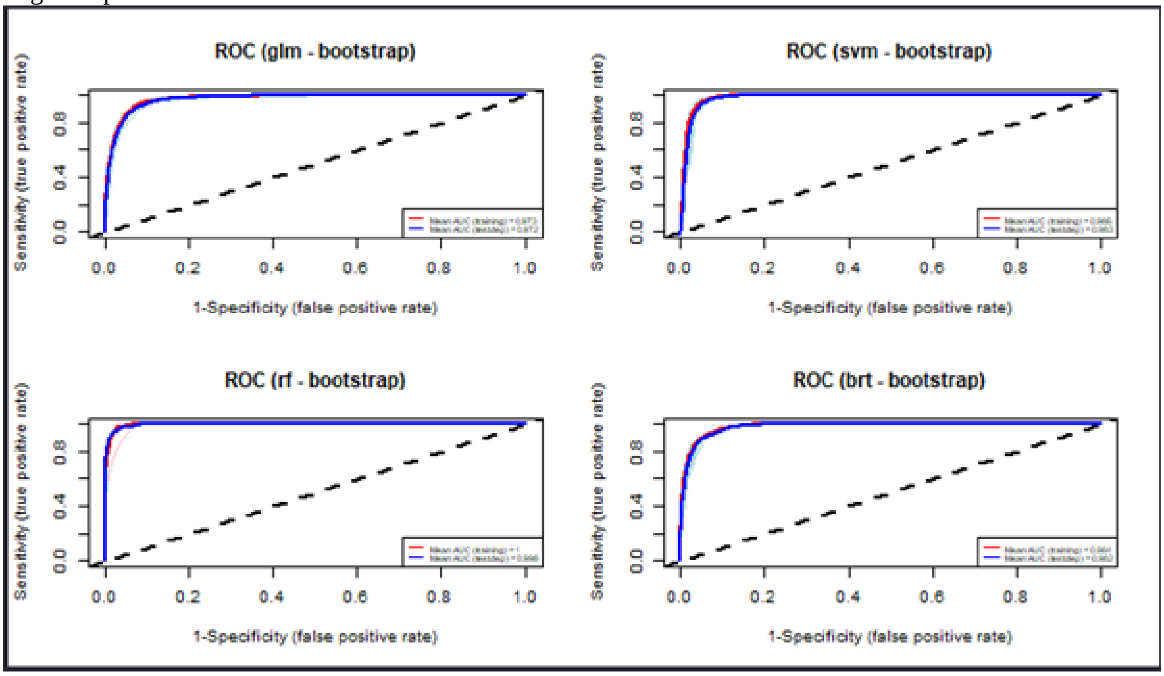

Figure 8: ROC-AUC value of the four SDM model

Figure 7 illustrates the model evaluation result based on ROC-AUC for the four sub-models included in the SDM model. All the models showed higher sensitivity (true positive rate) and lower specificity (false positive rate) for predicting the locust occurrence.

Table1. Model Mean performance (per species), using test dataset (generated using partitioning)

Methods	AUC	COR	TSS	Deviance
glm	0.97	0.87	0.86	0.41
svm	0.98	0.92	0.92	0.32
rf	1	0.96	0.97	0.13
brt	0.98	0.89	0.87	0.64

Based on the AUC and TSS statistics evaluation the four models performed very well (table 1). However, random forest (rf) and support vector machine (svm) machine learning algorithms outperformed compared to the other two regression methods.

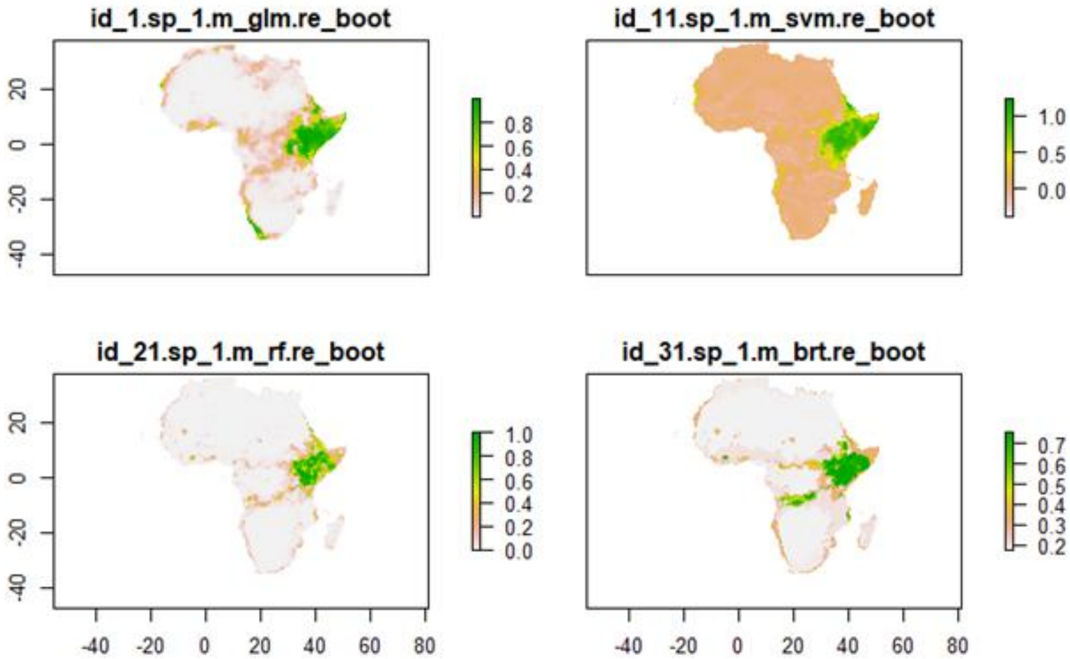

Figure9: Individual SDM models prediction map

The map (figure 9) shows the model prediction result from the individual models included in SDM. The models are generalized linear model (GLM) top right, support vector machine (SVM) top left, random forest (RF) bottom right , and boosted regression trees (BRT) bottom left. The result showed a typical difference in prediction capability in the models. The following model, BRT and GLM, had shown more prediction than the other two model types (SV and RF). A small patch of area suitable for locust invasion was predicted in GLM and BRT model, particularly in the border of Southern Africaand Democratic Republic of Congo.

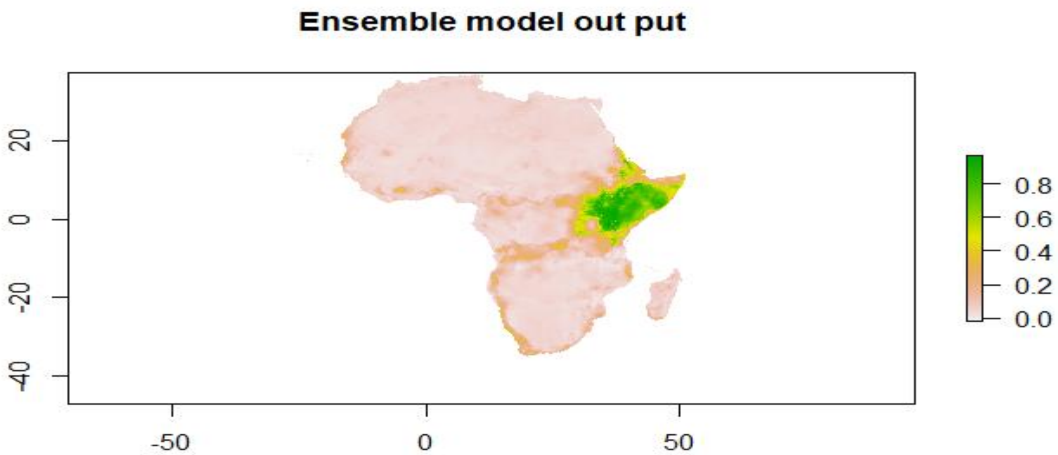

Figure 10: Predicted locust invasion map from ensemble model for 2020

The above figure represents the model output by using the ensemble approach. The ensemble approach uses the weighted average of the individual model output and finally produces one map from the several outputs obtained from the model. The result revealed most of the East African countries had experienced high locust infestation in the year 2020. During this period the highest locust attack was observed in Somalia, Kenya, South Sudan, and Southern Ethiopia. There is moderate area suitability for locust invasion in the countries such as Eritrea and Uganda.

Table 2: Environmental variables used for modelling the potential invasion of desert locust

Label	Predictor Variables	Units
bio1	Annual mean temperature	Degree Celsius
bio3	Isothermality	Degree Celsius
bio4	Temperature Seasonality (standard deviation ×100)	Degree Celsius
bio5	Max Temperature of Warmest Month	Degree Celsius
bio7	Temperature Annual Range (BIO5-BIO6)	Degree Celsius
bio9	Mean temperature of driest quarter	Degree Celsius
bio10	Mean Temperature of Warmest Quarter	Degree Celsius
bio12	Annual precipitation	Millimeter
bio13	Precipitation of Wettest Month	Millimeter
bio15	Precipitation Seasonality (Coefficient of Variation)	Millimeter
bio18	Precipitation of Warmest Quarter	Millimeter
bio19	Precipitation of Coldest Quarter	Millimeter

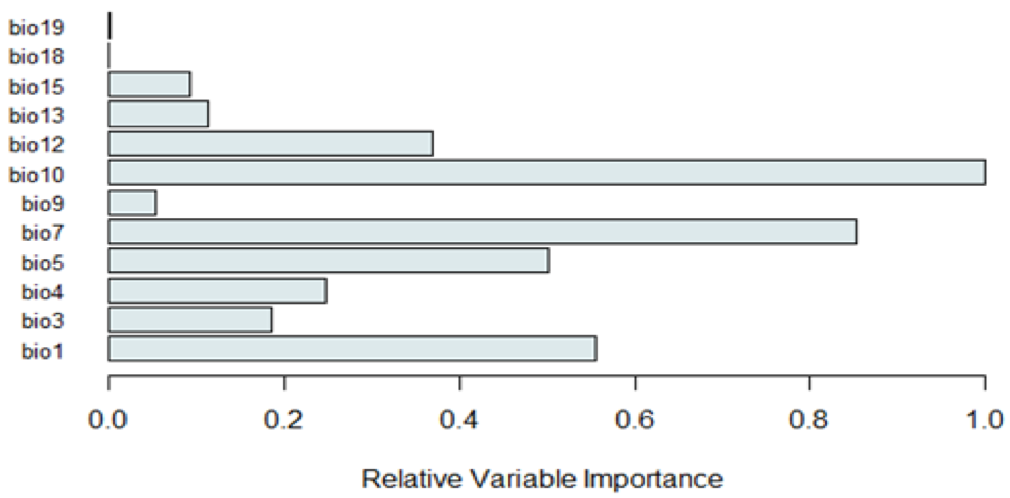

Figure 11: The relative importance of different environmental variables in the model

The relative contribution of environmental variables included in the SDM model is shown in the above bar graph(figure 11). The result revealed that bio10 (mean temperature of warmest quarter), bio7 (Temperature annual range), and bio1 (annual mean Temperature) have the greatest environmental variables favoring the occurrence of locust invasion. However, bio19 (precipitation seasonality), and (bio18) precipitation of the warmest quarter showed the least variables favoring locust attack.

V. CONCLUSION

Overall, the result revealed the potential of machine learning approaches for accurate estimation and predicting the location of possible locust invasion by fusion and analyzing different biophysical and satellite remote sensing products (NDVI).

In this study, the integration of machine learning is probably the most important and innovative part of the approaches. It is known that machine learning algorithms are robust to handle big data size and better results compared to other approaches, in this regard, the input dataset and quantity of information used here are bigger in size, and without it the analysis cannot be done.

The ongoing climate change and erratic weather patterns have not only tilted the scales for the environment but also for the farmers who are trying to play catch up with the new climatic changes.The results from the two approach permits us to understand what happened and the way forward on how to prevent the future invasion of desert locust. These data are important for the decision makers to get awareness quickly of the potential areas affected by the swarms as well as their breeding areas.

ACKNOWLEDGEMENTS

We would like to thank the Plan4All and SmartAfriHub Hackathon Challenge 4 team for their useful comments and suggestions on our manuscript.

REFERENCES

[1]. Alvarado-Serrano, D. F., & Knowles, L. L. (2014). Ecological niche models in phylogeographic studies: Applications, advances and precautions. Molecular Ecology Resources, 14(2), 233–248. https://doi.org/10.1111/1755-0998.12184
[2]. Baker, R., Sansford, C., Jarvis, C., Cannon, R., MacLeod, A., Walters, K., 2000. The role of climatic mapping in predicting the potential geographical distribution of non-indigenous pestsunder current and future climates. Agric., Ecosystem Environ. 82,57–71.
[3]. Dupe Dorcas (2020). Advancing remote sensing for food security. African Space Industry Annual Report 2020 edition
[4]. Landi, M., Giovanni, B., 2016. Protecting crop species from biotic and abiotic constraints in the era of global change: are we ready for this challenge? Am. J. Agric. Biol. Sci. 11, 51–53.
[5]. Naimi, B. (2015). On Uncertainty in Species Distribution Modelling (PhD dissertation). University of Twente,ITC. https://doi.org/10.3990/1.9789036538404
[6]. Scherm, H., Sutherst, R., Harrington, R., Ingram, J., 2000. Global networking for assessment of impacts of global change on plant pests. Environ. Pollution. 108, 333–341.
[7]. Srivastava, V., Lafond, V., &Griess, V. C. (2019). Species distribution models (SDM): Applications, benefits and challenges in invasive species management. CAB Reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources, 14(December). https://doi.org/10.1079/PAVSNNR201914020
[8]. http://www.fao.org/africa/news/detail-news/en/c/1260476/
[9]. https://www.unocha.org/east-africa-locust-infestation
[10]. https://www.theguardian.com/global-development/2020/apr/13/second-wave-of-locusts-in-east-africa-said-to-be-20-times-worse