

**DEVELOPMENT OF THE SSCM WEB AND MOBILE APP (CROMAP) USING AI AND MACHINE LEARNING**

1. **Introduction**

The JKUAT-KSA Small Scale Crop Mapping Project is a research plan to improve food security in Kenya, through boosting the production of small holder farms in the country. Access to safe and nutritious food is not a guarantee for people living in drought-prone areas such as Marsabit, West Pokot and other similar areas. Small holder farms are the basis of Agriculture in Kenya, with physical size ranging from 0.2 to 0.5 ha and 2 to 5 ha. Agriculture contributes to about 16% the Gross Domestic Product in Kenya. Improving decision making in this farms is an ultimate solution to food security.

Agricultural remote sensing can be used to monitor the status of a plant by observing the colors of its leaves or the general appearance of the plant. Remotely sensed photos from satellites provide a way to analyze field conditions to assess nutrient shortages, illnesses, water deficiency, weed infestations, among others. Remote sensing data can be used to create base maps for variable rate fertilizer and pesticide applications. Farmers can treat only the afflicted sections of a field using information from remotely sensed photographs. Problems in a field can be discovered remotely before they can be identified visually.

The project is aimed at integrating satellite imagery acquisition and processing skills to perform crop mapping and generate crop performance statistics to assess and develop a solution to improve food security in the country. The project will involve the development of a system to automatically acquire, process and classify satellite imagery using AI and ML for an information based decision making. A web and mobile app have been developed to provide users with various levels of expertise with a dashboard for accessing satellite imagery products, and provide equal access to technology and information for improved decision making.

* 1. **Vegetation Indices**

Several remote sensing indices have been used in the SSCM project to analyze crop performance and progress for the study area. Each vegetation index is intended to highlight a different aspect of the crops. For example, MSAVI is useful during the planting/ germination stage. NDVI is used to determine crop health, while NDRE is useful in assessing plant’s canopy density and greenness in the lower layers of the crops. NDRE factor range between 0.2 and 0.6 indicates an immature crop or sickly crop due to water stress. Values above 0.6 indicates healthy, mature, and ripening crop (Ceccato et al. 2001).

Since these indices may not quickly detect any stress in the crop, NDMI is useful a useful tool in quick detection of changes in the moisture content of the leaves, which correlates to water stress in crops. Low NDMI values indicate water stress, while figures above 0.4 indicate moisture adequacy (Jensen 2004). MNDWI is used to distinguish open water sources from dry land. MNDWI Values greater than 0.5 indicates presence of a water body.

Remote sensing indices used in the SSCM Project:

* Normalized Difference Vegetation Index (NDVI)
* Normalized Difference Red Edge Index (NDRE)
* Normalized Difference Moisture Index (NDMI)
* Red Edge Chlorophyll Index (ReCl)
* Modified Normalized Difference Water Index (MNDWI)
* Modified Soil Adjusted Vegetation Index (MSAVI)
  1. **Landsat 8 Band Combinations used in the SSCM project**

These combinations include natural color, healthy vegetation, land/water, Shortwave infrared, vegetation, false color and agriculture. They are summarized below.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Combination** | **Parameters** | | | | | |
|  | **NIR** | **SWIR1** | **SWIR2** | **Red** | **Green** | **Blue** |
| Natural Color |  |  |  | 4 | 3 | 2 |
| Color Infrared | 5 |  |  | 4 | 3 |  |
| False Color |  | 6 | 7 | 4 |  |  |
| Agriculture | 5 | 6 |  |  |  | 2 |
| Land/ Water | 5 | 6 |  | 4 |  |  |
| Shortwave Infrared Analysis | 5 |  | 7 | 4 |  |  |
| Healthy Vegetation | 5 | 6 |  |  |  | 2 |

1. **Mission background and Problem Statement**

The second goal of Sustainable Development is having reliable food production systems. Small scale farms support the livelihood of people in many regions, but are usually under-productive and at risk of experiencing losses due to adverse climatic conditions and natural hazards including pest outbreaks. They are usually equipped with little background information regarding such hazards and information about the crop types that suit certain soil properties, achievement and maintenance of maximum crop yield, and the nature of the local and international market for their produce. With the current agricultural practices such as reliance on unpredictable rains, millions continue to go hungry. The prevalence of the COVID-19 pandemic has also affected agriculture and jeopardized food security significantly. There arises a need to boost food production through agricultural remote sensing and Machine Learning for informed

Maximizing crop production using artificial intelligence involves efficient farming through informed decision making. The development of AI and the advancement of sensors can give real-time alerts to unusual conditions and plant stress. Indices such as NDVI can be calculated where lower values show crop areas under water stress or pests. Maps can be plotted to show affected areas and come up with cost-effective treatment for only the affected areas.

The JKUAT team seeks to combine Orthodox agricultural indices with space-based technology, Agricultural expert knowledge and AI/ML algorithms to help small-scale farmers to maximize food production with reduced farm inputs. To improve food security and the access to adequate safe and nutritious food, there is need to promote and support sustainable agriculture, small scale farmers and equal access to land, technology and markets (FAO Assembly,2015). Quality data, Data availability and usability is required.

1. **Research objectives for the SSCM Project**
   1. **General Objective:**

To promote food security by developing a user friendly tool to assess crop performance and promote information based decision making to boost farm productivity.

* 1. **Specific objectives:**

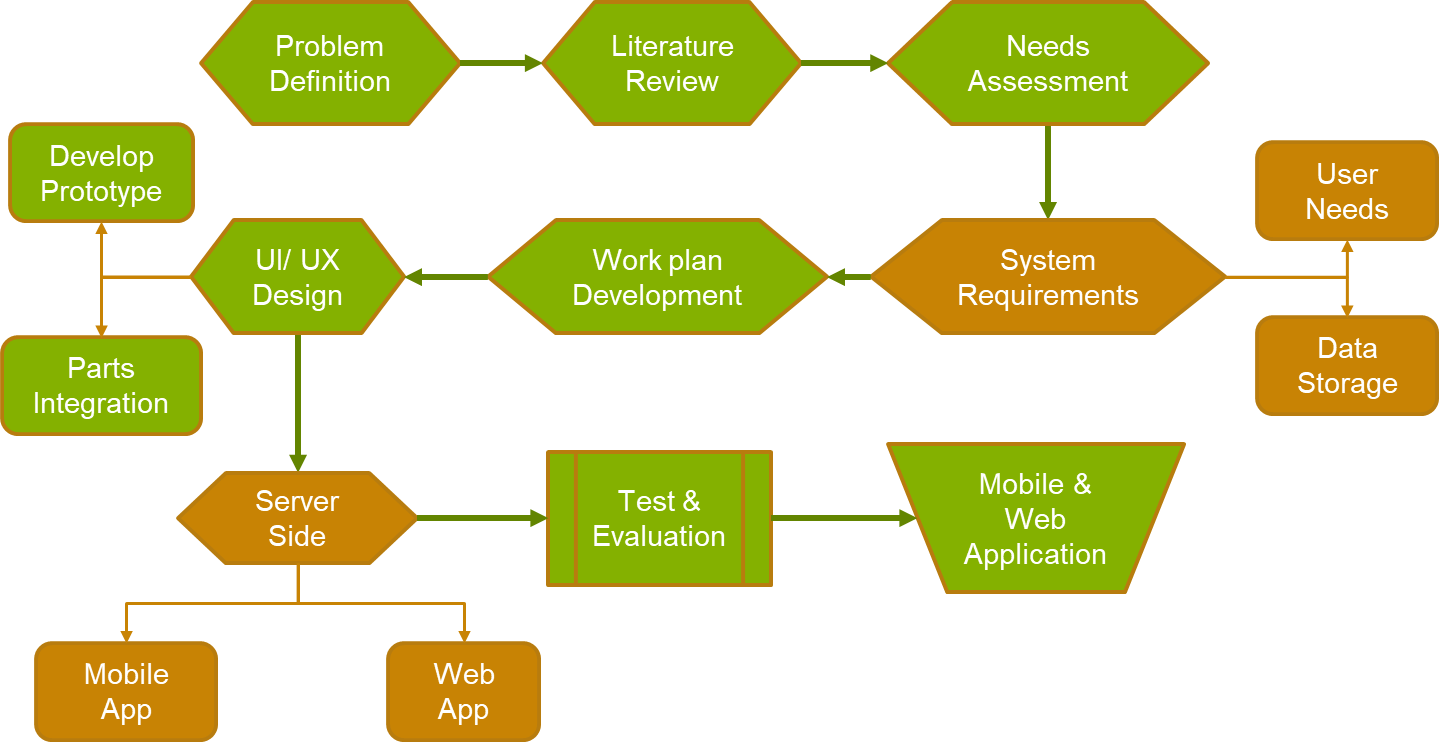
To develop AI/ML algorithms to map land cover, analyze crop performance and estimate crop yields.

To build a user friendly mobile and web application to perform near real-time crop type classification and assess crop performance to reduce production costs

To equip farmers with information for better decision making: weather and climate information to inform farm activities such as planting, top dressing and harvesting, and market information to prevent losses in times of glut.

1. **Project Outline**

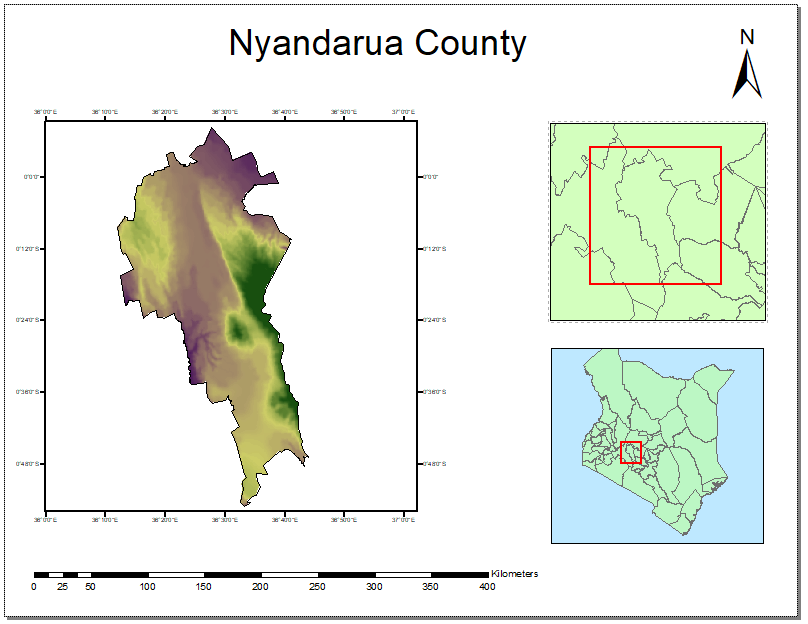
The steps towards developing a working crop mapping and analytics system (mobile application, web application and server side technologies) are presented below.



1. **Problem definition:** food security and information requirements
2. **Literature review and research:** reviewing existing farming applications such as <cropmonitoring.eos.com>, OneSoil app, Plantix and Agroptima
3. **User Needs assessment:** system requirements, user needs (farmers, GIS experts and policy makers), data and storage requirements.

* **Farmer Information Needs Assessment**
* Weather forecast and history: climate hazards, and weather patterns
* Market Information to avoid losses in times of glut.
* guidance on precise application of farm inputs to minimize production costs.
* Pest and disease infestation and control measures: precise application of pesticides.
* Education on best farming practices
* Informed decision making
* **Policy Makers**
* road networks, climate information and access to farms by agricultural extension officers
* Access to vast amounts of crop performance statistics for larger regions such as counties
* Highly processed visual data such as line graphs and charts to follow up on crop, soil and climate properties records
* Securing the long-term viability of the environment.
* **GIS Experts**
* Data collection and acquisition: high resolution satellite imagery.
* Data analysis, viewing vegetation indices, and climate properties.
* Extracting highly specific data from satellite images: calculate vegetation indices, process rainfall and climate data into simplified formats, etc)

1. **Establishing a work plan:** Dividing the project into GEE, App development and Research segments. A Gantt Chart was developed to guide project timelines.
2. **Developing the system model:** UI/UX designs and development of the actual product prototypes.
3. **Testing, Evaluation Refinement and Validation**: system evaluation to fulfill the TOR requirements, ground validation of the data from field work and user experience assessment from farmers’ interaction with the app prototypes.
4. **Finished product:** final web and mobile app.
5. **Results Presentation and Adoption:** presentation to the Kenya Space Agency.
6. **Materials and Methods**
   1. **STUDY AREA**



Nyandarua County is located in Central Kenya with a population of about 596,268 people and an area of 3,304 km2. The county lies between 0o 09’ N and 0 o 57’ S of the Equator and 36o 06’E and 36 o 52’ E of the Prime Meridian. Nyandarua has five electoral constituencies, including Kinangop, Kipipiri, Ndaragua, Ol Jorok and Ol’ Kalou. The latter serves as the county capital. Ol Kalou, Engineer and Ndunyu Njeru are the most popular urban centers in the county.

Nyandarua was a suitable choice for a study area because it is famous for agricultural productivity. Farming is the main economic activity, involving crop cultivation and dairy farming. The county’s climatic conditions are favorable for the growth of a wide variety of crops, with average annual temperatures of 23oC – 28oC, and well distributed rainfall of about 120mm per year. Potatoes, peas, cabbages, onions and carrots among other crops are grown in the region. Farming is mostly done on small scale, with farm sizes ranging from ¼ to 3 acres. The soils are generally deep well drained loam soils. Black cotton soils are also used for farming. Unreliable rains and insufficient water for irrigation is a major challenge affecting farmers in the region. Farms on the windward sides of the hills have access to water for irrigation, and therefore they do not depend on rain to carry out farming activities.

* 1. **Materials**

Google Earth Engine JS APIs

Google Earth Engine Code Editor

Google Earth Pro

React JS and React Native

Mongo db

* 1. **Data**

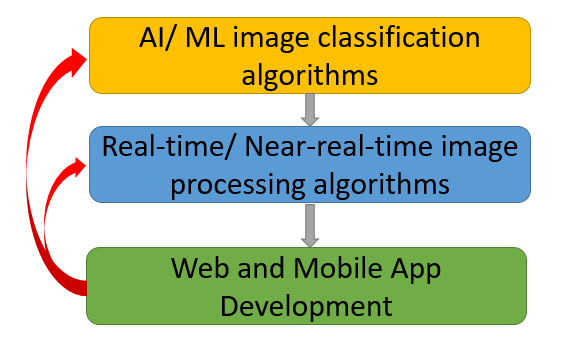
Landsat 8 imagery collection

Sentinel 2 imagery collection

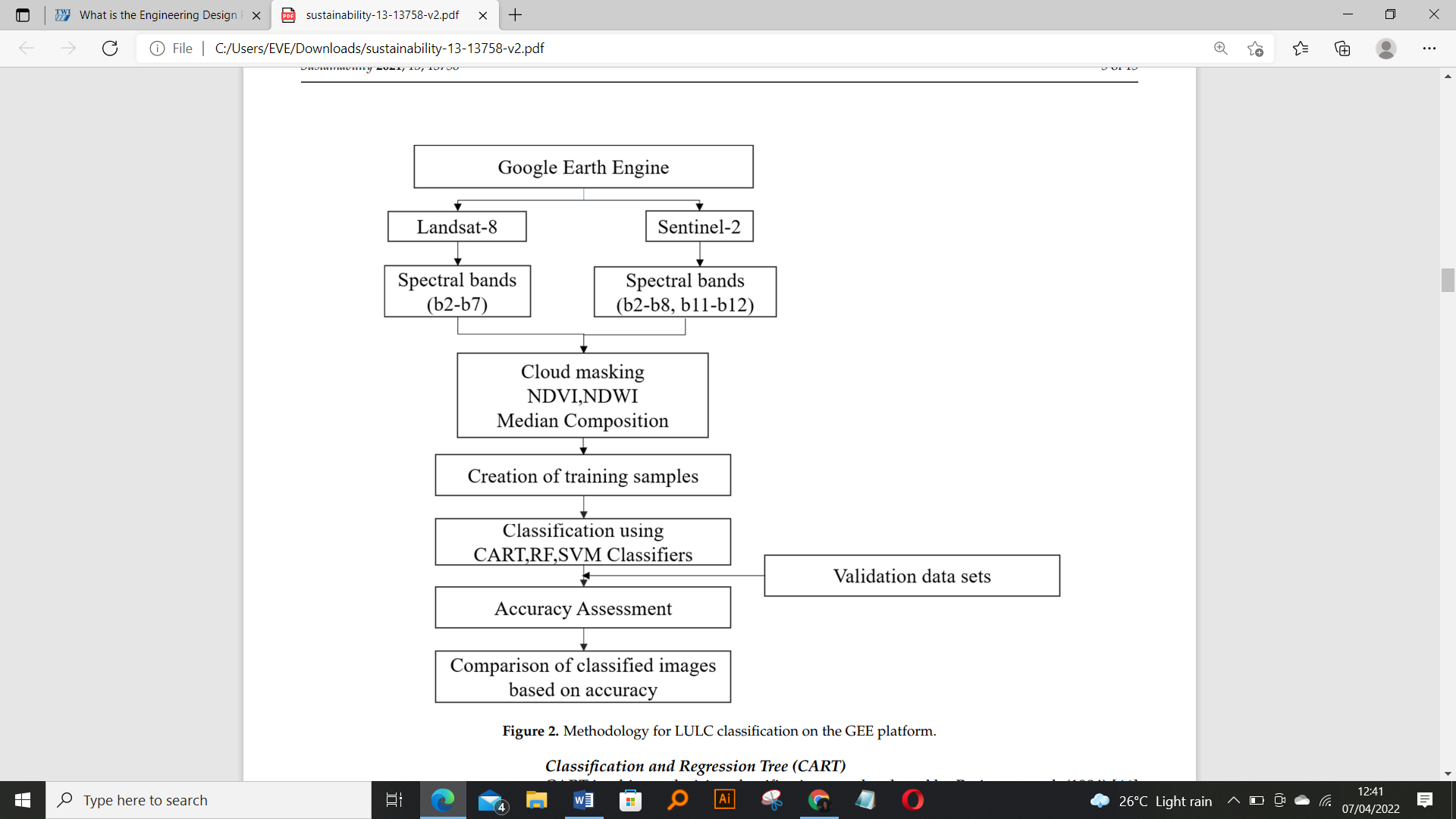
Validation data from Google Earth Pro

Field work data for ground trothing

1. **Methodology**



* 1. **Google Earth Engine, Classification and Classifiers**



Google earth engine is a platform that contains a public data archive, with Landsat and Sentinel imagery for a period of 30 years. It has JavaScript and python APIs, which enable users to utilize Google Cloud and perform complex Geospatial analysis tasks within very short time. The web-based code editor has been useful in the development of interactive classification algorithms within the SSCM project, with direct access to the vast Landsat and Sentinel Image Collections. Scripting on GEE was geared towards supervised classification of imagery for a time series between January 2020 to January 2021, isolation of cropland and performing crop type classification for integration into the mobile and web app functions. Calculation of indices such as NDVI, NDRE and MSAVI have been used to monitor in crop health statistics and crop yield pattern. Graphs have also been generated to show these trends for the specified time period. Procedures within the GEE platform are represented below:

* Data and Image acquisition: importing satellite imagery and relevant data
* Image preprocessing: cloud masking, mosaicking, sub setting, filtering by cloud cover, date, metadata and bounds and setting visualization parameters
* Calculation of indices: (NDVI, NDMI, MNDWI, NDRE, ReCl, MSAVI) for Sentinel 2 and NDVI, NDMI, MNDWI and MSAVI for Landsat 8.
* Generation of Time series graphs for the indices
* collection of training data on GEE through random sampling.
* Application of classifier algorithms: Random Forest, Support Vector Machine, CART
* Supervised classification
* Validation using data from Google Earth Pro and field work data.
* Accuracy Assessment
* Application of buffers.
* Area computations: Area by AOI (Area of Interest) and Area by Land Use Land Cover class
  + 1. **Land Cover Classes used on the GEE Platform**

The following land cover classes were settled upon to perform supervised classification for the SSCM project.

1. Cropland: these are areas under cultivated crops.
2. Transportation: used to identify tarmacked roads from the rest of the land cover features
3. Grassland and Shrubs: includes areas covered by long grasses and scattered short trees.
4. Natural Forests: these are areas under natural forest cover, such as in the Aberdare Ranges.
5. Built up Areas: these are areas with buildings such as settlements or schools.
6. Bare Areas: These are areas without buildings or artificial structures, and with less than 4% vegetation cover.
7. Artificial Water Bodies: These are areas covered by man-made water reservoirs such as dams, canals and artificial lakes.
8. Natural Waterbodies: These are areas with natural water sources such as lakes, rivers and their tributaries.
   * 1. **Classifiers Used in The SSCM Project**
9. **Random Forest Classifier**

Each tree in a random forest ensemble is constructed using a sample selected with replacement from the training set. Furthermore, the optimal split is selected from all input features or a random subset of size max features when dividing each node during tree construction. The goal of these two sources of randomness is to reduce the forest estimator's variance. Individual decision trees do, in fact, have a lot of diversity and are prone to overfitting. Forests with injected randomness produce decision trees with decoupling prediction errors. By merging distinct trees, random forests reduce volatility. The variance decrease is frequently large, resulting in a superior overall model.

Eight classes were used to train each classifier. Each class was given two training sets, each of which had more than 40 points. Training samples and reflectance data for bands 1 through 7 were used to train the classifier. This resulted in a total of above 350 trees.

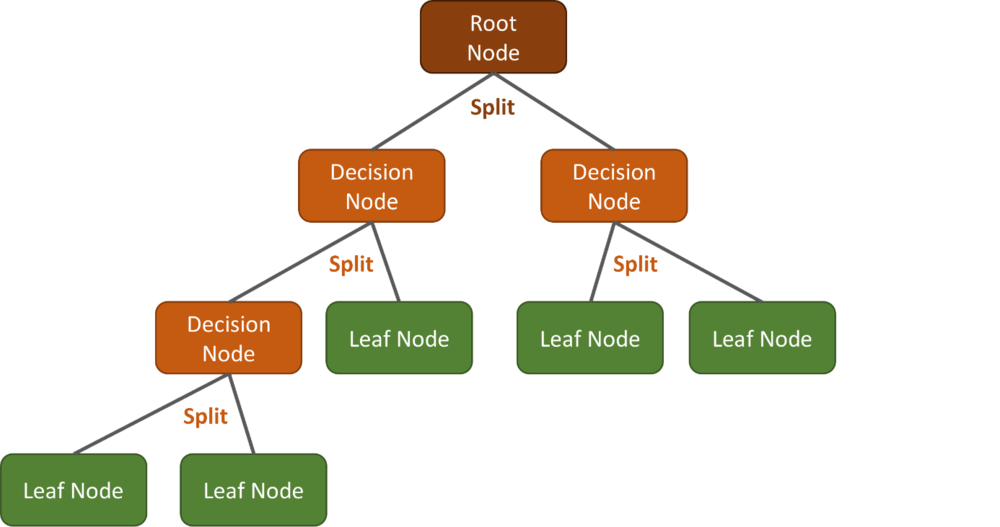
1. **Support Vector Machine Classifier**

Each data item is plotted as a point in n-dimensional space (where n is the number of features you have), with the value of each feature being the value of a certain coordinate in the SVM algorithm. Then, we accomplish classification by locating the hyper-plane that best distinguishes the two classes.

SVM classifier works exceptionally well with a clear margin of separation; it is effective in high-dimensional spaces; it is effective when the number of dimensions exceeds the number of samples; and it uses a subset of training points in the decision function, making it memory efficient.

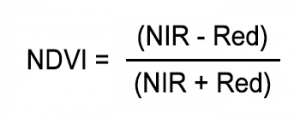
1. **Classification and Regression Tree Classifier**

A Classification and Regression Tree (CART) is a machine learning predictive technique. A Classification and Regression Tree (CART) is a predictive model that describes how the values of an outcome variable can be predicted based on previous values. A CART output is a decision tree, with each fork representing a split in a predictor variable and each end node representing a forecast for the outcome variable. This classifier process can be summarized in the diagram below.



* + 1. **Calculation of Indices used in the SSCM Project.**
* **Normalized Difference Vegetation Index**

It is used to estimate vegetation health by measuring the difference between the NIR band and Red light in the electromagnetic spectrum. NDVI varies from -1 to +1. Extremely low values of NDVI correlate to rocks, sand, or snow. Moderate levels (0.2 - 0.3) depict shrubs and meadows, while larger values (0.6 - 0.8) represent temperate and tropical forests. (Yang and Everitt 2011). NDVI calculated as shown below;



* **Normalized Difference Red Edge Index**

NDRE is a method for determining the quantity of chlorophyll in plants. It is computed from the NIR band and the Red Edge range between visible Red and NIR. The scale goes from bright red at -1 to saturated green at +1. NDRE is most effective when crops are mature or ripening. Values (0.6 – 1) are used to represent this. This happens near the end of the growing season, when levels below 0.6 indicate crop loss. At later growth stages, employing an NDRE map for variable-rate fertilization, spraying, irrigation, fertilization, and other field activities is more successful than NDVI. NDRE is calculated as shown below:

*NDRE = (NIR — Red Edge)/ (NIR + Red Edge)*

* **Normalized Difference Moisture Index**

It uses NIR and SWIR bands to display moisture in vegetation. NDMI value ranges from 0.685 to - 0.154. It can be computed as shown below:

*NDMI = (NIR - SWIR1)/ (NIR + SWIR1)*

* **Modified Normalized Difference Water Index**

MNDWI is used to distinguish water from dry land. In the visible and IR wavelength ranges, water bodies have minimal radiation and high absorbability. MNDWI value ranges from 0.146 to - 0.444. It is calculated as shown:

Sentinel-2 NDMI = (NIR - SWIR) / (NIR+ SWIR)

Landsat 8 NDMI = (NIR – SWIR1) / (NIR + SWIR1)

* **Modified Soil Adjusted Vegetation Index**

This index has low saturation value of 0.3. It is used when growing crops on the field for the first time or at a different elevation. It is also used during the seed germination and leaf development stages to detect uneven seed growth and monitor vegetation health. It ranges from -1 to 1, where 0.4 to 0.6 shows leaf development and above 0.6 apply NDVI. MSAVI is calculated as shown below:

MSAVI = ((NIR – RED) / (NIR + RED + L)) \* (1 + L)

Where L is the soil brightness correction factor.

* 1. **Fieldwork and Data Validation**

The field work exercise started with random selection of points from the initial classification data and a virtual reconnaissance on Google Earth Pro. Attribute information about the growth and progress of each crop and farming methods used was collected by sampling different farms, conducting farmer interviews and recording the information gathered on mobile phones and tablets. The SW maps application was particularly useful in the creation, management and collection of GIS data within the “Kinangop Project”, developed within the app. The SW Maps application used in SSCM fieldwork project is a free GIS and mobile mapping app that allows one to collect, store, present, and share geographic data. The team was able to capture, store, update, manipulate, analyze, and display geospatial data and information in the field. GPS enabled tablets were used to pick GPS points throughout the fieldwork exercise and collection of validation data. More validation data was also collected from Google Earth Pro.

1. **Web and mobile App development side**

**System Description**

Three parts of the applications were developed: front-end, back-end and the middleware (APIs). The APIs connect the front to the back.

**The Front-End**

The front-end is the interface where the client interacts with the system. It includes a mobile app and a web-application. The clients can use either of the platforms interchangeably. The front end collects, presents and does simple calculations on data. It also collects user passwords, location, commands and inputs. Data presented on the interface are results for GIS functions performed on the cloud.

The mobile app was written in React Native, while the web application was written in React Js in JavaScript. They applications are aesthetically and functionally equivalent. Both run on node.js. The mobile package was hosted on google playstore while the web package is hosted on heroku.com servers and may also be deployed on the Google App Engine.

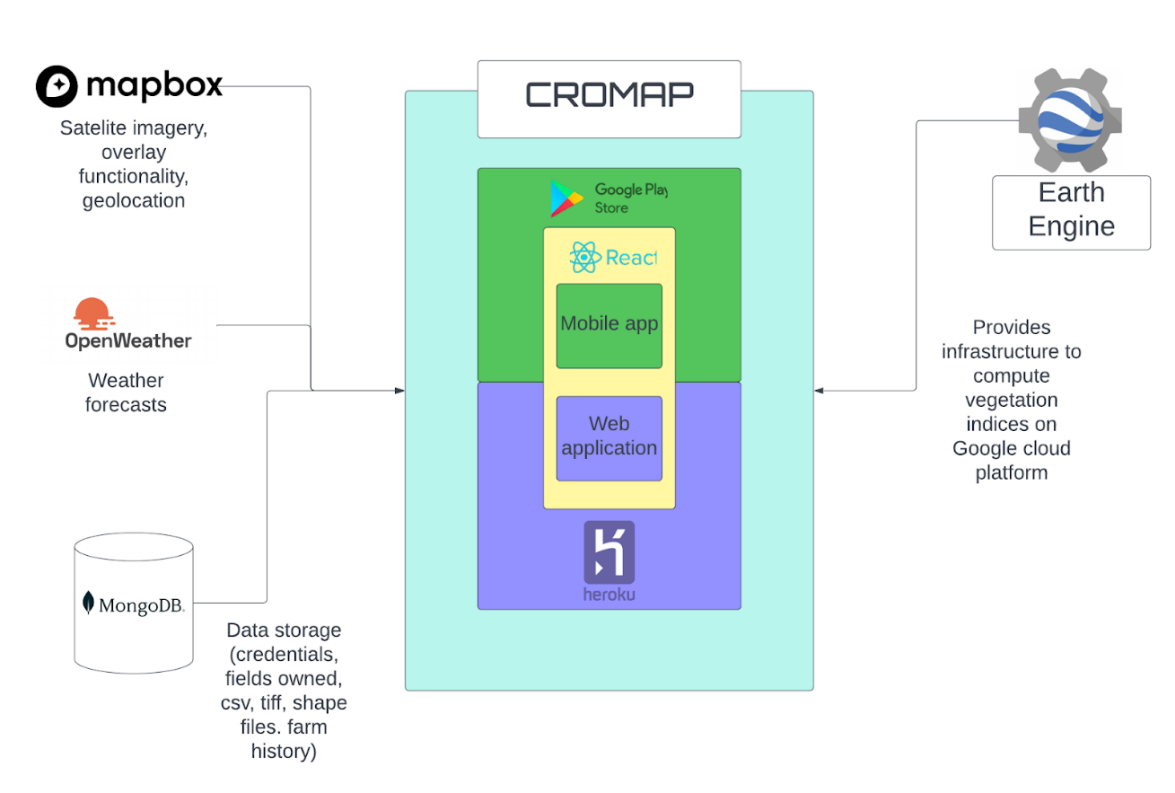
On the front-end, the following APIs were used to bring the interface together:

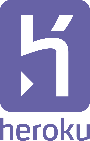
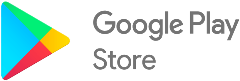
1. Mapbox.com to provide the scrollable map on the home page Mapbox provides maps for developers.
2. Openweathermap.org to provide the weather forecast feature.
3. Bugsnag to track errors, bugs and general analytics for the mobile platform.

**The Back-End**

The back-end is where the complex GIS tasks were performed and the system’s data computed. For example, calculation of indices and crop type classification. Google Earth Engine was used to perform these tasks. Although the scripts are run from the client-side, the data is ingested and manipulated on the cloud without having to go through a user’s device. Only the required results are sent to the client.

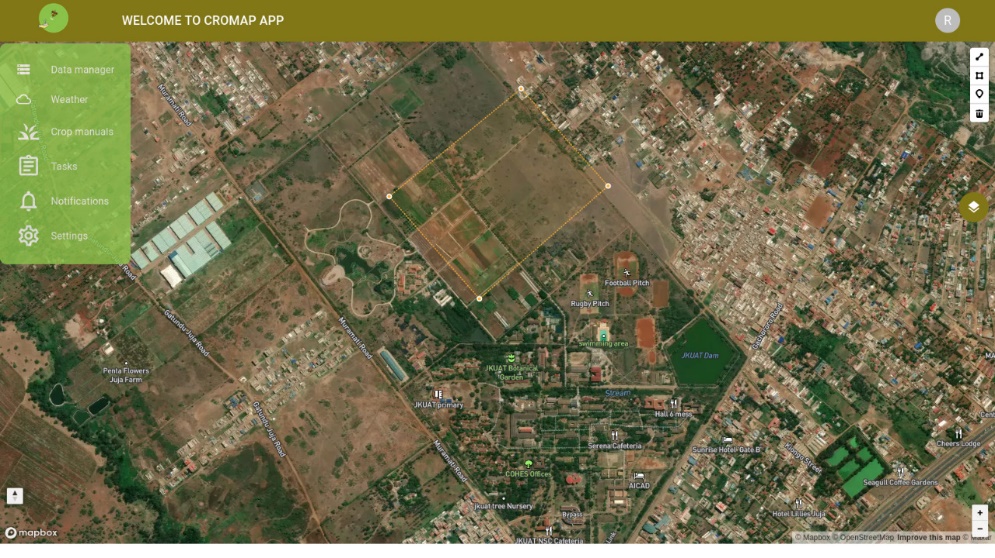
Data is stored on the Mongo DB platform (Mongodb.com). Mongo db is a documented oriented database. Data is stored in form of objects with instances, which may be related to other instances from the same or different objects. Most data in the SSCM project is in GeoJson format.

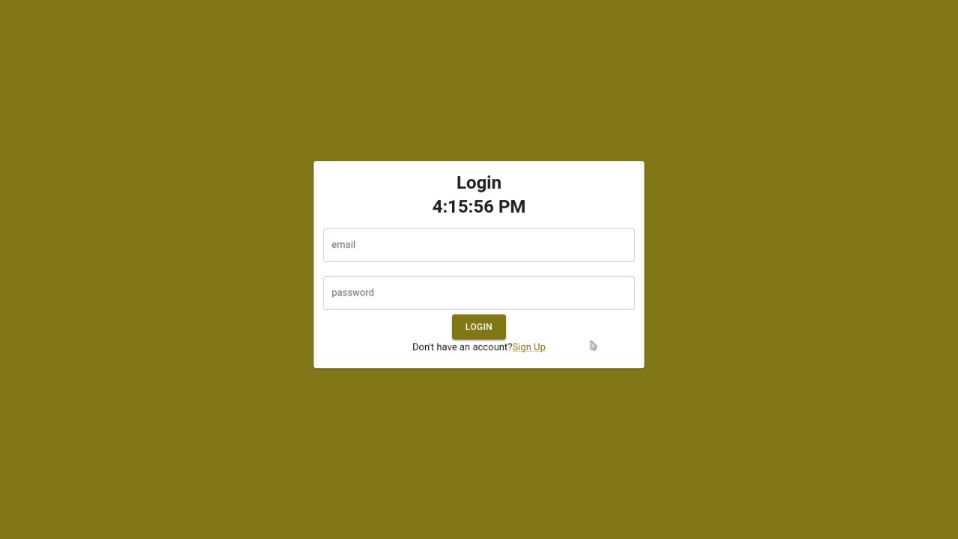


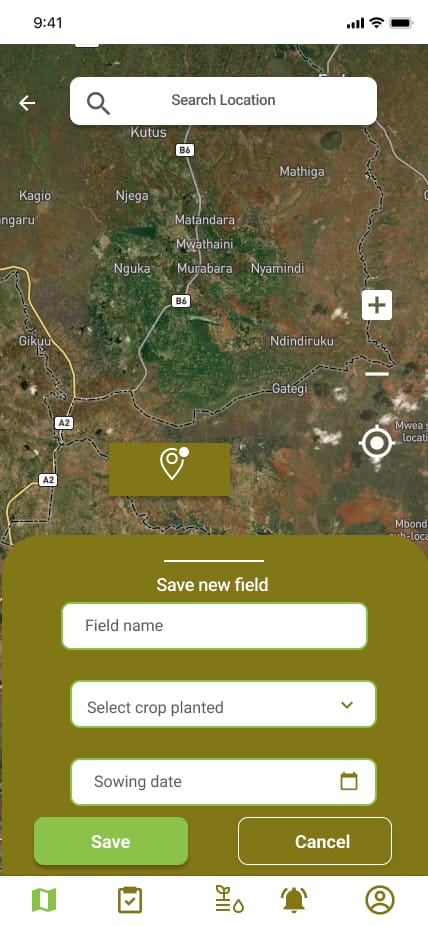
    

 https://lh5.googleusercontent.com/eCKIatf_m-1L-QhiM01nP7wqE49Q_fZbeJMSVUts5Za88CVqr3VFhkPFTLp05fc75A4_1td_q9sIXACNl1Yj-KxPd9eOfrGtaC37Vz1AflfvAZ3JxyVNovOyt8zbp0SKHkUqK8pH 

1. **Results (Mobile and Web Apps)**







1. **Challenges**

* Large scope of project. Farmers, GIS experts and Agricultural experts have different user needs that could not be covered by developing one system within the given time period.
* Before the fieldwork exercise, we lacked firsthand knowledge on what will be useful to a farmer on the ground.
* Crop Classification: The Machine Learning algorithms only classified the field-derived crops. It could not account for all crop types grown in Kenya
* Yield Estimation was challenging, given the varied farming practices of the small scale farmers.
* Farmers did not keep records of all the farm inputs and crop produce outputs from their farms. They also lacked standard measurement units for the yields, which made it difficult to unify the data collected during field work.
* The team could not do data validation using UAVs, due to the restrictions by the KCAA to operate the equipment over the area

1. **Recommendations**
   * + The time frame given to work on a similar project with the same scope should be increased to provide more efficient outputs.
     + User needs assessment is a crucial component of development. There should be interaction with the users as development continues.
     + Further research on crop classification should be carried out.
     + Crop variety, irrigation conditions, soil conditions, weather conditions and management practices are among the factors influencing crop yields. Further analysis should be carried out to enhance the efficiency of yield estimation.
     + The government should provide agricultural extension services to farmers, to advise them on how to keep proper records, and practice good farming methods to boost crop production.
     + The government should involve farmers in formulating farming policies to avoid imposing rules that affect the quality of their crops or send off buyers.
2. **Conclusion and Outlook**

The JKUAT-KSA Small Scale Crop Mapping Project was a great learning experience for the project team members. The team was able to fulfill the following requirements from the KSA terms of reference:

To develop AI/ML algorithms to map land cover, analyze crop performance and estimate crop yield using Google Earth Engine.

To build a mobile and web application with a few of the required functionalities: a user-friendly UI interface, functions for drawing fields, calculating NDVI, creating tasks for each field and managing their farm activities through the applications.

With continued support from the Kenya Space Agency, the team would be able to do the following:

Analyze the farmers needs more efficiently, in order to add functions that would be more useful to the farmers, through continuous monitoring during field work.

Connect the GEE side to the mobile and web applications for automatic classification, in order to handle client-server requests on demand.

Collect high resolution imagery using UAVs for more accurate crop type classification. The team was unable to fly UAVs in the study area, due to lack of the required licenses.

1. **Acknowledgement**

* We wish to thank the Kenya Space Agency for funding the Small Scale Crop mapping Project, the University of JKUAT for hosting the project, the Project lead Dr. E. Nduati for her continuous support and the student participants for the cooperation exercised so far.

1. **References**

Kustas, W. P.; Norman, J. M. (2009). "Use of remote sensing for evapotranspiration monitoring over land surfaces". Hydrological Sciences Journal.

Aguilar, R., Zurita-Milla, R., Izquierdo-Verdiguier, E., & A De By, R. (2018). A cloud-based multi- temporal ensemble classifier to map smallholder farming systems. Remote sensing, 10(5), 729.

Assembly, G. (2015). Sustainable development goals. SDGs Transform Our World, 2030.

Ben-Hur, Asa; Horn, David; Siegelmann, Hava; Vapnik, Vladimir N. ""Support vector clustering" (2001);" Journal of Machine Learning Research. 2: 125–137.

Bosc, P. M., Berdegué, J., Goïta, M., van der Ploeg, J. D., Sekine, K., & Zhang, L. (2013). Investing in smallholder agriculture for food security (No. 6). HLPE.

Cortes, Corinna; Vapnik, Vladimir N. (1995). "Support-vector networks" (PDF). Machine Learning. 20 (3): 273–297. CiteSeerX 10.1.1.15.9362.

Cortes, Corinna; Vladimir Vapnik (1995). "Support-Vector Networks". Machine Learning. 20 (3): 273–297.

Di Gregorio, A. (2005). Land cover classification system: classification concepts and user manual: LCCS (Vol. 2). Food & Agriculture Organization.

Gil, J. D. B., Reidsma, P., Giller, K., Todman, L., Whitmore, A., & van Ittersum, M. (2019). Sustainable development goal 2: Improved targets and indicators for agriculture and food security. Ambio, 48(7), 685-698.

Google Earth Engine Documentation and tutorials from <https://developers.google.com/earth-engine/tutorials/tutorials>

<http://gsp.humboldt.edu/olm/Courses/GSP_216/lessons/composites.html>

<https://gisgeography.com/landsat-8-bands-combinations>

<https://www.agck.or.ke/Downloads/NATIONAL-AGRICULTURE-INVESTMENT-PLAN-NAIP.pdf>

<https://www.knbs.or.ke/wp-content/uploads/2021/09/Economic-Survey-2021.pdf>

Huete, A.R (August 1988). "A soil-adjusted vegetation index (SAVI)". Remote Sensing of Environment.

Ibrahim, E. S., Rufin, P., Nill, L., Kamali, B., Nendel, C., & Hostert, P. (2021). Mapping Crop Types and Cropping Systems in Nigeria with Sentinel-2 Imagery. Remote Sensing, 13(17), 3523.

Joachims, Thorsten; "Transductive Inference for Text Classification using Support Vector Machines", Proceedings of the 1999 International Conference on Machine Learning (ICML 1999), pp. 200–209.

Kogan, F.N. (1995). "Application of vegetation index and brightness temperature for drought detection". Advances in Space Research. 15 (11): 91–100.

Liaw, A., Wiener, M. (2002): Classification and Regression by random Forest. Forest, 23.

Nduati, E., Sofue, Y., Matniyaz, A., Park, J. G., Yang, W., & Kondoh, A. (2019). Cropland Mapping Using Fusion of Multi-Sensor Data in a Complex Urban/Peri-Urban Area. Remote Sensing, 11(2), 207.

Pal., M. (2003): Random Forest Classifier for Remote Sensing Classification. In: International Journal of Remote Sensing, Volume 26, Issue 1. Pages 217-222.

Peters, A.J.; Walter-Shea, E.A.; Ji, L.; Vliia, A.; Hayes, M.; Svoboda, M.D. (2002). "Drought Monitoring with NDVI-Based Standardized Vegetation Index" (PDF). Photogrammetric Engineering & Remote Sensing. 68 (1): 71–75. Retrieved 16 May 2018.

Pinter, Jr., Paul J.; Hatfield, Jerry L.; Schepers, James S.; Barnes, Edward M.; Moran, M. Susan; Daughtry, Craig S.T.; Upchurch, Dan R. (2003). "Remote Sensing for Crop Management". Photogrammetric Engineering & Remote Sensing.

Zhang, X., Zhang, M., Zheng, Y., & Wu, B. (2016). Crop mapping using PROBA-V time series data at the Yucheng and Hongxing farm in China. Remote Sensing, 8(11), 915.