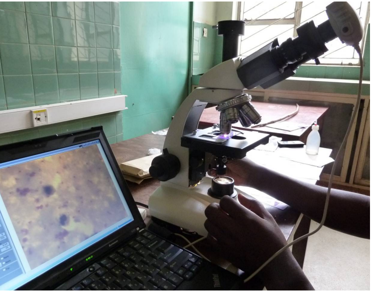
MLPR guest lecture: Machine learning and crop disease in Uganda

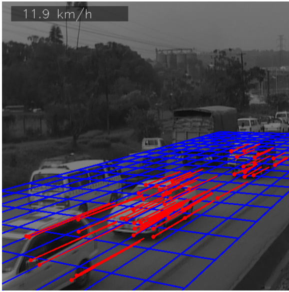
John Quinn

18.11.2016

Why machine learning in the developing world?

- ML can be used to automate the decision-making processes of scarce human experts.
- Can also be used to gather better information, e.g. about crop growth, infectious disease, resource usage.
- New deployment and data collection possibility: 84 million internet enabled phones in Africa.





Crop disease monitoring

Information about infectious crop diseases is vital in developing countries, to predict famine and plan interventions.


to stem devastating banana wilt

FEEDBACK EMAIL PRINT EASY READ SHARE

Photo: Kate Holt/IRIN


Uganda: Banana Wilt to Cause $4 Billion Loss

BY GRACE AGABA, 5 OCTOBER 2006

Kampala — UGANDA risks losing about $4b (sh7.4 trillion) to the banana bacterial wilt by 2010 unless efforts to prevent the fast-spreading disease are put in place, banana experts have warned.

The banana bacterial wilt spreads by drying banana plants. Researchers have discovered that the

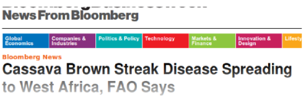

NEWS

New viral disease threatens region’s cassava farming

SHARE BOOKMARK PRINT RATING

Commercial cassava growers. Uganda scientists are currently doing research to understand the virus, and how to develop resistant varieties. Photo/FILE

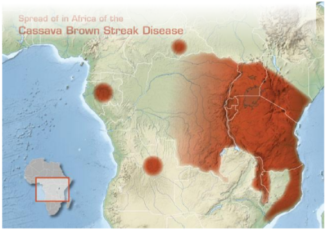
By HALIMA ABDALLAH


News From Bloomberg

Global Economics | Companies & Industries | Politics & Policy | Technology | Markets & Finance | Innovation & Design

Bloomberg News

Cassava Brown Streak Disease Spreading to West Africa, FAO Says




Home News

NEWS

Cassava farmers in Great Lakes facing double disease strike

SHARE BOOKMARK PRINT RATING

By ZAYNAB TURUKU

Posted Monday, November 23 2009 at 00:00

SHARE THIS STORY

Cassava farmers in the Great Lakes region are at risk of suffering losses following an outbreak of brown streak disease and cassava mosaic that are spread by white flies.

Crop disease monitoring

Information about infectious crop diseases is vital in developing countries, to predict famine and plan interventions.

- There is a shortage of agricultural experts to carry out diagnosis.
- There is scarce data about the spread of disease on a national scale.

to stem devastating banana wilt

FEEDBACK EMAIL PRINT EASY READ SHARE

Photo: Kate Holt/IRIN

Uganda: Banana Wilt to Cause $4 Billion Loss

BY GRACE AGABA, 5 OCTOBER 2006

Kampala — UGANDA risks losing about $4b (sh7.4 trillion) to the banana bacterial wilt by 2010 unless efforts to prevent the fast-spreading disease are put in place, banana experts have warned.

The banana bacterial wilt spreads by drying banana plants. Researchers have discovered that the

NEWS

New viral disease threatens region's cassava farming

SHARE BOOKMARK PRINT RATING

Commercial cassava growers. Uganda scientists are currently doing research to understand the virus, and how to develop resistant varieties. Photo/FILE

By HALIMA ABDALLAH

News From Bloomberg

Global Economics | Companies & Industries | Politics & Policy | Technology | Markets & Finance | Innovation & Design

Bloomberg News

Cassava Brown Streak Disease Spreading to West Africa, FAO Says

Home News

NEWS

Cassava farmers in Great Lakes facing double disease strike

SHARE BOOKMARK PRINT RATING

By ZAYNAB TURUKU

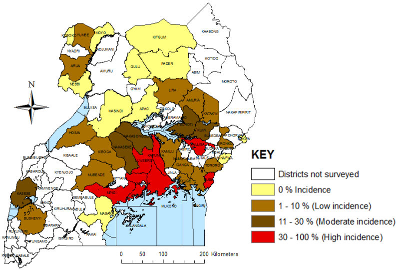
Posted Monday, November 23 2009 at 00:00

SHARE THIS STORY

Cassava farmers in the Great Lakes region are at risk of suffering losses following an outbreak of brown streak disease and cassava mosaic that are spread by white flies.

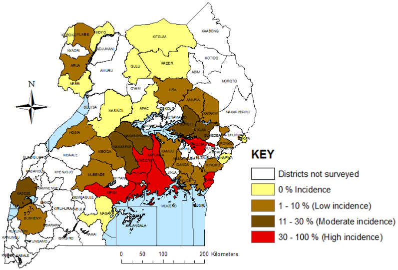
Crop disease monitoring

Current survey methods are expensive, limited, and slow: 2 months survey time + 3 months of data entry to produce a map:



Crop disease monitoring

Current survey methods are expensive, limited, and slow: 2 months survey time + 3 months of data entry to produce a map:



However, by collecting data on mobile devices instead of paper, real-time data can be obtained...


3:57 PM
ODK Collect > Cassava Surveillance Form
SEVERITY
CMD Severity
Cassava Mosaic Disease severity
One
Two
Three
Four
Five
3:50 PM
ODK Collect > Cassava Surveillance Form
PICTURE
Main Picture
Please take a snap shot of the cassava leaf
Take Picture
3:58 PM
ODK Collect > Cassava Surveillance Form
GEO Location
gps
Geographical Position System
Record Location
Date:Sunday March 27 2011 6:34:46 PM UTC
Disease Rate:CMD:4,CGM:5,CBS:1
GPS Accurracy:6.0

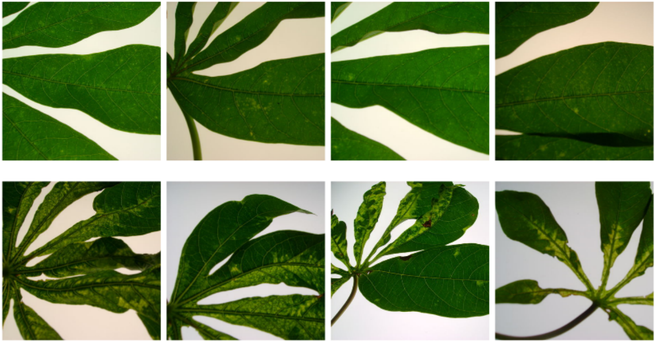
Overview

With real-time, digitised data, it's possible to use machine learning to achive some new results.

1. Automated diagnosis and symptom measurement.
2. Spatial and spatio-temporal modeling.

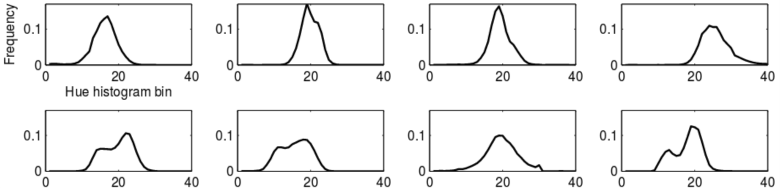
Image-based classification of disease

- One of the constraints on crop disease surveying currently is the scarcity of experts.
- When mobile data collection replaces paper-based surveying, camera-phones can be used to collect image data.
- If we can use these images to perform a diagnosis of disease automatically, then agricultural extension workers with basic training can carry out the survey.



Classification based on colour histogram

A simple way to do this is for each image to take instances $\mathbf{x}^{(i)}$ as histogram of hue values, used to predict the labels $y^{(i)} \in \{\text{diseased}, \text{healthy}\}$. Diseased leaves tend to have a bimodal hue distribution, where parts of the leaf affected by chlorosis add to the yellow range of the spectrum.



Normalised hue histograms of the leaf images (calculated from the corresponding images on the previous slide), with healthy plants on the top row, and those with cassava mosaic disease (CMD) on the bottom row.

Classification based on other features

Other possibilities for $\mathbf{x}^{(i)}$:

- Shape features – find contours in the images, and compute the areas, jaggedness, eccentricity.
- Local gradient descriptors.
- Image pixels directly – train convolutional nets to do the classification.

Disease classification: banana

With good features (discriminative and fast to compute) we can implement classification on a phone to handle these kinds of images. Different problems require different representations, e.g. banana leaves:

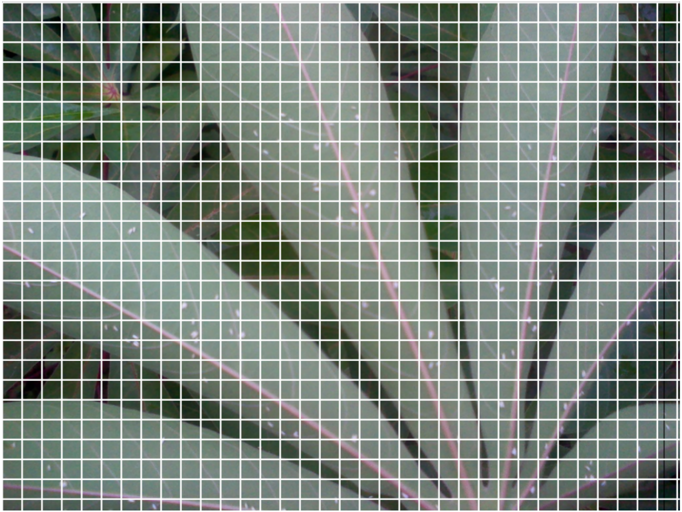
Left: healthy leaf, centre: banana bacterial wilt, right: black sigatoka disease

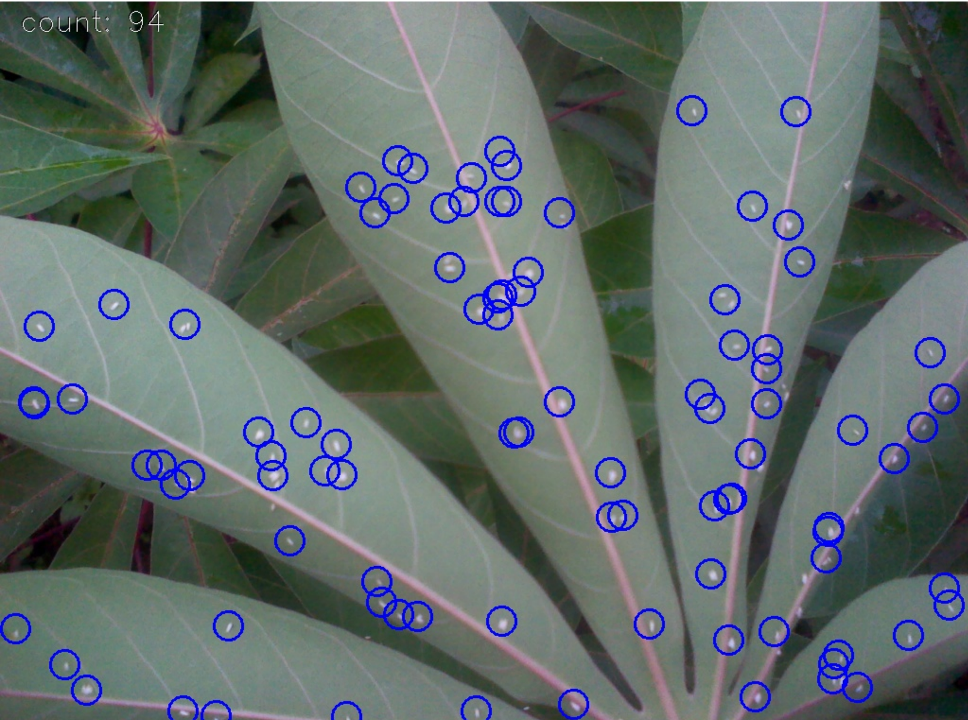
Classifying sub-images

We can do classification on parts of images too. To count the number of whiteflies in this image, cut it into a grid and classify each grid cell separately, $y^{(i)} \in \{\text{whitefly}, \text{no whitefly}\}$.

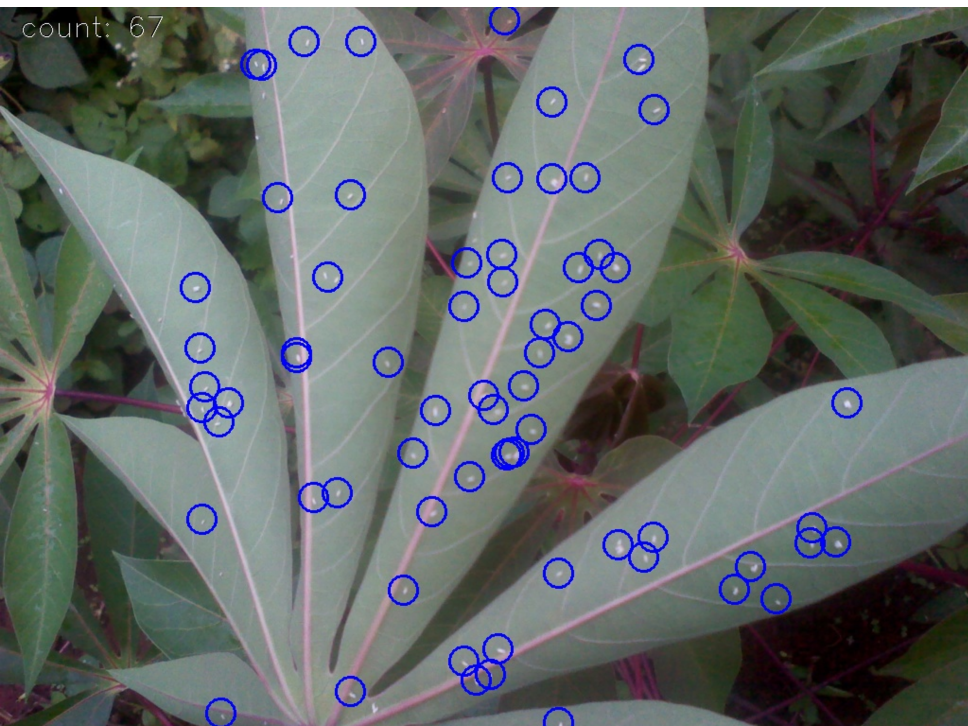
Classifying sub-images

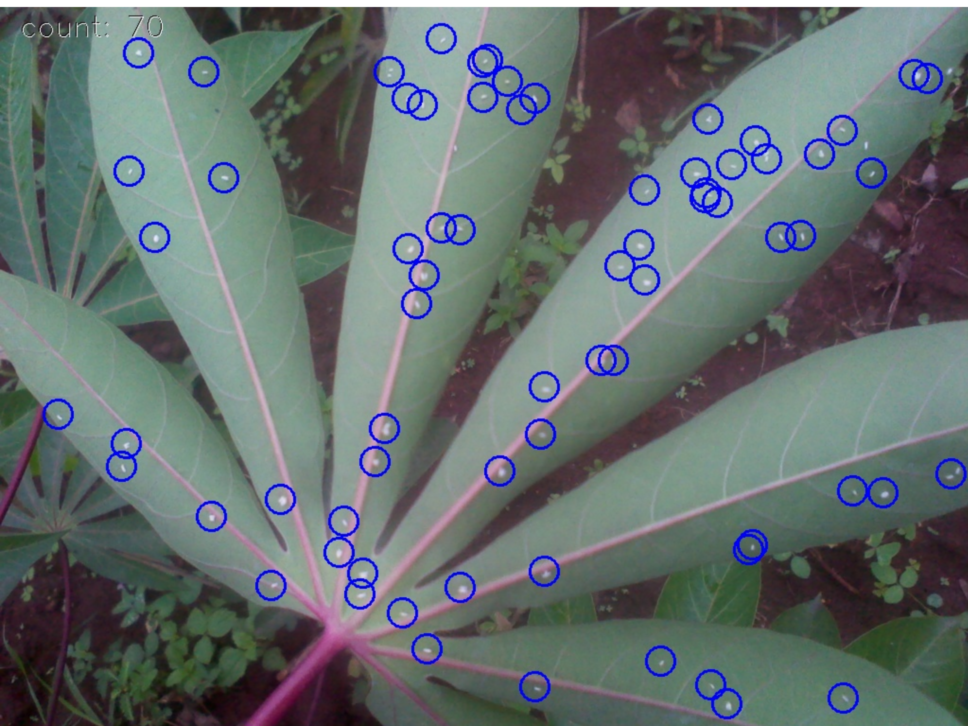
We can do classification on parts of images too. To count the number of whiteflies in this image, cut it into a grid and classify each grid cell separately, $y^{(i)} \in \{\text{whitefly}, \text{no whitefly}\}$.

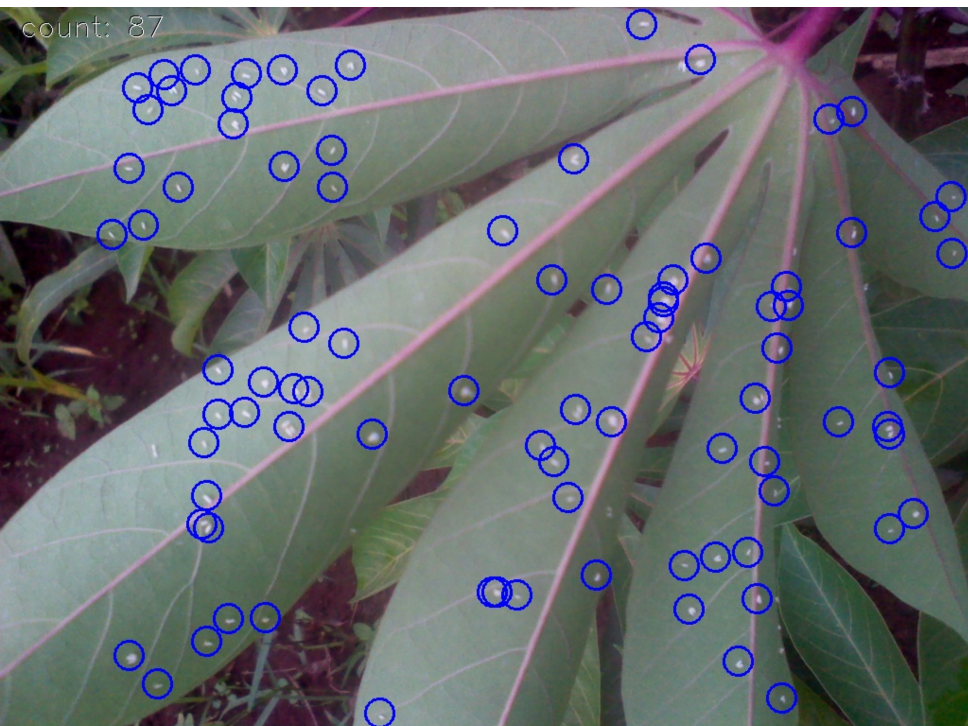



count: 94


count: 6


count: 67


count: 70


count: 87

Spatial modelling

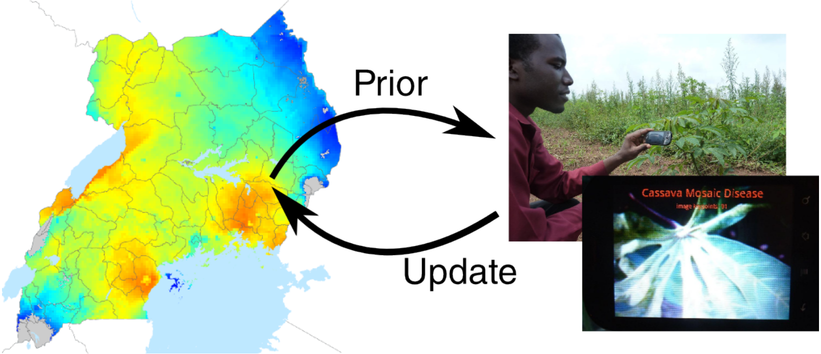
- From observations of disease at different locations, we want to estimate the underlying disease risk across the map.
- Now $\mathbf{x}^{(i)} \in \mathbb{R}^2$ is the location of the observation, and $y^{(i)}$ is the corresponding disease label.
- Notebook with demo: https://github.com/jqug/MLPR_crop_disease_mapping

Spatial modelling: other problems

- Can we decide where to send surveyors given limited resources? Given $\{\mathbf{x}^{(i)}, y^{(i)}\}_{i=1}^{N}$, how valuable do we think the information $\mathbf{x}^{(j)}, y^{(j)}$ would be, as a function of $\mathbf{x}^{(j)}$?
- Given the times that each observation was made, so that we have $\{\mathbf{x}^{(i)}, y^{(i)}, t^{(i)}\}_{i=1}^{N}$, can we learn the dynamics of the infection and make predictions?

Combined diagnosis and mapping

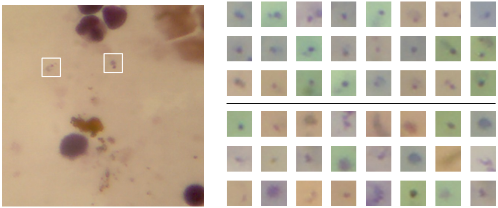
The tasks of diagnosis and mapping are mutually compatible:



Therefore we can try to construct multi-scale models which jointly model the uncertainty in both tasks.

Application to human diseases

Plasmodium in blood smear images:



Photomicroscopy with smartphone: