

# Data for Agriculture

*Challenges and Opportunities in East Africa*

**Dina Machuve, PhD**

**29th July 2023**

# Who am I?

Dina Machuve



# Agenda

1. Background
2. Data 4 Agriculture in East Africa
  - a. Use Case 1– Poultry Diseases Diagnostics
  - b. Use Case 2 – Cassava Diseases Detection
  - c. Use Case 3 – Lacuna Fund
3. Challenges and Opportunities

## Overview

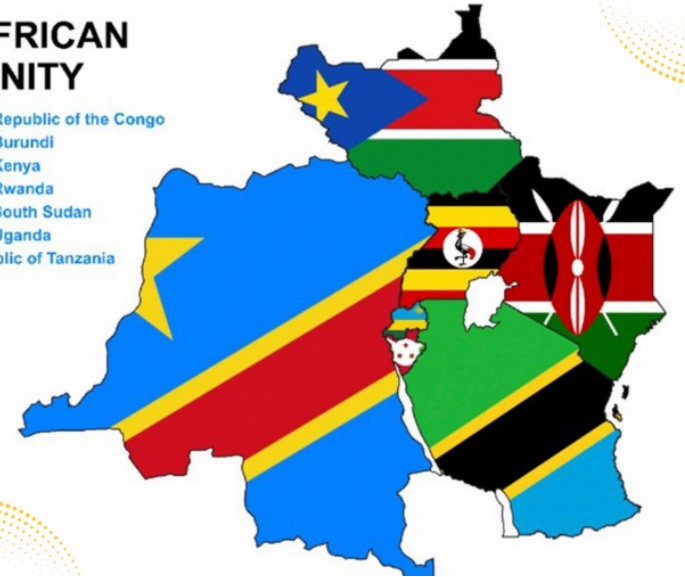
# Agriculture is the backbone of the economy

## EAST AFRICA

- 7 partner states
- Agriculture = Crop production + Livestock
- Population: 283.7 million (2021)
- 70% are smallholder farmers
- GDP: US\$ 305.3 bil (2021)

### EAST AFRICAN COMMUNITY

-  Democratic Republic of the Congo
-  Republic of Burundi
-  Republic of Kenya
-  Republic of Rwanda
-  Republic of South Sudan
-  Republic of Uganda
-  United Republic of Tanzania



Source: [www.eac.int](http://www.eac.int)

## Use Case 1

# Poultry health matters



## TANZANIA

- Low poultry productivity from diseases
- Lack of trusted poultry data
- Chickens: 36 mil
- Households with chickens: 4.6 mil
- Population: 56.32 mil (2018)
- GDP: US\$ 55.5 bil (2019)

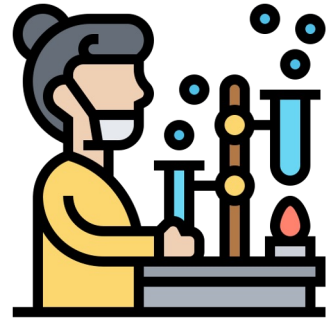


## Challenge

# Poultry diseases monitoring now

Salmonella, Newcastle and Coccidiosis poultry diseases:

- Diagnosed by lab procedures using droppings samples
- It takes 3 -4 days to get results
- Clinical signs
- Access to the services by farmers is expensive and limited
- Extension officers lack tools for rapid diagnostics



## Motivation

# Image data is better than lab data in developing countries

- **Ubiquity** of mobile phones in developing countries:
  - 60.3 mil mobile phone subscriptions in Tanzania
  - 31.1 mil internet subscriptions on mobile
  - 515 mil unique mobile subscriptions in Africa
  - Mobile phones as a sensor
- **Low levels of literacy** and **multiple languages**
  - Images form a universal data format



# Project Objectives



# Project goals

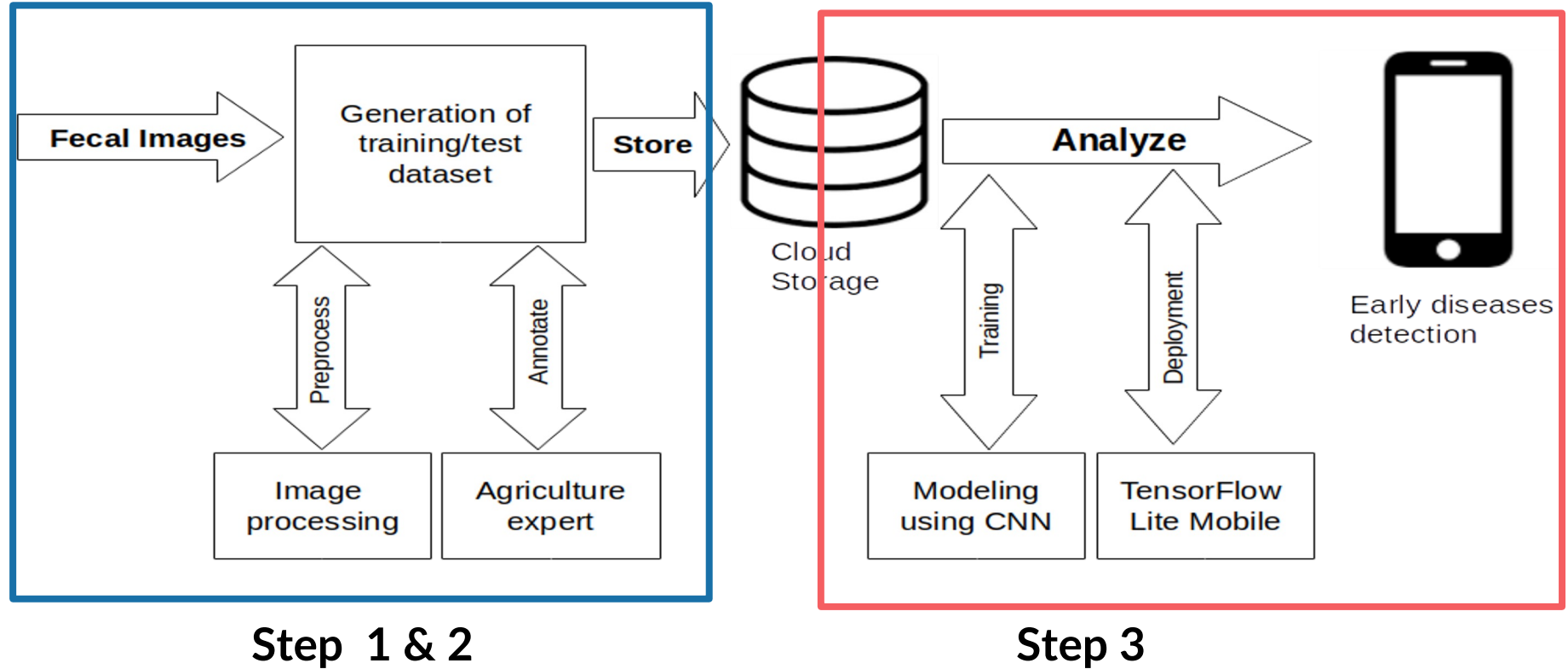
## Specific Objectives:

1. Generation of Training/Testing dataset for poultry diseases diagnostics
2. Dataset curation
3. Diagnostics/Modeling

Expected Outcome: Establishing an annotated dataset and a deep learning model for detecting three poultry diseases at the farm level

Task

Workflow

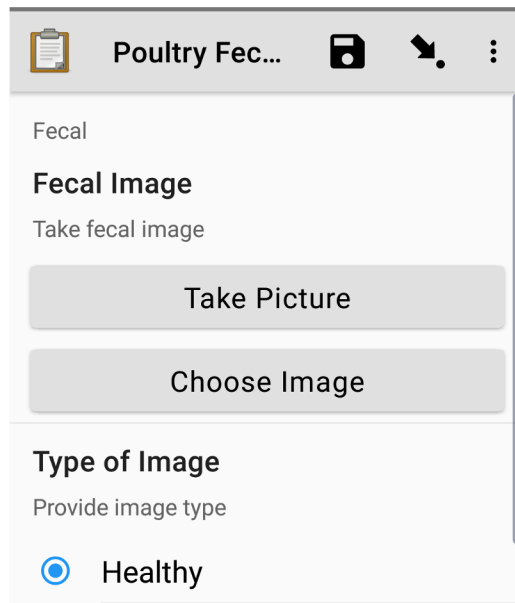


Generation of Training/Testing dataset for poultry diseases diagnostics

# Data collection tool

Used Open Data Kit (ODK)

Dataset on Google Sheets



Poultry Fec...

Fecal

**Fecal Image**

Take fecal image

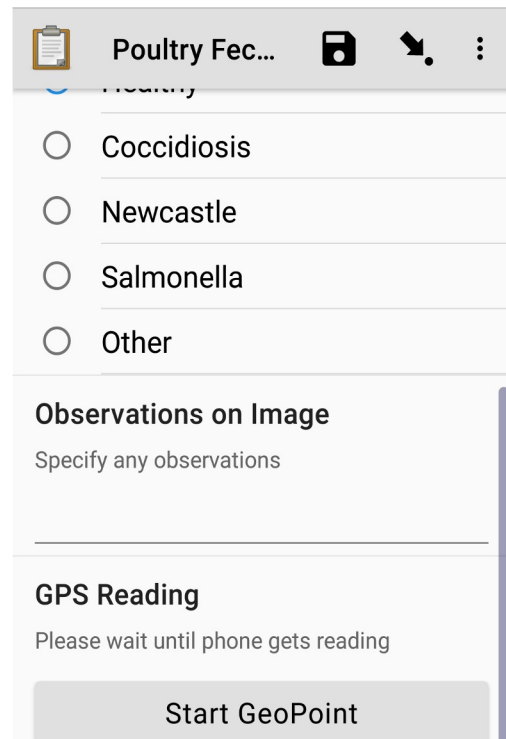
Take Picture

Choose Image

**Type of Image**

Provide image type

☒ Healthy



Poultry Fec...

Healthy

☐ Coccidiosis

☐ Newcastle

☐ Salmonella

☐ Other

**Observations on Image**

Specify any observations

**GPS Reading**

Please wait until phone gets reading

Start GeoPoint

729	23/09/20 09:21	23/09/20 09:22	23/09/20	355329114698347	<a href="https://drive.google.com/file/d/1...">https://drive</a>	coccid			-3.3577116666	931.3	50
730	23/09/20 09:22	23/09/20 09:22	23/09/20	355329114698347	<a href="https://drive.google.com/file/d/1...">https://drive</a>	coccid			-3.3577066666	931.8	50
731	23/09/20 09:22	23/09/20 09:23	23/09/20	355329114698347	<a href="https://drive.google.com/file/d/1...">https://drive</a>	coccid			-3.3576866666	931.7	4.9
732	23/09/20 09:23	23/09/20 09:24	23/09/20	355329114698347	<a href="https://drive.google.com/file/d/1...">https://drive</a>	coccid			-3.3576783333	933.9	4.8
733	23/09/20 09:24	23/09/20 09:24	23/09/20	355329114698347	<a href="https://drive.google.com/file/d/1...">https://drive</a>	coccid			-3.3576433333	934.1	50
734	23/09/20 09:25	23/09/20 09:25	23/09/20	355329114698347	<a href="https://drive.google.com/file/d/1...">https://drive</a>	coccid			-3.3575983333	933.3	4.9
735	23/09/20 09:25	23/09/20 09:26	23/09/20	355329114698347	<a href="https://drive.google.com/file/d/1...">https://drive</a>	coccid			-3.3576216666	935.3	50

# Image data collection at farms



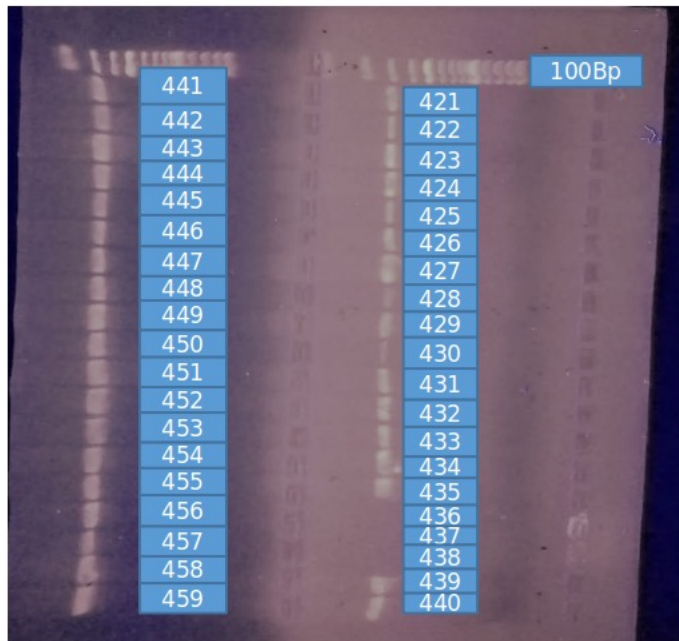


# Inoculation and monitoring



# Lab labelled data: PCR diagnostics

- Used Polymerase Chain Reaction (PCR) diagnostics to annotate the laboratory-labeled fecal samples.
- We collected fecal images and fecal samples from poultry farms.
- Established a high-quality labelled dataset of 1,255 poultry fecal images



**PCR RESULTS  
FOR  
SALMONELLA  
(38  
SAMPLES)  
USING InVA  
141 AND InVA  
139 PRIMERS**  
**Result:**  
**3 are -ve and  
35 are +ve lie  
between 200  
- 300Bp**

# Dataset: Chicken Droppings Images

Class	Farm-labelled Images	Lab-labelled Images
Healthy	2,057	347
Coccidiosis	2,103	373
Salmonella	2,276	349
Newcastle	376	186
<b>TOTAL</b>	<b>6,812</b>	<b>1,255</b>

Healthy



Coccidiosis



Salmonella





# Poultry Diseases Diagnostics

- Deep CNN model was developed to diagnose healthy and unhealthy poultry fecal images
- Architectures: baseline CNN, VGG16, InceptionV3, MobileNetV2 and Xception
- Dataset on training: Farm and laboratory labeled fecal images
- Dataset on testing: Farm labeled fecal images
- Goal: deployment of the model on smartphones to help farmers distinguish diseased from healthy poultry based on fecal images

# Crop Diseases Monitoring

## Use Case 2

# Crop diseases monitoring - Uganda

- Crop disease monitoring at farm level involves agricultural extension officers
  - Ratio (1:1,500) of extension officers to farmers is too low
- The trend on mobile apps for agriculture indicate the shift to automated solutions using ML and computer vision on crop diseases detection
  - The feedback to farmers is not instant
  - It takes 5-7 days before a farmer can get feedback
  - Lack of real-time feedback information to farmers leads to low productivity

## Challenge

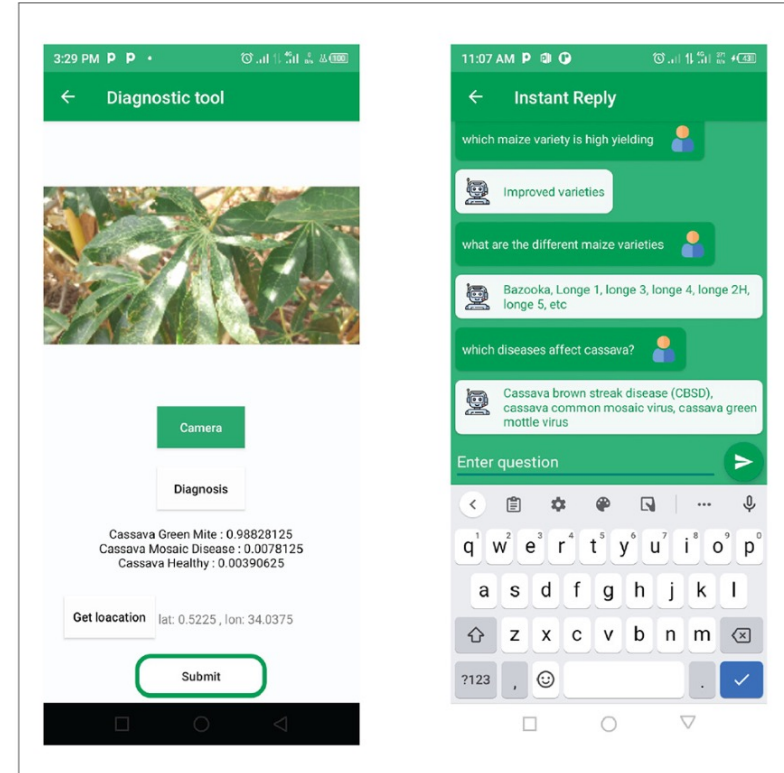
# Agronomic Q&A Dataset

- Pilot study engaging 100 farmers
- Q&A pairs on staple food crops such as cassava, maize and beans
- Questions obtained from farmers through interviews and mobile app
- Answers were given by agricultural experts
- A new dataset **3,939** question-answer pairs
  - Crowd-sourced from the farmers

## Challenge

# Agronomic recommender system

- Established using ML and NLP techniques
- Training on retrieval models: RetBERT, HayStack and Seq2seq
- Deployment of model on smartphone
  - Disease diagnosis interface
  - and Q&A interface



# Challenges

# Open Questions

- Published datasets for further research – what new knowledge will be generated that benefits the farmers?
- How do we incentivize farmers to use these digital tools?
- How do we reach farmers that don't own smartphones?
- Farmers in East Africa continue to have limited access to data:
  - Low level of literacy
  - Access to digital extension and advisory services dominantly through radio

# Opportunities



## Use Case 3

# Dataset Gap

- World's first collaborative fund to correct gaps and biases in data for AI
- ML Datasets in Agriculture:
  - Ground image data – smallholder farmers' fields in Kenya
  - Maize plot location and yield data in East Africa
  - IoT Fish Pond monitoring dataset
  - Crop Pest and Disease Diagnosis – Crop imagery and spectrometry data

# Opportunities on Data

- Robust end-user tools addressing the farmers needs
  - Using native languages – Swahili is dominant in East Africa
  - Used in production level and not ending on pilot stages
- Legal framework on data to protect contributors (farmers)
- Capacity building of local researchers and developers to develop solutions addressing local problems

**THANK YOU**