



Data Article

A labeled spectral dataset with cassava disease occurrences using virus titre determination protocol



Godliver Owomugisha^{a,*}, Joyce Nakatumba-Nabende^b,
Joshua Jeremy Dhikusooka^b, Estefania Taravera^c,
Ephraim Nuwamanya^d, Ernest Mwebaze^b

^a Faculty of Engineering, Busitema University, P. O. Box 236, Tororo, Uganda

^b College of Computing & IS, Makerere University, P.O. Box 7062, Kampala, Uganda

^c Faculty of Electrical Engineering, Data Management & Biometrics, University of Twente, P.O. Box 217 7500 AE, Enschede, the Netherlands

^d National Crops Resources Research Institute, P.O Box 7084, Kampala, Uganda

ARTICLE INFO

Article history:

Received 10 May 2023

Revised 4 July 2023

Accepted 4 July 2023

Available online 9 July 2023

Dataset link: [Cassava Spectral and Image Dataset \(Original data\)](#)

Keywords:

Spectral data protocol

Cassava diseases

Crop diagnosis

Smart agriculture

Early disease detection

ABSTRACT

In this work, we present a novel dataset composed of spectral data and images of cassava crops with and without diseases. Together with the description of the dataset, we describe the protocol to collect such data in a controlled environment and in an open field where pests are not controlled. Crop disease diagnosis has been done in the past through the analysis of plant images taken with a smartphone camera. However, in some cases, disease symptoms are not visible. Furthermore, for some cassava diseases, once symptoms have manifested on the aerial part of the plant, the root which is the edible part of the plant has been totally destroyed. The goal of collecting this multimodality of the crop disease is early intervention, following the hypothesis that diseased crops without visible symptoms can be detected using spectral information. We collected visible and near-infrared spectra captured from leaves infected with two common cassava diseases namely; Cassava Brown Streak Disease and Cassava

* Corresponding author.

E-mail address: ogodliver@eng.busitema.ac.ug (G. Owomugisha).

Social media: [@GodliverO](#) (G. Owomugisha), [@j_nabende](#) (J. Nakatumba-Nabende), [@jeremjosh1](#) (J.J. Dhikusooka), [@eTaraveraM](#) (E. Taravera), [@wamanya2](#) (E. Nuwamanya), [@emwebaze](#) (E. Mwebaze)

Mosaic Disease, as well as from healthy plants. Together, we also captured leaf imagery data that corresponds to the spectral information. In our experiments, biochemical data is collected and taken as the ground truth. Finally, agricultural experts provided a disease score per plant leaf from 1 to 5, 1 representing healthy and 5 severely diseased. The process of disease monitoring and data collection took 19 and 15 consecutive weeks for greenhouse and open field, respectively, until disease symptoms were visibly seen by the human eye.

© 2023 Published by Elsevier Inc.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

Specifications Table

Subject	Agronomy and Crop Science, Artificial Intelligence.
Specific subject area	Cassava Brown Streak Disease (CBSD), Cassava Mosaic Disease (CMD), Spectral Data)
Type of data	Raw spectral data Leaf Imagery data Biochemical lab test data Visual expert scoring
How the data were acquired	Spectral data was collected using visible and near-infrared handheld spectrometer [1]. Leaf image data was captured using a smartphone camera of 12-megapixel. The biochemical data was extracted using a 96-well Real Time PCR instrument [2]. Finally, agricultural experts score plants as healthy or diseased based on visual symptoms that appear on the plants.
Data format	1. Raw data 2. Analyzed data
Description of data collection	The dataset is in two major groups: greenhouse and open field experiment, collected for 19 and 15 consecutive weeks respectively.
Data source location	The dataset is in two major groups: greenhouse and open field experiment, collected for 19 and 15 consecutive weeks respectively.
Data accessibility	Repository name: Harvard Dataverse Data identification number: doi: 10.7910/DVN/R0KL7R Direct URL to data: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/R0KL7R
Related research article	Godliver Owomugisha, Ephraim Nuwamanya, John A. Quinn, Michael Biehl, and Ernest Mwebaze. 2020. Early detection of plant diseases using spectral data. In Proceedings of the 3rd International Conference on Applications of Intelligent Systems (APPIS 2020). Association for Computing Machinery, New York, NY, USA, Article 26, 1–6. https://doi.org/10.1145/3378184.3378222

Value of the Data

- The collected samples are a contribution to the field of smart farming and agriculture, more specifically to the growing area of early disease detection in asymptomatic plants.
- The findings of this dataset will contribute to the breeding research to develop the best crop breeds.
- To our best knowledge, this is the first spectral dataset that is publicly available on cassava diseases, a crop that is highly researched on.
- This dataset will also pave the way for the same research in other crops in early disease detection.

CRedit Author Statement

Godliver Owomugisha: Writing – review & editing, Conceptualization, Methodology; **Joyce Nakatumba-Nabende:** Conceptualization, Methodology; **Joshua Jeremy Dhikusooka:** Writing – original draft; **Estefania Taravera:** Writing – review & editing; **Ephraim Nuwamanya:** Data curation; **Ernest Mwebaze:** Conceptualization, Methodology.

Acknowledgments

This work was carried out with funding support from: (i) Lacuna Fund, an initiative co-founded by The Rockefeller Foundation, Google.org, and Canada's International Development Research Centre. The views expressed herein do not necessarily represent those of Lacuna Fund, its Steering Committee, its funders, or Meridian Institute.: 0328-S-001. (ii) Data Science Africa Research Award: DSA-RA-2020-035 The authors would like to thank the Directors of the Uganda National Crop Resources Research Institute (NaCRRI), for granting them permission to access cassava fields and screenhouses.

References

- [1] CID Bio-Science, Inc., "CI-710 MINIATURE LEAF SPECTROMETER," 2010. [Online]. Available: <http://www.cid-inc.com>.
- [2] Thermo Fisher Scientific, "StepOnePlus™ Real-Time PCR System." 2023 [Online]. Available: <https://www.thermofisher.com/order/catalog/product/4376600>.
- [3] L3Harris Geospatial, "Vegetation analysis: Using vegetation indices in envi." 2022. [Online]. Available: <https://www.l3harrisgeospatial.com/Support/Self-Help-Tools/Help-Articles/Help-Articles-Detail/ArtMID/10220/ArticleID/16162/Vegetation-Analysis-Using-Vegetation-Indices-in-ENVI>.
- [4] D. Mbanzibwa, Y. Tian, A. Tugume, S. Mukasa, F. Tairo, S. Kyamanywa, A. Kullaya, J. Valkonen, Simultaneous virus-specific detection of the two cassava brown streak-associated viruses by rt-pcr reveals wide distribution in east africa, mixed infections, and infections in manihot glaziovii, J. Virol. Methods 171 (2) (2011) 394–400 [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0166093410003496>.
- [5] G. Owomugisha, E. Nuwamanya, J.A. Quinn, M. Biehl, E. Mwebaze, Early detection of plant diseases using spectral data, in: Proceedings of the 3rd International Conference on Applications of Intelligent Systems, Association for Computing Machinery, New York, NY, USA, 2020 Ser. APPIS 2020 [Online], doi:[10.1145/3378184.33](https://doi.org/10.1145/3378184.33).