
This item was submitted to [Loughborough's Research Repository](#) by the author.
Items in Figshare are protected by copyright, with all rights reserved, unless otherwise indicated.

Validation of household sanitation classification using artificial intelligence software within a market-based sanitation intervention in Uganda

PLEASE CITE THE PUBLISHED VERSION

PUBLISHER

WEDC, Loughborough University

VERSION

VoR (Version of Record)

LICENCE

CC BY-NC-ND 4.0

REPOSITORY RECORD

Namakula, Patricia, Leland Smith, and Jonathan Annis. 2021. "Validation of Household Sanitation Classification Using Artificial Intelligence Software Within a Market-based Sanitation Intervention in Uganda". Loughborough University. <https://hdl.handle.net/2134/16918378.v1>.

42nd WEDC International Conference

ONLINE: 13 – 15 September, 2021

**EQUITABLE AND SUSTAINABLE WASH SERVICES:
FUTURE CHALLENGES IN A RAPIDLY CHANGING WORLD**

**Validation of household sanitation classification using
artificial intelligence software within a market-based
sanitation intervention in Uganda**

P. Namakula, L. Smith & J. Annis

Uganda

REFERENCE NO. 3308

Background

The Uganda Sanitation for Health Activity (USHA) is a five-year program financed by the United States Agency for International Development (USAID) working to improve household sanitation through market-based approaches in 20 districts in Uganda (USHA, 2019). Through March 2020, market-based sanitation efforts have reached 1,608 communities, resulting in over 20,051 latrines being constructed or upgraded, of which 65% percent meet the WHO/UNICEF Joint Monitoring Program definition of basic sanitation. USHA captures household sanitation status at baseline and monitors the outcomes of its interventions using digitized survey tools designed with Open Data Kit (ODK). USHA is using an image classification model powered by machine learning (ML) to accurately and efficiently analyse the characteristics (i.e. superstructure and floor) of thousands of images of completed toilets and will eventually compare the results from the ML algorithm to responses from observational questions manually inputted by enumerators.

Problem statement

Rural sanitation programs in the era of Community Led Total Sanitation (CLTS) have struggled to generate high quality data needed to track implementation effectiveness and associated outcomes (USAID, 2018). In the WASH sector, data is required to accurately track progress, inform planning, and management decisions (IRC, 2018). While geo-referencing sanitation facilities coupled with observational data collection pre- and post-intervention can be used to document a households investment, the process is prone to human error (i.e. improper classification) by enumerators during data collection (Jamie, *et al.*, 2014). Working with large data volumes further hinders the process of conducting data quality assessment checks to validate the correct classification of a household WASH status (Gine, *et al.*, 2013). Traditionally, machine learning and image classification has been the domain of trained data scientists and experts for accurate case classifications. There are however few documented applications of ML in the WASH sector.

Methodology and early results

The methodology involves verifying that the enumerator classification of toilet characteristics matches with the artificial intelligence algorithm run on sample WASH images and generating an output and confidence interval (Hubert, *et al.*, 2020). USHA is demonstrating this methodology using data from over 20,051 toilets constructed or upgraded across 20 districts in Uganda.

First, we acquired a randomized sample of toilet images from each of the districts to represent the complete spectrum of toilet facilities surveyed. We then imported these images into a simple training interface powered by Lobe.ai, an open source ML application developed by Microsoft. The team choose the application, which requires no initial software development, to demonstrate the growing capabilities of ML/AI software for practical application by of non-experts. The Lobe desktop program allows the user to tag individual images

with classifications. It then automatically carries out training and validation of a ML model, with immediate visual feedback to the user on accuracy and challenges, shown in Figure 1.

In this case, we can see that based on our training data set of 364 images of latrine superstructures, the ML algorithm correctly identified the type of structure 97% of the time. We then exported a Tensorflow model from Lobe and wrote a short Python script that fetches images directly from our database, applies the model, and records the output as a predicted classification and a set of confidence scores. This output is saved as a CSV file with references to the original survey items, so that we can rejoin the data for further validation and analysis.

At this time, we have run models superstructures with a high degree of accuracy. We've found challenges with edge cases, such as a wattle and daub structure with cracks that cause it to resemble bricks. Some images are also poorly framed, and the model is unable to accurately isolate the structure to be classified. Nonetheless, spot checks on a further 1,000 images analysed show a high degree of accuracy. We are using confidence outputs to target specific images to follow up manually. We can use these images as feedback to further adjust the model for improved accuracy. We can also use these scores for manual correction of outputs for final reporting. Accuracy for floor materials is significantly lower at the moment. The model is often unable to distinguish dirt floors from concrete floors with dust or dirt on them, though these cases are fortunately fairly few. We expect continued improvement in the model through additional feedback to the training dataset. By the end of March 2020, we will begin comparing the model results for superstructure and floor material to the data submitted by the enumerators.

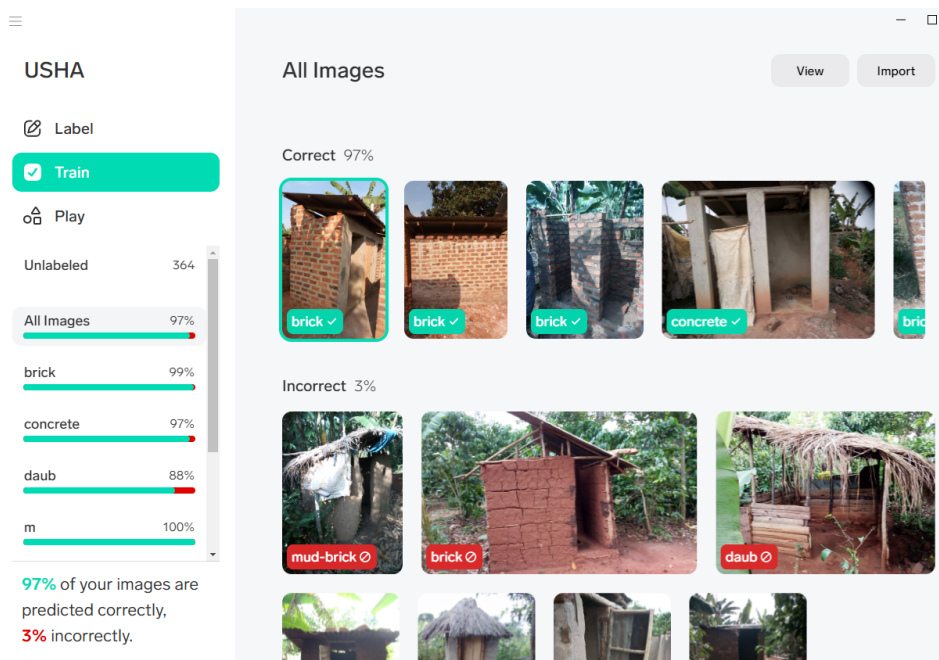


Figure 1: Lobe desktop training interface

References

- Giné Garriga R, Jiménez A, Pérez Foguet A. Water-sanitation-hygiene mapping: An improved approach for data collection at local level. *Sci. Total Environ.* 2013;463-464:700-711
- Hubert Jenny, Eduardo Garcia Alonso, Yihong Wang & Roberto Minguez. Using Artificial Intelligence for Smart Water Management Systems. (2020).
- IRC. A baseline of the strength of the WASH system in Uganda. (2018).
- Jamie Bartram, Clarissa Brocklehurst, Michael B. Fisher, Rolf Luyendijk, Rifat Hossain, Tessa Wardlaw and Bruce Gordon. Global Monitoring of Water Supply and Sanitation: History, Methods and Future Challenges. (2014).
- USAID, 2018. An Examination of CLTS's Contributions toward Universal Sanitation. Washington, DC., USAID Water, Sanitation, and Hygiene Partnerships and Sustainability (WASHPaLS) Project.
- USAID, 2019. Activity Monitoring, Evaluation, and Learning Plan, Uganda Sanitation for Health Activity [USHA].

Contact details

All three authors work for Tetra Tech, ARD. Patricia Namakula is USHA's Monitoring Evaluation Learning Advisor, based in Kampala, Uganda. Leland Smith is a Data Scientist based in Arlington VA. Jonathan Annis is USHA's Chief of Party, based in Kampala, Uganda.

Patricia Namakula:

Telephone: +256772861334.

Email: Patricia.Namakula@Uganda-Sanitation.org

Leland Smith:

Fax (802) 658.4247. Email: leland.smith@tetrattech.com

Jonathan Annis:

Telephone: +256779598038. Email: jonathan.annis@tetrattech.com