|  |  |  |
| --- | --- | --- |
| ITU Logo | INTERNATIONAL TELECOMMUNICATION UNION**TELECOMMUNICATIONSTANDARDIZATION SECTOR**STUDY PERIOD 2017-2020 | FG-AI4H-L-035 |
| **ITU-T Focus Group on AI for Health** |
| **Original: English** |
| **WG(s):** | TG-Outbreaks | E-meeting, 19-21 May 2021 |
| **DOCUMENT** |
| **Source:** | Institute for Technology & Global Health (ITGH) |
| **Title:** | TG-Outbreaks: Technology, Economics, & Policy: AI for Sanitation and Public Health |
| **Purpose:** | Discussion |
| **Contact:** | Khahlil A. LouisyHarvard University/Institute for Technology & Global HealthUnited States of America | Tel: +1-857-330-5341Email: klouisy@hks.harvard.edu |
| **Contact:** | Alexander RadunskyInstitute for Technology & Global HealthUnited States of America | Email: apr814@mail.harvard.edu |

|  |  |
| --- | --- |
| **Abstract:** | Sanitation is a major global health issue, which has a significant toll on human lives and economic output. About 50,000 litres of untreated sewage is released every second in South Africa and the country has one of the world’s highest rates of mortality for children under the age of five, with about 10% of those deaths resulting from diarrhoea. This proposal reflects details for an “AI for Sanitation” project to address the issue of poor sanitation as a public health threat and how it can be mitigated using advanced technologies such as Artificial Intelligence systems. The project will utilize terrestrial sensors in sewage systems combined with space technology from the European Space Agency (ESA), to collect and analyse data which will inform health workers of potential outbreaks as a result of disease-causing pathogens in excreta. The data collected will then be used to develop predictive models and train algorithms to detect the presence of pathogens in faecal waste. Outcomes from this study will be used to benchmark against manual approaches to testing for disease in sewage, associated costs, and potential economic benefits. |

**Overview and Feasibility of Project**

Early detection of disease-causing pathogens is critical to mitigating community spread and the possibility of pandemics, which could lead to global catastrophes as demonstrated by the COVID-19 pandemic. Today, sanitation continues to be a major global health issue, with approximately 3.8 trillion litres of waste left untreated or lost and 892 million people in the world still practicing open defecation (UNICEF/WHO JMP, 2017). In India, it is estimated that as a direct result of poor sanitation, about $54bn of GDP is lost annually and 200,000 human lives are lost.

Advances in technological innovation can not only help with the early detection of disease-causing pathogens, but the externalities from these early detections could have positive impacts on economic output. The use of AI technology to analyse waste products in sewage systems has the potential to mitigate the spread of disease by alerting health workers of impending outbreaks. The sanitation-related diseases, relevant to this proposal, are those that result in pathogen shedding in excreta and have an environmental transmission factor, e.g. Typhoid.

This implementation study, to be conducted in Durban, South Africa, will explore AI models trained on sensor data to predict sanitation-related user behaviour, environmental events and disease. This will be done by developing a complete system to collect, treat and process human waste, which will be managed by a single ICT system that can employ sensors, AI and machine learning to predict and manage the instance of certain pathogens and disease outbreaks as a result of poor sanitation management.

This project addresses goals six and three of the UN’s SDGs, that is to ensure access to water and sanitation for all and to ensure healthy lives and promote well-being for all at all ages, respectively. The level of health-related issues in this region highlights the need for 21st century smart sanitation solutions.

**South Africa’s Sanitation Issues**

In a survey in South Africa, the Department of Water Affairs: Green Drop Certification (2011) found that 55% of wastewater treatment plants, especially smaller ones, do not meet effluent standards and some do not even measure effluent quality. In 2009, when 449 wastewater treatment plants were assessed, according to official government data, 55% “did not perform within acceptable standards.” According to a World Bank report, Africa Infrastructure Country Diagnostic (AICD), the total expenditures are less than half of what would be required to achieve the Millennium Development Goals in Sub-Saharan Africa. The African Development Bank estimates that $12 billion is required annually to cover Africa’s needs in improved water supply and sanitation. Currently, 50,000 litres of untreated sewage is release every second in South Africa and the country has one of the world’s highest rates of mortality for children under the age of five. About 10% of those deaths are from diarrhea. Critically, the journal warned of the effects of E. coli and other sewage-related pollutants in a population with a high prevalence of HIV.

A feasibility assessment has been conducted, commissioned by the European Space Agency (ESA) and Woodco – a development and manufacturing company based in Ireland – on the viability of developing a human waste sewage treatment process using biomass waste to energy processes, which can be combined with sensor technology and artificial intelligence to monitor and predict disease occurrences, whilst rendering the waste into a pathogen-safe marketable carbon by-product.

The parties have entered into an agreement with the government of Durban, South Africa on the implementation of the system, data collection, and monitoring and evaluation. The first implementation is scheduled for early 2022, however work on the evaluation methodologies and frameworks will be developed in the interim with ITGH.

 **Data and Analysis**

Data from this project has not been made public and will be available for AI Developers and for training purposes.

The planning and overall impact analysis of the system incorporates terrestrial and EO based data which is merged with the terrestrial data derived from the sensors in the environs of the toilet blocks. The vacuum trucks are tracked using GNSS and can communicate over satellite short-burst data as well as terrestrial cellular communications.

EO data can be linked to terrestrial sensor data, waste sample analysis and toilet block monitoring. The data sets will be added to the sensor data collected from the IBM acoustic sensors developed under the CCN and the floating lab sensor under the AI kickstart activity to develop machine learning systems which can output artificial intelligence data. EO data is used to manage and monitor the system with the processing plants offering a solution as self-powered communications hubs in remote areas.

We will install a working CHP site and telemetry at the target toilet block location. The toilet block will be used to generate some of the faecal sludge with the majority of the sludge derived from the emptying of septic tanks and cesspools from domestic business environments. The main function of the toilet block telemetry is to test user usage profiles and to collect data which can be used to train predictive AI systems.

We will also place environmental sensors out into the field in the environs of the toilet blocks to pull data back on water and weather conditions to develop richer data sets for the predictive health monitoring. The location of the devices is critical to the success of the project as the data derived will inform the operating model. The acquired data will be used to reflect the monitoring of sanitation conditions and reflect the benefits of the system. The pilot set-up will therefore be of paramount importance.

The goal of the project is to deliver predictive health monitoring, which will require data from multiple sources. To correlate environmental data with detected disease events we need to physically collect faecal samples and process these. The date and location of the physical samples taken needs to be linked to the environmental data to allow the disease detection to be linked to human and environmental activity.

The pilot set-up process will include the physical sampling of faecal sludge at the target sites, therefore a sampling schedule and lab analysis plan will be developed as part of the pilot set-up process.

The trial will last for one calendar year. The initial deployment will be of the main plant and equipment which will include the combined heat and power pyrolysis units and the monitored toilet blocks. This will also include equipping the relevant service vehicles which are partaking in the trial period.

The second release will be the preliminary AI algorithms which will commence correlating the data from the initial installation and then this will be linked to the physical sludge sampling and ground truth EO data to widen the source of data which will be used to develop the health for sanitation analytic tools.

The final release will be of the floating lab with will autonomously detect pathogens in water courses in the region where the demonstration phase is occurring. Data from the floating labs will be used to generate data which will be input to the AI model.

The following space asset(s) will be used in the pilot-demonstration:

We will use Sentinel 1 and Sentinel 2 data to provide a “macro” view of the areas where the demonstration phase is occurring. We will amalgamate EO and GIS data to generate digital elevation maps to assess where the optimum location is for the plant and equipment and to look at the collection of data from ground sensors in the area. The EO data will be used to provide a wider geographical view of the specific area and will support the predictive health elements.

We will incorporate GNSS in all mobility and logistic elements and also to way-mark all sensors to allow us to add a GIS – ground truth layer to the overall data models. All data will be way-marked where possible. The vehicles used on the demonstration phase will use short-burst data on the L-Band and where applicable we will use L-band BGAN services (class2 terrestrial communications) to back-haul data from specific locations.

The pilot system deployment will require significant training and support and we will deploy a planning application to allow this to occur in order to maximize the efficacy of the pilot phase. We estimate that the release 1 deployment will take up to 5 months to complete, with release 2 taking 2 months and release 3 two months.

**Benchmarking**

We will be collecting data of varying types (both structured and unstructured) and from many different sources. We then plan to incorporate varying degrees of AI from simple if/else control statements to complex deep learning models for different purposes.

The general high-level approach would be to benchmark against:

1. What measurements are being taken at the moment in the areas where the solution is being deployed and what are they achieving.
2. What is the state of the art using AI and not using AI (not confined to what is being done on the ground at the moment).
* We will benchmark the AI-based solution against the existing largely manual and in-laboratory testing.
* We will also benchmark it against the state of the art AI and non-AI capabilities that are comparable to the solution.
* From a machine learning perspective, we have used transfer learning and have compared performance for the proof-of-concept using a large dataset of evenly distributed class samples with 80/20 training/validation splits with a number of machine learning architectures with the above constraints in mind. e.g. VGGNet, MobileNet and Resnet architectures.
* In terms of metrics, apart from accuracy, as the device is a remote sensing device, we are interested in cost, power consumption, deployability/embeddability, performance, and maintenance requirements.

**About the Institute for Technology & Global Health**

The Institute for Technology and Global Health (ITGH) is an applied research and policy organization, exploring the potential for novel technologies and their applications to improving public health outcomes and interrogating public policy questions related to their implementations. The non-profit organization is an independent initiative of PathCheck Foundation, which was founded at the Massachusetts Institute of Technology (MIT) to develop privacy-preserving technologies in response to health crises. The foundation and the institute are comprised of faculty and researchers from MIT and Harvard University, as well as professionals across a range of relevant industries and sectors, including government, public health, technology, and academia among others.

PathCheck Foundation developed and launched several pieces of technologies in response to the COVID -19 pandemic that are currently being utilized by several countries and U.S. States, including exposure notification systems and digital credentials. ITGH has conducted studies on the use of these applications including adoption and behavioural insights, implementation policies, and ethics and privacy. Further the institute serves in an advisory capacity to several multilateral organizations and international governments on the digital transformation of health systems and is currently undertaking studies to benchmark usage of novel technologies and digital systems transformations.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_