IST-Africa 2021 Conference Proceedings
Miriam Cunningham and Paul Cunningham (Eds)
IST-Africa Institute and IIMC, 2021
ISBN: 978-1-905824-67-0

Leveraging IoT and Machine Learning for Improved Monitoring of Water Resources - A Case Study of the Upper Ewaso Nyiro River

Jason KABI[1]*, Ciira MAINA*[2]
[1, 2] *Dedan Kimathi University of Technology, Department of Electrical and Electronic Engineering, Centre for Data Science and Artificial Intelligence. P.O. BOX - PRIVATE BAG – 10143, Dedan Kimathi - Nyeri, Kenya.*
Email: jason.kabi@dkut.ac.ke, ciira.maina@dkut.ac.ke

Abstract: This paper outlines the development of an Internet of Things (IoT) system, capable of water-level monitoring in rivers, with an aim of ensuring equitable distribution of water and sourcing of data to quantify unsustainable water usage, river catchment destruction and climate change. The paper also defines an application scenario in a specific hydrological region of the Ewaso Nyiro basin in Kenya, highlighting the characteristics of data collection and processing used. Fixed position node systems are described along with web-based data acquisition platform developments, integrated with IoT techniques to retrieve data. The developed architecture utilizes the LoRaWAN - LoRa protocol to send data packets from nodes deployed, to an IoT network server. From the server, data is transferred to a database, where it can be accessed and displayed through different customizable queries and graphical representations. Outlined characteristics are presented along with evidence of the deployment of different devices and of the IoT network infrastructure.

Keywords: Internet Of Things (IoT), water-level, Long Range Wide Area Network (LoRaWAN), machine learning.

1. Introduction

Competition for clean and safe water has, to a great extent, contributed to water management difficulties around the globe. Overexploitation of water resources is a considerable constraint on sustainable and safe agricultural development practices, thereby making it an essential factor to the alleviation of poverty. There has been a recognition of water as an essential component of food security [1], with more attention being drawn on the significance of management of water by the United Nations Conference on Sustainable Development 2012, in an attempt to meet the sustainable development goals (SDGs).

The upper Ewaso Nyiro (Ngare Ngiro), found in the Ewaso Nyiro basin, is one of the major rivers in Kenya. The Mt Kenya and Aberdare regions, also found in Kenya and are the main contributors to the Ewaso Nyiro river, have for a long time been the focus areas for water resource management and conservation practises. However, in recent years, there has been an experience of water crises of unknown extent in lower catchments and other areas along the rivers path [2] [3]. The water crises have been brought about by the intensified agriculture, reduced rainfall due to climate change and catchment degradation. These crises, in turn, cause conflicts between water-user communities along the river Ewaso Nyiro path in the lower catchment areas [4]. To ensure equitable distribution and sustainable usage of water available in river channels, effective monitoring is essential. River catchment area degradation in recent years has been severe due to encroachment by people. It is altering run-off and infiltration rates, accelerating soil erosion and increasing sediment transport and deposition. In a quest to protect the catchment from encroachment, monitoring water level can be an important source of data which can be used to quantify the rate of catchment degradation [4].

This paper describes the development of a water-level monitoring system for the Ewaso-Nyiro lower catchment that will be the first step in quantifying and discovering the justification statements stated. The rest of the paper is organised as follows. Section 2 presents the objectives. Section 3 presents the methodology used to design the sensor system and describes the main components of the system. Section 4 presents results of a test deployment of the system while section 5 describes the business benefits of the system. Section 6 concludes the paper.

2. Objectives

The main objectives of this work are:

1. To design a sensor system to monitor water-level in a river channel.
2. To deploy the LoRaWAN IoT network at Ol-Pejeta conservancy to facilitate data transmission.
3. To integrate the sensor system and the LoRaWAN Radio for long range low power data transmission.
4. To develop web infrastructure to visualize and store sensor data.
5. To utilize machine learning models/algorithms in performing anomaly detection on the water level data collected

3. Methodology

The IoT sensor system design is based on the Multitech mDot which is an Arm®[1] Mbed™ programmable LoRa module from Multitech[2] which is ideal for fast prototyping. The modules are programmed using the mbed[3] platform (mbed cli – mbed command line programming tool), which allow development of software in C/C++ and provide drivers/libraries for the peripheral devices connected to the MCU. The design also incorporates a Maxbotix[4] MB1010 ultrasonic sensor for river water-level measurement and a custom PCB designed and etched to house all the components.

Data transmission relies on a LoRa network server able to decode LoRa data packets from fixed position nodes and relay them to a database for storage, awaiting processing. The network server we used is provided by The Things Network[5] (TTN). Through the utilization of Python-MQTT (Message Queuing Telemetry Transport), data is transferred from TTN to an InfluxDB database located in a Google Cloud[7] (GCP) virtual machine instance (Compute Engine) for storage. From there, data is then visualized on a Plotly-Dash[6] web application WEB APP. Figure 2 shows a schematic diagram of the system.

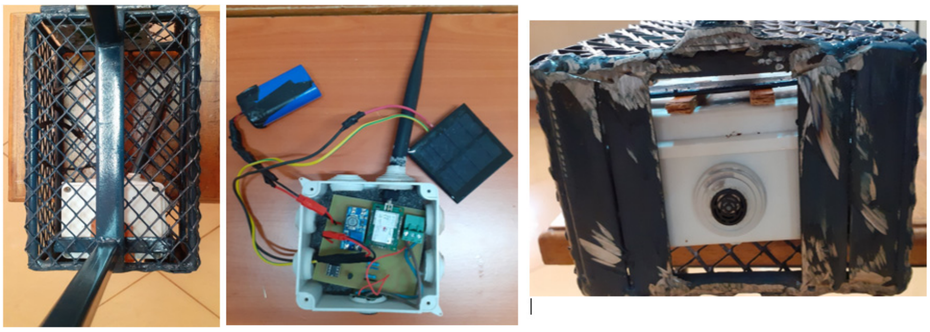

Figure 1: Prototype ready for deployment (development stages)

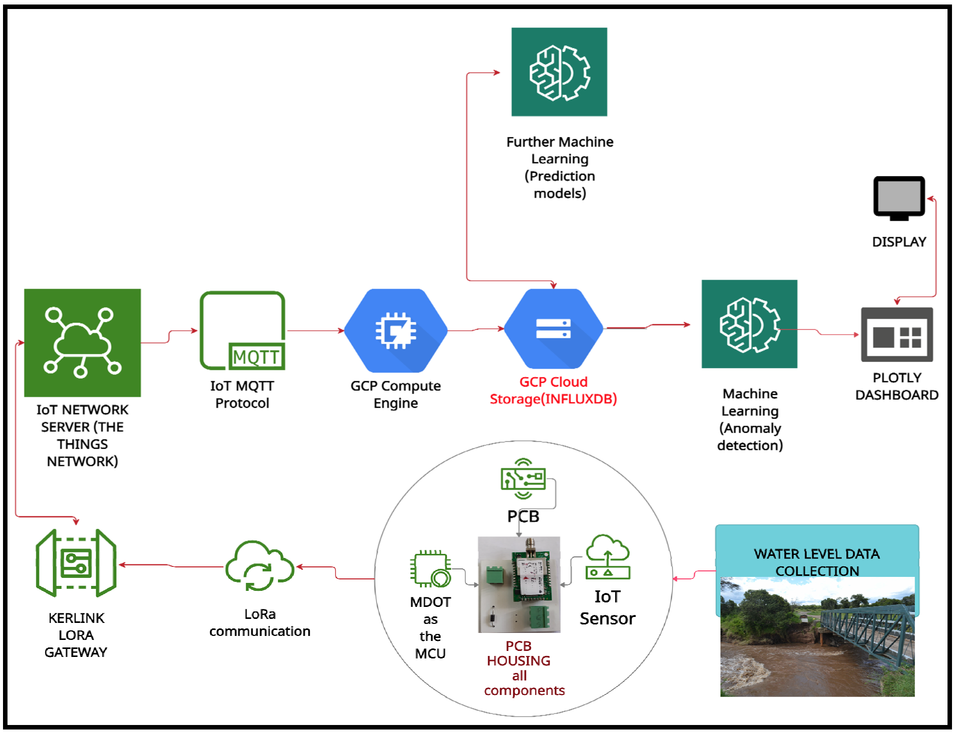

Figure 2: Schematic diagram of the system

The system was deployed along River Ewaso Nyiro at Ol-Pejeta conservancy, which is also home to a Wildlife Techlab setup to develop and test conservation technology. To avoid damage by primates in the conservancy, we designed a metallic cage, shown on Figure 3 and Figure 1, to house and secure the prototype and we have been receiving river water-level data since May 2020.

Figure 3: Deployment

1 https://developer.arm.com
2 https://www.multitech.com
3 https://www.mbed.com/en/
4 https://www.maxbotix.com/Ultrasonic_Sensors/MB1010.htm
5 https://www.thethingsnetwork.org/
6 https://plotly.com/dash/

3.1 Machine learning – Anomaly detection

There is high probability of erroneous readings (anomalies) from sensor prototypes deployed under potentially harsh weather conditions. This brings up the issue of detecting and eliminating these anomalies [5]. We used the clustering-based unsupervised approach – **KMeans,** to detect and eliminate anomalies from our time series, water level data. KMeans is able to cluster correct data points and anomalies in different clusters and with a few python operations we eliminated the anomalies [7].

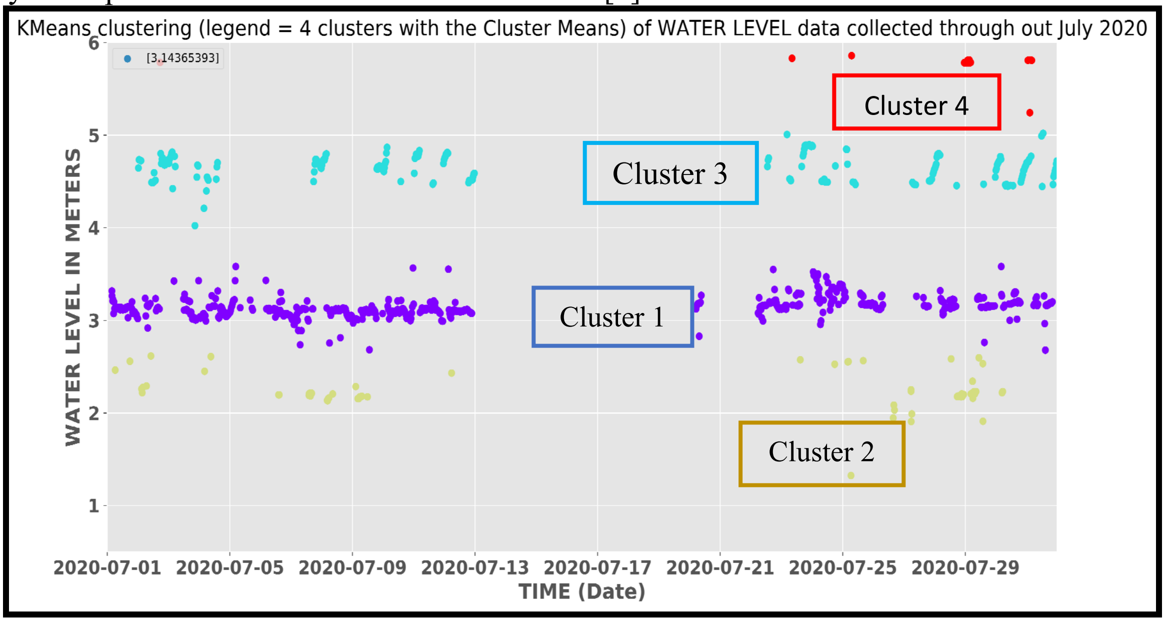

Figure 4: How the data is clustered by the KMeans clustering algorithm

Since **Purple Cluster 1 on figure 4** (cluster 2 between the 3 and 4 meter mark) has the accurate and highest number of data points it is retained by executing simple Python operations[7] and the points (anomalies), in the other clusters, Cluster 2, Cluster 3 and Cluster 4 are scrapped off the dataset.

3.2 Gateway installation and Radio Mapping

The LoRaWAN gateway, also known as the concentrator, is used to relay data packets between the end devices (nodes) and the network server via the Internet. It communicates over multi-channels with multi-spreading factors. With this technique, nodes communicate with the gateway using different channels and data-rates without pre-negotiation and enables the gateway to accommodate about 10000 end devices at a go. To facilitate data transmission at Ol-Pejeta, a Kerlink Outdoor gateway was configured and installed on a WIFI tower at a height of approximately 16 meters above the ground. The height provided adequate radio coverage of a large section of the upper Ewaso Nyiro River running through the conservancy.

The robust operation and efficient deployment of many IoT systems rely on the deployment of gateways and relays to ensure quality wireless coverage. Radio mapping aims to predict network coverage extent based on a small number of link measurements from sampled locations [6].

We conducted radio mapping at Ol-Pejeta conservancy[8] to determine transmission range of LoRa[9] enabled prototypes developed, and also to test the deployment along river Ewaso Nyiro within the conservancy. We mounted two LoRaWAN gateways at Ol-Pejeta house, a Kerlink Gateway at approximately 16 meters and LoRix One Gateway at approximately 13 meters.

We deployed 3 devices at various points within the conservancy. The devices and gateways were already connected to The Things Network and the network server was relaying radio propagation data to an InfluxDB[10] database for storage, awaiting processing.

3.3 Radio mapping results

The Received Signal Strength Indication (RSSI)

This refers to the signal power that is received in mill watts (mW), and it is measured in dBm. How clear/reliable the transmission of data is, between a sender and a receiver can be measured using this value. Received signal strength indication, i.e., RSSI, is usually a negative value; hence, the signal is better when it is more positive (closer to 0). The value ranges of typical LoRa RSSI is -140 dBm to -30dBm.

For our radio mapping experiments, we computed the mean RSSI for each of the 5 test locations that were used. At approximately 150m away from the gateway, the LORIX One gateway outperformed the Kerlink gateway by a margin of 10dBm as well as at the furthest distance (approximately 7.5km) by 13dBm dBm. The highest Received signal strength figure was recorded at the nearest test location (0.152km) and diminished as we approached test locations far away from the gateway, as depicted by **Figure 5.**

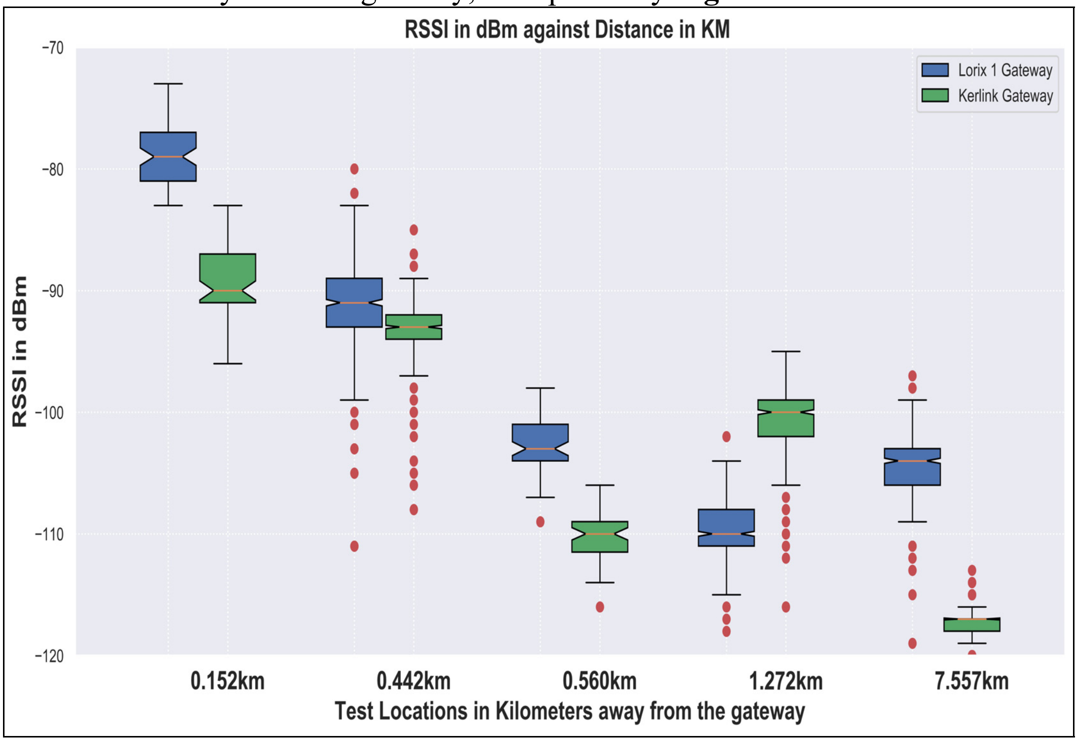

Figure 5: The Received Strength Plots for the 5 Test Locations.

Plots in **Figure 5** provides a quick graphical examination of the RSSI for each of five (5) test locations for each gateway. Outlier RSSIs were highly realized in test location 2 and they are plotted as individual points, while none were realized at location 1 (nearest to the gateways). However, location 1 depicts the highest notable degree of dispersion (spread) and skewness for both gateways. There is a general non-linear variation of the median positions of the RSSI, usually determined by various parameters, which include free space loss, shadowing, reflection and transmission, diffraction, among others.

3.4 Programming the Multitech mDot

Mbed enabled modules such as the Multitech mDot used in this work can be programmed in various ways including using an online program compiler and the offline mbed command line interface (CLI). I used the CLI to compile the program used. Code used in this can be accessed at https://github.com/ciiram/mdot-maji.

3.5 PCB Design

Since the components needed could not be soldered directly onto the Multitech mDot pins, we developed an etched circuit to harbour all the components and also facilitate deployment. The circuit design was developed using the KICAD PCB development and design software. The circuit consisted of the battery/power socket, a socket to harbour the Multitech mDot and an ultrasonic sensor socket.

3.5.1 Main Circuit

The mDot is powered by a lithium 4400mAH rechargeable battery which is charged by a 150mAh Solar panel through a diode. The analog pin **(AN)** of the MB1010 is connected to one of the analog pins on the mDot to facilitate data collection.

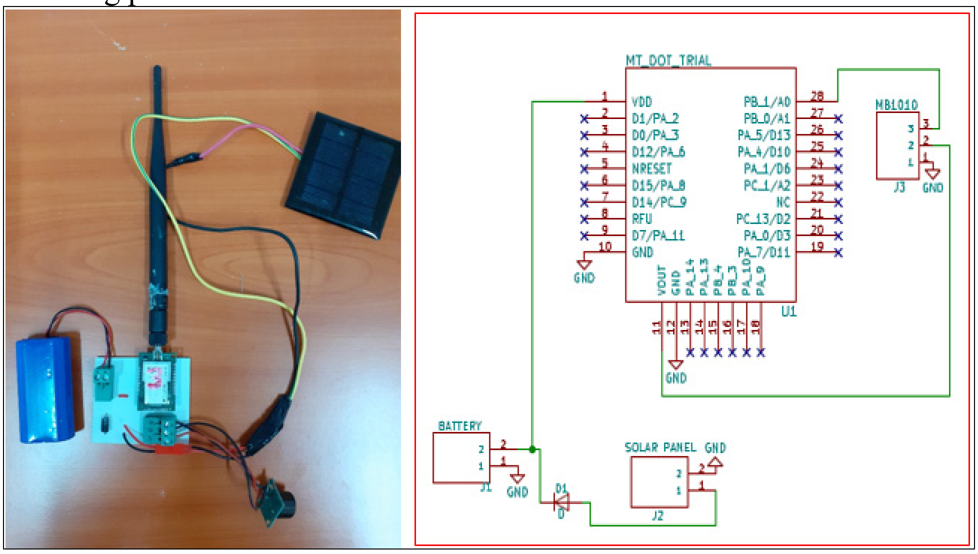

Figure 6: (LEFT) Complete system, (RIGHT) Circuit diagram of the system

4. Results

We deployed three prototypes along the river Ewaso Nyiro channel and have been getting data since May 2020. We also created a dash web application to display the water level data from the database. (DASH web app link: (https://water-monitoring-258811.wl.r.appspot.com). Figure 6 below shows comparison between water level daily values, for all days between (May 2020 - January 2021) and precipitation data from one weather station controlled by TAHMO[11] (Trans-African Hydro-Meteorological Observatory). We found out that there was a high degree of similarity between the rainfall profile and the water level profile. This meant that our dataset could be trusted. We also compared our data set to the data form other stations and the results were also similar. The water level means for every day were obtained after the collected data was passed through the anomaly detection process outlined in the methodology. Because our IoT network - LoRa was not working for some few days (5% of the total), we had to interpolate data points for those days. We used various interpolation algorithms to come up with suitable water level values for these days.

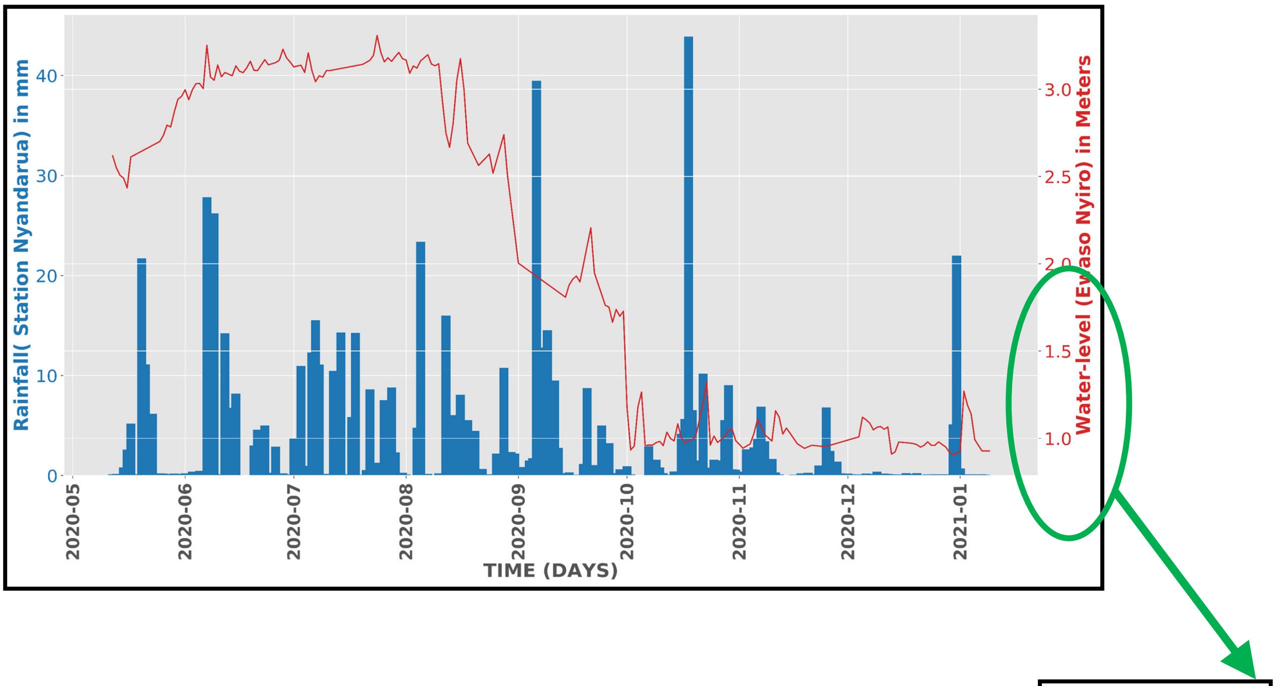

Figure 7: Rainfall profile comparison to the water level profile

4.1 Catchment analysis

To analyse the catchment, we decide to check how long a spike in water level takes to appear after a spike in rainfall occurs. According to the test part we considered on Figure 7, the spike in water level occurred two days after a spike in rainfall. This shows the catchment area is possibly not degraded to a large extent. If the spike in water level was picked up by the sensor hours after the spike in rainfall, it would have meant that the catchment is possibly degraded to a large extent and the water runoff had nothing to obstruct it. In the future we aim to develop prediction machine learning models for the water level profile with the help of the rainfall data from TAHMO. By predicting and collecting water level data various aspects of the river and the basin can be quantified.

5. Business Benefits

The system developed is capable of large-scale deployment along river channels in Kenya and across the globe. The system also has the potential to reduce the cost of monitoring water level in rivers and other artificial water storage facilities like dams, by replacing construction of stream gauges and manual measuring exercises using calibrated metal rods. The system also requires very little maintenance, which makes it cheap for a long term installation consideration. Also the system is small compared to other options available. This fact facilitates large-scale deployment. Since data acquisition and transfer does not require human presence, the system can be deployed in areas where accessibility is a problem. Some of disadvantages of the system are, high initial installation cost, intensive computer processing is needed, and the system is prone to vandalism. According to our assessment the advantages out-weigh the disadvantages, which makes the system viable for large-scale deployment.

6. Conclusion

This paper has outlined the real-time monitoring of water-level in rivers by leveraging Internet of Things (IoT) and Machine Learning. We collected useful data from the prototypes deployed and in turn we were able to track water levels in River Ewaso Nyiro channel. This data can be used by other governmental or non-governmental bodies to predict the future behaviour of the river, if some climatic and human conservation factors are considered. In the study, we learnt that small and cheap sensor systems can be used to quantify various natural phenomena. The initial cost of setting up the networks to handle data transfer is high but long term benefits of data collections and analysis can be very crucial in the making of various environment related decisions. Apart from the IoT network blackouts and very little vandalism cases, our systems were able to collect data for a long stretch of time. This fact means our system is capable of large-scale deployment.

In the near future, we aim to incorporate multiple data sources such as weather data from TAHMO to build water level - machine learning prediction models and use other machine learning algorithms in anomaly detection. Also, we plan to expand our sensor network by deploying more water level monitoring devices, turbidity monitoring devices and flow rate monitoring devices. The long term vision of this project is to collect enough data that can be used in the development of inundation models. Inundation models are used in flood forecasting in a river basin. The main input of inundations models is river water-level data collected over a long period of time. The other vital input is high resolution elevation maps (depth maps) of the river basin terrain. This means that work in water-level data collection is the first key step in the development of accurate inundation models for the Ewaso Nyiro River basin.

Acknowledgement

We would like to thank the management at Ol-Pejeta Conservancy (Kenya) for allowing us to test our systems at the conservancy and also use the Ol-Pejeta Tech Lab to prepare for deployments and field tests. We also thank Mr Stephen Mathenge from Dedan Kimathi University of Technology for his technical support during system development and deployment. We also thank TAHMO (The Trans-African Hydro-Meteorological Observatory) for giving us access to their weather station data from stations in the Ewaso Nyiro basin. Finally, we thank the NRF (National Research Fund) for funding this work through a multidisciplinary research award.

References

[1] UNESCO, “Facts and Figures from the United Nations World Water Development Report 4,” p. 16, 2012.
[2] J. Aeschbacher, H. Liniger, and R. Weingartner, “River Water Shortage in a Highland–Lowland System,” *Mountain Research and Development*, vol. 25, no. 2. pp. 155–162, 2005, doi: 10.1659/0276-4741(2005)025[0155: rwsiah] 2.0.co; 2.
[3] B. Notter, L. MacMillan, D. Viviroli, R. Weingartner, and H. P. Linger, “Impacts of environmental change on water resources in the Mt. Kenya region,” J. Hydrol., vol. 343, no. 3–4, pp. 266–278, 2007.
[4] R. Hirji SR, Ministry of Environment and Natural Resources, Ministry of Agriculture, National irrigation board, Ewaso Nyiro North and Ewaso Nyiro South Basin Authorities (2000) “Water Resources Management Issues, Challenges and Options: À Framework for a Strategy” Draft Issues Paper (Kenya).
[5] Giannoni F, Mancini M, Marinelli F. (2018) Anomaly Detection Models for IoT Time Series Data. ArXiv preprint arXiv: 1812.00890.
[6] O. Chipara, G. Hackmann, C. Lu, W. D. Smart, and G.-C. Roman, “Practical modelling and prediction of radio coverage of indoor sensor networks,” *Proceedings of the 9th ACM/IEEE International Conference on Information Processing in Sensor Networks - IPSN ’10*. 2010, doi: 10.1145/1791212.1791252.
[7] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2016.

7 https://kabi23.github.io/anomalydetection/
8 https://www.olpejetaconservancy.org
9 https://lora-alliance.org
10 https://www.influxdata.com
11 https://tahmo.org/weather-stations/

