Article

Sensor-Based River Monitoring System: A Case for Kikuletwa River Floods in Tanzania

Lawrence Mdegela [*] , Yorick De Bock , Esteban Municio , Edith Luhanga , Judith Leo , Steven Latré

Sensor-Based River Monitoring System: A Case for Kikuletwa River Floods in Tanzania

Lawrence Mdegela [1,4]*⬤, **Yorick De Bock** [1], **Esteban Municio** [2], **Edith Luhanga** [3], **Judith Leo** [4] and **Steven Latré** [1]

[1]   University of Antwerp - imec; lawrencenehemiah.mdegela@uantwerpen.be (L.M.);
     Yorick.DeBock@uantwerpen.be (Y.D.B.); Steven.Latre@uantwerpen.be (S.L.)
[2]   i2CAT Foundation; esteban.municio@i2cat.net
[3]   Carnegie Mellon University Africa; eluhanga@andrew.cmu.edu
[4]   The Nelson Mandela African Institution of Science and Technology; judith.leo@nm-aist.ac.tz
*   Correspondence: lawrencenehemiah.mdegela@uantwerpen.be; Tel.:+32494594736

Abstract: Reliable and accurate flood prediction is a challenging task in poorly gauged basins due to data scarcity. Data is an essential component of any AI/ML model today, and the performance of such models hugely depends on the availability of sufficient amount of trusted, representative data. However, unlike a few well-studied rivers, most of the rivers in developing countries are still insufficiently monitored, which significantly hinges the design and development of advanced flood prediction models and early warning systems. This paper presents a multi-modal, sensor-based and near-real time river monitoring system to produce a multi-feature data set for the Kikuletwa river in Northern Tanzania, an area that heavily suffers from frequent floods. Our deployed system, which gather information about river depth levels and weather at several locations, aims at widening the ground truth of the river characteristics and eventually improve the accuracy of flood predictions. We provide details on the monitoring system used to gather the data as well as report on the methodology and the nature of the data. Finally, we present the relevance of the data set in the context of flood prediction, discussing the most suitable AI/ML-based forecasting approaches, while also highlighting some applications of the data set beyond flood warning systems.

1. Introduction

The lower Kikuletwa sub-catchment is an area of about 6657 km2 located in the north-western part of the Pangani River basin, North-east of Tanzania, south of Mount Kilimanjaro (see Figure 1). River Kikuletwa has been subject to several flood events throughout its history, the main victims of the floods are the villages along the river especially the lower parts downstream. In April 2020, according to flood-list [1], more than 2,700 households in Moshi district were swept away by floods after a period of heavy rainfall leaving dozens homeless and destroying important infrastructures. Despite the fact that these events are inevitable, not much has been done in terms of mitigating the impacts to the communities around. Improving flood warning mechanisms, is without a doubt, one of the most straightforward things to do in order to reduce such impacts. Among the most successful technologies to identify, classify and predict events such as floods, are Artificial Intelligence (AI) and Machine Learning (ML) models [1–3]. Given an appropriate data set (large enough, representative, multi-featured, etc.) these models can significantly outperform traditional approaches. In this sense, data sets are the foundation of training, evaluating, and bench-marking machine learning models, and play a key role in the advancement of the field [4].

---

[1]   https://floodlist.com/africa/tanzania-floods-arusha-kilimanjaro-april-2020
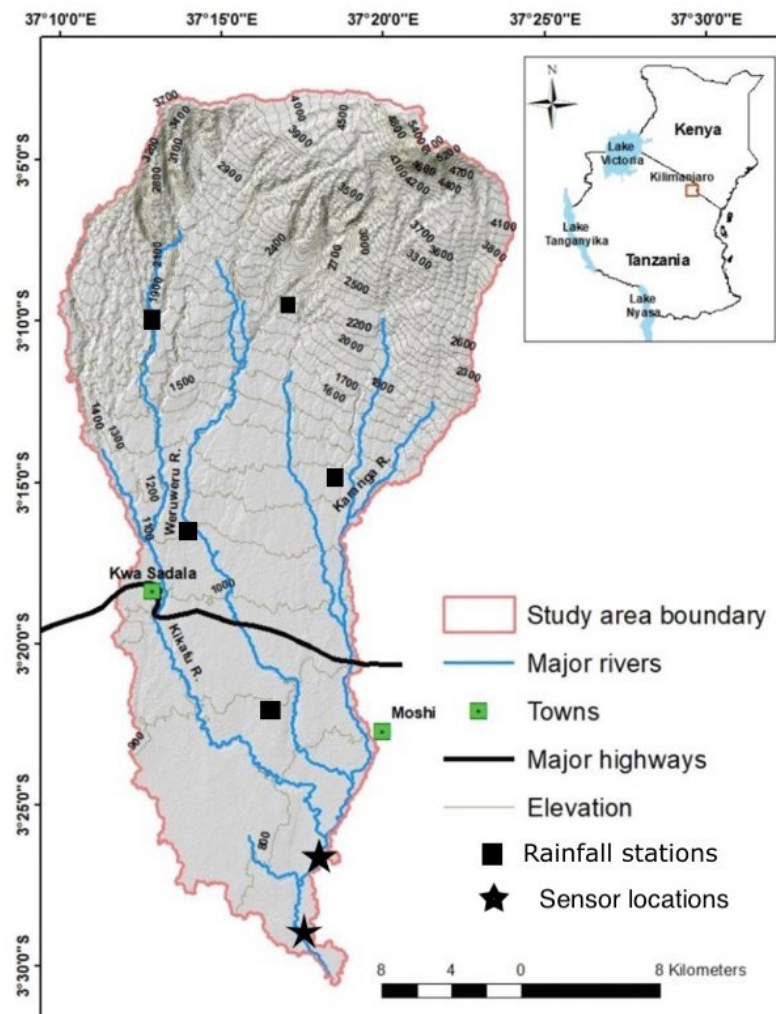
Figure 1. Lower Kikuletwa sub-catchment

However, data scarcity has proved to be one of the key challenges towards achieving better accuracy in flood warning systems. This is partly due to poorly gauged basins. Currently, there are manual water level staff gauges (as depicted in Figure 2) installed in some parts of the river for monitoring and recording river-levels. A staff gauge is like a big ruler, placed in or beside a watercourse, and from which it is possible to measure directly the height of the water. However, despite the fact that staff gauges are cheap to implement, there are several challenges that are associated with this type of monitoring when using them for river level measurement. Some of the challenges include errors due to bad angle reading, turbulence, or dirt on the scale due to debris. However, one of the most relevant challenge is missing out on the high flow readings due to when floods occur, no one is around to read the peak flows of the water until it is safe to do so. Missing out on the high flows is critical since they are the ones that enable definition of the threshold values at which flood warnings will be issued. This is especially the case at the Kikuletwa river.

Figure 2. Water level staff-gauges

In order to deal with these challenges, in this article we present a sensor-based river monitoring system that enables robust, autonomous, and large-scale data gathering in a low cost manner without the need of human intervention. We provide details on the system design and methodology, and demonstrate its monitoring capabilities by deploying a proof-of-concept in the Kikuletwa river. The in-field deployed monitoring system proves to be successful in producing continuous data series of river water level and weather data (e.g., river depth, rainfall, wind, etc.) and is able to export the information in an standardized manner. Finally we report and characterize the obtained data set and discuss its relevance for data-feeding early warning and forecasting systems based on AI/ML models.

The rest of the paper is structured as follows. In Section 2 we provide an overview of the state of the art. Section 3 details the design of the monitoring system and presents the methodology used for collecting and creating the data set. Next, we give details of the composition and nature of the data-set in Section 4, followed by a discussion in Section 5, where we comment on the most suitable early warning and prediction systems for the dataset. Finally we conclude the paper in Section 6.

2. Background and State of the Art

The implementation of sensor-based river monitoring systems has gained prominence in recent years due to the increasing need for accurate and real-time data for effective water management and flood prediction [5]. In the last decade, reports show that there has been an increase in the frequency of flood events all over Tanzania, with almost the same impacts each year; Loss of lives, infrastructures and other societal systems[2], [3], [4]. Worryingly, with the current climate change, such events are expected to be more frequent. This persistence observed in occurrence of such events with subsquent impacts is mainly because of limited knowledge in the manner in which they occur. This is partly due to data scarcity coming from poorly gauged or ungauged catchments. In order to prepare for future natural hazards such as river floods, it is necessary to better understand present and past events. Unfortunately, this is not the case as far as the status quo is concerned in Tanzania.

---

[2] Burundi and Tanzania – Floods Leave Homes Destroyed, Hundreds Displaced. https://floodlist.com/africa/burundi-tanzania-floods-late-february-2021

[3] Tanzania – Severe Flooding in Mtwara Region After Torrential Rainfall. https://floodlist.com/africa/tanzania-flood-mtwara-january-2021.

[4] Tanzania – 12 Killed in Dar Es Salaam Flash Floods. https://floodlist.com/africa/tanzania-daressalaam-floods-october-2020.

Kikuletwa River in Northern Tanzania is one of the many rivers that face challenges in terms of flash floods, which cause significant damage to infrastructure, agriculture, and loss of life. Implementing a sensor-based monitoring system is critical for mitigating the impact of floods and ensuring sustainable water management practices [6]. Traditional river monitoring systems employed, the manual staff-gauges, have shown little success, particularly, when it is time to provide early warnings for floods. The use of sensors to automatically monitor different river characteristics, is one of the promising mechanisms in overcoming this challenge. With enough and quality data, reliable and accurate early warning systems can be developed and deployed. Sensor-based monitoring systems use a network of sensors placed along the river to collect data on water levels, discharge, and rainfall, among other parameters [7]. This data is transmitted in real-time to a central server for analysis and integration with other data sources, such as satellite imagery and meteorological data [8]. The use of sensors provides a cost-effective and efficient solution compared to traditional manual monitoring methods [9]. The collected data can be used to generate early warning systems for flood events, allowing communities to prepare and evacuate in advance. In addition to providing real-time data for flood prediction and water management, sensor-based monitoring systems also offer several other benefits. For instance, they provide a continuous and comprehensive data record that can be used for scientific research and policy-making [10].

Further evidence from literature shows sensor-based monitoring systems are highly effective in providing real-time data for flood prediction and water management. This is the case, for example of the water quality management project reported in [11], where different types of wireless sensor nodes were employed to obtain continuous real time data for water quality control. Similarly, in projects such as [12], a wireless sensor network was installed in river Sitnica in Kosovo for real time surface water quality control. More examples can be found in [11,13–16]. There are also examples of works that use such networks for early warning systems. For example, in Argentina, a sensor-based early flood detection and warning system was developed by Libelium in 2018, this was after several costly and unsuccessful efforts to solve a persistent flood problem in some villages [17]. However their focus was mainly on flood detection solely based on the river level thresholds. In contrast, our work generates exportable monitoring data to systematically feed AI/ML algorithms and aim for more sophisticated flood warning system which include multi-feature flood forecasting. In this sense, we rather follow the trend paved by other works such as [1,3,18,19].

Unlike most of these deployments, in a developing country like Tanzania, there are several challenging factors such as cost (i.e., both in terms money and human resources) and available infrastructure that also need to be considered in the design. It is in this view that our work presents a low-cost, self-powered, automated monitoring system that aims to solve such challenges through a standalone system that uses low-cost devices and removes human dependency. The following Section 3 provides details on how the system is designed to achieve so.

3. Methodology and data sourcing

A preliminary study was carried out to find out which were the most relevant locations to have sensors installed along the river, from the source to the mouth in the Pangani river, with special emphasis on the riverside villages. Two locations were identified, the first is the bridge at Kikavu chini village and another one is found at the border of Kilimanjaro and Manyara regions, the Kikuletwa bridge. This was done with the help from the Pangani Water Board (PWB) authority, which is responsible to oversee all the aspects of ground and surface water in the area. Apart from the data being need for this research, PWB required a permanent water monitoring system for producing a satisfactory water flow in order to control its behavior.

Figure 3. Sensor installation at Kikuletwa bridge

3.1. Waspmote Plug&Sense Smart Agriculture PRO

The sensor are all attached to an Agriculture v30 Board called Waspmote Plug&Sense Smart Agriculture PRO (fig. 4). Multiple environmental parameters can be monitored using this device including, air and soil temperature, humidity, solar radiation, wind speed and direction, rainfall, atmospheric pressure, etc. In our case, main parameters being monitored are river water-level, rainfall, wind speed and direction. The sensor sockets are configured as shown in the table 1

Table 1. Sensor sockets configuration.

Sensor socket	Parameter	Reference
A	Weather station WS-3000 (anemometer+wind vane+pluviometer)	9256-P
F	Ultrasound(distance measurement)	9246-P

Two devices were installed in two different river locations, with each having an ultrasonic sensor for distance measurement, and a weather station containing pluviometer for rainfall measurement, wind vane for wind direction and anemometer for wind speed.

3.1.1. Weather station

The weather station consists of three different sensors: a wind vane, an anemometer and pluviometer. It is connected to Waspmote through six wires that are connected to the terminal block as can be seen in figure 4 where the anemometer is connected to the vane through an RJ11 socket.

Figure 4. Plug&Sense! Smart Agriculture PRO with all parts connected

Specifications for each of the sensors included in the weather station is summarized in table 2

Table 2. Weather station sensors specifications.

Sensor	Specifications
Anemometer	Sensitivity: 2.4km/h/turn, Wind Speed Range: $0 \sim 240$km/h, Height: 7.1cm, Arm length: 8.9cm, Connector: RJ11
Wind vane	Height: 8.9cm, Length: 17.8cm, Maximum accuracy: 22.5^0, Resistance range: 688 $\Omega \sim 120k\Omega$
Pluviometer	Height: 9.05cm, Length: 23cm, Bucket capacity: 0.28 mm of rain

3.1.2. Ultrasonic Sensor (MaxSonar®from MaxBotix$^{\text{TM}}$)

The ultrasonic sensor (fig. 5) measures the distance between itself and the surface of water. It emits ultrasonic waves which rebound in the water and offers the system with distance between the node and the water. The water level of the river is then calculated. Specifications for the ultrasonic sensor are listed in table 3

Figure 5. Ultrasonic I2CXL- MaxSonar®-MB7040 from Max- Botix^{TM} sensor

Table 3. Weather station sensors specifications.

Sensor	Specifications
Ultrasonic	Operation frequency: 42 kHz, Maximum detection distance: 765cm, Interface: Digital Bus, Power supply: 3.3 ~ 5V, Consumption: 2.1mA (powered at 3.3V) to 3.2mA (powered at 5V), Consumption(peak): 50mA (powered at 3.3V) to 100mA (powered at 5V), Usage: Indoors and outdoors (IP67)

3.1.3. 4G module

Waspmote Plug&Sense is integrated with a 4G radio module for wireless communication, which enables connectivity to high speed LTE, HSPA+, WCDMA cellular networks. It is oriented to work with Internet servers, implementing internally several application layer protocols, which make it easier for data to be sent to the cloud. HTTP navigation, downloading and uploading of different contents to a web server can also be done. Furthermore, secure connections using SSL certificates and setting TCP/IP private sockets can be set. In the same way, the FTP protocol is available which is, undoubtedly, very useful when file handling in an application is required.

3.1.4. Battery and Solar Panel

The devices are equipped with 2 types of batteries for the OEM line, a 6600 mAh, rechargeable lithium-ion battery (Li-Ion), with 3.7 V nominal voltage and a 52000 mAh, non-rechargeable battery, with 3.4 V nominal voltage. Waspmote is equipped with a control and safety circuit through which battery charge current is always made sure it is adequate. On the other hand in the power sources for the devices we have a solar panel. Waspmote comes with a rigid solar panel of 7 V , 500 mA which can allow up to 12V, and maximum charging current through the solar panel is 300 mA.

3.2. River Depth measurement

In order to have an initial river depth that will be used as a reference when determining river levels, we measured the depth of the river at both of the locations. The ultrasonic sensor used is measuring the distance between itself and the surface of water. The distance is later transformed to river levels having determined the initial depth. We applied bathymetry technique in determining initial depth of the river, which is a technique used in measurement of depth of water in oceans, rivers, or lakes [5]. Just like topographic maps, bathymetric maps uses lines to connect points of equal depth of the river. Echo sounding technology was employed (fig. 6). Echo sounding is a bathymetry technique of measuring depth using sonar. [6].

---

[5]   https://education.nationalgeographic.org/resource/bathymetry
[6]   https://en.wikipedia.org/wiki/Echo_sounding

The sonar (echo sounder) (fig 6), "pings" a beam of sound downward to the river floor. The time taken for the sound to travel through the water, bounce off the river floor and return to the echo sounder informs of the distance to the river floor. This was done in both locations before deploying the sensors.
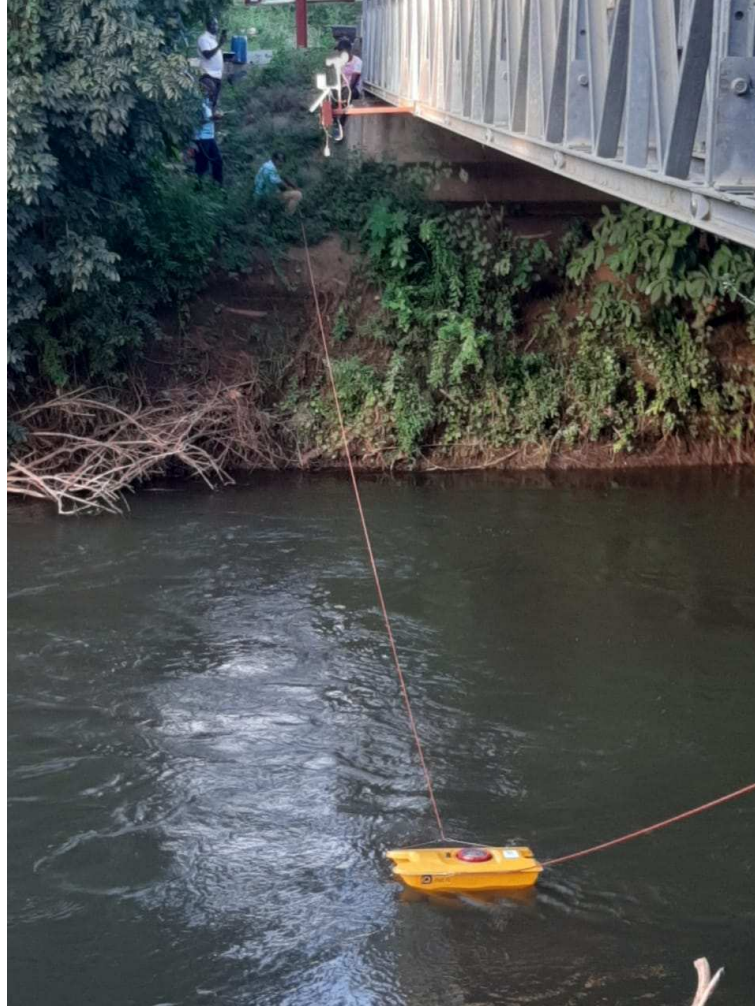
Figure 6. Echo sounding technique employed to determine initial river level

3.2.1. Level measurement at both bridges

Kikavu bridge is located at Latitude -3.44 and Longitude 37.30, it is more upstream of the two locations while Kikuletwa is at latitude - 3.55 and longitude 37.31 more downstream. The Ultrasonic sensor at Kikavu bridgre (kkv) was mounted at the height of 765 cm from the surface of water, which at the time of installation was at normal flow (flow during dry season). Initial river depth at Kikavu was measured at 94cm. Total distance from the sensor to the riverbed at Kikavu can now be written as:

$$d_t = d_0 + d_v = 765 + 94 = 859 cm \tag{1}$$

So, to get hourly river height from the incoming sensor data we use the following formula:

$$d_v = d_t - d_0 \tag{2}$$

Where d_v, is the river height at Kikavu bridge, d_t, distance from the sensor to the riverbed, and d_0, is the sensor-read distance or the distance from the sensor to the water surface.
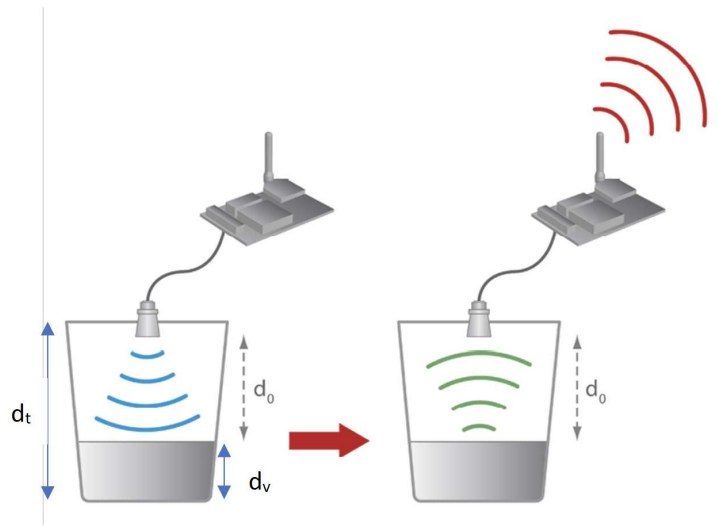
Figure 7. Ultrasound sensor measurement process.

The same procedure was was executed at Kikuletwa bridge with the sensors this time being mounted at the height of 400 cm from the surface of water. Initial river depth at Kikuletwa was measured at 130.5cm. Total distance from the sensor to the riverbed at Kikuletwa bridge can now be written as:

$$d_t = d_0 + d_u = 400 + 130.5 = 530.5 cm \tag{3}$$

So, to get hourly river height from the incoming sensor data we use the following formula:

$$d_u = d_t - d_0 \tag{4}$$

Where d_u, is the river height at Kikuletwa bridge, d_t, distance from the sensor to the riverbed, and d_0, is the sensor-read distance or the distance from the sensor to the water surface.

3.3. Programming of the sensor nodes

A programming cloud service provided by Libelium has a special application, P&S programmer, where each sensor node was programmed to send data to Libelium cloud bridge (figure 8, 9, 10, 11). Basic configurations for each sensor socket were done through P&S programmer, such configurations include communication and protocol of the destination block, in this case, 4G was set as a communication module. Other settings include sleep time, set at 3600 seconds (1 hour) for energy efficiency. This is the amount of time the device spends in sleep mode before a new cycle (sensor reading + transmission is performed). Furthermore, critical battery warning setting, where 3 thresholds (60%, 40%, and 20%) were set. A warning packet is sent upon reaching each threshold. After all the settings, we compiled to create binary files which are then uploaded to the devices through Smart Devices application. Meanwhile, a valid API key and encryption layer functions were transparently provided for each node. A valid API key and encryption functions are a must when programming the sensor nodes using Waspmote IDE. API keys authenticate calls to the Libelium cloud bridge service discussed in section 3.4.

Figure 8. Device selection by serial number

Figure 9. Sensor socket selection

Figure 10. Setting up communication and protocol destination block

Other settings

Sleep time* ⓘ

Days:	Hours:	Minutes:	Seconds:
− 0 +	− 1 +	− 0 +	− 0 +

Total sleep time for this node: 3600 seconds

GPS module ⓘ

GPS timeout ⓘ   − 240 +

☐ GPS location (latitude and longitude)
☐ GPS speed (km/h)
☐ GPS course (degree)
☐ GPS altitude (m)

Critical battery warning

A warning packet will be sent upon reaching each
threshold: ⓘ

Threshold 1   − 60 +   Range from 1 to 99%

Threshold 2   − 40 +   Range from 1 to 99%

Threshold 3   − 20 +   Range from 1 to 99%

RTC Watchdog mode ⓘ

RTC Watchdog period (minutes) ⓘ   − 60 +

Figure 11. Other settings

3.3.1. Smart Devices App

Smart devices App was one of the most important tools during the process of installing programs to Waspmote Plug & Sense. Binary files generated from the P&S programmer, were uploaded to each of the devices though Smart devices app. Latest version of Java Development Kit (JDK) is a requirement before using the Smart devices App, in our case during the time of installation we used Java SE 18. From the Smart Devices App, all the Plug & Sense devices are listed updates to programs were done from there after selecting a firmware that needed to be upgraded.

3.4. The cloud Bridge

Data gathered by the sensor is sent to Libelium cloud bridge were it can be visualized in it's raw state and it's status can be checked from time to time. The bridge is a service that allows users to send information from any IoT device to the main worldwide cloud platforms simultaneously and without having to implement each specific cloud protocol or authentication methodology. In our case we connected the bridge to Microsoft Azure. The Libelium cloud service has as buffer of limited size that is cleared after data is sent to the final cloud service. The bridge has three main functionalities:

- Configure cloud connectors
- Manage devices
- Configure gateways

All these functionalities are oriented to send data from the sensor nodes to the final cloud service.

3.4.1. Data Flow

Data sources are sensor nodes, which send data to cloud service. In the cloud service we have implemented a dashboard that is fed with data from the sensor nodes. The Libelium cloud bridge connects the sensor nodes with the cloud service, Microsoft Azure, as can be seen in figure 12
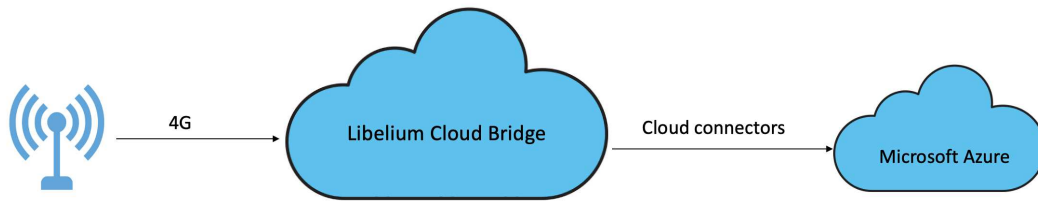
Figure 12. Libelium Cloud Bridge service data flow.

Sensor nodes, through 4G communication protocol, send data directly to the Libelium cloud bridge service. Libelium Cloud bridge service listens to HTTPS requests to receive data from sensor nodes. Valid requests must comply with the following requirements:

- The sensor node must be registered in the Libelium Cloud (Services Cloud Manager) user account
- A valid authentication API Key must be associated with the sensor node
- Integrity of the Libelium Cloud Bridge service encryption layer is respected

3.5. Data Storage and Visualisation

A virtual machine is configured on the data receiving end. Data monitoring and analytics were done via a InfluxDB database and Grafana. Grafana enables the visualization, querying and analysis of data. Libelium cloud bridge, through Microsoft Azure cloud connector, sends data to Azure storage container. A python script is implemented on the Virtual machine to check for new files every two hours and insert into InfluxDB.
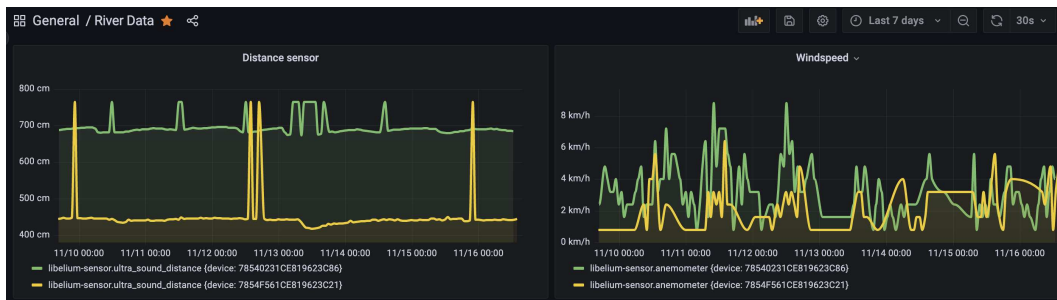
Figure 13. The dashboard on the virtual machine.

4. Data characterization

The data set consists of 6 variables measured hourly. The variable are river level in centimeters (cm) measured by a Ultrasound sensor, current hour rainfall (mm), previous hour rainfall(mm/h), last 24 hours rainfall (mm/day) all measured by the pluviometer, wind speed in kilometers per hour (km/h) by the anemometer and wind direction as direction by the wind vane. The most important features of these are rainfall and river level for they can directly say about the condition of the river and if there will be any flood risks. Rainfall patterns are the main natural factor affecting water levels with periods of wet conditions result in increasing water levels. On the other hand, wind speed and direction affect rainfall [patterns, therefore in this context, we have water level as a dependent variable while the rest are independent.]

4.1. River levels

River water levels are measured hourly in centimeters by the ultrasonic sensor. River level is then calculated based on the initial water level recorded at a corresponding point by applying formulas explained in section 3.2. Since deployment, there has been an almost constant distance reading with an average of 444.12 cm (table 4). This has mainly been due to below average rainfall in the last two rain seasons all over the country due to changing climate [20].

Table 4. ultrasound distance characteristics for November, 2022 at Kikuletwa bridge

| Characteristic	Ultrasound distance
Count	654
Mean	444.12
Standard Deviation	45.62
Minimum	345
25%	435.25
50%	442
75%	445
Max	765
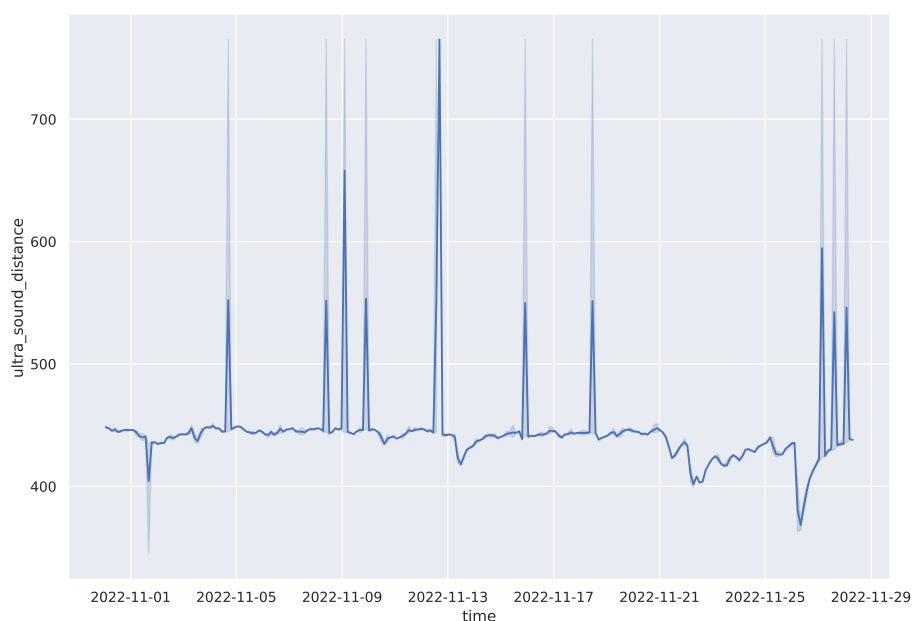
Figure 14. Ultrasound data plot for November 2022 at Kikuletwa bridge.

4.2. Rainfall

The pluviometer gives three types of readings, rainfall for the current hour in millimeters (mm), previous hour (mm/hour) and accumulated rainfall in the last 24 hours (mm/day). The pluviometer has a small bucket of approximately 0.28mm when full. During the measuring process, the switch is closed when the bucket is full and then it is emptied afterwards. Some characteristics of rainfall data are shown in table 5 with a corresponding plot (fig. 15) for the whole period since deployment in April.

Table 5. Characteristics of rainfall data from one location

| Characteristic	Daily rainfall
Count	4871
Mean	0.022
Standard Deviation	0.537
Minimum	0.00
25%	0.00
50%	0.00
75%	0.00
Max	17.323
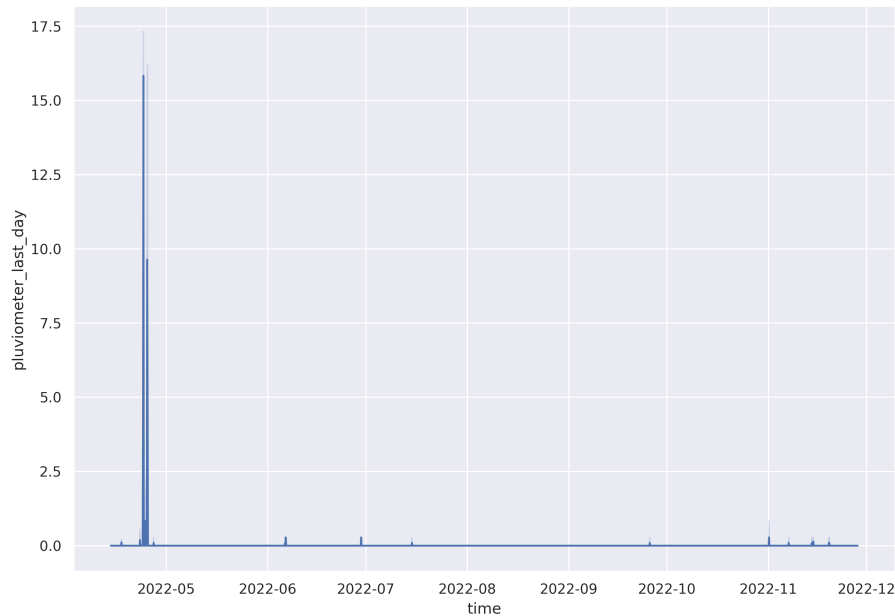
Figure 15. Daily rainfall trend from April to November, 2022.

4.3. Wind speed

The anemometer used consists of read switch that is normally open, and it closes only for a short time when the arms of the anemometer turn to complete a 180^0 angle. The reading of the anemometer is a digital signal, whose frequency is proportional to the wind speed in kilometers per hour (km/h). Example of the data from the anemometer for the month of November is shown in table 6 with a corresponding plot(fig. 16)

Table 6. Characteristics of wind-speed data from one location

Characteristic	Hourly wind speed (km/h) for November, 2022
Count	304
Mean	2.06
Standard Deviation	1.32
Minimum	0.80
25%	0.80
50%	1.60
75%	2.40
Max	8.00
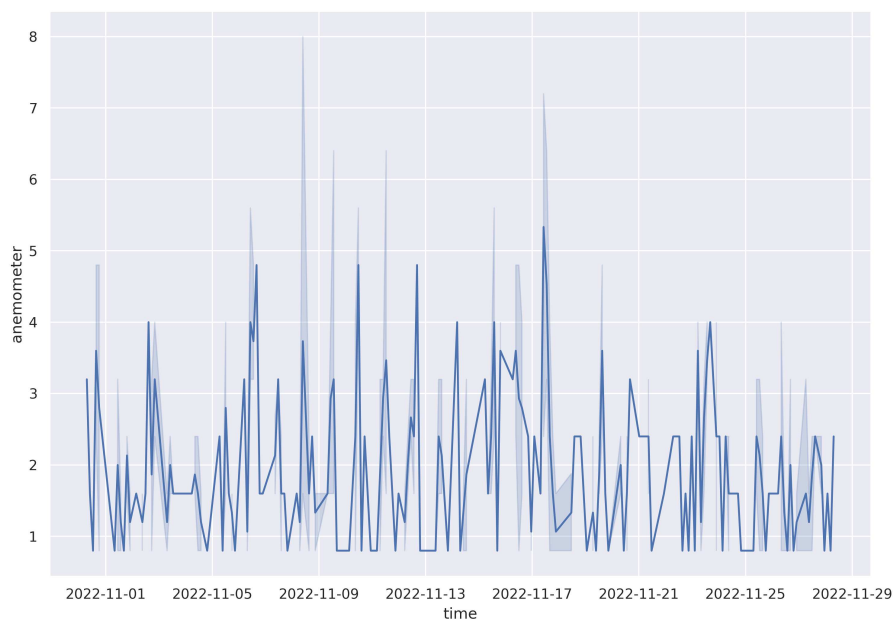
Figure 16. Hourly wind speed data for November, 2022.

4.4. Wind direction

The wind vane measures the direction. Table 7 shows different values that the equivalent resistance of the wind vane may take together withe the direction corresponding to each value.

Table 7. Wind direction value identification.

Direction(Degrees)	Resistance(k)	Voltage (V)	Identifier
0	33	2.53	SENS_AGR_VANE_N
22.5	6.57	1.31	SENS_AGR_VANE_NNE
45	8.2	1.49	SENS_AGR_VANE_NE
67.5	0.891	0.27	SENS_AGR_VANE_ENE
90	1	0.3	SENS_AGR_VANE_E
112.5	0.688	0.21	SENS_AGR_VANE_ESE
135	2.2	0.59	SENS_AGR_VANE_SE
157.5	1.41	0.41	SENS_AGR_VANE_SSE
180	3.9	0.92	SENS_AGR_VANE_S
202.5	3.14	0.79	SENS_AGR_VANE_SSW
225	16	2.03	SENS_AGR_VANE_SW
247.5	14.12	1.93	SENS_AGR_VANE_WSW
270	120	3.05	SENS_AGR_VANE_W
292.5	42.12	2.67	SENS_AGR_VANE_WNW
315	64.9	2.86	SENS_AGR_VANE_NW
337.5	21.88	2.26	SENS_AGR_VANE_NNW

5. Discussion

The use of data and machine learning in flood early warning systems has been widely researched in recent years due to the benefits it provides. The importance of data in these systems has been emphasized as the accuracy and completeness of the data determines the effectiveness of the warnings [21].

The solution presented in the paper focuses on improving the quality of the data used for flood early warning systems, which is crucial for the success of such systems [22]. The data collected should be of high precision, accuracy, and should be easily accessible to all stakeholders [23].

Furthermore, the paper highlights the importance of river basins as vital resources that support human and animal life [24]. The information gathered on river water levels can help in water management processes, which includes the utilization of flood waters for irrigation and other economic activities [25]. This information can be used to make better decisions regarding water resource management and allocation.

Recent studies have also shown that the integration of remote sensing data and machine learning algorithms can greatly enhance the accuracy and reliability of flood early warning systems [26,27]. In addition, the use of real-time monitoring systems and mobile technologies can improve the efficiency and speed of disaster response [28].

It is also important to note that this solution aligns with the United Nations Sustainable Development Goal 6, which aims to ensure availability and sustainable management of water and sanitation for all [29].

5.1. Experimental results and noise sources

Onsite setup of the devices had several things that need to be given some attention. Most important of them are sources of noise in the data. We were able to identify several sources of noise that must be put into consideration during further analysis of the data set. Since we employed a 4G connection as a communication standard, untimely recharging of internet-data package is one source of gaps in the data. However this is easily identifiable in the data set since it mostly occurs in the last or first days of the month. Another source of noise in the data is probably malfunctioning of the devices that brings values which are out of range. Such out of range readings can be seen from ultrasonic distances which are beyond the possible distance that can be recorded. For example, at Kikuletwa bridge, the distance sensor is hanging at 400 cm from the water surface, with and initial river depth of 130.5cm. It is expected that any distance recorded higher than 530.5 cm is out of range and it is one of those to be treated as outlier.

5.2. Flood prediction context

The data generated from this experiment is hourly multivariate time series data, which contains multiple variables observed over a period of time. Machine learning algorithms such as AutoRegressive (AR), Autoregressive Integrated Moving Average (ARIMA), Deep learning algorithms like Recurrent Neural Networks typically LSTMs can all be applied on the generated data set for different purposes. In the next phase of the study, LSTM auto encoders for anomaly detection will be applied on the river level data for detection of possible flood events. A key attribute of recurrent neural networks is their ability to persist information, or cell state, for use later in the network [30]. This makes them particularly well suited for analysis of temporal data that evolves over time.

5.3. Beyond flood prediction

There are several other areas that this data set might be of great use. Information about river water levels can be used by the water management decision makers in water management. For example, water harvesting for irrigation purposes can be simplified with the knowledge on how much water is flowing. Such knowledge can also be useful in drought detection, and thus inform on the preparedness in case the patterns will signal high probability of drought occurrence. For example for the year 2022, the numbers recorded in both rainfall amount and river levels reflect on the fact that there has been below average rainfall with widespread drought all over the country.

6. Conclusion

In order to create effective and precise river flood forecasting methods, we require data as the foundation of any AI/ML models. Many rivers in developing nations are inadequately researched, so the proposed solution aims to serve as a reliable data source for implementing disaster prevention strategies, such as identifying priority regions and taking corrective measures to prevent flooding damage. The data collected will not only expand the available information, leading to better early warning systems, but also, it can aid water management officials in making decisions such as water usage for irrigation and other purposes.

Abbreviations

The following abbreviations are used in this manuscript:

JDK	Java Development Kit
USB	Universal Serial Bus
CSV	Comma Separated Values
OEM	Original Equipment Manufacturer
AI	Artificial Intelligence
ML	Machine Learning
PWB	Pangani Water Board
HTTP	Hyper Text Transfer Protocol
TCP/IP	Transfer Control Protocol/Internet Protocol
FTP	File Transfer Protocol
API	Application Programming Interface
IDE	Integrated Development Environment

References

1. Chang, L.C.; Chang, F.J.; Yang, S.N.; Kao, I.F.; Ku, Y.Y.; Kuo, C.L.; Amin, I.M.Z.b.M. Building an intelligent hydroinformatics integration platform for regional flood inundation warning systems, 2018.
2. Huntingford, C.; Jeffers, E.S.; Bonsall, M.B.; Christensen, H.M.; Lees, T.; Yang, H. Machine learning and artificial intelligence to aid climate change research and preparedness. *Environmental Research Letters* **2019**, *14*, 124007.
3. Muñoz, P.; Orellana-Alvear, J.; Bendix, J.; Feyen, J.; Célleri, R. Flood Early Warning Systems Using Machine Learning Techniques: The Case of the Tomebamba Catchment at the Southern Andes of Ecuador. *Hydrology* **2021**, *8*, 183.

4. Sydorenko, I. What is a Dataset in Machine Learning: Sources, Features, Analysis. https://labelyourdata.com/articles/what-is-dataset-in-machine-learning.

5. Banai, F.; Sotoodeh, N. A review of the applications and challenges of wireless sensor networks for water resources management. *Journal of Ambient Intelligence and Humanized Computing* **2020**, *11*, 8857–8872.

6. Pandeya, B.; Uprety, M.; Paul, J.; Sharma, R.; Dugar, S.; Buytaert, W. Mitigating flood risk using low-cost sensors and citizen science: A proof-of-concept study from western Nepal. *J Flood Risk Management* **2021**, *14*, e12675. https://doi.org/10.1111/jfr3.12675.

7. Sarpong Adu-Manu, K.; Katsriku, F.A.; Abdulai, J.D.; Engmann, F. Smart River Monitoring Using Wireless Sensor Networks. *Wireless Communications and Mobile Computing* **2020**, *2020*, 19. https://doi.org/10.1155/2020/8897126.

8. Salazar Loor, J.; Fdez-Arroyabe, P. Aerial and Satellite Imagery and Big Data: Blending Old Technologies with New Trends. In *Big Data for Remote Sensing: Visualization, Analysis and Interpretation*; Dey, N.; Bhatt, C.; Ashour, A., Eds.; Springer: Cham, 2019. https://doi.org/10.1007/978-3-319-89923-7_2.

9. Zhang, L.; Leach, M.; Chen, J.; Hu, Y. Sensor cost-effectiveness analysis for data-driven fault detection and diagnostics in commercial buildings. *Energy* **2023**, *263*, 125577. https://doi.org/https://doi.org/10.1016/j.energy.2022.125577.

10. Zulkifli, C.Z.; Garfan, S.; Talal, M.; Alamoodi, A.H.; Alamleh, A.; Ahmaro, I.Y.Y.; Sulaiman, S.; Ibrahim, A.B.; Zaidan, B.B.; Ismail, A.R.; et al. IoT-Based Water Monitoring Systems: A Systematic Review. *Water* **2022**, *14*. https://doi.org/10.3390/w14223621.

11. Adu-Manu, K.S.; Katsriku, F.A.; Abdulai, J.D.; Engmann, F. Smart river monitoring using wireless sensor networks. *Wireless Communications and Mobile Computing* **2020**, *2020*.

12. Ahmedi, F.; Ahmedi, L. Dataset on water quality monitoring from a wireless sensor network in a river in Kosovo. *Data in Brief* **2022**, *44*, 108486. https://doi.org/https://doi.org/10.1016/j.dib.2022.108486.

13. Cheng, Z.; Lee, K.; Kim, D.; Muste, M.; Vidmar, P.; Hulme, J. Experimental evidence on the performance of rating curves for continuous discharge estimation in complex flow situations. *Journal of Hydrology* **2019**, *568*, 959–971. https://doi.org/https://doi.org/10.1016/j.jhydrol.2018.11.021.

14. Demetillo, A.T.; Japitana, M.V.; Taboada, E.B. A system for monitoring water quality in a large aquatic area using wireless sensor network technology. *Sustainability and Environmental Research* **2019**, *29*. https://doi.org/10.1186/s42834-019-0009-4.

15. Chopade, S.; Gupta, H.P.; Mishra, R.; Oswal, A.; Kumari, P.; Dutta, T. A Sensors-Based River Water Quality Assessment System Using Deep Neural Network. *IEEE Internet of Things Journal* **2022**, *9*, 14375–14384. https://doi.org/10.1109/JIOT.2021.3078892.

16. Zhang, X.; Chen, L. A review of the applications and challenges of wireless sensor networks for water resources management. *Journal of Ambient Intelligence and Humanized Computing* **2020**, *11*, 8857–8872.

17. Libelium. Early flood detection and warning system in Argentina developed with Libelium sensors technology. https://www.libelium.com/libeliumworld/success-stories/early-flood-detection-and-warning-system-in-argentina-developed-with-libelium-sensors-technology/.

18. Costache, R.; Bao Pham, Q.; Corodescu-Roșca, E.; Cîmpianu, C.; Hong, H.; Thi Thuy Linh, N.; Ming Fai, C.; Najah Ahmed, A.; Vojtek, M.; Muhammed Pandhiani, S.; et al. Using GIS, Remote Sensing, and Machine Learning to Highlight the Correlation between the Land-Use/Land-Cover Changes and Flash-Flood Potential. *Remote Sensing* **2020**, *12*. https://doi.org/10.3390/rs12091422.

19. Alfieri, L.; Salamon, P.; Pappenberger, F.; Wetterhall, F.; Thielen, J. Operational early warning systems for water-related hazards in Europe. *Environmental Science & Policy* **2012**, *21*, 35–49.

20. citizen news, T. Climate change has affected seasonal rains in Tanzania, says TMA. https://www.thecitizen.co.tz/tanzania/news/national/climate-change-has-affected-seasonal-rains-in-tanzania-says-tma-3783384.

21. Pang, L.; Li, C.; Li, X.; Li, X.; Wang, Y. A review of flood early warning systems: Methods, applications, and challenges. *Sensors* **2020**, *20*, 3318.

22. Zhang, Y.; Xiong, Y.; Liu, J.; Zhang, X. A comprehensive review of flood early warning systems. *Water* **2018**, *10*, 1186.

23. Sun, Y.; Liu, Y.; Li, Y. Development of flood early warning system based on improved gray model and particle swarm optimization. *Water* **2020**, *12*, 1053.

24. Wang, L.; Li, J.; He, Y. Climate change impacts on river basin water management and adaptation strategies: A review. *Environmental Science and Pollution Research* **2018**, *25*, 1815–1825.
25. Li, J.; Wang, L.; He, Y. A review of water management in river basins under the impact of climate change. *Environmental Science and Pollution Research* **2019**, *26*, 20517–20530.
26. Wang, Y.; Guo, Q.; Yin, Y.; Huang, L. A machine learning-based flood early warning system with remote sensing data. *Journal of Hydrology* **2019**, *570*, 191–201.
27. Guo, Q.; Yin, Y.; Wang, Y.; Huang, L. An integrated flood early warning system based on remote sensing and machine learning. *Sensors* **2020**, *20*, 2562.
28. Chen, Y.; Li, Y.; Li, X.; Zhang, Q. Big data analytics for flood monitoring and early warning system. *Information Sciences* **2019**, *465*, 1–12.
29. Nations, U. Sustainable Development Goals, 2021.
30. Yu, Y.; Si, X.; Hu, C.; Zhang, J. A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures. *Neural Computation* **2019**, *31*, 1235–1270, [https://direct.mit.edu/neco/article-pdf/31/7/1235/1053200/neco_a_01199.pdf]. https://doi.org/10.1162/neco_a_01199.