

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Behavioural Analysis of Water Consumption using IoT-based Smart Retrofit Meter

A. KUMAR LALL, A. TERALA, A. GOYAL, S. CHAUDHARI, K. S. RAJAN¹, S.S.CHOUHAN²

¹International Institute of Information Technology Hyderabad (IIIT-H), India (e-mail: {ayush.lall, aakash.t, archit.goyal}@research.iiit.ac.in, {sachin.c, rajan}@iiit.ac.in)

²Luleå University of Technology Luleå, Sweden (e-mail: shailesh.chouhan@ltu.se)

Corresponding author: S.S.Chouhan (e-mail: shailesh.chouhan@ltu.se).

This work was supported by grant from TIH Foundation for IoT & IoE at IIT Bombay under the TIH-IoT CHANAKYA Fellowship Program 2022-23 (Sanction Letter No. TIH-IoT/2023-03/HRD/CHANAKYA/SL/CFP-018)

ABSTRACT This paper presents the analysis of water supply behavior within an educational campus, serving as a use-case to demonstrate the broader applicability of an innovative IoT-based framework integrated with deep learning techniques. By retrofitting analog water meters with IoT devices, the study captures images of meter dials, which are then locally processed using a deep learning-based digit detection algorithm. This process converts the images into digits and transmits the data to the cloud for real-time analysis, thereby enhancing the accuracy and reliability of water usage data. Focusing on two key regions within the campus—student hostels and faculty/staff quarters—the analysis thoroughly examines the impact of water supply patterns on both a monthly and weekly basis. It reveals how the distinct characteristics of each month, such as holidays, exams, and class schedules, significantly influence water consumption in these areas. The study particularly highlights the variations in water usage in student hostels, driven by the academic calendar and student lifestyle, in contrast to the more stable water demand observed in faculty/staff quarters. The integration of the data refinement algorithm uncovers the underlying consumption patterns within these campus residence. The findings from this detailed investigation are instrumental in understanding the water distribution patterns, particularly within Integrated Water Systems (IWS), and set a precedent for the potential scalability and adaptability of the framework. This study not only sheds light on the specific water management needs of an educational campus but also suggests that the successful application of this system in such a dynamic and varied setting indicates its potential for broader application, thereby contributing to more informed decision-making and promoting sustainable water management practices in various contexts.

INDEX TERMS Consumption patterns, DL techniques, Informed decision-making, IoT-Based Framework, Retrofit solutions, Sustainable water management, Water supply behavior

I. INTRODUCTION

Water, an essential resource for sustaining life, is vital in various aspects of human existence. Reliable water distribution systems are crucial, ensuring equitable access to clean drinking water, promoting public health, and efficient resource management. Several initiatives have been taken to make the process smarter; for example, deploying smart IoT-based retrofit models on analog water meters to upgrade them smart [1]. By analysing the data collected, smart water meters can aid in understanding water consumption patterns and in leakage detection for efficient water management.

Limited literature exists in the domain of enhancing analog

water meters through digital recognition techniques exemplified by [2] and [3]. The work in [2] devised a convolutional neural network (CNN) rooted in the visual geometry group (VGG) architecture to recognise gas meter readings, achieving an 85.71% accuracy rate. In another study, [3] used a DL model with Raspberry Pi to recognize meter readings. The model, based on You Only Look Once (YOLO)-v3, showed 96% accuracy when tested on 100 water meter images. In our work [1], a retrofit solution was executed to digitise analog water meters with the help of the random forest (RF) model. Following post-processing refinements, an accuracy

of approximately 97% was attained. Nonetheless, a performance drop was observed in the RF model due to different illumination conditions at different locations. However, deep learning models for pattern detection use general features for training and can handle illumination fluctuations more effectively.

These IoT-based smart solutions provide us with a better understanding of water consumption patterns, helping us improve water utilisation. There are two major types of water distribution systems: CWS and intermittent water supply (IWS) [4]. CWS system is mostly adopted by developed countries, and it involves distributing water from a single point of source to different buildings. The pipelines used in this distribution system are controlled to deliver an uninterrupted and consistent water supply according to their residential, commercial, or industrial status. This system aims to overcome the limitations of supplying water at specific intervals or during certain hours of the day. In a continuous water supply system, water is available 24x7 without any interruptions, ensuring a steady water flow to consumers using the reliable system. However, implementing these systems is financially exhaustive and complex.

The IWS system supplies water from one source, stores it in a unit, and then distributes it to different places. The water is provided to consumers according to their timetable rather than being available continuously. IWS is commonly deployed in areas with limited water resources and infrastructure constraints that is commonly adopted by many developing countries [5] [6] [7]. Moreover, IWS systems are costly for consumers, requiring additional infrastructure like an overhead tank (OHT) and underground reservoir (sump) for water storage.

IoT can greatly enhance the reliability of the water distribution system, especially in the IWS-based system. IoT-based sensors are deployed across water distribution networks to monitor water flow, pressure, quality, and leak detection. The collected data from the sensors help to identify leaks or inefficient usage, enabling prompt action to reduce water wastage [1], [8]. IoT devices continuously monitor water quality parameters such as pH levels, dissolved oxygen, and any contamination. This ensures that authorities address water quality and proactively meet the required standards and issues. IoT-enabled smart meters monitor water consumption patterns in homes, businesses, and industries. IoT combined with various technologies like ML [1], DL [9], computer vision [10], etc., is used to create smart retrofit models for the existing solution which would be economical. Authorities can access this data to understand usage patterns, identify high-consumption areas, and implement water-saving measures.

Most CWS systems are integrated with advanced metering infrastructure (AMI) to collect information on consumption. One of the components of AMI is smart water meters that collect high-frequency data. The data collected by AMI helps us develop optimal water distribution system modelling. This enhances the planning and maintenance of the water distri-

bution system [11]. The IWS system does not support AMI because of the high operational and maintenance challenges. The additional infrastructure in IWS is also inadequate in many cases. The design of the additional infrastructure is based on the estimated water demand for the number of consumers. The literature work on high-frequency data collection, using smart water meters for CWS is available in [11], [12] and [13]. In [12], authors optimise the different stages of the urban water distribution cycle. Examining the supply-to-distribution process based on customer engagement. Efforts were made to disaggregate the various end-user events from the event stream data. The work in [11] explains the over-engineered design of Australia's existing water distribution system; this was done by calculating the water demand rates. The literature available on CWS-based systems analyses high-resolution data of water consumption patterns. However, these strategies don't work on the IWS system, as mentioned above. Thus, efforts to get high-frequency data for IWS systems remain largely unexplored due to deployment costs. Moreover, water usage demands are based on consumption areas in locations like colleges, schools, and institutions. Thus understanding the water distribution patterns becomes crucial to ensure sufficient water supply when needed and prevent unnecessary wastage. Hence, there is a significant need to utilise high-frequency data in IWS systems to enhance water management. The work proposed here is an extension of [9]. This work focuses on studying and analysing the high-resolution water flow meter data. A use case from the residential buildings of the IIIT-Hyderabad campus to understand the deployment suitable for the IWS system has been investigated. The proposed water smart water meters [9] were installed over the water distribution network, and data was collected. The study was conducted over the span of four months from the installed water meter. Of all smart retrofit meters deployed, only four were taken into this study because these four were strategically placed at the location where the consumption was expected to be higher by the user. These places were student accommodation (hostels) and faculty/staff residential block, providing insightful information about the water usage pattern. The specific contributions of this paper are¹:

- A IoT-based smart retrofit meter is implemented. The images of the analog water meter dial were taken. The meter reading was digitized using a DL-based digit detection algorithm, and the value was sent to the cloud.
- The detected values were refined using the Hamming distance-based technique.
- Four smart retrofit IoT nodes were strategically deployed within the IIIT-H campus over a span of four months and collected around 550,000 data points. Two off-the-shelf commercial digital meters (Shenitech Meters Models: 280W-R DN40 [14]) were also used for cross-referencing and analysis.

¹Initial results on DL-based detection were presented in [9], showing the DL-based method better than the ML-based method.

- Behavioural analysis is conducted in terms of weekly and monthly supply patterns on the collected data. Distinguishing water supply patterns within high-accommodation residential areas such as student hostels and faculty/staff quarters.
- For the IWS system, the limitations were demonstrated regarding time resolution and delay, which are usually not in the CWS system.
- The variation in the distribution of water supply patterns is also investigated during holidays, showing variations between the hostel and faculty/staff blocks.

This work presents an understanding of the water usage patterns using IoT-based smart retrofit meters to collect high-temporal data in an IWS distribution system. This enables us to gain valuable insights into the long-term behaviour of water consumption based on the changes in water supply to the OHT.

The paper is organized as follows: Section II outlines the experimental setup detailing hardware and software used as well as deployment details. Section III provides Data Processing, encompassing Data Collection, Error Analysis, and Data Correction. Section IV presents Results and Analysis, including Comparison between Commercial Digital and Proposed Smart Retrofit Meter, Time Series Data, Monthly Analysis, and Weekly Analysis. Finally, the conclusion is presented in Section V.

II. EXPERIMENTAL SETUP

The research utilizes IoT-based retrofit model, as cited in [1], to analyze water usage patterns across the campus. Four smart meters were strategically installed, integrating with the IWS system of the campus, enabling detailed monitoring of water storage patterns within the tanks. The study focuses on deciphering water consumption trends by assessing the data collected from these smart water meters within the IWS framework.

The investigation primarily targets the analysis of monthly and weekly water storage patterns, scrutinizing the variations in water usage during these periods. This graphical analytical approach is crucial for identifying the fluctuating water demands throughout different months and weekdays, aiding in better resource management and planning.

Campus architecture, including hostels and faculty/staff quarters, significantly influences water consumption patterns. The hostels, housing a large number of residents and incorporating student messes, are identified as the primary water consumers. Conversely, the faculty/staff quarters, accommodating relatively fewer inhabitants, primarily college faculty and staff families, contribute less to the overall water usage. This distinction in water consumption across different campus sectors provides a distinct understanding of water utilization, essential for optimizing the water distribution system.

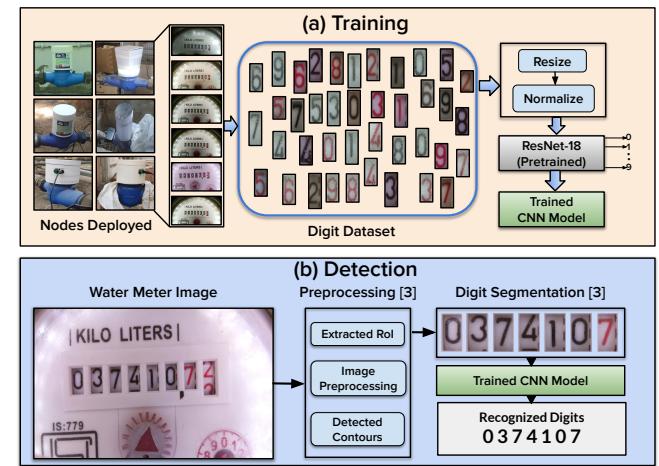


FIGURE 1. Algorithmic pipeline for the DL methodology proposed. The dataset collected from different nodes is used to train the ResNet-18 model (last layer having 10 nodes) using transfer learning mechanism. At the detection phase, preprocessing used in [1] was used to segment the digits from water meter image. Finally the digits are recognized using the trained CNN model.

A. HARDWARE SETUP

Fig. 2 shows the structure of the developed 3D printed smart retrofit model in [1] for collecting the data from the analogue water meters. Equipped with a PiCam, the retrofit model captures images of the analog water meter. It then extracts the meter reading, the region of interest (RoI), using a perspective transform from the OpenCV library [15]. The next step involves extracting individual digits from the RoI using contour forming. The segmented digits are processed by pre-trained ML [1] or DL models [9], embedded within the node. The detected digits are combined and transmitted to a cloud server for further analysis. A Raspberry Pi 3B+ microcontroller [16], connected to AC and battery power sources, controls the camera and executes the algorithms. Moreover, the retrofit model has an LED ring to ensure proper lighting, allowing image capture even during night-time. Capturing data at 1-minute intervals, the model provides densely sampled data points. This data abundance is invaluable for identifying significant water supply increases and understanding high-demand periods.

B. SOFTWARE SETUP

This section describes the digit detection process to obtain the meter's reading. The digit recognition relies on a state-of-the-art DL (Deep Learning)-based algorithm [9]. Although many CNN (Convolutional Neural Network) models are available for the object detection task, our use case required a CNN model with less computational complexity yet possessing sufficient accuracy to perform the digit recognition task. The main reason for choosing a computational complexity-aware model is to enable the performance of the inference directly on the hardware node. Following [17], a popular CNN model, ResNet-18 [18], was identified as a suitable choice considering computational complexity and accuracy.

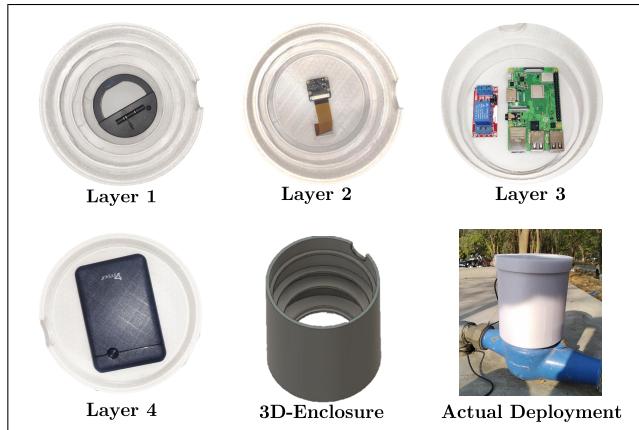


FIGURE 2. The retrofit model is being deployed at an analog water meter on the campus [1].

As evident from the name, ResNet-18 is an 18-layer deep network, and the model overall contains almost 11 million trainable parameters. Training the Resnet-18 model, where the weights are randomly initialized directly on the digit dataset, is an extremely challenging task. Hence, the transfer learning mechanism, a method where a model developed for one task is reused as the starting point for a model on a second task, is employed to address this challenge. [18] has trained the ResNet-18 model on a large dataset named ImageNet [19], which contains around 14 million images across 20,000 categories. We utilized this pre-trained model to further train on our digit dataset. The final fully connected layer of the pre-trained ResNet-18 model was replaced by ten neurons to match the number of outputs required for digit recognition.

Preprocessing: As the ResNet-18 model requires input of a fixed size, all digit images are initially resized to a resolution of 224×224 pixels to ensure compatibility with the pre-trained ResNet-18 model. Standard normalization (zero mean and unit standard deviation) is then applied to the input image. This step ensures that all inputs to the CNN model follow a similar data distribution, aiding in faster convergence during the network training process.

Training: The dataset was partitioned into training and validation sets, consisting of 90% and 10% of the data points, respectively, selected at random. This step is crucial for generalizing the model, as it is trained solely on the training dataset and evaluated on the validation dataset. The model underwent training for five epochs, achieving convergence. The optimizer chosen was Adam, a popular optimization algorithm, with a learning rate of 0.001, scheduled using a linear scheduler with gamma set to 0.1. Pytorch [20], a widely-used Python-based framework, was employed for all DL implementations.

The analyses throughout the paper are based on data collected using this DL-based digit detection algorithm. The proposed DL algorithm is built upon the ResNet-18 model [21] (utilizing Pytorch [20] libraries) with transfer learning. It was robustly trained on a dataset of approximately 160,000

images. A comprehensive overview of the training process is depicted in Figure 1. The captured images were processed by manually inputting the selected coordinates, width, and height to obtain a transformation matrix. This matrix facilitated the application of warp perspective to derive the Region of Interest (RoI).

After extracting the RoI, the subsequent step involved segmenting the digits (ranging from 0 to 9) from the RoI and organizing them into labeled folders, thus creating a dataset for supervised learning. Preference was given to independent and identically distributed data points to ensure an ideal dataset with minimal bias. Analysis revealed that only the right-most digit (as illustrated in Fig. 1 - the water meter image) exhibited significant changes over the time series data, while the other digits remained relatively unchanged due to the low water flow rate. This observation led to the presence of duplicate digits in the dataset. To address this issue, the images were converted into tensors by merging the respective RGB pixels and then resized to a 50-pixel format to minimize computational complexity during comparison. The comparison process utilized the mean square error (MSE) between the resized tensors of two images. A lower MSE value signified a higher resemblance between the images. Ideally, two images with an MSE of 0 would be identical; however, due to the resizing to 50 pixels, minor variations from the original images were expected. A maximum MSE threshold of 20 (determined through trial and error) was established to identify and eliminate duplicate images. Images exceeding this threshold were deemed duplicates and consequently removed from the dataset. The results obtained from the DL-based model were highly promising, enabling the expansion of the algorithm for data collection from eight different nodes for further analysis.

III. DEPLOYMENT SETUP

Fig. 3 illustrates the meter deployment setup across the campus, comprising two distinct regions for analysis: the student hostel block and the faculty/staff block. Each region has unique water storage and consumption patterns. The map offers a comprehensive overview of the buildings in each block along with their corresponding water supply arrangements. Both blocks are equipped with their own borewell and sump for water storage, from which water is distributed to an OHT for further use. The Government supplies water for drinking purposes. Moreover, the map delineates various types of water meters installed on the supply lines, enhancing understanding of the infrastructure.

In the student hostel block, only smart retrofit meters are utilized, signifying a uniform approach to metering. Conversely, the faculty/staff block employs a combination of digital meters (Shenitech meters) and smart retrofit meters, indicating a more diverse metering strategy. This block houses three meters in total: two digital and one smart retrofit meter. Notably, one digital and one analog meter are installed in series on the same pipeline, tasked with measuring the drinking water supply from the sump to the OHT. This strategic setup

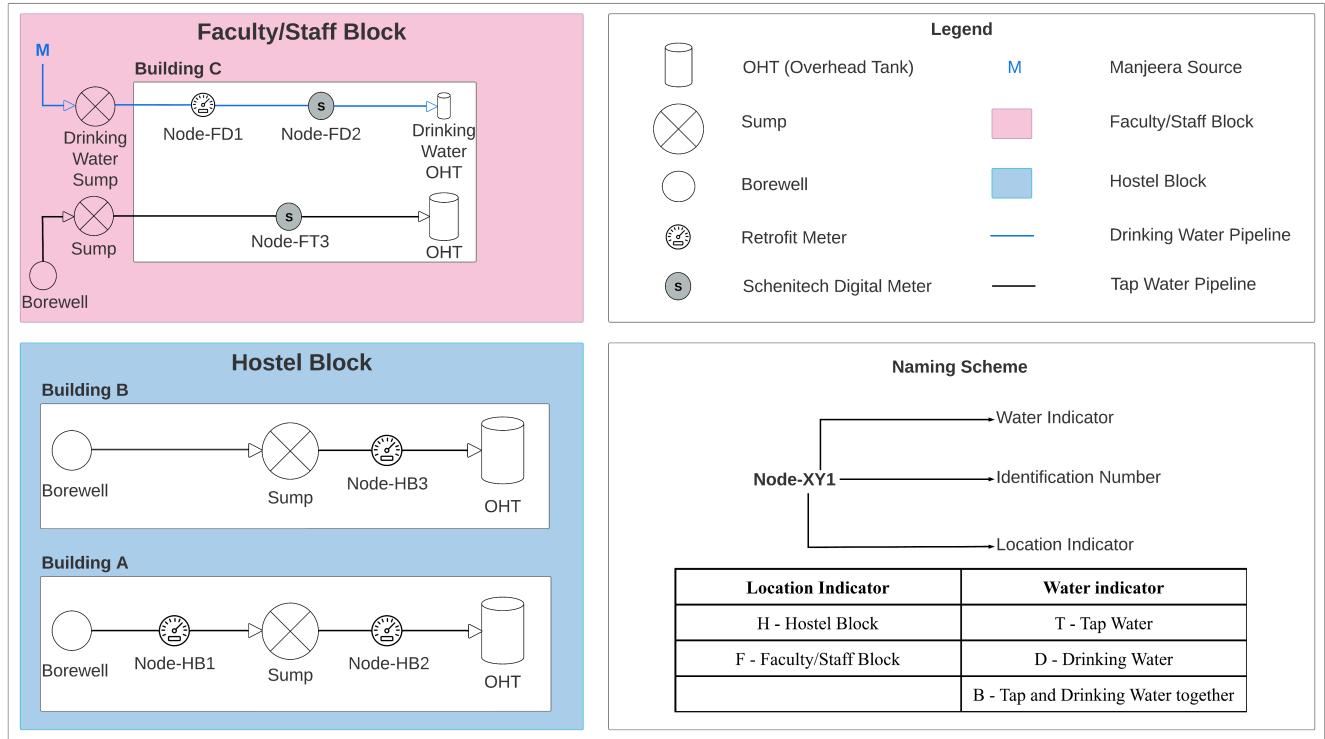


FIGURE 3. The region in blue color denotes Hostel Blocks, containing two student hostels, and the region shown in pink color is for residential block which shows water distribution for the building in the region.

serves a dual purpose: it allows for the verification of IoT-based smart retrofit meters against standard commercial digital meters, and it facilitates a detailed analysis, the findings of which are explored in subsequent sections. Additionally, an individual digital meter is dedicated to monitoring the tap water supply, ensuring comprehensive coverage of water usage metrics.

This careful meter deployment not only sheds light on the water consumption patterns across different campus regions but also plays a crucial role in evaluating various metering technologies, thereby contributing to the advancement of efficient water management practices.

IV. DATA PROCESSING

This section is divided into three subsections. The first subsection provides an overview of data collection, including the specific sensors or devices used, their placement, and the methodology for data collection. The second subsection

TABLE 1. Table showing missing data points of each node in respective regions.

Region	Node	August	September	October	November	Total
Hostel Region	Node-HB1	1	3	8	15	27
	Node-HB2	2	2	5	6	15
	Node-HB3	11	2	12	N/A	25
Faculty/Staff Region	Node-FD1	8	0	4	0	12

discusses challenges and errors related to the collected data, and the third subsection describes the corrective measures taken to address these errors.

A. DATA COLLECTION

Challenges were encountered with the data collected during the monsoon semester from August 2022 to November 2022, collected at every minute. These included missing data for certain days and delays in receiving the data. There were times when the deployed node had to be brought back to the lab for maintenance, leading to periods without data collection. On some days, heavy rain caused the circuit board powering the device to short-circuit, necessitating repairs and resulting in data loss for those periods. Additionally, there were delays in transmitting the data due to poor network connectivity and power outages. Table 1 provides an overview of the data loss occurrences, indicating the number of missing data points per node, which helps in assessing the extent and impact of the data collection challenges faced during this period.

The table indicates the number of data points missing in days for each node in the respective regions each month. It is important to note that the days mentioned in the table are not consecutive but discrete, and their total is provided. Most missing days occurred in October, mainly because many holidays fall during this month, leading to a reduced student presence and challenges in system maintenance. However, the low student occupancy resulted in decreased water con-

sumption. During holidays, with no classes in session, water consumption further declines. Therefore, collecting data at high frequency during this period has minimal impact on the analysis. The next sections further elaborate on the specific errors encountered during data collection and the measures taken to address them.



FIGURE 4. Issues Related to collected images.

B. ERRORS ANALYSIS

Fig. 4 illustrates the issues encountered while capturing images of the meters. The four images represent major issues related to environmental factors, such as dew deposition inside the meter glass, typically occurring during heavy rain, and smudges/scratches on the meter. There were instances when un-intentioanl human intervention with the deployed set-up like care-takers, while cleaning the area, accidentally impacted the device, causing issues with the camera's orientation or resolution. The climate of Hyderabad is semi-arid, featuring too little rain to feature the tropical savanna climate. Thus present of tiny insects is expected, these insects when entering the device led to errors in digit detection. These issues resulted in detection errors that produced abnormally high or low output values. The incorrectly detected value persisted as the output until the meter was cleaned. This repetitive error decreased detection accuracy. These errors typically manifest as sudden, highly significant changes over a very short period, which is implausible based on the pipe's capacity. This observation underscores the importance of considering logical inconsistencies in the data and serves as the basis for introducing a refinement algorithm in the next section.

C. DATA CORRECTION

To enhance detection accuracy, a refinement algorithm based on the concept of Hamming distance is proposed. The implementation of this algorithm involved software tools and steps specifically designed to address the unique challenges posed by the data collection process. The Hamming distance is a metric used to measure the difference between two equal-length strings, calculating the number of positions where the corresponding symbols (characters or bits) differ. In this case, the algorithm involves strings of digits. The algorithm defines a range beyond which the meter values cannot exceed in a given time-frame. Detected values are then checked to ensure they fall within this range. The dynamic nature of the range, based on the time difference between two detected values, adds an adaptive component to the algorithm, making it more

responsive to the varying conditions of data collection. The range of values can dynamically change based on the time difference between two detected values. Thus, the range is defined as $R = 10T_D$. Here, the number 10 represents the maximum range for the changing values of the meter's digit in one minute, determined based on the maximum rate of water flow supported by the motor in the specified time interval. T_D denotes the time difference between image captures. If there is a delay in receiving the image, the range expands based on the time difference between the initial value's timestamp and the currently received value. This mechanism ensures that the algorithm can accommodate delays in data transmission, a common challenge noted during the data collection phase.

The current detected value should fall between the previously detected value and the sum of the previously detected value and R (the range calculated in the equation). The current detected value is compared with each value within this range by examining individual digits, akin to how the Hamming distance operates. The number of differing digits between the current detected value and each value within the range is counted. The value with the fewest differing digits is deemed the correct value. Ideally, this difference count is 0, indicating the current detected value falls within the acceptable range. Fig. 5 provides a visual representation of this process, illustrating how the algorithm corrects for errors and refines the data.

T1 (Prev)	2	3	3	2	5	5
T2 (Detected)	2	3	3	1	5	9
Corrected	2	3	3	2	5	9

FIGURE 5. Illustration of refinement algorithm

In the illustration, the previous detection occurred at time T_1 . At time T_2 , the current detection took place, where one digit was incorrectly detected as 1 instead of 2, while the other digits were correctly detected. The Hamming distance between the values at these two time points is 2, as indicated by the two differing digits highlighted in orange. This illustration effectively demonstrates the algorithm's ability to identify and correct errors, thereby enhancing the accuracy of the data. The first highlighted digit is the incorrectly detected value of 2 as 1, and the second highlighted digit represents the difference between the detected digit 9 and the corresponding digit in the first value of the range, i.e., 5. Therefore, in this illustration, the Hamming distance is 2. As mentioned earlier, the range is set to $(P_D, P_D + 10T_D)$ where P_D is the previously detected value. Thus, the algorithm iterates from 233255 to 233265. When it reaches 233259, it records the least Hamming distance of 1, where only digits 1 and 2 mismatch, indicating this as the least Hamming digit value. Consequently, the final meter reading is determined to be 23325.9 kL. It is important to select the value with the least Hamming distance to avoid discrepancies with other digits having a similar Hamming distance. This method provides

a systematic approach to data correction, ensuring that the refined data is both accurate and reliable.

The rest of the analysis employed this algorithm. Table 2 not only presents the RMSE and MAE values of the detected DL-based algorithm before and after refinement but also offers a comparative analysis of the algorithm's performance, highlighting the significant improvements achieved through the refinement process. Approximately 4K data points were manually annotated to generate this table. The majority of detection errors were attributed to the reasons outlined in the previous subsection. The statistical analysis provided in this table underscores the effectiveness of the refinement algorithm in addressing the challenges posed by environmental factors and other sources of error. Notably, the original DL-based model boasted an accuracy of approximately 99%, as indicated in [9]. However, in the current case, the node was well-maintained for 20 days without any rain. The scenario involved capturing images for the entire semester, spanning approximately four months. These images, influenced by various climatic conditions such as rain, included defects mentioned earlier. Dust particles and smudges, particularly after heavy rain, and the intrusion of insects into the device significantly impaired image quality. Detection errors persisted until the dust was removed but were repetitive in nature. For instance, when dust particles settled on the Most Significant Digit (MSD) that rotated slowly, the same detection error would recur for an extended period. The refinement algorithm effectively addressed these errors, as evidenced by the improved data quality presented in the analysis.

Fig. 6 displays a graph for the node Node-HB2 (refer to Fig. 3), depicting both detected and refined values over the four months of data collection. This graph serves as a visual representation of the refinement process, illustrating the algorithm's impact on data quality and providing a clear comparison between detected and refined values. A total of 3,362 data points required correction by the refinement algorithm. The graph underscores the algorithm's role in ensuring data integrity and enhancing the overall quality of the study's findings.

TABLE 2. Table showing the improvement done by the refinement algorithm.

	Detected	Refined
Root Mean Square Error	8.60	0.24
Mean Absolute Error	0.67	0.02
Accuracy	97.01	97.37

Fig. 6 shows a graph for the node Node-HB2 (refer Fig. 3), depicting detected and refined values over the four months of data. Where 3,362 data points had to be corrected by the refinement algorithm.

V. RESULT AND ANALYSIS

This section is structured into four subsections for a comprehensive analysis. Firstly, it compares the Shenitech® digital water meter against smart analog water meters. The second

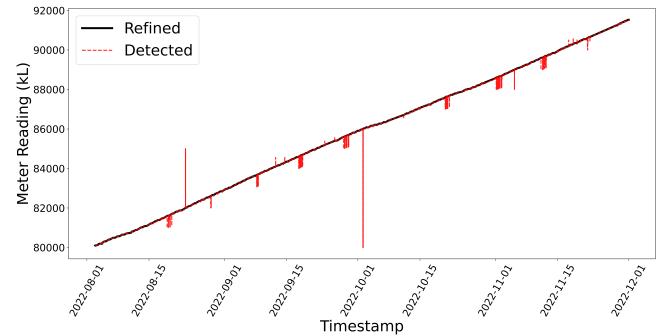


FIGURE 6. Graph of detected values and refined values for node Node-HB2, plotted over four months of data. Where on the X-axis timestamp is given and on the Y-axis meter readings in kL (Kilo Litres) is mentioned. (Best viewed in coloured)

subsection delves into time series data analysis across all smart water meter nodes. Notably, the analysis focuses on the IWS system, acknowledging that obtaining instantaneous consumption values isn't feasible. Instead, it emphasizes on accumulating supply pattern data to reveal consumption patterns over time. The third subsection conducts a detailed monthly analysis of water usage across all meters, whereas the fourth subsection explores a weekly breakdown of water usage patterns.

A. COMPARISON BETWEEN THE DIGITAL AND SMART RETROFIT METERS

Table 3, presents a comparison between two types of meters installed on the same pipeline: the DL-based smart retrofit meter (Node-FD1) and the off-the-shelf digital meter (Shenitech® Node-FD2). Remarkably, the values recorded by the smart retrofit meter on the analog device closely align with those obtained from the digital meter. The primary distinction lies in the precision of the values. However, this difference becomes negligible when we round off the precision of the digital meter's values from three digits to one decimal place, effectively aligning them with the readings of the smart analog meters.

B. TIME SERIES DATA

Fig. 7 shows the time series data collected from all the nodes for approximately a month (from 17 September 2022 to 15 October 2022). All the meter readings are plotted here by adjusting them to zero on 17 September 2022 for ease of visualization. The thin solid line indicates the best fit linear trend line for the plotted raw data (thick line of the same

TABLE 3. Table for comparison between net water volume flowing in a day for both analog and digital meter.

Date (DD/MM)	22/10	23/10	24/10	25/10	26/10	27/10	28/10	29/10	30/10	31/10	01/11
Retrofit Meter	0	11.4	0	0	0	0	2.1	0	0	2.3	0
Digital Meter	0	11.605	0.003	0.004	0.003	0.003	2.164	0.005	0.004	2.397	0

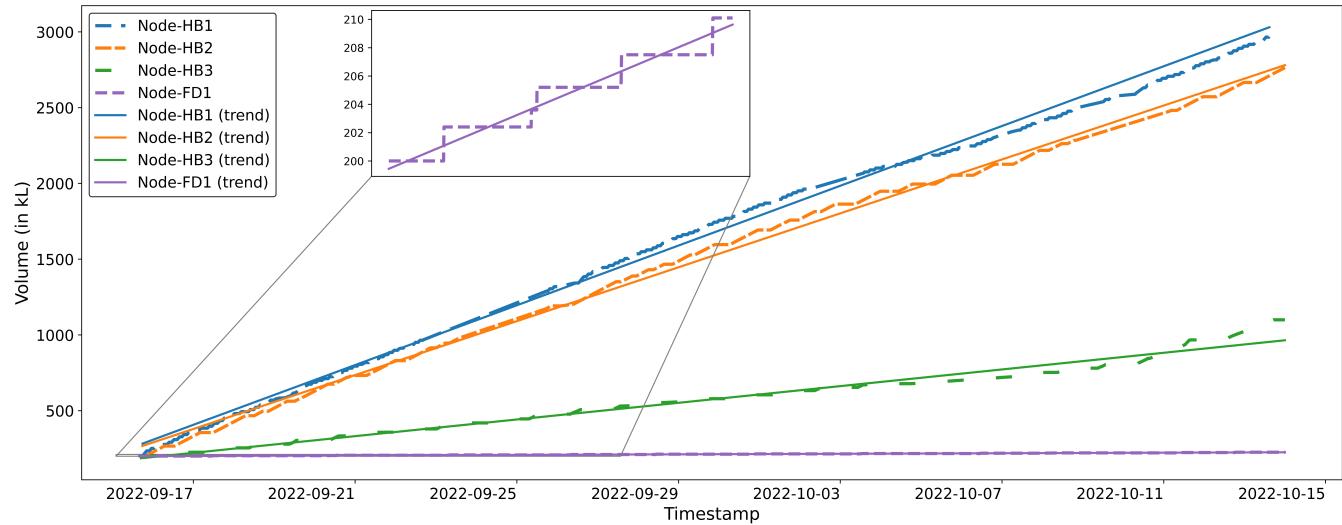


FIGURE 7. Time series plot for net water volume in kL. The dotted thick line shows the raw data for each node from September 17, 2022 to October 15, 2022. The solid thin lines of the same color shows the best fit linear trend line corresponding to each node. (Best view in coloured)

TABLE 4. Table for the Average Daily consumption for each node found using the best fitting linear trend line.

Node	Average Daily Consumption (kL/Day)
Node-HB1	98.60
Node-HB2	88.95
Node-HB3	27.52
Node-FD1	0.90

colour) for a node. It can be seen that the amount of water used is highest for Node-HB1 and Node-HB2 followed by Node-HB3, and Node-FD1 has the least water usage. Similar observations can also be noted from table 4.

Note that Nodes-HB1 and HB2 are for the same building A, where Node-HB1 is on the pipeline between the borewell to sump and Node-HB2 is from sump to OHT. This proximity accounts for their similar readings, except for a small difference which is attributable to the water storage in sumps. It can be observed that the Nodes-HB1/HB2 and HB3, present in the hostel blocks, have the highest volume of water flowing throughout the month along with the highest slopes (kL per day). This is expected as buildings A and B are hostel blocks with occupancy of around 800 and 425 students, respectively. Nodes HB1 and HB2 are much higher than Node-HB3 also because building A has two messes while building B does not have any. On the other hand, Node-FD1, responsible for supplying drinking water, displays the lowest slope, indicating the least volume of water flowing through it. It is intriguing to note the distinct steps in the curve, implying that the drinking water is supplied at intervals of a day or two, from the sump to the OHT. This suggests that the water in the storage tank remains unused for an extended period before being consumed.

C. ESTIMATING CONSUMPTION PATTERN IN IWS SYSTEM

The step nature of the water supply curves in Fig. 7 demonstrates the IWS nature, where the water is supplied to the OHTs by automated motors after certain intervals. Table 5 displays the average duration after which the motor pumps water to the OHT. Consequently, in these scenarios, given the absence of meters at the outputs of these tanks, the consumption pattern can only be inferred from the supply pattern after a delay, and the estimation is relatively coarse. For instance, at Node-HB2, water supply intervals occur every 18.8 hours on average. This indicates that the estimated consumption and supply closely align every 19 hours at water-filling instances. Moreover, it's noticeable that the estimation delay and resolution vary with the change in the average supply interval. Node-FT3 experiences water supply intervals every 11 hours.

For the subsequent analysis of monthly and weekly data, it's reasonable to infer that the average consumption and supply values are identical since the water supply durations are significantly shorter than the averaging intervals, as evidenced by the table.

D. MONTHLY ANALYSIS

Fig. 8 shows the monthly supplied water through all the nodes from the month of August to November 2022. In order to find these values, the meter reading at the end of the month was subtracted from the start of the month. From

TABLE 5. Table showing average time duration after which the water is supplied to the OHT. The nodes which are present at the input of the tank are considered in the table.

Node Name	Node-HB2	Node-HB3	Node-FD1	Node-FT3
Time Duration (hrs)	18.8	19.4	61	11

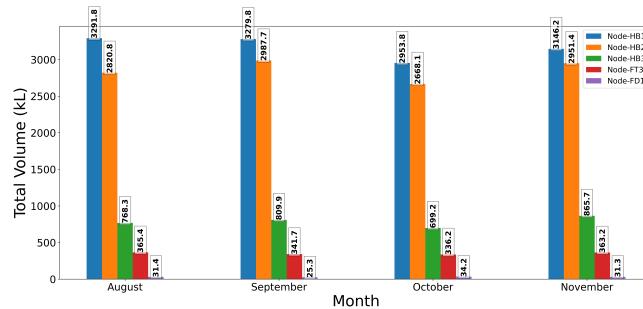


FIGURE 8. Monthly net water supply plot.

the graph, a similar observation as time series data can be observed, where nodes present at the hostels show higher monthly consumption compared to other nodes. From the figure, observations made are:

- Hostel nodes-HB2 and HB3 show higher water demand for the month of September and November. This trend can be attributed to the high occupancy of students during the semester exams conducted in those months.
- Conversely, Hostel nodes-HB2 and HB3 demonstrate reduced water demand in October, correlating with the decreased occupancy in hostels due to the numerous holidays in that month when many students return home.
- For Node-FT3 in the faculty and staff block, a consistent usage of tap water is observed throughout the entire data set. This consistency likely stems from the steady water demand of families, which is not significantly affected by exams or minor holidays during the semester.

E. WEEKLY ANALYSIS

Fig. 9 illustrates the total volume of water supplied on a weekly basis over a span of four weeks. Specifically, these four weeks were chosen to highlight the impact of consecutive holidays during the third week, from 3rd October to 9th October. During this period, a considerable number of students left the campus, leading to a decrease in active occupancy within the Hostel block. Consequently, the water consumption and supply in this block also decreased. The figure clearly indicates a noticeable dip in the weekly water supply for the Hostel Block nodes during the third week. In contrast, the Faculty/Staff Block experiences only a minimal dip during the same period. This observation underscores the fact that long holidays have a more pronounced effect on the active occupancy of the Hostel block compared to the Faculty/Staff block. The disparity in impact is attributed to the differing nature of occupancy within these blocks. The Hostel block serves as a temporary residence for students, leading to a more significant impact on water usage when they are away. On the other hand, the Faculty/Staff block represents a permanent residence for the faculty, resulting in a relatively stable occupancy even during extended holidays.

Fig. 10 shows average volume of the water supplied for each day of the week for different nodes over available data

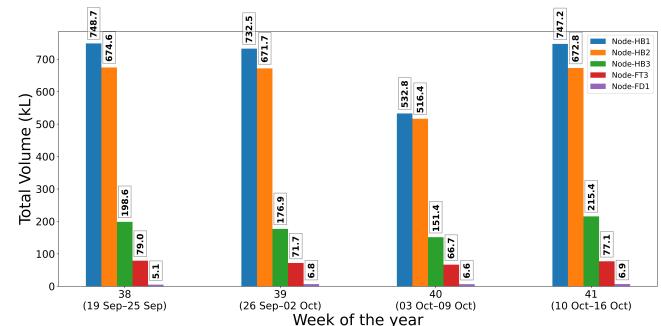


FIGURE 9. Weekly net water supply plot over 4 weeks.

in 4 months (August to November 2022). The idea behind this plot is that the average supply over a long period would be similar to the average consumption over a long period. The figure shows that the supply for each day of the week has little variation. Some intuitive observations made from the plot are:

- Water usage in the hostel block nodes is lower on weekends compared to other days. This trend is primarily due to the cleaning staff of hostels having a holiday on Sunday, and students starting their day later, spending more time resting, and some leaving the hostel, accounting for the reduced consumption. Conversely, water usage remains relatively consistent on the other days.
- In Node-FT3 of the faculty and staff block, water usage remains consistently stable across all days. This stability is largely attributed to the families residing in the building, as their water requirements remain constant, contrasting with the variable needs of students.

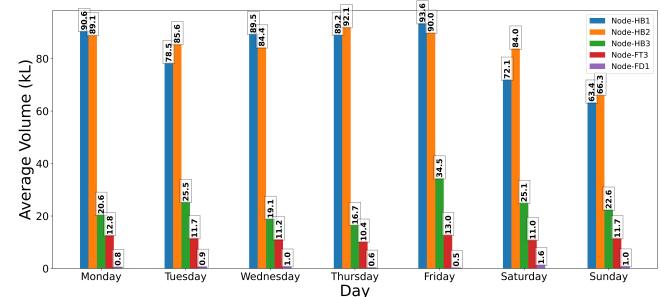


FIGURE 10. Weekday wise net water supply plot

In the analysis, graphs shown are accumulated average of the data points collected based on the time duration (monthly, weekly, day wise). The graphs show how each meter have their own characteristics based on the water usages. These characteristics are observable through out different graphs shown. For example in case of Node-FT3 water usage does not get affected due to factors like vacation or holidays. This characteristic of water usage can be observed in different graphs shown.

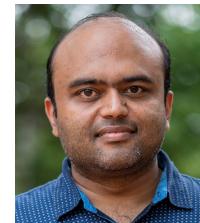
- elling & Software, vol. 125, p. 104633, Mar. 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1364815219303457>
- [14] “<http://shenitech.com/>,” accessed: Feb 22, 2024. [Online]. Available: <http://shenitech.com/>
- [15] OpenCV: Python library for real-time computer vision, accessed 10 Feb. 2021. [Online]. Available: <https://opencv.org>
- [16] Raspberry-Pi micro-controller module, accessed 21 Feb. 2021. [Online]. Available: <https://raspberrypi.org/>
- [17] S. Bianco, R. Cadene, L. Celona, and P. Napoletano, “Benchmark analysis of representative deep neural network architectures,” IEEE Access, vol. 6, pp. 64 270–64 277, 2018.
- [18] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” IEEE conference on computer vision and pattern recognition, pp. 770–778, 2016.
- [19] J. Deng, W. Dong, R. Socher, L.-J. Li et al., “Imagenet: A large-scale hierarchical image database,” IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255, 2009.
- [20] A. Paszke, S. Gross, F. Massa, A. Lerer et al., “Pytorch: An imperative style, high-performance deep learning library,” 33rd Conference on Neural Information Processing Systems, pp. 8024–8035, 2019.
- [21] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” CoRR, vol. abs/1512.03385, 2015. [Online]. Available: <http://arxiv.org/abs/1512.03385>
- [22] D. Arsene, A. Predescu, B. Pahontu, C. G. Chiru, E.-S. Apostol, and C.-O. Truică, “Advanced strategies for monitoring water consumption patterns in households based on iot and machine learning,” Water, vol. 14, no. 14, 2022. [Online]. Available: <https://www.mdpi.com/2073-4441/14/14/2187>
- [23] L. K. Narayanan, S. Sankaranarayanan, J. Rodrigues, S. Kozlov, L. Narayanan, S. Sankaranarayanan, J. Rodrigues, and S. Kozlov, “Ccc publications water demand forecasting using deep learning in iot enabled water distribution network,” International Journal of Computers, Communications Control (IJCCC), vol. 15, pp. 1–15, 11 2020.
- [24] R. Laroca, A. B. Araujo, L. A. Zanlorensi, E. C. De Almeida, and D. Menotti, “Towards image-based automatic meter reading in unconstrained scenarios: A robust and efficient approach,” IEEE Access, vol. 9, p. 67569–67584, 2021. [Online]. Available: <http://dx.doi.org/10.1109/ACCESS.2021.3077415>
- [25] I. S. Herath, “Smart water buddy: Iot based intelligent domestic water management system,” in 2019 International Conference on Advancements in Computing (ICAC), 2019, pp. 380–385.
- [26] A. Naim, A. Aaroud, K. Akodadi, and C. El Hachimi, “A fully ai-based system to automate water meter data collection in morocco country,” Array, vol. 10, p. 100056, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2590005621000047>
- [27] H. Fuentes and D. Mauricio, “Smart water consumption measurement system for houses using iot and cloud computing,” Environmental Monitoring and Assessment, vol. 192, no. 9, p. 602, Aug 2020. [Online]. Available: <https://doi.org/10.1007/s10661-020-08535-4>



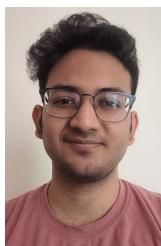
AAKASH TERALA is pursuing a Dual Degree (B.Tech + M.S.) in Computer Science at the International Institute of Information Technology, Hyderabad, India. He is a member of the Signal Processing and Communications Research Centre (SPCRC), where he focuses on IoT, Data Analysis, and Dashboard development. His interests include data analysis, machine learning, and DBMS.



ARCHIT GOYAL is pursuing a Master’s degree in Computer Science and Engineering at the International Institute of Information Technology, Hyderabad. He is also working as a researcher at the Signal Processing and Communication Research Center (SPCRC), focusing on research associated with the Internet of Things field. As a software engineer, he has contributed to projects like elm-dagre, an open-source package for drawing graphs.



DR. SACHIN CHAUDHARI received his B.E.(Electronics) from Visvesvaraya National Institute of Technology (VNIT), Nagpur, India in 2002; M.E.(Telecommunication) from the Indian Institute of Science (IISc), Bangalore, India in 2004 and his D.Sc. (Tech) from Aalto University (formerly TKK), Finland in 2012. Between August 2004 and May 2007, he was with Esquibe Communications, Bangalore, India, as a senior wireless communication engineer. During 2013–2014, he was a post-doctoral researcher at Aalto University. He joined the International Institute of Information Technology (IIIT), Hyderabad, India in Dec. 2014, where he is currently an Associate Professor. Sachin’s research interests are in the field of signal processing and machine learning for wireless communication and particularly, in the physical layer aspects of the internet of things (IoT), 5G, and cognitive radio. He is a senior IEEE member and the representative of IIITH to the Telecommunication Standards Development Society of India (TSDSI). He is the coordinator of the Center of Excellence on IoT for Smart Cities at IIITH, which is supported by the India-EU collaboration project, ETSI, and TSDSI. He is also actively involved in India’s First Living Lab for Smart City Research at IIITH.



AYUSH K. LALL is currently pursuing an integrated B-Tech and MS in research in Electronics and Communication Engineering (ECE) at the International Institute of Information Technology, Hyderabad. He is involved in research at the Signal Processing and Communication Research Center. His research focuses on applying machine learning to the Internet of Things. In addition, he also works with fields in computer vision, deep learning, and image processing.



PROF. K.S.RANJAN earned his Ph.D. from the University of Tokyo and holds a B.Tech. degree from IIT Kharagpur. He joined IIIT Hyderabad in 2006 and is currently a professor there and heads Lab for Spatial Informatics. His research areas span Geographic Information Systems (GIS) and remote sensing. He has published over 130 papers in peer-reviewed journals and top conferences.



DR. SHAILESH SINGH CHOUHAN received the D.Sc.(Tech.) degree in microelectronics from Aalto University. He is currently a Senior Researcher in CPS with the Luleå University of Technology. He works to bridge the gap between electronics and IT to achieve the vision of CPS. His research interest includes chiplet design and computational hardware integration into microsystems and microservice paradigms.

• • •