IEC
Institute of Electronics and Computer

Journal of Artificial Intelligence and Systems, 2023, 5, 1-22
https://iecscience.org/journals/AIS
ISSN Online: 2642-2859

Analyzing the Carro Pipa Operation with Geointelligence Techniques

Aloísio Vieira Lira Neto[1,*], Elias Paulino Medeiros [2], Filipe Maciel de Moura[3], Jose Wally Mendonça Menezes[1], Senthil Kumar Jagatheesaperumal[4], Victor Hugo C. de Albuquerque [2]

[1] Federal Institute of Ceará, Fortaleza/CE, Brazil
Email:aloisiolira.ie@gmail.com; wally@ifce.edu.br
[2] Department of Teleinformatics Engineering (DETI), Federal University of Ceará, Fortaleza/CE, Brazil
Email: elias.paulino@alu.ufc.br; victor.albuquerque@ieee.org
[3] State University of Ceará, Fortaleza/CE, Brazil
Email: filipemaciel92@yahoo.com
[4] Mepco Schlenk Engineering College, Sivakasi, India
Email: senthilkumarj@mepcoeng.ac.in
*Corresponding Author: Aloísio Vieira Lira Neto, Email: aloisiolira.ie@gmail.com

How to cite this paper: Aloísio Vieira Lira Neto, Elias Paulino Medeiros, Filipe Maciel de Moura, Jose Wally Mendonça Menezes, Senthil Kumar Jagatheesaperumal, Victor Hugo C. de Albuquerque (2023). Analyzing the Carro Pipa Operation with Geointelligence Techniques. Journal of Artificial Intelligence and Systems, 5, 1–22. https://doi.org/10.33969/AIS.2023050101.

Received: December 5, 2022
Accepted: February 21, 2023
Published: March 2, 2023

Abstract

Operation Carro Pipa (OCP) is a federal government action in Brazil with the objective of distributing drinking water to regions severely affected by long periods of drought and low rainfall using trucks. The region served has continental dimensions, covering an area of 688,064 km² and supplying 1703 cities and approximately 5.2 million people. Because it is a large-scale action that involves a lot of public resources, it is essential that all OCP activities are recorded in a safe, complete, and standardized way. This information can be used for potential benefits, audits, and analysis to propose improvements and ensure the provision of this essential service to society. To achieve this, this work analyzes data recorded from the OCP service offered in a Brazilian state and presents computational solutions that can improve the monitoring, registration, storage, and processing of the data generated by this action.

Keywords

Operation Carro Pipa; Internet of Things; Data analysis; Northeast of Brazil; Internet of Things

1. Introduction

Water is an essential resource for sustaining life and supporting numerous activities. Therefore, we can say that developing actions and projects that prioritize access to this natural resource is crucial. The northeast region of Brazil stands out for having a high population density and, due to climatic adversities, associated with other historical, geographic and political factors that go back hundreds of years, it is home to the poorest portion of the country's population [1]. This region is characterized by low rainfall rates, where the maximum annual precipitation reaches values around 200 to 960 mm, in addition to high insolation with an annual average of 2,800h.year, average annual temperatures around 23 °C to 27 °C, average evaporation of 2,000 mm. year and average relative humidity around 50% [2]. This set of factors causes this region to have a series of deficits in basic infrastructure, not just the lack of access to water, but to basic conditions such as food, education and leisure.

Historically, several measures were taken by the Brazilian government to alleviate the problems caused by water scarcity, such as: dam projects, integration of rivers, construction of cisterns, water distribution [3]. In 2005, the federal government established an action called Operation Carro Pipa (OCP) with the objective of distributing drinking water to the most affected population by means of trucks [2]. Currently, this action serves around 5.2 million people in 1703 cities [4]. According to a report by the Federal Court of Auditors (TCU), in 2012, this action already covered 782 cities, with an area of operation of 688,064 km², with approximately 70,000 supply points, serving a population of approximately 3.8 million people, which which corresponds to approximately 760 thousand families, and until today this action has been carried out with proportional dimensions.

Due to the size of this program, a lot of information is generated from the records of actions performed daily, which are later used for accountability with public bodies and for monitoring civil society regulated by the law on access to public information (Law No. 12,527 , of November 18 , 2011). However, some factors make it difficult to analyze these data, as they are structured in spreadsheets in a decentralized way by the management bodies in the respective regions of operation. This form of data treatment has already been signaled by the TCU as a point that can reduce the effectiveness and efficiency of the program.

As with OCP, data generation is possible in almost all sectors of human activity, especially after the advancement of the Internet of Things (IoT) concept, where this technology is incorporated into a wide spectrum of products, systems and sensor systems of network giving the possibility of offering society new resources and services that were not possible before. The IoT promises to interconnect anything, be it internet-enabled home appliances and medical devices, wearable devices monitoring vital signs of the human body, and vehicles sharing information helping for example to avoid congestion, that is, this technology offers the possibility to transform several sectors such as, agriculture, industry, health and safety, capturing, generating and transmitting information through sensors and network adapters [5].

Despite the benefits that IoT can bring, several issues and challenges need to be considered in the following points: infrastructure, interfaces, protocols, storage and data processing [6]. Currently bandwidth and storage is a challenge for real-time processing of big data in IoT networks, so several works have been considering the application of artificial intelligence (AI) techniques to summarize data captured by video sensors as a way to discard data redundant and facilitate the retrieval, indexing and navigation of information [7, 8, 9, 10, 11]. In the same line S. Liu et al. [12] works on building an algorithm for real-time remote monitoring in the context of multimedia data; Mehmood et al. [13] proposes a lightweight deep learning-based system running on power-constrained devices for retrieving big data from images generated in an IoT environment; Zhang et al. [14] works on a protruding object detection process in surveillance videos; Magaia et al. [15] presented the concept of IIoT and applications for smart cities, in addition to presenting security challenges faced by this area; Muhammad et al. [16] investigated communication methods that move complex computation from the cloud server towards the edge of the network; Silva et al. [17], Sodhro et al. [18] and Sodhro, Luo, et al. [19] addressed the Internet of Vehicles (IoVs) context by presenting algorithms, network architecture and listing their benefits and limitations.

Knowing how to deal with the data generated by the IoT and human activities in general is a challenge, because when treated in the right way, they can be used to identify patterns, behaviors, new ways of understanding the world, that is, to extract knowledge from a set of data and which it can be used to improve services offered to society [20]. These data were neglected for a long time, however, in recent years several studies have been carried out with the aim of seeking new ways to capture, persist, structure and summarize all the material generated from the computerization of human interactions [15, 16, 21]. For example, Liu et al. [22] used data analysis and visualization to assess the collective tactical behavior of team sports players in order to observe points such as fatigue, subsequent adaptation responses, performance, and reduce the risk of injuries and illnesses. Moalla et al. [23] used Data Warehousing techniques to analyze sentiments in texts extracted from twitter. Li et al. [24] analyzed massive data generated from the Internet of Things (IoT) of a smart city in order to make the smart city change towards good governance.

Database Management Systems (DBMS) is a set of software used to manage a database, responsible for controlling, accessing, organizing and protecting the information of an application that has the advantages of ease in data sharing. , efficient management, security, avoids redundancies and inconsistency in data, speed in data manipulation and access to information, and avoids problems related to data integrity [25]. The use of a DBMS in conjunction with a web system, replacing the use of spreadsheets, can avoid the aforementioned problems and centralize the information in a single database where it can be processed and analyzed in real time and updated as new information is received provided as input. According to Fayyad et al. [26] the process of discovering knowledge in data is carried out with the steps of selection, pre-processing, transformation, data mining and interpretation/evaluation, and if the data are centralized and well planned in a DBMS, the effort dedicated to the 3 first steps slows down considerably, both human and computational.

Due to the size of this program and because it is a public policy that charges a considerable financial resource, it is essential that the OCP contemplates efficient monitoring, management and inspection mechanisms, capable of determining data that serve as a basis for a more efficient management in order to to avoid not only fraud, but also to scale the expansion or reduction of vehicles and inputs when this is the case. Based on this understanding, an efficient management tool must include the identification of occurrences of deviations and inconsistencies between the planning and execution of OCP deliveries and, in addition, the implementation of new control mechanisms must be compatible, to guarantee r that the water was collected in the appropriate places, the regularity of the supply and the delivery to the right recipients, beneficiaries of the program. Therefore, the main contributions of this work are:

- Develop an algorithm capable of pre-processing and transforming OCP data.
- Analyze OCP data in the state of Ceará-Brazil.
- Generate a data visualization capable of changing the direction of good OCP governance.
- Propose a computational system capable of receiving OCP data as input, storing and processing this information in real time.
- Propose an Internet of Things (IoT) model capable of capturing the activities carried out by OPC service provider trucks using sensors.

2. Materials and Methods

2.1. Dataset

In this work, data present in a file in PDF (Portable Document Format) format with 34631 pages, built from the junction of hundreds of spreadsheets with information on the execution of the Carro-Pipa Operation in the state of Ceará, between the months of January and August 2021.

According to the way in which the file is structured, it is believed that the process involved in its construction is something similar to the details shown in Figure 3. It can be seen that it is composed with information from several spreadsheets from different contexts:

I. Spreadsheets with information on deliveries made by a water truck to a beneficiary;
II. Spreadsheets with the description of the delivery point of each beneficiary; and
III. Spreadsheets with the description of each source used by the operation.

We adopted to call each of these spreadsheet Detailed Delivery (DE), Detailed Delivery Points (DPE) and Detailed Supply Points (DPA), respectively. Figure 1 shows how the data was structured in the PDF file.

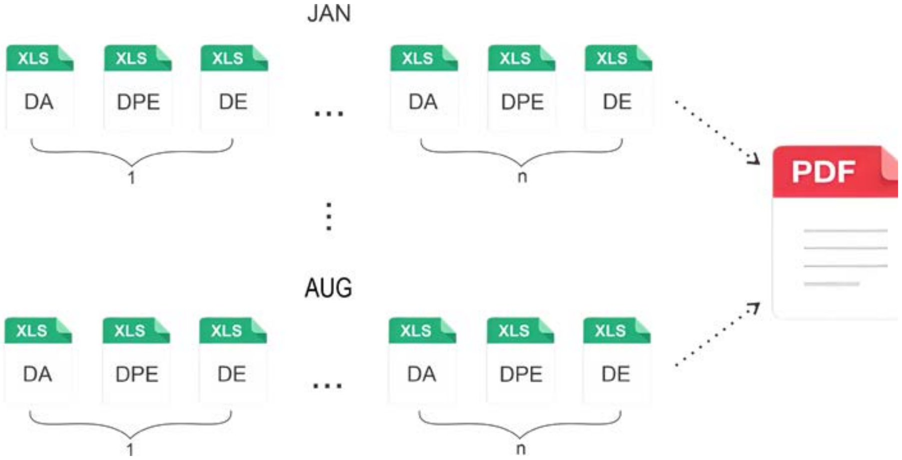

Figure 1. Construction methodology of the PDF file with data from the Carrro-Pipa operation.

2.2. Pre-Processing

In order to be able to work with the data, an algorithm was developed to carry out the inverse process of the PDF file, that is, transform it into several data sheets. However, a manual conversion process proves to be impracticable due to the characteristics of the file, since, as seen in the previous section, it has approximately 34,000 pages, the result of merging the content of hundreds of spreadsheets. Therefore, an alternative to the manual process was to build an algorithm that would allow the computer to read the PDF and make the file conversion an automated process. However, it was necessary to adopt a previous step in order to visualize and understand the way in which the data in the spreadsheets were arranged/structured after creating the PDF.

As a way of illustrating the visual perception obtained from a brief reading of the PDF, Figure 2 presents some screenshots of different pages of the file. For example, in screenshot 01, the beginning of a DE-type table can be seen, and this data continues until pages 10-11, illustrated in screenshot 02. On page 11, a DPE-type table begins, however, with only part of its columns. The information in these columns continues on page 12 (screenshot 03) to page 72. On page 73, presented in screenshot 04, continue the table started on page 12, with more columns of information, and continue to page 134. On page 135 (screenshot 05), we can see the beginning of another column of information from the same spreadsheet that ends only on page 196. Therefore, it was possible to notice that the data columns of the DE and DPA spreadsheets are most often on the same page, however, when the spreadsheets are of the DPE type, the data columns are usually presented in several ranges of pages. This understanding of the structure of the PDF was necessary to be able to identify these patterns, and thus allow the construction of an algorithm that could do the conversion

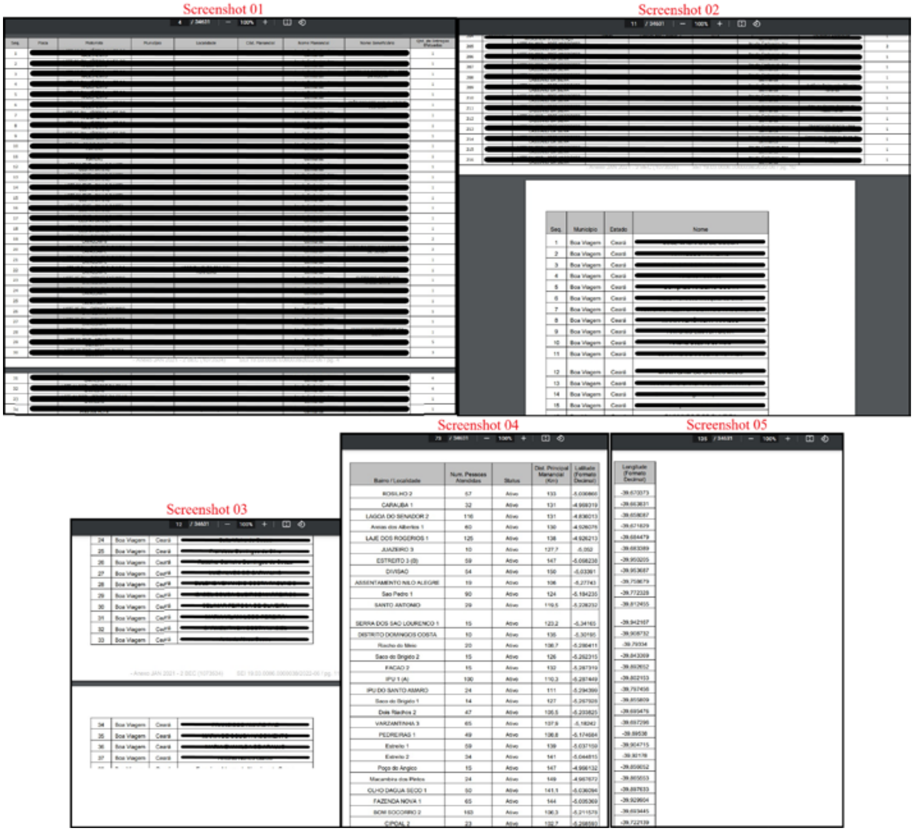

Figure 2. Screenshots taken during the visual analysis of the PDF file. Some fields in the figure have been crossed out to preserve personal information present in the data.

by dividing the data into files according to their context (DE, DPE and DPA).

Seeking to optimize the analysis, an algorithm was created in Python 3.9 programming language and the pandas and tabular libraries. Through functions of the tabular library, it was possible to read the PDF file page by page and load them into the computer's memory in table format of the DataFrame type - data type supported by the pandas library.

After processing the PDF by the algorithm, it was possible to extract a total of 269 data sheets, all identified in one of the three contexts (DE, DPA and DPE) and the reference month.

Algorithm 1 receives a *P file (PDF file)* as input, and provides three sets as output, where each one will be composed with files extracted from each context. Lines 4 – 6 show the vectors with the names of the columns that represent the data of each

```
Algorithm 1: convert pdf to DE, DPE and DPA spreadsheets
Input: P.
Output: F_dpe, F_dpa, F_de.
begin
  de: spreadsheets with details of deliveries;
  dpe: spreadsheets with details of delivery points;
  dpa: spreadsheets with details of supply points;
  for i from 1 to length(P) do
    h = header(P[i]);
    if h is different of null then
      Create file f;
      Add P[i] in f;
    end
    if h is equal to null then
      Add P[i] in f;
    end
    if header(P[i + 1]) is different of null then
      if h is equal to de then
        Add f in F_de
      end
      if h is equal to dpe then
        Add f in F_dpe
      end
      if h is equal to dpa then
        Add f in F_dpa
      end
    end
  end
end
```

Algorithm 1. Algorithm adopted in data analysis.

context. Line 7 generates an index counter associated with PDF pages, ranging from 1 to the number of pages in the input file. Line 8 retrieves the header of the accessed page and uses this information to verify the page corresponds to a beginning, middle or end of file (Lines 9 - 16). If at some point it is identified that the next page corresponds to the beginning of another file, the current one is saved in one of the sets of files F_{dpa}, F_{dpe} and F_{de}, according to its respective context.

Table 1 presents the quantities of converted spreadsheets per month and type. If we observe Table 2, there is no column in the data indicating the month in which the service was executed. However, on each page of the PDF, there is a watermark with the temporal information indicating when that data was manipulated. Therefore, by providing the page intervals to the algorithm, it was possible to assign to each extracted spreadsheet the reference month of that information. Thus, for each spreadsheet extracted from the PDF, the description of this new file was defined in the following pattern: SHEET_ID_TYPE_MONTH.xls, where ID is an integer value corresponding to the number of spreadsheets converted up to that moment, TYPE is the data context (DE, DPA or DPE), MES is the data reference month, and xls is the extension in which the file was saved.

Table 1 Number of spreadsheets extracted from the PDF through the execution of the constructed algorithm.

Type/Month	JAN	FEB	APR	May	June	JUL	AUG
DE	4	4	12	18	15	12	9
DPE	5	5	16	17	22	20	16
DPA	5	5	15	18	15	16	18
Total	14	14	43	52	52	48	43

Table 2 Information from the different spreadsheets that make up the PDF file with the Carro-Pipa operation data.

Spreadsheet-Type	Columns
DE	Plate, Driver, City, Locality, Cod. Source, Name Source, Beneficiary Name, Qty. of deliveries
DPE	Seq., City, State, Name, Locality, Num. Persons Served, Status, Dist. Spring, Latitude, Longitude
DPA	Seq., State, City, Status, Cod. GCDA Spring, Latitude, Longitude, Hierarchical Level

With this pattern of description of the generated files, it was possible to build another algorithm (Algorithm 2) to group, according to the context, the data of the 269 spreadsheets in 3 different files, and at the same time adding in each one, a

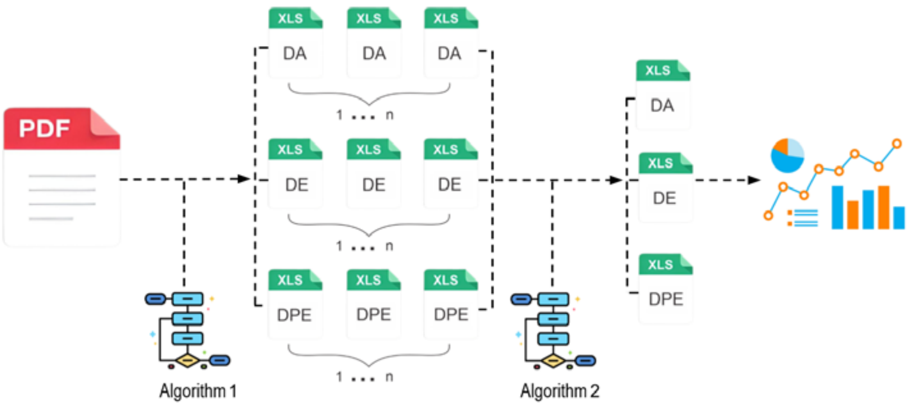

Figure 3. Approached methodology from pre- processing to data analysis.

column with the information of the month of data reference. Figure 3 illustrates the entire approach adopted in this work, from data pre-processing to the developed analysis.

After grouping the spreadsheets by context, two more pre-processing were performed:

I. All records of service information carried out in cities in the states of Pernambuco, Piauí and Bahia were discarded from the data;

II. Was added to the spreadsheet with information on the distance from the source to the delivery location. This information was extracted from the column 'Dist. Spring' present in the DPE spreadsheet.

2.3. Challenges found

During the execution of the PDF-spreadsheet conversion algorithm, it was possible to identify some problems in the PDF file:

I. Presence of some blank pages;

II. Some data from the DE spreadsheet does not have the grid that delimits the rows and columns of a spreadsheet, so when the function used to read data from the page is faced with this situation, it cannot recover the data in the structured form necessary for the functioning of the algorithm, so this data was discarded;

III. The range of pages referring to the month of March does not have information on the description of deliveries made (DE spreadsheet), therefore, data for the respective month were not included in the study.

3. Results and Discussion

Data from the water truck operation are distributed in 3 spreadsheets, according to the context: DE (description of deliveries made), DPE (description of delivery points) and DPA (description of supply points - springs). Therefore, the data were analyzed and will be discussed independently in this section, and together where appropriate. The DE spreadsheet is a type of file where information is stored regarding deliveries made by a water truck to a beneficiary. For every effective delivery, the following information is recorded: vehicle license plate and driver's name, beneficiary's name, city and location where the delivery took place, code of the source where the vehicle was supplied, number of deliveries made and the month in which the delivery took place.

When starting the process of analyzing the data in this spreadsheet, it was soon found that there may be several records of delivery to a beneficiary in the same month, that is, a beneficiary can receive more than one delivery of water by a water truck in the same month. A record is considered to be a complete line of information from the spreadsheet.

The first analysis carried out in this work was to verify the behavior of deliveries made by location and beneficiary, and they are detailed in the boxplots of Figure 4

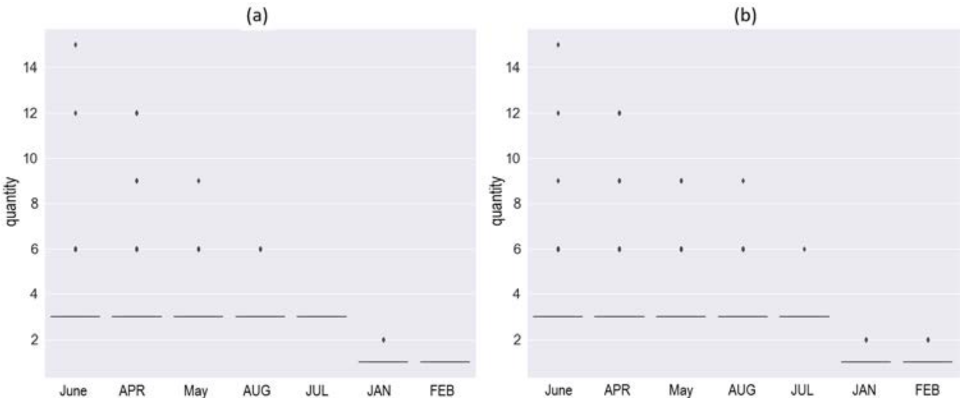

Figure 4. a) Delivery records by location. (b) Delivery records by Beneficiary.

(a) and (b), respectively. It can be seen that in both situations the data have practically the same behavior, indicating that possibly every beneficiary is associated with a single location. In both cases, the average of delivery records per location and beneficiary is around 3 between the months of APR and JUL, and 1 delivery in JAN and FEB, with a difference of just a few points of outliers in the months of JUN, JUL and AUG. These differences in (a) and (b) are motivated by cases in which the same beneficiary name is associated with more than one location, but it is not possible to state whether they are the same or different people, as there is no unique identifier in the data for each beneficiary.

Table 3 was created from the investigation into the data of beneficiaries and locations that represent the outliers common to Figure 4 (a) and (b). It is noticed that the cities common to these locations/beneficiaries are Campos Sales, Salitre and Parambu and that the location QUIXARIU (SEDE URBANA 03) stands out with high rates of quantity of deliveries in the months of APR, MAY and JUN. This location is located in the city of Campos Sales and received 15 deliveries in the month of June, that is, one delivery every two days for the consumption of 95 people. Knowing that this location is 112km away from the source, just on the one-way transfer, 1680 km were traveled in a single month, and the same water truck still made 18 deliveries in other locations in the same period.

Table 3. The top 10 locations and beneficiaries with the highest delivery records. Beneficiary names were coded in the Beneficiary column as a way to preserve personal information.

Location	City	Recipient	Month	Distance from spring	Amount
QUIXARIU (SEDE URBANA 03)	CAMPOS SALES	B1	June	112	15
SITIO ACUDE NOVO	SALITRE	B2	APR	60	12
BAIRRO BARREIROS A	SALITRE	B3	APR	55	12
HOSPITAL CICERO FERREIRA	PARAMBU	B4	June	76	12
ALTO ALEGRE 03	SALITRE	B5	APR	52	12
SITIO TANQUE NOVO	SALITRE	B6	APR	110	12
ALTO ALEGRE II B	SALITRE	B7	APR	18	12
QUIXARIU (SEDE URBANA 03)	CAMPOS SALES	B1	APR	112	9
QUIXARIU (SEDE URBANA 03)	CAMPOS SALES	B1	May	112	9
MAPIRINGA PAU DARCO	SALITRE	B8	APR	42.2	9

Another analysis carried out in this work was related to the behavior of the number of deliveries made by city. For this, Figure 5 helps to understand the results and it is possible to observe that AGO is the month with the highest average delivery in the cities, approximately 300, and those with the lowest averages are JAN and FEB.

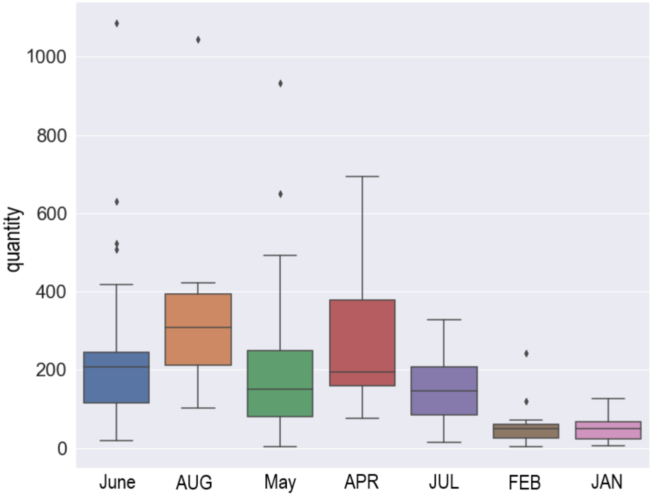

Figure 5. Records of deliveries made by city.

When carrying out the survey of the 10 largest quantities of delivery records, Table 4 shows that the cities that record the largest numbers are: Parambu, Morada Nova, Campos Sales, Salitre and Aiuaba. Parambu was the city that benefited from the highest number of deliveries in the months of ABR, MAY, JUN and AUG, however, if you add the distances from the locations to the water sources in this city only in the month of JUN, it was necessary to travel 82725 km to carry out the 1086 deliveries.

Table 4. The top 10 largest quantities of records of deliveries made by city.

City	Month	Quantity Deliveries
Parambu	June	1086
Parambu	AUG	1044
Parambu	May	933
Morada Nova	APR	693
Parambu	APR	687
Morada Nova	May	651
Morada Nova	June	630
Campos Sales	June	522
Salitre	June	507
Aiuaba	May	492

Figure 6 presents the boxplots (a) and (b), resulting from the analysis of the delivery records by information on the water truck plate and by driver's name. It is observed that the two have the same tendency, which gives a margin of safety in saying that generally a driver is associated with a water truck.

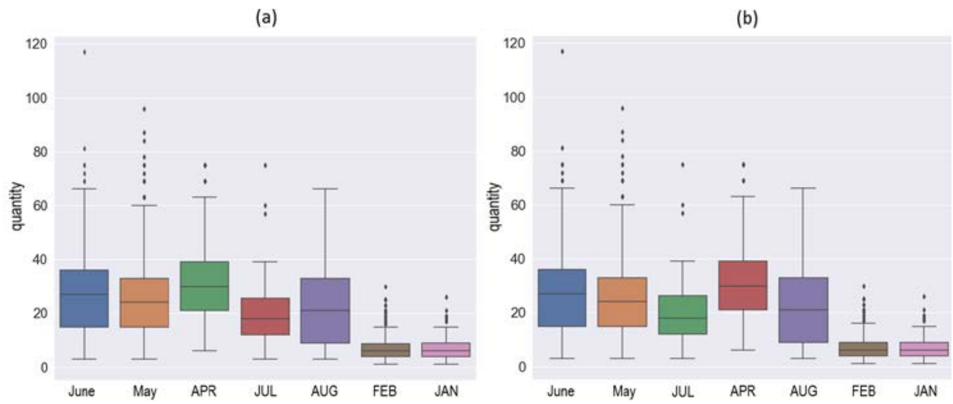

Figure 6. (a) Delivery records by water truck. (b) Delivery records by driver name.

In Figure 6 (b) it can be seen that in JAN and February (FEV) the average number of records is around 10 deliveries per water truck/driver, and in the period from APR to AGO the average rises to the range of 20 to 35 deliveries. However, in JUN, a single water truck/driver has 117 records of deliveries, providing support to 39 different locations in the city of Jaguaribe, and with supply always from the source called SAAE SEDE JAGUARIBE. If for each delivery record it is necessary to fill up at the source, this water truck/driver needed to travel 2874km to meet the 117 records, considering only the one-way distances from the source to the delivery point. Another study carried out in this work analyzes the behavior of the number of active springs, water trucks used, distances covered, locations and cities served in the state of Ceará between the months of JAN and AUG.

Figure 7 shows that while the lines of cities served and active springs remain stable over the months, the others show variations with large peaks and valleys, for example,

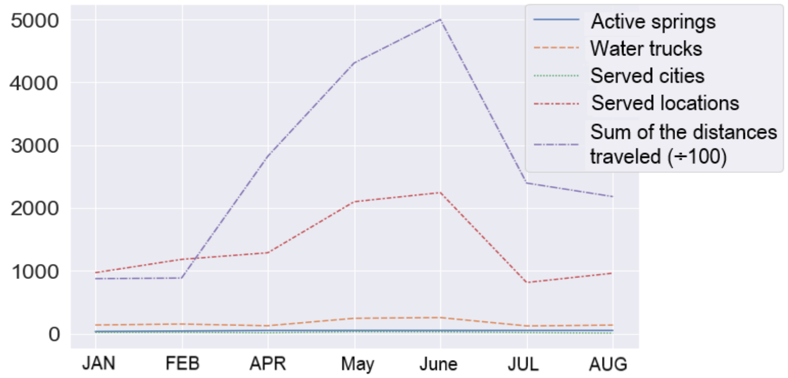

Figure 7. Number of water trucks used, from locations and cities served, active springs and sum of distances traveled (in Km) between the months of JAN to AGO.

the line of locations served and the sum of distances covered. This last index is the sum of all the distances from the source to the locations served in the delivery record sheet.

Comparing the numbers of water trucks and locations served, an inverse variation of the rates between FEV and ABR can be seen, that is, in the time that the number of locations increased, the number of water trucks had a slight decrease. On the other hand, the behavior becomes directly proportional between the months of ABR and AUG for the same lines. Another point that draws attention is the sharp growth in the number of locations served and the sum of the distances covered between the months of FEB and JUN, as this period corresponds to the winter season in the state. On the other hand, the drop in these indices in AGO is striking, as a dry period begins. It can also be seen in Figure 7 that in JUN was the peak of distances traveled, that is, to carry out all the deliveries recorded in the DE spreadsheet in that month, it was necessary to travel approximately 500,000km to supply locations in 28 cities. By analyzing the data from the DPA spreadsheet, it was possible to investigate the cities and regions of Ceará with the greatest supply of active and inactive springs.

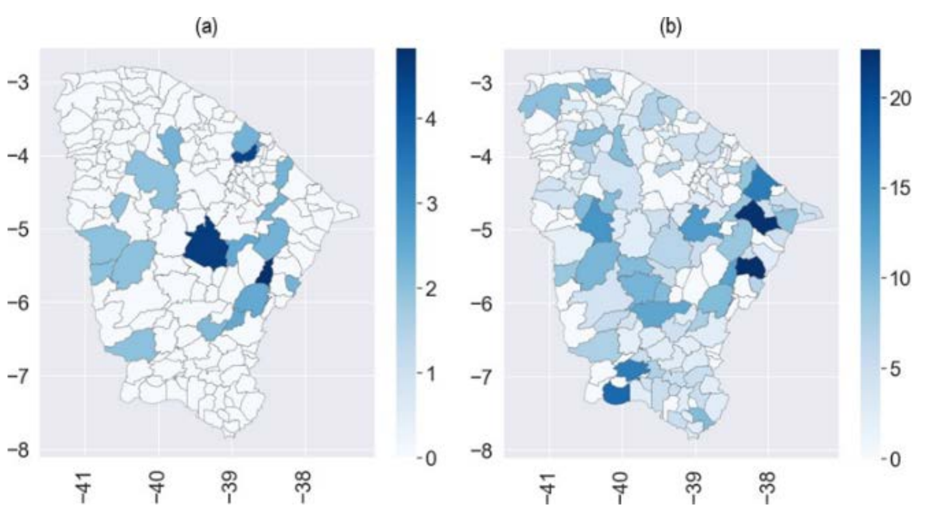

Figure 8. (a) Mapping of the average number of active springs. (b) Mapping of the average number of active and inactive water sources. Average calculated in the period from JAN to AGO only for cities in the state of Ceará.

Figure 8 helps to understand how these sources are distributed within the state, where (a) it takes into account only active sources in the period from JAN to AGO, and (b) it considers active and inactive sources. In Figure 8 (a) it can be seen that the springs are located in certain regions of the state such as: Vale Jaguaribe, Centro-Sul, Sertão Central, Inhamuns and in some cities close to the capital. Now, when the inactive are also considered, there is a better distribution of water sources within the state. Therefore, a possible way to reduce the distances traveled by water trucks would be to activate more water sources as the demand for deliveries increases in different regions. This effort is not seen in Figure 9, because as the number of locations served increases considerably, the number of active springs remains constant. With this, a question is pertinent: is the cost of maintaining an active source greater than keeping water trucks covering greater distances?

Finally, two more investigations were carried out on the data referring to the description of the delivery points, present in the DPE spreadsheet. In the first, the cities and regions with the highest amounts of demands for water deliveries were analyzed, therefore, in Figure 10 (a) was built based on the average number of registered locations, and (b), it was based on the average amount of people assisted in the registered locations. When the two figures are compared, it can be seen that (a) is very concentrated in the Centro-Sul, Sertão-Central, Vale Jaguaribe and Inhamuns regions, that is, they are the regions that have the largest number of registered locations, with emphasis on some cities described in Table 5. For example, based on this index, Quixeramobim and Tauá are considered the driest cities in Ceará.

In Figure 10 (b) one can see a uniform distribution of the driest cities within the state, calling attention to some cities on the coast. Table 6 shows the city of Capistrano, located in the Maciço do Baturité region, with an average of 149 people served, and the cities Camocim, Paracuru, Jijoca de Jericoacoara and Fortaleza, all located on the coast, with an average of over 121 people.

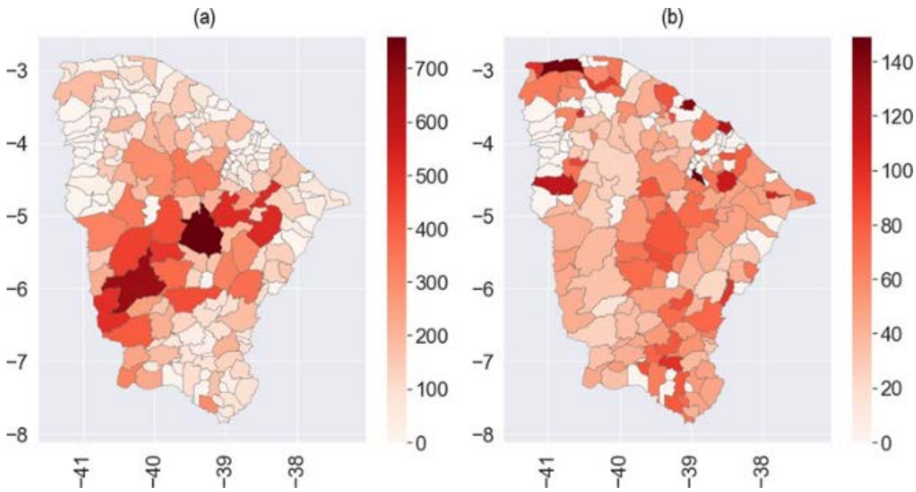

Figure 9. (a) Mapping of the average number of registered locations (active and inactive). (b) Mapping of the average number of people served in the registered locations. Average calculated in the period from JAN to AGO only for cities in the state of Ceará.

Table 5. The top 10 cities with the highest average number of registered locations.

City	Average number of registered locations
Quixeramobim	760
Taua	678
Morada Nova	517
Quixadá	515
Parambu	507
Pedra Branca	487
Independencia	465
Acopiara	437
Boa Viagem	434
Aiuaba	406

In the second and last analysis of data from the DPE spreadsheet, the distances between the delivery points of the municipalities and their respective supply sources were analyzed.

In Figure 10 (a) it can be seen that the average distance from the springs to the active delivery points is around 75km, and the cities with the highest averages are:

Itatira (156km), Canindé (142km), Boa Viagem (130km), Acopiara (130km) and Tauá (130km).

Table 6 The top 10 cities with the highest average number of people served.

City	Average number of people served
Capistrano	149
Camocim	144
Paracuru	138
Jijoca de Jericoacoara	121
Fortaleza	121
Ipueiras	117
Ocara	110
Itaiçaba	105
Alcantaras	101
Barroquinha	99

Finally, when analyzing the sum of the distances from all delivery points in a city to their respective water sources, that is, the mileage required to go at least once from the water source to all delivery points in the city. Figure 10 (b) shows that the average of these sums is around 10000km for the hotspots. The cities that stand out with this index are: Tauá (88918km), Parambu (34158), Campos Sales (21528km), Acopiara (18702km) and Aiuaba (16785km).

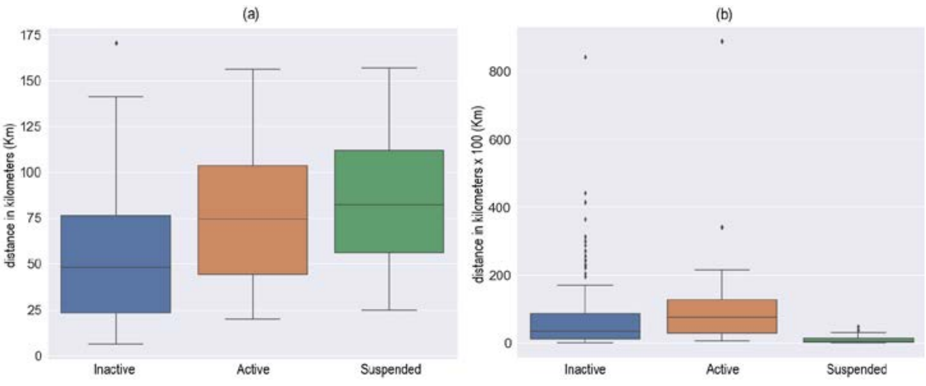

Figure 10 (a) Average distance between the delivery points of the cities and the water sources. (b) Sum of all distances between water sources and municipal delivery points.

3.1. Computing System Proposal

In view of the data and issues mentioned, as a result of the analyses, this work presents a proposal for a model based on optimizing the routing of OCP deliveries, considering the distances between the sources and delivery points. Another possibility would be the elaboration of a model of water sources alternative to those proposed, such as deep wells and other reservoirs not registered in the OCP database.

The proposed model suggests the development of a web application that can be fed with OCP information and centralize the data in a single database. This information can come from the registration of pipers, delivery points and springs, beneficiaries and deliveries made. With this, permanent and dynamic views, which are updated from the insertion of new data, can help in the periodic analysis and in the alert of possible irregularities in the OCP service. The GPS and flow and water quality sensors attached to the truck can daily update the truck's route, the amount of water dumped at each supply point and the quality of the water used.

The first type of visualization is of the map type, where all the sources and supply points will be geographically arranged, with the ability to suggest the best route between a starting point and destination, as well as indicating the best source to supply a location. The second type of visualization is graphics that can serve as an alert for possible irregularities in the service offered within a state, city, by a piper and delivery location, presenting information such as: number of cities served by state throughout the year, relation pipeiros X population served, number of people served per month, number of deliveries made by location, accumulated distance covered by

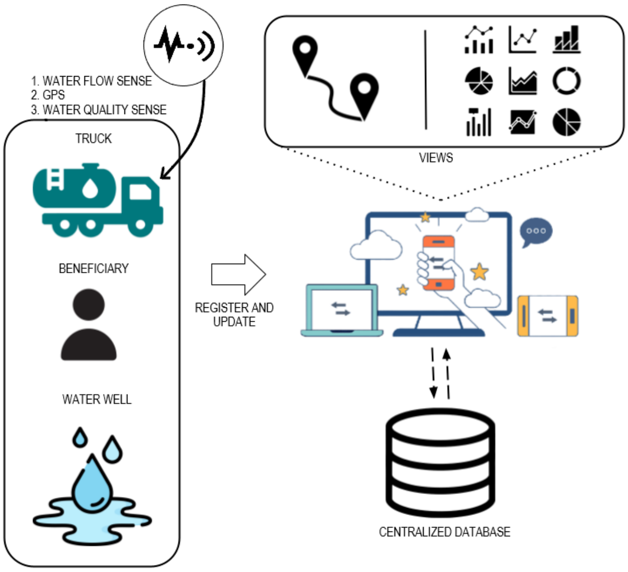

Figure 11 Diagram of the proposed computational system.

pipeiros, average water/person/day and percentage of deliveries made. The application design can be seen in Figure 11.

4. Conclusions

In view of the analyzes carried out based on the data and official documentation, they point to a complex structural situation in the OCP, which highlights a series of deficiencies, namely:

I. Insufficiency of rules that include, in exceptional cases, the flexibility of the goal of the amount of water to be distributed per person/day. It should be noted that, according to legal regulations, the distribution of water at OCP provides for a consumption of 20 liters of drinking water per person/day. However, it was verified in the available documentation that in some locations, this average of liters of water/person/day is dimensioned above 20 liters, as determined by the local Coordination of the Operation.
II. Persistence in the absence of strict norms and without legal “loopholes” by the OMEs (Military Executing Organizations) in defining the volume of drinking water per person/day, which could lead to an unequal distribution of water to the population benefited by the program.
III. Available data point to some states, such as Piauí, where previously established targets were 150% higher than planned.
IV. Submission of water potability reports after the deadline, which, according to Interministerial Ordinance 1/MI/MD, establishes that these are the responsibility of the cities, both carrying out and sending them to the managing body of the OCP, on a monthly basis.
V. There is no formal maximum limit on the number of loads per pointer.
VI. Absence of predefined criteria in the registration of pipeiros.
VII. Absence of norms that define a minimum frequency for carrying out inspection/survey on water trucks.

Seeking to obtain another source of information, the allegations of irregularities in the stages of execution of the OCP from TC 008.231/2010-9 were analyzed, the main problems pointed out, as stated in the document in question are:

I. Frauds in the measurement of the volume of water trucks and in the 'pipemen's queues'.
II. Stamps attesting non-existent truck supplies.
III. Ilegal sale of potable water or spilling it along routes, as a way to save fuel, refueling vehicles when approaching the final destination with water unfit for human consumption.

IV. Tampering with the mileage of vehicles between springs and filled cisterns benefiting 'pipemen', who earn per kilometer driven.

V. Use of program resources in activities outside its purpose.

VI. Negotiations with ghost companies, active corruption (payment for the silence of people who should have benefited from the program, but did not.

VII. External signs of wealth of 'pipeiros' registered in the program.

The adoption of a platform/application for real-time georeferenced monitoring of the routes of OCP vehicles is recommended, considering the following premises:

I. Registration of the beneficiary population, pointers and respective magnetic cards, collection points (registered springs), locations for water delivery, and vehicles used by the Operation.

II. Issuance of reports of confirmations of water deliveries.

III. Visualization of registered vehicles on georeferenced maps, including the possibility of plotting the route traveled and issuing a tracking report by kite car (distance traveled, stopped time, position history, references and trip summary, etc.);

IV. Suggestion permission/prior definition of vehicle traffic routes, issuing exit or entry alerts in these areas.

Official documentation points to an OCP logistics monitoring service, contracted through the bidding process (Electronic Auction 24/2012 - MI), which the TBK consortium has as its service to provide Embedded Monitoring Module (MEM) technology in each water truck contracted at OCP. However, these data are not available for analysis, we suggest the availability of such information.

In addition to this instrument, it was observed that inspection is carried out by the Executing Military Organization - OME and by the Northeast Military Command - CMNE. Even if in a preliminary way, it is possible to conclude that these problems related to OCP, can be regulated and mitigated from the adoption of appropriate means and technologies for the service in question, however, it should be noted that the integration between the planning teams, execution and supervision.

References

[1] Ab'Sáber, A. N. (1999). Nordeste sertanejo: a região semi-árida mais povoada do mundo. *Estudos avançados*, *13*(36), 60-61.

[2] Brasil. Ministério da Integração Nacional. Relatório final do grupo de trabalho interministerial para a redelimitação do Semiárido nordestino e do polígono das secas. Brasília, DF, 2005a. 1 CD-ROM.

[3] Duarte, R. (2002). *O estado da arte das tecnologias para a convivência com as secas no Nordeste* (Vol. 6). Banco do Nordeste.

[4] Silva, A. M. da; Brito, L. P. de; Trandade, U. B.; Silva, V. C. da (2022). *OPERAÇÃO CARRO PIPA: Uma abordagem da atuação do Exército Brasileiro como mão amiga e integração nacional*. Mauá, SP.

[5] Rose, K., Eldridge, S., & Chapin, L. (2015). The internet of things: An overview. *The internet society (ISOC)*, *80*, 1-50.

[6] Li, S., Xu, L. D., & Zhao, S. (2015). The internet of things: a survey. *Information systems frontiers*, *17*, 243-259.

[7] Hussain, T., Muhammad, K., Ullah, A., Del Ser, J., Gandomi, A. H., Sajjad, M., ... & de Albuquerque, V. H. C. (2020). Multiview summarization and activity recognition meet edge computing in IoT environments. *IEEE Internet of Things Journal*, *8*(12), 9634-9644.

[8] Muhammad, K., Hussain, T., Rodrigues, J. J., Bellavista, P., de Macêdo, A. R. L., & de Albuquerque, V. H. C. (2020). Efficient and privacy preserving video transmission in 5G-enabled IoT surveillance networks: Current challenges and future directions. *IEEE Network*, *35*(2), 26-33.

[9] Hussain, T., Muhammad, K., Ding, W., Lloret, J., Baik, S. W., & de Albuquerque, V. H. C. (2021). A comprehensive survey of multi-view video summarization. *Pattern Recognition*, *109*, 107567.

[10] Hussain, T., Ullah, F. U. M., Khan, S. U., Ullah, A., Haroon, U., Muhammad, K., ... & de Albuquerque, V. H. C. (2022). Deep Learning Assists Surveillance Experts: Toward Video Data Prioritization. *IEEE Transactions on Industrial Informatics*.

[11] Hussain, T., Muhammad, K., Del Ser, J., Baik, S. W., & de Albuquerque, V. H. C. (2019). Intelligent embedded vision for summarization of multiview videos in IIoT. *IEEE Transactions on Industrial Informatics*, *16*(4), 2592-2602.

[12] Liu, A., Mahapatra, R. P., & Mayuri, A. V. R. (2021). Hybrid design for sports data visualization using AI and big data analytics. *Complex & Intelligent Systems*, 1-12.

[13] Mehmood, I., Ullah, A., Muhammad, K., Deng, D. J., Meng, W., Al-Turjman, F., ... & de Albuquerque, V. H. C. (2019). Efficient image recognition and retrieval on IoT-assisted energy-constrained platforms from big data repositories. *IEEE Internet of Things Journal*, *6*(6), 9246-9255.

[14] Zhang, J., Xu, C., Gao, Z., Rodrigues, J. J., & de Albuquerque, V. H. C. (2020). Industrial pervasive edge computing-based intelligence IoT for surveillance saliency detection. *IEEE Transactions on Industrial Informatics*, *17*(7), 5012-5020.

[15] Magaia, N., Fonseca, R., Muhammad, K., Segundo, A. H. F. N., Neto, A. V. L., & de Albuquerque, V. H. C. (2020). Industrial internet-of-things security enhanced with deep learning approaches for smart cities. *IEEE Internet of Things Journal*, *8*(8), 6393-6405.

[16] Muhammad, K., Del Ser, J., Magaia, N., Fonseca, R., Hussain, T., Gandomi, A. H., ... & de Albuquerque, V. H. C. (2022). Communication Technologies for Edge Learning and Inference: A Novel Framework, Open Issues, and Perspectives. *IEEE Network*.

[17] Silva, L., Magaia, N., Sousa, B., Kobusińska, A., Casimiro, A., Mavromoustakis, C. X., ... & De Albuquerque, V. H. C. (2021). Computing

paradigms in emerging vehicular environments: A review. *IEEE/CAA Journal of Automatica Sinica*, *8*(3), 491-511.

[18] Sodhro, A. H., Rodrigues, J. J., Pirbhulal, S., Zahid, N., de Macedo, A. R. L., & de Albuquerque, V. H. C. (2020). Link optimization in software defined IoV driven autonomous transportation system. *IEEE Transactions on Intelligent Transportation Systems*, *22*(6), 3511-3520.

[19] Sodhro, A. H., Luo, Z., Sodhro, G. H., Muzamal, M., Rodrigues, J. J., & De Albuquerque, V. H. C. (2019). Artificial Intelligence based QoS optimization for multimedia communication in IoV systems. *Future Generation Computer Systems*, *95*, 667-680.

[20] Allam, Z., & Dhunny, ZA (2019). On big data, artificial intelligence and smart cities. *Cities* , *89* , 80-91.

[21] Wang, S., Liu, X., Liu, S., Muhammad, K., Heidari, A. A., Del Ser, J., & de Albuquerque, V. H. C. (2021). Human Short Long-Term Cognitive Memory Mechanism for Visual Monitoring in IoT-Assisted Smart Cities. *IEEE Internet of Things Journal*, *9*(10), 7128-7139.

[22] Liu, S., Wang, S., Liu, X., Gandomi, A. H., Daneshmand, M., Muhammad, K., & De Albuquerque, V. H. C. (2021). Human memory update strategy: a multi-layer template update mechanism for remote visual monitoring. *IEEE Transactions on Multimedia*, *23*, 2188-2198.

[23] Moalla, I., Nabli, A., Bouzguenda, L., & Hammami, M. (2017). Data warehouse design approaches from social media: review and comparison. *Social Network Analysis and Mining*, *7*(1), 5.

[24] Li, X., Liu, H., Wang, W., Zheng, Y., Lv, H., & Lv, Z. (2022). Big data analysis of the internet of things in the digital twins of smart city based on deep learning. *Future Generation Computer Systems*, *128*, 167-177.

[25] Zygiaris, S. (2018). *Database management systems: A business-oriented approach using ORACLE, MySQL and MS access*. Emerald Group Publishing.

[26] Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine*, *17*(3), 37-37.

[27] Muhammad, K., Hussain, T., Del Ser, J., Palade, V., & De Albuquerque, V. H. C. (2019). DeepReS: A deep learning-based video summarization strategy for resource-constrained industrial surveillance scenarios. *IEEE Transactions on Industrial Informatics*, *16*(9), 5938-5947.