

Research Article

3D Modeling Technology of Mobile Big Data Application in the Electric Vehicle Condition Monitoring System

Ziyang Wang and Gangjun Li 

School of Intelligent Manufacturing, Chengdu Technological University, Chengdu, 610000 Sichuan, China

Correspondence should be addressed to Gangjun Li; tkzy@cdu.edu.cn

Received 21 May 2022; Revised 12 July 2022; Accepted 21 July 2022; Published 27 August 2022

Academic Editor: Chia-Huei Wu

Copyright © 2022 Ziyang Wang and Gangjun Li. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In recent years, the modern information society has entered the era of big data. Big data has rapidly developed into a popular category favored by academia and industry and has been widely used. This paper is aimed at studying the application of the 3D modeling technology of mobile big data to the monitoring of the state of electric vehicles. It proposes related concepts of big data algorithm and 3D modeling technology, introduces the trajectory matching algorithm based on mobile big data, and finally introduces the related concepts of electric vehicles and the design of the big data analysis platform for condition monitoring of substation equipment. On the basis of discussing the application of the three-dimensional modeling technology of mobile big data in the electric vehicle condition monitoring system, the experimental research was carried out with the big data in the condition monitoring of the intelligent substation. The experimental results of this paper showed that the storage overhead of the NJLS model was reduced by 34% to 40% compared with the conventional star schema, which could reduce the space overhead.

1. Introduction

Big data is the most popular direction of information science research nowadays. Relevant technology has gradually developed into a database system since the mid-1980s. It revolves around all the data and continuously excavates some data information with potential value, which has great application value. This is a dynamic and interactive process. With the rapid popularization of automobiles and the increase in the number of automobiles, energy consumption and environmental pollution are becoming increasingly serious. Electric vehicles are attracting more and more attention due to their low pollution, clean, and efficient characteristics. The electric vehicle drive system is mainly composed of a high-efficiency motor, a controller, and an energy storage device. The development goal of electric vehicles is zero emissions.

Compared to conventional cars, electric vehicles are still in their infancy. The working environment and the control technology of electric vehicles are more complex. If an electric car breaks down, it will not work properly and heavy electric vehicles will cause serious road accidents. Therefore,

mobile big data is needed to monitor the status of electric vehicles. The main purpose of monitoring this data is to provide early warning of impending motor failures, to perform preventive regular maintenance on possible future problems, and to provide a diagnostic plan. Therefore, it is of great practical significance to apply the 3D modeling technology of mobile big data to monitor the status of electric vehicles.

The innovations of this paper are as follows: (1) The use of big data to monitor the state of electric vehicles is innovative and practical. (2) The monitoring of the state of electric vehicles can effectively protect people's lives, health, and safety.

2. Related Work

With the advancement of economy and innovation, portable large information has continuously assumed a significant part and numerous researchers have done investigation on it. Qing-Chao et al. installed spark on the THREAD platform, built a more complete macroanalysis model based on mobile data, and used the monthly data of more than 29 million users for ETL and mining processing, proving that the model could

be applied to the mining of urban macrotraffic characteristics [1]. Mobile cloud computing (MCC) and IoT technology based on wireless network technology are developing rapidly. Stergiou and Psannis combined the above two technologies (i.e., MCC and IoT) with big data technologies to examine their common characteristics and found the advantages of MCC and IoT that could improve the use of big data applications [2]. In order to study how to mine valuable information in low-complexity multimedia big data, Guo et al. proposed a big data object detection method with a compressed measurement domain under a mobile distributed computing architecture [3]. In recent years, the research field of mobile big data has risen rapidly but somewhat fragmented. Xiang et al. aimed to provide a complete picture of this emerging field, build multidisciplinary bridges, and hopefully inspire future research [4]. Using big data analysis technology, Parwez et al. contributed in two ways. First, mobile network data (big data)—call detail records—was used to analyze the abnormal behavior of the mobile wireless network. Second, a neural network-based predictive model was trained using anomalous and non-anomaly data to highlight the impact of anomalies in the data while training/building an intelligent model [5]. Qiao et al. proposed a mobile big data framework, called FMBD, which provided massive data traffic collection, storage, processing, analysis, and management functions to cope with massive data traffic [6]. Hu and Yan solved the private dot product calculation problem for mobile big data applications, and the calculation efficiency was very ideal [7]. However, the shortcoming of these studies is the uncertainty of the data quality; the calculation and analysis of massive data are very complicated, so the construction of the model still needs to be improved.

3. Method of 3D Modeling Technology for Mobile Big Data

3.1. Big Data Algorithms

3.1.1. Data Mining and Machine Learning Framework. Data mining refers to the process of using certain algorithms to automatically search for information and laws with special relationships in a dataset from a large amount of data. Today's data not only is huge but also has incomplete and ambiguous characteristics [8]. Finding out the relevant patterns is the task of data mining. It has two main characteristics: one is that it can obtain valuable information in large-scale data. The second is that it can respond according to the changes in the data and adapt and collect the changes in the data in real time. With the rapid improvement of computer computing and data processing capabilities, more and more machine learning algorithms are applied in practice. The corresponding weight parameters are obtained through the training data, the accuracy is obtained by inputting the model with the test dataset, and the final output result is obtained by making a decision based on the experience generated by the historical training data [9]. As the technical support and theoretical basis of artificial intelligence, machine learning not only can process and analyze data to obtain laws and parameters quickly but can also classify and predict related data problems. In the current

trend of increasing data volume, machine learning is particularly important. Today, there are numerous frameworks, libraries, and tools available for machine learning, which facilitate the use of machine learning techniques.

As per Wikipedia's definition, huge information alludes to informational indexes that surpass the capacity, the board, and handling abilities of conventional programming or consume additional time than OK. With respect to idea of large information, it is for the most part accepted that there are 4Vs, to be specific volume, speed, assortment, and worth. At the present transformative phase, giving a compelling and exact meaning of large data is remarkably difficult. And when a new technical concept is proposed, it usually needs to go through a process [10].

The first is volume. This is the biggest feature that distinguishes big data from traditional data, that is, the amount of data is large. This feature is inseparable from the development of technology. First of all, the data storage capacity in the past was very limited. According to the development of Moore's law, the performance of hardware has been continuously improved and the price has gradually decreased, making large-scale data storage possible. In addition, the proliferation of technologies such as social networking, e-commerce, and the Internet of things has created a flood of data.

The second is velocity. It refers to the constant influx of data flow at an unprecedented rate and must be processed within an acceptable time frame. This is the main challenge faced by big data. Traditional data storage and processing methods simply cannot achieve the efficiency that can be used.

The third is variety. It refers to the variety of forms of data, mainly because of the abundance of data sources. The storage and analysis of these data are also two of the main challenges faced by big data.

3.1.2. Big Data Storage. In the earliest days, information was straightforwardly put away in records. Such a file storage method has many defects, cannot be stored for a long time, and has a small amount of storage. With the development of disseminated record frameworks, information can be put away as appropriated documents, which enjoys the accompanying benefits: It can oblige a lot of information [11]. The distributed file system represented by HDFS can store a large file in multiple machines, and let each machine store a part of the file to dissipate the pressure of a single machine. And redundant backup of data is supported. By default, HDFS saves three copies of each data and distributes them to different machines in the cluster, so that even if a machine in the cluster goes down or is completely destroyed, the data will not be lost. It has good scalability. When the amount of data increases gradually, the data storage capacity can be improved by adding cluster machines. There are three key components in HDFS, namely, DataNode, NameNode, and Clié. The DataNode sends the information of the data block to the NameNode [12], and the Clié is responsible for initiating read and write requests.

3.1.3. Correlation Analysis. Correlation analysis is an exploratory attempt before data analysis, which aims to explore the relationship and nature of input parameters and variables.

The results of correlation analysis can effectively guide what method to take in the next step. For instance, in displaying, the information overt repetitiveness of the model can be diminished by the consequences of relationship investigation, which is the essential work before information mining. During the time spent information demonstrating, there are many sorts of boundaries and the connection between every information is unique. All kinds of parameters are directly used as model input, which increases model redundancy and reduces model calculation efficiency. In order to simplify the model input and improve the calculation efficiency of the model, it is urgent to perform a correlation analysis on the original parameters and eliminate the variables with greater correlation [13].

The commonly used correlation analysis methods are as follows:

(1) Covariance and the covariance matrix

Covariance can be utilized to gauge the general mistake of two factors. In the event that it is positive, the two factors are decidedly corresponded; in any case, they are adversely related; assuming that the factors are free of one another, the covariance is 0.

$$\text{COV}(R, T) = \frac{\sum_{i=1}^n (R_i - \bar{R})(T_i - \bar{T})}{n-1}. \quad (1)$$

Correlation analysis is performed on the two groups of data in equation (1). When there are more than two groups of data, the covariance matrix needs to be used. Taking three sets of data as an example, the equation for calculating the covariance matrix is as follows:

$$\begin{aligned} D &= \text{COV}(R, R) \quad \text{COV}(R, T) \quad \text{COV}(R, L), \\ D &= \text{COV}(T, R) \quad \text{COV}(T, T) \quad \text{COV}(T, L), \\ D &= \text{COV}(L, R) \quad \text{COV}(L, T) \quad \text{COV}(L, L) \end{aligned} \quad (2)$$

(2) Correlation coefficient

The correlation coefficient is a statistical indicator of the degree of closeness between variables [14].

$$R_{RT} = \frac{S_{RT}}{S_R S_T}. \quad (3)$$

R is the sample correlation coefficient, S_{RT} is the sample covariance, S_R is the sample standard deviation of R , and S_T is the sample standard deviation of T

(3) Mutual information analysis

When $(R, T) \subset P(r, t)$, the mutual information between variables R and T is defined as follows:

$$\text{MI}(R, T) = G(R) + G(T) - G(R, T) \quad (4)$$

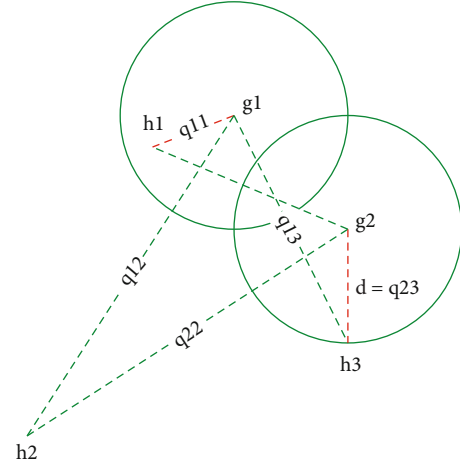


FIGURE 1: Hausdorff distance algorithm.

In equation (4), $G(R, T)$ is the joint entropy of variables X and Y ; $G(R)$ and $G(T)$ are the unconditional entropy of variables R and T , respectively. In order to facilitate the evaluation, the mutual information is normalized, which can be expressed as follows:

$$\text{NMI}(R, T) = 2 \frac{I(R, T)}{G(R) + G(T)}. \quad (5)$$

The value of $\text{NMI}(R, T)$ will reflect the magnitude of the correlation between the variables R and T .

Although the commonly used Pearson coefficient method and regression analysis method can effectively quantify the linear relationship between variables, they cannot characterize the correlation between nonlinear-related variables. In order to facilitate the numerical representation of the relationship between nonlinear correlated variables and realize quantitative statistics, the correlation can be analyzed by applying mutual information.

3.1.4. Research Status of 3D Imaging Technology. A 3D data model is an abstract-simulated representation of real things that connect the computer world and the real world. Three-dimensional modeling methods mainly include vector data structure, splicing, mixing, and analysis.

3.2. Basic Concept of Trajectory

3.2.1. Trajectory Data Classification. According to the different requirements for collecting trajectory information, the trajectory sampling method can be adjusted. Trajectory data can be roughly divided into three categories: time-based sampling trajectory data, location-based trajectory data sampling, and trigger event-based sampling trajectory data. This section will introduce these three types of trajectory data.

(1) *Sampling Trajectories Based on Time.* Trajectory-based temporal sampling, represented by equation (6), refers to discrete trajectory points preserved by sampling the

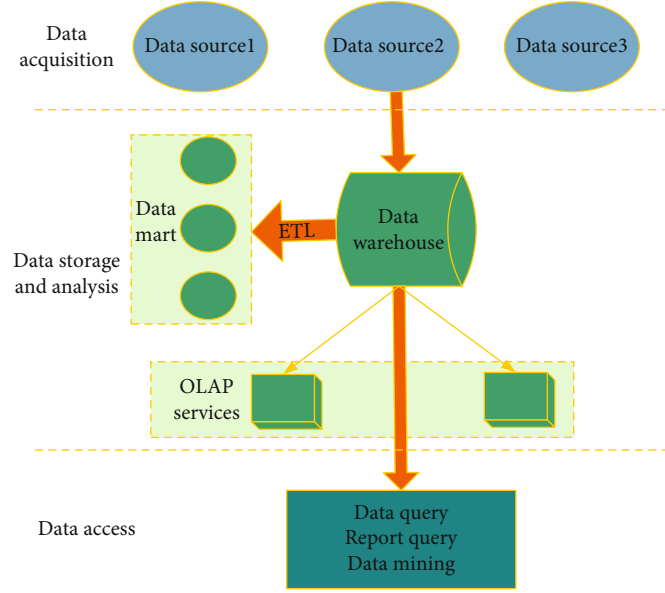


FIGURE 2: Architecture diagram of traditional substation equipment condition monitoring platform (reproduced from Zhang et al. 2018).

information of a moving object in the same time period and recording its position and time information [15].

$$\text{Traj} = \{ (\vec{s}_1, d_1), (\vec{s}_2, d_1 + \Delta d), \dots, (\vec{s}_v, d_1 + (n-1) \bullet \Delta d) \}. \quad (6)$$

(2) *Sampling Trajectory Based on Location.* A location-based sampling path is an independent trajectory point reserved for sampling data, and record its position, time, and other properties as the position of the moving object changes. It can be represented by equation (7) as follows:

$$\text{Traj} = \{ (\vec{s}_1, d_1), (\vec{s}_2, d_2), \dots, (\vec{s}_u, d_u), \dots, (\vec{s}_v, d_v) \}. \quad (7)$$

(3) *Sampling Trajectory Based on Trigger Events.* Based on the trigger event, the sampling path triggers the reception of sensor information through some specific behaviors of the moving object by recording the current position, time, and other attribute information of the moving object. For example, using mobile phones for daily calls, surfing the Internet, and swiping credit cards, citizen cards, etc. will trigger the work of some sensors such as surrounding base stations and card swiping machines, so that the corresponding data will be recorded and saved to the database. It can be represented by equation (8) as follows:

$$\text{Traj} = \{ (\vec{s}_1, d_1), (\vec{s}_2, d_2), \dots, (\vec{s}_u, d_u), \dots, (\vec{s}_v, d_v) \}. \quad (8)$$

Time sampling trajectories are generally used for vehicle GPS positioning, animal migration research, etc.; location sampling trajectories are generally used for individual travel and group migration research; trigger event sampling trajectories are often used for user check-in and user hotspot research.

When the trajectory points collected based on the time sampling method express trajectory information, it may cause the lack of important trajectory features, which is not conducive to trajectory analysis research. At the same time, a large amount of redundant trajectory information will also be collected, which increases the complexity of the later trajectory analysis. Most of the location-based sampling methods require manual screening, which not only increases the labor cost but also makes the trajectory data sorting and updating cycle longer, and the acquisition process is more complicated. The collection of trajectory points based on the trigger event sampling method can obtain a large amount of trajectory data, the data is automatically collected by the sensor, the sorting and updating cycle is short, and the sampled trajectory data can also better express the trajectory characteristics. However, this collection method also has problems, because data is collected only when the corresponding event is triggered, which may cause the situation that the complete trajectory cannot be expressed.

3.2.2. *Track Similarity Calculation.* Use the corresponding function to obtain the distance between trajectories and use it as the metric basis. The smaller the distance is, the higher the similarity between trajectories is. The main similarity measurement methods are as follows: Euclidean distance method, Hausdorff distance method, longest common subsequence method, edit distance method, and dynamic time warping method. These intertrack similarity measures are described in detail below.

(1) *Euclidean Distance Method.* Euclidean distance is the most commonly used method to measure the similarity of objects. For the same dimension, the Euclidean distance between two trajectory points is calculated and the Euclidean distance

between the trajectories can be obtained by adding the distances in each dimension.

Equation (9) is as follows:

$$T(G, H) = \sum_{u=1}^v \text{distance}(g_u, h_u), \quad (9)$$

$$\text{Distance}(g_u, h_u) = \sqrt{(g_{ua} - h_{ua})^2 + (g_{ub} - h_{ub})^2}.$$

(2) *Hausdorff Distance Method.* A common metric to measure the distance between trajectories is the Hausdorff distance method.

The main idea of the algorithm is that, given the target trajectory point set $G = \{g_1, g_2, \dots, g_m\}$ and the target trajectory point set $H = \{h_1, h_2, \dots, h_n\}$ to be matched, Hausdorff distance calculation equation (10) of the two trajectories is as follows:

$$D(G, H) = \max(d(G, H), d(H, G)). \quad (10)$$

The calculations for $d(G, H)$ and $d(H, G)$ are shown in equation (11) as follows:

$$d(G, H) = \max_{g_u \in G} \left(\min_{h_n \in H} q(g_u - h_n) \right), \quad (11)$$

$$d(H, G) = \max_{h_n \in H} \left(\min_{g_u \in G} q(h_n - g_u) \right).$$

Among them, 1 and 2 express the Euclidean distance between the trajectory points in trajectory point set G and trajectory point set H . Among them, 3 is the one-way Hausdorff distance between trajectory point set G and trajectory point set H , which is the minimum and maximum distances from the trajectory point in trajectory point set G to the trajectory point in trajectory point set H . It expresses the degree of the least similarity between trajectory point set G and trajectory point set H , and similarly, 4 can be obtained. Since the Hausdorff distance is directional, in general, 5 and 6 are not equal and the maximum value is selected as the final Hausdorff distance to express the degree of difference between trajectory point set G and trajectory point set H . That is, the distance from each point of trajectory point set G to each point of trajectory point set H is not greater than the Hausdorff distance, as shown in Figure 1.

(3) *Longest Common Subsequence Method.* The main idea of the longest common subsequence method is to find the longest common subsequence between two trajectories [16]. First of all, the definition of a subsequence is to ensure that the order of the original sequence remains unchanged and there is a new sequence obtained by selecting any element from the sequence. Thus, the subsequences may not necessarily be contiguous. Furthermore, through the length of the longest common subsequence, it is converted into a corresponding distance metric to represent the similarity between the two tracks, as shown in equation (12).

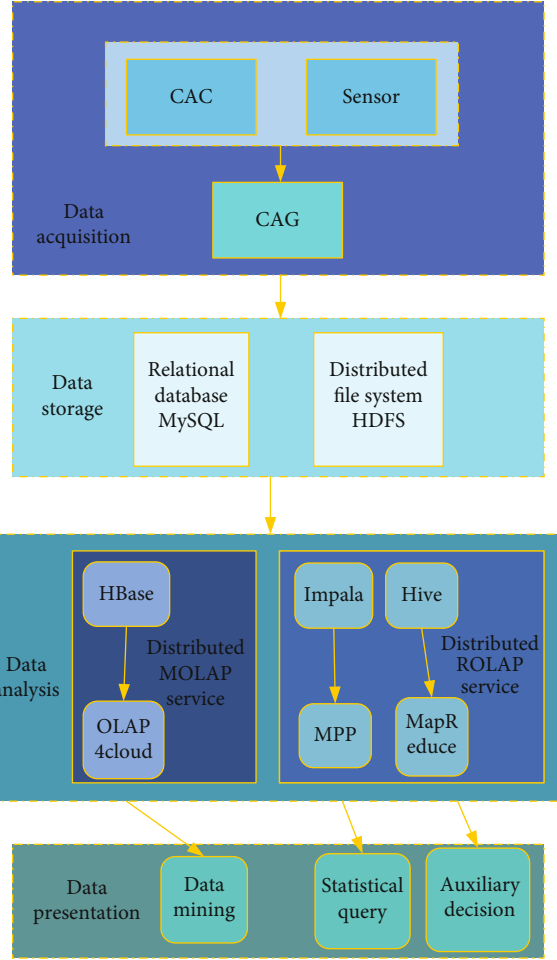


FIGURE 3: Architecture diagram of substation equipment condition monitoring platform under big data (reproduced from Zhang et al. 2018).

$$\text{LCS}(G, H) = \begin{cases} 0, & \text{if } n = v = 0, \\ \text{LCS}(\text{rest}(G), \text{rest}(H)) + 1, & \text{if } |g_{1a} - h_{1a}| \leq \alpha \text{ and } |g_{1b} - h_{1b}| \leq \beta, \\ \max\{\text{LCS}(\text{rest}(G), H), \text{LCS}(G, \text{rest}(H))\}, & \text{otherwise,} \end{cases}$$

$$Q_{\text{LCS}}(G, H) = 1 - \frac{\text{LCS}(G, H)}{\min(n, v)}. \quad (12)$$

Among them, G and H are two trajectories and the numbers of trajectory points of the two trajectories are n and v . 1 is the length of the longest common subsequence between G and H , and 2 and 3 represent the thresholds in the horizontal and vertical directions, respectively. When the difference between the two track points in the corresponding directions is less than 4 and 5, it means that the two track points are similar and the value of 6 is increased by 1. Converting the longest common subsequence length into a distance metric, the relevant equation (13) is as follows:

$$Q_{\text{LCS}}(G, H) = 1 - \frac{\text{LCS}(G, H)}{\min(n, v)}. \quad (13)$$

$\min(n, v)$ represents the smaller of n and v .

TABLE 1: Comparison of characteristics of main types of new energy vehicles in China.

Type	Hybrid electric vehicle	Blade electric vehicles	Fuel cell vehicle
Driving mode	Internal combustion engine + motor drive	Motor drive	Motor drive
Energy system	Battery, internal combustion engine power generation unit	Battery	Fuel cell
Use energy	Fuel, electric energy	Electric energy	Hydrogen, formaldehyde or gasoline, etc.
Infrastructure	Gas station, grid charging facility (optional)	Grid charging facilities	Storage and loading equipment of fuel cell raw materials
Main features	The emission is very low, the energy conversion efficiency is related to the mixing degree and mixing mode, the driving range is long, the vehicle function is not limited, and part of it depends on crude oil, so the structure is complex	Zero emission, high energy conversion efficiency, short driving range, limited vehicle function, independent of crude oil, relatively simple structure	Zero emission or ultralow emission, high energy conversion efficiency, long driving range, unlimited vehicle functions, can not rely on crude oil, poor adaptation to the working environment
Key points of technological development	Multi-energy management, optimal control	Storage battery and its management, charging facilities	Fuel cell life and cost, hydrogen storage, hydrogenation facilities

The Euclidean distance algorithm has a simple structure and is easy to use. However, the algorithm requires the trajectory to have the same number of trajectory points, which causes a relatively large limitation for the input of experimental data, and does not consider other factors in the process, because the Euclidean distance is very sensitive to noise. When there is noise in the trajectories, it is easy to affect the similarity judgment between trajectories and affect the accuracy of the experimental results.

The Hausdorff distance method is more widely used in practical scenarios and also has higher efficiency and accuracy. However, this method is sensitive to noise track points. Even if there are a small number of noise points, it will have a significant impact on the calculation of similarity distance. Therefore, when using this method, it is necessary to pay attention to the removal of track noise points.

Due to the discontinuity of subsequences, when noise occurs in the trajectory, it can be avoided as much as possible. Therefore, the longest common subsequence algorithm has higher robustness and accuracy. However, due to the huge amount of data points in the trajectory, the algorithm requires a large amount of calculation and is not efficient. Even if the idea of dynamic programming is used, it still consumes a lot of computing time. Therefore, this method is less used in projects with massive data.

3.3. Design of Big Data Platform for Condition Monitoring of Substation Equipment

3.3.1. Traditional Substation Equipment Condition Monitoring Platform Architecture. The information procurement layer is the establishment for laying out the substation condition

checking information stage. The recorded information sources and substation gear condition observing information in the conventional social dataset are transferred to the information stockroom through ETL (extraction, change, cleaning, and stacking). ETL will extricate, clean, change, and burden the information dispersed in different business frameworks (like security creation, projection of the board, and condition checking frameworks), with the goal that this essential information becomes top notch and significant information for brilliant substations. Its consistent construction is displayed in Figure 2 [17].

3.3.2. Design of the Condition Monitoring Platform for Substation Equipment under Big Data. The data acquisition layer mainly collects data of substation equipment through CAC (state access controller), sensors, etc. and transmits it to CAG (state access gateway machine) in the form of Web service. According to the advantages of the distributed file system HDFS and relational database MySQL in various fields, the data storage layer will integrate HDFS and MySQL and play their respective roles in their areas of expertise. Condition monitoring big data with unified specifications is stored in the distributed file system HDFS. MySQL is mainly used to store various model information of substation equipment condition monitoring and manage Hive metadata. The tables, fields, and spacers created by Hive will be stored in MySQL. While performing data operations, the MySQL engine needs to be started to verify the existence of metadata. Impala can share metadata information with Hive.

Most of the traditional solutions use conventional data storage and analysis methods, resulting in poor system scalability and high cost, and cannot meet the requirements of

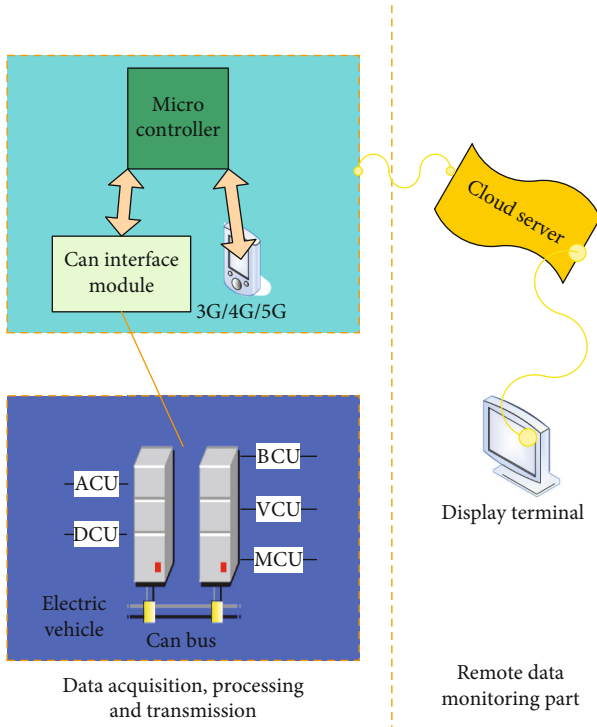


FIGURE 4: Overall structure diagram of electric vehicle motor status and fault remote monitoring system.

substations for complex analysis and trend prediction of the collected monitoring data. The integration of distributed data storage technology and big data analysis technology into the condition monitoring data platform brings a new idea to the data storage and analysis of condition monitoring. Considering the immense measure of information of substation hardware, it is important to relocate the information of conventional the social dataset (like MySQL) to nonsocial dataset. Figure 3 shows the design of the substation hardware condition checking stage under enormous information [18].

3.4. Development Trend of Electric Vehicles

3.4.1. *Characteristics of New Energy Vehicles.* The characteristics of the main types of new energy vehicles in China are compared as in Table 1.

3.4.2. *System Scheme Design.* From the analysis of functional requirements, it can be seen that the system needs to have the function of remote monitoring of electric vehicle motor status and faults: data acquisition function, data processing function, data remote transmission function, and remote monitoring data function. The first three of these four functions need to be realized on the vehicle end, and the remote monitoring part is realized on the cloud server end and the display terminal. The overall structure diagram of the system composed of each function is shown in Figure 4.

The system design scheme is given as follows:

(1) Data acquisition function

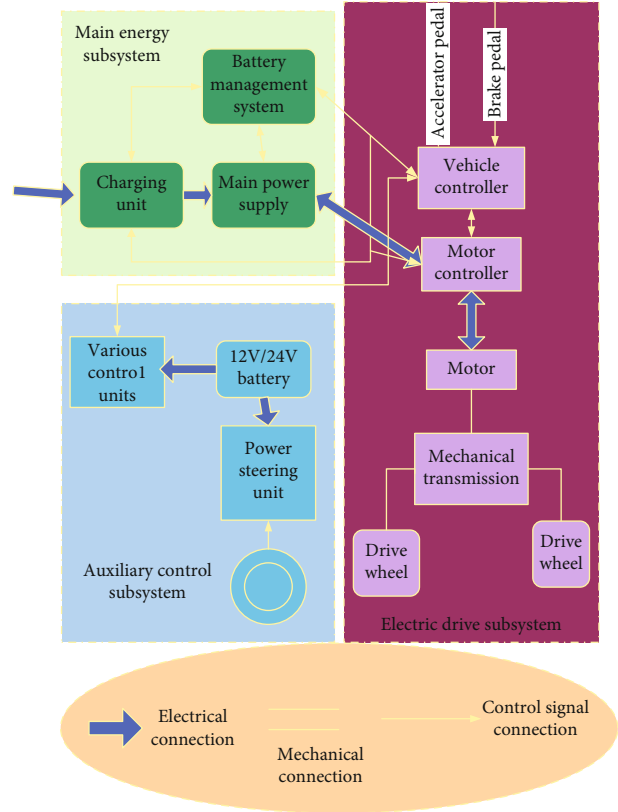


FIGURE 5: Basic structure of electric vehicles.

The data acquisition part of the system can be connected to the CAN bus as a node device to interpret the communication protocol of the application layer and obtain the data required by the system [19].

(2) Remote monitoring data function

The data monitoring function needs to be realized by building a monitoring system that displays the data engine. Therefore, the mobile phone application is selected as the monitoring interface for displaying the data engine.

3.4.3. *Classification and Application Status of the Electric Vehicle Bus System.* The structure of electric vehicles is very different from traditional vehicles, but the communication category of the bus is still the same. Figure 5 shows the basic structure of an electric vehicle (in series). It can be seen that both the vehicle controller and the motor controller are connected to the high-speed CAN bus of the system [1].

The vehicle controller (VCU) is designed with a modular circuit consisting of a high-performance microcontroller, CAN and peripheral circuits, and special sensor acquisition circuits. The VCU receives various data on the CAN bus (control signals such as an accelerator gear and vehicle status such as vehicle speed, etc.), stores the required information after analyzing and judging, and sends control commands to the corresponding control system. The motor controller (MCU) is designed based on a high-performance main

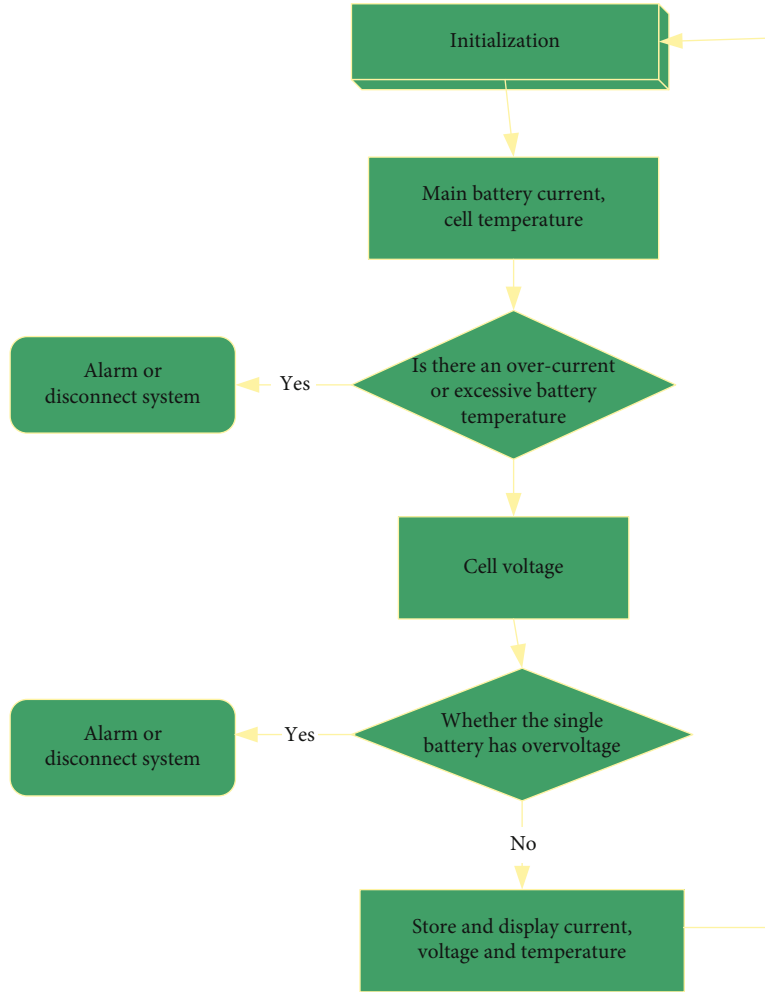


FIGURE 6: Main control flow chart.

TABLE 2: Experimental dataset size.

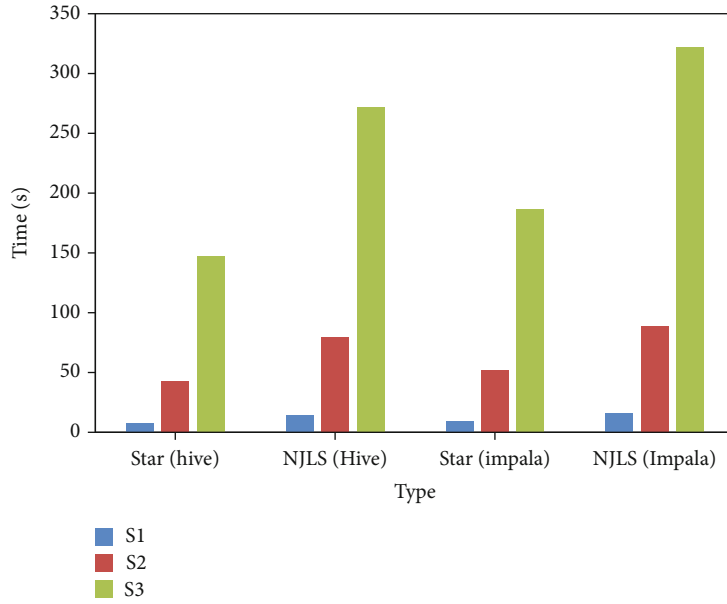
Dimension	S_1	S_2	S_3
Date	1 month	1 month	1 month
Equipment	2 kinds of equipment	4 kinds of equipment	4 kinds of equipment
Monitor type	4 types	7 types	7 types
Size ($S_1S_2S_3$)	1.6×10^7 line	10^8 line	4×10^8 line

processor and is the unique core control electronic unit for electric vehicles. Therefore, it can receive the control instructions sent by the VCU, can control the motor to run according to the control instructions of the VCU, and also has the functions of motor fault diagnosis, protection, and storage [20].

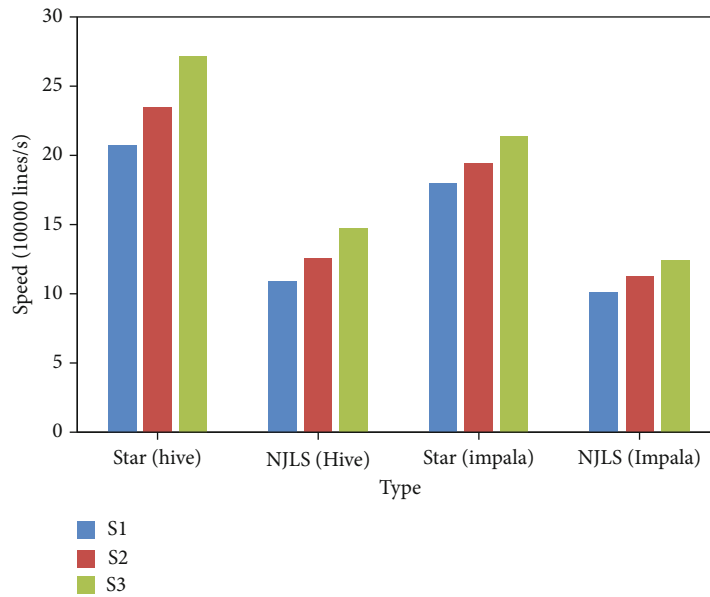
As can be seen, the engine speed, torque, voltage, temperature, and other data collected by this system come from the vehicle controller and the engine controller, which communicate with each other through the high-speed carbon canister. Therefore, the data acquisition and transmission module should also be connected to the high-speed CAN bus of the electric vehicle, and then, collect the data required by the system through the CAN bus. The data acquisition

hardware circuit is composed of the CAN transceiver, microcontroller containing CAN controller, and its peripheral circuits. This hardware circuit can be directly added to the CAN bus of electric vehicles as a CAN node device. Through the software realization of the data acquisition circuit function by using C language, the data acquisition module can receive the data frame composed of the motor speed and current and other data on the CAN bus of the electric vehicle and realize the data acquisition function.

3.4.4. Introduction to the Principle of CAN Bus. CAN (controller area network) is a serial communication protocol of the International Organization for Standardization (ISO). The main



(a) Data loading time (s)



(b) Data loading speed

FIGURE 7: Data loading time and speed.

TABLE 3: Run times for rollup operations.

	S ₁	S ₂ Hive	S ₃	S ₁	S ₂ Impala	S ₃
Star	58.6	77.9	115	18.86	25.66	37.53
NJLS	25.6	34.16	46.57	8.22	12.88	18.55

control process is a continuous loop program for real-time monitoring of battery voltage, current, temperature, and other information, as well as surge and high-temperature alarms. The real-time tracking information is stored for easy collection by the host computer [21]. The flowchart is shown in Figure 6.

4. Big Data Monitoring in Smart Substation Status Experiment

This paper discussed the application of the 3D modeling technology of mobile big data in the electric vehicle condition monitoring system and proposed an experimental analysis of the data model of electric equipment condition monitoring based on connectionless hierarchical coding.

The Not Join Level-encoding Schema is abbreviated as NJLS. NJLS applies this model to the condition monitoring of substation equipment, which brings a different method to condition monitoring big data storage and big data analysis [22].

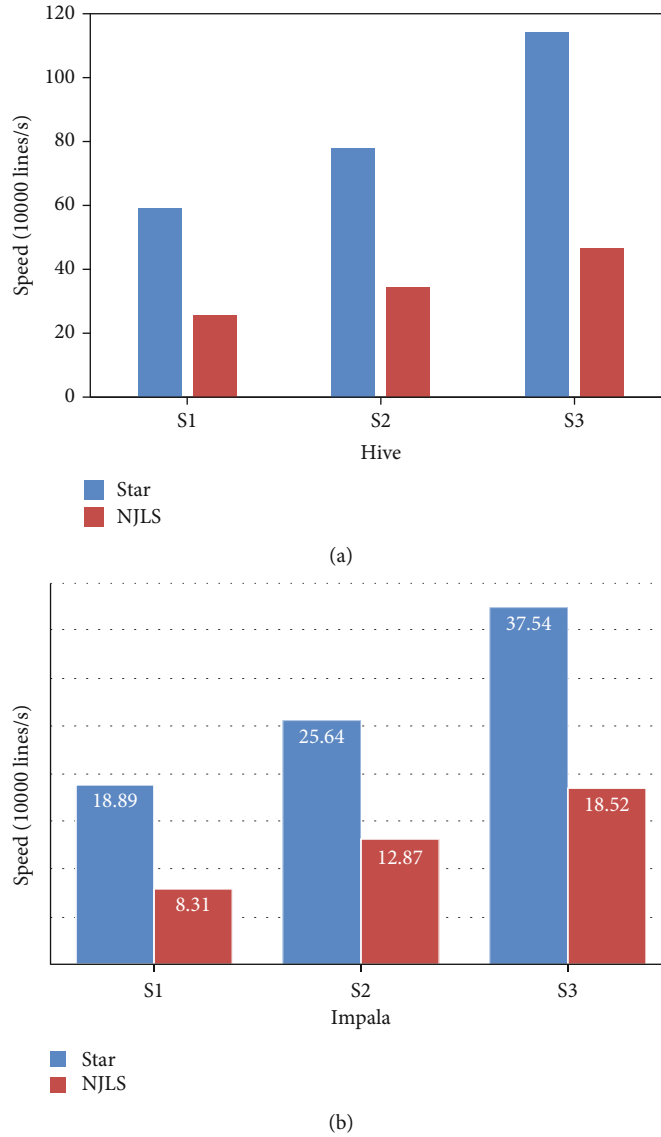


FIGURE 8: Variation of rollup performance with dataset.

4.1. Monitoring Data Preparation

4.1.1. Data Preparation. Three sets of monitoring large datasets ($S_1S_2S_3$) were used in this section, and three test groups participated in a comprehensive evaluation of the model, including data loading monitoring, aggregation operations, and overall storage costs. The individual datasets are shown in Table 2.

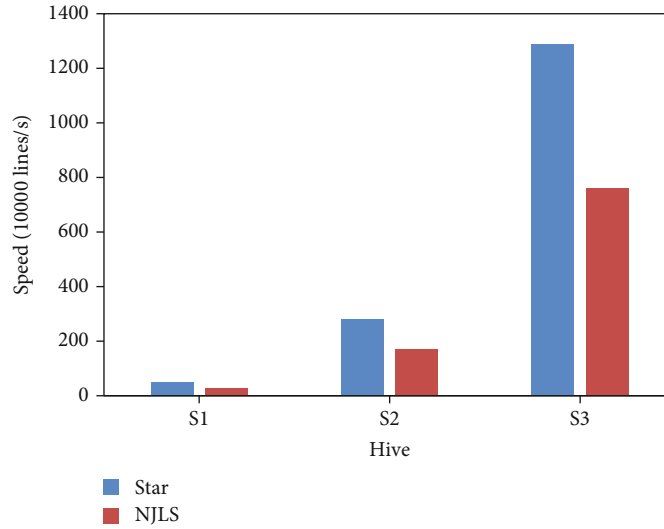
4.1.2. Monitoring Data Loading. In the test to compare the data charging performance, this work used single-cable data to charge Hive and Impala. Figure 7 shows the loading time and loading speed of the monitoring data loading experiment.

The following conclusions can be drawn from the data loading speed in Figure 7(b):

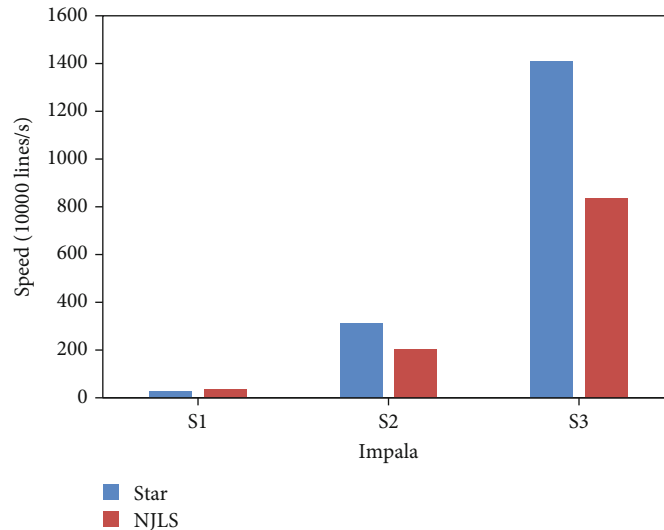
- (1) Compared with the conventional star model, the NJLS model for condition monitoring of substations

has no advantages in loading. This is due to the need for preprocessing and hierarchical encoding settings for the NJLS model state monitoring dataset, and the monitoring data loading speed is about 42% of the conventional model

- (2) The data loading speed of Hive and Impala is relatively stable and will not decrease significantly with the doubling of condition monitoring data. Because when loading data, the data to be loaded is stored in HDFS and the metadata correction is completed at the same time
- (3) Figure 7(b) shows that the loading speed of Impala monitoring data is slightly slower than that of Hive. Due to the limitations of the data file format supported by Impala, Hive needs to be used to load the data first and then perform other operations



(a)



(b)

FIGURE 9: Dataset storage overhead comparison.

4.2. Operation Details. Table 3 shows the running season of the rollup activity between the NJLS model and the star construction with various checking dataset sizes. Figure 8 shows the presentation patterns of NJLS and star blueprint in Hive and Impala frameworks with various condition observing datasets.

Combining Table 3 and Figure 8, it tends to be seen that (1) the roll-up execution season of the NJLS model is 40% to 49% more limited than that of the star model and (2) the running season of Impala's roll-up activity for each condition checking informational index is more limited than that of Hive.

4.3. Storage Overhead. This group of experiments mainly analyzed the storage overhead of the NJLS model and the conventional star model from the aspect of data storage and proved through theory and experiments that the storage overhead of the NJLS model was still smaller than that of the star model even under large-scale condition monitoring data volume.

In this set of experiments, the number of backups was the default value of 3 for Hadoop. Figure 9 compares the storage overhead of the NJLS model and the regular star model in different condition monitoring datasets of Hive and Impala.

Taking the S_2 condition observing informational collection in Figure 9 for instance, it tends to be seen that (1) the capacity upward of the NJLS model is decreased by 34% to 40% contrasting and the ordinary star model and (2) among numerous huge information investigation frameworks, Hive, Impala stockpiling upward is little.

5. Discussion

With the widespread application of wireless communication technology and sensor technology, the coverage of communication equipment is expanding [23]. Base stations set up in cities can receive radio waves from devices with moving objects and convert them into mobile data [24].

With the quick advancement of imaging innovation and Internet innovation, varying backgrounds have advanced higher necessities for the instinct and realness of data show and the examination of 3D imaging innovation has likewise gotten increasingly more consideration. The genuine working information of the electric vehicle motor is vital for the support and innovative work of the electric vehicle motor, and guaranteeing the security of the electric vehicle is additionally vital [25].

The 3D scene constructed by 3D imaging technology is not only realistic but also convenient for spatial analysis. With the explosive growth of the number of electric vehicles in China, the technical core power battery of electric vehicles has gradually attracted more attention. The application basis for the development of the electric vehicle industry is the data analysis of the operation of the power battery. Therefore, online monitoring and evaluation of electric vehicle power batteries are of great significance.

6. Conclusions

Combined with people's expectations for electric vehicles and the development trend of the electric vehicle industry, this paper used Hadoop technology to build a substation equipment condition monitoring experimental platform and briefly described the construction process of the platform. The NJLS-based substation condition monitoring data model proposed in this paper was compared and analyzed in Hive and Impala from three aspects of monitoring data loading, summary operation, and storage cost. The test results showed that the NJLS-based substation condition monitoring data model proposed in this paper could solve the connection operation problem distributed in multiple troublesome tables in ROLAP. So, it is more suitable for performing big data analysis on large-scale and distributed clusters.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Acknowledgments

This work was supported by the Research on Design Theory and Control Technology of High-Speed, Light-Load and Limited-dof Robots for Electronic Information Manufacturing and Light Industrial Production (Project approval number: 2018JY0116; project name: application of 3D modeling technology of mobile big data in the state monitoring system of electric vehicles).

References

- [1] S. Qing-Chao, D. Honghui, J. Li-Min, and Z. Hui, "Modeling the macro-characteristic of city passenger flow based on mobile big data," *Boletin Tecnico/Technical Bulletin*, vol. 55, no. 17, pp. 847–854, 2017.
- [2] C. Stergiou and K. E. Psannis, "Recent advances delivered by mobile cloud computing and internet of things for big data applications: a survey," *International Journal of Network Management*, vol. 27, no. 3, pp. 1–12, 2017.
- [3] J. Guo, B. Song, F. R. Yu, Z. Yan, and L. T. Yang, "Object detection among multimedia big data in the compressive measurement domain under mobile distributed architecture," *Future Generation Computer Systems*, vol. 76, pp. 519–527, 2017.
- [4] C. Xiang, L. Fang, X. Hong, and L. Yang, "Exploiting mobile big data: sources, features, and applications," *IEEE Network*, vol. 31, no. 1, pp. 72–79, 2017.
- [5] M. S. Parwez, D. B. Rawat, and M. Garuba, "Big data analytics for user-activity analysis and user-anomaly detection in mobile wireless network," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 4, pp. 2058–2065, 2017.
- [6] Y. Qiao, Z. Xing, Z. M. Fadlullah, J. Yang, and N. Kato, "Characterizing flow, application, and user behavior in mobile networks: a framework for mobile big data," *IEEE Wireless Communications*, vol. 25, no. 1, pp. 40–49, 2018.
- [7] C. Hu and H. Yan, "Efficient privacy-preserving dot-product computation for mobile big data," *IET Communications*, vol. 11, no. 5, pp. 704–712, 2017.
- [8] Z. Zhou, "Research on improving intelligent inspection efficiency of substation based on big data analysis," in *2020 5th Asia-Pacific Conference on Intelligent Robot Systems (ACIRS)*, Singapore, 2020.
- [9] S. Srinivasan, "Mobile big data: a roadmap from models to technologies," *Computing Reviews*, vol. 60, no. 8, pp. 314–314, 2019.
- [10] A. Enayet, M. A. Razzaque, M. M. Hassan, A. Alamri, and G. Fortino, "A mobility-aware optimal resource allocation architecture for big data task execution on mobile cloud in smart cities," *IEEE Communications Magazine*, vol. 56, no. 2, pp. 110–117, 2018.
- [11] F. Kong and X. Lin, "The method and application of big data mining for mobile trajectory of taxi based on MapReduce," *Cluster Computing*, vol. 6, pp. 1–8, 2018.
- [12] A. W. Yusuf-Asaju, Z. M. Dahalin, and A. Ta'A, "Framework for modelling mobile network quality of experience through big data analytics approach," *Journal of Information and Communication Technology*, vol. 17, no. 1, pp. 79–113, 2018.
- [13] J. Lee and L. Sangeun, "Using the mobile big data for the smart river space management : data validation and water-friendly space indicators," *Planning Review*, vol. 101, no. null, pp. 3–18, 2019.
- [14] Q. Jia, "Research on mobile learning in a teaching information service system based on a big data driven environment," *Education and Information Technologies*, vol. 26, no. 5, pp. 6183–6201, 2021.
- [15] K. K. Lwin, Y. Sekimoto, W. Takeuchi, and Institute of Industrial Science, The University of Tokyo 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan, "Development of GIS integrated big data research toolbox (BigGIS-RTX) for mobile CDR data processing in disasters management," *Journal of Disaster Research*, vol. 13, no. 2, pp. 380–386, 2018.
- [16] R. Gorli, "Interlinking OF IoT, big data, smart mobile app with smart garbage monitoring," *International Journal of Computerences & Engineering*, vol. 5, no. 1, pp. 70-71, 2017.

- [17] F. Luoyang and C. Xiang, "Location privacy in mobile big data: user identifiability via habitat region representation," *Networks*, vol. 3, no. 3, pp. 31–38, 2018.
- [18] P. Hui, Y. Li, J. Ott, S. Uhlig, B. Han, and K. Tan, "Mobile big data for urban analytics," *IEEE Communications Magazine*, vol. 56, no. 11, pp. 12–12, 2018.
- [19] X. Cheng, L. Fang, and L. Yang, "Mobile big data based network intelligence," *IEEE Internet of Things Journal*, vol. 5, no. 6, pp. 4365–4379, 2018.
- [20] G. Xu, Y. Bai, Q. Pan, Q. Huang, and Y. Yang, "Data verification tasks scheduling based on dynamic resource allocation in mobile big data storage," *Computer Networks*, vol. 126, no. oct. 24, pp. 246–255, 2017.
- [21] S. Wan, J. Lu, P. Fan, and K. B. Letaief, "Toward big data processing in IoT: path planning and resource management of UAV Base stations in mobile-edge computing system," *IEEE Internet of Things Journal*, vol. 7, no. 7, pp. 5995–6009, 2020.
- [22] J. Antonio, M. Garcia, and P. Toril, "Big data analytics for automated QoE management in mobile networks," *IEEE Communications Magazine*, vol. 57, no. 8, pp. 91–97, 2019.
- [23] Y. Sun, H. Song, A. J. Jara, and R. Bie, "Internet of things and big data analytics for smart and connected communities," *IEEE Access*, vol. 4, pp. 766–773, 2016.
- [24] C. Li, H. J. Yang, F. Sun, J. M. Cioffi, and L. Yang, "Multiuser overhearing for cooperative two-way multiantenna relays," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 5, pp. 3796–3802, 2016.
- [25] L. Jian Jian and W. Pan, "Three-dimensional landscape rendering and landscape spatial distribution of traditional villages based on big data information system," *Mobile Information Systems*, vol. 2022, 13 pages, 2022.