Classification of Urban Functional Areas From Remote Sensing Images and Time-Series User Behavior Data

Chen Chen, Jining Yan, Lizhe Wang, *Fellow, IEEE*, Dong Liang, and Wanfeng Zhang

***Abstract*—Urbanization is accelerating at a rapid rate, which has introduced many challenges, especially in the field of urban planning. Under the backdrop of global urbanization, some cities are particularly vulnerable to climate change and natural disasters that are influenced by unplanned urban expansion. Rational planning of urban functional areas needs to be strengthened to improve the scientific approach of urban planning and urbanization. In this study, the classification of urban functional areas based on dual-modal data (i.e., remote sensing image and user behavior data) was implemented using machine learning (ML) algorithms. After the set test, the classification accuracy of urban functional areas reached 82.45%. Through analysis, it could be concluded that the use of data of two modalities achieved a higher classification accuracy than that achieved by using data of a single modality. The data of the two modalities complement each other, and the use of ML algorithms to train such data can yield good results.**

***Index Terms*—Classification of urban functional areas, convolutional neural network (CNN), light gradient boosting machine (LightGBM), multimodal data.**

I. Introduction

THE United Nations has proposed the 2030 Agenda for Sustainable Development to achieve sustainable development goals (SDGs) [1] in three dimensions: the society, economy, and environment. Urbanization plays a key role in each of these dimensions. Cities present a challenging and inspiring scope for achieving the SDGs, and urban planning is the key factor in realizing sustainable construction [2], [3]. With the development of the social economy, the influence of spatial pattern of functional zones in cities on the efficiency of life is increasing. To cater to the various needs of the residents of a city, the geographical area of the city is divided on the basis of the functions of the city to form a development model with different functions in different areas [4], [5]. Identifying and classifying the functional areas of cities can improve the scientific nature of urban planning and promote the construction of new cities [6]–[8].

In recent years, many studies on the functional zones of cities have been conducted. Chen *et al.* used building-level social media data and a dynamic time warping distance based *k*-medoids method to group buildings with similar social media activities into functional areas [9]. Xing and Meng adopted the random forest (RF) method to process and analyze the landscape and socioeconomic characteristics data for classifying urban functional areas [10]. Arnaud used remote sensing images, three-dimensional city models, and other data to classify urban-structure types using support vector machine methods [11].

Remote sensing images are rich in information and have played a vital role in various studies [12]–[15]. However, they only reflect the actual characteristics of the surface features [16], [17]. Because of the lack of semantic information regarding the functions of geographic regions, pure remote sensing images have certain limitations with respect to the classification of functional areas [18], [19]. The exploration of human social activities by advanced information technologies has demonstrated that the travel and habitual activities of people are highly regular [20]. Furthermore, human behavior and urban spatial structures have been found to be closely related [21], [22]. Therefore, the travel characteristics of people can be used as an important basis for urban functional zoning [4], [23], [24]. However, because functional areas are defined by a combination of human activities and urban regional space, simple human activity characteristics cannot be used to explore the actual distribution of functional areas [25].

In this study, remote sensing image data and online user behavior data are used to classify urban functional areas. The user behavior statistical data indicate that the travel direction and duration of a crowd are related to the type of functional area. The core functions of different areas of a city can be judged by the regularity in behavior of the crowd. For remote sensing image data, a convolutional neural network (CNN) is used as an algorithm for image recognition to obtain the probabilistic features of each category. This is because a CNN enables good image recognition and classification and can accurately extract

Manuscript received September 14, 2020; revised November 6, 2020 and December 6, 2020; accepted December 9, 2020. Date of publication December 14, 2020; date of current version January 6, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant U1711266, Grant 41701468 and in part by the Fundamental Research Funds for the Central Universities, China University of Geosciences (Wuhan) under Grant 1910491B18. *(Corresponding author: Jining Yan.)*

Chen Chen, Jining Yan, and Lizhe Wang are with the School of Computer Science, China University of Geosciences, Wuhan 430074, China (e-mail: cc676190@163.com; yanjn@cug.edu.cn; Lizhe.Wang@gmail.com).

Dong Liang is with the Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China, and also with the University of Chinese Academy of Sciences, Chinese Academy of Sciences, Beijing 100094, China (e-mail: liangdong@radi.ac.cn).

Wanfeng Zhang is with the Key Laboratory of Space Utilization, Technology and Engineering Center for Space Utilization, Chinese Academy of Sciences, Beijing 100094, China (e-mail: wfzhang@csu.ac.cn).

Digital Object Identifier 10.1109/JSTARS.2020.3044250

image characteristics, such as color, texture, shape, and spatial relationships [26]. Finally, light gradient boosting machine (LightGBM) is used to combine the processed data of the two modalities for the final classification. LightGBM is a type of ensemble learning method. Ensemble learning can complete learning tasks by constructing and combining multiple learners, which can achieve significantly better generalization performance than that achieved with a single learner [27].

With the dataset and algorithm used in this study, after five-fold cross-validation, a model with an accuracy of 82.45% is obtained. The results show that the classification using the combined data of the two modalities achieves significant improvement over the classifications using the two modalities separately, indicating that the data of the two modalities complement each other and offer certain advantages in functional area classification. In addition, a comparison between LightGBM and other algorithms can show that the former has certain advantages in terms of classification accuracy and training time. The results of this study show that the proposed method is effective and feasible. Therefore, with its definite theoretical and practical value, the proposed method is suitable for use in urban functional areas, employing multimodal data.

The two major contributions of this study are as follows.

1) Multimodal data were used. Data from different modalities complement each other and offer better classification than that offered by single-modal data.
2) The LightGBM machine learning (ML) algorithm was used, which has higher classification accuracy and shorter training time than those of other ML algorithms.

II. Related Work

Remote sensing images are effective in reflecting the regional characteristics of land, which significantly improves the ability of remote sensing in the research on the changes in urban land use and cover [28]. Zhou *et al.* used the SO-CNN method to classify urban functional areas based on remote sensing images. They found that it is highly significant in the study of small-scale functional spatial structures [29]. Cheng *et al.* proposed discriminative CNNs to classify remote sensing image scenes, which solved the problems of within-class diversity and between-class similarity. This method improved the performance of remote sensing image scene classification [30]. In addition, remote sensing images are often used in region segmentation (i.e., spatial division of functional regions in an image). Zhang *et al.* used VHR satellite images to perform multiscale geological segmentation of the city and effectively divided the functional areas of the city [31]. The remote sensing image reflects the actual features of the ground objects and lacks semantic information related to human activities. Therefore, a pure remote sensing image is not ideal for identifying functional areas.

Social software are being widely applied, and the generated data are also increasing significantly. People extract more and more effective information from a large amount of data. The use of location-related information generated by social networks to study the behavior of people has become a major trend [24]. The behavior of people reflected in social network data improve the classification of urban functional areas. Zhi *et al.* used social media sign-in data based on the low rank approximation model to study the regular behavior of people. They identified and classified functional areas based on the correlation between time and space in the city [32]. Toole *et al.* conducted research on urban land use classification based on human spatiotemporal activities and big data conditions [33]. Han *et al.* used bus smart card data to obtain commutation information and identify functional areas in Beijing [34]. Yuan *et al.* used taxi global positioning system (GPS) information and city point-of-interest (POI) data to extract features through data mining. They studied different functions undertaken in different areas of the city [35]. Yao *et al.* proposed a method based on the Word2Vec model to extract semantic information from POI data for classifying functional areas [36].

TABLE I
Statistics of Various Types of Data

Functional Area	Number
Residential area	9542
School	7538
Industrial park	3590
Railway station	1358
Airport	3464
Park	5507
Shopping area	3517
Administrative district	2617
Hospital	2867

ML is a method of imitating the human learning behavior. Through automatic adjustment, it can judge the current learning situation and make improvements without relying on manpower to perform autonomous learning. ML generates models by analyzing more complex data automatically, quickly, and accurately, with high recognition accuracy and low cost [37]. Owing to its favorable characteristics, ML plays a key role in various fields, including engineering, medicine, and geosciences. Wang *et al.* detected transportation modes based on the LightGBM classifier from the GPS trajectory data [27]. Chen *et al.* used the LightGBM algorithm to predict the role of proteins for realizing biological applications [38]. Ju *et al.* used a model combining a CNN and LightGBM to study wind power forecasting [39]. Sun *et al.* adopted the LightGBM algorithm to forecast the price trend (falling or not falling) of the cryptocurrency market [40].

III. Datasets

The dataset used in this study comprises multimodal data. Remote sensing data are represented in the form of images, and user behaviors are represented in the form of texts. The data include nine functional areas: residential area, school, industrial park, railway station, airport, park, shopping area, administrative district, and hospital. They are expressed by providing the time information of several user visits corresponding to the scene of the remote-sensing image for each area. We evaluated the amount of data contained in each functional area, and the statistical results are presented in Table I. The first column lists each functional area, and the second column lists the number of files

Fig. 1. Remote sensing images of various functional areas. (a) Residential area. (b) School. (c) Industrial park. (d) Railway station. (e) Airport. (f) Park. (g) Shopping area. (h) Administrative district. (i) Hospital.

corresponding to each functional area. This study takes 1/10 of the dataset as the test set to quantify the training effect.

A. Remote Sensing Data

High-resolution remote sensing images have been widely used for processes, such as urban function classification. This study utilized 40 000 satellite images with a pixel size of 100×100. Examples of remote sensing images are shown in Fig. 1, where Fig. 1(a)–(i) display residential area, school, industrial park, railway station, airports, park, shopping area, administrative district, and hospital, respectively.

B. User Data

This study used 40 000 user files, each of which comprises user access records in a given area. The data recording time was from October 1, 2018 to March 31, 2019. The data recording frequency was 1 h, that is, the data were recorded every hour. Each piece of data includes the user ID, date, and specific time. The data format is shown in Table II.

Table II lists the data of several users. As an example, the user with the user name 0d4fe92bc13a2525 was counted at 8:00, 9:10, etc., on October 9, 2018, and at 8:00, 9:00, 10:00, etc., on October 10, 2018.

TABLE II
EXAMPLES OF USER BEHAVIOR DATA

ID	Data	Time 1	Time 2	Time 3	.....
0d4fe92bc13a2525	20181009	08	09	10	.....
	20181010	08	09	10	.....
cd94bfc45402a2da	20181010	08	09	10	.....
	20181019	14			
	20181107	13	16		
0f3c66013f897f49	20190101	00	01	02	.....
	20190102	00	01		
ba698efaf0f8f15a	20190113	01	02	03	.....
	20190114	00	01	02	.....
	20190115	10	11	12	.....

Note: The ellipsis omits part of the time, which is the same as the previous three times.

IV. METHODS

In this study, a functional area recognition model was developed based on dual-modal data of remote sensing images and text recorded from online user visits. In the case of text modality data, we use feature engineering for data mining to extract the travel characteristics of people. In the case of image modality

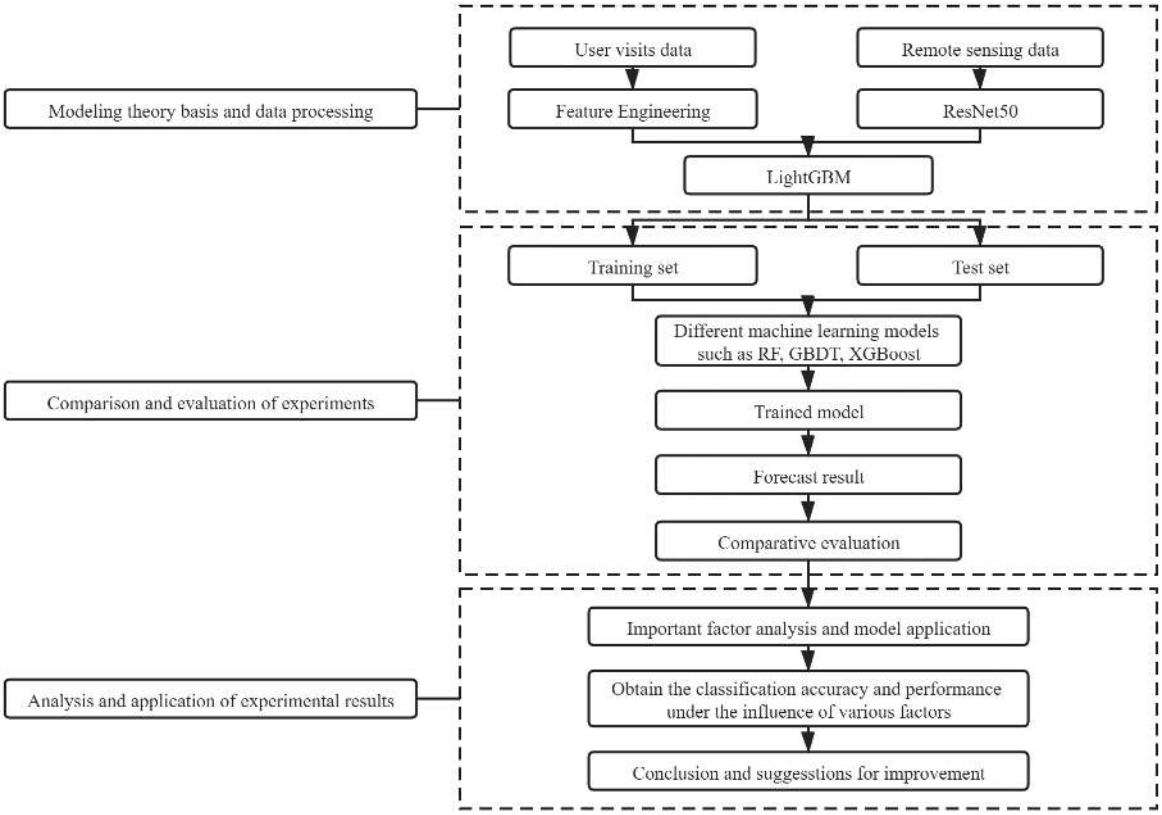


Fig. 2. Algorithm flowchart.

data, we use a residual neural network (ResNet), an architecture of a CNN, for model training. LightGBM is used to combine the processed data of the two modalities for training and learning. Fig. 2 shows a flowchart of the algorithm described in this article.

A. Probability Characteristics of the CNN Remote Sensing Image Classification

CNNs are feedforward neural networks that include convolution calculations and a deep structure. They achieved promising results in image recognition [41], [42]. The training process is described as follows.

1) The input data are first extracted from the convolution and filter kernels.
2) The nonlinear operation of the data after the convolution through the activation function makes the model more adaptable.
3) The data are input into the pooling layer for subsampling, reducing the complexity of the model, speeding up the calculation, reducing the possibility of overfitting, and making the image characteristics after sampling invariant.
4) Finally, the fully connected layer combines features to perform prediction classification [43].

As a CNN feature extraction network, ResNet solves the problem of network degradation after the increase in network depth through a deep residual learning framework. As shown in Fig. 3, ResNet fits a residual mapping, the input is x, the mapping to be solved is $H(x)$, and the residual learning aims to learn the residual between the input and the output (i.e., to learn $F(x) = H(x) - x$). Thus, the initial problem turns into $F(x) + x$. It would be easier to push the residual to zero than to fit an identity mapping by a stack of nonlinear layers if identity mappings are optimal. The formulation of $F(x) + x$ can be realized by feedforward neural networks with "shortcut connections", which are those connections that skip one or more layers. In addition, when the dimensions do not match, we can perform a linear projection W_s by the shortcut connections to match the dimensions. We consider a building block defined as $y = F(x, W_i) + x$ [44].

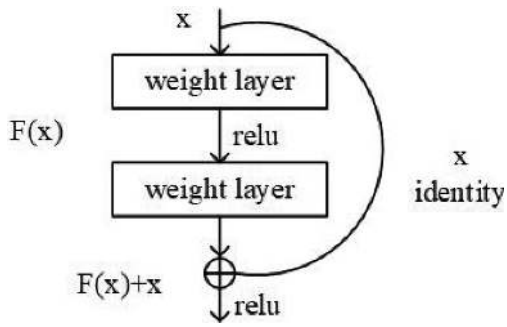


Fig. 3. ResNet.

B. Classification of Multimodal Data Based on LightGBM

LightGBM [45] is an efficient implementation of the gradient lifting decision tree (GBDT), proposed by Microsoft in 2017, which offers superior performance with respect to classification,

prediction, and sorting. Therefore, it is widely used in ML and data mining tasks [38], [40], [46].

1) Foundation of LightGBM—Gradient Lifting Decision Tree: GBDT [47], [48] uses an addition model and a forward distribution algorithm to combine several decision trees through voting and other methods. The key aspect of GBDT is the use of the negative gradient value of the loss function as an approximation of the residual to fit a regression tree [49].

For a given supervised training set $X = \{(x_i, y_i)\}_{i=1}^{n}$, where X is the input space, Y is the output space, and $T(x; \theta_m)$ is the basic learner. GBDT iteratively constructs M regression trees, $T(x, \theta_1), T(x, \theta_2) \ldots, T(x, \theta_m)$, and $f(x)$ refers to the final model.

The expression is as follows [50]:

$$f(x) = \sum_{m=1}^{M} f_m(x) = \sum_{m=1}^{M} \beta_m T\left(x; \theta_m\right) \tag{1}$$

where $T(x; \theta_m)$ is the regression tree and β_m is the weight of the corresponding regression tree. The input space is divided into J disjoint regions $R_1, R_2, R_3, \ldots, R_J$, which determine the output constants c_j in each region. Then, the regression tree is expressed as

$$T\left(x; \theta_m\right) = \sum_{j=1}^{J} c_j I\left(x \in R_j\right). \tag{2}$$

In the mth step of the forward distribution algorithm, given the current model $f_{m-1}(x)$, the least squares method and the gradient are used to calculate the residual direction to obtain the model parameters θ_m. The formula is as follows:

$$\begin{aligned} \theta_m &= \arg\min_{\theta,\beta} \sum_{i=1}^{N} \left[y_i^* - \beta T(x; \theta)\right]^2 \\ &= \arg\min_{\theta,\beta} \sum_{i=1}^{N} \left[y_i^* - \beta \sum_{j=1}^{J} c_j I\left(x_i \in R_j\right)\right]^2. \end{aligned} \tag{3}$$

The formula for the gradient y^* is

$$y_i^* = -\left[\frac{\partial L\left(y_i, F\left(x_i\right)\right)}{\partial F\left(x_i\right)}\right]_{f(x)=f_{m-1}(x)}. \tag{4}$$

The optimal value is determined according to $T(x; \theta_m)$, that is, the weight of the current model

$$\begin{aligned} \beta_m &= \arg\min_{\beta} \sum_{i=1}^{N} L\left(y_i, f_{m-1}\left(x_i\right) + \beta T\left(x_i; \theta_m\right)\right) \\ &= \arg\min_{\beta} \sum_{i=1}^{N} L\left(y_i, f_{m-1}\left(x_i\right) + \beta \sum_{j=1}^{J} c_{jm} I\left(x_i \in R_{jm}\right)\right). \end{aligned} \tag{5}$$

This algorithm has a high prediction accuracy and is widely used. However, when the original data are input into the GBDT for analysis, the feature selection node split needs to traverse all possible division points, calculate information gain, and find the optimal division point, which are time consuming and computationally complex [51], [52].

2) LightGBM: With the drastic growth in the amount of data, the GBDT is encountering increasing problems in terms of efficiency. The LightGBM algorithm was proposed to solve these problems. It can reduce time and memory utilization without a reduction in accuracy. LightGBM includes two new technologies to resolve the issues faced by the GBDT [53], [54].

a) Gradient-based one-side sampling (GOSS): The data from different gradients have different roles in the calculation of information gain. The smaller the gradient of the sample, the smaller the training error and weight. Therefore, data with a larger gradient have a greater impact on the calculation of information gain. Samples with larger gradients can be retained, whereas those with smaller gradients can be removed. However, after the sample removal, the original data distribution changes. Therefore, GOSS retains the samples with larger gradients, sets the sampling ratio, and randomly draws the samples with smaller gradients. First, the gradient values are sorted according to the absolute values. The sample with the largest gradient is drawn according to the set ratio $a \times 100\%$, after which the remaining samples are drawn according to the ratio $b \times 100\%$. When calculating the information gain, the multiplying factor of the sample with a small gradient is amplified by $\frac{(1-a)}{b}$, which can cause the samples with larger training errors to not have a significant impact on the original data distribution [55].

b) Exclusive feature building: In practical applications, high-dimensional datasets are typically highly sparse, and most are mutually exclusive, that is, they almost never take nonzero values at the same time. Hence, a bundling strategy can be adopted for such data, thereby reducing the feature dimensions [55]. LightGBM can package mutually exclusive features into a single feature. It uses the feature scanning algorithm to build the same feature histogram from these feature packages. Hence, the complexity of the algorithm can be reduced from O(#data × #feature) to O(#data × #bundle). The value of the bundle is significantly lower than the value of the feature.

V. Results

A. Processing and Results of Remote Sensing Image and User Data

The CNN model after training can fit the training dataset well. To reduce overfitting of the training model, the CNN model was optimized through five-fold cross-validation. This study uses the deep CNN models ResNet50 and ResNet101 for training classification.

The ResNet50 model is a 50-layer residual network with five parts. The first part is a $7 \times 7 \times 64$ convolutional layer. The data then pass through $3 + 4 + 6 + 3$ building blocks, with three layers for each block, and finally a layer of fc. The ResNet101 model is a 101-layer residual network with five parts. The first part is a $7 \times 7 \times 64$ convolutional layer. The data then passes through $3 + 4 + 23 + 3$ building blocks, with three layers for each block, and finally a layer of fc. The difference between the two models is the conv4_x layer, where ResNet50 has 6 blocks, whereas ResNet101 has 23 blocks.

For remote sensing images, we removed the black occluded pictures and performed data enhancement. For the text modal

TABLE III
RESNET50'S PREDICTION OF THE NUMBER OF EACH FUNCTIONAL AREA

Resnet50	Number of predictions								
	Residential area	School	Industrial park	Railway station	Airport	Park	Shopping area	Administrative district	Hospital
Residential area	809	61	9	4	8	31	24	15	20
School	127	522	18	14	5	28	9	21	18
Industrial park	23	20	230	10	14	34	15	13	6
Railway station	16	6	3	70	3	7	3	7	1
Airport	3	1	5	1	298	14	2	5	0
Park	88	40	21	13	13	337	15	16	9
Shopping area	49	14	26	5	8	21	215	5	11
Administrative district	36	14	22	6	8	12	4	139	20
Hospital	61	17	5	1	2	10	6	9	169
Accuracy	82.47%	68.50%	63.01%	60.34%	90.58%	61.05%	60.73%	53.26%	60.36%

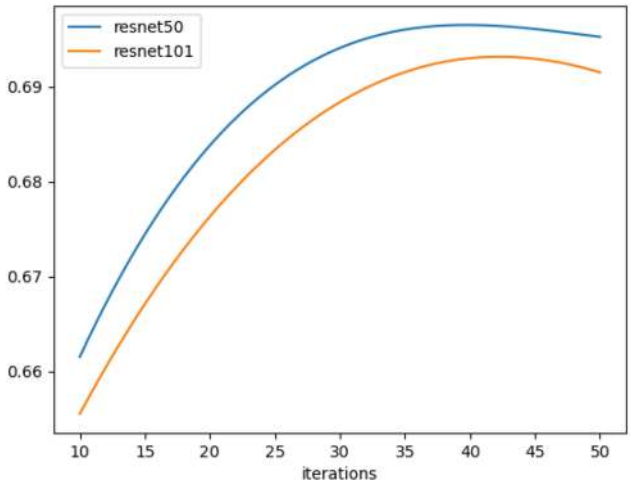


Fig. 4. Relationship between iterations and accuracy.

data, we simply count the number of people in each region at each moment. The classification accuracy of the models on the test set using two models at different iteration times is shown in Fig. 4. According to the results, the classification accuracy of the model increases first and then decreases, reaching a maximum value when the number of iterations is 40. At this point, the accuracy of ResNet50 was 69.72%, and that of ResNet101 was 69.30%.

Table III lists the prediction number of the ResNet50 algorithm for each functional area and the classification accuracy of each of the nine functional areas. The numbers in the first to ninth rows represent the numbers of corresponding files, and the values in the last row represent the classification accuracies of the different categories. According to the data, the correct number of classifications in the residential area is 809, accounting for 82.47% of the total residential area. The number of areas correctly classified as school is 522, accounting for 68.50% of the total number of school. The number of areas correctly classified as industrial park is 230, accounting for 63.01% of the total number of industrial park. The number of areas correctly classified as railway station is 70, accounting for 60.34% of the total number of railway station. The number of areas correctly classified as airport is 298, accounting for 90.58% of the total number of airport. The number of areas correctly classified as park is 337, accounting for 61.05% of the total number of park. The correct number of categories in the shopping area is 215, accounting for 60.73% of the total shopping area. The number of correct classifications in the administrative district is 139, accounting for 53.26% of the total number of administrative district. The number of areas correctly classified as hospital is 169, accounting for 60.36% of the total number of hospital.

Taking parks as an example, the classification using ResNet50 was evaluated. The correct classification accuracy of this functional area is 61.05%, and the images shown in Fig. 5 are examples of the correct classification of a park. Fig. 6 displays images of parks that were misclassified as residential area .

B. *Extraction of User Behavior Features*

User behavior data contains a large amount of information, consisting of both user information and access time information. Human activities and travel have certain characteristics in terms of time. Based on this, information regarding the time and frequency of occurrence in the data can be studied to determine the type of functional area.

Considering that most people go out to work or school in the morning on weekdays, the starting point is generally the residential area and the destination point is a school, industrial park, administrative district, etc. The starting and destination points for travelling at night are the opposite to those in the morning. The main activities engaged in weekends and holidays are shopping, leisure, and entertainment, so the frequency of user visits in park and shopping area will increase relatively. Therefore, the data on working days and rest days (weekends and holidays) for the various functional areas should be considered separately. The average number of people appearing in different categories of dates and functional areas each hour was counted, and the results were displayed in the form of a line chart. The statistical results are shown in Fig. 7, where Fig. 7(a)–(i) represent the residential area, school, industrial park, railway station, airport, park, shopping area, administrative district, and hospital, respectively.

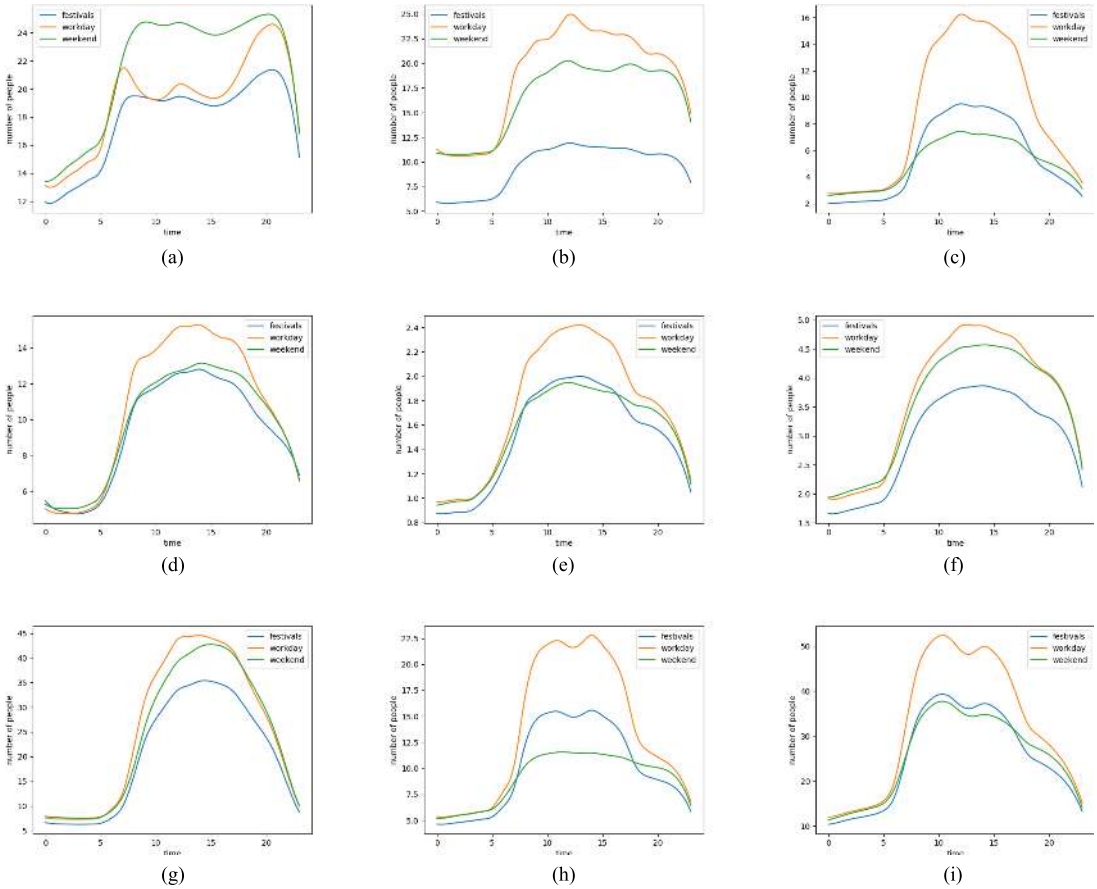
Fig. 7 shows the difference in the statistical characteristics of the number of people at different times in the functional area of the same category, where the abscissa refers to the specific time of day, and the ordinate refers to the average number of people per hour. Because the amount of data in each category is

(a) (b) (c)

Fig. 5. Images correctly classified as parks by using the ResNet50 algorithm. (a)–(c) Parks.

(a) (b) (c)

Fig. 6. Images of park that were misclassified as residential area by using the ResNet50 algorithm. (a)–(c) Parks.



(a) (b) (c) (d) (e) (f) (g) (h) (i)

Fig. 7. Statistics of user behavior characteristics by category. (a) Residential area. (b) School. (c) Industrial park. (d) Railway station. (e) Airport. (f) Park. (g) Shopping area. (h) Administrative district. (i) Hospital.

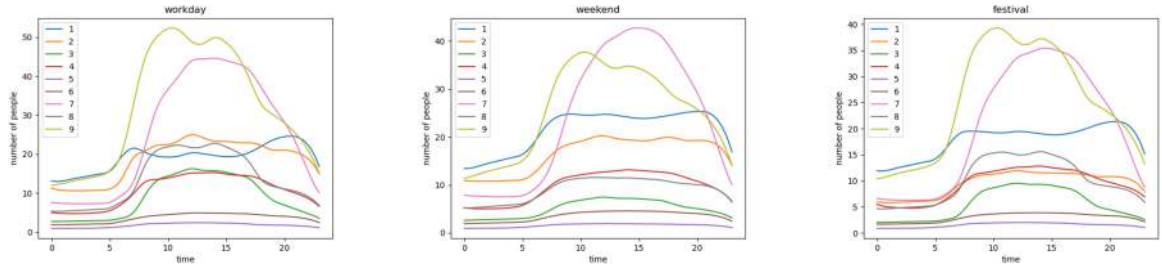


Fig. 8. Statistics of user data on various dates. (a) Workday. (b) Weekend. (c) Festival.

different, the displayed data are the average number of people corresponding to the range of each remote sensing image in a category.

It can be seen from Fig. 7 that the number of people who appear in different functional areas at different times varies greatly. The data reveal that there is some regularity in people's travel. During the time period 8:00–18:00, the number of people present on working days in the residential area is significantly smaller than the number on ordinary weekends. The number of people in the residential area during holidays is also relatively low. In the school area, the number of people present on working days and ordinary weekends was significantly higher than the number of people on holidays. The number of people present in the industrial park on working days is also significantly higher than the number of people on weekends and holidays. In addition, in the time period from 0:00 to 5:00, although the number of people in all functional areas using the social network at night decreased, it is still possible to observe the difference in the number of people in different functional areas. The number of people in residential area, hospital, and other areas at night is more than the number of people in administrative district, shopping area, and other areas. The number of people in schools on weekdays and ordinary weekends differs greatly from the number of people on holidays.

To compare the different characteristics of each functional area in terms of time more clearly, the characteristics of the number of people in each functional area on weekdays, weekends, and holidays were separately counted.

Numbers 1–9 in Fig. 8 represent residential area, school, industrial park, railway station, airport, park, shopping area, administrative district, and hospital, respectively. The abscissa refers to the specific 24 h of the day, and the ordinate refers to the number of people. Fig. 8(a) shows the statistical characteristics of the nine functional areas on weekdays. Fig. 8(b) shows the statistical characteristics of the nine functional areas on weekends. Fig. 8(c) shows the statistical characteristics of the nine functional areas on holidays. The curve in the figure demonstrates the difference in the number of people in each functional area over different time periods.

The difference reflects people's different travel purposes, that is, the difference in the functional area in the city where people go. To make use of ML and deep learning algorithms to obtain good prediction classification for the features observed in the data, irregular and redundant data should be converted into rules by statistical methods, highlighting the data features that can be used as partial input data for the algorithms. Therefore, the traffic characteristics of each category for each hour were estimated for working days and rest days. According to the information in Fig. 8, there is a considerable difference in the number of people present before and after 0:00, 8:00, and 18:00. Therefore, these specific times were used as the dividing points to partition one day into three time periods, and the flow characteristics of each time period were estimated.

In addition, the user ID also provides a wealth of information, with a high user overlap rate in the dataset. The same user can appear in different areas multiple times, so the total duration that the same user appeared in each functional area was counted.

TABLE IV
LIGHTGBM PARAMETER SETTINGS

Parameter	Parameter value
learning_rate	0.11
num_leaves	31
max_depth	–1
early_stopping_rounds	30

C. Classification Results Using LightGBM Algorithm

1) Experimental Environment and Parameter Settings: Python 3.7 was employed to implement the algorithms. The experiments were performed on a computer with a 2.50 GHz, 8-GB RAM, and 64-b Windows 10 operating system. The metric used was classification accuracy, which is the number of correct classifications divided by the total amount of data. The parameter settings used in the experiment are listed in Table IV.

After repeated experiments, the variation of classification accuracy with learning rate is shown in Fig. 9 after five-fold cross-validation.

It can be seen from Fig. 9 that different learning rates have an impact on the classification accuracy. LightGBM performed best when the learning rate was 0.11 and the maximum accuracy was 82.45%. Therefore, the learning rate used in this study was 0.11.

2) Prediction Accuracy of Various Functional Areas: Table V lists the prediction number from the LightGBM algorithm for each functional area and the classification accuracy of each of the nine functional areas. The numbers in the first to ninth rows

TABLE V
LIGHTGBM'S PREDICTION OF THE NUMBER OF EACH FUNCTIONAL AREA

LightGBM	Number of predictions								
	Residential area	School	Industrial park	Railway station	Airport	Park	Shopping area	Administrative district	Hospital
Residential area	814	62	10	5	2	31	25	21	11
School	93	609	8	2	0	24	5	8	13
Industrial park	17	8	313	0	2	12	9	2	2
Railway station	1	1	0	107	1	3	1	2	0
Airport	0	0	0	1	318	9	0	1	0
Park	48	27	8	3	1	446	8	5	6
Shopping area	32	5	8	1	2	15	280	8	3
Administrative district	26	13	7	1	1	8	3	194	8
Hospital	38	11	1	0	0	4	5	4	217
Accuracy	82.98%	79.92%	85.75%	92.24%	96.66%	80.80%	79.10%	74.33%	77.50%

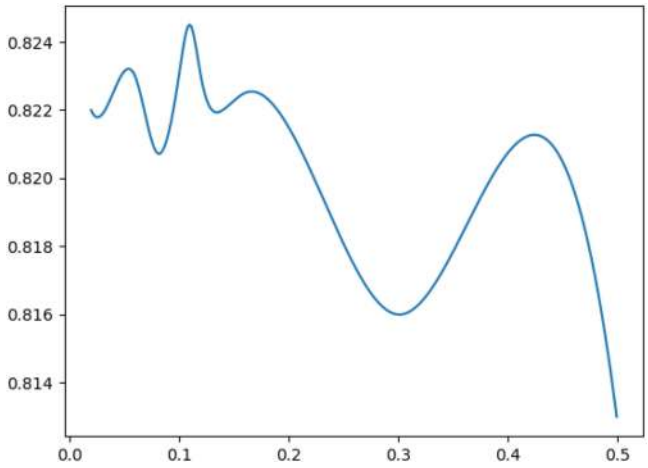


Fig. 9. Relationship between learning rate and accuracy.

Fig. 10. LightGBM's prediction of the number of each functional area.

represent the numbers of corresponding files, and the values in the last row represent the classification accuracies of the different categories. According to the data, the correct number of classifications in the residential area is 814, accounting for 82.98% of the total residential area. The number of areas correctly classified as school is 609, accounting for 79.92% of the total number of school. The number of areas correctly classified as industrial park is 313, accounting for 85.75% of the total number of industrial park. The number of areas correctly classified as railway station is 107, accounting for 92.24% of the total number of railway station. The number of areas correctly classified as airport is 318, accounting for 96.66% of the total number of airport. The number of areas correctly classified as park is 446, accounting for 80.80% of the total number of park. The correct number of categories in the shopping area is 280, accounting for 79.10% of the total shopping area. The number of correct classifications in the administrative district was 194, accounting for 74.33% of the total number of administrative district. The number of areas correctly classified as hospital is 217, accounting for 77.50% of the total number of hospital. The use of the LightGBM algorithm shows relatively high accuracy in the classification of railway station and airport, reaching 92.24% and 96.66%, respectively, whereas the accuracy in classification of administrative district and hospital is relatively low, at 74.33% and 77.50%.

Fig. 10 shows the classification of various types using the LightGBM algorithm, where numbers 0–8 represent residential area, school, industrial park, railway station, airport, park, shopping area, administrative district, and hospital, respectively. The grid in the figure is the proportion of the functional area when the functional area is represented by the ordinate, and the LightGBM classification is represented by the abscissa. A dark grid color indicates comparatively large proportions. The number in the grid is the ratio of the actual number of classifications to the total number of functional areas, and the units are percentage. Taking the ordinate of 0 as an example, the value of the intersection with the abscissa 1 is 6.3, which means that all the residential areas in the test set are classified and the actual number of schools is 6.3% of the total number of residential area.

The classification of each functional area can be seen intuitively from the image. In the diagonal part, the diagonal grid as a whole is significantly darker than the nondiagonal line, indicating that the proportion of correct classification of each functional area is much higher than the proportion of incorrect classification, of which the colors of the railway station and airport are the darkest, and the classification accuracy is the highest. The lower right corner of the grid is lighter, and the classification accuracy is relatively low. In the nondiagonal part,

(a) (b) (c)

Fig. 11. Images correctly classified as parks by using the LightGBM algorithm. (a)–(c) Parks.

(a) (b) (c)

Fig. 12. Images of park that were misclassified as residential area by using the LightGBM algorithm. (a)–(c) Parks.

(a) (b) (c)

Fig. 13. Images that were correctly classified after adding data with time features. (a)–(c) Parks.

there is a high proportion of misclassification of functional areas of school, park, shopping area, administrative district, and hospital as residential area, and the overall color of this part is darker.

Taking the park as an example, the classification using LightGBM is evaluated. The correct classification accuracy of this functional area is 80.80%, and the images shown in Fig. 11 are examples of the correct classification as a park. The images shown in Fig. 12 are examples of images in a park that have been misclassified as a residential area.

Fig. 13 shows an example of data that was originally misclassified as a residential area and now correctly classified as a park after adding data containing the time characteristics of user access instead of using only image data.

Fig. 14 shows an example of parks that were originally misclassified as residential area and were still misclassified as residential area after adding data containing the time characteristics of user access instead of using only image data.

One possible reason for this result is that the remote sensing image can only display the actual appearance of the landform, and the functional areas of different categories have overlapping features, such as architecture and greening, so that when only the remote sensing image is available, the type of remote sensing image cannot be clearly identified. The urban spatial structure has a close relationship with human activities, so after adding human-activity information, information that cannot be extracted from remote sensing data can be supplemented, and the classification accuracy is improved.

(a) (b) (c)

Fig. 14. Images that were still misclassified after adding data with time features. (a)–(c) Parks.

TABLE VI
COMPARISON OF RESULTS BETWEEN MULTIMODAL DATA AND SINGLE-MODAL DATA

Modal	Accuracy
image	48.28%
text	80.43%
image and text	82.45%

TABLE VII
CLASSIFICATION ACCURACY AND TRAINING DURATION OF ML ALGORITHMS

Machine learning algorithm	Accuracy	Training time
RF	80.40%	163 s
GBDT	81.05%	10532 s
XGBoost	81.98%	2794 s
LightGBM	82.45%	372 s

D. Comparative Test

1) Comparison Between Multimodal Data and Single-Modal Data: Table VI lists the classification accuracy achieved using the data of each modality. It shows that the classification accuracies achieved using image and text modality data separately are 48.28% and 80.43%, respectively, whereas the classification accuracy achieved using the combined image and text modality data is 82.45%.

Table VI indicates that the classification accuracy achieved by combining the data of the two modalities is the highest (34.17% and 2.02% higher than the classification accuracies achieved using image and text modality data, respectively). In other words, the table indicates that combining the data of the image and text modalities improves the classification accuracy. The improvement over image modality is higher than that over text modality. One possible reason is that the image modality dataset used in this study has low pixels, and thus, a relatively low amount of information can be extracted from it.

2) Comparison of LightGBM and Other ML Algorithms: Table VII lists the classification accuracy and training duration of each algorithm. The data in Table VII show that the classification accuracy and training time of the RF algorithm are 80.40% and 163 s, respectively. The classification accuracy and training time of the GBDT are 81.05% and 10532 s, respectively. The classification accuracy and training time of the extreme gradient boosting (XGBoost) algorithm are 81.98% and 2794 s, respectively. The classification accuracy and training time of the LightGBM algorithm are 82.45% and 372 s, respectively.

From the perspective of accuracy, the LightGBM algorithm has the highest accuracy among the ML algorithms, at 82.45%, and the classification accuracy of the RF algorithm is relatively low, at 80.40%. In terms of time, the training time of the RF algorithm is the shortest, at 163 s, and the training time of the LightGBM algorithm is also short. Because the LightGBM method has been improved, the training time is much shorter than those of the GBDT and XGBoost methods. Overall, LightGBM is the most suitable algorithm for this process.

3) Prediction Accuracy of Different ML Algorithms for Each Functional Area: Table VIII lists the predicted number of each functional area using the RF algorithm and the classification accuracy of the nine functional areas. The numbers in the first to ninth rows represent the numbers of corresponding files, and the values in the last row represent the classification accuracies of the different categories. From the data, the correct number of classifications of residential area is 829, accounting for 84.51% of the total number of residential area. The number of correctly classified school is 594, accounting for 77.95% of the total number of school. The correct number of classifications of industrial park is 306, accounting for 83.84% of the total number of industrial park. The number of areas correctly classified as railway station is 100, accounting for 86.21% of the total number of railway station. The number of areas correctly classified as airport is 316, accounting for 96.05% of the total number of airport. The number of areas correctly classified as park is 421, accounting for 76.27% of the total number of park. The correct number of classifications of shopping area is 263, accounting for 74.29% of the total shopping area. The number of correct classifications of administrative district is 180, accounting for 68.97% of the total number of administrative district. The number of areas correctly classified as hospital is 207, accounting for 73.93% of the total number of hospital. In conclusion, the accuracy of classification of airport using the RF algorithm is relatively high, reaching 96.05%, whereas the accuracy of classification of administrative district is relatively low (68.97%). The structures of Tables IX and X are similar to the structure of Table VIII. Tables IX and X present the number of each functional area

TABLE VIII
RF'S PREDICTION OF THE NUMBER OF EACH FUNCTIONAL AREA

RF	Number of predictions								
	Residential area	School	Industrial park	Railway station	Airport	Park	Shopping area	Administrative district	Hospital
Residential area	829	53	12	3	3	35	21	16	9
School	109	594	9	2	0	23	6	6	13
Industrial park	29	6	306	0	3	14	4	1	2
Railway station	2	3	0	100	1	5	2	3	0
Airport	2	0	0	1	316	9	0	1	0
Park	65	27	11	2	4	421	10	8	4
Shopping area	44	9	6	1	1	21	263	4	5
Administrative district	36	13	9	1	1	10	5	180	6
Hospital	47	10	2	1	0	3	5	5	207
Accuracy	84.51%	77.95%	83.84%	86.21%	96.05%	76.27%	74.29%	68.97%	73.93%

TABLE IX
GBDT'S PREDICTION OF THE NUMBER OF EACH FUNCTIONAL AREA

GBDT	Number of predictions								
	Residential area	School	Industrial park	Railway station	Airport	Park	Shopping area	Administrative district	Hospital
Residential area	811	60	12	5	2	33	26	19	13
School	99	598	8	3	0	24	6	8	16
Industrial park	19	8	312	0	2	12	7	3	2
Railway station	5	7	0	98	1	3	1	1	0
Airport	1	1	0	1	316	9	0	1	0
Park	50	28	12	4	2	430	11	7	8
Shopping area	36	9	8	1	1	19	271	5	4
Administrative district	22	12	11	1	1	9	5	194	6
Hospital	36	12	1	2	0	6	6	4	213
Accuracy	82.67%	78.48%	85.48%	84.48%	96.05%	77.90%	76.55%	74.33%	76.07%

TABLE X
XGBOOST'S PREDICTION OF THE NUMBER OF EACH FUNCTIONAL AREA

XGBoost	Number of predictions								
	Residential area	School	Industrial park	Railway station	Airport	Park	Shopping area	Administrative district	Hospital
Residential area	816	64	11	3	2	32	22	20	11
School	95	606	6	2	0	29	5	7	12
Industrial park	17	10	309	0	2	16	8	2	1
Railway station	1	1	0	106	1	3	1	3	0
Airport	1	0	0	1	317	9	0	1	0
Park	49	28	10	2	1	442	9	5	6
Shopping area	38	6	6	1	2	14	276	7	4
Administrative district	26	15	7	1	1	9	3	192	7
Hospital	38	11	1	1	0	6	4	4	215
Accuracy	83.18%	79.52%	84.66%	91.38%	96.35%	80.07%	77.97%	73.56%	76.79%

predicted by the GBDT and XGBoost algorithms, respectively, and the classification accuracy of each of the nine functional areas.

Figs. 15–17 show the classification of various types using different algorithms. The format is the same as in Fig. 10. For school, park, shopping area, administrative district, and hospital, the five functional areas are misclassified as residential area, the overall color of which is darker. Among them, the XGBoost, GBDT, and LightGBM algorithms are lighter than RF, indicating that the classification error rate of RF in this part is higher. For the rest, the XGBoost and LightGBM algorithms are shallower than those of the GBDT and RF algorithms, and the classification error rate of the first two is generally lower.

The line chart in Fig. 18 is a comparison of the accuracy of classification by each algorithm in each category. The abscissas 0–8 in the figure represent residential area, school, industrial park, railway station, airport, park, shopping area, administrative district, and hospital, respectively, and the ordinate refers to the accuracy. Combining Tables V, VIII–X, and Fig. 17, it can be seen intuitively that the accuracy of classification using RF for most functional areas is slightly lower than that of other algorithms, and it is significantly lower in shopping area, administrative district, and hospital, respectively, at 74.29%, 68.97%, and 73.93%. The accuracy of classification for residential area is higher than that of other algorithms, at 84.51%, and the accuracy of classification by GBDT on railway station is obviously low,

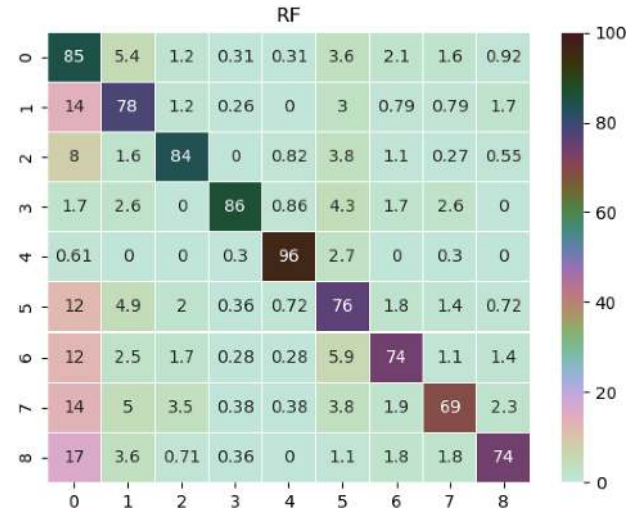


Fig. 15. RF's prediction of the number of each functional area.

Fig. 16. GBDT's prediction of the number of each functional area.

Fig. 17. XGBoost's prediction of the number of each functional area.

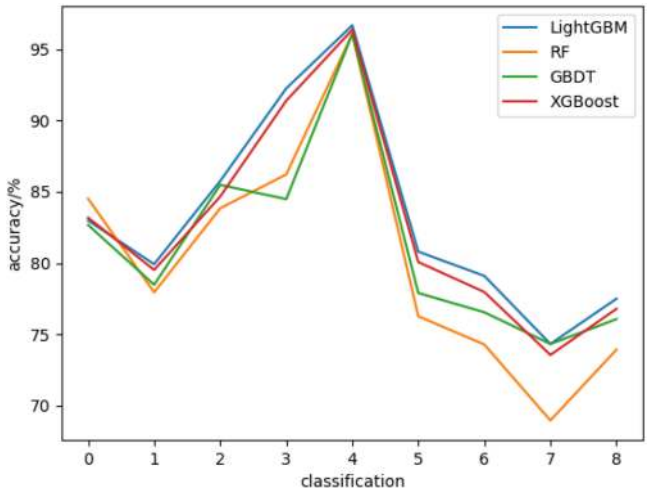


Fig. 18. Classification accuracy of each ML algorithm.

at only 84.48%. In conclusion, the LightGBM algorithm has a high classification accuracy for various functional areas.

VI. Conclusion

In this article, we propose a method based on the data from two modalities and ML algorithms to classify urban functional areas. The main technical contributions of this study include the following: first, the data from two modalities were processed and combined. Second, LightGBM ML algorithm was used to classify urban functional areas, and it achieved a relatively high accuracy and a short training time. The effective combination of image recognition technology and statistical technology provides new ideas for the classification of urban functional areas and new solutions for the sustainable development of cities. The use of big data technology is an inevitable choice for urban scientific planning.

There are still some technical problems associated with urban functional area classification based on the remote sensing images used in this study. The images used were multiple local images in each category, which are conducive to the recognition and classification of different functional areas by their unique landforms but cannot directly reflect the distribution of different types of functional areas in the city and their interrelationships. Hence, it is impossible to draw a functional area planning map of the city using these images, which leads to certain limitations in studying the geographical distribution of functional areas. In addition, through experiments, it was found that increasing the time information regarding a user visit greatly improved the recognition accuracy. The reason for this result is that adding the time feature adds another aspect to the separate image dataset. However, the classification accuracy can still be improved. One of the reasons for this is that the mining of user behavior information is insufficient. Therefore, the focus of future research is to determine ways to fully express the time information in the dataset.

REFERENCES

[1] W. Colglazier, "Sustainable development agenda: 2030," *Science*, vol. 349, no. 6252, pp. 1048–1050, 2015.

[2] S. Parnell, "Defining a global urban development agenda," *World Develop.*, vol. 78, pp. 529–540, 2016.

[3] S. Kumar, N. Kumar, and S. Vivekadhish, "Millennium development goals (MDGs) to sustainable development goals (SDGs): Addressing unfinished agenda and strengthening sustainable development and partnership," *Indian J. Community Med.: Official Publication Indian Assoc. Preventive Social Med.*, vol. 41, no. 1, pp. 1–4, 2016

[4] M. Li, Z. Shen, and X. Hao, "Revealing the relationship between spatio-temporal distribution of population and urban function with social media data," *GeoJournal*, vol. 81, no. 6, pp. 919–935, 2016.

[5] N. Longbotham, C. Chaapel, L. Bleiler, C. Padwick, W. J. Emery, and F. Pacifici, "Very high resolution multiangle urban classification analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 4, pp. 1155–1170, Apr. 2012.

[6] L. Jiao *et al.*, "Proximity expansion index: An improved approach to characterize evolution process of urban expansion," *Comput., Environ. Urban Syst.*, vol. 70, pp. 102–112, 2018.

[7] P. S. Chasek, L. M. Wagner, F. Leone, A.-M. Lebada, and N. Risse, "Getting to 2030: Negotiating the post-2015 sustainable development agenda," *Rev. Eur., Comparative Int. Environ. Law*, vol. 25, no. 1, pp. 5–14, 2016.

[8] A. Tayyebi, B. C. Pijanowski, and A. H. Tayyebi, "An urban growth boundary model using neural networks, GIS and radial parameterization: An application to Tehran, Iran," *Landscape Urban Planning*, vol. 100, no. 1/2, pp. 35–44, 2011.

[9] Y. Chen *et al.*, "Delineating urban functional areas with building-level social media data: A dynamic time warping (DTW) distance based k-medoids method," *Landscape Urban Planning*, vol. 160, pp. 48–60, 2017.

[10] H. Xing and Y. Meng, "Integrating landscape metrics and socioeconomic features for urban functional region classification," *Comput., Environ. Urban Syst.*, vol. 72, pp. 134–145, 2018.

[11] A. P. Montanges, "Classification of urban structural types (UST) using multiple data sources and spatial priors," École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, Tech. Rep., 2014, doi: 10.1590/S1519-566X2009000100011.

[12] T. Pei, S. Sobolevsky, C. Ratti, S.-L. Shaw, T. Li, and C. Zhou, "A new insight into land use classification based on aggregated mobile phone data," *Int. J. Geographical Inf. Sci.*, vol. 28, no. 9, pp. 1988–2007, 2014.

[13] M. C. Pagliarella *et al.*, "From one-to two-phase sampling to reduce costs of remote sensing-based estimation of land-cover and land-use proportions and their changes," *Remote Sens. Environ.*, vol. 184, pp. 410–417, 2016.

[14] F. Pacifici, M. Chini, and W. J. Emery, "A neural network approach using multi-scale textural metrics from very high-resolution panchromatic imagery for urban land-use classification," *Remote Sens. Environ.*, vol. 113, no. 6, pp. 1276–1292, 2009.

[15] G. Cheng, X. Xie, J. Han, L. Guo, and G. Xia, "Remote sensing image scene classification meets deep learning: Challenges, methods, benchmarks, and opportunities," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, no. 99, pp. 3735–3756, Jun. 2020.

[16] P. Longley, "Geographical information systems: A renaissance of geodemographics for public service delivery," *Prog. Human Geography*, vol. 29, no. 1, pp. 57–63, 2005.

[17] M. Li, A. Stein, W. Bijker, and Q. Zhan, "Urban land use extraction from very high resolution remote sensing imagery using a Bayesian network," *ISPRS J. Photogrammetry Remote Sens.*, vol. 122, pp. 192–205, 2016.

[18] X. Zhang, S. Du, and Q. Wang, "Hierarchical semantic cognition for urban functional zones with VHR satellite images and POI data," *ISPRS J. Photogrammetry Remote Sens.*, vol. 132, pp. 170–184, 2017.

[19] D. Martin, S. Cockings, and A. Harfoot, "Development of a geographical framework for census workplace data," *J. Roy. Statist. Soc.: Ser. A (Statist. Soc.)*, vol. 176, no. 2, pp. 585–602, 2013.

[20] S. Jiang, J. Ferreira, and M. C. Gonzalez, "Activity-based human mobility patterns inferred from mobile phone data: A case study of Singapore," *IEEE Trans. Big Data*, vol. 3, no. 2, pp. 208–219, Jun. 2017.

[21] C. Zhong, X. Huang, S. M. Arisona, and G. Schmitt, "Identifying spatial structure of urban functional centers using travel survey data: A case study of Singapore," in *Proc. 1st ACM SIGSPATIAL Int. Workshop Comput. Models Place*, Oct. 2013, pp. 28–33.

[22] X. Chen, "A tale of two regions in China: Rapid economic development and slow industrial upgrading in the Pearl river and the Yangtze river deltas," *Int. J. Comparative Sociol.*, vol. 48, no. 2/3, pp. 167–201, 2007.

[23] S. L. Hedden, D. E. Whitaker, S. Von Thomsen, S. G. Severtson, and W. W. Latimer, "Latent patterns of risk behavior in urban African-American middle school students in Baltimore city," *J. Child Adolescent Substance Abuse*, vol. 20, no. 1, pp. 34–47, 2010.

[24] Y. Zheng, L. Capra, O. Wolfson, and H. Yang, "Urban computing: Concepts, methodologies, and applications," *ACM Trans. Intell. Syst. Technol.*, vol. 5, no. 3, pp. 1–55, Sep. 2014.

[25] Y. Ran, X. Li, L. Lu, and Z. Li, "Large-scale land cover mapping with the integration of multi-source information based on the Dempster–Shafer theory," *Int. J. Geographical Inf. Sci.*, vol. 26, no. 1, pp. 169–191, 2012.

[26] Z. Shao and J. Cai, "Remote sensing image fusion with deep convolutional neural network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 5, pp. 1656–1669, May 2018.

[27] B. Wang, Y. Wang, K. Qin, and Q. Xia, "Detecting transportation modes based on LightGBM classifier from GPS trajectory data," in *Proc. IEEE 26th Int. Conf. Geoinformat.*, 2018, pp. 1–7.

[28] G. Cheng, J. Han, and X. Lu, "Remote sensing image scene classification: Benchmark and state of the art," *Proc. IEEE*, vol. 105, no. 10, pp. 1865–1883, Oct. 2017.

[29] W. Zhou, D. Ming, X. Lv, K. Zhou, and H. Bao, "SO–CNN based urban functional zone fine division with VHR remote sensing image," *Remote Sens. Environ.*, vol. 236, 2019, Art. no. 111458.

[30] G. Cheng, C. Yang, X. Yao, L. Guo, and J. Han, "When deep learning meets metric learning: Remote sensing image scene classification via learning discriminative CNNs," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 5, pp. 2811–2821, May 2018.

[31] Z. Xiuyuan, D. Shihong, W. Qiao, and Z. Weiqi, "Multiscale geoscene segmentation for extracting urban functional zones from VHR satellite images," *Remote Sens.*, vol. 10, no. 2, 2018, Art. no. 281.

[32] Y. Zhi *et al.*, "Latent spatio-temporal activity structures: A new approach to inferring intra-urban functional regions via social media check-in data," *Geo-Spatial Inf. Sci.*, vol. 19, pp. 94–105, 2016.

[33] J. L. Toole, M. Ulm, D. Bauer, and M. C. Gonzalez, "Inferring land use from mobile phone activity," in *Proc. ACM SIGKDD Int. Workshop Urban Comput.*, pp. 1–8, 2012.

[34] H. Han, X. Yu, and Y. Long, "Discovering functional zones using bus smart card data and points of interest in Beijing, " in, *Geospatial Analysis to Support Urban Planning in Beijing*. Berlin, Germany: Springer, 2015.

[35] N. J. Yuan, Y. Zheng, and X. Xie, "Discovering regions of different functions in a city using human mobility and POIs," in *Proc. 18th SIGKDD Conf. Knowl. Discovery Data Mining*, Aug. 2012. [Online]. Available: https://www.microsoft.com/en-us/research/publication/discovering-regions-of-different-functions-in-a-city-using-human-mobility-and-pois/

[36] Y. Yao *et al.*, "Sensing spatial distribution of urban land use by integrating points-of-interest and Google Word2Vec model," *Int. J. Geographical Inf. Sci.*, vol. 31, pp. 825–848, 2016.

[37] D. Praveen Kumar, T. Amgoth, and C. S. R. Annavarapu, "Machine learning algorithms for wireless sensor networks: A survey," *Inf. Fusion*, vol. 49, pp. 1–25, 2019.

[38] C. Chen, Q. Zhang, Q. Ma, and B. Yu, "LightGBM-PPI: Predicting protein-protein interactions through LightGBM with multi-information fusion," *Chemometrics Intell. Lab. Syst.*, vol. 191, pp. 54–64, 2019.

[39] Y. Ju, G. Sun, Q. Chen, M. Zhang, H. Zhu, and M. U. Rehman, "A model combining convolutional neural network and LightGBM algorithm for ultra-short-term wind power forecasting," *IEEE Access*, vol. 7, pp. 28309–28318, 2019.

[40] X. Sun, M. Liu, and Z. Sima, "A novel cryptocurrency price trend forecasting model based on LightGBM," *Finance Res. Lett.*, vol. 32, 2020, Art. no. 101084.

[41] J. Gu *et al.*, "Recent advances in convolutional neural networks," *Pattern Recognit.*, vol. 77, pp. 354–377, 2018.

[42] Z. Deng, H. Sun, S. Zhou, J. Zhao, and H. Zou, "Toward fast and accurate vehicle detection in aerial images using coupled region-based convolutional neural networks," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 8, pp. 3652–3664, Aug. 2017.

[43] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[44] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
[45] Q. Meng *et al.*, "A communication-efficient parallel algorithm for decision tree," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 1279–1287.
[46] X. Ma, J. Sha, D. Wang, Y. Yu, Q. Yang, and X. Niu, "Study on a prediction of P2P network loan default based on the machine learning LightGBM and XGboost algorithms according to different high dimensional data cleaning," *Electron. Commerce Res. Appl.*, vol. 31, pp. 24–39, 2018.
[47] C. Becker, R. Rigamonti, V. Lepetit, and P. Fua, "Supervised feature learning for curvilinear structure segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*, 2013, pp. 526–533.
[48] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Statist.*, vol. 29, pp. 1189–1232, 2001.
[49] H. Rao *et al.*, "Feature selection based on artificial bee colony and gradient boosting decision tree," *Appl. Soft Comput.*, vol. 74, pp. 634–642, 2019.
[50] G. Ridgeway, "Generalized boosted models: A guide to the GBM package," *Update*, vol. 1, no. 1, 2007, Art. no. 2007.
[51] J. Son, I. Jung, K. Park, and B. Han, "Tracking-by-segmentation with online gradient boosting decision tree," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 3056–3064.
[52] M. Schonlau, "Boosted regression (boosting): An introductory tutorial and a Stata plugin," *Stata J.*, vol. 5, no. 3, pp. 330–354, 2005.
[53] S. Ranka and V. Singh, "CLOUDS: A decision tree classifier for large datasets," in *Proc. 4th Knowl. Discovery Data Mining Conf.*, pp. 2–8, vol. 2, no. 8, 1998.
[54] R. Jin and G. Agrawal, "Communication and memory efficient parallel decision tree construction," in *Proc. SIAM Int. Conf. Data Mining*, 2003, pp. 119–129.
[55] G. Ke *et al.*, "LightGBM: A highly efficient gradient boosting decision tree," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 3146–3154.

Chen Chen is currently working toward the B.E. degree with the China University of Geosciences, Wuhan, China.

Her research interests include remote sensing data processing, data mining, and classification of urban functional areas.

Jining Yan received the Ph.D. degree in signal and information processing from the University of Chinese Academy of Sciences, Beijing, China, in 2017.

He is currently an Associate Professor with the School of Computer Science, China University of Geosciences, Wuhan, China. His research interests include remote sensing data processing, time-series analysis and change detection, cloud computing in remote sensing, and applied oceanography.

Lizhe Wang (Fellow, IEEE) received the B.E. and M.E. degrees in electrical engineering and automation from Tsinghua University, Beijing, China, in 1998 and 2001, respectively, and the D.E. degree (magna cumlaude) in applied computing science from the Karlsruhe Institute of Technology, Karlsruhe, Germany, in 2007. He is currently a "ChuTian" Chair Professor with the School of Computer Science, China University of Geosciences, Wuhan, China.

His main research interests include HPC, e-Science, and remote sensing image processing.

Prof. Wang is a Fellow of IET and British Computer Society. He serves as an Associate Editor for the IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, IEEE TRANSACTIONS ON CLOUD COMPUTING, and IEEE TRANSACTIONS ON SUSTAINABLE COMPUTING.

Dong Liang received the B.Sc. degree in applied mathematics from the University of Hull, Hull, U.K., in 2006, and the M.Sc. degree in applied mathematics from Malardalen University, Västerås, Sweden, in 2009. He is currently working toward the Ph.D. degree in cartography and GIS with the University of Chinese Academy of Sciences, Beijing, China, with a focus on Big Earth Data and polar remote sensing.

Wanfeng Zhang received the bachelor's degree and master's degree in geography information system from Jilin University, Changchun, China, in 2005 and 2007 respectively, and the Ph.D. degree in multi-datacenters parallelized computing from the University of Chinese Academy of Sciences, Beijing, China, in 2014.

He is currently an Associate Researcher with the Technology and Engineering Center for Space Utilization, Chinese Academy of Sciences, Beijing, China. His current research interests include space science data processing on high-performance cluster, cloud computing, and ground data processing system.