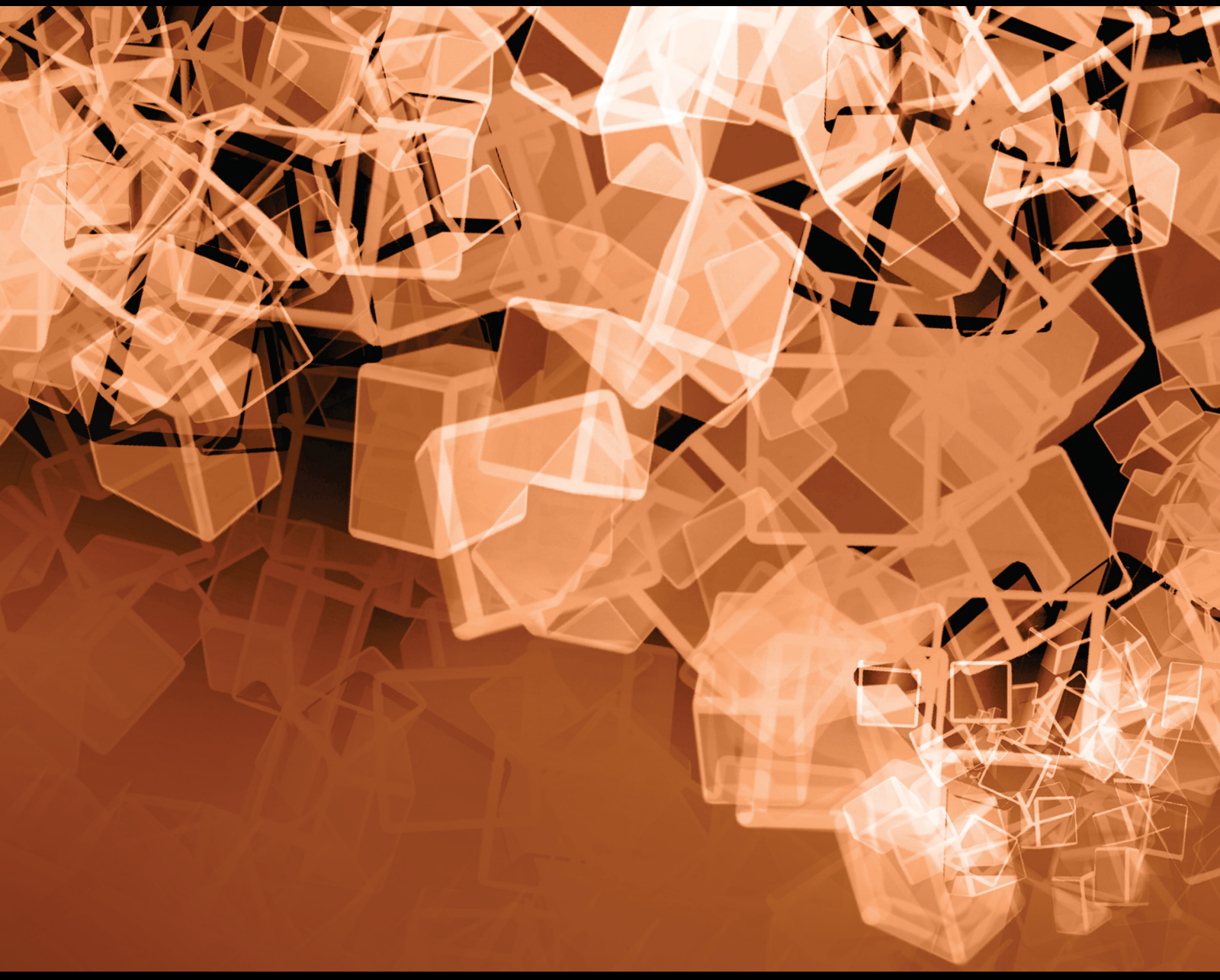
Discrete Dynamics in Nature and Society

Discrete Dynamics of Complex Interactions between Natural and Artificial Systems 2021

Lead Guest Editor: Jorge E. Macias-Diaz
Guest Editors: Qin Sheng and Stefania Tomasiello



Discrete Dynamics of Complex Interactions between Natural and Artificial Systems 2021

Discrete Dynamics in Nature and Society

Discrete Dynamics of Complex Interactions between Natural and Artificial Systems 2021

Lead Guest Editor: Jorge E. Macias-Diaz
Guest Editors: Qin Sheng and Stefania Tomasiello

Chief Editor

Associate Editors

Academic Editors

Contents

Hindawi
Discrete Dynamics in Nature and Society
Volume 2022, Article ID 8417295, 13 pages
https://doi.org/10.1155/2022/8417295

Research Article

Improved CenterNet for Accurate and Fast Fitting Object Detection

Huimin He,[1] Qionglan Na,[1] Dan Su,[1] Kai Zhao,[2] Jing Lou,[1] and Yixi Yang[3]

[1]*State Grid Jibei Information and Telecommunication Company, Beijing 100053, China*
[2]*North China Electric Power University, Department of Electronic and Communication Engineering, Baoding 071003, China*
[3]*State Grid Information and Telecommunication Branch, Beijing 100761, China*

Correspondence should be addressed to Qionglan Na; 81885883@qq.com

Received 14 January 2022; Accepted 4 May 2022; Published 30 May 2022

Academic Editor: Jorge E. Macias-Diaz

Accurate and fast detection of typical fittings is the prerequisite of condition monitoring and fault diagnosis. At present, most successful fitting detectors are anchor-based, which are challenging to meet the requirements of edge deployment. In this paper, we propose a novel anchor-free method called HRM-CenterNet. Firstly, the lightweight MobileNetV3 is introduced into CenterNet to extract multi-scale features of different layers. In addition, the lightweight receptive field enhancement module is proposed for the deep layer features to further enhance the characterization power of global features and generate more accurate heatmaps. Finally, the high-resolution feature fusion network with iterative aggregation is designed to reduce the loss of spatial semantic information in subsampling and further improve the accuracy of small and occlusion objects. Experiments are carried out on the TFITS and PASCAL VOC datasets. The results show that the size of the network is more than 60% lower than that of CenterNet. Compared with other detectors, our method achieves comparable accuracy with all accurate models at a much faster speed and meets the performance requirements of real-time detection.

1. Introduction

Transmission line fault is an important cause of power grid blackout. Inspection periodically can significantly reduce the workload of operation and maintenance personnel on pole inspection, which is an essential means to ensure the safe operation of the power system [1]. Fittings are iron or copper metal accessories widely used in transmission lines, mainly used to support, fix, and connect bare conductors, conductors, and insulators [2]. Because the fittings run outdoors all year round, it is easy to produce corrosion, deformation, damage, and other phenomena. Therefore, the realization of high-precision automatic detection of fittings can predict their faults in advance, which is of great significance in ensuring the safe operation of the power grid [3]. In recent years, the power system has promoted unmanned aerial vehicle transmission line inspection for its high security and efficiency. It can also be combined with object detection technology to realize intelligent processing [4]. Besides, the use of computer vision and image processing technology for aerial images and automatic video processing can realize the automatic fault location of transmission line fittings, which significantly improves the efficiency of power maintenance.

With the successful application of deep convolutional neural network (DCNN) [5], object detection performance has improved significantly. At present, most of the existing fitting detectors are anchor-based [6]. Researchers are committed to improving the accuracy and efficiency of anchor-based methods. Although the stability of these algorithms has been improved, it has high requirements for hardware computing resources due to too many model parameters. In addition, the detection speed is slow, which cannot meet the needs of real-time detection. At the same time, the large size of the model also makes it unable to apply to the operating platform with relatively limited hardware resources.

To solve these problems, we put forward anchor-free detectors to improve the flexibility of the fitting detection. Anchor-free methods do not depend on the preset anchors

but adapt different fitting objects through regression. However, these detectors represented by CenterNet [7] also face some problems. First, they cannot achieve a good trade-off between efficiency and accuracy in practical applications. The accuracy of the faster network is ordinary, while the efficiency of the network with higher accuracy cannot meet real-time detection requirements. Secondly, the anchor-free detectors adopt a simple design and no feature fusion operation, which leads to the problem of mutual interference between objects of different scales widely, especially in the detectors using an hourglass network as the backbone. Finally, these detectors only generate a single-scale feature map. The insufficient feature extraction leads to low accuracy of small fittings (such as hanging board, u-type hanging ring) and those with occlusion (such as shockproof hammer and yoke plate). In addition, the low-resolution feature map also causes objects with huge scale gaps to be mixed into one feature map, so the accuracy is ordinary for complex aerial fitting images [8].

The performance of fitting object detection can be further improved by designing a novel backbone network and integrating the feature fusion method. Compared with anchor-free methods, anchor-based detection accuracy is better than the former, mainly due to the feature fusion network (FPN) [9]. FPN integrates the features of various scales to reduce feature loss in the upsampling process. Inspired by this idea, we propose a novel anchor-free method by introducing the lightweight backbone network and designing the feature fusion structure to CenterNet. Our method is called high-resolution MobileNet-based CenterNet (HRM-CenterNet), which can achieve better detection accuracy with high efficiency. Figure 1 shows the diagram of our method.

The main contributions of this paper are listed as follows. Firstly, it proposes the lightweight MobileNetV3 as the backbone to extract convolution features of different layers, reducing the size of the model and detection time-consuming while maintaining high accuracy. Secondly, it improves the lightweight feature receptive field enhancement module, which can expand the receptive field of high-layer features, enhance the feature expression ability of high-layer semantics, and strengthen its context information. Thirdly, it designs the high-resolution feature fusion network based on iterative aggregation. It can reduce the spatial features lost due to continuous down-sampling, maximizing the use of feature information extracted from the backbone network and generating heat maps more accurately.

We note that a shorter conference version of this manuscript appeared in Zhao et al. [10]. Our initial conference paper did not address the problem of HRM-CenterNet's effectiveness on mutual interference between fittings of different scales. This manuscript proposes the lightweight receptive field enhancement module to address this issue and provides additional performance comparison analysis with other state-of-the-art detectors. Our method effectively improves the fitting detection performance of CenterNet and achieves the best trade-off between accuracy and efficiency.

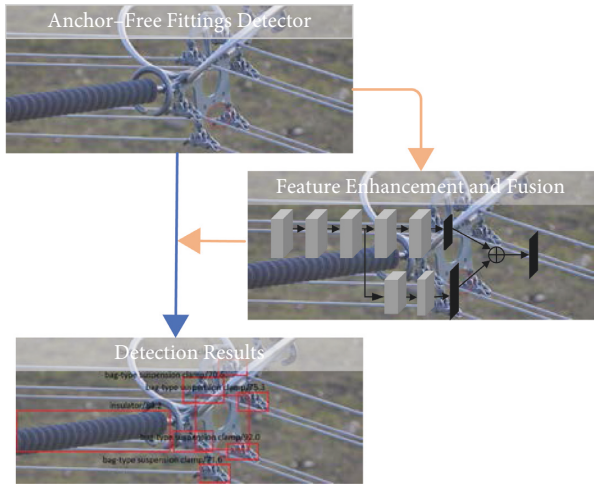


Figure 1: Diagram of our method.

2. Related Work

Object detection has always been the focus of research in computer vision, and it is also one of the difficulties in this field. Traditional detection technology uses artificially designed features, such as edge detection symbols, AdaBoost, or combines shallow features such as color, shape, and texture for recognition [11, 12]. These methods are rigid, with low accuracy and poor generalization ability. Thanks to the powerful feature extraction function of convolutional neural network(CNN), deep learning has gradually replaced traditional machine vision as the mainstream methods in image classification, object detection, and semantic segmentation [13–16]. Object detectors based on CNN can extract image features more effectively and perform end-to-end training. The existing methods can be divided into anchor-based detectors and anchor-free detectors.

2.1. Anchor-Based Detectors. The anchor-based detectors inherit the traditional idea of sliding window and region proposal strategy. Most detectors take multiple width-to-height ratio anchors as reference points for object positioning and then calculate the intersection over union (IOU) of bounding boxes between ground truth and prediction to select the most accurate one [17]. Whether there is a region proposal process can be divided into single-stage detectors and two-stage detectors.

2.1.1. Single Stage. The single-stage object detection algorithm arranges the possible bounding boxes on the image in a complex way and classifies them using the sliding window method without generating proposal regions. Moreover, feature maps are directly generated through the convolution network to predict category probability and position coordinates. Two typical single-stage detectors are YOLO [18] and SSD [19].

YOLO uses a single network for detection, which improves detection speed and has strong generalization ability. YOLOv2 [20] introduces the anchor box idea of Fast R-CNN

and uses the k-means clustering method to generate anchors. YOLOv3 [21] uses multi-label classification and cross-scale prediction methods, and it introduces Darknet-53 [22] as the feature extraction network to improve the detection accuracy of the model. YOLOv4 [23] uses CSPDarknet-53 as the backbone network, and the neck adopts SPP additional blocks and PANet [24] path aggregation blocks. It also achieves an excellent trade-off between detection speed and accuracy. SSD combines the regression idea of YOLO with the anchor mechanism of Faster R-CNN [25] to extract feature maps with different resolutions for detection [26].

2.1.2. Two Stage. In the two-stage detectors, the image is passed through a pretrained CNN to extract high-level features firstly. The region proposal network (RPN) is applied to attain two outputs: the probability that the region has an object and the coordinates of the bounding box. The RPN is trained to efficiently extract a predefined number ($k = 2000$) regions from images. R-CNN [8], one of the first successful detectors, selects the region of interest and then sends it to the subsequent convolutional neural networks (CNNs) for classification. Fast R-CNN [27] cuts the image features of different scales to reduce the amount of calculation and then sends the features to the classification and regression layer to calculate the final results. However, these two methods have the disadvantage of relying on manual feature extraction to obtain the region of interest. Fast R-CNN proposes the RPN, which replaces the traditional region of interest generation method. Mask R-CNN [28] uses ROIAlign instead of ROIPooling and adds mask branch and corresponding loss based on Fast R-CNN to achieve better classification performance. Cascade R-CNN [29] designs the cascade detection network and detects based on different IOU thresholds.

The anchor-based detectors improve the accuracy but bring many disadvantages, such as too many super-parameters and an imbalance of positive and negative samples. Besides, it needs much practical experience to design abundant anchors and assign them to specific objects. When the anchors use interest over union (IOU) as the evaluation criterion to determine the object, different IOU thresholds will significantly fluctuate the algorithm performance. The tuning processes of anchors are usually time-consuming and laborious. For complex transmission line inspection scenarios with diversified objects, the applicability of the anchor-based method is limited.

2.2. Anchor-Free Detectors. Anchor-free detectors use key point estimation for object detection. They do not need NMS postprocessing operation and preset anchors [30], which provide a new idea for achieving high-precision real-time fitting detection.

CornerNet [31] applies key points to object detection for the first time. It transforms the object position detection into detecting key points in the upper left corner and lower right corner of the object bounding box. The introduction of anchor-free approaches greatly simplifies the output, but at the same time, the accuracy is better than the one-stage anchor-based detectors. The ExtremeNet [32] detects four extreme points and one center point through a standard key point estimation network and groups the key points using the geometric relationship. A group of extreme points corresponds to a detection result. The CenterNet (key point triplets for object detection) [33] constructs triple key points based on CornerNet. Each object is represented by a center key point and a diagonal point. The CenterNet (objects as points) simplifies the problem of detecting paired key points into the estimation problem of center key points and uses the regression method to obtain objects' category, width, and height. The model is more concise and speeds up the overall speed of the algorithm [34]. CenterNet requires only one central point to locate the objects, which is one of the best anchor-free networks. FCOS [35] obtains the detection result of the input image by regressing the distance between the pixel and the left, top, right, and bottom edges of the object bounding box through the regression operation of the feature map pixel level. The pixel-by-pixel operation also brings the problem of slow detection speed.

In general, object detection by center point estimation can effectively adapt to the fitting object with variable size. In addition, it can meet the needs of real-time detection efficiency. However, the accuracy in complex fitting scenes is ordinary. To deal with that, we introduce feature receptive field enhancement and high-resolution methods to improve fitting detection accuracy.

2.3. Fitting Object Detectors. Driven by the wave of deep learning, transferring the detectors that perform well in general object detection to power field detection has become a research hotspot of power system recognition and detection. In addition, some of the early research used nondeep learning algorithms, which mainly use the color, gradient, and contour of the fitting object to recognize the metal tools. [36] uses saliency detection based on color and gradient features to locate insulators and then uses adaptive morphology to detect self-exploding defects. [37] applies F-PISA clustering to locate insulators based on color and structural characteristics and established a color model to identify damaged areas. [38] proposes GrabCut segmentation algorithm extract insulator contour and calculated the changes of gap and overhang of insulator string umbrella cap based on the convex defects of insulator contour, to quantitatively analyze the ice coating condition of insulators. [39] adopts multi-layer perceptron to detect the insulators and shock hammers based on the detected object's location correlation feature and local contour.

At present, the transmission line fitting detectors based on deep learning algorithm are mainly anchor-based method. Single-stage models such as SSD and Yolo series and two-stage model such as Faster R-CNN are classic algorithms. [40] applies the Faster R-CNN to the detection of grading ring and shockproof hammer. Aiming at the insulator detection task in aerial images, [41] proposes an object decomposition and aggregation algorithm based on YOLOv3. The above algorithms have achieved good results in fittings with a large proportion of grading ring, insulator,

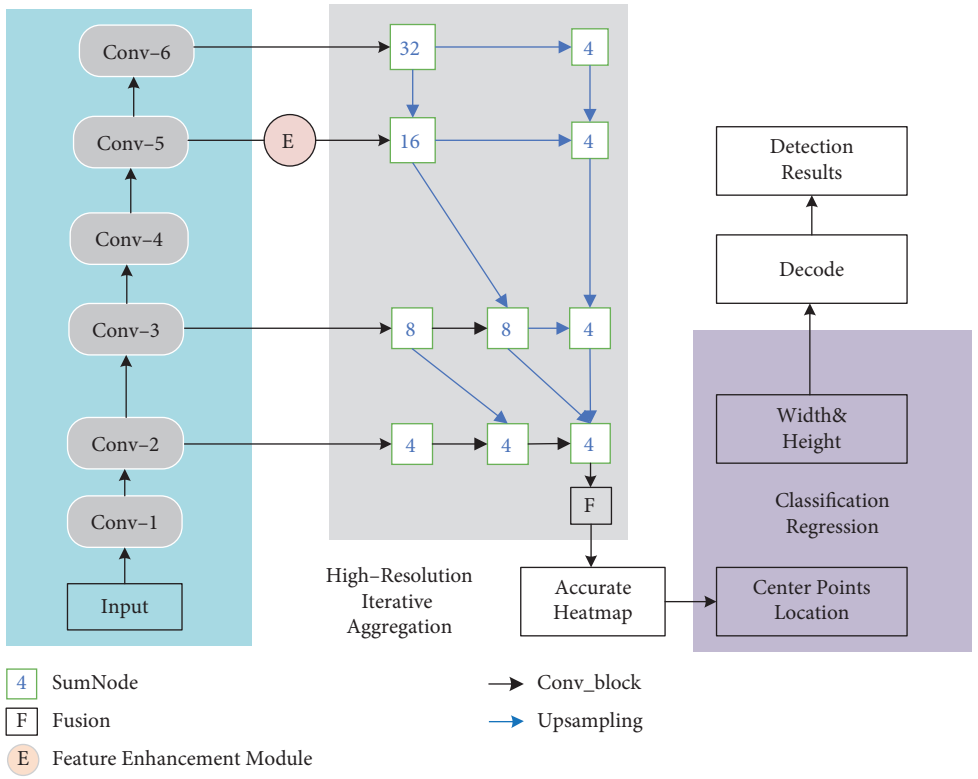


Figure 2: Structure of HRM-CenterNet.

and shockproof hammer in the inspection image and single shape. However, the complex environment of the transmission line leads to the changeable background of aerial images and multi-scale fitting objects. Hence, the detection effect of directly using the above algorithms is poor. [42] puts forward the description method of occlusion relationship between fitting objects and designs the occlusion relationship module to improve the accuracy of occlusion fittings. [43] improves the performance of SSD on dense occluded fittings by expanding the sample of metalware and introducing effective repulsion loss for dense object detection. [44] proposes a typical Faster R-CNN fitting detector combining KullbackLeibler divergence and shape constraints, which solves the problem of inaccurate object bounding box regression in fitting detection to a certain extent.

Recently, anchor-free detectors have been widely used in various industrial scenarios, including the electric power system. [45] adopts the CenterNet combined with structured positioning to realize the accurate identification and positioning of different substation equipment and components. It shows that the anchor-free method effectively improves the detection accuracy of power images and provides a new method for intelligent detection of fitting images. [46] uses DLANet backbone network, deep layer aggregation, sequence and exception module, and deformable convolution to design an efficient deep feature extraction network DLA-SE on the CenterNet to achieve the real-time detection of three common inspection faults: insulator self-explosion, shockproof hammer falling off, and bird's nest.

3. HRM-CenterNet Approach

The proposed HRM-CenterNet is an anchor-free detection method. We redesign MobileNetV3 [47] as the backbone of CenterNet. Built on M-CenterNet, the lightweight receptive field enhancement module and high-resolution feature fusion network are proposed to extract features better. Figure 2 shows the structure of HRM-CenterNet.

The key idea of the CenterNet is to predict the center point of the object through the key point heatmap and then regress the object's size, 3D position, and pose attributes from other feature maps corresponding to the key points. Compared with other anchor-free detectors, CenterNet locates the objects through only one center point with fewer parameters and a faster detection speed.

We assume that the input image is $I \in R^{W\times H\times 3}$, and W and H are the width and height of the fitting image, respectively. After passing through the backbone network, the key point heatmap $\widehat{Y} \in [0,1]^{W/R\times H/R\times c}$ is generated, R is the scale of the heatmap size, and c is the number of fittings. Then, three prediction branches are generated from the heatmap: the key point prediction branch is used to detect the key points of the heatmap and the object center point. The object size prediction branch generates a prediction bounding box based on the center point to detect the width

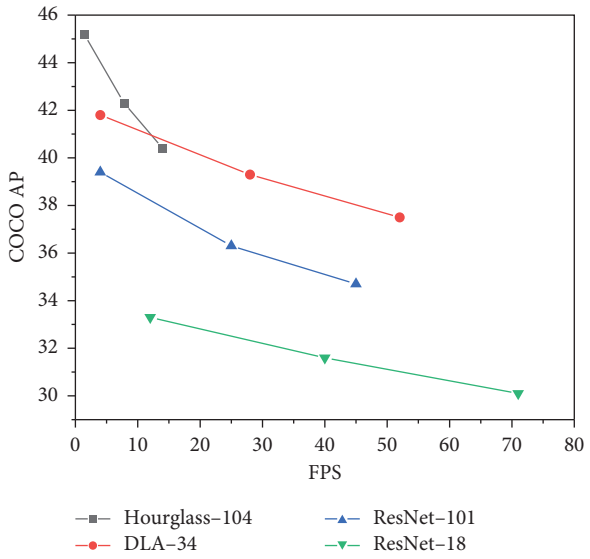


Figure 3: Performance comparison of origin backbone.

and height. The local offset prediction branch regresses the offset of the key point and the center point to detect the object accurately.

3.1. M-CenterNet. CenterNet adopts the design method of the backbone and downsampling path as a whole, which has robust scalability. The author uses four different backbone networks, but the performance has apparent differences. As shown in Figure 3, the high-speed network has general accuracy, while the high-precision network speed cannot meet the real-time requirements.

The transmission line inspection platform usually adopts a small-size edge-end device to ensure flexibility and portability. Such systems are relatively scarce of computing and storage resources, and they are more sensitive to the size and speed of the model. Choosing a lightweight network is the only way to achieve real-time detection. However, the original four backbone networks cannot effectively balance speed and accuracy, so it is necessary to use a novel backbone network that takes both into account.

The MobileNetV3 is a lightweight network focusing on mobile terminals and embedded devices. It improves the existing deep separable convolution (DSC) [48] inverted residuals and proposes a squeeze-and-excitation block (SE block). DSC decomposes the standard convolution operation into depthwise convolution and pointwise convolution, so the parameters are greatly reduced. The inverted residuals first expand the channels of the input feature map, then use DSC for downsampling to reduce the size of the feature map, and finally build the channels to enhance the expression ability of the model. SE block uses global pooling to generate channel statistical information for compression operation, compresses the global spatial information into a channel descriptor, passes through two fully connected layers, and finally uses sigmoid activation function for activation operation. Because of the advantages of MobileNetV3 in lightweight and speed, it is proposed as the backbone network of our model to extract useful features from fitting images. In this paper, MobileNetV3-large is selected as the feature extraction network for fitting detection.

In order to obtain more effective feature maps for high-precision detection, we have made further improvements to MobileNetV3. We remove the average pooling layer and three 1×1 convolution layers of the last bottleneck layer of the MobileNetV3 and then add three upsampling transposed convolution layers to restore the semantic and location image information. The feature maps generated by transposed convolution are sent to the three sub-networks of CenterNet for key point, offset, and size prediction. The network is called M-CenterNet, and the structure is shown in Figure 4.

3.2. Lightweight Receptive Field Enhancement Module. In the detection task of fittings, there are different scales and different kinds, and the same kind of fittings also has different scales. Because of the different distances and angles between aerial images and cameras, even the same fittings have large differences. It is difficult to solve mutual interference between fittings of different scales only by relying on single feature extraction and simple network design. In addition, in the top-down decoding and fusion feature process of M-CenterNet, the deep abstract features representing the semantic meaning of significant objects will be gradually diluted by the shallow representation information, thus losing the guiding spatial information.

Based on the idea of RFBNet [49], we propose the lightweight receptive field enhancement module (LRFEM) captures different levels of context information by designing multi-branch and multi-scale perforated convolution to enhance the robustness of global features.

The RFBNet simulates the relationship between the size and eccentricity of receptive fields (RFs) in the human visual system, enhances the feature extraction ability and robustness, and achieves high detection accuracy while considering the efficiency. It draws on the idea of the inception algorithm and introduces three dilated convolution layers, which effectively increases the receptive field of the network. The structure is shown in Figure 5.

The dilated convolution adds the dilation rate to each conventional convolution layer. The dilation rate determines the distance between pixels when the convolution kernel processes data, expands the kernel to the specified scale, and fills the unoccupied pixel area in the original kernel with 0. Therefore, the receptive field of dilated convolution will be improved compared with conventional convolution without increasing the amount of calculation. The calculation formula of dilated convolution receptive field is shown in formula (1): K is the receptive field of dilated convolution, rate is the dilation rate, and k is the size of the convolution kernel.

$$K = (\text{rate} - 1) \times (k - 1) + k. \tag{1}$$

In order to reduce the amount of calculation, RFB-s is improved based on RFB. The structure is shown in Figure 6(a). On the one hand, 3×3 convolution layer is used

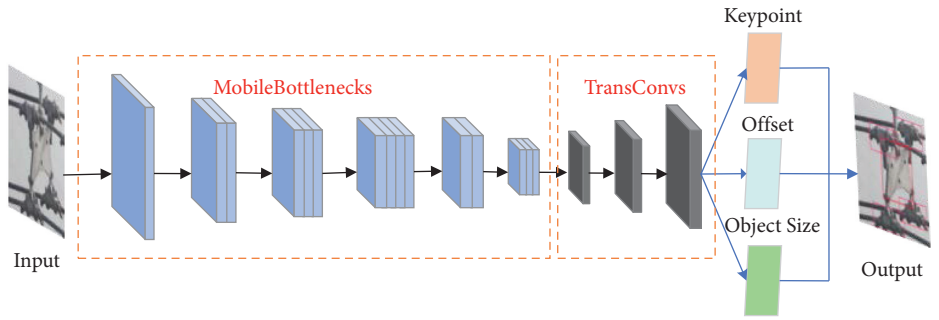


Figure 4: Structure of M-CenterNet.

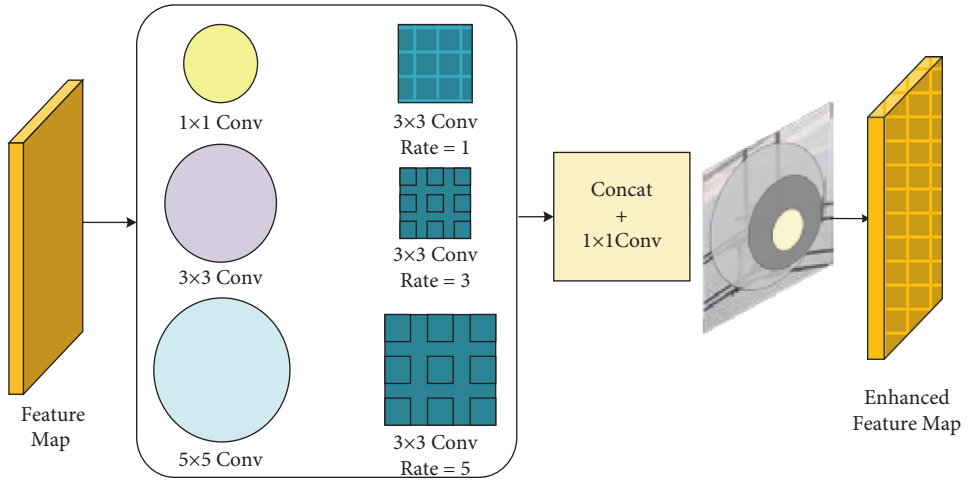


Figure 5: Structure of RFB.

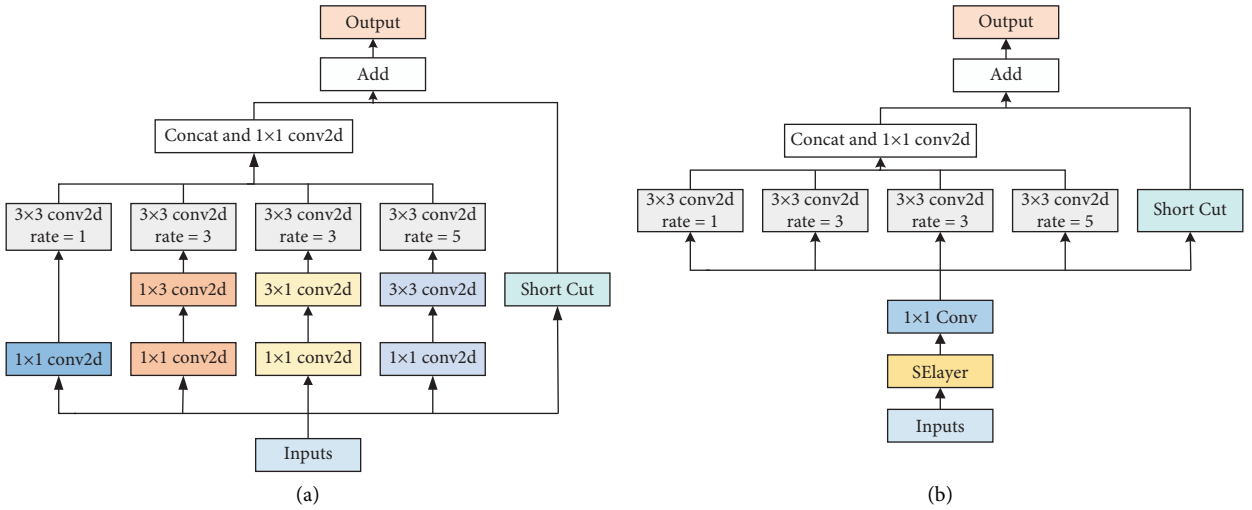


Figure 6: Structure comparison of RFB-s and our improved lightweight RFB module. (a) Structure of RFB-s. (b) Structure of improved LRFBM.

to replace 5×5 convolution layer, and on the other hand, 1×3 and 3×1 convolution layers are used to replace 3×3 convolution layer. Considering that MobileNetV3 has too many convolution layers, we appropriately delete some convolution layers in the RFB-s to reduce the calculation consumption of the network and avoid the feature map being too small. The structure is shown in Figure 6(b). The input feature map first passes through the squeeze-and-excitation layer (SE layer) to select the channel and then goes through the 1×1 convolution layer to reduce the dimension and connect four 3×3 dilated convolution layers in parallel. After passing through the 1×1 convolution layer, they are

integrated with the previous convolution layer for the second time. The results show that after the module deletes these convolution layers, the detection accuracy of the model is equivalent to that before, and the computational efficiency is significantly improved. In our work, to not excessively increase the amount of calculation, we mainly apply the LRFEM to the 16 times downsampling feature map.

3.3. High-Resolution Feature Fusion Method. The network usually connects the feature extraction module composed of convolution operation in series and generates a single-scale feature map among the anchor-free detectors. Although the low-resolution feature map output after multiple downsampling contains rich high-level semantic information, it also loses much spatial information.

For example, when ResNet18 is used as the backbone of CenterNet, the ResNet18 performs 32 times downsampling on the input image, resulting in two adjacent fittings in the original image which will become the same pixel or even disappearing. Therefore, it is difficult to achieve high-precision detection for small-size fittings and fittings with mutual occlusion.

In addition, the author only uses the largest feature map to generate the heatmap in the CenterNet, but the loss of image features leads to inaccurate generation. In order to make full use of the feature map generated after convolution operation, we propose a high-resolution feature fusion network based on iterative aggregation by referring to the high-resolution representation network.

HRNet [50] maintains high-resolution representation and gradually adds parallel subnets in feature extraction, using the feature maps extracted from the subnets for multi-scale feature fusion. It effectively utilizes feature maps with different resolutions and has a better prediction effect on heatmaps. Figure 7 shows the feature fusion method of the original HRNet. Firstly, the low-resolution feature map is sampled to the same scale as the high resolution; then, the four feature maps are concatenated. Compared with other networks characterized by aggregating layers by upsampling from low to high, HRNet has higher parameter efficiency and lighter weight.

Although the original HRNet utilized features with different resolutions, the fusion method was too simple to fully use features with different resolutions. Therefore, we designed the feature fusion method of iterative aggregation, as shown in Figure 8.

It integrates low-resolution and high-resolution features through iterative aggregation, maximizing the use of feature information extracted from the backbone network and generating heat maps more accurately. The formula of iterative aggregation is equation (2). $x_1, \ldots, x_n$ represents the aggregation node and is the input of the aggregation node.

$$I(x_1, \ldots, x_n) = \begin{cases} x_1, & \text{if } n = 1, \\ I(N(x_1, x_2), \ldots, x_n), & \text{otherwise.} \end{cases} \quad (2)$$

We remove the transpose convolution of the M-CenterNet and then select four feature maps obtained from 4

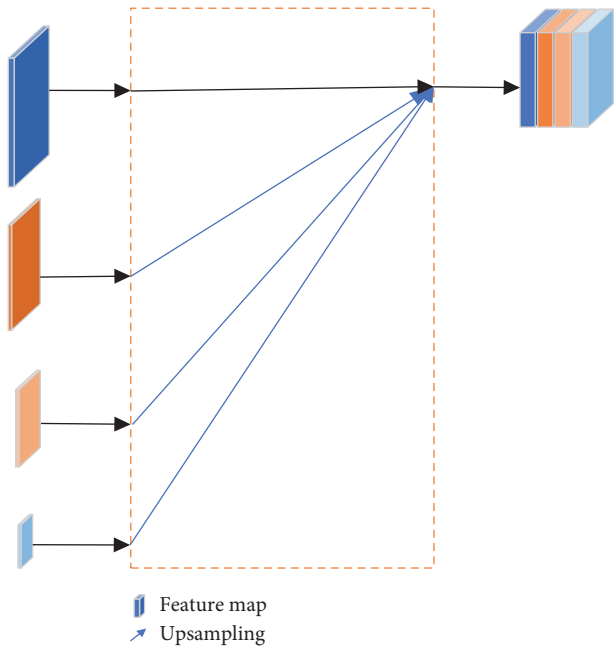


Figure 7: Feature fusion of original HRNet.

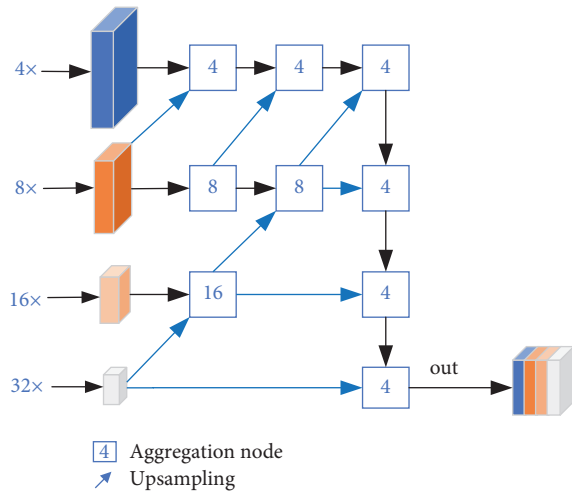


Figure 8: High-resolution feature fusion method.

times of downsampling feature map C2, 8 times of downsampling feature map C3, 16 times of downsampling feature map C5, and 32 times of downsampling feature map C6 for high-resolution feature fusion. The structure of HRM-CenterNet is shown in Figure 2. It increases the depth of the network and improves the learning ability of difficult samples. Due to the different sizes of the input feature maps of each aggregation node, we sample the low-resolution feature maps to the same size as the high-resolution feature maps through transpose convolution.

4. Experiments

In this section, we evaluate the performance of the HRM-CenterNet on TFITS and PASCAL VOC. The experiments

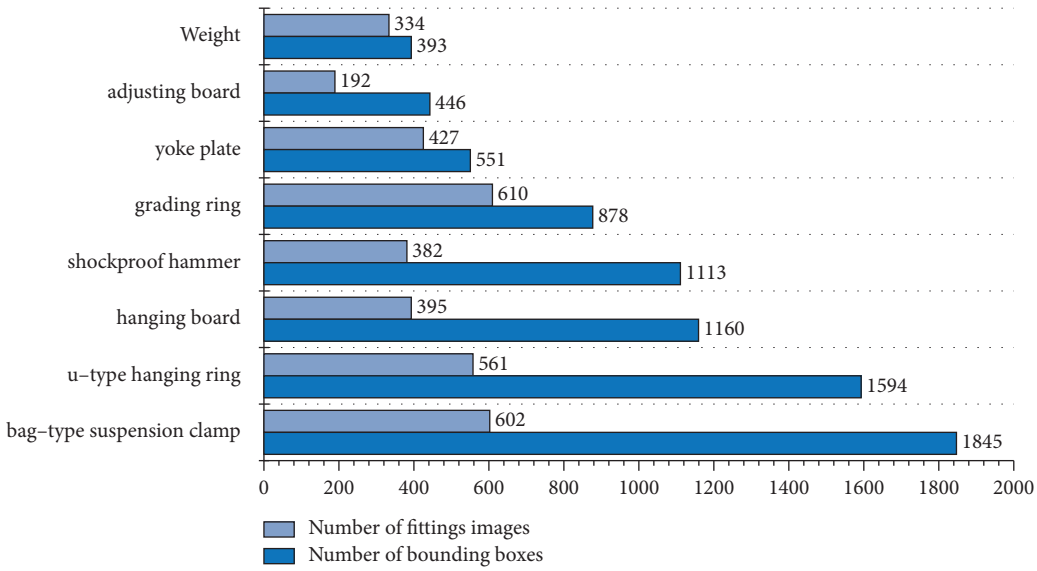


Figure 9: Quantity distribution of images and bounding boxes.

Figure 10: Aerial image of complex transmission lines in TFITS.

are implemented in the PyTorch on a machine with GeForce GTX1080Ti GPUs, CUDA 9.0, and cuDNN v7.

4.1. Implementation Details

4.1.1. Experiments on TFITS. Referring to the construction method of the PASCAL VOC dataset, we select the aerial images of UAV according to the requirements of the standard for shooting position, exposure degree, and focusing accuracy, and then construct the professional Typical Fittings Dataset (TFITS). The TFITS Dataset includes 8 types of fitting objects, consisting of 3503 images and 7980 boxes. The specific number of images and bounding boxes is shown in Figure 9. Some examples are shown in Figure 10.

We adopt the mini-batch stochastic gradient descent (SGD) momentum method to train HRM-CenterNet. The input images are uniformly scaled to 512×512. The initial learning rate is set to 1.25×10^{-4}, and the batch size is 8. The same learning rate is adopted for all layers. In the training process, when the detection accuracy of the verification set is no longer improved, the learning rate is reduced to 10% of the current learning rate by cosine annealing until the accuracy is no longer improved by adjusting the learning rate. It is trained for 140 epochs, and flip augmentation is used in testing.

4.1.2. Experiments on PASCAL VOC. The PASCAL VOC dataset contains 20 categories, which are quite different in object size, direction, posture, brightness, and occlusion position. We focus on the categories prone to false and missing detection, such as chairs, dining tables, and potted plants.

We experiment HRM-CenterNet in a small training resolution. The input images are 384×384, while all other hyper-parameters in the loss function are the same as the TFITS experiments. The network is trained on two GPUs with the batch size of 32. Adam optimizer is used with an initial learning rate of 1.25×10^{-4}. It is trained for 160 epochs, and flip augmentation is also used in testing.

4.2. Evaluation Index. In order to analyze the performance of the fitting detectors, we need to consider their accuracy and efficiency. At the same time, considering the limited

Table 1: Results on TFITS test set.

Methods	Backbone	mAP/%	FPR	FNR	Size/Mb	FPS
Single-stage detectors						
SSD [19]	ResNet101	74.1	28.4	23.5	90.6	13.1
YOLOv3 [21]	DarkNet53	75.0	26.5	20.5	59.6	20.0
YOLOv4 [23]	DarkNet53	78.6	23.1	19.2	30.0	25.0
RetinaNet [51]	ResNet101	79.1	22.3	19.3	65.3	10.2
Two-stage detectors						
Fast R-CNN [27]	ResNet101	77.1	22.8	18.4	98.0	8.8
Faster R-CNN [25]	ResNet101	80.2	20.0	15.3	122.0	7.0
Anchor-free detectors						
ExtremeNet [32]	Hourglass104	80.9	19.6	14.7	150.6	16.8
CornerNet [31]	Hourglass104	77.9	23.2	19.8	160.0	9.2
EfficientDet [52]	EffcientNetB0	74.0	26.1	20.8	80.6	47.0
CenterNet						
CenterNet-HG	Hourglass104	78.1	23.8	19.4	220.3	14.7
CenterNet-Res18	ResNet18	69.9	30.2	25.9	60.3	55.0
CenterNet-Res50	ResNet50	74.6	27.6	23.5	128.0	48.5
CenterNet-Res101	ResNet101	76.0	25.1	20.6	180.0	22.5
CenterNet-DLA34	DLA-34	76.4	24.0	20.1	77.0	27.2
M-CenterNet	MobileNetV3	75.9	25.3	20.9	18.5	43.2
HRM-CenterNet	MobileNetV3	80.3	19.8	13.6	24.6	32.5

HRM-CenterNet							
HRFF	LRFEM	Backbone	mAP/%	FPR	FNR	Size/Mb	FPS
		MobileNetV3	75.9	25.3	20.9	18.5	43.2
√		MobileNetV3	79.0	22.1	16.6	24.0	33.0
√	√	MobileNetV3	80.3	19.8	13.6	24.6	32.5

hardware devices of the edge, we also need to compare the computer memory occupied by the model. Therefore, we quantitatively evaluate the comprehensive performance of the detectors from the three dimensions: detection accuracy, efficiency, and model size.

We adopt the average accuracy, false detection rate, and missed detection rate as the evaluation indexes of the detection accuracy. Model parameters and size can better reflect the occupation of hardware computing and storage resources. The calculation methods of false and missed detection rates are shown in equations (3) and (4) The FP indicates that a nonexistent object is predicted. FN indicates that no existing object is predicted. TP indicates that the existing object is correctly predicted, and TN indicates that the algorithm correctly predicts the background.

$$\text{false detection rate} = \text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}}. \tag{3}$$

$$\text{missed detection rate} = \text{FNR} = \frac{\text{FN}}{\text{TP} + \text{FN}}. \tag{4}$$

4.3. Structure Experiments on TFITS Validation Set. We horizontally compare the performance of the detectors before and after improvement in this section. Table 1 shows the performance of different detectors. Table 2 shows the average accuracy of CenterNet-HG, M-CenterNet, and HRM-CenterNet.

It can be seen that our methods have improved the detection accuracy of each category. Among all kinds of fittings, the AP of the hanging board, u-type hanging ring, yoke plate, shockproof hammer, and adjusting board increases most obviously. Besides, the plate fittings such as yoke plate, adjusting board, and hanging board have large shape differences due to different shooting angles and are easily confused with tower materials. After adding the multiscale feature receptive field enhancement module, the network can significantly improve the accuracy of such fittings. In addition, some small fittings such as the hanging board, u-type hanging ring, and shockproof hammer are densely distributed in the image, resulting in a large number of occlusions among the fittings. The high-resolution feature fusion method can significantly improve the accuracy of such fittings.

We can see that the false detection rate and missed detection rate of our HRM-CenterNet are 19.8% and 13.6% in Table 1, which are better than before. Besides, compared with baseline M-CenterNet, the detection accuracy is improved by 4.4%. The false detection rate and the missed detection rate are lower by 5.5% and 7.3%, which shows that our method can improve the feature extraction ability to distinguish the fitting object and background better. Compared with the CenterNet-HG, which has the highest accuracy in the four original backbone networks, our method improves the accuracy by 2% and the efficiency by more than twice, meeting the requirements of real-time detection.

4.4. Ablation Experiments on TFITS Validation Set. The baseline for all experiments in this section is the basic M-CenterNet, which only introduces the MobileNetV3 as the backbone network. We explore the effectiveness of

Table 2: Precision comparison on TFITS.

Category	CenterNet-HG	M-CenterNet	HRM-CenterNet
Grading ring	88.3	86.3	89.3
Shockproof hammer	85.8	83.7	89.6
Bag-type suspension clamp	89.6	89.0	91.8
Yoke plate	70.8	68.4	73.9
u-type hanging ring	60.7	59.5	64.2
Hanging board	59.6	56.1	60.7
Adjusting board	78.3	76.2	80.4
Weight	92.0	88.6	92.2
mAP	78.1	75.9	80.3

introducing lightweight MobileNetV3, the lightweight receptive field enhancement module (LRFEM), and the high-resolution feature fusion (HRFF) network. The experiments are carried out on the TFITS Dataset.

Firstly, we verify the influence of different backbones on the performance of the detectors and the superiority of adopting MobileNetV3. The results in Table 1 show that the detection speed of the M-CenterNet network can reach 43.2FPS, far exceeding the CenterNet with Hourglass104 and DLA-34 as the backbone network. In addition, compared with the CenterNet with Hourglass104, DLA-34, ResNet101, ResNet50, and ResNet18 as the backbone, the size of M-CenterNet is reduced by 91.6%, 75.9%, 69.3%, 85.5%, and 89.7%, respectively, greatly reducing the occupation of computer hardware and storage resources. As a matter of fact, the detection efficiency of M-CenterNet has been slightly improved, although it brings a small decrease in accuracy. It achieves an excellent performance balance in speed and accuracy.

Secondly, to verify the influence of embedding the lightweight receptive field enhancement module (LRFEM) in different positions for the feature extraction ability, we design the ablation experiment of the HRM-CenterNet network without embedding LRFEM and with embedding LRFEM in different positions on performance. Table 3 shows the AP of plate fittings with large shape difference among yoke plate, adjusting board and hanging board, and the mAP of 8 types of fittings. The experimental results show that the module embedded in different positions all improves the ability of the model to capture different layers of context information. Performance improvement is the largest when the module is applied to 16 times downsampling feature map. When applied to the 32 times downsampling feature map, the receptive field of conventional convolution is large enough, so the accuracy is slightly improved after the module is embedded. In addition, as the number of layers embedded in the module deepens, the amount of calculation brought by the module increases gradually. Therefore, the module is applied to 16 times downsampling feature map in the HRM-CenterNet.

Thirdly, to verify the effectiveness of the lightweight receptive field enhancement module, we designed the performance experiment based on HRM-CenterNet using RFB module, RFB-s module, and our improved module, respectively. Table 4 shows that the amount of parameters is reduced by 40% compared with the RFB-s module after deleting part of the convolution layer. However, the detection accuracy is the same as before.

Table 3: The influence of LRFEM embedding position.

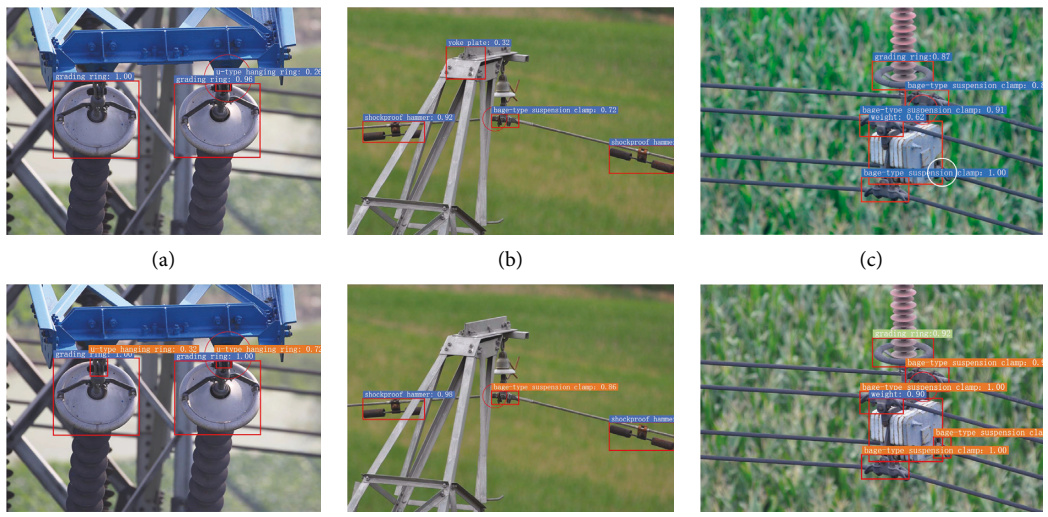
Position	AP/%			mAP
	Yoke plate	Adjusting board	Hanging board	
No LRFEM	69.4	77.2	57.2	79.0
With 4X	69.8	77.4	57.2	79.3
With 8X	71.3	79.5	58.9	79.7
With 16X	73.9	80.4	60.7	80.3
With 32X	73.5	80.0	58.6	79.8

Table 4: The influence of different RFB modules.

Methods	mAP/%	Parameter/M
RFB block	80.0	0.7
RFB-s block	80.3	1.0
Ours	80.3	0.6

Then, the results of ablation experiments for HRFF and LRFEM are shown in the last column of Table 1. When both methods are introduced, the detection accuracy of the fittings reaches the best. The detection accuracy is improved by 1.3% and 3.1% for LRFEM and HRFF, respectively, and it shows that the latter improves the accuracy of the model more significantly than the former.

Finally, we conduct comparative visual experiments. Figure 11 shows qualitative examples of HRM-CenterNet for fitting detection on the TFITS validation set. The above is the detection result of CenterNet, and the below is the result of HRM-CenterNet. False and missed detections have been marked with white boxes. For example, although the grading ring is detected in the CenterNet in Figure 11(a), the u-type hanging ring on the left grading ring was missed. In Figure 11(b), the tower material is mistakenly detected as a yoke plate. The bag-type suspension clamp blocked by the weight in the right corner in Figure 11(c) is not detected by the CenterNet. In Figures 11(d)–11(f), the detection results of HRM-CenterNet embedded with high-resolution feature fusion and multi-scale receptive field enhancement module have greatly improved these problems. Small fitting u-type hanging ring and bag-type suspension clamp due to dense occlusion are detected. In addition, the confidence of the bounding boxes has been significantly improved, representing the improvement of the feature extraction ability of



(a) (b) (c)

(d) (e) (f)

Figure 11: Qualitative examples of HRM-CenterNet. The above represents the detection results of CenterNet, and the below represents the detection results of HRM-CenterNet.

Table 5: The validation results on the PASCAL VOC dataset.

Methods	AP/%				mAP	FPS
	Potted plants	Cups	Chairs	Dining tables		
CenterNet-Res18	49.6	63.1	59.2	72.8	75.7	96.0
CenterNet-Res101	54.0	67.3	63.2	75.6	78.7	27.0
CenterNet-DLA	58.2	71.4	65.6	77.9	80.1	30.0
CenterNet-HG	58.7	71.5	67.2	78.2	81.5	9.5
M-CenterNet	53.1	66.4	62.0	76.6	78.2	35.5
HRM-CenterNet	59.3	71.6	67.1	79.1	81.6	26.0

the detectors. The experiments show that HRM-CenterNet reduces the probability of false and missed detection and verifies the effectiveness of the innovations.

4.5. Comparisons with Other Approaches. In order to compare with other state-of-the-art approaches, we trained our HRM-CenterNet in the TFITS Dataset and then submitted the results to the leaderboards. The results are shown in Table 1. The M-CenterNet and HRM-CenterNet have the smallest size, greatly reducing computer hardware and storage resources. For detection efficiency, although the HRM-CenterNet is lower than EfficientDet, it is 6.3% higher than EfficientDet in accuracy, and both of them have real-time performance. For detection accuracy, HRM-CenterNet ranks second, but its detection efficiency is more than twice that of rank 1 ExtremeNet. Although our HRM-CenterNet cannot achieve the best performance in every category, it achieves the best speed-accuracy trade-off among all the detectors shown in Table 1.

4.6. Experiments on PASCAL VOC 2007. We further compare our HRM-CenterNet with different backbone networks on PASCAL VOC 2007. The results are shown in Table 5. We selected four categories prone to missed and false detection, such as potted plants and chairs. The potted plants have large-scale changes and many small-scale targets, but the performance of HRM-CenterNet is much higher than that of M-CenterNet. The accuracy is improved by 6.3%, indicating that the LRFEM can better retain spatial semantic information, and the iterative aggregation feature fusion can make full use of this retained information. Compared with chairs, cups, dining tables, and other categories vulnerable to occlusion, HRM-CenterNet is superior to other algorithms. It shows that our network has a stronger feature extraction ability and good robustness. In summary, our HRM-CemterNet also achieves the best trade-off between accuracy and efficiency on PASCAL VOC 2007 dataset.

5. Conclusion

In conclusion, aiming to improve the performance of the anchor-free detectors for fitting object detection, we introduce feature enhancement and iterative aggregation to CenterNet. The detection accuracy is efficiently improved for introducing lightweight feature enhancement modules and high-resolution feature fusion, verified by the experimental results on the TFITS Dataset. The results on the TFITS and PASCAL VOC 2007 demonstrate that our HRM-CenterNet achieves the best speed-accuracy trade-off. Besides, HRM-CenterNet is suitable for deployment on the outdoor inspection platform. In further research, we will optimize the

detection performance of difficult samples. The next step of the research will be to extract richer features, smaller structures, and more efficient methods and then make further model performance improvements in embedded devices to achieve real-time fitting detection on the mobile terminal.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Science and Technology Project of State Grid Jibei Power Company Limited (no. B3018E200000).

References

[1] J. Toth and A. G. Jackson, "Smart view for a smart grid—unmanned Aerial Vehicles for transmission lines," in *Proceedings of the International conference on applied robotics for the power industry*, pp. 1–6, IEEE, Montreal, Canada, October 2010.

[2] W. Tong, J. Yuan, and B. Li, "Application of image processing in patrol inspection of overhead transmission line by helicopter," *Power System Technology*, vol. 34, no. 12, pp. 204–208, 2010.

[3] V. N. Nguyen, R. Jenssen, and D. Roverso, "Automatic autonomous vision-based power line inspection: a review of current status and the potential role of deep learning," *International Journal of Electrical Power & Energy Systems*, vol. 99, no. 2, pp. 107–120, 2018.

[4] Z. Zhao, H. Qi, Y. Qi, K. Zhang, Y. Zhai, and W. Zhao, "Detection method based on automatic visual shape clustering for pin-missing defect in transmission lines," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 9, pp. 6080–6091, 2020.

[5] A. Krizhevsky, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 54, no. 6, pp. 1097–1105, 2012.

[6] S. Fan, F. Zhu, S. Chen et al., "FII-CenterNet: an anchor-free detector with foreground attention for traffic object detection," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 1, pp. 121–132, 2021.

[7] X. Zhou, D. Wang, and P. Krähenbühl, "Objects as points," 2019, https://arxiv.org/abs/1904.07850.

[8] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 580–587, IEEE, Columbia, Oh, USA, June 2014.

[9] H. Kuang, B. Wang, J. An, M. Zhang, and Z. Zhang, "Voxel-FPN: multi-scale voxel feature aggregation for 3D object detection from LIDAR point clouds," *Sensors*, vol. 20, no. 3, pp. 704–710, 2020.

[10] K. Zhang, K. Zhao, and X. H. Feng, "HRM-CenterNet: a high-resolution real-time fittings detection method," in *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 564–569, IEEE, Melbourne, Australia, October 2021.

[11] D. Hang, X. Qiu, and C. Cao, "Algorithm of vibration damper detection combined with aggregation channel and complex frequency domain features," *Computer Technology and Development*, vol. 36, no. 3, pp. 147–151, 2020.

[12] S. Han, R. Hao, and J. Lee, "Inspection of insulators on high-voltage power transmission lines," *IEEE Transactions on Power Delivery*, vol. 24, no. 4, pp. 2319–2327, 2009.

[13] Q. Zhou, X. Wu, S. Zhang, B. Kang, Z. Ge, and L. Jan Latecki, "Contextual ensemble network for semantic segmentation," *Pattern Recognition*, vol. 122, pp. 108290–108301, 2022.

[14] Q. Zhou, Y. Wang, Y. Fan et al., "AGLNet: towards real-time semantic segmentation of Self-driving images via attention-guided lightweight network," *Applied Soft Computing*, vol. 96, no. 11, pp. 106682–106694, 2020.

[15] Q. Zhou, W. Yang, G. Gao et al., "Multi-scale deep context convolutional neural networks for semantic segmentation," *World Wide Web*, vol. 22, no. 2, pp. 555–570, 2019.

[16] Q. Zhou, J. Wang, and J. Liu, "RSANet: towards real-time object detection with residual semantic-guided attention feature pyramid network," *Mobile Networks and Applications*, vol. 26, no. 4, pp. 77–87, 2021.

[17] Z. Zheng, P. Wang, and W. Liu, "Distance-IoU loss: faster and better learning for bounding box regression," in *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 12993–13000, New York, NY, USA, February 2020.

[18] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779–788, IEEE, Las Vegas, NV, USA, June 2016.

[19] W. Liu, D. Anguelov, and D. Erhan, "SSD: single shot multibox detector," in *Proceedings of the European conference on computer vision*, pp. 21–37, Springer, Amsterdam, Netherlands, October 2016.

[20] J. Redmon and A. Farhadi, "YOLO9000: better, faster, stronger," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7263–7271, IEEE, Hawaii, Hl, USA, July 2017.

[21] J. Redmon and A. Farhadi, "Yolov3: an incremental improvement," 2018, https://arxiv.org/abs/1804.02767.

[22] B. Benjdira, T. Khursheed, and A. Koubaa, "Car detection using unmanned aerial vehicles: comparison between Faster R-CNN and Yolov3," in *Proceedings of the 2019 1st International Conference on Unmanned Vehicle Systems-Oman (UVS)*, pp. 1–6, Muscat, Oman, February 2019.

[23] A. Bochkovskiy, C. Wang, and H. Liao, "Yolov4: optimal speed and accuracy of object detection," 2020, https://arxiv.org/abs/2004.10934.

[24] S. Liu, L. Qi, and H. Qin, "Path aggregation network for instance segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 8759–8768, IEEE, Salt Lake City, UT USA, June 2018.

[25] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R.-C. N. N.: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2016.

[26] C. Fu, W. Liu, and A. Ranga, "DSSD: deconvolutional single shot detector," 2017, https://arxiv.org/abs/1701.06659.

[27] R. Girshick, "Fast R-CNN," in *Proceedings of the IEEE international conference on computer vision*, pp. 1440–1448, IEEE, Santiago, Chile, May 2015.

[28] K. He and G. Gkioxari, "Mask R-CNN," in *Proceedings of the IEEE international conference on computer vision*, pp. 386–397, IEEE, Venice, Italy, October 2020.

[29] Z. Cai and N. Vasconcelos, "Cascade R-CNN: delving into high-quality object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 6154–6162, Salt Lake City, UT, USA, June 2018.

[30] C. Zhu, Y. He, and M. Savvides, "Feature selective anchor-free module for single-shot object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 840–849, Long Beach California, USA, June 2019.

[31] H. Law and J. Deng, "CornerNet: detecting objects as paired keypoints," in *Proceedings of the European conference on computer vision*, pp. 734–750, Munich, Germany, September 2018.

[32] X. Zhou, J. Zhuo, and P. Krahenbuhl, "Bottom-up object detection by grouping extreme and center points," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 850–859, California, USA, June 2019.

[33] K. Duan and S. Bai, "CenterNet: keypoint triplets for object detection," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 6569–6578, Seoul, Korea, October 2019.

[34] I. Ahmed, M. Ahmad, J. J. P. C. Rodrigues, and G. Jeon, "Edge computing-based person detection system for top view surveillance: using CenterNet with transfer learning," *Applied Soft Computing*, vol. 107, no. 6, pp. 107–114, 2021.

[35] Z. Tian, C. Shen, H. Chen, and T. He, "FCOS: fully convolutional one-stage object detection," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 9627–9636, Seoul, Korea, October 2019.

[36] Y. Zhai, D. Wang, M. Zhang, J. Wang, and F. Guo, "Fault detection of insulator based on saliency and adaptive morphology," *Multimedia Tools and Applications*, vol. 76, no. 9, pp. 12051–12064, 2017.

[37] Y. Zhai, H. Cheng, R. Chen, Q. Yang, and X. Li, "Multi-saliency aggregation-based approach for insulator flashover fault detection using aerial images," *Energies*, vol. 11, no. 2, pp. 340–349, 2018.

[38] Y. Hao, J. Wei, X. Jiang et al., "Icing condition assessment of in-service glass insulators based on graphical shed spacing and graphical shed overhang," *Energies*, vol. 11, no. 2, pp. 318–328, 2018.

[39] Y. Liu, J. Li, and W. Xu, "A method on recognizing transmission line structure based on multi-level perception," in *Proceedings of the International Conference on Image and Graphics*, pp. 512–522, Shanghai, China, September 2017.

[40] Y. Tang, J. Han, and W. Wei, "Research on part recognition and defect detection of transmission line in deep learning," *Electronic Measurement Technology*, vol. 41, no. 6, pp. 60–65, 2018.

[41] Q. Gao and Q. Lian, "Research on target detection of the insulator in the aerial image," *Electrical Measurement & Instrumentation*, vol. 21, no. 6, pp. 119–123, 2019.

[42] Z. Zhao, A. Jiang, and Y. Qi, "Fittings detection of transmission line image based on SSD model embedded with occlusion relation module," *Journal of Intelligent Systems*, vol. 14, no. 8, pp. 343–348, 2020.

[43] Y. Qi, A. Jiang, and Z. Zhao, "Fittings detection method in patrol images of transmission line based on improved SSD," *Electrical Measurement & Instrumentation*, vol. 15, no. 4, pp. 656–662, 2020.

[44] Z. Zhao and Y. Li, "Typical fittings detection method with faster R-CNN combining KL divergence and shape constraints," *High Voltage Engineering*, vol. 46, no. 9, pp. 123–125, 2020.

[45] R. Zhao and G. Zhao, "A real-time fault detection method for high voltage transmission line based on CenterNet Improved Algorithm," *Computer Engineering and Applications*, vol. 34, no. 8, pp. 34–37, 2020.

[46] M. Zhao and Y. Lu, "Object detection technology in the complex environment based on CenterNet algorithm," *Journal of China Academy of Electronics*, vol. 64, no. 6, pp. 654–660, 2021.

[47] A. Howard, M. Sandler, and G. Chu, "Searching for mobilenetv3," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1314–1324, Seoul, Korea, October 2019.

[48] F. Chollet, "Xception: deep learning with depthwise separable convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1251–1258, Hawaii, Hl, USA, July 2017.

[49] S. Liu and D. Huang, "Receptive field block net for accurate and fast object detection," in *Proceedings of the European Conference on Computer Vision*, pp. 385–400, Munich, Germany, September 2018.

[50] K. Sun, B. Xiao, and D. Liu, "Deep high-resolution representation learning for human pose estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5693–5703, California, USA, June 2019.

[51] T. Lin, P. Goyal, and R. Girshick, "Focal loss for dense object detection," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2980–2988, Venice, Italy, October 2017.

[52] M. Tan, R. Pang, and Q. Le, "Efficientdet: scalable and efficient object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 10781–10790, Seattle, WA USA, June 2020.

Hindawi
Discrete Dynamics in Nature and Society
Volume 2022, Article ID 3954598, 17 pages
https://doi.org/10.1155/2022/3954598

Research Article

Modelling the Impact of Media-Induced Social Distancing on the Containment of COVID-19 in Beijing

Yinjiao Gong,[1] **Aili Wang**,[1] **Jin Guo**,[1] **and Stacey R. Smith?**[2]

[1]*School of Mathematics and Information Science, Baoji University of Arts and Sciences, Baoji 721013, China*
[2]*Department of Mathematics and Faculty of Medicine, the University of Ottawa, Ottawa, ON K1N 6N5, Canada*

Correspondence should be addressed to Aili Wang; aily_wang83@163.com

Received 4 January 2022; Revised 23 February 2022; Accepted 18 March 2022; Published 20 April 2022

Academic Editor: Jorge E. Macias-Diaz

With the multiple waves of COVID-19 in China and other countries, there is an urgent need to design effective containment, especially nonpharmaceutical interventions, to combat the transmission. Media reports on COVID-19—which can induce precautionary behaviour such as social distancing, by providing disease-related information to the public—are thought to be effective in containing the spread. We include the media-reporting data collected from authoritative and popular websites, along with the corresponding IP-visiting data, to study the effects of media reports in curbing the outbreak of COVID-19 in Beijing. To quantify how social distancing affects the spread of COVID-19, we differentiate the fully susceptible from those susceptibles who are media aware and practice social distancing or are quarantined. We propose a discrete compartment model with the fully susceptible, the media-aware susceptible, and the quarantined susceptible as three separate classes. We adopt functions dependent on the media reports and the contacts of media-aware susceptibles to describe the progression rate of susceptibles to media-aware susceptibles. By fitting the targeted model to data on the two Beijing outbreaks, we estimated the reproduction numbers for the two outbreaks as $R_0 = 1.6818$ and $R_0 = 1.3251$, respectively. Cross-correlation analysis on our collected data suggests a strong correlation between the media reporting and epidemic case data. Sensitivity and uncertainty analysis show that even with the intensified interventions in force, reducing either the social distancing uptake rate or the average duration of social distancing for media-aware susceptibles could aggravate the severity of the two outbreaks in Beijing by magnifying the final confirmed cases and lengthening the end time of the pandemic. Our findings demonstrate that enhancing social distancing and media reporting alone, if done in sufficient measures, are enough to alleviate the COVID-19 epidemic.

1. Introduction

The COVID-19 pandemic is highly transmittable, with specific features, including the emergence of multiple variants, incubation period, and asymptomatic carriers, that differentiate it from other coronaviruses, such as SARS and MERS. Although the epidemic in China was essentially controlled in the early stage, many cities in China, including Guangzhou, Nanjing, and Nei Monggol, have experienced a second epidemic wave [1, 2]. This is also true for other countries and regions, where new outbreaks or multiple epidemic waves occurred despite quarantine and social-distancing policies. The number of confirmed cases has continued to grow, with 264,815,815 confirmed cases and 5,249,793 deaths as of 7 December, 2021 [3]. Controlling the spread of COVID-19 is of vital importance for the whole world. Even with an effective vaccine or specific antivirals, nonpharmaceutical interventions are integral in reducing the probability of contracting the disease for an individual and avoiding new waves for a country or a region [4].

Human behaviour plays a vital role in the transmission of epidemics [5, 6], and an important factor influencing behavior is the amount of attention an individual places on the virus [7]. A major factor influencing such attention is media awareness [8]. In case of COVID-19, individuals are

media aware if (a) they receive virus-related information via media and (b) trust the information. Such individuals are more likely to practice social distancing and take precautions to remove themselves as much as possible from exposure to the virus [9–15]. Social-distancing practices may range from moderate to extreme total isolation and is one of the main reasons why the spread of COVID-19 in China is under control. However, some key questions remain about social distancing: specifically, how it alters the contact patterns of the public and how it interacts with other strict containment measures to reduce the transmission. It is thus necessary to understand how social distancing may help in curbing the transmission of COVID-19 and how it might be affected by the media.

Many mathematical models have been presented to describe the dynamics of the evolution of COVID-19 since the outbreak of COVID-19 [16–33]. Wu et al. [16–18] estimated the size of the epidemic in Wuhan based on the initial data. Gatto et al. [19–21] examined the effects of nonpharmaceutical interventions (NPIs) by modelling the unfolding epidemic with the laboratory surveillance data, illustrating the critical contribution of NPIs in reducing transmission. Laxminarayan et al. [22–26] revealed the features as well as transmission pathways of COVID-19 and identified the key factors affecting the contact pattern and the total number of infections. Zhu et al. [27–29] designed dynamic models to mimic the data on the epidemic and hospital bed capacities, studying the role of hospital capacity in helping curb the outbreak in certain regions. Khajanchi et al. [31–33] proposed compartmental models to explain the transmission dynamics of COVID-19 and forecast the case numbers in India. Although these studies contribute understanding COVID-19 transmission, the impact of media reports on the pandemic has not been quantified.

Some mathematical models have been proposed to assess the effects of mass media on the containment of other infectious diseases by decreasing the contacts of susceptible individuals [14, 15, 34, 35]. Heffernan et al. [14] formulated a stochastic agent-based model by including social distancing levels in their modelling, resulting in three susceptible compartments corresponding to three different social-distancing levels. They quantified the effects of mass-media reports in the 2009 H1N1 pandemic using this model and found that the report rate affected the variability surrounding public-health interventions. Rai et al. [15] formulated a deterministic compartment model by setting the media-aware susceptibles as a separate class that cannot contract the disease in order to mimic the COVID-19 outbreak in India. However, quantifying the variability in the contact rates of media-aware susceptibles as the media reports vary has been ignored in these studies, and no media-reporting data was used in the analysis or to calibrate the proposed model. Zhou et al. and Guo et al. [34, 35] considered the variability in the contacts of the population by adopting media-dependent contact rates and including media reporting data in the parameterization of the targeted model for the COVID-19 outbreak. However, they did not reflect the difference in the degrees to which different individuals undertake social distancing. To quantitatively assess the media effect on the contacts and illustrate the different degrees of social distancing induced by media reports, we use a compartment model and take the two COVID-19 outbreaks in Beijing as case study.

We present a discrete dynamic model in order to quantify the effect of media-induced social distancing on curbing the transmission of COVID-19. We use the model to mimic the data of the two outbreaks in Beijing: the outbreak from 20 January to 28 April, 2020, and the outbreak from 11 June to 3 September, 2020. We introduce two functions in terms of media reports, $\kappa M(t)$ and $c_{\min} + (c - c_{\min})e^{-\kappa M(t)(t-t_0)}$, to describe the effect of media-induced social distancing. We quantify the effect of social-distancing using two parameters: the social-distancing uptake rate κ and the average time $1/\lambda_f$ that susceptibles practice social distancing.

2. Methods

2.1. Data Collection and Analysis. We collected the data of COVID-19 cases from 20 January to 28 April 2020 from Beijing Municipal Health Commission. The data include the cumulative number of confirmed cases and cumulative number of recovered cases, as shown in Figure 1(a).

As of 28 April 2020, the cumulative number of confirmed cases climbed to 415, and the cumulative number of recovered cases increased to 406. No newly confirmed cases were reported until 11 Jun 2020. Therefore, we also collected the data of COVID-19 cases in Beijing from 11 June to 3 September 2020, as shown in Figure 1(e). The cumulative number of confirmed cases was 335, and the cumulative number of recovered cases was 335 between 11 June and 3 September 2020. For convenience, the outbreak between 20 January and 28 April is referred as the first outbreak and the outbreak between 11 June and 3 September is referred as the second outbreak in the rest of this work. We also obtained the daily number of media reports on COVID-19 from four authoritative and popular websites, including xinhuanet.com, huanqiu.com, cnr.cn, and news.sina.com.cn during the first outbreak, using the key word "COVID-19" included in the title or the full text, as shown in Figure 1(b). In addition, we obtained the daily number of IP addresses that visited each website, as shown in Figure 1(c). We similarly obtained the daily number of media reports on COVID-19 from seven websites, including china.com.cn, people.com.cn, xinhuanet.com, news.sina.com.cn, cyol.com, chinanews.com, and gmw.cn during the second outbreak, using the key words "The epidemic in Beijing" included in the title or the full text, as shown in Figure 1(f). We collected the daily number of IP addresses that visited these seven websites from 11 June to 3 September 2020, as shown in Figure 1(g). Finally, we calculated the average daily number of media reports for the first outbreak by defining an index "hotness" as the weighting coefficients to average the number of media reports collected in the first four websites, as shown in Figure 1(d). If the daily number of media reports collected in each website is

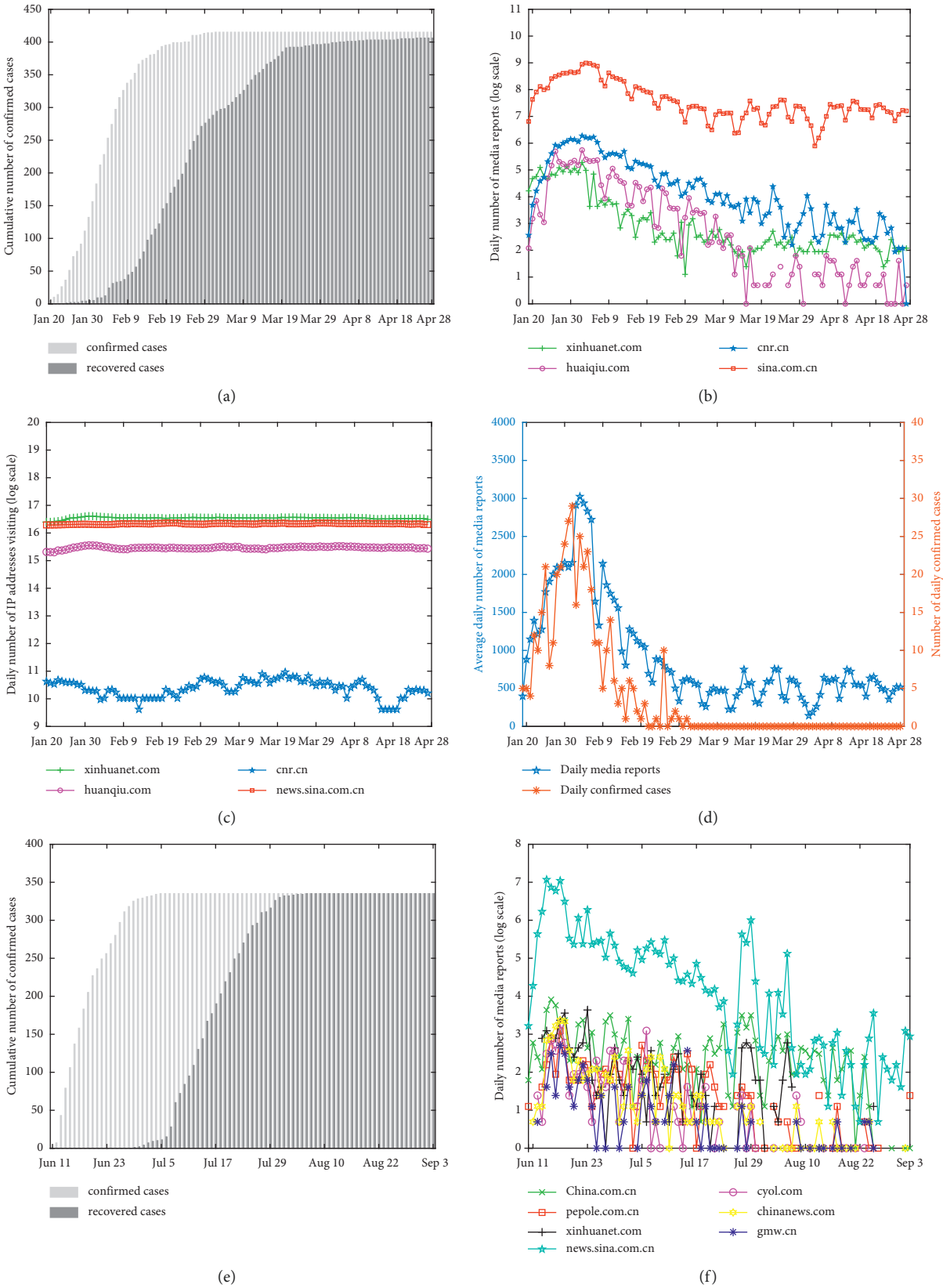
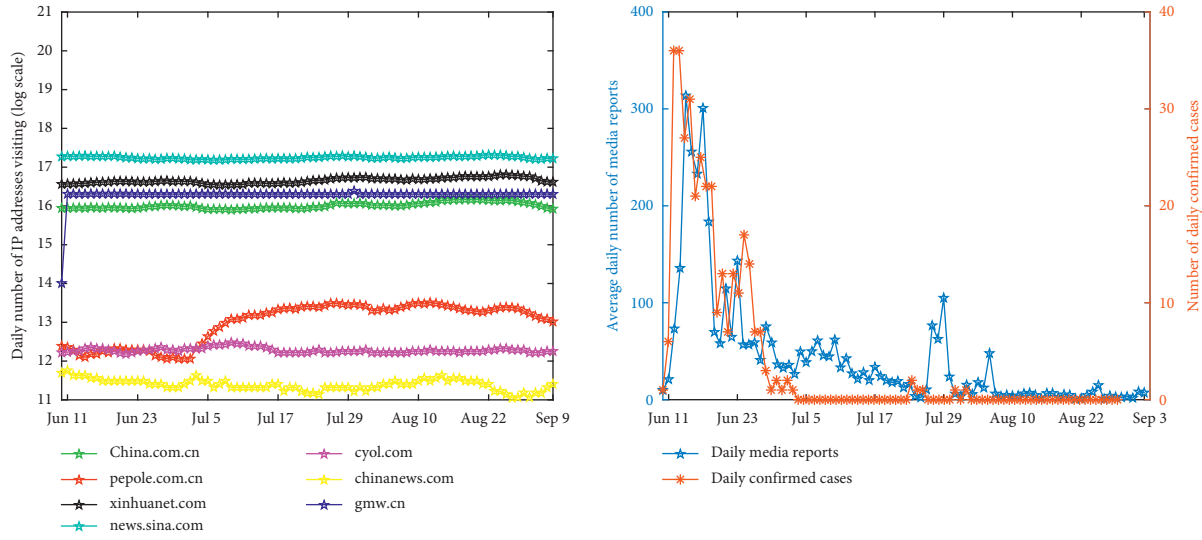


Figure 1: Continued.



(g) (h)

FIGURE 1: Cumulative number of confirmed cases and recovered cases from 20 January to 28 April 2020 (a) and from 11 June to 3 September 2020 (e); Number of daily media reports (b) and IP addresses (c) visiting xinhuanet.com, huanqiu.com, cnr.cn, and news.sina.com.cn, respectively; Number of daily media reports (f) and IP addresses (g) visiting China.com.cn, people.com.cn, xinhuanet.com, news.sina.com.cn, cyol.com, chinanews.com, and gmw.cn, respectively. Number of daily confirmed cases and the average daily number of media reports from 20 January to 28 April (d) and from 11 June to 3 September (h).

denoted by x_i and the daily number of IP addresses visiting the corresponding website by p_i, $i = 1, 2, 3, 4$, representing xinhuanet.com, huanqiu.com, cnr.cn, and news.sina.com.cn, respectively, then the average daily number of media reports x takes the form

$$x = h_i x_i, \tag{1}$$

where $h_i = p_i / \sum_{i=1}^{4} p_i$ represents the index hotness of the first four websites, respectively. By the same method, we defined the average daily number of media reports for the second outbreak in Beijing based on the last seven websites, as shown in Figure 1(h).

2.2. Cross-Correlation Coefficients. It follows directly from Figure 1(d) and 1(h) that the average number of daily media reports relates closely to the number of daily confirmed cases. The cross-correlation analysis method is adopted to provide qualitative insights on the causal temporal interaction between the daily number of media reports and the number of daily confirmed cases in the two outbreaks in Beijing. Let x_{1i} denote the number of daily media reports and y_1 denote the number of daily confirmed cases from 20 January to 28 April 2020, where $i = 1, 2, 3, 4$ represent xinhuanet.com, huanqiu.com, cnr.cn, and news.sina.com.cn, as shown in Figure 2.

We can see from Figures 2(a)–2(d) that there are statistically significant cross-correlations between the daily number of media reports $x_{11}, x_{12}, x_{13}, x_{14}$ and the number of daily confirmed cases y_1; the local maximal cross-correlation coefficient occurs at $lag = 3, -1, -1, -1$, respectively. We also calculated the cross-correlation function (CCF) between the average daily number of media reports (x_1) and the number of daily confirmed cases (y_1) from 20 January to 28 April 2020 at specific lags. It shows that x_1 correlated with y_1 significantly at lags ranging from −9 to 5, and this cross-correlation coefficient achieves its maximum value at $lag = 0$, as shown in Figure 2(e). Similarly, we denote the number of daily media reports by x_{2i} and the number of daily confirmed cases by y_2 from 11 June to 3 September 2020, where $i = 1, 2, 3, 4, 5, 6, 7$ represent china.com.cn, people.com.cn, xinhuanet.com, news.sina.com.cn, cyol.com, chinanews.com, and gmw.cn, respectively. The results are summarized in Figure 3.

From Figure 3, we can see that there also exist statistically significant cross-correlations between x_{2i} and y_2, with the local maximal cross-correlation coefficient occurring at $lag = -2, -2, -2, -2, -2, -2, -3$ days. Finally, we calculated the cross-correlation function between the average daily number of media reports (x_2) and the number of daily confirmed cases (y_2) at specific lags. The result suggests statistical cross-correlation between x_2 and y_2, with the cross-correlation coefficient achieving its maximum value at $lag = -2$, as shown in Figure 3(h). This demonstrates that the average daily number of media reports has the strongest correlation with the daily number of confirmed cases two days prior.

2.3. The Model. Based on the disease progression of COVID-19 and the intervention measures, we established a discrete compartment model. To study the effects of media-induced social distancing, we classified the total population in the natural transmission process into eight compartments, including fully susceptible (S), media-aware susceptible (S_M), quarantined susceptible (S_q), exposed (E), quarantined

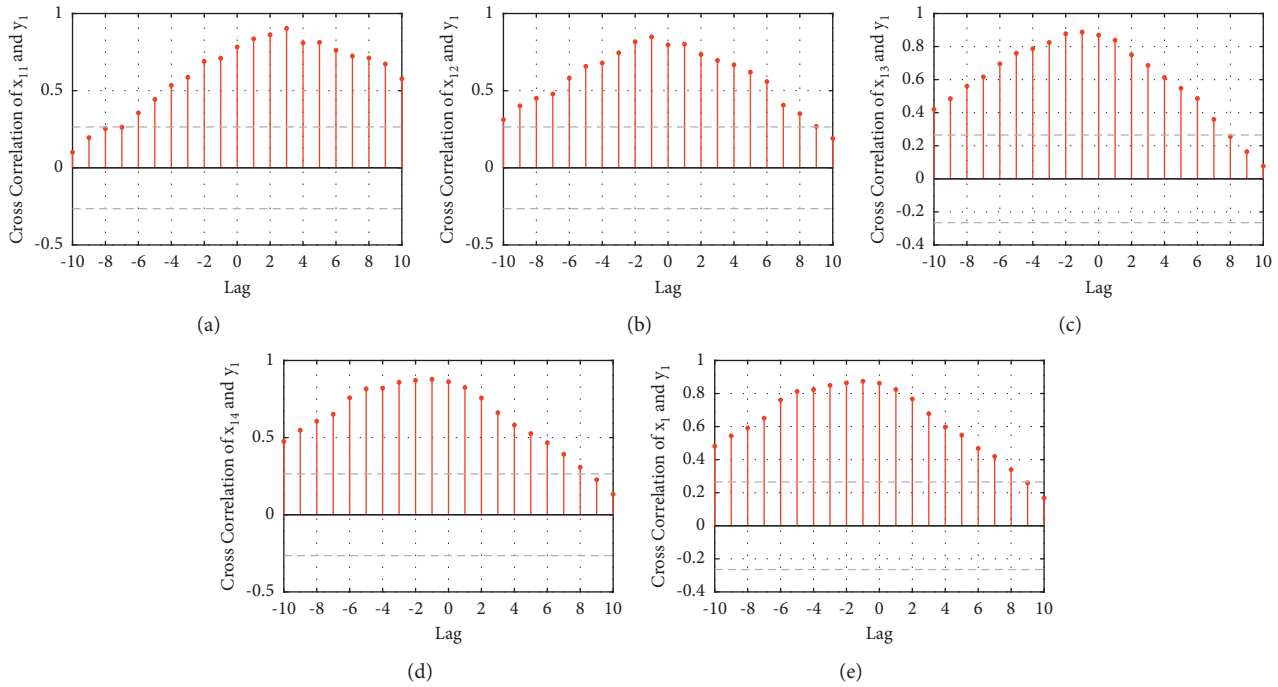


Figure 2: Cross-correlation coefficients between the number of daily confirmed cases of COVID-19 in Beijing from 20 January to 28 April, 2020, and the daily number of media reports of websites (a) xinhuanet.com, (b) huanqiu.com, (c) cnr.cn, and (d) news.sina.com.cn. (e) Cross-correlation coefficients between the average daily number of media reports and the number of daily confirmed cases.

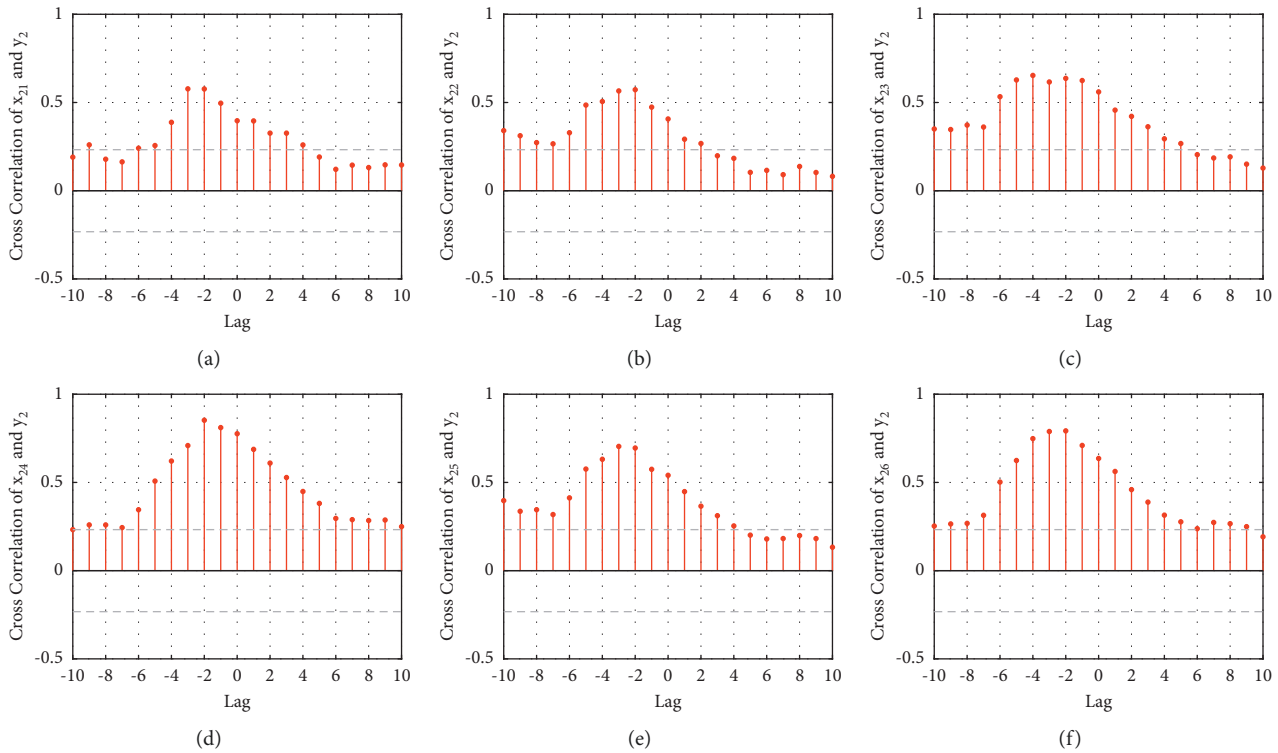


Figure 3: Continued.

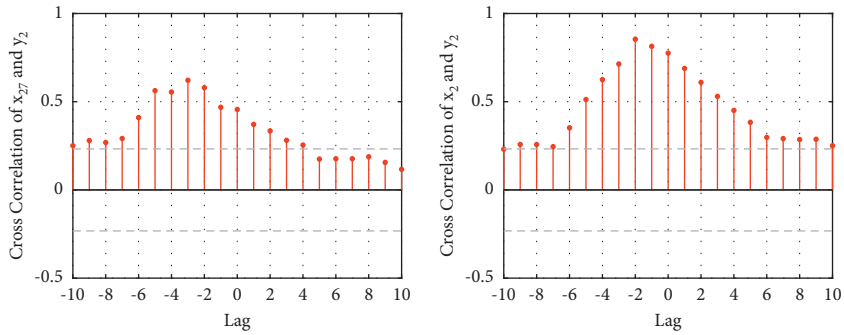


FIGURE 3: Cross-correlation coefficients between the number of daily confirmed cases of COVID-19 in Beijing from 11 June to 3 September, 2020 and the daily number of media reports of websites (a) China.com.cn, (b) people.com.cn, (c) xinhuanet.com, (d) news.sina.com.cn, (e) cyol.com, (f) chinanews.com, and (g) gmw.cn. (h) Cross-correlation coefficients between the average daily number of media reports and the number of daily confirmed cases.

exposed (E_q), symptomatic infectious (I), hospitalized (H), and recovered (R), where M represents media reports (Table 1).

To clearly illustrate the disease transmission, we plotted the flow diagram of the model in Figure 4.

Based on the flowcharts shown in Figure 4, we established the following model equations:

$$
\begin{aligned}
S_t + 1 &= S_t - \frac{\left(1 - e^{-\beta c I_t}\right) S_t}{N_t} - \frac{\left(1 - e^{-(1-\beta) c q I_t}\right) S_t}{N_t} - \left(1 - e^{-\mu(t)}\right) S_t + \left(1 - e^{-\lambda_f}\right) S_M, t + \left(1 - e^{-\lambda_q}\right) S_{q,t},\\
S_{M,t+1} &= S_{M,t} + \left(1 - e^{-\mu(t)}\right) S_t - \frac{\left(1 - e^{-(1-\beta) c_f(t) I_t}\right) S_{M,t}}{N_t} - \frac{\left(1 - e^{-\beta c_f(t) I_t}\right) S_{M,t}}{N_t} - \left(1 - e^{-\lambda_f}\right) S_{M,t},\\
S_{q,t+1} &= S_{q,t} + \frac{\left(1 - e^{-(1-\beta) c q I_t}\right) S_t}{N_t} + \frac{\left(1 - e^{-(1-\beta) c_f(t) I_t}\right) S_{M,t}}{N_t} - \left(1 - e^{-\lambda_q}\right) S_{q,t},\\
E_{t+1} &= E_t + \frac{(1-q)\left(1 - e^{-\beta c I_t}\right) S_t}{N_t} - \left(1 - e^{-\sigma}\right) E_t,\\
E_{q,t+1} &= E_{q,t} + \frac{q\left(1 - e^{-\beta c I_t}\right) S_t}{N_t} + \frac{\left(1 - e^{-\beta c_f(t) I_t}\right) S_{M,t}}{N_t} - \left(1 - e^{-\sigma}\right) E_{q,t},\\
I_{t+1} &= I_t + \left(1 - e^{-\sigma}\right) E_t - \left(1 - e^{-\delta}\right) I_t - \left(1 - e^{-\alpha}\right) I_t,\\
H_{t+1} &= H_t + \left(1 - e^{-\sigma}\right) E_{q,t} + \left(1 - e^{-\delta}\right) I_t - \left(1 - e^{-\gamma H}\right) H_t - \left(1 - e^{-\alpha}\right) H_t,\\
R_{t+1} &= R_t + \left(1 - e^{-\gamma_H}\right) H_t,\\
M_{t+1} &= M_t + \eta\left(\left(1 - e^{-\sigma}\right) E_{q,t} + \left(1 - e^{-\delta}\right) I_t\right) - \left(1 - e^{-\mu_M}\right) M_t + \overline{m}.
\end{aligned}
\tag{2}
$$

In model (1), the contact rate of susceptible individuals is denoted by c, the transmission probability per contact is β, and the quarantined proportion of exposed individuals is q. If individuals in one compartment move to the other compartment at rate a, they stay in their own compartment with probability e^{-a} and move to the other compartment with probability $1 - e^{-a}$. In model (1), susceptible individuals exposed to the virus are effectively infected with probability $1 - e^{-\beta c I_t}/N_t$. They move to the (quarantined) exposed compartment with probability $q(1 - e^{-\beta c I_t})/N_t$ or $(1-q)(1 - e^{-\beta c I_t})/N_t$; those not infected but quarantined will move to the quarantined susceptible compartment with probability $(1 - e^{-(1-\beta) c q I_t})/N_t$. We assume that a

Table 1: Estimated initial values of variables and parameters for model (1). LS = least squares method.

Variables	Description	Initial value (I)	Resource	Initial value (II)	Resource
S	Susceptible population	21536000	Data	21536000	Data
E	Exposed population	40	LS	30	LS
I	Infectious population	41	Data	58	Data
S_M	Media-aware susceptibles	0	Assumed	0	Assumed
S_q	Quarantined susceptible population	13	LS	2	Data
E_q	Quarantined exposed population	10	LS	0	Data
H	Hospitalized population	5	Data	1	Data
R	Recovered population	0	Data	0	Data
M	Media reports	396.5927	Data	9.8516	Data
Parameters	**Description**	**Value (I)**	**Resource**	**Value (II)**	**Resource**
c	Contact rate (per person per day)	5.0061	LS	14.1108	LS
$c_{\min}$	Minimum contact rate of the media-aware individuals	1.9968	LS	5.5671×10^{-5}	LS
β	Probability of transmission from I to S per contact	0.0897	LS	0.0801	LS
q	Quarantined proportion of latent individuals	0.3001	LS	0.6382	LS
λ_f	Relaxation rate of social-distancing practices	0.0055	LS	0.0356	LS
κ	Social-distancing uptake rate of media-aware susceptibles	1.3706×10^{-5}	LS	0.0032	LS
λ_q	Release rate of quarantined individuals	1/14	[17]	1/14	[17]
σ	Progression rate of exposed individuals to infectives	1/5	[36, 37]	0.2328	LS
α	Disease-induced death rate	9.9328×10^{-4}	LS	0	Data
δ	Hospitalization rate	1/4.8644	Data	0.3690	LS
γ_H	Recovery rate of hospitalized individuals	0.0644	LS	0.0924	LS
η	Media-reporting rate of the number of new hospital notifications	107.6820	LS	2.3685	LS
μ_M	Media-waning rate	0.0013	LS	0.1510	LS
$\overline{m}$	Basic number of media reports	456.8387	LS	0	Assumed

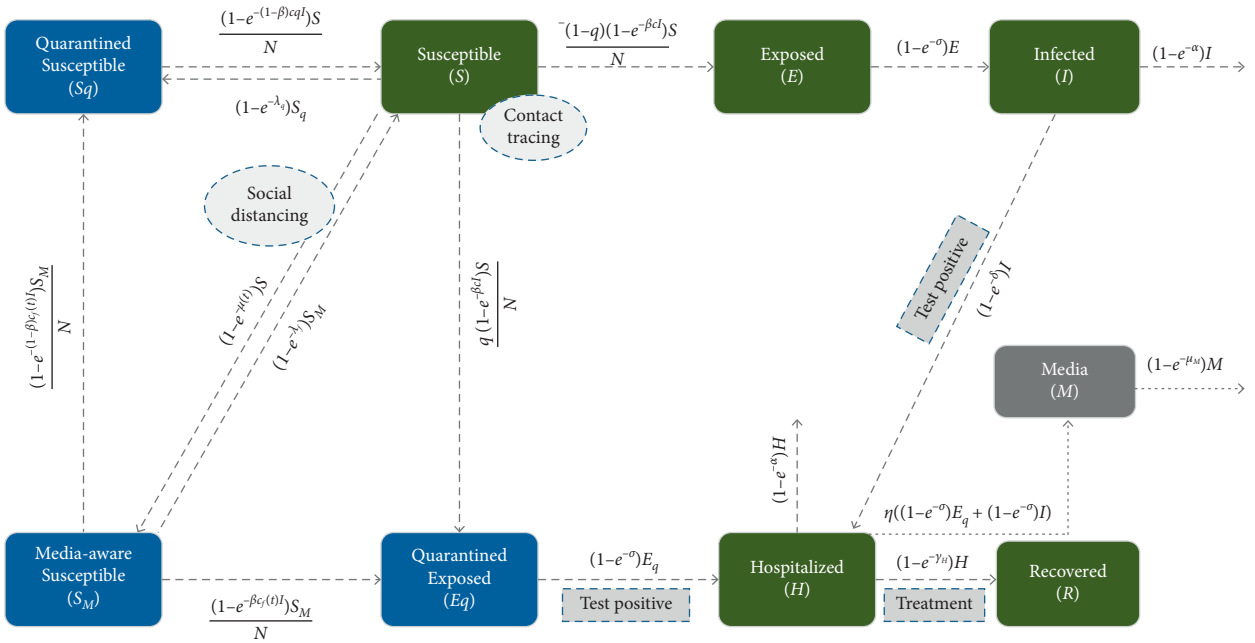


Figure 4: Flow diagram of the model for illustrating the COVID-19 infection dynamics in Beijing city. Intervention measures—including contact tracing, quarantine, and isolation—are illustrated. The media-aware susceptibles have lower probability of transmission and will eventually relax from social distancing and move back into susceptible compartment.

susceptible person would practice social distancing after receiving disease information reported via media, moving to the media-aware susceptible compartment (S_M). Media-aware susceptibles are less social and take more precautions, and therefore would have a lower contact rate ($c_f(t)$) compared with susceptible individuals (c). Media-aware susceptibles exposed to the virus can either move to the quarantined exposed compartment with probability $(1 - e^{-\beta c_f(t) I_t})/N_t$ or the quarantined susceptible compartment with probability $(1 - e^{-(1-\beta) c_f(t) I_t})/N_t$. $1 - e^{-\mu(t)}$

represents the probability by which susceptibles move to the media-aware susceptible compartment directly or after being aware of the disease. $1 - e^{-\lambda_f}$ represents the probability by which media-aware susceptibles (S_M) relax social distancing. $1 - e^{-\lambda_q}$ represents the probability by which quarantined susceptibles (S_q) are released. $1 - e^{-\sigma}$ represents the probability by which exposed individuals (E) progress to the infected compartment (I) or quarantined exposed individuals (E_q) progress to the hospital (H). $1 - e^{-\delta}$ represents the probability by which infected individuals progress to hospital. $1 - e^{-\gamma_H}$ represents the probability by which hospitalized individuals recover (R). $1 - e^{-\alpha}$ represents the probability by which the infected and the hospitalized die. $1 - e^{-\mu_M}$ represents the probability by which the media (M) wanes due to ineffectivenes, story staleness, and other factors. Finally, η is the response intensity of awareness programs on the number of newly confirmed cases, and $\overline{m}$ is the basic number of media reports.

The social-distancing uptake rate for media-aware susceptibles is

$$\mu(t) = \kappa M(t), \tag{3}$$

where $\kappa = \kappa_1\kappa_2\kappa_3$, where $\kappa_3 M(t)$ stands for the fraction of susceptible people who are exposed to media, κ_2 stands for the fraction of exposed people who trust media, and κ_1 stands for fraction of media-trusting people who practice media-induced social distancing. Here, $M(t)$ represents the media reports. It follows from Section 2.2 that the local maximal cross-correlation coefficient between the number of daily confirmed cases and the average daily number of media reports occurs at lag = 0 days for the first outbreak and at lag = −2 days for the second outbreak. Because the cross-correlation at lag = 0 is not very different from that at lag = −2 during the second outbreak, we do not take this time delay into account in our modelling.

The lower contact rate $c_f(t)$ would decrease with the increasing of the number of media-aware susceptibles, which we represent as

$$c_f = c_{\min} + (c - c_{\min})e^{-\mu(t)(t-t_0)}, \tag{4}$$

where $c_{\min} < c$ is the minimum contact rate of media-aware susceptibles with self-isolation and interventions, c is the contact rate without social-distancing practice, $\mu(t)$ is the decreasing rate of contact, and t_0 is the starting day of the data of the two outbreaks we used for model fitting.

The relaxation of social-distancing practices in media-aware susceptibles λ_f is inversely proportional to the average duration of social-distancing practices T; i.e.,

$$\lambda_f = \frac{1}{T}, \tag{5}$$

where T is the average time an individual spends practicing social distancing.

Because the first recovered cases of the initial outbreak in Beijing was reported on 24 January 2020, as shown in Figure 1(a), we formulate the recovery rate γ_H using a piecewise-defined function, which is 0 before 24 January 2020 and is a constant since January 24 2020; i.e.,

$$\gamma_H = \begin{cases} 0, & t \le 4, \\ \gamma_{H_1}, & t > 4. \end{cases} \tag{6}$$

For the second outbreak in Beijing, the first recovered cases was reported on 29 June 2020, as shown in Figure 1(e), so we have

$$\gamma_H = \begin{cases} 0, & t \le 19, \\ \gamma_{H_2}, & t > 19. \end{cases} \tag{7}$$

The detailed definitions and values of variables and parameters have been listed in Table 1.

It is worth emphasizing that the targeted model we propose in this work is novel and quite distinct to the ones found in the literature. Specifically, we differentiate the fully susceptibles (i.e., no social-distancing practices) from the media-aware susceptibles (i.e., taking moderate social-distancing practices) and the quarantined susceptibles (i.e., total isolation), which is distinct from existing models, where only fully susceptible and quarantined susceptibles are considered [15, 34, 35]. A media-dependent function $c_f = c_{\min} + (c - c_{\min})e^{-\mu(t)(t-t_0)}$ with $\mu(t) = \kappa M(t)$ is defined to mimic the contact rate of media-aware susceptibles in our model, which is quite different from the modelling explained by Collinson et al. [14], where a constant is adopted to represent the contact rate of media-aware susceptibles.

3. Results

3.1. Parameter Estimation. The reproduction number R_0 represents the average number of new infections generated by one infected individual in the population during the average infection period [38]. The reproduction number can therefore be regarded as a threshold value, from which we can determine whether COVID-19 spreads or not. In particular, the COVID-19 pandemic can be eradicated from the population for $R_0 < 1$; conversely, it will spread if $R_0 > 1$. Using the next-generation method [39, 40], the reproduction number was calculated for our targeted model (1) as

$$R_0 = \frac{c(1-q)\beta}{2 - e^{-\delta} - e^{-\alpha}}. \tag{8}$$

In the first outbreak of the COVID-19 epidemic in Beijing, almost everyone was susceptible to the virus, so we set $S(0) = 21,536,000$, the population of Beijing, and assumed $S_M(0) = 0$. Quarantined individuals were required to be isolated for 14 days, so $\lambda = 1/14$ [17]. According to the data from the Beijing Municipal Health Commission of the People's Republic of China, the total number of quarantined individuals was 23 as of 20 January, so we set $S_q(0) + E_q(0) = 23$. The number of hospitalized individuals was $H(0) = 5$, and recovered cases was $R(0) = 0$ on 20 January. The average daily number of media reports on 20 January was $M(0) = 396.59$. We calculated $I(0) = 41$ and the average period from symptom onset to hospitalization as $1/\delta = 4.8644$ days based on the detailed information of 178 reported cases. It is worth mentioning that we collected the daily number of media reports by searching the key word

"COVID-19" included in the title or the full text, so when the first outbreak of the epidemic was over in Beijing but not over in other provinces or cities, there were still media reports, so it is natural to set the basic number of media reports $\overline{m}$ of this outbreak as a nonzero constant.

According to the data of the second outbreak of the epidemic in Beijing, there were two quarantined individuals as of 11 June, and both were susceptible individuals, so we set $S_q(0) = 2$ and $E_q(0) = 0$. We also had the number of hospitalized individuals $H(0) = 1$, the number of recovered cases $R(0) = 0$, and the disease-induced death rate $\alpha = 0$. We calculated $I(0) = 58$ based on the detailed information of 285 cases. The average daily number of media reports was calculated as $M(0) = 9.85$ on 11 June. For this outbreak, we collected the number of media reports by searching the key words "epidemic" and "Beijing" included in the title or the full text, so when the epidemic was over, there were no media reports, which resulted in $\overline{m} = 0$.

By simultaneously fitting our targeted model to the data of cumulative number of confirmed cases, cumulative number of recovered cases, and average daily number of media reports from 20 January to 28 April and from 11 June to 3 September, 2020, we first estimated the values of unknown parameters and the initial conditions of variables using the nonlinear least-squares method. The best-fitting results were marked as black asterisks in Figure 5, showing that our targeted model captures the data well. In Figure 5, the red asterisks are the data, whereas the black asterisks are the fitted result. The reproduction number of the first and second outbreaks were estimated as $R_0 = 1.6818$ and $R_0 = 1.3251$. It follows that a single infected individual can infect more susceptibles during the first outbreak than the second outbreak. The final size of infected individuals in the first outbreak is hence larger than that in the second outbreak, resulting in 415 cases in the first outbreak and 335 cases in the second outbreak.

To obtain the confidence intervals, we assumed that the cumulative numbers of confirmed cases, recovered cases, and average daily number of media reports follow Poisson distributions with the observed data on each day being the respective means of 500 randomly generated samples of data sets. After 500 stochastic fittings of model (1), we derived 95% upper confidence limits, 95% lower confidence limits, and 95% confidence intervals of the cumulative number of confirmed cases, recovered cases, and daily number of media reports, which are shown by blue curves, green curves, and grey regions, respectively, in Figure 5. It should be noted that the number of media reports around 3 February fell outside the confidence interval. This was mainly due to the official delivery of the Huoshenshan Hospital to the Huibei medical support team of PLA on February 2, and the overnight construction of three Fangcang shelters in Wuhan on 3 February, which resulted in an unusually high number of media reports around 3 February.

In Table 1, initial value (I) and parameter value (I) (resp. initial value (II) and parameter value (II)) refer to the initial value and parameter value of the first outbreak (resp. the second outbreak). It follows from Table 1 that the media-reporting rate η and the media-induced social distancing uptake rate κ of the first outbreak are much greater than that of the second outbreak. This is because the COVID-19 outbreak had spread across the country during the first outbreak in Beijing, when all media reports about COVID-19 had an impact on the people in Beijing. So we searched the keyword "COVID-19" to collect the data of media reports, which led to a comparatively large size of the number of media reports. However, the second outbreak occurred only in Beijing, so we collected media reports by searching the key words "epidemic" and "Beijing", which resulted in a relatively small size. From Table 1, the contact rate c of the second outbreak is significantly greater that of the first outbreak. The probability of transmission from I to S per contact β of the first outbreak is greater than that of the second outbreak, and the recovery rate γ_H of the first outbreak is less than that of the second outbreak, whereas the quarantine rate of latent individuals q and the progression rate of infectives to hospital δ of the first outbreak are less than those of the second outbreak. This was because the Newland Market began to close gradually starting 12 June, and all the employees of the market and their close contacts were undergoing nucleic acid testing, which helped shorten the time from the infected population with symptomatic I to the hospitalized population H.

3.2. Uncertainty and Sensitivity Analysis. We conducted a sensitivity analysis of the peak size and peak time of the epidemic in Beijing from 20 January to 28 April and from 11 June to 3 September, 2020 with respect to the key parameters by performing Latin hypercube sampling (LHS) and calculating partial rank correlation coefficients (PRCCs) [41]. This allows us to assess whether there is significant effect of one parameter on the peak time and peak size of the daily number of confirmed cases. Latin hypercube sampling was conducted with 5000 bins and 500 simulations per sampling. It can be seen from Figures 6(a) and 6(b) that $\beta, \sigma, \kappa, \delta, c$ are the most sensitive parameters of the first outbreak.

In particular, Figures 6(a) and 6(b) demonstrate that decreasing the transmission probability β and the contact rate c could lower the peak size of the number of newly confirmed cases as well as bring forward the peak time of the first outbreak significantly. Reducing the hospitalized period $1/\delta$ could lower the peak size as well as advance the peak time, whereas shortening the incubation period $1/\sigma$ could lead to an increase in the peak size and advance in the peak time. It is worth emphasizing that the parameters related to social-distancing practices can also affect the peak size and peak time significantly: increasing the average time an individual spends practicing social distancing $1/\lambda_f$ and the social distancing uptake rate for S_M class κ could greatly lower the peak size while bringing forward the peak time. This further suggests the vital role of media reports in curbing the disease transmission. We also derive from Figures 6(a) and 6(b) that enhancing the quarantine rate q could bring forward the peak time significantly. Figures 6(c) and 6(d) show that $\beta, \kappa, c, \lambda_f, \delta, \sigma$ are the most sensitive parameters to the peak size and peak time of the daily number of confirmed cases in the second outbreak. Besides

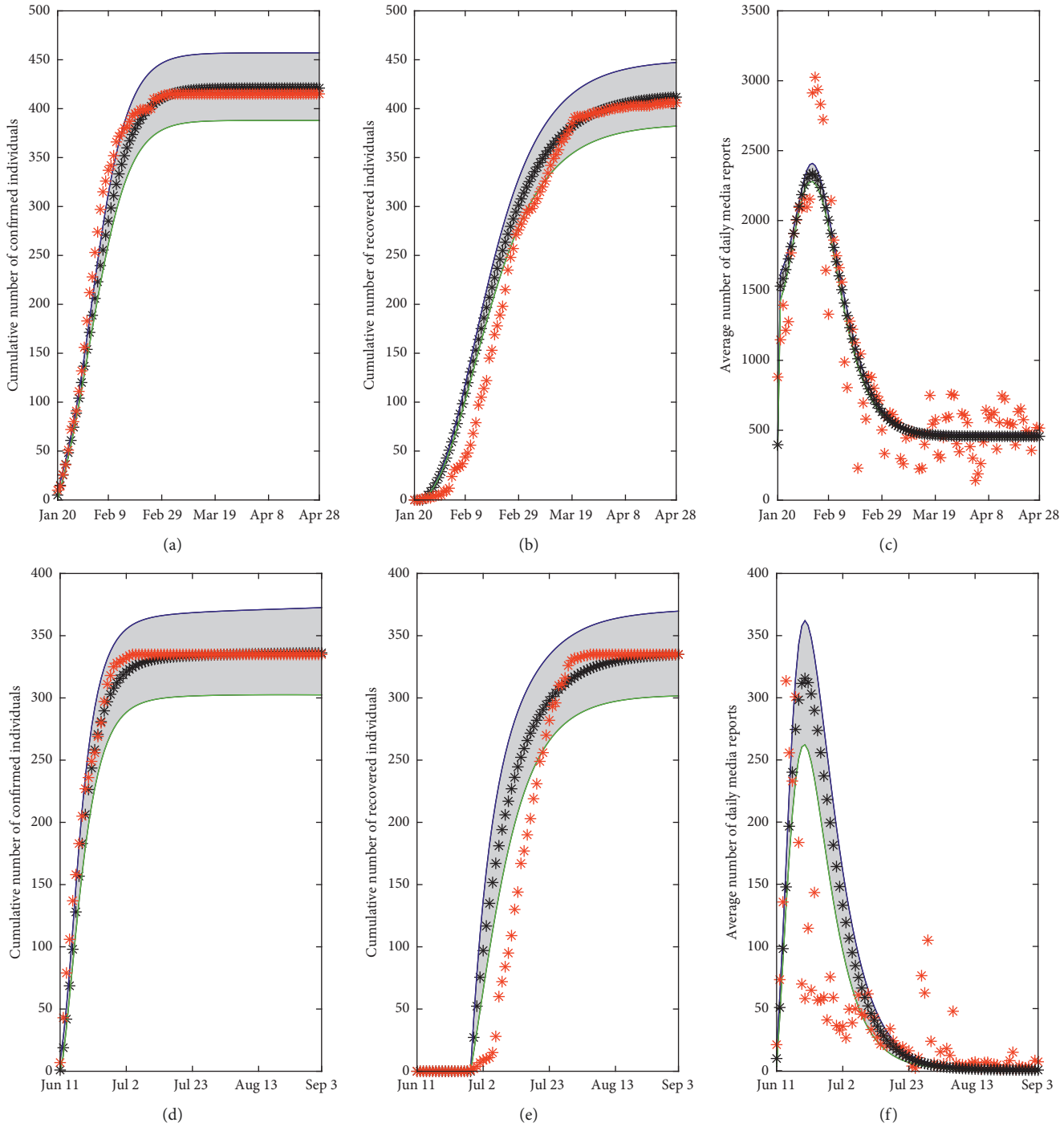


FIGURE 5: Fitting result for the Beijing data from (a)–(c) 20 January to 28 April 2020 and (d)–(f) 11 June to 3 September 2020. The red asterisks are the cumulative number of confirmed cases, recovered cases, and average daily number of media reports, respectively. The black asterisks are the best-fitting result. The grey region represents the 95% confidence intervals with the blue/green curves representing the 95% upper/lower confident limits.

these parameters, strengthening the quarantine rate q can significantly advance the peak time of the second outbreak as well as lower the peak size, enhancing the response intensity of media reporting on the number of newly confirmed cases η or declining the minimum contact rate of media-aware susceptibles $c_{\min}$ could decrease the peak size and bring the peak time forward.

Although the parameters $\beta, \kappa, c, q, \lambda_f, \delta, \sigma$ are significant to both the first and second outbreaks, the magnitude of the impact is different. The transmission probability β, the contact rate c, and the progression rate of exposed individuals to infectives σ have more effect on the peak size in the second outbreak than in the first one, whereas the transmission probability β, the contact rate c, the quarantine rate

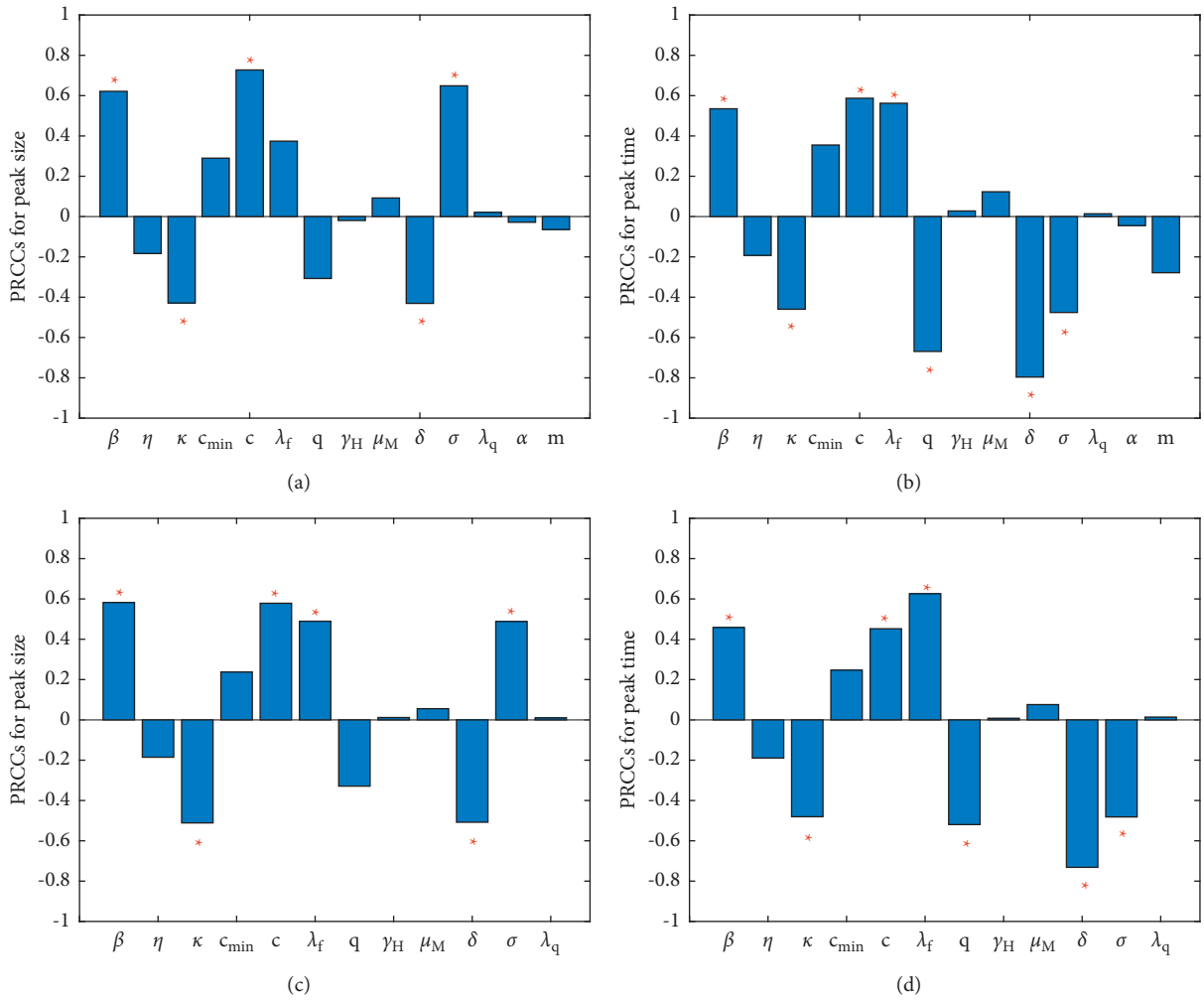


Figure 6: Sensitivity analysis of the peak size and peak time of the daily number of confirmed cases to the first ((a)–(b)) and second ((c)–(d)) outbreaks.

q, and the progression rate of infectives to hospital δ have more effect on the peak time in the first outbreak than in the second. It is worth noting that the social-distancing uptake rate κ and the relaxation rate of social-distancing practices λ_f have more effect on the peak size and peak time in the second outbreak than in the first, which indicates a vital role that the social-distancing practices induced by media coverage played in combating the second outbreak.

To further investigate the dependence of the peak time and peak size of the two outbreaks of COVID-19 in Beijing, 2020, we plotted the contour plots of the peak time and peak size with respect to the quarantine rate q and the contact rate c (Figures 7(a) and 7(b) for the first outbreak and Figures 7(e) and 7(f) for the second outbreak), the social distancing uptake rate κ and the relaxation rate of social distancing λ_f (Figures 7(c) and 7(d) for the first outbreak and Figures 7(g) and 7(h) for the second outbreak), respectively.

We find that increasing the quarantine rate q and decreasing the contact rate c would bring the peak times of both outbreaks forward (Figures 7(a) and 7(e)), whereas the peak sizes would be lowered significantly (Figures 7(b) and 7(f)). If the social-distancing uptake rate κ was enhanced, the peak time would be advanced (Figures 7(c) and 7(g)), while the peak size would decline (Figures 7(d) and 7(h)). If the social-distancing uptake rate κ was fixed, decreasing the relaxation rate of social-distancing practices could help advance the peak time as well as lower the peak size, which is not obvious. The results indicate that in the early stage of the outbreak, enhancing quarantine, strengthening the media reporting to induce more social-distancing practices, and reducing contacts could reduce the severity of the epidemic significantly. In particular, the peak size of the second outbreak could be greatly reduced with higher quarantine. To explore the explicit effectiveness of the social distancing uptake rate induced by mass media and quarantine on both COVID-19 outbreaks in Beijing, we examined how the cumulative number of confirmed cases vary with different values of the contact rate c, the quarantine rate q, the progression rate of infectives to hospital δ, the media-reporting rate η, the social distancing uptake rate κ, and average social distancing time T. We performed a sensitivity

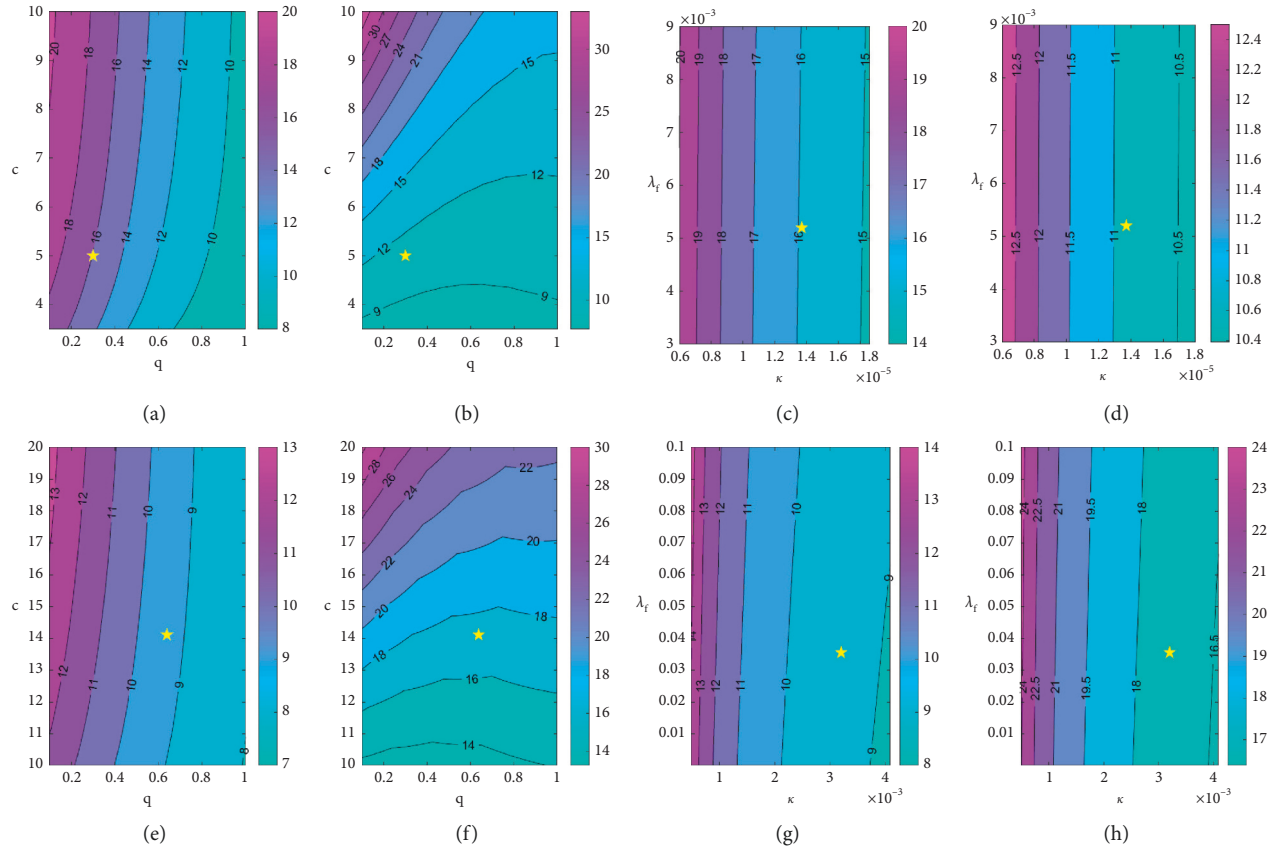


FIGURE 7: Contour plots of the peak time and peak size of the number of newly confirmed cases with respect to q and c from 20 January to 28 April (a, b) and from 11 June to 3 September, 2020 (e, f). Contour plots with respect to λ_f and κ from 20 January to 28 April (c, d) and from 11 June to 3 September, 2020 (g, h). The yellow star represents the position of (q, c) or (λ_f, κ) we have estimated using the data to fit the targeted model (1). (a) Peak time. (b) Peak size. (c) Peak time. (d) Peak size. (e) Peak time. (f) Peak size (g) Peak time. (h) Peak size.

analysis to quantify what would happen if c were reduced to $0.8c, 0.6c, 0.4c$ (Figures 8(a) and 8(d)).

We also investigated the cases if q were decreased to $0.7q, 0.4q, 0.1q$ (Figures 8(b) and 8(e)), if δ were reduced to $0.8\delta, 0.6\delta, 0.4\delta$ (Figures 8(c)–8(f)), if η were diminished to $0.5\eta, 0.3\eta, 0.1\eta$ for the first outbreak (Figure 8(g)) and $0.5\eta, 0.3\eta, 0.1\eta$ for the second outbreak (Figure 8(j)), or if κ and T were decreased to $0.4\kappa, 0.2\kappa, 0.1\kappa$ and $0.4T, 0.2T, 0.1T$ (Figures 8(h) and 8(k), 8(i) and 8(l)). It is known from Figure 8 that decreasing the quarantine rate q, the progression rate of infectives to hospital δ, the response intensity of awareness programs η, the social distancing uptake rate κ, and the average time of social distancing T could increase the cumulative number of confirmed cases; however, the cumulative number of confirmed cases could be reduced with a decreasing of the contact rate c. The social-distancing uptake rate κ had a more significant impact on containing the first outbreak than it did on the second. Figures 8(i) and 8(l) indicate a more significant impact of the average time of social distancing on curbing the second outbreak than it did on the first, which illustrates the importance of social-distancing practices in the post-epidemic period. Note that the progression rate of infectives to hospital of the second outbreak ($\delta = 0.3690$) is higher than that of the first ($\delta = 0.2056$), although the contact rate during the second outbreak ($c = 14.1108$) is also larger than that of the first ($c = 5.0061$). In fact, the Newland Market, where the second outbreak occurred, began to close the day after the outbreak (12 June 2020), and all individuals related to the Newland Market were tested, which resulted in a lower contact rate and a higher progression rate of infectives to hospital. We thus examined the effect of the contact rate and the progression rate of infectives to hospital during the second outbreak on mitigating the transmission of COVID-19, as shown in Figure 9.

Figure 9(a) shows that around 200 fewer cumulative confirmed cases would be reported if the contact rate of the second outbreak were reduced to the contact rate of the first outbreak (i.e., $c_2 = c_1$), where c_1 represents the estimated value of the contact rate for the first outbreak and c_2 is defined similarly. If the contact rate of the second outbreak only dropped to the average value of c_1 and c_2, around 100 infections would be avoided. Figure 9(b) shows that the cumulative number of confirmed cases of the second outbreak would increase by around 200 cases if the progression rate of infectives to hospital of the second outbreak remained the same as that of the first (i.e., $\delta_2 = \delta_1$), where δ_1 represents the estimated value of the confirmation rate for the first

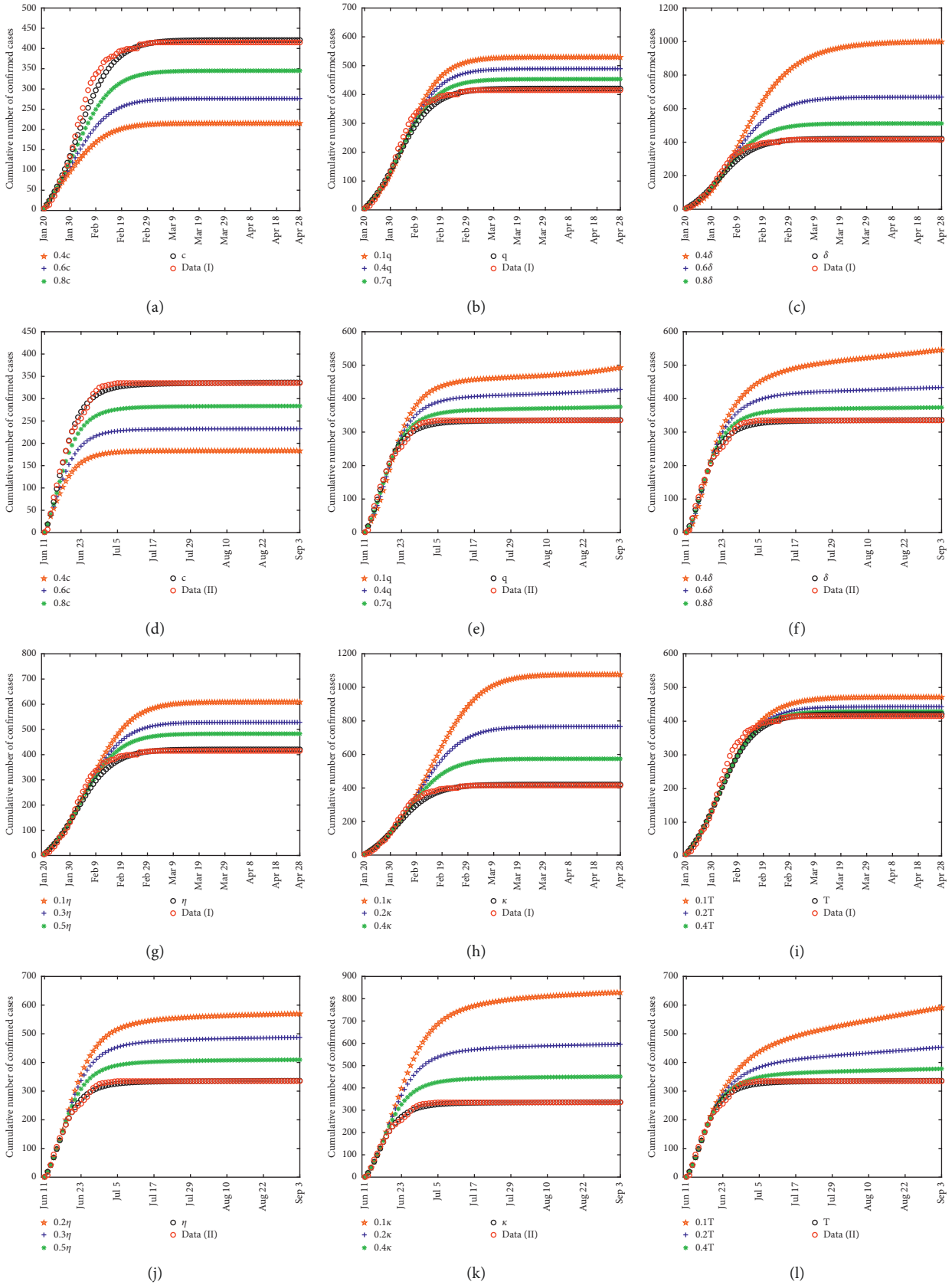


FIGURE 8: Variation of the cumulative number of confirmed individuals with various values of the contact rate c, quarantine rate q, progression rate of infectives to hospital δ, media-reporting rate η, social-distancing uptake rate κ, and the average time of social distancing T for both the first outbreak ((a)–(c) and (g)–(i)) and the second outbreak ((d)–(f) and (j)–(l)).

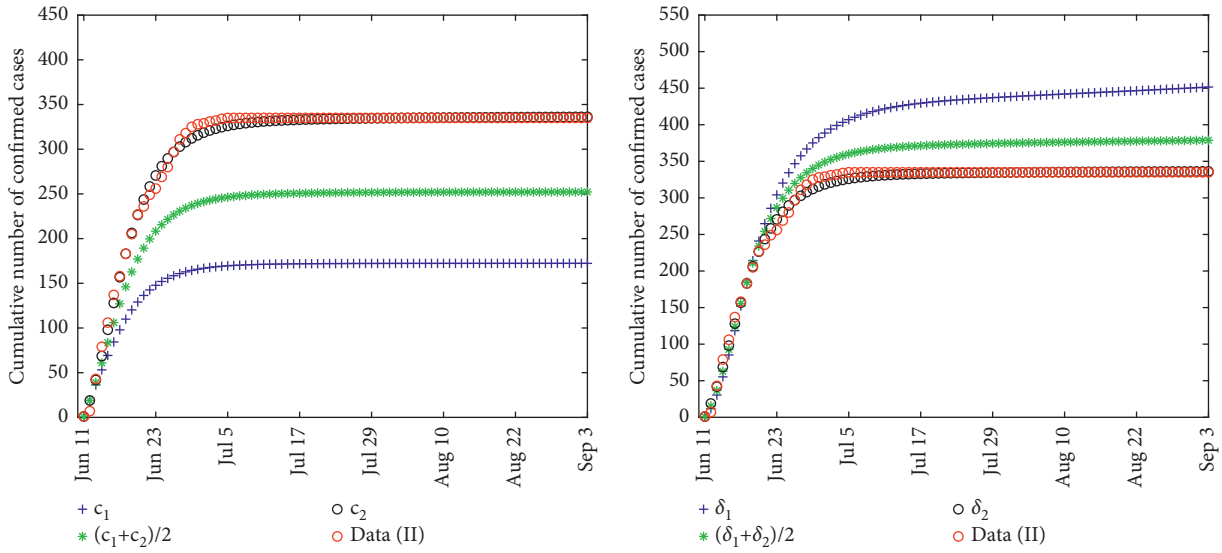


Figure 9: Cumulative number of confirmed cases of the second outbreak in Beijing from 11 June to 3 September, 2020 with various values of contact rate c or progression rate of infectives to hospital δ.

outbreak of Beijing and δ_2 is defined similarly. If the progression rate of infectives to hospital of the second outbreak only dropped to the average value of δ_1 and δ_2, the cumulative number of confirmed cases would increase by around 80 cases. These findings demonstrate the effect of hospitalization (characterised by δ_2) in the containment of the second outbreak and the impact of contacts (characterised by c_2). To quantify the effect of social-distancing practices induced by media coverage, contact tracing, and quarantine measures, we conducted calculations on the number of confirmed cases and the length of the epidemic for both outbreaks in Beijing with different values of κ, T, q and δ with other parameters fixed. The results are listed in Table 2.

These results demonstrate that if the social-distancing uptake rate were reduced to $0.4\kappa, 0.2\kappa$, and 0.1κ, the final confirmed cases would increase by 159, 351, and 661 cases (from 415 cases to 574, 766, and 1076 cases), respectively, for the first outbreak, whereas final confirmed cases would increase by 116, 260, and 493 cases (from 335 cases to 451, 595, and 828 cases), respectively, for the second outbreak. If the average time of social-distancing practices were decreased to $0.4T, 0.3T$ or $0.2T$, then 429, 434, or 443 confirmed cases (instead of 415 confirmed cases in reality) would be reported for the first outbreak, whereas 378, 403, or 453 confirmed cases (instead of 335 confirmed cases in reality) would be reported for the second outbreak. It is worth emphasizing that the length of the epidemic would be prolonged by 14% and 28% for the first and second outbreaks, respectively, even if the social-distancing uptake rate were reduced to 0.4κ; it would be prolonged by 5% and 80% for the first and second outbreaks, respectively, if the average time of social-distancing practices were decreased to $0.2T$. This demonstrates that social-distancing practices induced by media, quarantined, and confirmation efficiency could all alleviate outbreaks of COVID-19 in Beijing. We can conclude that enhancing media reporting to induce more social-distancing practices of susceptible individuals, strengthening contact tracing and quarantine, and improving the confirmation efficiency can help to effectively control the spread of COVID-19.

Table 2: The impact of media awareness, contact tracing, and confirmation measures on the length of the epidemic and the final number of confirmed cases.

Parameters	Length of the epidemic (I)	Length of the epidemic (II)	Final confirmed cases (I)	Final confirmed cases (II)
Real data	44	25	415	335
0.4κ	50	32	574	451
0.2κ	58	35	766	595
0.1κ	68	45	1076	828
$0.4T$	44	32	429	378
$0.3T$	45	41	434	403
$0.2T$	46	45	443	453
$0.7q$	44	29	453	376
$0.4q$	46	32	490	427
$0.1q$	47	36	530	493
0.8δ	48	29	512	374
0.6δ	57	34	669	434
0.4δ	75	56	999	546

4. Conclusion and Discussion

The COVID-19 pandemic will remain a potential threat to many countries globally because of its nature of transmission and emergence of mutated viruses. Various non-pharmaceutical interventions have helped to mitigate the epidemic. We proposed a discrete compartment model to

explore how media-induced social-distancing practices, coupled with contact tracing and quarantine measures, helped contain the COVID-19 transmission in Beijing from 20 January to 28 April and from 11 June to 3 September, 2020. In fact, one of the reasons why China has been so successful in containing the COVID-19 outbreak is the social-distancing practices of the media-aware susceptibles, which has led to a considerable reduction of contacts. In this work, we differentiated the fully susceptible from susceptibles who practice social distancing and revealed the vital role of social distancing in containing COVID-19.

To study the effect of the initial transmission of COVID-19 in Beijing, we computed the basic reproduction number as $R_0 = 1.6818$ for the first outbreak and $R_0 = 1.3251$ for the second outbreak. This suggests a more severe outbreak of the first wave compared with the second. This is because the progression rate of infectives to hospital δ and the quarantined rate q of the second outbreak $(\delta = 0.3690, q = 0.4382)$ are higher than those of the first outbreak $(\delta = 0.2056, q = 0.3001)$. That suggests a significant effect that the more intense nonpharmaceutical interventions in the second outbreak had compared with the first outbreak.

We conducted a sensitivity analysis of the peak time and peak size of the cumulative number of confirmed cases with respect to the model parameters, as shown in Figure 6. The results illustrated that both the first and the second outbreaks are significantly sensitive to the parameters $\beta, \kappa, c, c_{\min}, \lambda_f, \delta$, and σ, although each of them had a different magnitude of impact on the two outbreaks. It revealed the vital role of the social-distancing practices of the media-aware susceptibles (i.e., κ and λ_f) besides the control measures in mitigating the severity of the epidemic in Beijing, which suggested that increasing the social-distancing uptake rate κ and the average time spent practicing social distancing $1/\lambda_f$ could greatly lower the peak size as well as bring forward the peak time of the two outbreaks. The finding also demonstrated that the second outbreak is much more sensitive to the media-reporting rate η than the first outbreak. This illustrates the important role of timely media reporting in the outbreak.

The results presented in Figure 8 indicate that the key factors related to media-induced social-distancing practices had a significant impact on the cumulative number of confirmed cases. Table 2 quantitatively revealed the impact of social-distancing practices, quarantine and confirmation efficiency on the final confirmed cases, and the length of the epidemic. It suggested that 159 infections (resp. 116 infections) and 14 infections (resp. 43 infections) are avoided using the social-distancing uptake rate for media-aware susceptibles κ and the average time of practicing social distancing T that we reported in Table 1 for the first (resp. the second) outbreak, compared with the infections at 0.4κ and $0.4T$. These results demonstrated that strengthening media-induced social-distancing practices, enhancing contact tracing and quarantine measures, and improving confirmation efficiency could all help alleviate the severity of the outbreak significantly.

The three media-related factors—the social distancing uptake rate for media-aware susceptibles κ, the average duration of social distancing T, and the media-reporting rate η—will significantly affect the outcome of the COVID-19 outbreak in Beijing. The media-reporting rate has previously been shown to affect the outcome of the 2009 H1N1 pandemic and the outcome of COVID-19 in India, Wuhan, and Shaanxi, China [14, 15, 34, 35]. Media fatigue was related to producing two waves of the 2009 H1N1 pandemic [14]. However, the second outbreak in Beijing was not triggered by the first outbreak because there were no new cases between the two outbreaks. It follows that strengthening media reports to enhance social distancing is a critical tool in containing the COVID-19 outbreak.

We focused on the effect of media-induced social-distancing practices in mitigating the transmission of COVID-19 using Beijing as an example. By fitting multisource data, including the epidemic data and media data, to our targeted model incorporating a media-aware susceptible class and quarantined susceptibles, we found that media coverage and quarantine measures had a significant effect in containing the outbreak in Beijing. Our findings may aid in policymaking in combating COVID-19 for China and other regions or countries considering nonpharmaceutical measures.

Data Availability

The website we used to collect the cumulative number of confirmed cases, deaths, and recovered was http://wjw.beijing.gov.cn/. The websites we used to collect the media reports were https://news.sina.com.cn/ and http://www.cnr.cn/ and http://www.xinhuanet.com/ and https://www.huanqiu.com and http://www.gmw.cn/ and http://www.chinanews.com/ and http://www.cyol.com/ and and http://www.people.com.cn/ and http://www.china.com.cn. The website we used to collect the IP addresses that visited each website was https://lexa.chinaz.com/.

Conflicts of Interest

The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Acknowledgments

This research was funded by the National Natural Science Foundation of China (grant number: 11801013 (AW)), special scientific research project of emergency public-health security of education department of Shaanxi province (grant number: 20JG002), the innovative research projects for graduate students (grant number: YJSCX21ZD05 (YG)), and by NSERC Discovery and Alliance Grants in Canada (SRS?).

References

[1] NHC, "National Health Commission of the People's Republic of China," 2021, http://www.nhc.gov.cn/xcs/yqtb/202112/efdac7bd514b45b8b3b2012d04aefd2a.shtml.

[2] Z. Li, K. Nie, and K. Nie, "Genome characterization of the first outbreak of COVID-19 delta variant B 1.617. 2-Guangzhou city, guangdong province, China, may 2021," *China CDC Weekly*, vol. 3, no. 27, pp. 587–589, 2021.

[3] WHO, "World Health Organization," 2021, https://www.who.int/emergencies/diseases/novel-coronavirus-2019.

[4] S. M. C. Abo and S. R. Smith?, "Is a COVID-19 vaccine likely to make things worse?" *Vaccines*, vol. 8, no. 4, p. 761, 2020.

[5] S. Del Valle, H. Hethcote, J. M. Hyman, and C. Castillo-Chavez, "Effects of behavioral changes in a smallpox attack model," *Mathematical Biosciences*, vol. 195, no. 2, pp. 228–251, 2005.

[6] I. S. Kristiansen, P. A. Halvorsen, and D. Gyrd-Hansen, "Influenza pandemic: perception of risk and individual precautions in a general population. Cross sectional study," *BMC Public Health*, vol. 7, no. 1, pp. 1–7, 2007.

[7] M. A. Riva, M. Benedetti, and G. Cesana, "Pandemic fear and literature: observations from Jack London's The Scarlet Plagu," *Emerging Infectious Diseases*, vol. 20, no. 10, pp. 1753–1757, 2014.

[8] J. M. Tchuenche, N. Dube, C. P. Bhunu, R. J. Smith?, and C. T. Bauch, "The impact of media coverage on the transmission dynamics of human influenza," *BMC Public Health*, vol. 11, no. Suppl 1, 2011.

[9] Y. Kim, W. Zhong, M. Jehn, and L. Walsh, "Public risk perceptions and preventive behaviors during the 2009 H1N1 influenza pandemic," *Disaster Med. Public*, vol. 9, no. 2, pp. 145–154, 2015.

[10] B. J. Cowling, D. M. W. Ng, D. K. M. Ip et al., "Community psychological and behavioral responses through the first wave of the 2009 influenza A (H1N1) pandemic in Hong Kong," *The Journal of Infectious Diseases*, vol. 202, no. 6, pp. 867–876, 2010.

[11] P. Poletti, M. Ajelli, and S. Merler, "The effect of risk perception on the 2009 H1N1 pandemic influenza dynamics," *PLoS One*, vol. 6, no. 2, Article ID e16460, 2011.

[12] N. Perra, D. Balcan, B. Gonçalves, and A. Vespignani, "Towards a characterization of behavior-disease models," *PLoS One*, vol. 6, no. 8, Article ID e23084, 2011.

[13] J. M. Epstein, J. Parker, D. Cummings, and R. A. Hammond, "Coupled contagion dynamics of fear and disease: mathematical and computational explorations," *PLoS One*, vol. 3, no. 12, Article ID e3955, 2008.

[14] S. Collinson, K. Khan, and J. M. Heffernan, "The effects of media reports on disease spread and important public health measurements," *PLoS One*, vol. 10, no. 11, Article ID e0141423, 2015.

[15] R. K. Rai, S. Khajanchi, P. K. Tiwari, E. Venturino, and A. Misra, "Impact of social media advertisements on the transmission dynamics of COVID-19 pandemic in India," *J. Appl. Math. Comput.* vol. 68, no. 1, pp. 19–44, 2022.

[16] J. T. Wu, K. Leung, and G. M. Leung, "Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study," *The Lancet*, vol. 395, no. 10225, pp. 689–697, 2020.

[17] B. Tang, X. Wang, Q. Li et al., "Estimation of the transmission risk of the 2019-nCoV and its implication for public health interventions," *Journal of Clinical Medicine*, vol. 9, no. 2, p. 462, 2020.

[18] S. Zhao, S. S. Musa, Q. Lin et al., "Estimating the unreported number of novel coronavirus (2019-nCoV) cases in China in the first half of January 2020: a data-driven modelling analysis of the early outbreak," *Journal of Clinical Medicine*, vol. 9, no. 2, 2020.

[19] M. Gatto, E. Bertuzzo, L. Mari et al., "Spread and dynamics of the COVID-19 epidemic in Italy: effects of emergency containment measures," *P. Natl. Acad. Sci USA*.vol. 117, no. 19, Article ID 10484, 2020.

[20] C. C. Kerr, R. M. Stuart, D. Mistry et al., "Covasim: an agent-based model of COVID-19 dynamics and interventions," *PLoS Computational Biology*, vol. 17, no. 7, Article ID e1009149, 2021.

[21] R. E. Baker, S. W. Park, W. Yang, G. A. Vecchi, C. J. E. Metcalf, and B. T. Grenfell, "The impact of COVID-19 nonpharmaceutical interventions on the future dynamics of endemic infections," *P. Natl. Acad. Sci. USA*.vol. 117, no. 48, Article ID 30547, 2020.

[22] R. Laxminarayan, B. Wahl, S. R. Dudala et al., "Epidemiology and transmission dynamics of COVID-19 in two Indian states," *Science*, vol. 370, no. 6517, pp. 691–697, 2020.

[23] J. Zhang, M. Litvinova, Y. Liang et al., "Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China," *Science*, vol. 368, no. 6498, pp. 1481–1486, 2020.

[24] P. Shi, Y. Dong, H. Yan et al., "Impact of temperature on the dynamics of the COVID-19 outbreak in China," *The Science of the Total Environment*, vol. 728, Article ID 138890, 2020.

[25] X. Hao, S. Cheng, D. Wu, T. Wu, X. Lin, and C. Wang, "Reconstruction of the full transmission dynamics of COVID-19 in Wuhan," *Nature*, vol. 584, no. 7821, pp. 420–424, 2020.

[26] D. He, S. Zhao, Q. Lin et al., "The relative transmissibility of asymptomatic COVID-19 infections among close contacts," *International Journal of Infectious Diseases*, vol. 94, pp. 145–147, 2020.

[27] J. Li, P. Yuan, and J. Heffernan, "Fangcang shelter hospitals during the COVID-19 epidemic, Wuhan, China. B," *World Health Organ*, vol. 98, no. 12, pp. 830–841D, 2020.

[28] X. Wang, Q. Li, X. Sun et al., "Effects of medical resource capacities and intensities of public mitigation measures on outcomes of COVID-19 outbreaks," *BMC Public Health*, vol. 21, no. 1, pp. 1–11, 2021.

[29] G. Sun, S. Wang, M. Li et al., "Transmission dynamics of COVID-19 in Wuhan, China: effects of lockdown and medical resources," *Nonlinear Dynam*, vol. 101, no. 3, pp. 1981–1993, 2020.

[30] B. Tang, F. Xia, S. Tang et al., "The effectiveness of quarantine and isolation determine the trend of the COVID-19 epidemics in the final phase of the current outbreak in China," *International Journal of Infectious Diseases*, vol. 95, pp. 288–293, 2020.

[31] S. Khajanchi and K. Sarkar, "Forecasting the daily and cumulative number of cases for the COVID-19 pandemic in India," *Chaos*, vol. 30, no. 7, Article ID 071101, 2020.

[32] P. Samui, J. Mondal, and S. Khajanchi, "A mathematical model for COVID-19 transmission dynamics with a case study of India," *Chaos, Solitons & Fractals*, vol. 140, Article ID 110173, 2020.

[33] S. Khajanchi, K. Sarkar, J. Mondal, K. S. Nisar, and S. F. Abdelwahab, "Mathematical modelling of the COVID-19 pandemic with intervention strategies," *Results in Physics*, vol. 25, Article ID 104285, 2021.

[34] W. Zhou, A. Wang, F. Xia, Y. N. Xiao, and S. Y. Tang, "Effects of media reporting on mitigating spread of COVID-19 in the early phase of the outbreak," *Mathematical Biosciences and Engineering*, vol. 17, no. 3, pp. 2693–2707, 2020.

[35] J. Guo, A. Wang, W. Zhou, Y. Gong, and S. R. Smith, "Discrete epidemic modelling of COVID-19 transmission in Shaanxi Province with media reporting and imported cases," *Mathematical Biosciences and Engineering*, vol. 19, no. 2, pp. 1388–1410, 2022.

[36] J. Backer, D. Klinkenberg, and J. Wallinga, "Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20–28 January 2020," *Euro Surveillance*, vol. 25, no. 5, Article ID 2000062, 2020.

[37] Q. Li, X. Guan, and P. Wu, "Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia," *New England Journal of Medicine*, vol. 382, pp. 1199–1207, 2020.

[38] J. M. Heffernan, R. J. Smith, and L. M. Wahl, "Perspectives on the basic reproductive ratio," *Journal of The Royal Society Interface*, vol. 2, no. 4, pp. 281–293, 2005.

[39] P. van den Driessche and J. Watmough, "Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission," *Mathematical Biosciences*, vol. 180, pp. 29–48, 2002.

[40] S. Khajanchi, S. Bera, and T. K. Roy, "Mathematical analysis of the global dynamics of a HTLV-I infection model, considering the role of cytotoxic T-lymphocytes," *Mathematics and Computers in Simulation*, vol. 180, pp. 354–378, 2021.

[41] S. Marino, I. B. Hogue, C. J. Ray, and D. E. Kirschner, "A methodology for performing global uncertainty and sensitivity analysis in systems biology," *Journal of Theoretical Biology*, vol. 254, pp. 178–196, 2008.

Hindawi
Discrete Dynamics in Nature and Society
Volume 2022, Article ID 8412895, 9 pages
https://doi.org/10.1155/2022/8412895

Research Article

Optimized Multivariate Adaptive Regression Splines for Predicting Crude Oil Demand in Saudi Arabia

Eman H. Alkhammash,[1] **Abdelmonaim Fakhry Kamel**,[2] **Saud M. Al-Fattah**,[3] **and Ahmed M. Elshewey**[4]

[1]*Department of Computer Science, College of Computers and Information Technology, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia*
[2]*Faculty of Graduate Environmental Studies, Ain Shams University, Cairo, Egypt*
[3]*Saudi Aramco, Dhahran, Saudi Arabia*
[4]*Faculty of Computers and Information, Computer Science Department, Suez University, Suez, Egypt*

Correspondence should be addressed to Eman H. Alkhammash; hms_1406@hotmail.com and Ahmed M. Elshewey; elshewy86@gmail.com

Received 4 November 2021; Revised 16 December 2021; Accepted 24 December 2021; Published 10 January 2022

Academic Editor: Jorge E. Macias-Diaz

This paper presents optimized linear regression with multivariate adaptive regression splines (LR-MARS) for predicting crude oil demand in Saudi Arabia based on social spider optimization (SSO) algorithm. The SSO algorithm is applied to optimize LR-MARS performance by fine-tuning its hyperparameters. The proposed prediction model was trained and tested using historical oil data gathered from different sources. The results suggest that the demand for crude oil in Saudi Arabia will continue to increase during the forecast period (1980–2015). A number of predicting accuracy metrics including Mean Absolute Error (MAE), Median Absolute Error (MedAE), Mean Square Error (MSE), Root Mean Square Error (RMSE), and coefficient of determination (R^2) were used to examine and verify the predicting performance for various models. Analysis of variance (ANOVA) was also applied to reveal the predicting result of the crude oil demand in Saudi Arabia and also to compare the actual test data and predict results between different predicting models. The experimental results show that optimized LR-MARS model performs better than other models in predicting the crude oil demand.

1. Introduction

The development of prediction techniques and machine learning models is a critical task for crude oil demand [1]. The prediction techniques can predict different features in oil [2] including oil price, oil demand, oil viscosity, etc. Prediction models and techniques can present many advantages in energy sector such as energy planning, strategy formulation, and energy advancement. The design of prediction models and techniques is a complex task which has huge impacts for the economic trajectories of countries, energy companies, and other industrial sectors [3]. According to the International Energy Agency (IEA), the global demand for crude oil accounted for about 41% of the total fuel share in 2016. According to the Organization of the Petroleum Exporting Countries (OPEC), Saudi Arabia is one of the world's largest oil consumers, ranking fifth after Russia with a 3.4% share of global oil consumption in 2016.

There are numerous models that support the crude oil demand prediction, including autoregressive conditional heteroscedasticity (ARCH) model [4], other time series models, artificial neural networks [5], and fuzzy theory predictions [6, 7].

Machine learning models play an important role in the evaluation and prediction tasks. The features included in the dataset can be used to perform predictions. Machine learning models can also perform future predictions based on the available in the dataset [8]. Regression analysis is a

statistical process used to assess the relationship between various variables. In the field of machine learning, regression analysis models are widely used for predictions. The idea of the regression analysis is to show how the dependent variable (predicted variable) changes when one of the independent variables changes and other independent variables remain constant [9]. When the independent variables are restricted, regression analysis is used to obtain the average value of the dependent variable. There are three main processes for regression analysis which are (1) determining the strength of the predictors, (2) predicting an effect, and (3) trend prediction [10]. Many techniques have been presented in the field of regression analysis, which can be divided into parametric method and nonparametric method. In parametric method, the parameters contain all information about the data. The parameters contain all of the information required to predict the value of future data from the model. For example, in linear regression with a single variable, two parameters (intercept and coefficient) must be known in order to predict a new value. In nonparametric method, because more information is available, the ability to predict new values is more flexible because the parameters in the nonparametric method have infinite dimensions, and the data characteristics are superior to parametric models.

The purpose of this paper is to propose the LR-MARS model for predicting the demand for crude oil in Saudi Arabia. To improve the accuracy of the MARS model, social spider optimization is applied to improve the hyperparameters of the MARS model.

2. Related Work

This section outlines relevant studies in regard to Artificial Intelligent (AI) models for predicting the demand for crude oil. In [11], the authors proposed wavelet method to predict oil price in the long term. The proposed model can forecast the Brent oil price one year ahead. Several time series prediction approaches were compared to [11] model such as ARIMA, GARCH, and Holt-Winters. Result has shown that [11] model provides better prediction models than the other models. China's crude oil demand was predicted using soft and hard computing [12]. In [13], three estimated models for the price of petroleum called theories model, simulation model, and informal model were used. The informal estimate model performs better results than the other two models. The authors in [14] make use of eight artificial neural networks (ANN) and fuzzy regression (FR) for oil price prediction. The analysis of variance (ANOVA) and Duncan's multiple range test (DMRT) are then used to test the forecast produced by ANN and FR. The mean absolute percentage error (MAPE) was calculated for ANN models and the results have shown that ANN models outperform the FR models. For verification and validation purposes, the author have applied Spearman correlation test. The authors in [15] studied the factors that play a role in affecting the demand for oil in thirty developed countries using cointegration functions model. The variables used in the study were energy prices and national income. The result has shown strong relationship between income and the demand of energy and oil. In [16], distinct nine oil models were studied and compared. Oil price, gross domestic product (GDP), and time trend for improvement were considered among the most influencing factors of the models. A method that estimates coefficients was used in the comparison of econometric response of these models [16]. Another study focused on the markets of global crude oil and natural gas in the period 1918–1999 [17]. This study predicted price and income elasticities for crude oil, demand models, and natural gas supply [17].

Panel quantification analysis techniques were used to estimate long-term income and price elasticities in crude oil demand in the Middle East [18]. Data employed in the study covered the period 1971–2002. The result has shown high price inelasticity and slight income elasticity [18]. A prediction model for crude demand based on cointegration and a vector error correction model (VEC) is introduced [19]. Four main factors that affect the crude oil demand were considered: GDP, population growth, oil price, and the share of industrial sector in GDP. Both error correction model (ECM) and Johansen cointegration test were applied for the estimation of elasticities.

In [20], the International Energy Agency (IEA) proposed the scenarios for future oil demand for China in 2006 World Energy Outlook. The study concluded that the minimum statistical (lower bound) annual oil consumption in developed countries is 11 barrels per capita. [21]. Another study in [21] developed crude oil demand models that combines variance analysis and a flexible fuzzy regression model. The results demonstrated the superiority of fuzzy regression over the conventional model. The data used covered the period 1990–2005 for different countries: Japan, Canada, Australia, and United States [21]. In [22], the authors used data that covers the period 1981–2005. Input variables include population, GDP, oil imports, and export of oil. The study demonstrated the benefits of the optimization of particulate swarm (PSO) versus GA in estimating and predicting Iran's crude oil demand. In the domain of energy consumption prediction, another study [23] compared the performance of energy consumption prediction using conventional econometric and artificial intelligence-based models. The result reflected that AI-based models are robust and scalable for prediction. The results also showed that, in national level, the prediction of yearly energy consumption is preferred using conventional models. Moreover, nonlinear regression models obtain the lowest average MAPE (1.79%) for long-term prediction.

SSO has been successfully used to solve the continuous optimization problems [24]. In [24] the researchers adopted SSO and support vector regression as short-term electric load forecasting model. Results showed that SSO helps to achieve good results [24]. Another study in [25] used SSO algorithm to search for optimal cluster centers in fuzzy c-means clustering algorithm. The results showed that SSO improved the performance of fuzzy c-means clustering algorithm among other optimization algorithms [25]. Another study in [26] used SSO algorithm to solve discrete optimization problems. SSO was used for the problem of traveling salesman [26]. SSO was compared to eighteen

algorithms and the experimental results revealed that the performance of SSO algorithm in solving discrete problems was very useful for both low and middle-scale TSP datasets [26].

3. Materials and Methods

3.1. Linear Regression Model. On real-world data, linear regression model works perfectly. There are numerous advantages to using linear regression, such as the fact that the linear regression model in training is faster than many predictive models [27]. Linear regression is used to compute the strength of the relationship between the dependent variable and the independent variables, as well as to determine which independent variables have no relationship with the dependent variable and which independent variables contain redundant information about the dependent variable. Furthermore, linear regression models are simple to implement and use a small amount of memory [28]. If there is only one independent variable in a linear regression model, the regression function is a straight line; if there are two independent variables, the regression function is plane; and if there are n independent variables, the regression function is hyperplane with $n-$ dimensions [10]. If the actual values and predicted values are fitted, then the actual values will be similar to the predicted values. However, if there is a difference between the actual and predicted values, this difference is referred to as a cost, loss, or error. The regression function $\widehat{y}$ dependent on n independent (predictor) variables x_1,x_2, …, x_n is calculated using the following equation:

$$\widehat{y} = w_0x_0 + w_1x_1 + \cdots + w_nx_n + b. \tag{1}$$

Equation (1) represents how the value of $\widehat{y}$ varies with the independent x_1,x_2, …, x_n. w_0, w_1, …, w_n, where x_1, x_2, …, x_n. w_0, w_1, …, w_n are known as feature weights (model coefficients) and b is called a constant bias term (intercept).

3.2. Ridge Regression Model. Ridge regression is a model for multiple regression in order to perform data analysis. In ridge regression, the independent variables are highly correlated. Ridge regression model is used to avoid overfitting and to reduce the complexity of the model. New values that are predicted by ridge regression model give better results when the predictor variables are correlated [10]. Ridge regression model learns two parameters w, b by using the same standard of the least squares with adding a penalty term to make an appropriate variation for the parameter w. The penalty term in ridge regression is known as regularization in order to perform restriction to the model and reduce the overfitting, and also the coefficients of the regression are controlled using the regularization methods; this will reduce the sampling error and minimize the variance [29]. Also, L2 regularization is used for ridge regression model to minimize the residual sum of square (RRS) of the coefficients [29]. RSS for ridge regression can be expressed as in the following equation:

$$\mathrm{RSS}(w,b) = \sum_{i=1}^{N} \left(y_i - (wx_i + b)\right)^2 + \alpha \sum_{j=1}^{p} w_j^2, \tag{2}$$

where α is the penalty term. When the value of α is high, this means that the model is simple and more regularization. The penalty term α adjusts the parameters when the values of the parameters are high, so ridge regression minimizes the parameters to make the model simple and reduce the complexity of the model.

3.3. Multivariate Adaptive Regression Splines Model. MARS model is a nonlinear and nonparametric regression approach that uses piecewise linear splines to simulate the nonlinear relationship between the dependent and independent variables [30]. The MARS model is built as a linear combination of the following basis functions BFs showed in the following equation:

$$f(x) = \beta_0 + \sum_{i=1}^{m} \beta_i \mathrm{BF}_i, \tag{3}$$

where β_i, $i = 1, 2, \ldots..m$ are unknown coefficients that can be estimated using the least square method and m is the number of terms found in the final model using a forward backward stepwise process. BF_i is the $i-th$ basis function defined from piecewise linear basis functions and based on knot C. BF_i is calculated from the following set functions that is showed in the following equation:

$$\mathrm{BF}_i = \left\{ \left|x - C_i\right|^+, \ \left|C_i - x\right|^+ \right\}, \tag{4}$$

where $|x - C_i|^+$ and $|C_i - x|^+$ are given by

$$\left|x - C_i\right|^+ = \max\left(0, x - C_i\right), \tag{5}$$

$$\left|C_i - x\right|^+ = \max\left(0, C_i - x\right). \tag{6}$$

Finally, the predicted model is built with m numbers of BFs to provide the lowest generalized cross validation (GCV) value that is calculated by the following equation:

$$\mathrm{GCV} = \frac{\mathrm{SSE}i}{\left(1 - vmi/n\right)^2}, \tag{7}$$

where SSEi is the sum of square error, where SSE$i = \sum (O - f(x))^2$ and v is the smoothing parameter.

3.4. Analysis of Variance (ANOVA). ANOVA is a statistical analysis technique which is developed by R.A. Fisher in the 1920s. ANOVA can be used for many purposes such as comparing group mean. Two hypotheses are applied to determine the output of the comparison, namely, null hypothesis and alternative hypothesis. ANOVA is also known as analysis of an analysis of variance because it compares two variance estimations, namely, variation within groups and variation between groups. In this paper we perform a one-way ANOVA. The purpose of a one-way between-groups ANOVA is to show if there are any differences among the means of two or more groups/models. When at least two of

the groups/models have means that are significantly different from each other, the ANOVA test is significant in this case. However, it does not reveal which of the groups/models are different.

3.5. Social Spider Optimization Algorithm. The social spider optimization (SSO) is swarm intelligence-based metaheuristic algorithm [31]. SSO is chosen in this study because it is a new heuristic algorithm that solves difficult optimization problems. It is a vital model to search for the global optimum through performing a simulation to the social spider behavior. SSO mimics the behaviors of spiders. Spiders identify the position of prey via the vibration that occurred on the spider web. Any unusual vibration is a sign for the social spider to search for food and move into the source of vibration. The search area of SSO uses chain-like social spider structure. The direction of the food is determined by insects through signals generated through vibrations from the spider web. Equations (8) and (9) define the SSO operation.

The vibration intensity [32] at position x is calculated by the following equation:

$$I(x, x, \text{iter}) = \log\left(\frac{1}{F(x) - C} + 1\right), \tag{8}$$

where $F(x)$ denotes the cost function and C denotes a constant number.

The iteration attenuation is given by the following equation:

$$I(x_1, x_2, \text{iter}) = I(x_1, x_2, \text{iter}) \times \exp\left(-\frac{D(x_1, x_2)}{\sigma \times r_a}\right), \tag{9}$$

where $D(x_1, x_2)$ indicates the distance between x_1 and x_2. The standard deviation of all members along one searched dimension is indicated by σ. The free parameter is r_a.

3.6. The Proposed Prediction Model. This paper combines both LR model and MARS model based on SSO to develop an optimized LR-MARS prediction model that predicts crude oil demand. The proposed LR-MARS model is developed based on five main stages as demonstrated in Figure 1. There are five stages used to develop the LR-MARS model which are (1) data collection and data preprocessing stage, (2) determining training and testing sets, (3) LR model and MARS model, (4) using SSO, and (5) performance evaluation.

3.6.1. Data Collection and Preprocessing. The process of data collection starts with collecting different features for crude oil demand from different sources. Data are tracked and verified for any externality or inconsistency. For example, the gross domestic product (GDP) feature is gathered from the sources: OPEC, IEA, International Monetary Fund (IMF), Saudi Statistics Authority, and World Bank. The data used in this article come from various sources and cover the period 1980–2015 [3]. Features such as year, oil demand, GDP, population, Brent prices, Light-Duty Vehicles (LDV), and Heavy-Duty Vehicles (HDV) are shown in Table 1.

Table 1 describes a number of statistical metrics such as mean, standard error, median, standard deviation, etc., of the features of the dataset which are oil demand, GDP, population, LDV, and HDV. For instance, the maximum value of the oil demand is 3318.656317, the minimum value is 602, and the standard deviation is 774.0563839.

In statistics, the correlation matrix shows the correlation coefficients between variables. The correlation matric of the features of Saudi Arabia oil demand dataset is shown in Table 2. Each cell represents the correlation value between two variables. As can be seen in Table 2, the correlation coefficient of the features is closer to 1 which means that we have strong positive correlation between each two features in the dataset.

Data preprocessing stage is an essential step in machine learning [33]. The quality of the data can directly affect the ability of the models to learn; thus, it is critical that we preprocess our data before using data as inputs into the proposed model. In this paper, preprocessing is done using normalization. If the data contains input values with varying scales, normalization can be used to scale these values. Normalization scales each input value separately through subtracting the mean (centering) and dividing by the standard deviation in order to change the distribution's mean to zero and standard deviation to one [33]. Normalization is calculated using the following equation:

$$z = \frac{x - \mu}{\sigma}, \tag{10}$$

where x is the input value, μ is the mean value, and σ is the standard deviation value. Mean value (μ) is calculated using the following equation:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i. \tag{11}$$

Standard deviation (σ) is calculated using the following equation:

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (x_i - \mu)^2}. \tag{12}$$

3.6.2. Training and Testing Sets. The crude oil demand dataset is split into train data (90%) and test data (10%). Following that, the train data is split further into training set (50% of train data) and validation set (50% of train data).

3.6.3. LR Model and MARS Model. The training set (50% of train data) is trained by LR model and the validation set (50% of train data) is used as an input to the LR model to make predictions through LR model. LR model provides two predictions (validation prediction set and test prediction set). Finally, the validation prediction set will be trained with MARS model to create LR-MARS model. This LR-MARS is

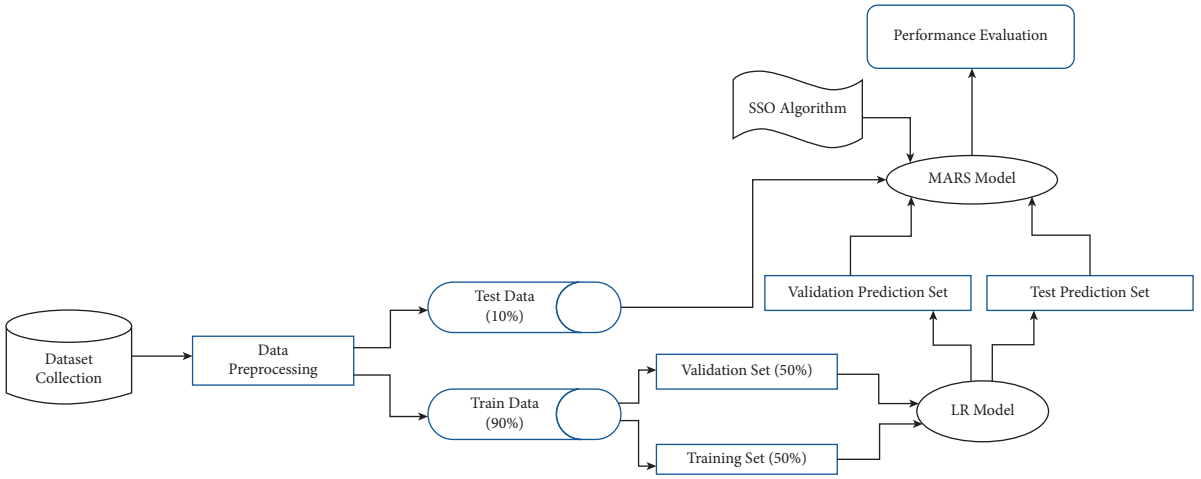


Figure 1: The stages for the proposed LR-MARS model.

Table 1: Statistical data analysis of features for Saudi Arabia oil demand.

Statistic	Oil demand (MBD)	GDP (bill SAR)	Population (MM)	Brent price ($/Bbl)	LDV (M)	HDV (M)
Mean	1539.087014	1481.705056	19.57855192	41.68838889	5390.042144	3587.345317
Standard error	129.0093973	79.40556539	1.079092403	5.189740084	578.5509528	288.0411458
Median	1230.609065	1359.74	19.2705	28.5755	4214.092	3085.9235
Standard deviation	774.0563839	476.4333924	6.474554418	31.13844051	3471.305717	1728.246875
Sample variance	599163.2855	226988.7774	41.91985491	969.6024771	12049963.38	2986837.26
Kurtosis	−0.247425069	−0.292910639	−1.010727991	0.214674645	−0.121742024	−0.158338512
Skewness	0.962853111	0.655422825	0.157041289	1.232271204	0.950014993	0.857854418
Minimum	602	778.227	9.32	12.713	1268.38	1199.523
Maximum	3318.656317	2545.24	31.016	111.62	13749.2784	7713.0792
Confidence level (95.0%)	261.9030003	161.2018679	2.190674043	10.53573249	1174.520876	584.7546137

Table 2: Correlation matrix of features for Saudi Arabia oil demand dataset.

	Oil demand (MBD)	GDP at 2010	Population	Brent prices	LDV	HDV
Oil demand (MBD)	1					
GDP at 2010	0.945006302	1				
Population pop	0.952639467	0.924158747	1			
Brent prices	0.84962017	0.834478322	0.747342982	1		
LDV	0.996837432	0.954042182	0.962077275	0.828231293	1	
HDV	0.994288957	0.944879459	0.970942636	0.807996732	0.998513235	1

used to make final predictions on the test prediction set to obtain the final predicted output that is in turn compared with the actual test data.

3.7. Performance Metrics. To validate the performance and effectiveness of the prediction models proposed, five error analysis criteria are introduced to evaluate the proposed models, as shown in equations (13)–(17), where y_{real_i} is the actual values, y_{pred_i} is the predicted values, and $\overline{y}$ is the mean value of actual values [24, 34]. For each model, the performance is evaluated using the Mean Absolute Error (MAE), Median Absolute Error (MedAE), Mean Square Error (MSE), Root Mean Square Error (RMSE), and R-squared (R^2).

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|y_{\text{real}_i} - y_{\text{pred}_i}\right|, \tag{13}$$

$$\text{MedAE} = \text{median}\left(\left|y_{\text{real}_1} - y_{\text{pred}_1}\right|, \ldots\ldots, \left|y_{\text{real}_N} - y_{\text{pred}_N}\right|\right), \tag{14}$$

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}\left(y_{\text{real}_i} - y_{\text{pred}_i}\right)^2, \tag{15}$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_{\text{real}_i} - y_{\text{pred}_i}\right)^2}, \tag{16}$$

Table 3: Comparison of prediction performances using machine learning models.

Models	LR-MARS	LR	Ridge regression
MAE	0.024	0.042	0.055
MedAE	0.023	0.047	0.054
MSE	0.0007	0.0026	0.0036
RMSE	0.02	0.05	0.06
R^2	99.9%	99.6%	99.4%

Table 4: Comparison of prediction performances using LR-MARS model with different cases.

Models	Case 1	Case 2	Case 3
MAE	0.024	0.034	0.046
MedAE	0.023	0.041	0.048
MSE	0.0007	0.0038	0.0041
RMSE	0.02	0.031	0.036
R^2	99.9%	99.3%	99.1%

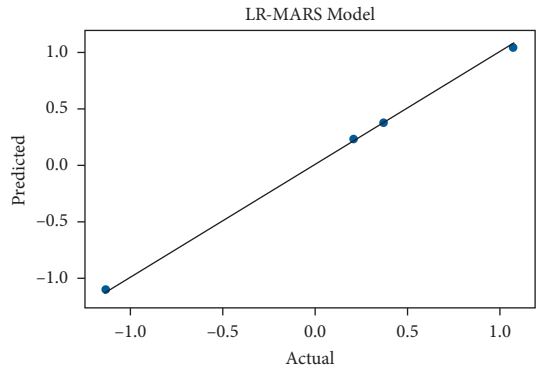


Figure 2: A cross-plot of the actual and predicted crude oil demand using LR-MARS model.

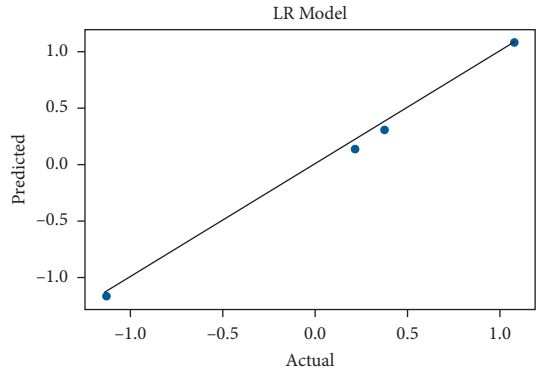


Figure 3: A cross-plot of the actual and predicted crude oil demand using LR model.

$$R^2 = 1 - \frac{\sum_{i=1}^{N}\left(y_{\text{real}_i} - y_{\text{pred}_i}\right)^2}{\sum_{i=1}^{N}\left(y_{\text{real}_i} - \overline{y}\right)^2}. \tag{17}$$

4. Results and Discussion

The implementation of the models is done using Google Colab notebook. Google Colab notebook helps to write and execute python in the browser, where it is an open source and widely used for the implementation of machine learning algorithms such as regression, classification, and clustering. To evaluate the performance of the optimized LR-MARS model in crude oil demand prediction more effectively, other models are chosen for comparison. Furthermore, the models commonly used in machine learning are chosen. SSO has been used to perform tuning to the two hyperparameters (penalty term and maximum number of basis functions (BFs)). The population of SSO metaheuristic algorithm consists of 30 members and the

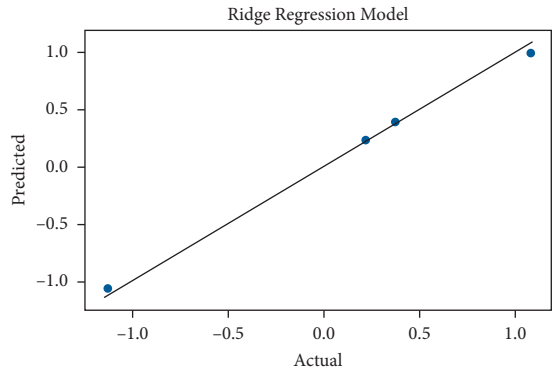


Figure 4: A cross-plot of the actual and predicted crude oil demand using ridge regression model.

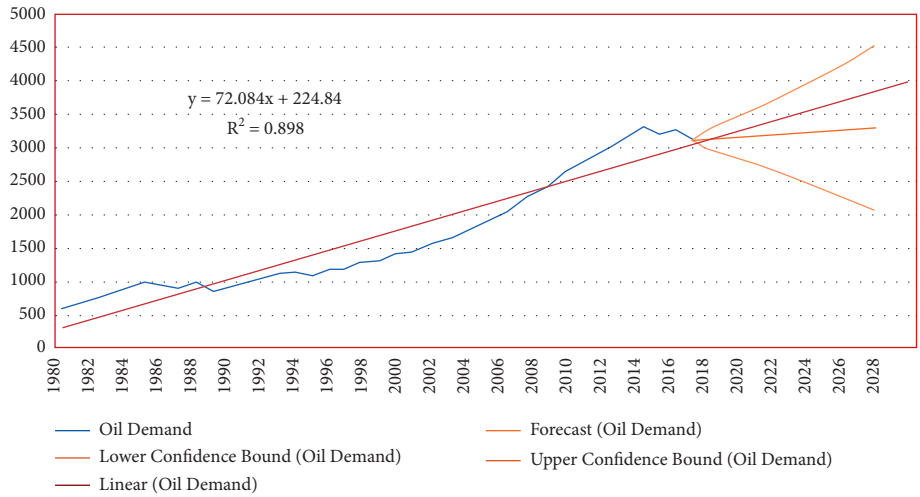


Figure 5: A cross-plot of the actual and predicted crude oil demand using ANOVA model.

Table 5: Comparison of prediction performance for LR-MARS optimized model and ANOVA model, respectively.

Models	LR-MARS	ANOVA
MSE	0.0007	0.0493
RMSE	0.02	0.2374
R^2 (%)	99.9%	89%

Table 6: Analysis of the source of variation.

Source of variation	Sum of square SS	Degree of freedom DF	Mean of square MS	F-value	P value	$F_{critical}$
Between groups	0.001588403	3	0.000529468	10.00246025	0.00982156	3.490294819
Within groups	2.582510317	12	0.215209193			
Total	2.58409872	15				

maximum number of generations is 100. The output of the optimization process is that the maximum number of basis functions (BFs) is 42 and the penalty term is 1.46. The prediction model proposed in this paper, which combines linear regression model with multivariate adaptive regression splines model, has shown high prediction accuracy when predicting crude oil demand in Saudi Arabia. To effectively evaluate the performance of LR-MARS in crude oil demand prediction, traditional prediction models of machine learning are used in this paper as comparative experiments. During the experiment, LR model and ridge regression model are used for crude oil demand prediction as comparative tests. To objectively evaluate and describe the performance of the three prediction models, the prediction error values of each model are calculated according to equations (13)–(17). The experimental results of MAE, MedAE, MSE, RMSE, and R^2 of the test data are shown in Table 3.

Among all the experimental models in Table 1, ridge regression model has the largest error, and its MAE, MedAE, MSE, RMSE, and R^2 are 0.055, 0.054, 0.0036, 0.06, and 99.4%, respectively. The MAE, MedAE, MSE, RMSE, and R^2 of LR model are 0.042, 0.047, 0.0026, 0.05, and 99.6%, respectively. The error of LR-MARS model with optimizing the two hyperparameters (penalty term and maximum number of basis functions (BFs)) using SSO algorithm is the smallest; its MAE, MedAE, MSE, RMSE, and R^2 are 0.024, 0.023, 0.0007, 0.02 and 99.9%, respectively, which is significantly lower than the other two models. It can be seen from Table 3 that LR-MARS model with optimizing the two hyperparameters (penalty term and maximum number of basis functions (BFs)) using SSO algorithm has a high accuracy in predicting crude oil demand and is more effective than the other models. Table 4 demonstrates a comparison of LR-MARS model with different cases: case 1, optimizing the two hyperparameters (penalty term and maximum number of basis functions (BFs)) using SSO algorithm, case 2, optimizing the one hyperparameter (penalty term) using SSO algorithm, and third, without optimizing any hyperparameter.

Figures 2–4 show a cross-plot of the actual and predicted crude oil demand using LR-MARS model, LR model and ridge regression model, respectively.

4.1. Analysis of Variance (ANOVA). In this section, we use ANOVA for two purposes. The first purpose is to predict the crude oil demand in Saudi Arabia. The second purpose is using ANOVA to compare the actual test data and the predicted data results between LR-MARS, LR, and ridge regression model, respectively. R^2 which is also known as coefficient of determination, is used to calculate how close the data are to the fitted regression line. The value ($R^2 = 0.898$) indicates a better fit for the model as shown in Figure 5.

4.2. ANOVA Predicting Result. ANOVA is used as a prediction model. Table 5 provides a comparison of ANOVA prediction model and the proposed LR-MARS optimized model. The results show that LR-MARS optimized model gives a high performance comparing to ANOVA model.

In Table 6, the analysis of the source of variation is carried out in two ways: between groups and within groups. Between-groups analysis determines the source of variance of LR-MARS, LR, and ridge regression models, respectively. Within-groups analysis identifies the experimental error between the group and itself. From the ANOVA results in Table 6, SS = 2.582510317, while Mean Square MS = 0.215209193. Therefore, we can conclude that the null hypothesis was rejected because $F_{\text{critical}} = 3.490294819$ and $F = 10.00246025$, where $F_{\text{critical}} < F$. Moreover, since the P value is less than 0.05 (i.e., $0.00982 < 0.05$), this is another indication of the significant differences in the attribute (crude oil demand) between LR-MARS, LR, and ridge regression models, respectively, and therefore is another evidence to reject the null hypothesis.

5. Conclusion

In this paper, a hybrid model called LR-MARS is developed for predicting the crude oil demand in Saudi Arabia. This paper used historical data of one of the world's largest oil producers (Saudi Arabia) to demonstrate the applicability and effectiveness of the proposed LR-MARS model. The dataset used in the LR-MARS consists of seven features: time, oil demand, GDP, population, Brent crude prices, LDV, and HDV. The LR-MARS model is a combination of linear regression model and multivariate adaptive regression splines (MARS) model. We also used SSO algorithm for optimizing two hyperparameters, namely, penalty term and maximum number of basis functions (BFs) for the MARS model. To evaluate the performance of LR-MARS optimized model, we used MAE, MedAE, MSE, RMSE, and R^2 to examine and test the predictions performance for the LR-MARS model that are 0.024, 0.023, 0.0007, 0.02, and 99.9%, respectively. We have also compared LR-MARS optimized model to other machine learning prediction models. The optimized LR-MARS model is more accurate in predicting crude oil demand in Saudi Arabia than other models. Moreover, we have used ANOVA as prediction model to predict the crude oil demand in Saudi Arabia and also to compare the actual test set and predicted results between LR-MARS, LR, and ridge regression models. This paper will be useful for oil demand planning, setting strategies, and future oil investments. Due to the limitation in obtaining some features and the inconsistency of scaling some data, these limitations of features will lead to a certain range of errors in data-processing process and prediction process. Therefore, other possible influencing features can be considered as input variable. As a direction of future work, as splines can be modelled by adding more knots, this will help in increasing the model flexibility. Moreover, cubic spline model and natural cubic spline model can be used to enhance the results.

Data Availability

The data used in this paper were obtained from different sources (OPEC, IEA, International Monetary Fund (IMF), Saudi Statistics Authority, and World Bank) and cover the period 1980 to 2015 [3].

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors would like to acknowledge Taif University Researchers Supporting Project number (TURSP-2020/292), Taif University, Taif, Saudi Arabia, for funding this research.

References

[1] H. Duan, G. R. Lei, and K. Shao, "Forecasting crude oil consumption in China using a grey prediction model with an optimal fractional-order accumulating operator," *Complexity*, vol. 2018, Article ID 3869619, 12 pages, 2018.

[2] Y. He, S. Wang, and K. K. Lai, "Global economic activity and crude oil prices: a cointegration analysis," *Energy Economics*, vol. 32, no. 4, pp. 868–876, 2010.

[3] S. M. Al-Fattah, "Application of the artificial intelligence GANNATS model in forecasting crude oil demand for Saudi Arabia and China," *Journal of Petroleum Science and Engineering*, vol. 200, Article ID 108368, 2021.

[4] R. F. Engle, "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," *Econometrica*, vol. 50, no. 4, pp. 987–1007, 1982.

[5] S. Aiguo and L. Jiren, "Evolving Gaussian RBF network for nonlinear time series modelling and prediction," *Electronics Letters*, vol. 34, no. 12, pp. 1241–1243, 1998.

[6] A. Hatami-Marbini and F. Kangi, "An extension of fuzzy TOPSIS for a group decision making with an application to Tehran stock exchange," *Applied Soft Computing*, vol. 52, pp. 1084–1097, 2017.

[7] F. Gaxiola, P. Melin, F. Valdez, and O. Castillo, "Interval type-2 fuzzy weight adjustment for backpropagation neural networks with application in time series prediction," *Information Sciences*, vol. 260, pp. 1–14, 2014 Mar 1.

[8] P. Garrard, V. Rentoumi, B. Gesierich, B. Miller, and M. L. Gorno-Tempini, "Machine learning approaches to diagnosis and laterality effects in semantic dementia discourse," *Cortex*, vol. 55, pp. 122–129, 2014.

[9] P. Aleksandar, P. Silvana, and Z. P. Valentina, "Multiple linear regression model for predicting bidding price," *Technics Technologies Education Management*, vol. 10, no. 3, pp. 386–393, 2015.

[10] A. M. Elshewey, "Machine learning regression techniques to predict burned area of forest fires," *International Journal of Soft Computing*, vol. 16, no. 1, pp. 1–8, 2021.

[11] Q. Liang, Y. Fan, and Y. M. Wei, "A long-term trend forecasting approach for oil price based on wavelet analysis," *Chinese Journal of Management Science*, vol. 13, no. 1, pp. 30–36, 2005.

[12] L. Wu, S. Liu, L. Yao, S. Yan, and D. Liu, "Grey system model with the fractional order accumulation," *Communications in Nonlinear Science and Numerical Simulation*, vol. 18, no. 7, pp. 1775–1785, 2013.

[13] J. Tuo and S. Yanbing, "Summary of world oil price forecasting model," in *Proceedings of the 2011 Fourth International Symposium on Knowledge Acquisition and Modeling*, pp. 327–330, IEEE, Sanya, China, October 2011.

[14] A. Azadeh, M. Moghaddam, M. Khakzad, and V. Ebrahimipour, "A flexible neural network-fuzzy mathematical programming algorithm for improvement of oil price estimation and forecasting," *Computers & Industrial Engineering*, vol. 62, no. 2, pp. 421–430, 2012.

[15] I. B. Ibrahim and C. Hurst, "Estimating energy and oil demand functions," *Energy Economics*, vol. 12, no. 2, pp. 93–102, 1990.

[16] H. G. Huntington, "OECD oil demand," *Energy Economics*, vol. 15, no. 1, pp. 49–56, 1993.

[17] N. Krichene, "World crude oil and natural gas: a demand and supply model," *Energy Economics*, vol. 24, no. 6, pp. 557–576, 2002.

[18] P. K. Narayan and R. Smyth, "A panel cointegration analysis of the demand for oil in the Middle East," *Energy Policy*, vol. 35, no. 12, pp. 6258–6265, 2007.

[19] J. Xiong and P. Wu, "An analysis of forecasting model of crude oil demand based on cointegration and vector error correction model (VEC),"in *Proceedings of the 2008 International Seminar on Business and Information Management*, vol. 1, pp. 485–488, IEEE, Wuhan, China, December 2008.

[20] W. P. Nel and C. J. Cooper, "A critical review of IEA's oil demand forecast for China," *Energy Policy*, vol. 36, no. 3, pp. 1096–1106, 2008.

[21] A. Azadeh, M. Khakestani, and M. Saberi, "A flexible fuzzy regression algorithm for forecasting oil consumption estimation," *Energy Policy*, vol. 37, no. 12, pp. 5567–5579, 2009.

[22] E. Assareh, M. A. Behrang, M. R. Assari, and A. Ghanbarzadeh, "Application of PSO (particle swarm optimization) and GA (genetic algorithm) techniques on demand estimation of oil in Iran," *Energy*, vol. 35, no. 12, pp. 5223–5229, 2010.

[23] N. Wei, C. Li, X. Peng, F. Zeng, and X. Lu, "Conventional models and artificial intelligence-based models for energy consumption forecasting: a review," *Journal of Petroleum Science and Engineering*, vol. 181, Article ID 106187, 2019.

[24] A. Sina and D. Kaur, "Short term load forecasting model based on kernel-support vector regression with social spider optimization algorithm," *Journal of Electrical Engineering & Technology*, vol. 15, no. 1, pp. 393–402, 2020.

[25] Q.-T. Bui, Q.-H. Nguyen, V. M. Pham et al., "A novel method for multispectral image classification by using social spider optimization algorithm integrated to fuzzy C-mean clustering," *Canadian Journal of Remote Sensing*, vol. 45, no. 1, pp. 42–53, 2019.

[26] B. A. Emine and E. Ülker, "Dıscrete socıal spıder algorıthm for the travelıng salesman problem," *Artificial Intelligence Review*, vol. 54, no. 2, pp. 1063–1085, 2021.

[27] A. K. Marandi and D. A. Khan, "An impact of linear regression models for improving the software quality with estimated cost," *Procedia Computer Science*, vol. 54, pp. 335–342, 2015.

[28] N. Aghdaei, G. Kokogiannakis, D. Daly, and T. McCarthy, "Linear regression models for prediction of annual heating and cooling demand in representative Australian residential dwellings," *Energy Procedia*, vol. 121, pp. 79–86, 2017.

[29] J. M. Pereira, M. Basto, and A. F. D. Silva, "The logistic lasso and ridge regression in predicting corporate failure," *Procedia Economics and Finance*, vol. 39, pp. 634–641, 2016.

[30] B. Keshtegar, C. Mert, and O. Kisi, "Comparison of four heuristic regression techniques in solar radiation modeling: kriging method vs. RSM, MARS and M5 model tree," *Renewable and Sustainable Energy Reviews*, vol. 81, pp. 330–341, 2018.

[31] D. T. Vu, X.-L. Tran, M.-T. Cao, T. C. Tran, and N.-D. Hoang, "Machine learning based soil erosion susceptibility prediction using social spider algorithm optimized multivariate adaptive regression spline," *Measurement*, vol. 164, Article ID 108066, 2020.

[32] J. Q. James and V. O. Li, "A social spider algorithm for global optimization," *Applied Soft Computing*, vol. 30, pp. 614–627, 2015.

[33] D. Singh and B. Singh, "Investigating the impact of data normalization on classification performance," *Applied Soft Computing*, vol. 97, Article ID 105524, 2020.

[34] M. Y. Shams, O. M. Elzeki, L. M. Abouelmagd, A. E. Hassanien, M. A. Elfattah, and H. Salem, "HANA: a healthy artificial nutrition analysis model during COVID-19 pandemic," *Computers in Biology and Medicine*, vol. 135, Article ID 104606, 2021.

Hindawi
Discrete Dynamics in Nature and Society
Volume 2021, Article ID 8166376, 14 pages
https://doi.org/10.1155/2021/8166376

Research Article

A New Method for Identifying Key and Common Themes Based on Text Mining: An Example in the Field of Urban Expansion

Yanwei Zhang,[1] **Xinhai Lu**,[1,2] **Chaoran Lin**,[3,4] **Feng Wu**,[1] **and Jinqiu Li**[4]

[1]*College of Public Administration, Huazhong University of Science and Technology, Wuhan 430074, China*
[2]*College of Public Administration, Central China Normal University, Wuhan 430079, China*
[3]*College of Computer Science and Technology, Harbin Engineering University, Harbin, Heilongjiang 150001, China*
[4]*School of Economics & Management, Harbin Engineering University, Harbin 150001, China*

Correspondence should be addressed to Chaoran Lin; hcab@qq.com

Received 3 August 2021; Accepted 26 September 2021; Published 14 October 2021

Academic Editor: Jorge E. Macias-Diaz

Urban land use is a core area of multidisciplinary research that involves geography, land science, and urban planning. With the rapid progress of global urbanization, urban expansion has become a research focus in recent years. Therefore, how to scientifically and accurately identify key and common themes in the urban expansion literature has become crucial for scientific research institutions in various countries. This paper proposes a new framework for identifying such themes based on an analysis of scientific literature and by using text mining and thematic evolutionary analysis. First, the latent Dirichlet allocation algorithm is used to capture the thematic clustering of scientific literature. Second, the key degree of the thematic node in the thematic evolution transfer network is used to represent the key feature of a theme, and the PageRank algorithm is employed to measure the critical score of this theme. When recognizing common themes, the common features of various themes are digitized and mapped to a specially selected quadratic function to measure the degree of commonness. Finally, the hidden Markov model is used to build a thematic prediction model. This method can efficiently identify key and common themes from the literature and provide theoretical and technical support for future research in related fields.

1. Introduction

The increasingly drastic land-use changes during the process of urbanization are important factors that affect the global social economy and ecological stability [1, 2]. For half a century, many areas in the world have undergone rapid urbanization, thereby resulting in the continuous emergence of large cities and megacities [3, 4]. Accordingly, urban expansion has become a research hotspot in the fields of geography, economics, ecology, environmental science, and sociology [5]. Urban expansion refers to an increase in the total area of urban land and the outward development of land use under the influence of economic development, population growth, urban planning, and urbanization [6]. However, due to the lack of reasonable planning and guidance for urbanization, the process of urban expansion is often associated with an excessive demand for urban land and the disorderly development of urban fringe areas [7], both of which have a negative impact on regional economic growth, the layout of production and living spaces, residential types, urban morphology, and ecological environment [8–10]. Therefore, how to guarantee a continuous demand for urban land for social-economic development in the new era and how to clarify the formation mechanism of urban growth in order to reasonably control the urban scale, delimit the urban growth boundary, and optimize the spatial pattern of urban land have become the working emphases of current urban management [11, 12]. Scholars have carried out extensive research to clarify those factors that drive urban expansion [13, 14], strengthen the development and application of dynamic monitoring technologies for urban expansion [15, 16], and accelerate sustainable urban spatial planning [17].

The scientific literature is an important and authoritative knowledge carrier. Using bibliometrics and text mining methods to study the thematic evolution and thematic prediction of the massive urban expansion literature can help trace the development trajectory and grasp the flow of knowledge in the field of urban expansion. In recent years, some scholars have qualitatively combed the findings of urban expansion research from multiple levels, dimensions, and perspectives [18, 19]. Some scholars have also used knowledge maps and bibliometric methods to quantify and visualize the results and topics in urban expansion research [20, 21]. However, as the number of documents has increased exponentially, the types of these documents have also become increasingly abundant. Accordingly, thematic identification is increasingly being used in big data scientific literature identification. When automatic thematic identification is faced with high data dimensions and complex data types, traditional thematic identification methods may become ineffective.

This article divides themes into key and common themes. Key themes play relatively important roles in the urban expansion field. These themes have attracted significant concern, are mature, and have development potential in each time window. The evolutionary process of key themes plays an important role in describing the future development of a theme [22]. In addition, in the globalization context, cities have increasingly become the power centers of global social-economic development. Whether in developed or developing countries, the status and role of cities in national development have become increasingly important and have begun to represent countries in a global competition. Therefore, urban development and expansion play key roles in the competition among countries in the context of globalization. This article then defines common themes as "themes that have received equal amounts of attention from scholars in developed and developing countries."

How to accurately identify themes has always been a challenge in the field of bibliometrics. Statistical methods based on word frequency and co-occurrence frequency are widely used in thematic identification. However, these methods only simulate the literature as a language package and do not fully consider the relationship among themes, and revealing the rich thematic information contained in the literature is not an easy task [23]. The topic model represented by latent Dirichlet allocation (LDA) uses Dirichlet distribution to describe the literature generation process and obtains vocabulary clusters by maximizing the co-occurrence probability of keywords. This model can avoid parameter explosion and overfitting problems and can effectively extract hidden themes from the literature [24]. However, this method requires predefined empirical values and can only reveal the potential semantic relationships among themes. To solve these problems, scholars have recently examined thematic identification by constructing networks and comprehensively evaluating certain indicators, such as network centrality. For example, by constructing a citation network, Shibata et al. [25] demonstrated the novelty of themes from the time and function dimensions and detected the emerging themes in regenerative medicine. Small [26] and Lee and Choe [27] not only considered the novelty of themes in their identification method but also employed network time series analysis and structural hole theory to measure the characteristics of thematic growth and influence. However, these studies conduct static analysis based on historical literature data, which are unable to reflect the dynamic development characteristics of a network in real-time. Therefore, how to construct a dynamic network of themes in the field of urban expansion and how to accurately identify key and common themes that can help scholars achieve breakthroughs in this field are crucial.

Identifying themes is not the end. By expanding research on thematic evolution, this article aims to predict the future development of these themes. Researchers have divided thematic prediction methods into qualitative and quantitative analyses based on different theoretical foundations. While qualitative methods are often limited by subjective judgment, quantitative methods are highly scientific. The most commonly used forecasting methods include the gray forecasting method [28], the life cycle method [29], time series analysis [30], and the neural network forecasting method [31]. However, uncertainty, ambiguity, and randomness are essential phenomena in scientific research, and the above models generally ignore the randomness in the development of technological innovation. The hidden Markov model (HMM) with a double stochastic process can describe the Markov stochastic process of mutual transfer among various themes, reveal the potential evolution path, and provide a basis for predicting future thematic development [32].

This article first applies the LDA model for topic modelling based on the title and abstract of urban expansion literature to obtain a detailed thematic classification. Second, in identifying key themes, the key degree of the thematic node in the thematic evolutionary transfer network is regarded as the key feature of themes. Third, the PageRank algorithm is employed to measure the criticality score of each thematic node in the thematic relationship network. When identifying the common theme, the common features of various themes are digitized and mapped to a specially selected quadratic function to measure the degree of commonality. Finally, by using the HMM, the future development trend of each theme is predicted from the microcosmic angle of the thematic evolution, and a visual display is given.

2. Materials and Methods

2.1. Data Sources. In order to ensure the quality and completeness of the sample, this paper selects the Web of Science core collection database for retrieval. Web of Science includes articles, reviews, editorials, letters, and other document types. Considering that the article is more creative and the results are more complete, this article only selects the article for retrieval. The search formula is TI = "Urban expansion" or "Urban extension" or "Urban Growth Boundaries" or "Urban land growth" or "Urban land expansion" or "Urban sprawl" [21, 33], and the time span is

"1985–2020" (November 4, 2020). A total of 1,933 papers are retrieved, and the bibliographic information is downloaded and summarized in the form of full records (including references). The bibliographic items used in this article include article titles, abstracts, publication years, and reprint addresses, which provide the name, organization, and country information of the corresponding author. Given that international collaborative papers contain information from multiple countries, the corresponding author is used to determine the country of each paper (when the corresponding author is not specified, the first author is used). Each article is attributed to only one country to prevent international co-authored papers from affecting the accuracy of national distinction. After the data deduplication, cleaning, and sorting, a total of 969 documents from developed countries and 1,045 documents from developing countries are obtained.

2.2. Research Methods

2.2.1. Thematic Extraction Module. Scholars have investigated the concept of thematic identification by using the topic model. The current mainstream model adopted in thematic identification is the LDA model proposed by Blei et al. [34]. As a text mining method based on unsupervised machine learning, LDA can dig out potential themes from documents while overcoming the shortcomings of traditional methods in calculating text similarity. In addition, the LDA model can express scientific literature in the form of thematic probability vectors, thereby greatly reducing the dimensionality of the literature data and improving the accuracy of text classification and thematic identification. LDA and its improved models have been widely used in text analysis. The output of these models is usually obtained based on the distribution of words under each theme in order to extract high-frequency keywords to describe the themes and achieve excellent thematic classification results [35]. The hidden themes in the urban expansion literature are assumed to follow the distribution

$$\mathrm{Dir}\left(\theta_d \mid \alpha\right) = \frac{\Gamma\left(\sum_{k=1}^{k} \alpha_k\right)}{\prod_{k=1}^{k} \Gamma\left(\alpha_k\right)} \prod_{k=1}^{k} \theta_{dk}^{\alpha_k - 1}, \tag{1}$$

where θ_{dk} represents the distribution of theme k in scientific literature d. The thematic term distributions $\varnothing_k \sim \mathrm{Dir}(\beta)$ and $\theta_d \sim \mathrm{Dir}(\alpha)$ are generated for themes k and d, respectively, and the thematic term $Z_{dn} \sim \mathrm{Multinomial}(\varnothing_{Z_{dn}})$ is generated for the n-th term in each literature. Therefore, the LDA likelihood model can be described as follows:

$$p(W|\alpha,\beta) = \prod_{d=1}^{D} \int p(\theta_d|\alpha) \prod_{n=1}^{N_d} \sum_{Z_d} p(Z_{dn} \mid \theta_d) p\left(w_{dn} \mid \varnothing_{Z_{dn}}\right) \mathrm{d}\theta_{\mathrm{d}}. \tag{2}$$

This paper uses Heinrich's parameter estimation method, where $\alpha = 50/k$ and $\beta = 0.1$, and Gibbs sampling to obtain the theme set $K = \{k_1, \ldots, k_h\}$ and theme attribution set $D_k = \{j_1, \ldots, j_n\}$ of each paper.

2.2.2. Key Theme Identification Model. Based on the connotation of key themes, during the model construction, the thematic evolution is regarded as a hidden Markov process to obtain the thematic transfer network. There are two dynamics for thematic evolution in the field of urban expansion research: one is the inspiration of historical research results and the emergence of new ideas in the process of thematic evolution. However, due to the lack of a record carrier, this process is an unobservable hidden sequence; the second is that under the impetus of the first kind of driving force, as the research environment changes and unexpected research results appear, scholars constantly adjust their research thinking and then change the research direction. The professional literature effectively records the research results into an observable sequence. The latter constitutes the microfoundation of the former, and the former is the macroscopic manifestation of the latter. Therefore, the thematic evolution in the field of urban expansion can be seen as the superposition of these two processes. This research uses HMM to describe the evolution process of urban expansion theme. By inferring the state transition matrix and the probability distribution of the initial state in the HMM, the confusion and transition matrices between the themes in the evolution of themes are determined, and then the evolution history and future evolution trends of the themes are determined. Afterward, the criticality of these themes is measured based on their network relationship. The PageRank algorithm is then used to calculate the scores of network nodes in the thematic transfer network and serve as the foundation of key theme identification. This process is specifically described as follows:

(1) Set the hidden state random transition sequence set of HMM to $S = \{s_1, \ldots, s_h\}$, where h is the number of themes generated in the LDA model. Suppose that the hidden state sequence generated by the random process is $Q = \{q_1, \ldots, q_t\}$, where $q_t \in S$.

(2) The probability distribution of the transition state is $A = \{\alpha_{ij}\}$, where $\alpha_{ij} = P\{(q_{t+1} = S_j | q_t = S_i)\}$, $1 \le i$, and $j \le N$, and satisfies $\alpha_{ij} \ge 0$, $\sum_{j=1}^{N} \alpha_{ij} = 1$, which suggests that during the development of the urban expansion field, the themes will shift from state S_i to S_j.

(3) When the state is S_i, the probability distribution of the observed variable is $B = \{b_i(v)\} = \{f(Q_t = v|q_t = S_i)\}$, where Q_t is the t-th observation variable. The observation sequence is $O = \{O_1, \ldots, O_t\}$ or the proportion of each theme over the years.

(4) The probability distribution of the initial state of the system is $\pi = \{\pi_i, 1 \le i \le N\}$, where π_i is the occurrence probability of state S_i. Given that a higher frequency of theme co-occurrence will facilitate the shift and evolution among themes, this paper uses the thematic co-occurrence matrix as the initial iteration value of the state transition matrix π_i.

(5) Set the initial value of model training to $O = Q = \pi$. This paper uses the Baum–Welch algorithm [36] to

obtain the following single optimal state transition matrix:

$$U = \pi' = \begin{bmatrix} \pi'_{11} & \cdots & \pi'_{1h} \\ \vdots & \ddots & \vdots \\ \pi'_{h1} & \cdots & \pi'_{hh} \end{bmatrix}, \quad (3)$$

where π'_{nm} is the probability of transition from theme n to theme m. By extracting all π'_{nm} that exceed a certain threshold, a directed graph of the topological relationship between themes in the transfer network can be established. Key theme identification has always been an important research problem in thematic network analysis. PageRank processes the search results of thematic matching based on a web page link analysis. As the most famous web page ranking algorithm, the PageRank algorithm has been widely used to monitor key nodes in various directed, undirected, weighted, or unweighted networks [37, 38]. Applying this algorithm to compute for the centrality of thematic network nodes presents a very meaningful research problem.

When calculating the PageRank value of theme i at each moment in a dynamic thematic network, the network topology structure of the current snapshot and the influence of the previous centrality on the existing network should both be considered. One effective method for achieving this goal is network reconstruction, where the previous network topology relationships are weighed into the current network to construct a new network. To describe the dynamic network G, this dynamic network needs to be sampled at different times. The sampling results are then arranged in a time sequence to obtain the time sequence network $G = <G_1, G_2, \ldots, G_{t-1}, G_t, \ldots, G_n>$, where G_t represents the sampling result at time t or the snapshot at time t. The analysis of the dynamic network is transformed into the analysis of the sequential network [39]. Let $G'_1 = G_1$, $G'_i = \alpha * G_{i-1}{}' + (1-\alpha) * G_i$, $i \geq 2$. In this case, $G' = <G'_1, G'_2, \ldots, G'_n>$ can be obtained. The parameter $\alpha \in [0, 1]$ is used to balance the contribution of the current and previous topologies to the centrality of the network node. The PageRank centrality of node i at time t is treated as the PageRank value of node i in the construction network G'_t. The key theme in this article refers to any mature theme that has development potential in scientific research. Therefore, key themes can easily achieve migrating power in the process of thematic network transfer. Given that the G score can measure the importance of nodes in the process of directed network migration, this paper takes the standardized score G' as the critical score for each theme.

2.2.3. Common Theme Identification Model. To identify the degree of commonness of themes, the selected model should be able to measure the common skewness of different themes in the field of urban expansion. Skewness refers to the numerical characteristics of the asymmetric degree in the statistical data distribution [40–42]. Common skewness in this article refers to the measurement of the direction and degree of skewness of each thematic distribution. Let $k \in K = \{k_1, \ldots, k_m\}$, and theme k corresponds to the number of documents p_k. A type documents (developed countries) correspond to a_k themes, and B type documents (developing countries) correspond to b_k themes. Let $p_k = a_k + b_k$, and define U_{ka} and U_{kb} as the common skewness of A and B type documents in theme k as follows:

$$U_{ka} = \frac{a_k / \sum_{k=1}^{m} a_k}{\left(a_k / \sum_{k=1}^{m} a_k\right) + \left(b_k / \sum_{k=1}^{m} b_k\right)}, \quad (4)$$

$$U_{kb} = \frac{b_k / \sum_{k=1}^{m} b_k}{\left(a_k / \sum_{k=1}^{m} a_k\right) + \left(b_k / \sum_{k=1}^{m} b_k\right)}, \quad (5)$$

where $U_{ka} + U_{kb} = 1$. Formulas (4) and (5) eliminate the influence of the number of A and B documents on common skewness to prevent the difference in the number of documents from affecting the calculation of the co-occurrence degree of themes.

When the common skewness is $U_{ka} = U_{kb} = 1/2$, that is, theme t comprises the themes of A and B type documents, such skewness indicates the highest degree of commonness. By mapping U_{ka} and U_{kb} to the $[0, 1]$ range of the inverted quadratic function, the common function can be monotonized. Without loss of generality, this quadratic function relationship is set to

$$\begin{aligned} C_{ka} &= 1 - 4 * \left(U_{ka} - 1/2\right)^2, \\ C_{kb} &= 1 - 4 * \left(U_{kb} - 1/2\right)^2. \end{aligned} \quad (6)$$

By learning $C_{ka} = C_{kb}$ from the symmetry of the quadratic function, $C_k = C_{ka} = C_{kb}$ can be obtained. The C_k function is then used to measure the degree of commonness. The highest and lowest degrees of commonness are measured when $C_k = 1$ and $C_k = 0$, respectively. The logic structure is shown in Figure 1.

3. Results and Analysis

3.1. Data Preprocessing and Thematic Extraction from the Urban Expansion Literature. The preprocessing work in this article mainly involves word segmentation, removal of stop words, root restoration, and marked information removal. In view of the language characteristics of English articles, the words in a text can be directly divided by spaces and punctuation. Removal of stop word removes those words that do not provide useful information for the text analysis, such as auxiliary words, pronouns, conjunctions, and adverbs. According to the characteristics of the collected urban expansion literature, this article expands the stop words to include some additional words (e.g., "data," "study," and "use") that are unique in the field and appear frequently yet have no effects on the experimental results. Root restoration restores words to their corresponding roots. After such processing, the number of feature items in the sample set can be greatly reduced, and the efficiency of thematic extraction can be improved.

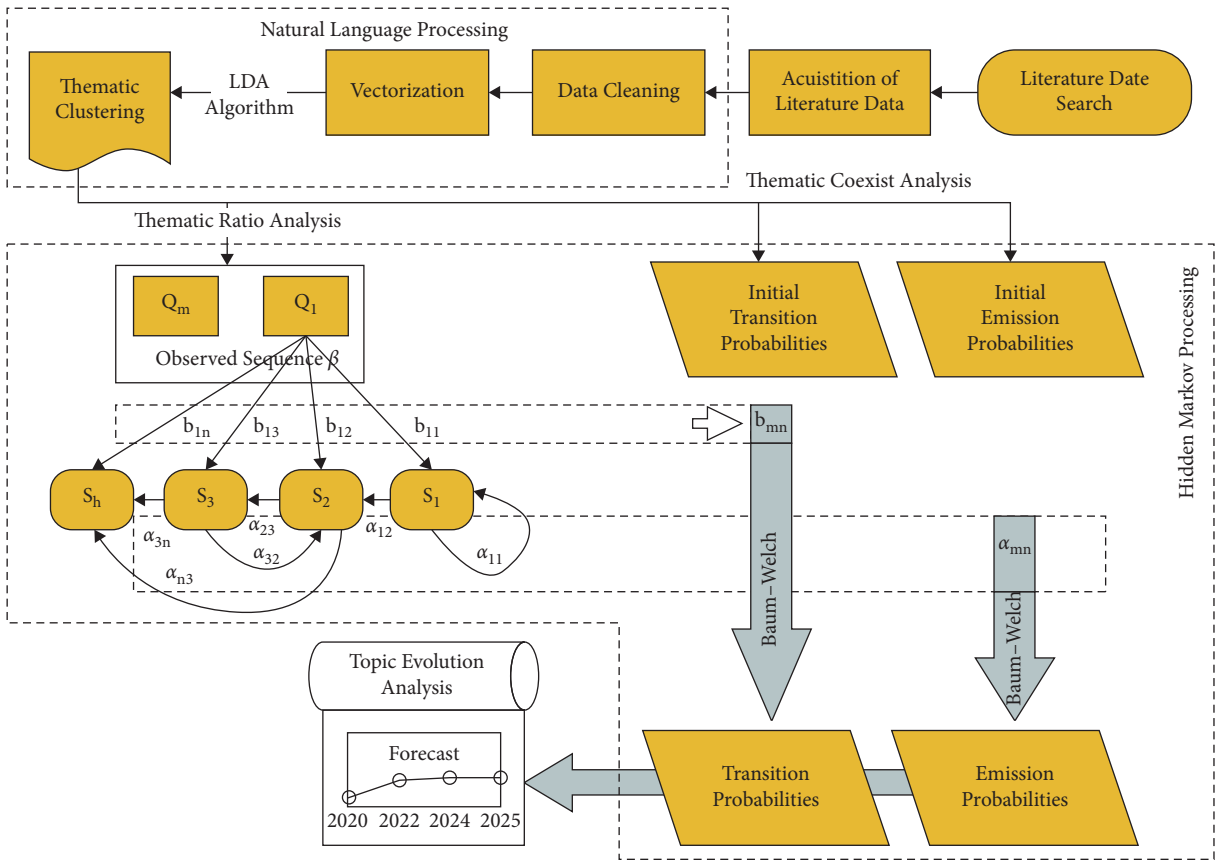


Figure 1: Logic structure flowchart.

Themes are abstract concepts, and the number of themes in the corpus can be quantified by dividing them into different granularities. The number of themes in the LDA model should be specified in advance. A larger corpus corresponds to a greater number of themes, and such number dynamically changes across different time windows. This article uses perplexity to determine the optimal number of themes [43]. Perplexity gradually decreases along with an increasing number of themes. A lower perplexity corresponds to a stronger generalization ability and better performance of the model.

By calculating the perplexity of each theme, the optimal number of themes in the LDA model employed in this work is 29. Experts in the field of urban expansion have read the sample of thematic classification literature and observed a relatively high accuracy (with a classification error rate of less than 3%). The boundaries between the themes are clear, and the division effect is ideal. For ease of reference, these themes are named based on keywords (Table 1).

3.2. Identification of Key Themes in the Urban Expansion Literature. The confusion matrix in the HMM indicates the possibility of transforming a hidden state into an observable state. This probability in turn can measure the threshold barriers for the transition of 29 themes in urban expansion research and can characterize the direction and extent of thematic evolution. The dark (light) squares in the confusion matrix heat map represent those themes that are easy (difficult) to transfer in the innovation process (Figure 2). Most themes in the field of urban expansion show limited movement, and the thematic evolution is relatively stable. Varying degrees of transfer possibilities are also observed among different themes. To highlight the transfer relationship among these themes, this paper draws a confusion relationship network diagram (Figure 3) where the direction of arrows indicates the direction of thematic transfer.

Figure 3 shows that certain themes, including themes 26 (temperature), 15 (urban agglomeration), and 11 (economic development), have a high proportion of transfer inflow and a small proportion of transfer outflow. These themes are identified as core themes in the field of urban expansion. Therefore, transfer inflow and outflow are important manifestations of the criticality of a theme. To measure such criticality, this paper uses the PageRank link analysis algorithm, which obtains the critical evaluation of each node based on PageRank scores. A higher score corresponds to a higher criticality of a theme. The results are shown in Table 2.

Table 2 shows the key themes in the field of urban expansion, including themes 26 (temperature), 15 (urban agglomeration), 11 (economic development), 13 (housing development policy), 17 (surface change), and 9 (population density), of which temperature is the most critical. One obvious feature of urban expansion is the continuous increase in the area and density of various buildings in urban construction, which leads to the transformation of many

Table 1: Keywords and theme naming.

No.	KW1	KW2	KW3	KW4	KW5	ID
1	expansion	urban	china	urbanization	city	urban expansion
2	growth	urban	spatial	pattern	model	spatial pattern
3	city	growth	urban	infrastructure	state	urban infrastructure
4	urban	plan	planning	growth	management	urban planning management
5	urban	scenario	future	growth	prediction	scenario prediction
6	land	change	agricultural	development	urban	agricultural land change
7	sprawl	urban	measure	study	index	urban sprawl
8	urban	study	remote	city	datum	remote
9	population	city	density	large	high	population density
10	metropolitan	region	urban	period	paper	metropolitan region
11	development	spatial	urban	pattern	economic	economic development
12	factor	urban	distance	variable	datum	distance variable
13	policy	development	housing	effect	local	housing development policy
14	public	social	cost	population	economic	social cost
15	urban	level	regional	agglomeration	beijing	urban agglomeration
16	space	development	green	indicator	paper	green space
17	datum	surface	image	change	map	surface change
18	urban	dynamic	analysis	coastal	result	coastal urban
19	transportation	effect	transport	emission	center	transportation emission
20	model	simulation	simulate	cellular	automaton	cellular automata model
21	change	increase	climate	effect	result	climate change
22	associate	health	form	location	physical	physical health
23	forest	increase	zone	urbanization	loss	forest loss
24	landscape	pattern	metric	soil	patch	landscape pattern
25	risk	flood	level	present	environment	flood risk
26	increase	effect	temperature	decrease	degree	temperature
27	urban	global	carbon	time	project	urban carbon
28	ecosystem	natural	service	loss	river	ecosystem service
29	water	impact	quality	environmental	concentration	water quality

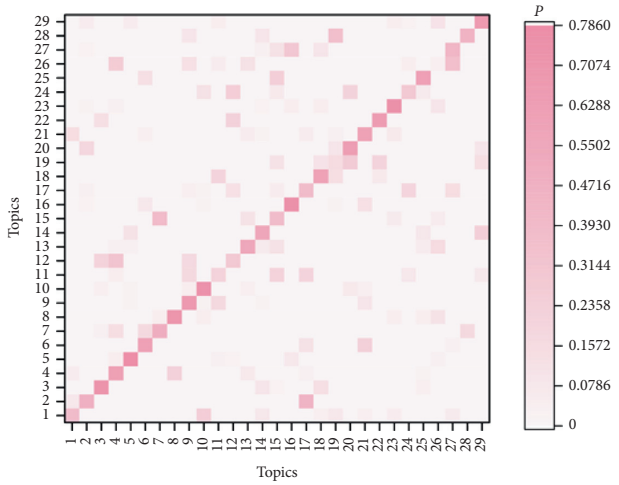


Figure 2: Confusion matrix estimated by HMM.

natural surfaces into impervious surfaces. The changes in the type and spatial structure of land cover affect the storage and transmission of surface temperature, thereby generating urban heat island effects [44]. Using remote sensing technology in analyzing surface thermal infrared information makes the result of urban spatial temperature distribution more accurate than the traditional calculations based on surface meteorological data. Therefore, such information provides a reliable basis for quantitatively studying the spatial distribution of urban thermal environments [45]. Studies on the relationship between urban expansion and surface radiant temperature based on remote sensing technology are of great significance for improving urban thermal environments.

Scholars have also investigated those factors that drive urban expansion and find that economic development (theme 11), population density (theme 9), and housing development (theme 13) have important effects on urban expansion [46, 47]. Urban expansion and economic development conform to the Kuznets curve. During the initial stage of urbanization, economic development requires the development of a large amount of construction land and infrastructure land, thereby resulting in the outward expansion of cities. However, with the adjustment of the industrial structure, the improvement of infrastructure, and the increasing intensiveness of land use, the rate of urban expansion will decline [48]. Moreover, urban land is the main place that supports human life, work, and study. An increase in the urban population will inevitably increase the pressures on housing, transportation, and public facilities. Therefore, the demand of the urban population for space will generate momentum for urban expansion. For example, by studying the law of urban expansion and population growth in the metropolitan regions of the USA, Marshall [49] found that the average land area needed to support a new urban population is twice larger than the per capita land area of the existing city. Moreover, due to the agglomeration economy

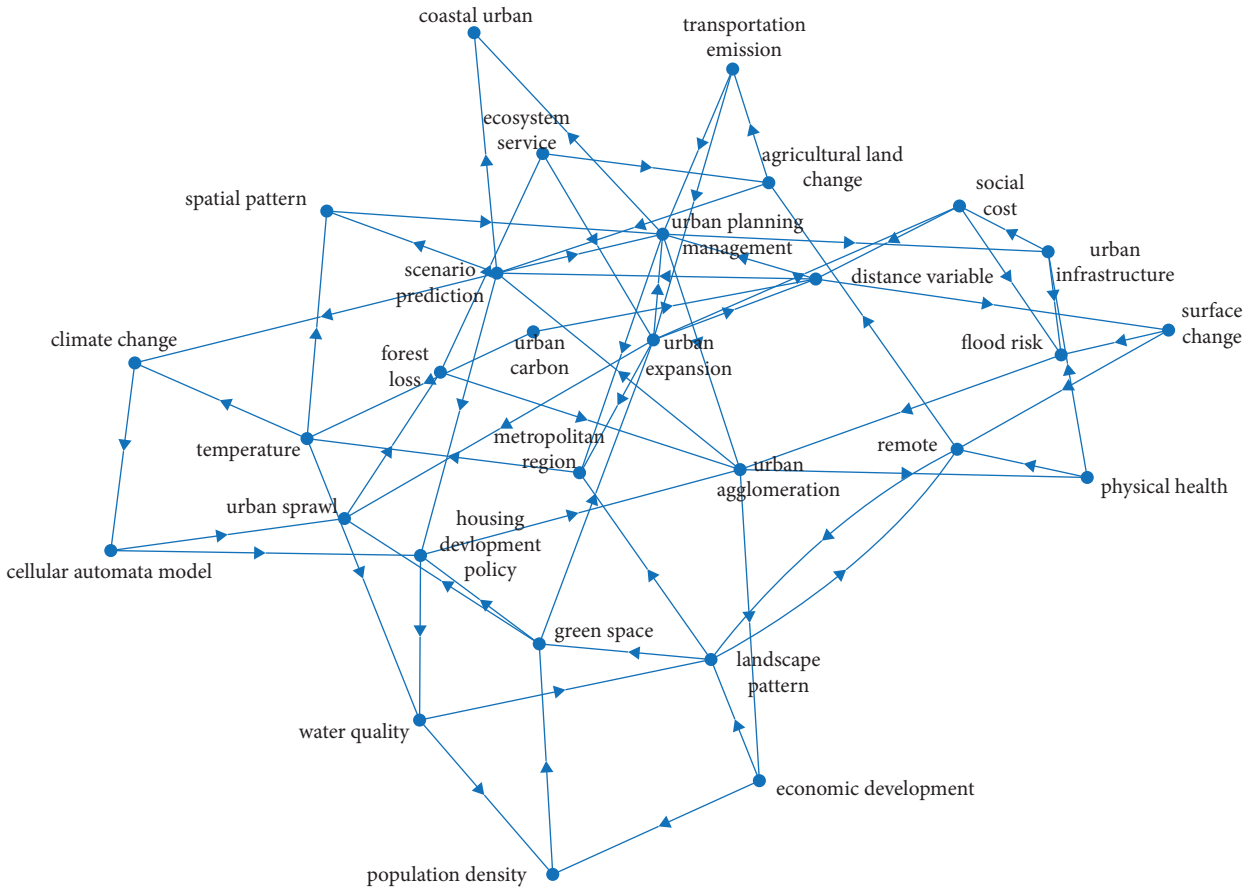


Figure 3: Confusion relationship network diagram among themes in the field of urban expansion.

Table 2: Critical scores of each theme.

No.	ID	Critical score	No.	ID	Critical score
26	temperature	1.0000	18	coastal urban	0.3262
15	urban agglomeration	0.8994	16	green space	0.2226
11	economic development	0.7654	2	spatial pattern	0.2154
13	housing development policy	0.6528	24	landscape pattern	0.2012
17	surface change	0.6197	3	urban infrastructure	0.1993
9	population density	0.6014	6	agricultural land change	0.1835
8	remote	0.5052	20	cellular automata model	0.1624
4	urban planning management	0.5035	28	ecosystem service	0.1449
10	metropolitan region	0.5025	29	water quality	0.1182
27	urban carbon	0.4330	22	physical health	0.0796
7	urban sprawl	0.4210	5	scenario prediction	0.0386
1	urban expansion	0.4025	14	social cost	0.0000
21	climate change	0.3833	23	forest loss	0.0000
19	transportation emission	0.3778	25	flood risk	0.0000
12	distance variable	0.3565			

of sharing, matching, and learning in urban areas, enterprises and laborers are constantly attracted to these areas. However, the urban space is limited, and the constant gathering of the labor force has increased both housing prices and living costs. People are also forced to settle further away from the city center and pay high commuting costs. When the costs of living and commuting are high enough, these laborers will move elsewhere due to the low net utility

of living in urban areas. In this case, the government invests in the conversion of land into urban transportation infrastructure [50]. By substituting commuting and housing costs [51], the negative impact of rising housing costs is weakened, thereby facilitating a continued urban expansion.

3.3. Identification of Common Themes in the Urban Expansion Literature. Based on the abovementioned thematic distribution, the proportion of each theme in the documents of developed and developing countries after unitization is calculated, and the degree of commonness of these themes is measured using formulas (4) and (5). The above results are then used to plot the degree of commonness of each theme in a graph as shown in Figure 4. The red and blue bars indicate the proportion of relevant documents in developing and developed countries after unitization, respectively, whereas the folding line indicates the degree of commonness of themes. The common themes in the field of urban expansion include themes 16 (green space), 26 (temperature), 4 (urban planning management), 2 (spatial pattern), 18 (coastal urban), and 5 (scenario prediction). With the transformational improvement of research data and technical research methods over the last few years, the available methods for urban expansion research have further expanded to scenario prediction [52], 3S spatial analysis [53], spatial econometrics [54], cellular automata [55], and multiagent simulation [56]. Using these methods to explore the spatial-temporal pattern distribution of urban expansion and effectively describe, simulate, analyze, and predict the process of urban evolution can provide decision-making support for urban planning and management. In addition, urban expansion research in developed and developing countries has mainly focused on coastal urban areas [57, 58] because compared with other cities, coastal cities have unique geographical locations and resource advantages, and urban expansion is highly susceptible to economic development, land-use policies, and regional development policies. Some significant differences in future land-use change are also observed under different development strategies. To expand living and development spaces, coastal areas are reclaiming land from the sea to address the increasingly serious problem of scarcity of land resources [59]. Reclaiming land from the sea is a large-scale human process that greatly disturbs the geographic processes of coastal zones. On the one hand, such land reclamation can increase food supply, attract more investments, and provide a new development space for urban areas. On the other hand, this reclamation can also reduce the service functions of marine ecosystems, destroy the ecological security of bay landscapes, result in marine sedimentation, degrade the quality of marine environments and habitats, and reduce coastal biodiversity [60, 61]. Therefore, how to protect coastal zones during their development has become a research hotspot.

By combining the aforementioned key and commonness indices, a key and commonness bubble for themes in the urban expansion field can be drawn (see Figure 5). This bubble chart is divided into the following quadrants based on the mean values of key and commonness: high degrees of key and commonness (first quadrant), high degree of key and low degree of commonness (second quadrant), low degrees of key and commonness (third quadrant), and low degree of key and high degree of commonness (fourth quadrant). The first quadrant has eight themes, namely, themes 26 (temperature), 15 (urban agglomeration), 11 (economic development), 17 (surface change), 4 (urban planning management), 7 (urban sprawl), 27 (urban carbon), and 19 (transportation emission), of which themes 7 (urban sprawl) and 4 (urban planning management) have more documents than the median. In other words, these themes have received much attention in urban expansion research and are considered key research directions in this field. The connotations of urban sprawl include the following: (1) urban sprawl is a unique way of urban growth that usually occurs when the land development rate exceeds the population growth rate; (2) urban sprawl is characterized by low density, fragmentation, unsustainability, single-form development, excessive reliance on motor vehicles, and massive consumption of agricultural and ecological lands [62]; and (3) urban sprawl has a series of negative effects on traffic flow, plant and animal habitat, the ornamental nature of natural landscapes, and water circulation mechanisms [63, 64]; An in-depth study of urban sprawl has resulted in the formulation of three main theories in the field of urban expansion, namely, compact city theory [65], smart growth theory [66], and new urbanism theory [67]. Urban sprawl control methods can also be divided into two categories. The first category includes the urban planning methods that are implemented by the government and have attributes of administrative orders, such as urban growth boundary, zoning, planned unit development, transfer of development rights, traditional neighborhood development, and transit-oriented development [12, 68, 69]. These measures are based on the best spatial structure and scale of urban areas and directly affect the development decisions of landowners and developers. The second category includes guided regulation measures that are based on market orientation, including land development, fuel, property, and split-rate taxes. These measures do not compulsorily regulate the behavioral choices of people and have indirect control over the urban sprawl. In curbing urban sprawl, the pure market mechanism has a very limited influence on the development of compact cities. Therefore, the government needs not only to formulate various urban sprawl control measures but also to ensure that the relevant policies match the legal and political environment while restraining rapid urban expansion [70, 71].

3.4. Forecast on Thematic Evolution in the Field of Urban Expansion. This paper uses 2020 as the forecast base period and imports the confusion matrix parameters into the HMM module to obtain the hidden Markov forecast results for the evolution of themes in the urban expansion literature from 2020 to 2025 (Figure 6). The proportion of landscape patterns in the prediction results has rapidly increased from 3.09% to 5.14%. The natural landscape is an important environmental resource in the urban ecosystem that has

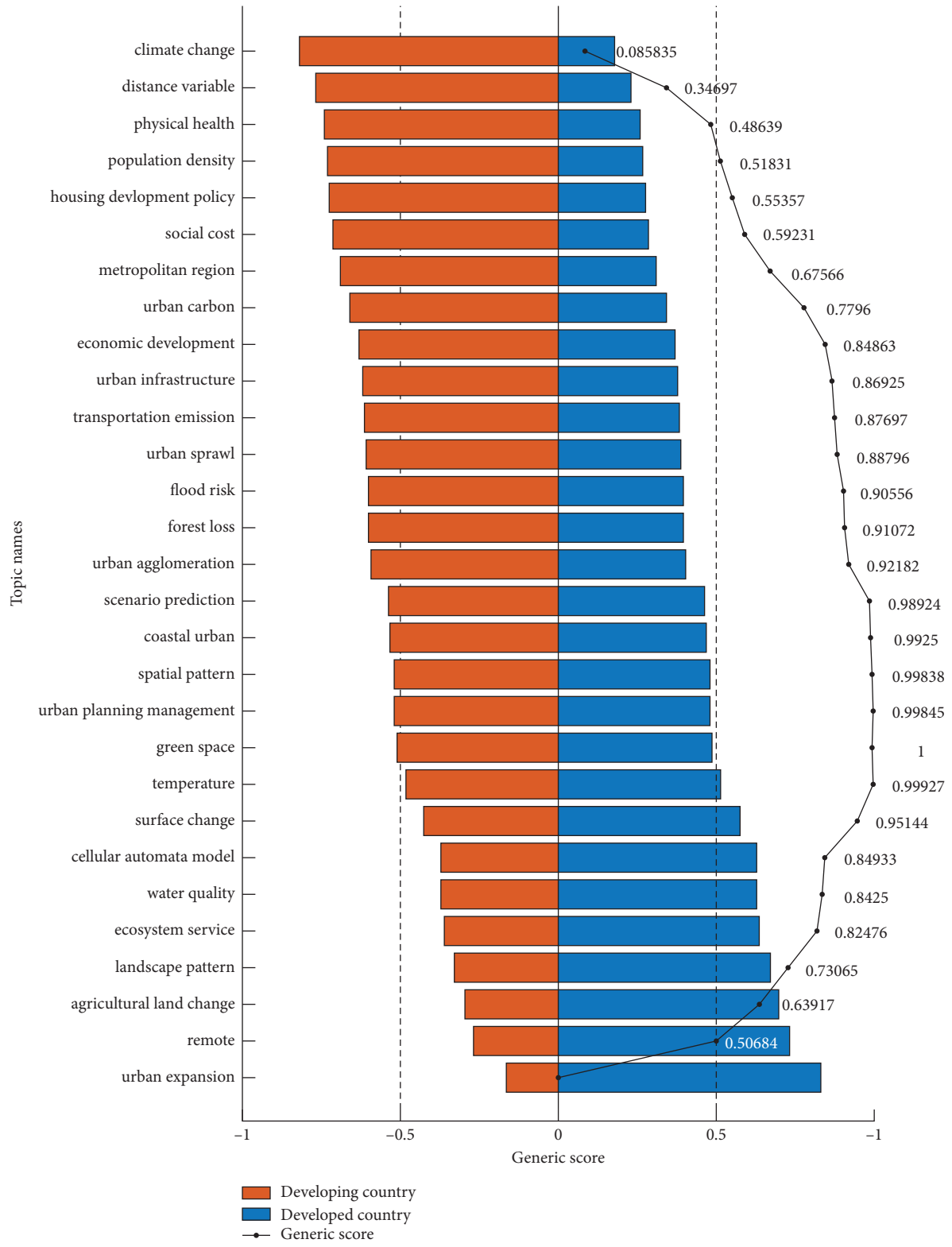


Figure 4: Degree of commonality of each theme.

significant ecological and social functions. Meanwhile, rapid urban expansion is a process in which man-made landscapes gradually erode, occupy, and transform natural landscapes, including forest land, cultivated land, lakes, and grassland, under the influence of human disturbance. Therefore, rapid urban expansion not only reduces the natural landscape area but also results in the fragmentation of natural landscape patterns. A landscape tends to be a complex, heterogeneous,

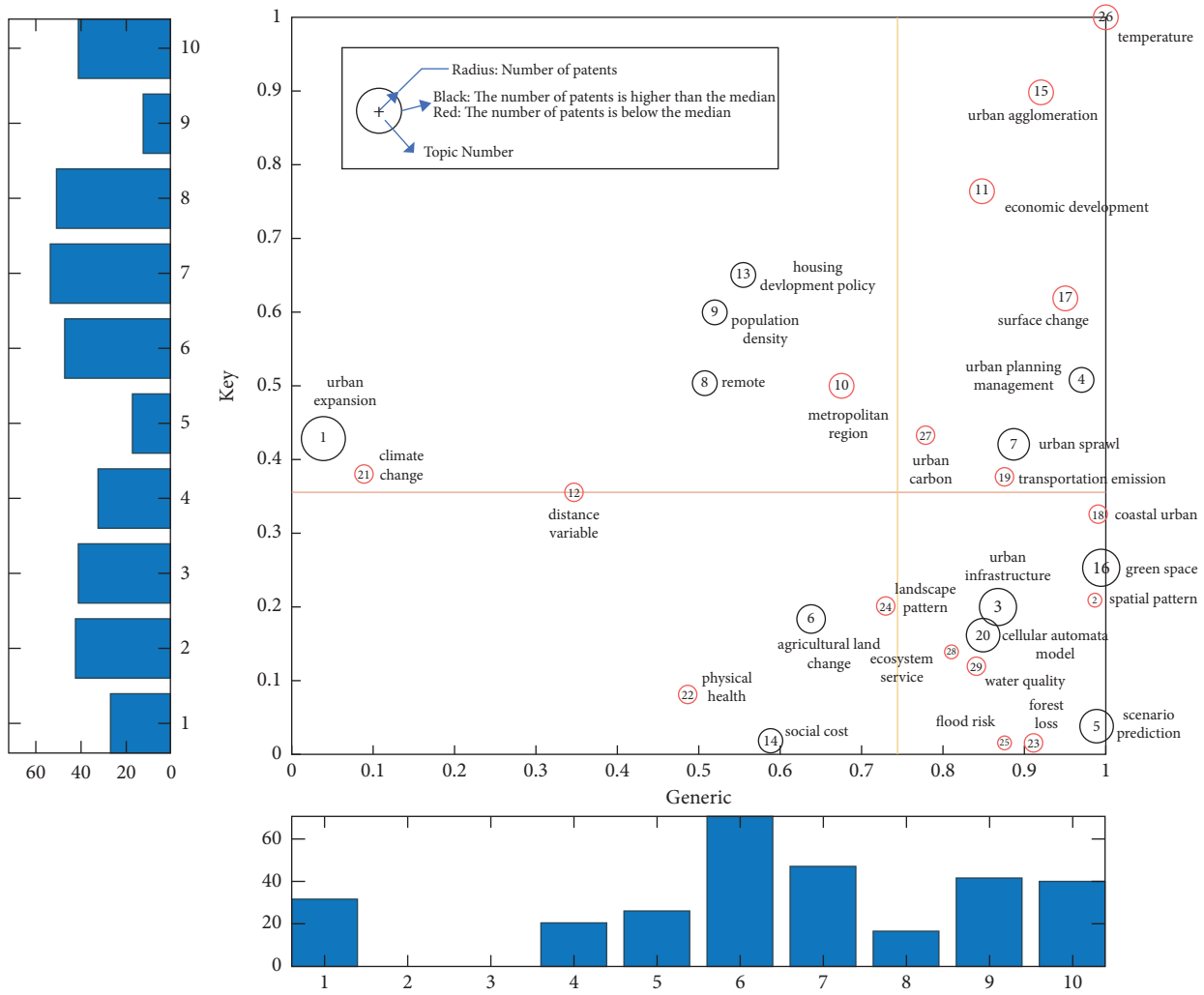


Figure 5: Thematic strategic diagram based on critical score and degree of commonness.

and discontinuous patch mosaic from a single, homogeneous, and continuous relative whole [72, 73]. The fragmentation of the urban landscape not only reduces the quality of the living environment of residents but also seriously endangers the urban ecosystem and urban sustainable development. Therefore, quantitatively identifying the urban landscape based on a remote sensing index (e.g., vegetation, impervious, and water indices) and exploring the responses of natural landscapes to urban expansion have become important ways of understanding the ecological effects of urban landscape evolution [74, 75] and provide valuable references for regional urban planning and ecological construction.

Agriculture land change remains the main direction in urban expansion research. The cultivated land occupied by urban expansion faces not only a decreasing quantity but also changes in its quality. Those areas that surround cities have excellent conditions, topography, water conservancy, and transportation. Urban expansion often encroaches on high-quality cultivated land [76] and affects cultivated land-use intensity in two ways. On the one hand, urban expansion easily results in the scarcity of cultivated land resources. The intensity of cultivated land increases along with the continuous growth of population and demand for food. In addition, the rapid increase in the degree of intensification of agricultural production also promotes the application of chemical fertilizers and pesticides per unit area of cultivated land, thereby bringing agricultural nonpoint source pollution and ecological damage to the environment [77]. On the other hand, the increasingly open labor market promotes the transfer of agricultural labor and consequently reduces or abandons agricultural labor input. After the abandonment of cultivated land, the natural succession of farmland ecosystems destroys species habitats and degrades traditional agricultural landscapes with a high conservation value [78]. In addition, some species that live in the farmland system, especially birds and arthropods, will begin to disappear. The natural succession after abandonment also homogenizes the vegetation on abandoned land, thereby increasing the risk of fire and reducing biodiversity by promoting the growth of pyrophytes [79]. Therefore, studying the evolution of the spatial-temporal pattern of cultivated land occupied by urban expansion can provide technical support and a decision-making basis for handling the relationship between

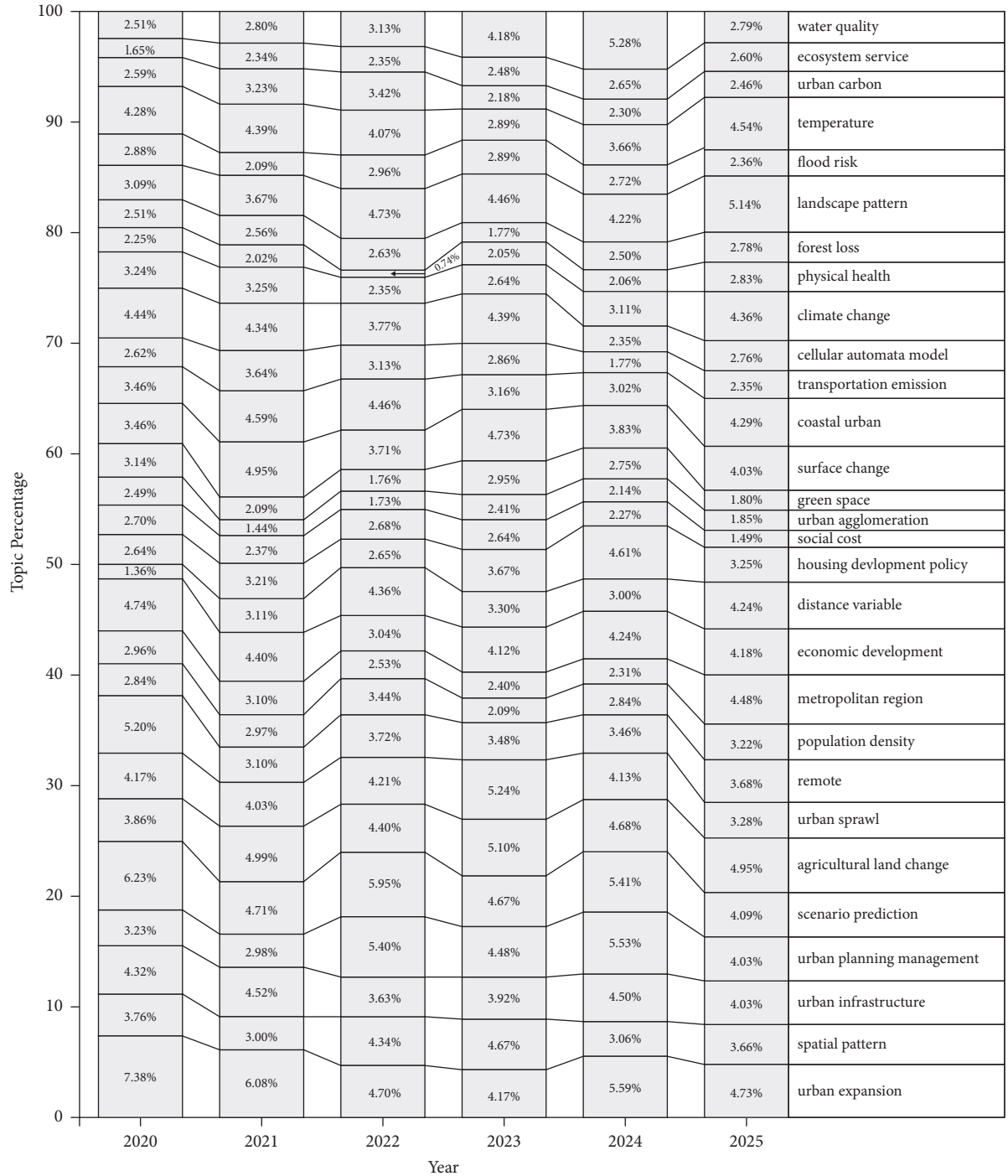


Figure 6: Forecasted thematic evolution trend from 2020 to 2025.

urban expansion and cultivated land protection and for scientifically coordinating urban development. Examining such evolution also has important practical significance in realizing sustainable land use.

4. Conclusion

Our study combines the LDA topic model with HMM to develop a new method for identifying key and common themes from the urban expansion literature. This method overcomes the subjectivity of traditional methods. By applying text mining in a large number of studies in the field of urban expansion, an accurate thematic classification can be achieved, and the identified themes meet the empirical expectations. This study provides theoretical and operational support for identifying key and common themes in the field of bibliometrics.

To study the development trends in the field of urban expansion, this paper divides the scientific literature into 29 themes. By considering both the critical score and degree of

commonness, a total of eight important themes for developed and developing countries are identified, of which six themes (i.e., temperature, urban agglomeration, economic development, surface change, urban carbon, and transportation emission) have documents less than the median number. Future works should focus on these themes in light of the practical problems being faced in the urban expansion field.

The key and common theme identification methods proposed in this paper have good clustering effects, clear thematic boundaries, and accurate recognition results, all of which fully demonstrate their effectiveness and practicability. Future research may consider increasing the scope of the literature collection and including multisource heterogeneous documents to achieve a more comprehensive identification of key and common themes. However, this article also has shortcomings. The data only come from the core database of Web of Science, so the comprehensiveness of the data cannot be guaranteed. This may have a certain impact on the accuracy of the analysis results. Therefore, in future research, various databases should be combined to broaden the data sources in order to more accurately identify the key themes and common themes in the field of urban expansion.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this study.

Authors' Contributions

Yanwei Zhang and Xinhai Lu contributed equally to this work and share the first authorship.

Acknowledgments

The authors would like to thank Jingwen Liao for her contribution to the photo editing in the article. This research was supported by the National Natural Science Foundation of China (no. 71673096), the National 985 Project of Nontraditional Security at Huazhong University of Science and Technology, P.R. China, the Fundamental Research Funds for the Central Universities, HUST (no. 2021WKZDJC001) and the Social Science Foundation of Heilongjiang Province (no. 20JYC156).

References

[1] H. Wang, Q. He, X. Liu, Y. Zhuang, and S. Hong, "Global urbanization research from 1991 to 2009: a systematic research review," *Landscape and Urban Planning*, vol. 104, no. 3-4, pp. 299–309, 2012.

[2] P. Zhao, "Too complex to be managed? New trends in peri-urbanisation and its planning in Beijing," *Cities*, vol. 30, pp. 68–76, 2013.

[3] B. Anasuya, D. Swain, and V. Vinoj, "Rapid urbanization and associated impacts on land surface temperature changes over Bhubaneswar Urban District, India," *Environmental Monitoring and Assessment*, vol. 790, 2019.

[4] Y. Anker, V. Mirlas, A. Gimburg et al., "Effect of rapid urbanization on Mediterranean karstic mountainous drainage basins," *Sustainable Cities and Society*, vol. 51, Article ID 101704, 2019.

[5] S. Angel, J. Parent, D. L. Civco, A. Blei, and D. Potere, "The dimensions of global urban expansion: estimates and projections for all countries," *Progress in Planning*, vol. 75, pp. 53–107, 2011.

[6] S. M. Richter, "Revisiting urban expansion in the continental United States," *Landscape and Urban Planning*, vol. 204, Article ID 103911, 2020.

[7] H. Wang, B. Zhang, Y. Liu et al., "Urban expansion patterns and their driving forces based on the center of gravity-GTWR model: a case study of the Beijing-Tianjin-Hebei urban agglomeration," *Journal of Geographical Sciences*, vol. 30, no. 2, pp. 297–318, 2020.

[8] Y. Tu, B. Chen, and L. Yu, "How does urban expansion interact with cropland loss? a comparison of 14 Chinese cities from 1980 to 2015," *Landscape Ecology*, vol. 36, 2020.

[9] L. Tang, X. Ke, Y. Chen et al., "Which impacts more seriously on natural habitat loss and degradation? Cropland expansion or urban expansion," *Land Degradation and Development*, vol. 32, 2020.

[10] W. H. Lee, "How to identify emerging research fields using scientometrics: an example in the field of Information Security," *Scientometrics*, vol. 76, no. 3, pp. 503–525, 2008.

[11] D. Huang, J. Huang, and T. Liu, "Delimiting urban growth boundaries using the CLUE-S model with village administrative boundaries," *Land Use Policy*, vol. 82, pp. 422–435, 2019.

[12] S. Chakraborti, D. N. Das, B. Mondal, H. Shafizadeh-Moghadam, and Y. Feng, "A neural network and landscape metrics to propose a flexible urban growth boundary: a case study," *Ecological Indicators*, vol. 93, pp. 952–965, 2018.

[13] Y. Feng, J. Wang, X. Tong et al., "Urban expansion simulation and scenario prediction using cellular automata: comparison between individual and multiple influencing factors," *Environmental Monitoring and Assessment*, vol. 191, p. 291, 2019.

[14] A. Colsaet, Y. Laurans, and H. Levrel, "What drives land take and urban land expansion? A systematic review," *Land Use Policy*, vol. 79, pp. 339–349, 2018.

[15] A. Awotwi, G. K. Anornu, J. A. Quaye-Ballard, and T. Annor, "Monitoring land use and land cover changes due to extensive gold mining, urban expansion, and agriculture in the Pra River Basin of Ghana," *Land Degradation & Development*, vol. 29, pp. 3331–3343, 2018.

[16] C. Zeng, M. Zhang, J. Cui, and S. He, "Monitoring and modeling urban expansion-A spatially explicit and multi-scale perspective," *Cities*, vol. 43, pp. 92–103, 2015.

[17] J. Yao, A. T. Murray, J. Wang, and X. Zhang, "Evaluation and development of sustainable urban land use plans through spatial optimization," *Transactions in GIS*, vol. 23, pp. 705–725, 2019.

[18] Z. Zhang, F. Liu, X. Zhao et al., "Urban expansion in China based on remote sensing technology: a review," *Chinese Geographical Science*, vol. 28, no. 5, pp. 727–743, 2018.

[19] A. Abu Hatab, M. E. R. Cavinato, A. Lindemer, and C.-J. Lagerkvist, "Urban sprawl, food security and agricultural systems in developing countries: a systematic review of the literature," *Cities*, vol. 94, pp. 129–142, 2019.

[20] V. Saini and R. K. Tiwari, "A systematic review of urban sprawl studies in India: a geospatial data perspective," *Arabian Journal of Geosciences*, vol. 13, pp. 1–21, 2003.

[21] H. Xie, Y. Zhang, and K. Duan, "Evolutionary overview of urban expansion based on bibliometric analysis in Web of Science from 1990 to 2019," *Habitat International*, vol. 95, Article ID 102100, 2020.

[22] M. Lacey-Barnacle, R. Robison, and C. Foulds, "Energy justice in the developing world: a review of theoretical frameworks, key research themes and policy implications," *Energy for Sustainable Development*, vol. 55, no. 138, 2020.

[23] G. W. Ryan and H. R. Bernard, "Techniques to identify themes," *Field Methods*, vol. 15, no. 1, pp. 85–109, 2003.

[24] D. M. Blei and J. D. Lafferty, "A correlated topic model of science," *Annals of Applied Statistics*, vol. 1, pp. 17–35, 2007.

[25] N. Shibata, Y. Kajikawa, Y. Takeda, I. Sakata, and K. Matsushima, "Detecting emerging research fronts in regenerative medicine by the citation network analysis of scientific publications," *Technological Forecasting and Social Change*, vol. 78, no. 2, pp. 274–282, 2011.

[26] H. Small, "Tracking and predicting growth areas in science," *Scientometrics*, vol. 68, no. 3, pp. 595–610, 2006.

[27] D. Lee and H. Choe, "Estimating the impacts of urban expansion on landscape ecology: forestland perspective in the greater Seoul metropolitan area," *Journal of Urban Planning and Development*, vol. 137, no. 4, pp. 425–437, 2011.

[28] Z. Zhang, W. Sun, and Y. Yu, "Research on the development of marine regional E-commerce based on the analysis of cloud computing and grey prediction method," *Journal of Coastal Research*, vol. 115, pp. 333–337, 2020.

[29] Y. P. Tsang, W. C. Wong, G. Q. Huang, C. H. Wu, Y. H. Kuo, and K. L. Choy, "A fuzzy-based product life cycle prediction for sustainable development in the electric vehicle industry," *Energies*, vol. 13, no. 15, 3918 pages, 2020.

[30] J. Wang, G. Nie, and C. Xue, "Landslide displacement prediction based on time series analysis and data assimilation with hydrological factors," *Arabian Journal of Geosciences*, vol. 13, no. 460, 2020.

[31] F. Almomani, "Prediction the performance of multistage moving bed biological process using artificial neural network (ANN)," *The Science of the Total Environment*, vol. 744, Article ID 140854, 2020.

[32] L. Baum and T. Petrie, "Statistical inference for probabilistic functions of finite state Markov chains," *The Annals of Mathematical Statistics*, vol. 37, pp. 554–1563, 1966.

[33] C. Zeng, Y. Liu, Y. Liu, and L. Qiu, "Urban sprawl and related problems: bibliometric analysis and refined analysis from 1991 to 2011," *Chinese Geographical Science*, vol. 24, no. 2, pp. 245–257, 2014.

[34] D. Blei, A. Ng, and M. Jordan, "Latent Dirichlet allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 2003.

[35] H. Jelodar, Y. Wang, C. Yuan et al., "Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey," *Multimedia Tools and Applications*, vol. 78, no. 11, pp. 15169–15211, 2018.

[36] R. Welch, "Hidden Markov models and the Baum–Welch algorithm," *IEEE Information Theory Society Newsletter*, vol. 53, pp. 194–211, 2003.

[37] K. M. Frahm and D. L. Shepelyansky, "Ising-PageRank model of opinion formation on social networks," *Physica A: Statistical Mechanics and Its Applications*, vol. 526, Article ID 121069, 2019.

[38] F. A. Massucci and D. Docampo, "Measuring the academic reputation through citation networks via PageRank," *Journal of Informetrics*, vol. 13, no. 1, pp. 185–201, 2019.

[39] P. Holme and J. Saramäki, "Temporal networks," *Physics Reports*, vol. 519, no. 3, pp. 97–125, 2012.

[40] G. Abramo, C. A. D'Angelo, and A. Soldatenkova, "An investigation on the skewness patterns and fractal nature of research productivity distributions at field and discipline level," *Journal of Informetrics*, vol. 11, no. 1, pp. 324–335, 2017.

[41] L. Bornmann and L. Leydesdorff, "Skewness of citation impact data and covariates of citation distributions: a large-scale empirical analysis based on Web of Science data," *Journal of Informetrics*, vol. 11, no. 1, pp. 164–175, 2017.

[42] J. Ruiz-Castillo and R. Costas, "The skewness of scientific productivity," *Journal of Informetrics*, vol. 8, no. 4, pp. 917–934, 2014.

[43] D. M. Blei, "Probabilistic topic models," *Communications of the ACM*, vol. 55, no. 4, pp. 77–84, 2012.

[44] I. Buo, V. Sagris, I. Burdun, and E. Uuemaa, "Estimating the expansion of urban areas and urban heat islands (UHI) in Ghana: a case study," *Natural Hazards*, vol. 105, pp. 1299–1321, 2021.

[45] J. Du, X. Xiang, B. Zhao, and H. Zhou, "Impact of urban expansion on land surface temperature in Fuzhou, China using Landsat imagery," *Sustainable cities and society*, vol. 61, 102346 pages, 2020.

[46] W. Admasu, S. Van Passel, A. Minale, E. Tsegaye, H. Azadi, and J. Nyssen, "Take out the farmer: an economic assessment of land expropriation for urban expansion in Bahir Dar, Northwest Ethiopia," *Land Use Policy*, vol. 87, Article ID 104038, 2019.

[47] M. Jia, Y. Liu, S. Lieske, and T. Chen, "Public policy change and its impact on urban expansion: an evaluation of 265 cities in China," *Land Use Policy*, vol. 97, Article ID 104754, 2020.

[48] Y. Zhang and H. Xie, "Interactive relationship among urban expansion, economic development, and population growth since the reform and opening up in China: an analysis based on a vector error correction model," *Land*, vol. 8, 153 pages, 2019.

[49] J. Marshall, "Urban land area and population growth: a new scaling relationship for metropolitan expansion," *Urban Studies*, vol. 44, 2007.

[50] P. Zhao, B. Lü, and G. de Roo, "Urban expansion and transportation: the impact of urban form on commuting patterns on the city fringe of Beijing," *Environment and Planning: Economy and Space*, vol. 42, no. 10, pp. 2467–2486, 2010.

[51] R. J. Arnott and J. E. Stiglitz, "Aggregate land rents, expenditure on public goods, and optimal city size," *Quarterly Journal of Economics*, vol. 93, no. 4, pp. 471–500, 1979.

[52] Z. Lei, Y. Feng, X. Tong, S. Liu, C. Gao, and S. Chen, "A spatial error-based cellular automata approach to reproducing and projecting dynamic urban expansion," *Geocarto International*, vol. 16, pp. 1–21, 2020.

[53] H. Lamphar, "Spatio-temporal association of light pollution and urban sprawl using remote sensing imagery and GIS: a simple method based in Otsu's algorithm," *Journal of Quantitative Spectroscopy and Radiative Transfer*, vol. 253, Article ID 107068, 2020.

[54] S. Mathur, "Impact of an urban growth boundary across the entire house price spectrum: the two-stage quantile spatial regression approach," *Land Use Policy*, vol. 80, pp. 88–94, 2019.

[55] S. Chen, Y. Feng, Z. Ye et al., "A cellular automata approach of urban sprawl simulation with Bayesian spatially-varying transformation rules," *GIScience & Remote Sensing*, vol. 57, 2020.

[56] A. Mustafa, M. Cools, I. Saadi, and J. Teller, "Coupling agent-based, cellular automata and logistic regression into a hybrid urban expansion model (HUEM)," *Land Use Policy*, vol. 69, pp. 529–540, 2017.

[57] Y. Deng, W. Qi, Bo. Fu, and K. Wang, ""Geographical transformations of urban sprawl: exploring the spatial heterogeneity across cities in China 1992–2015," *Cities*, vol. 105, Article ID 102415, 2019.

[58] Y. Kim and G. Newman, "Climate change preparedness: comparing future urban growth and flood risk in Amsterdam and Houston," *Sustainability*, vol. 11, 1048 pages, 2019.

[59] Y. Li, Y. Shi, X. Zhu, H. Cao, and T. Yu, "Coastal wetland loss and environmental change due to rapid urban expansion in Lianyungang, Jiangsu, China," *Regional Environmental Change*, vol. 14, no. 3, pp. 1175–1188, 2014.

[60] D. S. Hammond, V. Gond, C. Baider, F. B. V. Florens, S. Persand, and S. G. W. Laurance, "Threats to environmentally sensitive areas from peri-urban expansion in Mauritius," *Environmental Conservation*, vol. 42, no. 3, pp. 256–267, 2015.

[61] T. M. Coxon, B. K. Odhiambo, and L. C. Giancarlo, "The impact of urban expansion and agricultural legacies on trace metal accumulation in fluvial and lacustrine sediments of the lower Chesapeake Bay basin, USA," *The Science of the Total Environment*, vol. 568, pp. 402–414, 2016.

[62] F. Bidandi and J. Williams, "Understanding urban land, politics, and planning: a critical appraisal of Kampala's urban sprawl," *Cities*, vol. 106, Article ID 102858, 2020.

[63] S. Hamidi and R. Ewing, "Is sprawl affordable for Americans?" *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2500, no. 1, pp. 75–79, 2015.

[64] S. M. Mccoshum and M. A. Geber, "Land conversion for solar facilities and urban sprawl in southwest deserts causes different amounts of habitat loss for Ashmeadiella bees," *Journal of the Kansas Entomological Society*, vol. 92, no. 2, pp. 468–478, 2020.

[65] M. Jun, "The effects of polycentric evolution on commute times in a polycentric compact city: a case of the Seoul metropolitan area," *Cities*, vol. 98, Article ID 102587, 2020.

[66] Å. Gren, J. Colding, M. Berghauser-Pont, and L. Marcus, "How smart is smart growth? Examining the environmental validation behind city compaction," *Ambio*, vol. 48, no. 6, pp. 580–589, 2018.

[67] A. Stanislav and J. Chin, "Evaluating livability and perceived values of sustainable neighborhood design: new Urbanism and original urban suburbs," *Sustainable Cities and Society*, vol. 47, Article ID 101517, 2019.

[68] A. H. Whittemore, "The new communalism," *Journal of Planning History*, vol. 14, no. 3, pp. 244–259, 2015.

[69] W. Lang, E. Hui, T. Chen, and X. Li, "Understanding livable dense urban form for social activities in transit-oriented development through human-scale measurements," *Habitat International*, vol. 104, Article ID 102238, 2020.

[70] B. Fernandez Milan and F. Creutzig, "Municipal policies accelerated urban sprawl and public debts in Spain," *Land Use Policy*, vol. 54, pp. 103–115, 2016.

[71] D. Coq-Huelva and R. Asián-Chaves, "Urban Sprawl and Sustainable Urban Policies. A review of the cases of Lima, Mexico City and Santiago de Chile," *Sustainability*, vol. 11, 5835 pages, 2019.

[72] P. Hien, N. Men, P. Tan, and M. Hangartner, "Impact of urban expansion on the air pollution landscape: A case study of Hanoi, Vietnam," *Science of The Total Environment*, vol. 702, Article ID 134635, 2020.

[73] B. Rimal, H. Keshtkar, R. Sharma, N. Stork, S. Rijal, and R. Kunwar, "Simulating urban expansion in a rapidly changing landscape in eastern Tarai, Nepal," *Environmental Monitoring and Assessment*, vol. 191, p. 255, 2019.

[74] Y. Al-husban, "Urban expansion and shrinkage of vegetation cover in Al-Balqa Governorate, the Hashemite Kingdom of Jordan," *Environmental Earth Sciences*, vol. 78, 620 pages, 2019.

[75] S. Bonilla-Bedoya, A. Mora, A. Vaca, A. Estrella, and M. Herrera, "Modelling the relationship between urban expansion processes and urban forest characteristics: an application to the Metropolitan District of Quito," *Computers, Environment and Urban Systems*, vol. 79, Article ID 101420, 2019.

[76] B. Güneralp, M. Reba, B. U. Hales, E. A. Wentz, and K. C. Seto, "Trends in urban land expansion, density, and land transitions from 1970 to 2010: a global synthesis," *Environmental Research Letters*, vol. 15, no. 4, Article ID 044015, 2020.

[77] H. Xie, Y. Zhang, and Y. Choi, "Measuring the cultivated land use efficiency of the main grain-producing areas in china under the constraints of carbon emissions and agricultural nonpoint source pollution," *Sustainability*, vol. 10, Article ID 1932, 2018.

[78] J. Fischer, T. Hartel, and T. Kuemmerle, "Conservation policy in traditional farming landscapes," *Conservation Letters*, vol. 5, no. 3, pp. 167–175, 2012.

[79] C. Vega-Garcia and E. Chuvieco, "Applying local measures of spatial heterogeneity to Landsat-TM images for predicting wildfire occurrence in Mediterranean landscapes," *Landscape Ecology*, vol. 21, pp. 595–605, 2006.