

Hanshan Huang¹Jianbo Zhang¹Liu Gan¹Ning Ning^{1*}

Analysis of Urban Population Flow and Spatial Distribution Patterns: A Study Based on Cluster Analysis Algorithm



Abstract: - The research offers insightful information about how China's industrial structure and socioeconomic variables affect population flows between cities. It is not without restrictions, though. These include the possibility of biases and incompleteness in the data utilized, difficulties in understanding the intricate models that are used, and the restricted applicability of the results outside of China. In this research Analysis of urban population flow and spatial distribution patterns: a study based on cluster analysis algorithm (UPF-SDP-CAA) is suggested. At first, the Landsat satellite information is used to gather the input data. The image is provided to preprocessing phase. During the pre-processing phase, the NIOF is used to remove noise for image. Then the preprocessed data are fed to feature extraction phase. Self-Supervised Nonlinear Transform (SSNT) is used to choose urban characteristics like Ward boundaries, geographic features, population density, socioeconomic characteristics, and environmental features. Extracted features were sent to the Progressive Graph Convolutional Network (PGCN) and the urban population flow and spatial distribution prediction used to classify like Low density and fragmented built-up land, High density built-up land greenery, Water bodies, crop land, and grassland wetlands, bare territory. In order to accurately classify, the PGCN classifier is optimized using the HSWOA. The proposed technique is implemented to Python. The effectiveness of the suggested UPF-SDP-CAA approach is evaluated using a number of performance criteria, including Accuracy, Recall, Precision, RMSE and AUC. The proposed UPF-SDP-CAA method covers 28.36%, 23.42% and 33.27% higher precision, 17.42%, 25.36% and 17.27% higher recall and of 19.36%, 26.42% and 23.27% higher accuracy compared with existing Simulating inter-city population flows based on graph neural networks (SIC-PF-GNN), An integrated simulation approach to the assessment of urban growth pattern and loss in UGS in Kolkata, India: A GIS-based analysis (UGP-UGS-ANN) and The effects of sample size and sample prevalence on cellular automata simulation of urban growth (CAS-UG-SVM) respectively.

Keywords: Harbor Seal Whiskers Optimization Algorithm, Landsat satellite dataset, Non-Integer Order Generalized Filters, Python, Progressive Graph Convolutional Network, Population flow, Self-Supervised Nonlinear Transform.

I. INTRODUCTION

The movement of people into, out of, and within urban regions is referred to as urban population flow. It includes a range of variables, including immigration from foreign nations, internal mobility within cities, and migration from rural to urban areas [1]. Since urban population flow affects housing, social services, infrastructure, and urban growth, it is essential knowledge for researchers, politicians, and urban planners [2]. Urban population flow is influenced by a number of important elements, including government regulations, social networks, housing availability and cost, infrastructure and amenities, economic opportunities, natural disasters, and climate change. Policymakers and urban planners can better anticipate future population trends, plan infrastructure investments, handle housing demands, and devise methods to support sustainable urban growth by analysing patterns of urban population flow [3-5]. Additionally, making sure that every citizen has fair access to resources and services requires a grasp of the demographic makeup of metropolitan regions. While linear distribution happens along lines like rivers or highways, clustered distribution is caused by human preferences or environmental variation [6, 7]. Radial dispersion is the outer extension of a central source. Comprehending these patterns facilitates resource allocation, conservation initiatives, and regional development strategies, and helps decision-makers in domains such as ecology, economics, and urban planning. Developing models to predict the migration of individuals across cities over time is necessary to simulate inter-city population fluxes [8-10]. These models replicate how variables like housing availability, employment possibilities, and infrastructure affect population movements. They do this by utilizing data on demographics, economics, and transportation [11, 12]. Data gathering, model building, parameter estimation, scenario planning, validation, and policy analysis are important processes. Urban planners and policymakers can use these simulations to forecast future population patterns, assess the effects of various policies, and make well-informed

¹ ¹ Guangxi Natural Resources Vocational And Technical College, Fusui, Guangxi 532199, China

*Corresponding author e-mail: 13977794924@163.com

Copyright © JES 2024 on-line : journal.esrgroups.org

decisions about infrastructure and regional development [13-15]. The intervening opportunities model was an additional early model used to examine population migration, in addition to the gravity model. According to this concept, the overall amount of options available at each location and the accessibility of intermediate destinations determine how many people go between two areas [16, 17].

It suggests that a destination's appeal is inversely correlated with its distance from other locations and directly correlated with its opportunities. There is scientific analogy to this idea of intervening opportunities, which led to the creation of the radiation model [18]. According to this model, visitors choose a destination in two steps: first, the attractiveness of each destination is distributed according to a certain set of parameters, and then, based on their distance from the origin, they select the destination that is the closest and has enough attraction [19, 20]. While there is potential for using GNNs to predict population movements between cities, there are a number of obstacles to overcome, including the need for high-quality data, the inability to capture dynamics across spatial scales, limited interpretability even with explainability tools, and high processing demands. Similar challenges, such as over fitting endangering dependability and complexity impeding result interpretation, face Artificial Neural Networks (ANNs) in spatial distribution and urban population flow research. High computational complexity, interpretability limitations, and difficulty capturing temporal dynamics in urban population flow problems are some of the other obstacles that Support Vector Machines (SVMs) confront. In order to overcome these constraints, specialized methods such as pre-processing, model selection, and customized feature engineering are needed. Additionally, alternative or complementary approaches to assessing urban dynamics may need to be investigated.

The aim of this project was to urban population flow and spatial distribution, many models out now that successfully integrate various network data types for thorough urban population flow and spatial distribution prediction. The new model should show better results by utilizing data from multiple sources at the same time. This model based on the Landsat satellite dataset with input data and PGCN.

Important contribution of this paper as follows;

- In this research work, Analysis of urban population flow and spatial distribution patterns: a study based on cluster analysis algorithm (UPF-SDP-CAA) is proposed.
- By utilizing PGCN techniques to conduct comprehensive experiments on the Landsat satellite dataset, the work on urban population flow and spatial distribution on the dataset extended.
- Evaluates the variation in green space availability and per-capita allocation within cities from a spatiotemporal approach.
- The obtained results of proposed UPF-SDP-CAA algorithm is comparing to the existing models such as SIC-PF-GNN, UGP-UGS-ANN and CAS-UG-SVM methods respectively.

The remaining manuscript is organized as follows: Part 2 outlines the Literature Survey, Part 3 Displays the suggested method, Part 4 presents the results with conversations, and Part 5 concludes the manuscript.

II. LITERATURE SURVEY

A literature presents a number of research projects on deep learning-based Classification of urban population flow and spatial distribution; this section evaluated some of the most recent studies.

Lou and Chen, [21] has presented SIC-PF-GNN. In order to better understand population dynamics and their complex relationships to industrial structure, this paper uses the technique to model China's intercity population flows. This method effectively includes Ten cent migration big data with demographic and socioeconomic data, taking into account the spatial relationships between cities. According to the findings, the model's forecast accuracy for air travel was only moderate, while for road and rail traffic it was high using the CPC index. The model's ability to forecast the regional aggregation of flows and the urban hierarchy was confirmed through comparison with empirical data. Utilizing Explainer, the findings showed that as manufacturing advanced, population mobility for air travel increased while it dropped for road and rail traffic, population size positively influenced population flow. Through scenario modeling in Northeast China, were discovered that improving the business and consumer service sectors in the area could lessen the negative effects of population exodus. It provides high accuracy, and it provides low precision.

Dinda, et al. [22] has suggested an integrated simulation method for evaluating Kolkata, India's urban growth pattern and shrinkage of its urban green space: a GIS-based study. The present investigation makes use of a multi temporal land utilization transition scenario to analyze the loss of urban green space (UGS) in relative to

LULC alteration and to forecast future UGS loss in Kolkata. Data from Landsat satellites were analyzed. Three comparative land categorization techniques have been established in order to evaluate the trend of urban expansion. A combined Markov chain and cellular automata model was used to represent the status of UGS in 2025 and 2035, as well as the dynamics of LULC. The land use dynamic degree index, UGS change intensity index, UGS land index, rate and speed of urban expansion and the ratio of change in green space and UGS coverage with respect to the total area of all land were all measured. It provides high Sensitivity and it provides low accuracy.

Zhang and C Xia, [23] has presented the implications of sample prevalence and size on urban development modeling using cellular automata. In general, the CA example based on the approaches produce consistent results. Because of the high level of uncertainty, it is best to scrap the sampling strategy with a small sample size and low sample prevalence. While sample prevalence influences a CA examples performance when there were enough samples, sample size establishes CA examples robustness. Specifically, the complexity and fragmentation of the simulated urban layouts increase with simulation precision. decrease when the sample prevalence approaches the population prevalence. We suggest that the optimal sampling strategy would have a sample rate of sample prevalence that is equal to the population prevalence. Populations that a various research areas represent do not influence the choice of the best sampling plan. It provides high recall, and it provides low Sensitivity.

Wang, et al. [24] has proposed Studying the network structures and variables that affect migration and population movement in China's Yangtze River Delta metropolitan area. This examines the features of urban networks' spatial organization and the factors that influence them from the standpoint of migration and interprovincial population flow. The findings indicate that Shanghai, being a major metropolis, has a significant siphon effect, resulting in an influx of people into Suzhou, Nanjing, Hangzhou, Ningbo, Wuxi, and Changzhou. The migration and population mobility network of the Yangtze River Delta urban agglomeration blatantly displays hierarchical tendencies. During the Spring Festival travel rush, the primary migration path was the secondary network relationship, whereas the primary flow path was the first-level network relationship during the regular period. Three indicators show a persistent strong link state between the Yangtze River Delta urban agglomeration network and the establishment of a local cluster structure: the network density, mean centrality, and control force based on population movement and migration. In terms of migration and population movement, these indicators also emphasize the dual characteristics of city tightness, which were linked to "overcoming the friction of space" and "the restriction of the geographic space effect." Both great accuracy and low recall are offered by it.

Yang and Wong, [25] have presented Spatial transfer of tourist flows to China's cities. The present research makes an effort to look at the spatial distribution of domestic and inbound visitor flows to Chinese cities, as well as their growth rates, using exploratory spatial data analysis. This approach was a collection of GIS spatial statistical tools that can be used to identify hotspot patterns, describe and visualize the spatial distribution, and propose spatial regimes. There was a strong positive and considerable spatial autocorrelation, according to the worldwide Moran's I data for inbound and inland tourist flows. Additionally, four large inbound tourist hotspot locations, as well as five significant domestic tourism hotspot sites identified by the Moran importance maps. A number of conclusions were made in light of the findings, including the need to prioritize allocating resources to hotspot areas and to take advantage of their spill over effects. It provides high accuracy, and it provides high RMSE.

Zhen, et al. [26] has suggested analysing urban development patterns have to be examined using the flow analysis approach as a basis. taking into account the network structure between cities and how each city's function and role within the network were formed. Urban network and space of flow theories serve as the foundation for the flow analysis of different elements across cities. China's Hebei Province served as the study's example. Through the simulation of financial, informational, traffic, and economic flows between cities, every city in the regional urban network was assessed in terms of its general condition. A development pattern, having a network topology that has several cores and nodes, can be categorized into three tiers, according to the results. Furthermore, the urban developmental level can be predicted using the flow analysis method. Additionally, a few policy suggestions were made to help Hubei's regional integrated growth. It provides high precision, and it provides low Sensitivity.

Peng, et al. [27] have presented the distribution of land use and urban population density in Guangzhou, China: A spatial spill over viewpoint. The current study examined the spatial evolution and temporal distribution features of urban population density using methods from spatial correlation analysis and the spatial Durbin model. It next examined the spatial spill over impact of land use on urban population density. The findings demonstrate that the circulation of urban population thickness was generally defined by aggregation and differs between weekdays and weekends. The population density fluctuations over the course of a day exhibit a "rapid growth-gentle decline-rapid growth-rapid decline" pattern. Additionally, there was spatial spill over effects of land use that were just as significant in the distribution of density of the population as the immediate impacts. Furthermore, at dissimilar times, different land utilize forms have varied direct effects and spatial spill over effects. These results imply that the indirect effects of surrounding areas ought to be considered when distributing population density in a balanced manner. The results of this study should have consequences for legislators and urban designers. regarding the optimization of urban spatial structure and the wise use of public resources. It provides high Sensitivity and it provides high RMSE.

III. PROPOSED METHODOLOGY

In this section, Analysis of urban population flow and spatial distribution patterns: a study based on cluster analysis algorithm (UPF-SDP-CAA) is proposed. There are five phases in this process: capturing images, pre-processing, extracting features, categorization, and optimization. In the suggested were collected and pre-processing to prepare them for further analysis. Following Extraction characteristics including the ward line, geographic features, population density, socioeconomic characteristics, and environmental characteristics are extracted. These features are then organized into a feature vector. The final step involves employing a PGCN for classification, The HSWOA method is introduced for training the PGCN. The block diagram of suggested UPF-SDP-CAA method is represented in Fig 1. Accordingly, detailed description of all step given as below,

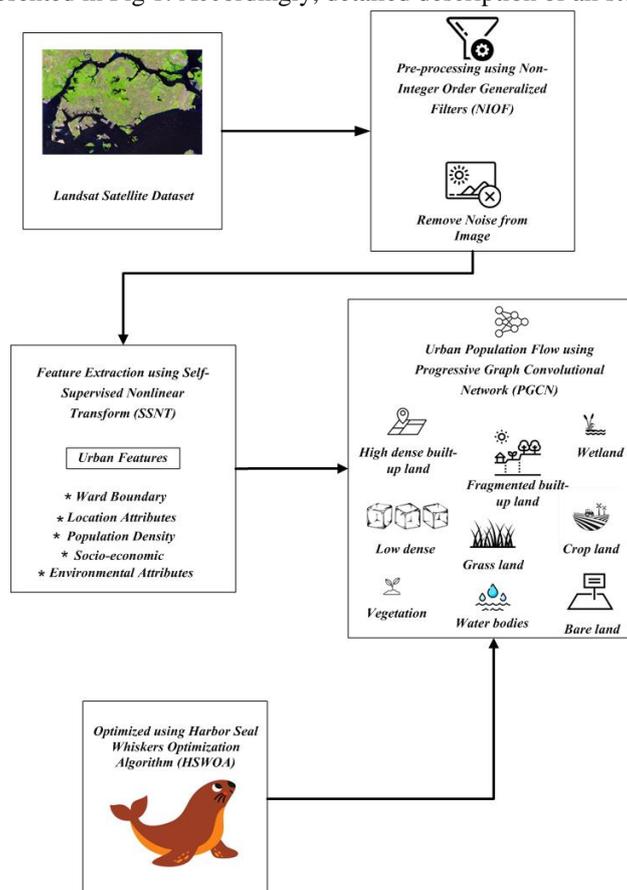


Figure 1: Block Diagram of UPF-SDP-CAA methods

A. Image Acquisition

To begin with, the input images are gathered from Landsat satellite dataset [28], was detecting the urban population flow and spatial distribution Classification. The satellite datasets are from Landsat-3 Multi-Spectral

Scanner (MSS) for 1980, Landsat-5 Thematic Mapper (TM) for 1991, Landsat-7 Enhanced TM Plus (ETM+; path/row = 138/44) for 2001 and 2011, and Landsat-8 Working Land picture (OLI; path/row = 138/44) for 2018.

B. Pre-processing using Non-Integer Order Generalized Filters (NIOF)

Pre-Processing using NIOF [29] is discussed. A NIOF method has used to remove noise for image. Due to NIOF's ability to adapt, predictions and analyses are more accurate since it can better handle the complexities and anomalies seen in urban data. NIOF helps researchers, policymakers, and urban planners make better decisions by providing enhanced insights into population migrations and spatial distributions by capturing the intricacies of real-world urban circumstances. Their usefulness is further increased by their capacity to reveal hidden dynamics and interact with cutting-edge methods like machine learning and spatial analysis. Furthermore, NIOF's resistance to noise and scalability guarantee their applicability in a variety of urban planning Refine urban population flow analysis by improving spatial resolution, forecasting trends, and optimizing service distribution by utilizing NIOF. Find anomalies, analyze the impact on the environment, determine equity, and guide transportation planning. Encourage interdisciplinary cooperation to achieve thorough comprehension and creative solutions in urban dynamics and research situations, ranging from neighbourhood-level studies to city-wide evaluations in equation (1),

$$I_{MQ}(t) = H_0 \left[\frac{1}{(\pi)^\alpha + 1} \right]^\gamma \tag{1}$$

Where, I_{MQ} represents the transfer function, H_0 is a constant, π represent the time constant; α represent the integer order filter. Noise reduction features are available in many photo editing software packages, and they can efficiently reduce noise while maintaining image detail. Thus it is given equation (2).

$$I'_{IQ}(t) = H_0 \frac{[(\tau t)^\alpha + 1]^\gamma}{(\tau t)^\alpha} \tag{2}$$

Where, I'_{IQ} is represents the transfer function of the related inverse filter. τt is represent the value of the slope transition, This can assist in reducing noise and enhancing the clarity of the image as a whole. Thus, it is given equation (3)

$$I_{IQ.QM}(t) = H_0 \left(\frac{\tau t}{\pi + 1} \right)^\gamma \tag{3}$$

Where, I is represent the power-law inverse transfer function; γ has represent the phase angle range. The slope of the transition among the filter's two typical bands and the range that defines its phase response are also, the image by this equation (4).

$$|I_{IQ.QM}(\omega)| = H_0 \frac{(\omega \tau)^\gamma}{[1 + (\omega \tau)^2]^{1/2}} \tag{4}$$

Where, H_0 is represent the maximum gain and factorial-order filter; $\omega \tau$ is represent the input image; Pre-processing the satellite photos is crucial in order to remove aberrations like sensor noise is in equation (5),

$$I_{CQ.QM}(t) = H_0 \left[\frac{(\pi)^\beta}{\pi + 1} \right]^\gamma \tag{5}$$

Where, $(\tau t)^\beta$ is represent the located at a relatively; $(\tau t)^\alpha$ is represent the change from the stop-band to the pass-band of the filter. Finally, the NIOF has to remove noise for image, and then Pre-processed information is given to the feature extraction for removing the features.

C. Feature Extraction using Self-Supervised Nonlinear Transform (SSNT)

In this section, Self-Supervised Nonlinear Transform (SSNT) [30] is discussed. This SSNT method is extracting the features like ward boundary, location attributes, population thickness, socio-economic and environmental qualities. It reduces the dimensionality of the data without sacrificing important information by automatically learning pertinent features. It guarantees trustworthy insights into population dynamics and is robust against

noise. It allows for the real-time examination of urban trends and is scalable to big datasets. Cross-domain transfer of SSNT-learned representations is beneficial for public health, transportation, and urban planning. It provides interpretable insights through visualization even though it is nonlinear. Through constant adaptation to shifting urban dynamics, its relevance endures over time. To sum up, SSNT offers a strong, scalable, and flexible framework for comprehending and forecasting patterns of urban population flow and geographical distribution. The following are the objectives for SSNT in urban population flow: extract relevant features, minimize dimensionality, group areas, forecast patterns, identify anomalies, integrate with GIS, facilitate transfer learning, guarantee interpretability, maximize scalability, and specify assessment criteria. By offering insights into intricate population dynamics and spatial distribution, you can aid in urban planning ward boundary is expressed as the following equation (6)

$$\omega_j(y) = \sigma(y \times_3 X_j) \tag{6}$$

Where, $\omega_j(y)$ denotes the production of the layer, σ denotes a nonlinear activation purpose, X_j denotes the learnable matrix and y denotes the number of layers. Ward boundaries are drawn to guarantee that every area under a jurisdiction is fairly represented and that local government is effective and responsive to community concerns. The population density is defined as equation (7),

$$g(y) = \omega_q \circ \omega_{q-1} \circ \dots \circ \omega_1(y_0) \tag{7}$$

Where, $g(y)$ denotes notation for a function; y_0 denotes the initialization function; \circ denotes the composition of functions; ω_q denotes the number of layers in f ; $\omega_1(y_0)$ denotes the output of the layer. Historical and cultural significance includes cultural sites, landmarks, and heritage conservation initiatives. The location attributes is defined as equation (8),

$$h(y) = \omega_{q+r} \circ \omega_{q+r-1} \circ \dots \circ \omega_{q+1}(y) \tag{8}$$

Where, $h(y)$ denotes a linear transform function; r denotes the number of layers in g ; q and q denotes a multilayer transform function; $\omega_{q+1}(y)$ number of layers in X . It's a crucial demographic indicator for analysing community dynamics, allocating resources, and socio-economic defined as equation (9),

$$L = L_1 + L_2 \tag{9}$$

Where, L denotes value of the transform function; L_1 and L_2 denotes layers of the linear function. Environmental characteristics provide information on the ecological health and sustainability of an environmental attributes is defined as equation (10),

$$\frac{\partial \|(g(o))^{(l)}\|_*}{\partial (g(o))^{(l)}} \ni \tilde{V}_l \tilde{W}_l^U \tag{10}$$

Where, $\partial (g(o))^{(l)}$ denotes the matrix singular value decomposition; \tilde{V}_l and \tilde{W}_l^U denotes the number of non-zero elements. These characteristics include biodiversity, climate patterns, air and water quality, and natural resources. Finally, essential features like such ward boundary, site attributes, population thickness, socio-economic and environmental attributes are extracted, after completing feature extraction the extracted features are fed to PGCN.

D. Urban Population Flow using Progressive Graph Convolutional Network (PGCN)

Urban PGCN [28] is discussed. PGCN is proposed for Urban Population Flow and it classifies like Low density and fragmented built-up land, High density built-up land greenery, Water bodies, crop land, and grassland wetlands, bare territory. They provide transferability between comparable metropolitan areas, scale well to big datasets, and effectively represent spatial interdependence. Urban planning and policy decisions are aided by the interpretable insights that PGCNs offer into the elements influencing population dynamics. Through the addition of past information and the consideration of geographical fundamentals, they provide a thorough understanding of urban systems, thereby advancing sustainable development. All things considered, PGCNs provide a strong, effective, and understandable framework for scalable and flexible analysis of intricate urban phenomena. Urban population analysis using PGCN aims to model population dynamics, anticipate spatial distribution, evaluate impacts, identify communities, find anomalies, integrate data sources, guarantee scalability, and improve

interpretability. Through an effective understanding and visualization of patterns of population flow and distribution, these goals support well-informed decisions about urban planning and policy it is given in equation (11)

$$t_{jk}^u = \tilde{y}_u^{j(U)^T} \cdot \tilde{y}_u^{k(U)} \tag{11}$$

Where, t_{ab} denotes the graphs between the tasks of all MEC nodes, s is a unit course, $\tilde{x}_s^{b(T)}$ is a minimized data of each nodes b at time T and $\tilde{y}_s^{a(T)}$ is the similarities learn randomly created a learnable a at adjuster, They are typically found in urban areas when there is a high demand for space and a scarcity of land is given in equation (12)

$$B_{Q_{jk}}^t = \text{soft max}(\text{ReLU}(\tilde{y}_u^{j(U)^U} \omega_{bek} \tilde{y}_u^{k(U)})) \tag{12}$$

Where, the function soft max accepts a vector of arbitrary real-valued scores for offloading and turns them into probabilities that offloading. ReLU , is an activation function, ω_{bek} is an adjacency matrix and $B_{Q_{jk}}^t$ denotes the different progressive adjacency bandwidth. These regions may feature a combination of urban, suburban, and rural development patterns and often have low population densities and a weight parameter is given equation (13),

$$Y_u * h gX = \sum_{l=0}^{l-1} Q^l Y_u X_{l,1} + Q^{U^l} Y_u X_{l,2} \tag{13}$$

Where, h expression indicates main function, Y_u step diffusion process with picking the edge servers s , $X_{l,1}$ are learnable parameters, Q^l and Q^{U^l} are used to model the forward, In ecosystems, vegetation is essential because it affects the climate, creates habitats for a variety of creatures, stops soil erosion, and improves the general health of the surrounding areas given in equation (14),

$$y_u^{j(U)} * U \gamma = \sum_{Q=0}^Q \gamma(Q) y_u^{j(U)} (u - e \times Q) \tag{14}$$

Where y indicates the data storing point location in the edge server, γ denotes the decision of Numerous animals, including huge herbivores like bison, zebras, and antelope, as well as predators like lions, wolves, and cheetahs, can be found in grasslands. It is given in equation (15)

$$I_u = \tan(Y_u^U * U \Gamma_1) \Theta \sigma(Y_u^{(U)} * U \Gamma_2) \tag{15}$$

Where, Γ_1 and Γ_2 are the kernels for dilated casual difficulties, Θ denotes the element-wise multiplication, $\tau Y_u^{(U)}$ is a sigmoid beginning function, I_u dilated convolution operation with kernel and $\tan l$ values in parentheses represent vector indices. Finally PGCN classified as Low density and fragmented built-up land, High density built-up land greenery, Water bodies, crop land, and grassland wetlands, bare territory. Utilizing HSWOA is used to accurately enhance PGCN weight and bias parameters I and U . Here HSWOA is used to turning the PGCN weight and bias parameters.

E. Optimized using Harbor Seal Whiskers Optimization Algorithm (HSWOA)

In this section, the optimization using HSWOA [31] is discussed. The PGCN weight parameter I and U is optimized by HSWOA. Its flexibility mimics seals' skill in changing conditions, and its effective navigation eases traffic. Resilience to urban changes is ensured via robustness. Large-scale issue resolution is made possible by parallelism, and spatial configurations are optimized via exploration and exploitation techniques. Constant improvement based on experience is made possible by adaptive learning. The expertise of HSWOA in non-linear optimization is appropriate for intricate urban dynamics. Urban transportation, disaster response, resource distribution, social services, growth control, tourism, and information-driven choice making are all optimized by HSWOA. Through an optimization of layouts, infrastructure, and services, it improves the resilience, sustainability, and efficiency of cities. In addition to promoting economic growth and ensuring fair access to facilities, this strategy places a high priority on environmental sustainability.

Step 1: Initialization

The starting population of HSWOA is generated randomly. Then the initialization is derived in equation (16).

$$X_i = \begin{bmatrix} x_1^1 & x_2^1 & \dots & x_i^1 \\ x_1^2 & x_2^2 & \dots & x_i^2 \\ \cdot & \cdot & \cdot & \cdot \\ x_1^n & x_2^n & \dots & x_i^n \end{bmatrix} \tag{16}$$

Where, x denotes the total population of seals whiskers in the tracks; n denotes the n^{th} number of HSWOA while attacking towards its prey and X represents the distance between the prey and HSWOA.

Step 2: Random generation

Weight parameters are created at random after setup. Best fitness values are chosen based on a conditional explicit hyperparameter scenario.

Step 3: Fitness Function

Using initialized parameters, the fitness function generates a random solution. It calculated using optimizing parameter. Thus it is shown in equation (17),

$$Fitness\ function = optimizing [I,U] \tag{17}$$

Where, I represents the increasing accuracy and U represents the decreasing the RMSE.

Step4: Exploration Phase [I]

Harbor seals hunt and attack at a certain detection velocity using their whiskers. The seal follows its senses and lifts its whiskers away from its face beneath the surface. The water stirs as a prey moves. A seal can follow the movements of its prey by using its whiskers to detect the hydrodynamic trails the prey leaves behind. The seal can now identify the direction and proximity thanks to this. Then it is shown in equation (18),

$$U_j = \frac{N (2y_j^2 - E^2) + I}{2\pi (y_j^2 + E^2)^{5/2}} \tag{18}$$

Where, y_j^2 denotes the position of HSWOA in seal; U_j denotes the total amount of seals whiskers on surrounding area; N indicates the high sensing capacity of seal whiskers; π denotes the parameter of seal and n denotes the inspect objects or tracing its prey.

Step5: Exploitation Phase [U]

An anatomical structural diagram of the whisker is produced during the exploitation phase. There is a nerve in the cheek of a harbour seal that transmits information to the brain when its whiskers move in unison. By doing this, the seal is able to take in and comprehend intricate circumstances, such as the pathways that obstacles and prey leave behind. The ability to differentiate between attack angle and water flow is attributed to the oval cross section of the harbour seal's whiskers. Once their velocity for whisker sensing has been updated, the seals exploit the possibly advantageous poses of their prey. Then it is shown in equation (19),

$$w_j^{l+1} = Ms_1 w_j^l + cRs_2 (HQ_{best} - y_j^l) + bRs_3 (MQ_{best,j} - y_j^l) + U \tag{19}$$

Where, w_j^{l+1} denotes the total number of seals whiskers in the surrounding area; H indicates the distance among the seal also its target; s denotes a random number in the range $[0,1]$; HQ indicates a harbour seal to notice perturbations in its prey's underwater environment; l is the angle of attack for moving water and a_i^l denotes the cross sections there are in a single whisker.

Step 6: Termination Condition

The weight parameter values (I,U) of producer from PGCN is enhanced with the help of HSWOA, will iteratively repeat the step 3 until fulfil the hesitant criteria $x = x + 1$ is met. Then PGCN has managed the urban population flow and spatial distribution Classification by assessment with higher accuracy. Flowchart of HSWOA for optimizing PGCN parameter is shown in figure 2.

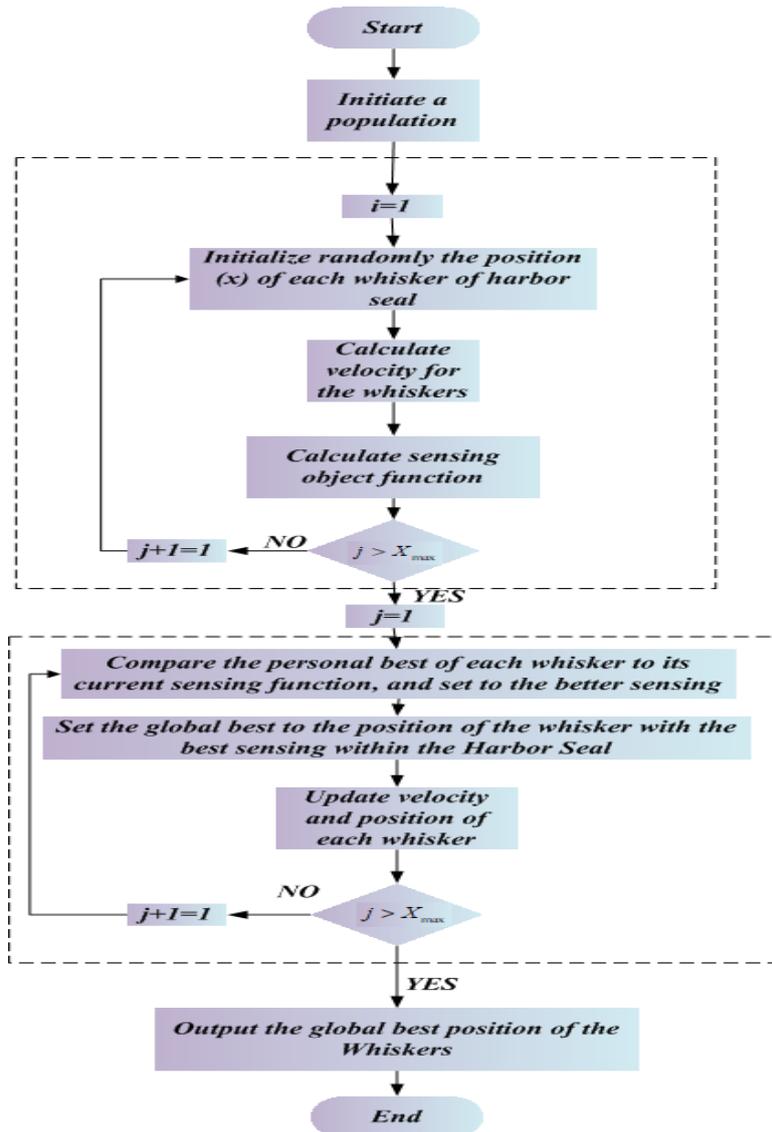


Figure 2: Flowchart of HSWOA for optimizing PGCN parameter

IV. RESULT AND DISCUSSION

Urban population flow and spatial distribution is one of the experimental results of the proposed UPF-SDP-CAA approach. In Implementation work was carried Python and evaluated by using several performance analysing ethics like RMSE, AUC, accuracy, recall, and precision are examined. A results of the proposed UPF-SDP-CAA methodology are compared to the current methods such as SIC-PF-GNN [21],UGP-UGS-ANN [22] and CAS-UG-SVM [23].

A. Performance Measures

Performance measures include accuracy, recall, precision, AUC and RMSE. The misperception matrix will used to scale the presentation parameters, it is decided.

- True Positive (TP) : Perfectly predictive urban population flow and spatial distribution system into positive class.
- True Negative (TN): Perfectly predictive urban population flow and spatial distribution system into negative class.
- False Positive (FP): Imperfectly predictive urban population flow and spatial distribution system into positive class.

- False Negative (FN): Imperfectly predictive urban population flow and spatial distribution system into undesirable class.

1) Accuracy

Accuracy is the capacity to measure an exact value. A metric called accuracy can be used to characterize the examples performance in all classes. It is quantified by the following calculation (20)

$$Accuracy = \frac{(TP+TN)}{(TP+FP+TN+FN)} \tag{20}$$

2) Recall

A machine learning model's recall quantifies its ability to find uplifting instances. Put another way, it measures the likelihood of getting a favourable result. That's provided in equation (21)

$$Recall = \frac{TP}{(TP+FN)} \tag{21}$$

3) Precision

Precision, or how well a machine learning model generates positive predictions, is one indicator of the algorithm's efficacy. The following stated equation (22) is used to measure it.

$$Precision = \frac{TP}{(TP+FP)} \tag{22}$$

B. Performance Analysis

The simulation results of the suggested UPF-SDP-CAA method are shown in Figure 3 to 7. The proposed UPF-SDP-CAA techniques linked to the SIC-PF-GNN, UGP-UGS-ANN) and CAS-UG-SVM techniques, in that order.

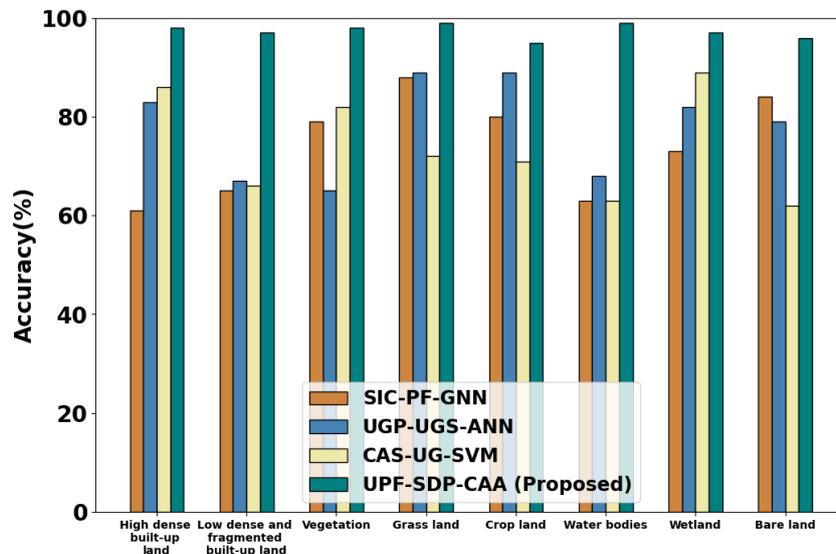


Figure 3: Performance analysis of Accuracy

Figure 3 shows presentation analysis of correctness. UPF-SDP-CAA shows especially notable accuracy gains in difficult-to-distinguish land cover types such as crop land, wetlands, and water bodies. These improvements are ascribed to the novel approaches integrated into UPF-SDP-CAA, which comprise advanced feature extraction methods, fine-tuned classification algorithms, and painstaking data pre-treatment procedures. These developments not only improve accuracy but also strengthen the technique's practicality. The proposed UPF-SDP-CAA technique reaches in the range of 19.36%, 26.42% and 23.27% higher accuracy for High dense built-up land, 16.42%, 23.36% and 19.27% higher accuracy for Low dense and fragmented built-up land, 33.26%, 17.26% and 20.41% higher accuracy for Vegetation, 22.36%, 15.42% and 18.27% higher accuracy for Grass land, 22.36%, 35.42% and 28.27% higher accuracy for Crop land, 24.36%, 15.42% and 23.27% higher accuracy Water bodies, 34.26%, 15.32% and 33.37% higher accuracy Wetland, 28.36%, 35.42% and 18.27% higher accuracy Bare land, Compared with existing techniques such as (SIC-PF-GNN), (UGP-UGS-ANN) and (CAS-UG-SVM) respectively.

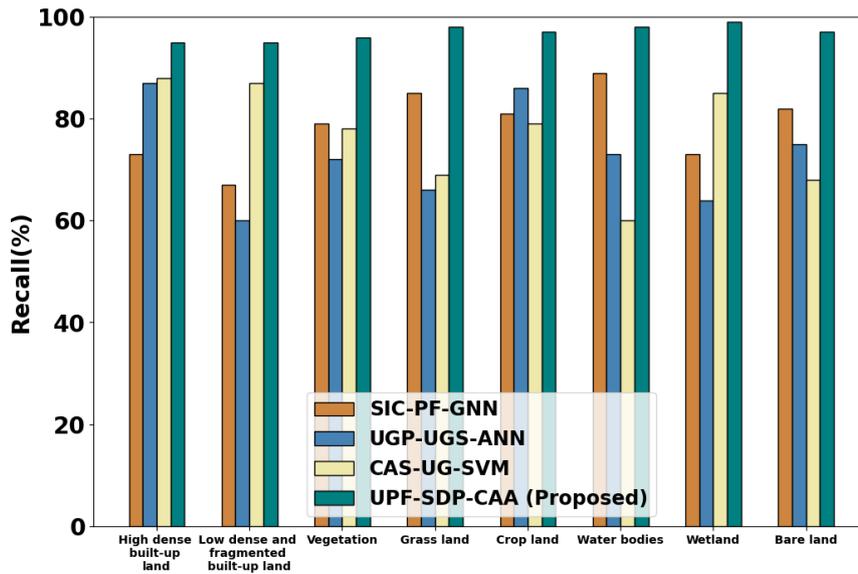


Figure 4: Performance analysis of Recall

Figure 4 shows Presentation analysis of Recall. These improvements demonstrate how well UPF-SDP-CAA can distinguish between various land cover types, which makes it a viable method for environmental monitoring and remote sensing applications. UPF-SDP-CAA represents a substantial development in land cover classification approaches with its higher performance across several categories, providing researchers and practitioners with a more dependable tool for managing and comprehending diverse landscapes. The proposed UPF-SDP-CAA technique reaches in the range of 18.36%, 16.42% and 28.27% higher Recall for High dense built-up land, 17.42%, 25.36% and 17.27% higher Recall for Low dense and fragmented built-up land, 32.26%, 18.26% and 24.41% higher Recall for Vegetation, 26.36%, 33.42% and 30.27% higher Recall for Grass land, 18.26%, 32.41% and 29.26% higher Recall for Crop land, 18.42%, 29.36% and 17.27% higher Recall for Water bodies, 15.44%, 35.36% and 24.47% higher Recall for Wetland, 19.42%, 26.36% and 22.27% higher Recall for Bare land Compared with existing techniques such as (SIC-PF-GNN), (UGP-UGS-ANN) and (CAS-UG-SVM) respectively.

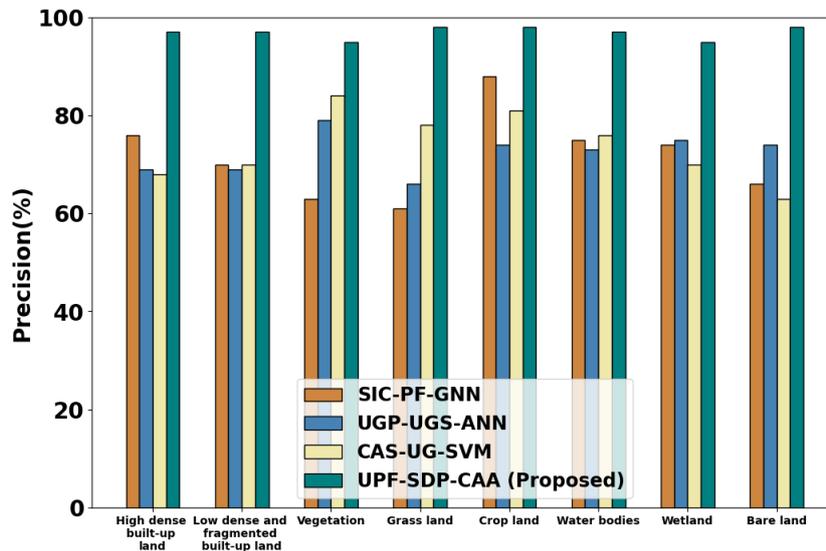


Figure 5: Performance analysis of Precision

Figure 5 shows Performance analysis of Precision. The significant gains in Precision measures across the board demonstrate its excellence. Specifically, as compared to its predecessors, UPF-SDP-CAA achieves notably higher Precision for high density developed land, vegetation, grassland, crop land, low-density and fragmented developed land, water bodies, wetland, and bare ground. UPF-SDP-CAA's performance improvements, which range from 15.36% to 38.26%, highlight the technology's potential to transform remote sensing applications by providing unparalleled precision and dependability in land cover mapping assignments. The proposed UPF-

SDP-CAA technique reaches in the range of 15.36%, 18.42% and 29.27% higher Precision for High dense built-up land, 18.26%, 32.41% and 29.26% higher Precision for Low dense and fragmented built-up land, 32.26%, 18.26% and 24.41% higher Precision for Vegetation, 28.36%, 23.42% and 33.27% higher Precision for Grass land, 38.26%, 16.26% and 29.41% higher Precision for Crop land, 34.26%, 16.26% and 27.41% higher Precision for Water bodies, 31.26%, 16.26% and 33.41% higher Precision for Wetland, 31.26%, 15.26% and 29.41% higher Precision for Bare land Compared with existing techniques such as (SIC-PF-GNN), (UGP-UGS-ANN) and (CAS-UG-SVM) respectively.

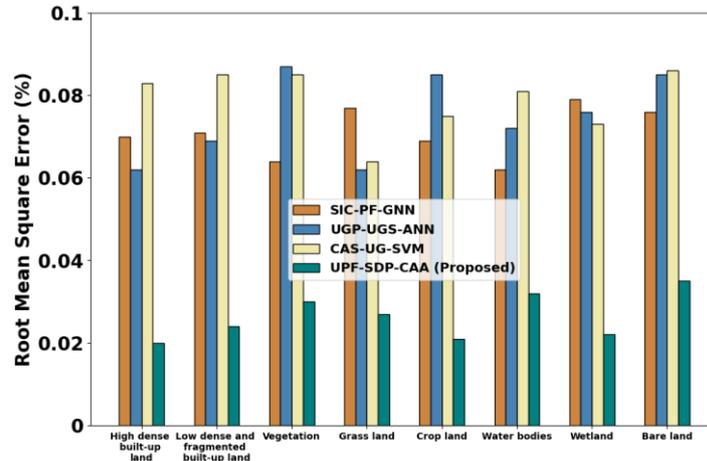


Figure 6: Performance Analysis of RMSE

Figure 6 shows Performance examination of RMSE. When compared to other strategies such as SIC-PF-GNN, UGP-UGS-ANN, and CAS-UG-SVM, the UPF-SDP-CAA technique shows notable improvements in performance across different types of land cover. In particular, it accomplishes significant reductions in the root mean square error (RMSE) for the following land uses: vegetation, grassland, cropland, and water bodies, bare land, low density and fragmented built-up land, and high density built-up land. UPF-SDP-CAA clearly demonstrates its superiority by improving accuracy across a wide range of land cover categories, with reductions ranging from 19.22% to 32.45%. This provides a potential approach for accurate land cover analysis and classification. The proposed UPF-SDP-CAA technique reaches in the range of 29.40%, 25.33% and 29.37% lower RMSE for High dense built-up land, 32.44%, 19.31% and 24.36% lower RMSE for less dense and fragmented built-up land, 28.46%, 27.38% and 20.45% lower RMSE for Vegetation, 22.40%, 27.32% and 25.24% lower RMSE for Grass land, 29.22%, 22.45% and 29.46% RMSE for Crop land, 19.22%, 26.45% and 19.46% RMSE for Water bodies, 23.22%, 32.45% and 29.46% RMSE for Wetland, 24.22%, 27.45% and 19.46% RMSE for Bare land, Compared with existing techniques such as (SIC-PF-GNN), (UGP-UGS-ANN) and (CAS-UG-SVM) respectively.

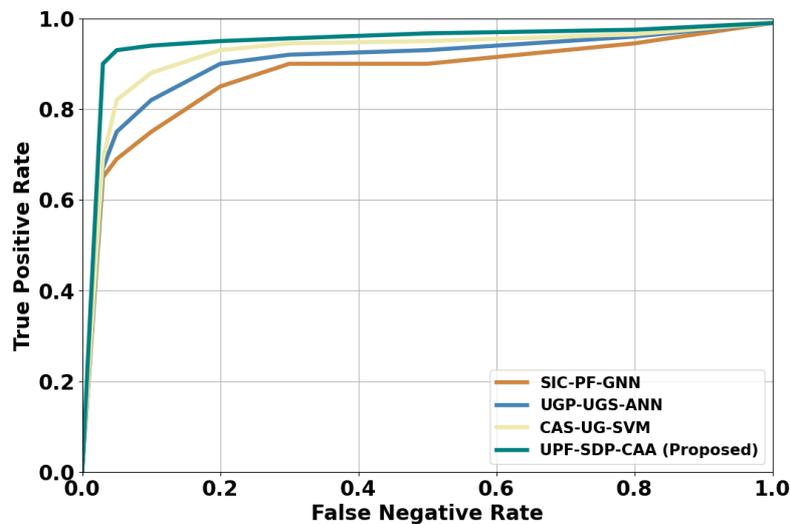


Figure 7: Performance analysis of AUC

Figure 7 shows Performance examination of AUC. With 25.26%, 22.26%, and 16.41% better AUC than (SIC-PF-GNN), (UGP-UGS-ANN), and (CAS-UG-SVM) techniques, respectively, the UPF-SDP-CAA technique shows notable performance benefits over current methods. This demonstrates its exceptional capacity for result prediction, suggesting that it could improve decision-making processes in pertinent fields. The proposed UPF-SDP-CAA technique reaches in the range of 25.26%, 22.26% and 16.41% higher AUC compared with existing techniques like(SIC-PF-GNN), (UGP-UGS-ANN) and (CAS-UG-SVM) respectively.

C. Discussion

This study develops the UPF-SDP-CAA model's initial step toward a urban population flow and spatial distribution. PGCN was contrasted with earlier findings from different investigations. Also found that the models were far better at urban population flow and spatial distribution prediction. Ultimately, UPF-SDP-CAA classifies the urban population flow and spatial distribution prediction. The benefits of PGCN and HSWOA optimization methods were combined to create a proposed urban population flow and spatial distribution. Were used the Landsat satellite dataset to examine the suggested Information from the urban population flow and spatial distribution methods. The dataset is first pre-processed the remove noise for image. Second, includes methods for extraction, the features extraction process includes the following: ward boundaries, geographic characteristics, population density, socioeconomic, and environmental factors.. Using various assessment metrics, the results confirmed the proposed method's outstanding performance and it has improved the traditional HSWOA exploitation and exploration phases, here infer that the proposed HSWOA performs much better than the traditional HSWOA. As a result, from an economic perspective, the proposed method is less costly than the comparison procedures. The cross-validation processes achieved overall accuracies.

V. CONCLUSION

In this paper, Analysis of urban population flow and spatial distribution patterns: a study based on cluster analysis algorithm (UPF-SDP-CAA) was successfully implemented. Here, Landsat satellite dataset were used in thorough evaluation tests to assess the presented technique. The proposed UPF-SDP-CAA technique is executed in Python. The presentation of suggested UPF-SDP-CAA method cover 32.44%, 19.31% and 24.36% lower RMSE and of 25.26%, 22.26%, and 16.41% higher AUC compared with existing (SIC-PF-GNN), (UGP-UGS-ANN) and (CAS-UG-SVM)methods. Future work takes into account several HSWOA optimizers for Landsat satellite dataset. In works includes for creating models using various techniques and combining multiple machine learning or deep learning algorithms. The outcomes acquired here additionally reveal and validate the superiority and advantage of the suggested method over the most popular approaches in the literature. The findings of this study could be examined and refined further to classify items more accurately and quickly utilizing deep learning neural networks. Future research can think about enlarging the data set to include industrial structures in order to give a more thorough evaluation of their impact on regional population flows. Furthermore, combining economic theories with process simulation may provide a comprehensive understanding of the dynamics of changes in population flow.

Acknowledgement

Guangxi Education Science's "14th Five-Year Plan" 2021 Special Project "Research on the Preparation of Education Special Planning in Land and Space Planning" (Project No. 2021ZJY1866)

REFERENCES

- [1] Sapena, M., Wurm, M., Taubenböck, H., Tuia, D., & Ruiz, L.A. (2021). Estimating quality of life dimensions from urban spatial pattern metrics. *Computers, environment and urban systems*, 85, 101549.
- [2] Li, C., Liu, M., Hu, Y., Wang, H., Zhou, R., Wu, W., & Wang, Y. (2022). Spatial distribution patterns and potential exposure risks of urban floods in Chinese megacities. *Journal of Hydrology*, 610, 127838.
- [3] Lin, Q., Xiang, M., Zhang, L., Yao, J., Wei, C., Ye, S., & Shao, H. (2021). Research on urban spatial connection and network structure of urban agglomeration in Yangtze River Delta—Based on the perspective of information flow. *International Journal of Environmental Research and Public Health*, 18(19), 10288.
- [4] Do, T.A.T., Do, A.N.T., & Tran, H.D. (2022). Quantifying the spatial pattern of urban expansion trends in the period 1987–2022 and identifying areas at risk of flooding due to the impact of urbanization in Lao Cai city. *Ecological Informatics*, 72, 101912.

- [5] Ning, F., Ou, S.J., Hsu, C.Y., & Chien, Y.C. (2021). Analysis of landscape spatial pattern changes in urban fringe area: A case study of HunheNiaodao Area in Shenyang City. *Landscape and Ecological Engineering*, 17(4), 411-425.
- [6] Sun, X., Zhang, B., Ye, S., Grigoryan, S., Zhang, Y., & Hu, Y. (2024). Spatial Pattern and Coordination Relationship of Production–Living–Ecological Space Function and Residents’ Behavior Flow in Rural–Urban Fringe Areas. *Land*, 13(4), 446.
- [7] Cengiz, S., Görmüş, S., & Oğuz, D. (2022). Analysis of the urban growth pattern through spatial metrics; Ankara City. *Land use policy*, 112, 105812.
- [8] Zhu, X.H., Lu, K.F., Peng, Z.R., He, H.D., & Xu, S.Q. (2022). Spatiotemporal variations of carbon dioxide (CO₂) at Urbanneighborhood scale: Characterization of distribution patterns and contributions of emission sources. *Sustainable Cities and Society*, 78, 103646.
- [9] Shan, L., Jiang, Y., Liu, C., Zhang, J., Zhang, G., & Cui, X. (2022). Conflict or coordination? spatiotemporal coupling of urban population–land spatial patterns and ecological efficiency. *Frontiers in Public Health*, 10, 890175.
- [10] Ariffin, N.A., Wan Ibrahim, W.M.M., Rainis, R., Samat, N., Mohd Nasir, M.I., Abdul Rashid, S.M.R., Ahmad, A., & Zakaria, Y.S. (2024). Identification of Trends, Direction of Distribution and Spatial Pattern of Tuberculosis Disease (2015-2017) in Penang. *Geografia-Malaysian Journal of Society and Space*, 20(1), 68-84.
- [11] Lei, W., Jiao, L., Xu, Z., Zhou, Z., & Xu, G. (2021). Scaling of urban economic outputs: Insights both from urban population size and population mobility. *Computers, Environment and Urban Systems*, 88, 101657.
- [12] Gao, S., Li, C., Rong, Y., Yan, Q., Liu, W., & Ma, Z. (2022). The places–people exercise: understanding spatial patterns and the formation mechanism for urban commercial fitness space in changchun city, China. *Sustainability*, 14(3), 1358.
- [13] Ding, R., Fu, J., Zhang, Y., Zhang, T., Yin, J., Du, Y., Zhou, T., & Du, L. (2022). Research on the evolution of the economic spatial pattern of urban agglomeration and its influencing factors, evidence from the Chengdu-Chongqing urban agglomeration of China. *Sustainability*, 14(17), 10969.
- [14] Li, T., Liu, K., Tang, R., Liang, J.R., Mai, L., & Zeng, E.Y. (2023). Environmental fate of microplastics in an urban river: Spatial distribution and seasonal variation. *Environmental Pollution*, 322, 121227.
- [15] Wang, W., Yang, Q., Gan, X., Zhao, X., Zhang, J., & Yang, H. (2022). Spatial Distribution Pattern and Influencing Factors of Homestays in Chongqing, China. *Applied Sciences*, 12(17), 8832.
- [16] Arantes, B.L., Castro, N.R., Gilio, L., Polizel, J.L., & da Silva Filho, D.F. (2021). Urban forest and per capita income in the mega-city of Sao Paulo, Brazil: A spatial pattern analysis. *Cities*, 111, 103099.
- [17] Zhang, P., Li, W., Zhao, K., & Zhao, S. (2021). Spatial pattern and driving mechanism of urban–rural income gap in Gansu province of China. *Land*, 10(10), 1002.
- [18] Isinkaralar, O., Isinkaralar, K., & Bayraktar, E.P. (2023). Monitoring the spatial distribution pattern according to urban land use and health risk assessment on potential toxic metal contamination via street dust in Ankara, Türkiye. *Environmental Monitoring and Assessment*, 195(9), 1085.
- [19] Lin, J., Zhang, W., Wen, Y., & Qiu, S. (2023). Evaluating the association between morphological characteristics of urban land and pluvial floods using machine learning methods. *Sustainable Cities and Society*, 99, 104891.
- [20] Zhang, H., Zong, R., He, H., Liu, K., Yan, M., Miao, Y., Ma, B., & Huang, X. (2021). Biogeographic distribution patterns of algal community in different urban lakes in China: insights into the dynamics and co-existence. *Journal of Environmental Sciences*, 100, 216-227.
- [21] Luo, M., & Chen, Y. (2024). Simulating inter-city population flows based on graph neural networks. *Geocarto International*, 39(1), 2331223.
- [22] Dinda, S., Chatterjee, N.D., & Ghosh, S. (2021). An integrated simulation approach to the assessment of urban growth pattern and loss in urban green space in Kolkata, India: A GIS-based analysis. *Ecological Indicators*, 121, 107178.
- [23] Zhang, B., & Xia, C. (2022). The effects of sample size and sample prevalence on cellular automata simulation of urban growth. *International Journal of Geographical Information Science*, 36(1), 158-187.
- [24] Wang, X., Ding, S., Cao, W., Fan, D., & Tang, B. (2020). Research on network patterns and influencing factors of population flow and migration in the Yangtze River Delta urban agglomeration, China. *Sustainability*, 12(17), 6803.
- [25] Yang, Y., & Wong, K.K. (2013). Spatial distribution of tourist flows to China's cities. *Tourism Geographies*, 15(2), 338-363.
- [26] Zhen, F., Qin, X., Ye, X., Sun, H., & Luosang, Z. (2019). Analyzing urban development patterns based on the flow analysis method. *Cities*, 86, 178-197.
- [27] Peng, Y., Liu, J., Zhang, T., & Li, X. (2021). The relationship between urban population density distribution and land use in Guangzhou, China: A spatial spillover perspective. *International Journal of Environmental Research and Public Health*, 18(22), 12160.
- [28] Nako, J., Psychalinos, C., Elwakil, A.S., & Minaei, S. (2023). Non-Integer Order Generalized Filters Designs. *IEEE Access*.
- [29] Luo, Y.S., Zhao, X.L., Jiang, T.X., Chang, Y., Ng, M.K., & Li, C. (2022). Self-supervised nonlinear transform-based tensor nuclear norm for multi-dimensional image recovery. *IEEE Transactions on Image Processing*, 31, 3793-3808.

- [30] Shin, Y., & Yoon, Y.J. Pgcn: Progressive graph convolutional networks for spatial-temporal traffic forecasting. arXiv 2022. *arXiv preprint arXiv:2202.08982*.
- [31] Zaher, H., Al-Wahsh, H., Eid, M.H., Gad, R.S., Abdel-Rahim, N., & Abdelqawee, I.M. (2023). A novel harbor seal whiskers optimization algorithm. *Alexandria Engineering Journal*, 80, 88-109.