Data-Driven Identification of Industrial Clusters: a Patent Analysis Approach

Wenguang Lin, Ting Wang, Zhizhen Chen, and Renbin Xiao

***Abstract*—Accurate Identification of Industrial Clusters (IIC) serves as a reference for regional economic policymaking and enterprise development decision-making. Although data-driven methods have been extensively used in previous studies to support objective and effective work, both the data sources and research algorithms have significant shortcomings for IIC. To address these challenges, this paper proposes a novel research framework that integrates patent mining and machine learning. Patents, with their quantifiable knowledge attributes and accessibility from public databases, are particularly suited for macro-level analysis of innovation activities, providing robust support for identifying and analyzing clusters on a national scale, especially knowledge-intensive ones. This article introduces an improved density-based parameter adaptive algorithm designed to effectively carry out IIC based on the geographical location of patent applicants. Based on spatial cluster types defined by Markusen (1996), target clusters are classified using patent analysis. Four quantitative indexes–scale, output, efficiency, and quantity–are proposed to evaluate clusters based on their spatial structure and industrial organization. The practical application is demonstrated through a case study of China's flexible electronics industry (FEI). Additionally, the Silhouette Coefficient (SC) index is employed to compare the effectiveness of the proposed algorithm against other methods. This article advances the theory of IIC, and provides foundation for scholars, calling for empirical research on industrial clusters from the perspective of individual enterprises. It also provides practical guidance for enterprises and policymakers on the application of IIC.**

***Index Terms*—Flexible electronics industry, identification, industrial clusters, patent analysis, spatial distribution.**

I. INTRODUCTION

As an effective form of spatial production organization, industrial clusters, also known as industrial districts, play an important role in promoting regional economic and social development. They have emerged as a popular policy tool for economic development strategies and plans [1]. For example, China's Ministry of Industry and Information Technology (MIIT) has published a list of advanced manufacturing clusters, such as Yangtze River Delta (YRD), Pearl River Delta (PRD), and Beijing-Tianjin-Hebei (BTH). It emphasizes the deepening of the optimization of urban industrial spatial distribution and clearly demonstrates the role of clusters in economic development [2]. The IIC is a common tool for governments in formulating cluster development policies. The key to adjusting the direction of the clusters development to adapt to the ever-changing development environment lies in thoroughly analyzing the category characteristics of industrial clusters [3]. For example, Brenner [4] proposed IIC in Germany and discusses the characteristics of domestic clusters' distributions, providing important support for the implementation of industrial policies.

Manuscript received 16 December 2023; revised 11 February 2024, 13 May 2024 and 8 September 2024; accepted 31 October 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 52275249, in part by the Social Science Foundation of Fujian Province under Grant FJ2021B128 and Grant FJ2023B077. Review of this manuscript was arranged by Department Editor Dr. Tugrul Daim.(*Corresponding author: Renbin Xiao*.)

Wenguang Lin and Ting Wang are with the School of Mechanical and Automotive Engineering, Xiamen University of Technology, Xiamen 361024,China (e-mail: linwg@xmut.edu.cn; wt22992022@163.com).

Zhizhen Chen is with the School of Business, University of Greenwich, London SE10 9LS, UK; (e-mail:z.chen@gre.ac.uk).

Renbin Xiao is with the School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan 430074, China (e-mail:rbxiao@hust.edu.cn).

As an interdisciplinary field composed of economics, management, and geography, industrial cluster research has extensive studies on various aspects such as concepts, mechanisms, characteristics, and related policies [5], [6], [7], [8]. However, due to the different research paradigms adopted by scholars from diverse disciplinary backgrounds, it has been challenging to harmonize the research themes. This significantly constrains the breakthrough of cluster theoretical research and the guidance cluster policy formulation and implementation. In the context of high-tech clusters, innovation entities form the foundation of cluster competitiveness. Patents, representing the largest source of technical information globally [9] are vital for knowledge dissemination, particularly in high-tech industries, such as electronics, biotechnology, and artificial intelligence [10]. Patent information comprises structured data such as patent numbers, citations, classifications, and dates, as well as unstructured data containing extensive information about applicants and enterprise technologies, including technical content, abstract, and specifications [11].

Global scholars extensively use patents to study the spatial diffusion of innovation, the complexities of geographic spaces, innovation networks, and the relationship between technologyand regional innovation performance. Patents are crucial in the technology innovation process, granting inventors exclusive rights to specific creative ideas or designs for a limited period [12], [13], [14], thereby fostering innovation. Additionally, patent output is closely linked with other innovation metrics, such as Research and Development (R&D) expenditures and the number of active researchers [15], [16]. Building on this perspective, Alcácer and Zhao [17] argued that the spatial concentration of patents serves as a clear indicator of the existence of technology clusters.

The determination of traditional cluster types has primarily relied on expert judgment. While expert judgment offers the advantages of generality and expertise, it often comes with subjective biases. Currently, the primary reliance on government economic data, which often lacks a higher granularity and rarely involves specific enterprises, poses challenges in data acquisition. Through data-driven approaches, with the in-depth analysis of the technology and the dynamic process of industrial clusters, enabling a deeper understanding of the persistent technological competitive advantages of industrial clusters.

Current research on IIC mainly focuses on geography concentration, and overlooks other crucial characteristics, such as output, efficiency, and scale [18]. These characteristics are essential for the identification and analysis of technology-based industrial clusters. Insufficient analysis of industrial clusters can hinder the development of the theoretical framework and the comprehensive understanding of cluster-related issues. Therefore, enhancing the IIC is of significant importance for optimizing the accuracy and effectiveness of governance. This is crucial for the development of industrial clusters. In order to address the research gaps, this study aims to introduce a fusion data-driven recognition method of IIC and subsequently identify their spatial characteristics.

II. Research Overview

This section presents the current state of research on data sources as well as on IIC methods and categories.

A. Patent Analysis

In practice, understanding the spatial dynamics of innovation activities at the national regional level faces significant challenges. National economic census data on microenterprises are difficult to obtain and are not generalizable, and statistics on technology and innovation tend to be more limited. The lack of comprehensive data makes it difficult to explore the true characteristics of a region.

In recent years, with the increasing importance of patents in economic development, countries have been strengthening the protection, management, and utilization of intellectual property rights [19]. Scholars have categorized industries with high patent density or volume, as patent-intensive industries, which are also increasingly supported by policies in various countries [20]. Patents often used as a proxy for innovation output, provide formalized and quantifiable data that reflect the knowledge capabilities of firms, independent of human factors [21].

Patents, as an important data source for studying innovation activities on a national scale. Due to their quantifiable knowledge properties and the fact that they can be widely accessed from public databases. They are also suitable for analyzing explicit knowledge transfer within industries at the macro level.

This article delves into the identification of flexible electronics (FE) industry in China and the classification of industrail clusters (CIC) using patents. The patents are sourced from the patent information service platform and micro-level data of these enterprises, including enterprise size, type, latitude, and longitude, are obtained from the open data platform that belongs to the Ministry of Natural Resources of China (https://www.mnr.gov.cn/sj/sjfw/) and Baidu open data platform (https://api.map.baidu.com/).

B. IIC

Industrial clusters serve as an important carrier for fostering local endogenous growth and facilitating the development of regional innovation [22], and are an organizational form of industry where numerous enterprises are clustered within a region and supported by universities, research institutions, and financing organizations. Industrial clusters aim to foster new knowledge production, technological advancement, and product innovation through enterprise networks. The independent innovation ability serves as the core component and developmental foundation of industrial clusters.

The essential feature of industrial clusters lies in the intricate network connections among enterprises. From the enterprise's perspective, when a certain number of enterprises gather and operate within the same value chain, they constitute an industrial cluster. IIC refers to the process of detecting and delineating clusters by identifying patterns of industrial linkages and geographical proximity among enterprises [23]. This is considered the initial step in industrial cluster research and is crucial for subsequent empirical analyses. The knowledge structure in the field of industrial clusters is depicted in Fig. 1. Moreover, clusters offer potential advantages as catalysts for perceiving the demand and opportunities for innovation [24].

There are many definitions in current literature of industrial clusters, which can be categorized into three main categories based on the characteristics [33]:

1) Industry clusters formed by industrial linkages: In these clusters, the goods, and services of a group of enterprises within an industrial cluster are intricately connected, creating an interdependent production network [34].

2) Industry clusters formed by geographical proximity: These clusters represent the institutional and geographic concentrations of similar or related enterprises [35].

3) Industry clusters formed by both industrial linkages and geographical proximity: In these clusters, enterprises and supporting institutions are interconnected, fostering a more competitive network [36].

Compares to other characteristics, more scholars emphasize that spatial concentration is a key feature of industrial clusters [37],[38],[39]. Essentially, cluster reflects proximity, which refers to the spatial clustering of enterprises from a micro-geographical perspective.In this context, we choose the second type of cluster as the object of IIC research.

In practice, objective and accurate IIC is critical for supporting regional development policymaking and enterprise investment decisions [40]. Any inclusion or exclusion decisions in the IIC process, possibly influenced by subjective factors, can have significant consequences. These decisions may also impact the allocation of subsidies and the effectiveness of policies, and even force enterprises inside and outside the cluster to adjust their development strategies [41]. For example, Chinese government often introduces many incentives such as tax reduction and public subsidies to attract enterprises

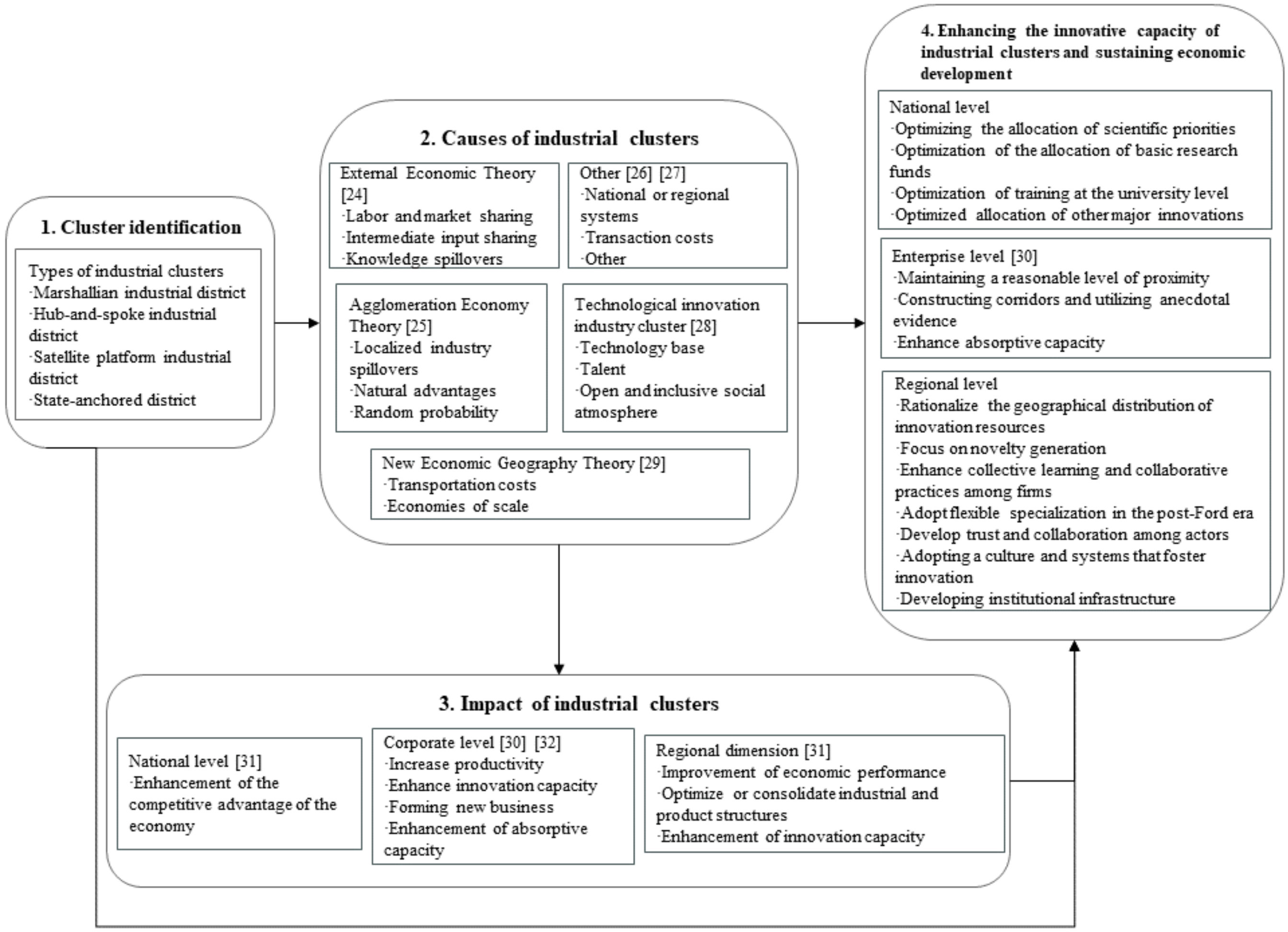


Fig. 1. Knowledge structure in the field of industrial clusters.

to invest in specific clusters, such as the Zhongguancun Electronic industrial cluster and the Ningbo Chemical industrial cluster [42], [43]. Under this condition, many firms willing to reconsider their location decisions, and choose to open subsidiaries within clusters to obtain those benefits. In addition, enterprises may proactively join clusters to reduce costs and access richer knowledge resources, a good example is that many multinational corporations in the US relocating headquarters to Austin [44].

Therefore, it is necessary to employ an unsupervised approach to achieve the objective and accurate cluster identification. Current research primarily relies on quantitative methods [45]. For instance, Rosenfel [46] employed the location entropy method to identify industrial clusters in Germany, while Benita proposed principal component analysis to assess industrial clusters dynamic [47]. Brachert and Kubis used qualitative input-output analysis to identify the production and knowledge generation activities within industrial clusters in Saxony, Germany [48]. However, these methods primarily focus on measuring input-output connections between regional industries without defining the spatial scope of industrial clusters. Moreover, the lack of spatial analysis and the reliance on statistical data or qualitative descriptions limit the discovery of various spatial elements within the region.

In addition, clustering algorithms have also been introduced to replace manual statistics during the process of IIC. These algorithms are categorized based on clustering criteria into division-based, hierarchical-based, model-based, and density-based clustering [49].

Division-based clustering algorithms detect local clusters in small neighborhoods obtained by recursive spatial tessellation, and then merge them hierarchically, following the Divide-and-Conquer paradigm [50]. However, division-based clustering requires determining the final classification of the dataset [51], and since the number of industrial clusters is typically uncertain, this algorithm may not be suitable for IIC.

Hierarchical-based clustering algorithms discover the underlying structure of a dataset in unsupervised learning scenarios where the ground truth is unknown and classical machine learning classifiers are not suitable. However, hierarchical-based clustering is only suitable for discovering spherical or nearly spherical clusters [52]. As industrial clusters often exhibit diverse forms, this algorithm may not be helpful for non-spherical clusters. Model-based clustering algorithms assume that the input dataset follows a potential distribution [53]. Each cluster within this algorithm is modeled to find the best fit of the data to the given distribution. This algorithm mainly refers to probabilistic model-based methods and neural network model-based. However, if the assumptions underlying the model are not valid, it may affect the outcome. Since the distribution of enterprise data is often random, model-based clustering algorithms may not be suitable for clustering such data.

Density-based clustering algorithms use density as the primary criterion for clustering, enabling the detection of clusters with arbitrary shapes without the need to preset the number of clustering. One representative algorithm is Density-Based Space Clumping of Applications with Noise (DBSCAN), initially proposed by Ester et al. [54]. DBSCAN is a representative algorithm in density-based spatial clustering, capable of segmenting high-density areas into clusters while effectively filtering out low-density areas. Moreover, it facilitates the clustering of arbitrary shapes within datasets containing noise, identifying outliers.

While Zhao et al. [55] employed a standard developmental ellipses algorithm to identify geographic industry clusters. This method may obtain clusters with compact internal elements, but struggles to identify irregularly shaped clusters, diverging from real-world scenarios. Alternatively, Zhou et al. [7] used patent information to analyze the geographical distribution of enterprises, its spatial cluster identification relies on the Louvain algorithm. However, the Louvain algorithm generates different cluster results based on the number of iterations.

In contrast, DBSCAN used in our research offers a more practical approach by generating clusters based on the number of elements and spatial distance. In addition, DBSCAN is more suitable for identifying clusters of arbitrary shapes in space. It has found extensive application in processing spatial data obtained from platforms such as Twitter, Google Maps, or GPS-based systems [40]. However, its application in geographic economics remains relatively uncommon [56].

The traditional DBSCAN algorithm effectively addresses the recognition of an unknown number of industrial clusters, irregular shapes, and the need to filter out noise. However, two defects exist in the algorithm. Firstly, it is quite sensitive to the parameter of neighborhood radius *Eps* and the minimum density threshold *MinPts*, both of which must be set manually. Secondly, the cluster results may be inaccurate when clustering datasets with significant density variations [57].

To mitigate these issues, adaptive determination of DBSCAN algorithm parameters has been proposed. Yue et al. [58]. Introduced an algorithm based on the statistical information of data to determine the *Eps* parameter, allowing for a wider search range. However, the *MinPts* parameter remains a fixed value, limiting its ability to accurately reflect the dataset's distributional characteristics. Jahirabadkar et al. [59]. Dynamically set the *Eps* parameter according to the density distribution of the dataset in each dimension, but the *MinPts* parameter still requires manual input. Yang et al. [60] introduced the concept of distance distribution, setting *MinPts* as a certain value, calculating and sorting the distance between two objects, and then obtaining *Eps*. However, the number of clusters still needs to be pre-set. Moreover, Song et al. [61] calculated the probability density distribution curve of distance values based on the statistical characteristics of the dataset, determining the *Eps* from the steepening area through manual observation.

In addition, Zhou et al. [62] proposed the adaptive and fast DBSCAN algorithm, fitting a distance curve corresponding to the input dataset with a polynomial to address the problem of the inflection point of the curve. Determined the optimal *Eps* values as the maximum distance corresponding to the inflection point. An adaptive DBSCAN based on the inflection point of the K curve has been proposed [63],[64]. The algorithm defines density layers of the dataset based on the quadtree to identify clusters with varying densities, although it still needs pre-setting of relevant parameters.

Furthermore, Campello et al. [65] introduced HDBSCAN algorithm. An extension of DBSCAN that does not require pre-defined parameters like *MinPts* and *Eps* and can handle noise and outliers without setting the number of clusters. Instead, HDBSCAN requires parameters like min_cluster_size and min_samples, and the setting of min_cluster_size and min_samples parameter affects the accuracy and number of final clustering results [66]. Moreover, HDBSCAN has a high computational complexity, which may not suitable for dealing with large-scale datasets.

To address the limitations of traditional DBSCAN algorithm in clustering datasets with varying densities, this article uses an improved density-based parameter adaptive DBSCAN algorithm based on K-Average Nearest Neighbor (KANN), referred to as KANN-DBSCAN [67]. Furthermore, a sensitivity formula is integrated to further refine the accuracy of the KANN-DBSCAN algorithm, resulting in the development of the KANNs-DBSCAN. This algorithm automatically optimizes clustering parameters based on the distance distribution characteristics of the dataset, enabling adaptive selection of *Eps* and *MinPts* parameters. Subsequently, the DBSCAN algorithm is applied to cluster enterprise data according to the latitude and longitude distribution of enterprise. This application effectively clustering industrial clusters, and robustly supporting subsequent analysis.

C. CIC

In practice, influenced by internal enterprises, industrial clusters are demonstrated in various forms; CIC is a process to categorizing different types of industrial clusters based on various factors such as scale, industry, industrial level, structural form, and regional distributions. The spatial and industrial layout observed in cities like Beijing, Shanghai, Guangdong [68], [68], [70], China, and Lille, France [71]. Industrial clusters play various roles and have significant impacts on enterprises. In China, for example, clusters located in Beijing tend to have strong political affiliations, whereas those in Guangdong and Shanghai are more market-oriented. Consequently, members of clusters in Beijing tend to rely more on government support, whereas those in Guangdong and Shanghai focus more on market expansion and technological innovation [72]. CIC is a process of categorizing different types of industrial clusters based on various factors such as scale, industry, industrial level, structural form, and regional distributions; The CIC is crucial as it based on a rational and objective categorization of clusters. In other words, the purpose of cluster classification is to provide more accurate targets for enterprises' market strategies, relocation decisions, investment, and other business activities.

This classification is particularly crucial for knowledge-intensive clusters, where enterprises are exposed to greater risks and are easily influenced by the internal cluster environment. This dynamic often results in frequent relocations between clusters.

Knowledge-intensive industries primarily rely on core production factors such as intelligence, knowledge, technology, information, and skills. These factors play a core role in industrial economic activities, with knowledge resources constituting a significant proportion compared to traditional factors like labor and capital. Patents, as a representative of knowledge resources [73], are therefore highly suitable for studying the characteristics of knowledge-intensive industry clusters.

Moreover, patents serve as efficient tools for strategic management, particularly at the national level where they inform public policy development. For example, detailed studies on patent applicants and countries involved in electrochemical energy storage technologies guide future technology policies. Similarly, analyses of patent time horizons, distribution of technology types, and analysis of technology trends in the field of solar thermal utilization inform governmental energy policies [74]. Global wind turbine companies' patent analyses have also provided direction for Asian and European countries in developing new markets and corporate innovation policies [75]. Therefore, this article proposes a study on CIC based on patent analysis.

Marshall [24] was the first scholar to discover the concept of industrial cluster and proposed two important concepts: Marshallian Industrial Cluster (MIC) and Italian industrial district (flexible specialized clusters). These concepts underwent extensive debates among academics regarding their impacts on academia and regional economies. Markusen [76], for instance, believes that these districts represent a specialized form of industrial clusters rather than a universal phenomenon. More widely acknowledged types of industrial districts include MIC, Hub-and-Spoke Industrial Cluster (HSIC), Satellite Platform Industrial Cluster (SPIC), and State-Anchored Industrial Cluster (SAIC), each characterized distinctly, as shown in Table I.

In the framework of classifying the spatial structure of industrial clusters, the analysis of patent numbers within each cluster, categorized by enterprise nature and size, aims to identify distinct categories based on enterprise microdata. This approach facilitates a deeper understanding of the relationship between the industrial technology environment, cluster evolution, technical knowledge conditions, and economic growth.

This study adopts Markusen's [76] perspective on the spatial structure of industrial clusters, with a particular focus on the pervasive industrial district type classification.

III. Research Framework and Methodology

In order to fill the current research gaps, this study aims to propose a fusion data-driven IIC method to identify its spatial organization characteristics. This method enables overcoming the limitations of rough management of industrial clusters and realizing the development and growth of industrial clusters. First, based on the characteristics and development cycle of the regional industry, we formulated the search terms for technological patents, encompassing the patent application region, keywords, classification numbers, and application dates. Subsequently, the data noise is addressed through procedures such as duplications-merging and eliminating rejected patents. In addition, geographic information of the patent applicants is obtained by map APIs. The spatial clustering of enterprises are identified using the KANNs-DBSCAN algorithm in conjunction with the composite extraction rules derived from the Markusen model for various cluster categories. Lastly, a set of indices comprising output, scale, efficiency, and quantity index are proposed to evaluate four categories of industrial clusters.

A. KANNs-DBSCAN Clustering Algorithm

As mentioned in the introduction section, current research highlights the significance of algorithms in the IIC. However, existing clustering algorithms often suffer from the shortcomings of requiring predefined parameters. To address this issue, this article adopted an improved density-based clustering algorithm, KANNs-DBSCAN. The clustering effect of DBSCAN is related to the density parameter, signifying the presence of *MinPts* data points within a circle with *Eps* as the radius [67], [77]. Therefore, this article introduces the density parameter as an indicator to evaluate the clustering effect of DBSCAN. The specific calculation method is outlined as follows:

$$Density = \frac{MinPts}{\pi Eps^2}. \tag{1}$$

Given that the DBSCAN algorithm lacks an adaptive method for calculating *MinPts* and *Eps*, this article uses the KANNs-DBSCAN algorithm. It deterministically obtains the *MinPts* and *Eps* values by assessing the closest distance between each data point and its Kth data point, thereby computing the density value. Li et al. [67] argued that the clustering result is considered stable when the same number of clusters is generated three times consecutively. However, this criterion is deemed subjective and lacks empirical support, prompting an enhancement in this article. They also suggested the number of clusters of the clustering results usually tends to be stable as the density parameter decreases, and the optimal results are obtained when a certain density interval remains stable. This study employs a sensitivity formula to assess the stability of clustering results. The magnitude of sensitivity change reflects the rate of density parameter variation with increasing K, with stable sensitivity indicating the optimal number of clusters. Accordingly, the optimal combination of *MinPts* and *Eps* value is determined as the input parameter of DBSCAN. The specific algorithm is outlined as follows:

$$Sensitivity = \frac{\triangle Density}{\triangle K}. \tag{2}$$

The advantage of the KANNs-DBSCAN algorithm is evident in its ability to determine the appropriate density threshold EPs and MinPts. It enables the balance between these

TABLE I
CLASSIFICATION OF REGIONAL INDUSTRIAL CLUSTERS

	MIC	HSIC	SPIC	SAIC
Enterprise size	Small and medium-sized enterprises (SMEs)	Single large enterprise	Branch of parent company	State monopoly
Enterprise source	Local company	Local company	Parent company	Local company
key feature	SMEs-dominated; low level of specialization; a dense network of competitive and cooperative relationships; based on trust relationships.	Coexistence and symbiosis between large firms and SMEs; pronounced internal hierarchical structure.	SMEs-dominated; strong dependence on external enterprises (parent companies); the low-end segment of the industry based on low labor costs.	National institutions (e.g., weapons research laboratories, universities, government offices) dominated; based on the high level of cooperation and connection among external companies in the region.
Market structure in the cluster	Perfect competition	Oligopoly	Incomplete monopoly	Complete monopoly

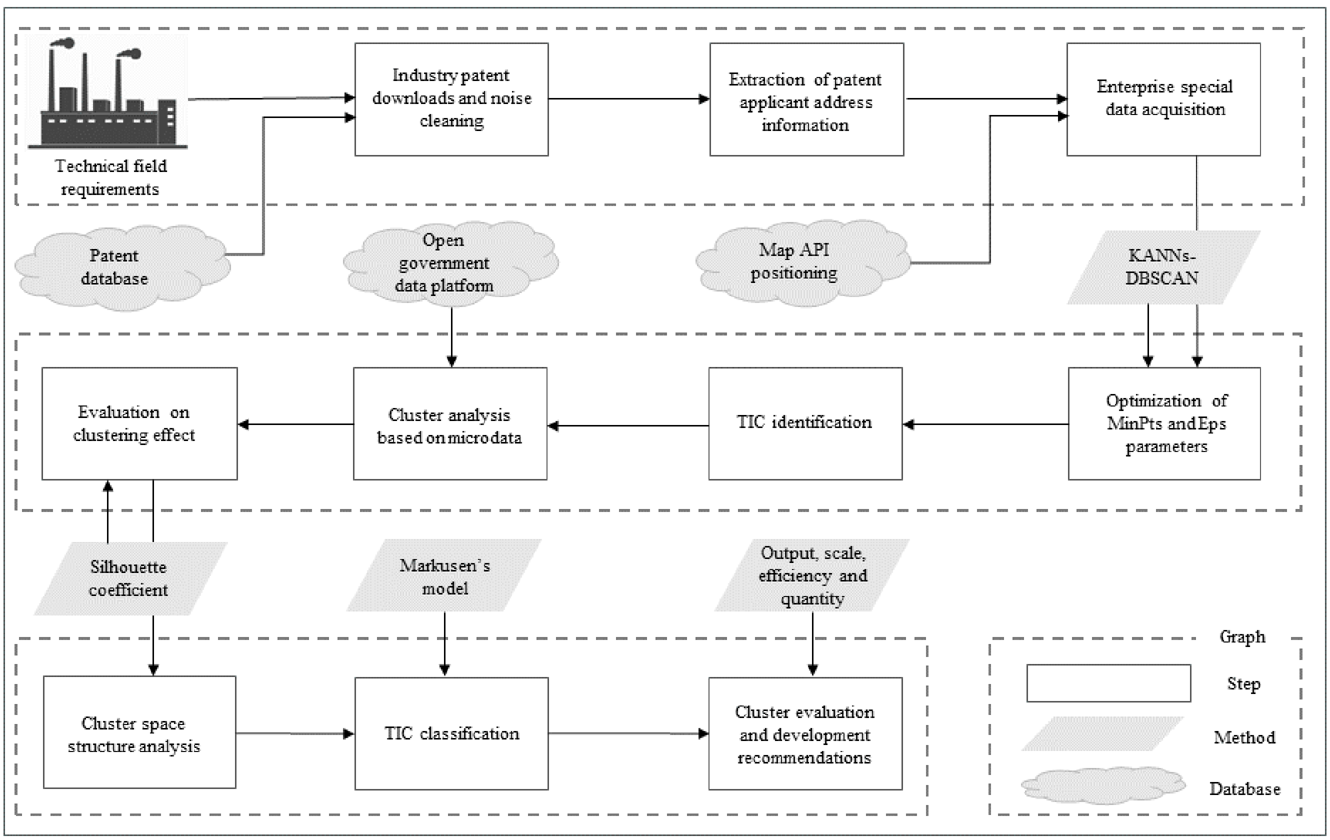


Fig. 2. Research framework.

two parameters to obtain the optimal parameter pair with an unsupervised approach, then achieving high accuracy and low computational complexity. To leverage this advantage, this article utilized the spatial distribution characteristics of the patent applicant's data, translated into longitude and latitude data, to generate the density threshold list based on KANN and the mathematical expectation method. The process of identifying clusters using the algorithm is shown in Fig. 3.

1) List of Eps parameters generation: The KANN algorithm and the mathematical expectation method are used to generate the *Eps* list. KANN is an extension of the mean K-Nearest Neighbor algorithm, by calculate all K values, and lastly obtaining the KANN distance of the dataset. The specific steps involve calculating the distance distribution matrix and arranging each row of this matrix in ascending order to obtain the vector D_K. This involves averaging the elements in vector D_K to obtain KANN distance $\overline{D_K}$ of vector D_K, which can be used as a candidate *Eps* parameter. By calculating all the K values, the list of *Eps* parameters D_{Eps} is obtained, defined as follows:

$$D_{Eps} = \{\overline{D_K} | 1 \leq K \leq n\}. \quad (3)$$

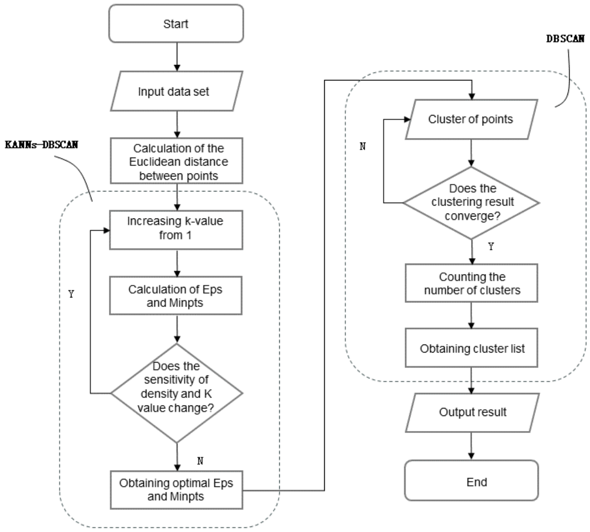


Fig. 3. Process of identifying clusters.

2) List of MinPts parameter generation: The mathematical expectation method is used to generate the *MinPts* list. For a given *Eps* list, the number of neighboring objects corresponding to each *Eps* is sequentially determined. In addition, the mathematical expectation value of the number of neighboring objects for all objects is calculated and set as the neighborhood density threshold *MinPts* for the dataset *D*, defined as follows:

$$MinPts = \frac{1}{n}\sum_{n=1}^{\infty}(P_i) \tag{4}$$

where P_i is the number of *Eps* neighboring objects of the *i* object and n is the total number of objects in the dataset D.

3) Adaptive determination of optimal parameters: According to KANN, the candidate *Eps* and D_{Eps} are derived for the data set D. The KANN distance corresponding to different *K* values (*K* = 1, 2, ..., n) are selected sequentially, i.e., the elements in set D_{Eps} are selected as the candidate *Eps*, and *MinPts* parameters are obtained from the formula (4). Based on the generated *Eps* and *MinPts* parameter lists, the corresponding parameters can be calculated using the equation to obtain the density parameter list and sensitive parameter list. When the change in sensitivity begins to stabilize, the clustering results are considered to be stabilized and the number is considered optimal. The value corresponding to the optimal number of clusters is selected as the optimal *K* value, while the *MinPts* and *Eps* corresponding to the optimal *K* value are the optimal *Eps* parameter and *MinPts* parameter.

B. Cluster Category Identification Metrics

In the previous section, this study proposes a data-driven method for identifying industrial clusters, which obtains the optimal clustering through spatial element analysis. Given that geographic clustering and spatial correlation are key features of industrial clusters, it is essential to systematically analyze the interrelations among cluster members. This also enables researchers to gain deeper insight into the complex internal structure of industrial clusters.

Currently, there are several tools available for studying the internal characteristics of clusters, with scientometric analysis being the most widely used method [78]. This approach objectively maps the scientific knowledge domain and identifies research themes and corresponding challenges through the mining and visualization of patent documents and papers [79], [80]. Since patents are closely related to R&D activities and readily accessible online, they serve as measurement tools for firms' innovative activities and a key tool in research. [81]. In this regard, this article employs patent analysis as the key means to support IIC research.

Based on the results of spatial clusters on patent application firms, this section employs a patent analysis method to assess the innovation ability, nature, and scale of enterprises within the cluster. Additionally, it formulates classification rules according to the characteristics of the four clusters listed in Table I.

1) MIC: MIC was proposed by Marshall in 1890 [24], which is formed by SMEs with strong regional ties, and they are mostly local enterprises. Over time, with the increased process of openness, their constituent firms tend to expand cooperation and develop industrial chains. MIC is not static entities. With the accumulation of knowledge and skilled talents, some traditional industrial clusters evolve towards innovative clusters for higher profits [82]. Characterized by small-sized firms, intense competition among firms, and a high degree of marketization, led to more efficient resource utilization [83]. However, enterprises within these clusters typically display limited R&D capabilities. Therefore, we identify MIC based on two key indicators: the number of patents, and the size of their employees. According to the criteria set by the National Bureau of Statistics of China, enterprises with fewer than 300 employees are SMEs [84]. Thus, the rules for identifying MIC is as follows:

$$C_o \in C_m, s.t. \frac{\sum_{j=n,A_j\in C_m}^{j} P(A_j^S)}{P(C_O)} \geq 0.8. \tag{5}$$

Where C_o is the object cluster, C_m is MIC, A_j^S is the patent applicant j who is a SME, $P(A_j^S)$ is the patent number of A_j^S, and $P(C_o)$ is the patent number of all applicants in C_o. The Pareto Principle also known as the 80/20 rule, it was propounded by Vilfredo Pareto when he observed that 20% of the people of Italy owned 80% of the wealth. The name "Pareto Principle" was suggested by American management consultant Joseph Juran [85], who recognized that Pareto's observation was a "universal" principle. For example, Koch [86] realizes that 80% of their sales come from 20% of their products, and 80% of their profits come from 20% of their customers. Therefore, based on the Pareto principle and with 80% as the dividing line, MIC is defined as a group of small and medium-sized enterprises whose patents account for more than 80% of all patents in the cluster.

2) HSIC: The HSIC model typically revolves around one or more large vertically integrated enterprises, which serve as the focal point of the cluster and often form the core of the regional economy. These enterprises cover the entire spectrum of the industrial chain, encompassing product research, manufacturing, and sales. They wield significant control over vital resources such as capital, technology, and services. Moreover, they play a pivotal role in guiding the development of other enterprises within the cluster [87]. In the HSIC model, numerous auxiliary enterprises, often SMEs, cluster around the production needs of these core enterprises. By participating in the supply chain of the core entities, SMEs can mitigate their individual weaknesses and concentrate on specialized areas, thereby benefiting from industry growth [88]. In turn, the core enterprises leverage the capabilities of these auxiliary firms to enhance their competencies, facilitating technological development and industrial upgrading collaboratively. Core enterprises within HSIC has significant advantages in terms of employee size and R&D. Accordingly, we use patent numbers and employee size as key indicators [84].

Therefore, the rules for identifying HSIC is as follows:

$$C_o \in C_h, s.t. \frac{\sum_{j=n,A_j\in C_h}^{j} P(A_j^L)}{P(C_o)} \geq 0.8. \quad (6)$$

Where C_h is HSIC, A_j^L is the patent applicant j who is a large-sized enterprise with more than 300 employees, $P(A_j^L)$ is the patent number of A_j^L.

3) SPIC: The SPIC is composed of branches of the parent company in other regions or overseas. Each branch maintains close contact with its parent company, suppliers, and customers outside the region, fostering a vertically integrated relationship among them. In terms of R&D, subsidiaries typically expand new technological fields based on the technology of the parent company, sharing patented innovations with the parent company. This dynamic often forms a technological monopoly within the cluster [89]. To define SPIC, we utilize government open data platforms to analyze the investment relationships among different enterprises with their parent companies within the cluster. Subsequently, we calculate the proportion of patents held by enterprises and their parent companies within the cluster as the indicator. The specific calculation rule is shown as follows:

$$C_o \in C_s, s.t. \frac{\sum_{j=n,A_j\in C_s}^{j} [P(A_j) + P(A_j^P)]}{[P(C_o) + \sum_{j=n,A_j\in C_s}^{j} P(A_j^P)} \geq 0.8. \quad (7)$$

Where C_s is SPIC, A_j is the patent applicant j in the cluster C_o, $P(A_j)$ is the patent number of A_j, A_j^P is the parent company of A_j , and $P(A_j^P)$ is the patent number of A_j^P. Through this formula, it can be seen that the patents of subsidiaries and their parent companies account for more than 80% of the cluster, which is the technical core of the cluster.

4) SAIC: SAIC is dominated by one or several large state-owned entities (e.g., military bases, defense factories, research institutes, laboratories, universities, and government offices) rather than private entities. State power represents the largest non-market entity in both the economy and society, possessing a binding force and redistributive capacity unparalleled by any other organization. Scholars like Cooke [90] and Aranguren [91] focused on the relationship between government behavior and cluster innovation, asserting that the advancement of regional innovation capacity within industrial clusters is intricately linked to the national innovation system, where the government plays a vital role. Therefore, those state-owned entities have the advantage in innovation. To identify SAIC, we examine both the number of patents and the nature of the enterprises involved. The identification rule is shown as follows:

$$C_o \in C_c, s.t. \frac{\sum_{j=n,A_j\in C_c}^{j} P(A_j^C)}{P(C_o)} \geq 0.8. \quad (8)$$

Where C_c is SAIC, A_j^C is the patent applicant j in the cluster C_c, $P(A_j^C)$ is the patent number of A_j^C, and $P(A_j^C)$ is the patent number of A_j^C. As for the background of patent application enterprises, they can also be queried through open data platforms.

C. Evaluation of Four Categories Industrial Clusters

In the development process of one or more industries, multiple spatial structure clusters often emerge simultaneously in various regions with different influences. Therefore, evaluating clusters with different spatial structures is very important for formulating cluster development strategies. While Wong et al. [92] solely employed patent data for cluster output evaluation, without considering the IIC. In addition, technological maturity is used as the only indicator, which is unable to capture the competitiveness of the entire cluster. To overcome the shortcomings mentioned above, leveraging patent data and employing rigorous econometric techniques, this article integrates four indicators average output, scale, efficiency, and number of clusters to evaluate spatial structures.

1) The average output is obtained by calculating the average number of patents across all clusters within the same spatial structure, which can analyze the technological achievements, especially innovative clusters.

2) The average size is determined by calculating the average number of applicants across all clusters within the same spatial structure, to analyze the size of such institutional clusters in terms of enterprises.

3) The average efficiency is assessed by calculating the average number of patents of applicants across all clusters within the same spatial structure, thus analyzing the research efficiency of enterprises.

4) The number of clusters refers to the number of all clusters within the same spatial structure and studies the popularity of different spatial structures throughout the industry.

IV. Examples of Research

A. Patent Data Processing

Flexible circuits (FE) is a new type of electronic technology that uses flexible materials such as plastics and fibers to manufacture electronic devices, achieving high flexibility in applications such as bending, folding, and compression of electronic devices. At present, compared to developed countries, China's FEI is still in early development stages [93]. It is necessary to analyze the distribution of China's existing FEI enterprises in order to provide reference for the formulation of subsequent industry support policies and enterprise development decisions. Due to the fact that FE belongs to the technology industry and has a large number of related patents, this article selects Chinese patents applied for the past 20 years

TABLE II
SEARCH FORMULA OF CHINA FE PATENT

Candidate	Search formula
FE	TACD:('Flexible electronic' OR 'flexible printed circuit' OR 'flex circuits' OR 'flexible hybrid electronics' OR 'flexible displays' OR 'flexible printed sensor' OR 'stretchable electronic') OR CPC:(G09F9/301 OR H01L51/0097 OR G06F1/1652 OR H01L2251/5338 OR H04M1/0268 OR G09G2380/02 OR G05B2219/25321 OR G05B2219/25439 OR H01H2229/038 OR H01R12/59 OR H01R12 /61 OR H01R12/77 OR H01R12/78 OR H05K1/028 OR H05K1/118 OR H05K1/147 OR H05K3/361 OR H05K2201/2009 OR H05K2201/046 OR H05K2201/2027 OR H05K3/4635 OR h05k2201/09445 OR g09g3/035 OR g02f1/133305 OR g06f2203/04102 OR g06f1/1616 OR h01f2017/006 OR h01h2001/5816 OR h01h2001/5827 OR h01g9/2095 OR H01L23/4985 OR G05B2219/23358 OR H05K2201/05 OR H05K1/148) AND APD:[20000101 TO 20201231]

as the research object. The search formula is implemented in Table II. Low quality patents that have been repeatedly applied, rejected, or withdrawn are excluded. Subsequently, a total of 860,701 patents were obtained, partially shown in Table III. The table includes the title, application number, applicant, applicant's address, longitude, and latitude.

As for the applicants, to improve the precision of the analysis results and avoid the influence of individual applicants who only apply for one patent, we choose to analyze the applicants who have applied for more than two patents, and a total of 5,610 applicants are detected. On this basis, Baidu map API (https://lbsyun.baidu.com/) is introduced to obtain the latitude and longitude information according to the applicant's name and address.

B. Cluster Recognition Based on KANNs-DBSCAN Clustering Algorithm

As stated in Section III, the selection of *Eps* and *MinPts* parameters significantly influences the clustering results of density-based algorithms. According to the research flowchart shown in Fig. 3, the KANNs-DBSCAN algorithm is used to cluster the patent applicants distributed across China. First, the relationship between the K value and the density threshold is determined, as shown in Fig. 4. Subsequently, the sensitivity is calculated, where $\triangle K$ is set 1. The density is stabilized at 41.5 where sensitivity falls below 0.1, it indicates that the former is not highly sensitive to K value ($K \geq 25$), as depicted in Fig. 5. Therefore, the KANND $D_K = 24$ corresponding to $K = 25$ is identified as the optimal *Eps* parameter, and then the optimal *MinPts* parameter is obtained using formula (4), yielding *Eps* = 0.4 and *MinPts* = 20.

The KANNs-DBSCAN algorithm is then used to assess the spatial distribution of FEI in China, and the results are shown in Fig. 6. Different colors are assigned to differentiate

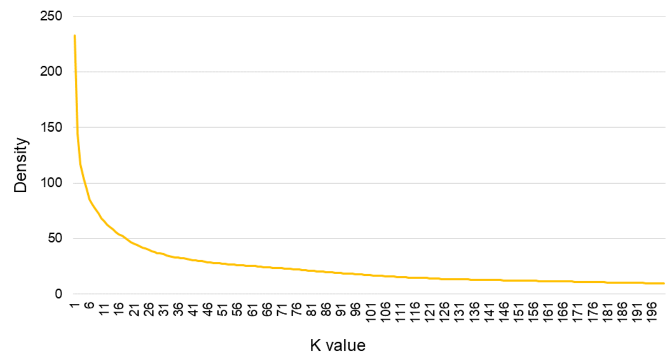


Fig. 4. Plot of the Density and K value.

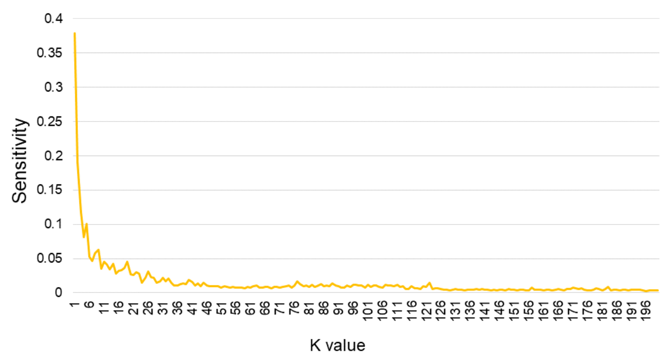


Fig. 5. Plot of the Sensitivity and K value.

between various clusters, resulting in a total of 24 clusters obtained based on the K value, as shown in Table IV. In practice, the size of these clusters varies greatly, ranging from 21 to 1,532 patent applicants in a single cluster. Notably, larger clusters in China's FEI are mainly located in the YRD, PRD, and BTH regions. In addition, these larger clusters are not only characterized in the scale, but also encompass multiple administrative regions. For example, the FEI in Shenzhen extends to neighboring cities such as Guangzhou and Dongguan, while the Shanghai cluster expands into Suzhou, further confirming the phenomenon of cross-region industrial clustering [94]. Consequently, development industrial clusters requires joint efforts across multiple administrative regions to effectively improve the competitiveness of the cluster. Particularly, the YRD cluster is highly concentrated in the core areas of Shanghai, South Jiangsu, and North Zhejiang, exhibiting distinct V-shaped distribution characteristics.

In order to verify the rationality of the clustering algorithm used in this article, several commonly used density-based clustering algorithms, such as Kernel DBSCAN and I-DBSCAN, were compared by the SC index. The results are shown in Table V.

The SC index integrally reflects the degree of compactness within clusters and the degree of separation between classes [95], and the SC value is ranging from -1 to 1. A higher SC value signifies better clustering results. As presented in Table IX, KANNs-DBSCAN algorithm yielded the highest SC value, which indicates that the elements within the clusters delineated by the algorithm are more compact, and the different clusters

TABLE III
INFORMATION OF PATENT APPLICANT

Application No.	Patent name	Applicant	Applicant address	Longitude and latitude
CN202011550473.0	A method of making a conductive circuit board	Shenzhen Bairou New Material Technology Co.	4F and 1F, Building A, Shuanghuan Industrial Zone, No.8 Baoqing Road, Baolong Community, Baolong Street, Longgang District, Shenzhen, Guangdong, China (Guangdong,Shenzhen,Longgang District).	118.09644, 24.48541
CN202010943557.4	The production method of circuit board with embedded conductive lines	Pengding Holdings (Shenzhen) Co.	Building A1 to Building A3, Peng Ding Park, Song Luo Road, Yan Luo Community, Yan Luo Street, Baoan District, Shenzhen City, Guangdong Province (with business premises in Building 1 to Building 3, No. 1 Yanshan Avenue, Yan Chuan Community, to engage in business activities) (Guangdong, Shenzhen, Baoan District).	113.86367, 22.79640
CN202010195388.0	A circuit board and its manufacturing method	Yancheng Wixin Electronics Co.	No.999, Yandu Road, Yandu District, Yancheng City, Jiangsu Province (Jiangsu,Yancheng,Yandu District).	120.18987, 33.34369
...	...	...	...	...
CN201410149990.5	Preparation of copper-zinc-tin-sulfur films on flexible substrates using magnetron sputtering	Guangdong University of Technology	No.100 Waihuan West Road, Guangzhou University City, Panyu District, Guangzhou City, Guangdong Province, China.	113.39960, 23.04570

TABLE IV
BASIC INFORMATION ON REGIONAL CLUSTERS

No.	Total number of patent applicants	Total number of patent applications
1	1109	9754
2	1532	15323
3	64	763
4	79	1102
5	101	853
6	121	1976
7	96	2917
8	21	595
9	419	10176
10	27	533
11	96	1512
12	27	417
13	49	472
14	24	104
15	53	702
16	30	276
17	28	118
18	49	504
19	68	464
20	26	293
21	28	317
22	25	790
23	22	181
24	22	165

TABLE V
SEGMENTATION RESULTS AND EVALUATION METRICS OF DIFFERENT CLUSTERING ALGORITHMS

Clustering algorithm	*Eps*	*MinPts*	SC
Kernel-DBSCAN	1.269	59	0.379
I-DBSCAN	0.138	6	0.208
KANNs-DBSCAN	0.4	20	0.391

TABLE VI
MIC CHARACTERS

No.	Total number of patent applicants	Total number of patent applications	Percentage of SMEs quantity	Percentage of patent quantity of SMEs
3	64	763	0.73	0.84
8	21	595	1	1
11	96	1512	0.68	0.9
16	30	276	0.67	0.83
17	28	118	0.93	0.93
18	49	504	0.78	0.81

are more dispersed from each other, indicating a more rational delineation compared to other algorithms.

C. Cluster Structure Identification Results

Using extraction rules, an analysis was conducted on 24 industrials clusters, incorporating different clustering classifications. The characteristics and spatial distribution of different cluster classifications are presented in Table VI-IX and Fig. 7-10. Notably, clusters 20, 21, and 23 exhibit a significant presence of patents from large enterprises, accounting for over 80% of the total patents, indicative of typical HSIC. Clusters 8, 9, 10, 15, 16, and 22 exhibit state-owned enterprises and research institute patents exceeding 80%, all of which are affiliated with SAIC. The proportion of small business patents in clusters 3, 8, 11, 16, 17, and 18 is over 80%, belonging to the MIC. Most of the companies in clusters 4, 12, and 19 belong to the same parent company in SPIC.

D. Cluster Evaluation and Development Recommendations

The evaluation results of the characteristics of different types of clusters in China's FEI are shown in Fig. 10. It can be seen that SAIC is superior to the other three types of clusters in terms of output, scale, efficiency and number of clusters. This indicates that the current technology R&D of China's

Fig. 6. IIC result of China's FE.

TABLE VII
HSIC CHARACTERS

No.	Total number of patent applicants	Total number of patent applications	Percentage of large enterprises	Percentage of patents by large enterprises
20	26	293	0.62	0.82
21	28	317	0.54	0.8
23	22	181	0.55	0.81

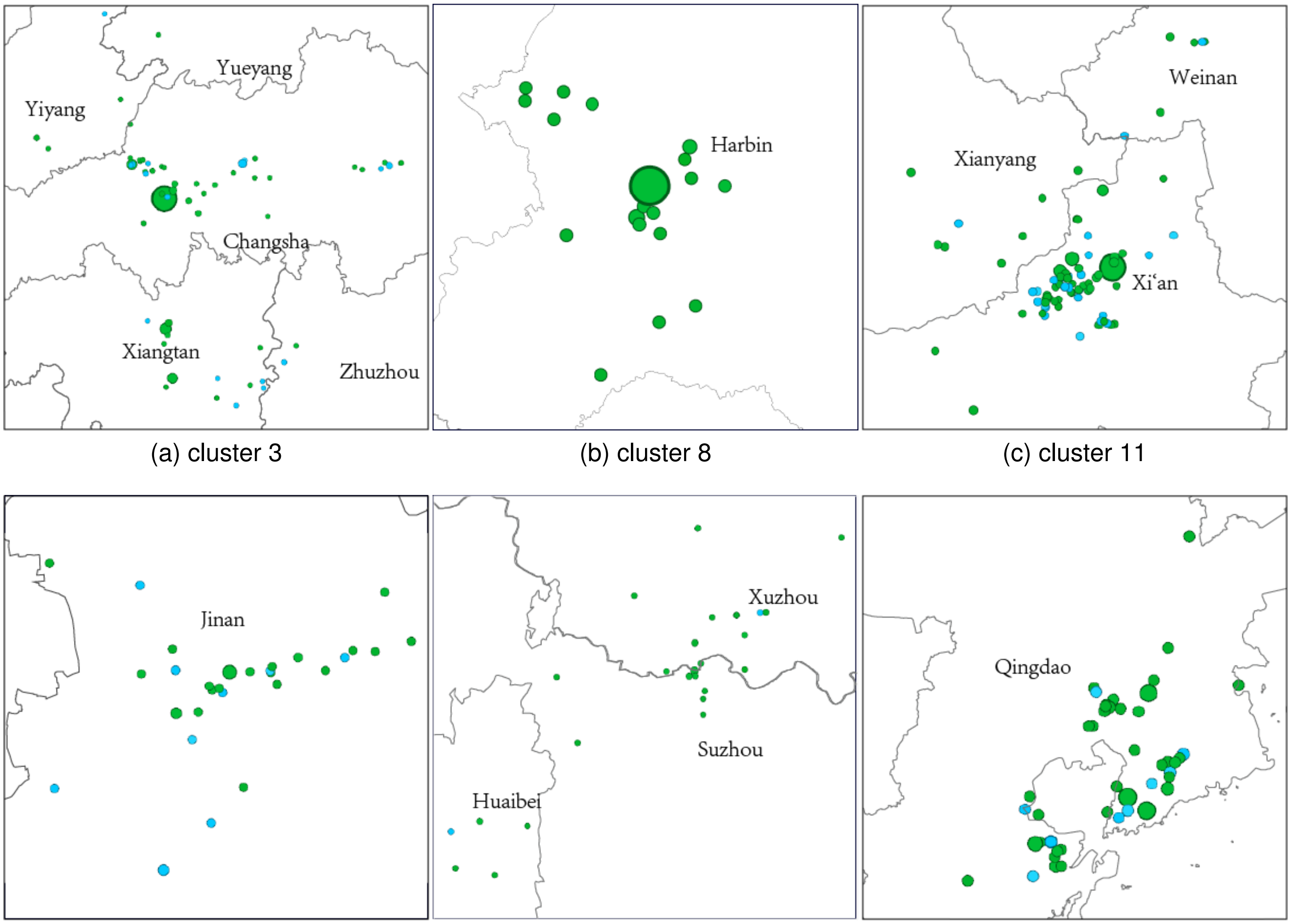
TABLE VIII
SPIC CHARACTERS

No.	Total number of patent applicants	Total number of patent applications	Percentage of subsidiaries	Percentage of total patents of subsidiary and parent companies
4	79	1102	0.1	0.92
12	27	417	0.07	0.87
19	68	464	0.06	0.93

TABLE IX
SAIC CHARACTERS

No.	Total number of patent applicants	Total number of patent applications	Percentage of state-owned enterprises	Percentage of patents by state-owned enterprises
8	21	595	0.48	0.93
9	419	10176	0.42	0.8
10	27	533	0.48	0.9
15	53	702	0.45	0.81
16	30	276	0.47	0.84
22	25	790	0.64	0.93

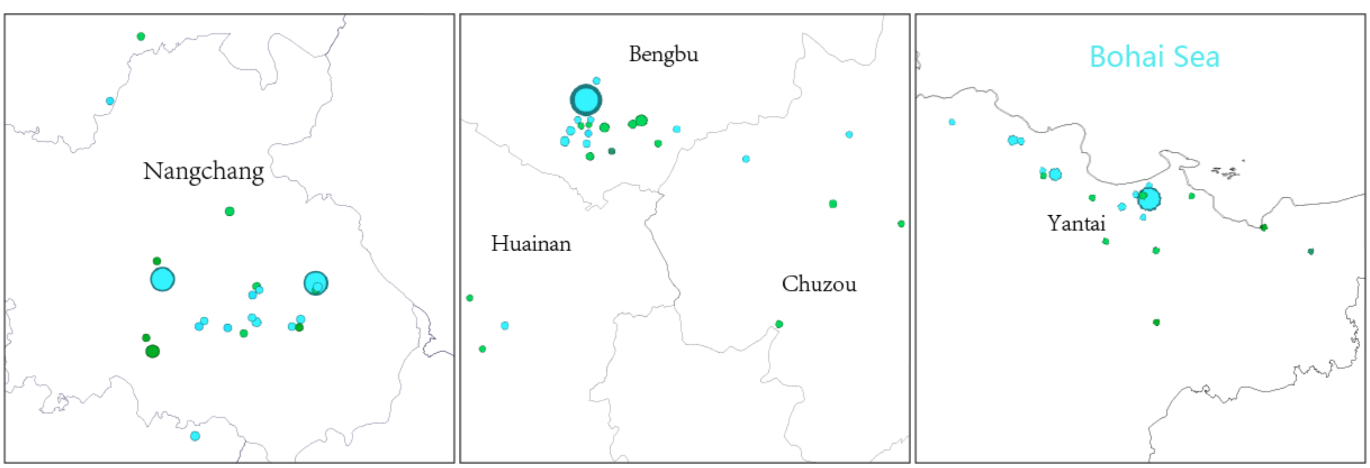
FEI mainly relies on government forces, including state-owned enterprises, universities and research institutes. The output value of the HSIC is significantly lower than that of the other types of industry clusters, and the number of HSIC is also smaller, only three. The main reason is that the R&D strength of core enterprises has not yet reached a high level. Although the number of clusters in MIC is higher, other indicators are relatively lower, indicating that although the cluster structure is more common, the R&D strength of the cluster is relatively low [66]. The main reason is that most of the clusters are small and SMEs with low R&D investment. The number of SPIC



(a) cluster 3 (b) cluster 8 (c) cluster 11

(d) cluster 16 (e) cluster 17 (f) cluster 18

Fig. 7. MIC identification result (blue for large enterprises, green for medium and small enterprises).



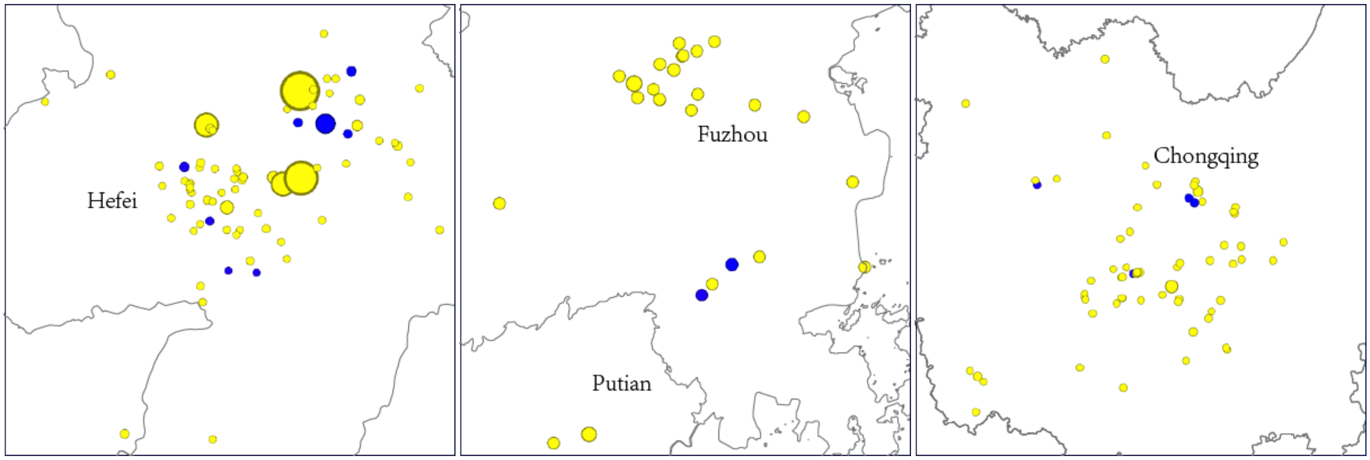
(a) cluster 20 (b) cluster 21 (c) cluster 23

Fig. 8. HSIC identification result (blue for large enterprises, green for medium and small enterprises).

is the same as that of the HSIC, while the output, scale and efficiency indicators rank in the middle of the four clusters. This suggests that the core enterprises are still predominantly local and rarely expand their industries to other regions.

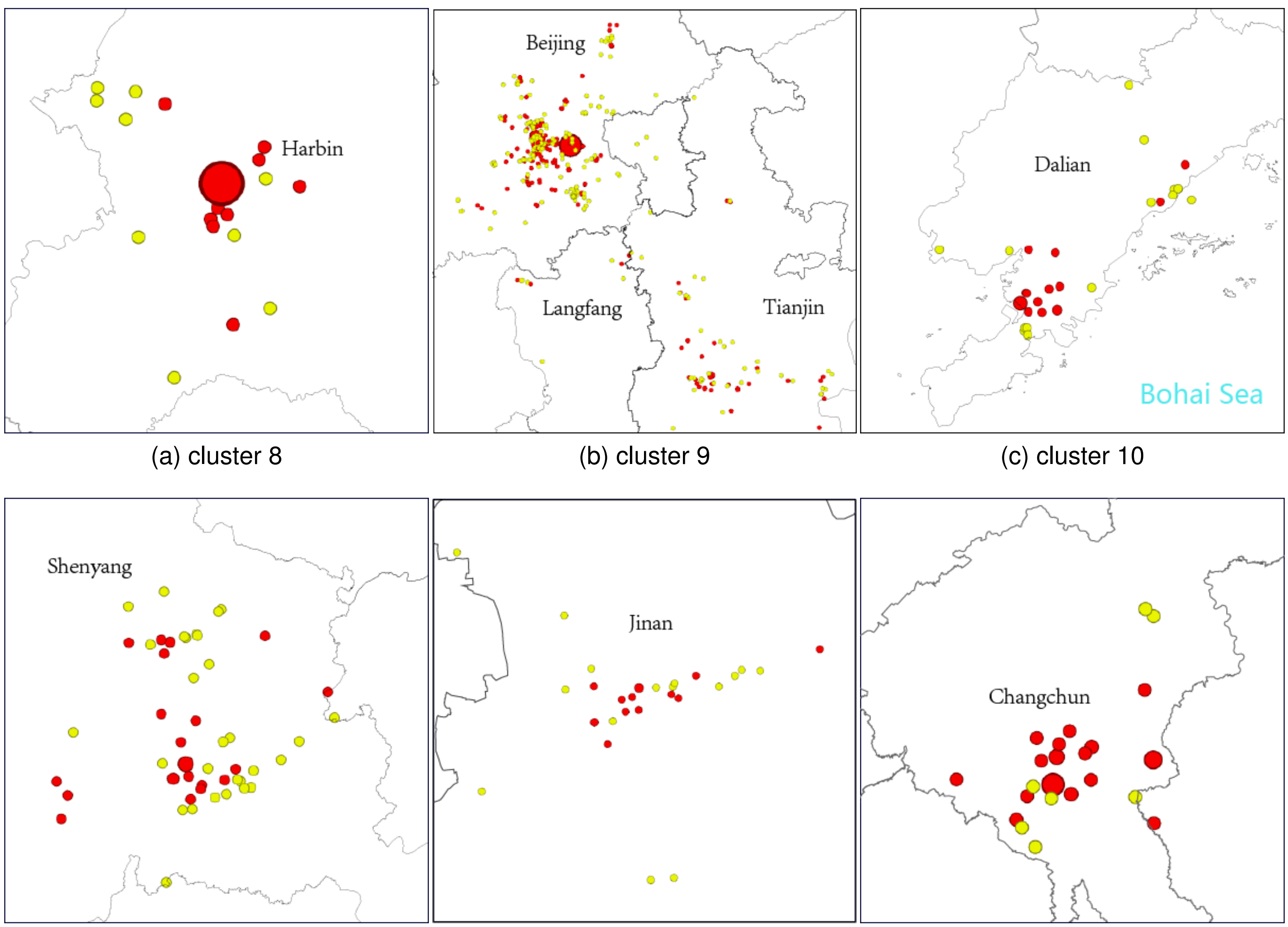
Overall, from the perspective of market economy, the FEI should still belong to the primary development stage in China. Constrained by the technological threshold, it still mainly relies on universities and research institutes with strong research

This article has been accepted for publication in IEEE Transactions on Engineering Management. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TEM.2024.3493627



(a) cluster 4 (b) cluster 12 (c) cluster 19

Fig. 9. SPIC identification result (blue for subsidiaries of external firms, yellow for local firms).



(a) cluster 8 (b) cluster 9 (c) cluster 10

(d) cluster 15 (e) cluster 16 (f) cluster 22

Fig. 10. SAIC identification result (red for state-owned enterprises, yellow for private enterprises).

strength to engage in R&D. Most of the non-state-owned enterprises have weak R&D strength, making government-led clusters the main form. However, these clusters rely too much on government power and are less able to resist market shocks than other clusters. In this regard, we give the following suggestions: (I) We suggest that the government should encourage cooperation between universities and institutes and enterprises, especially private enterprises, to accelerate the development

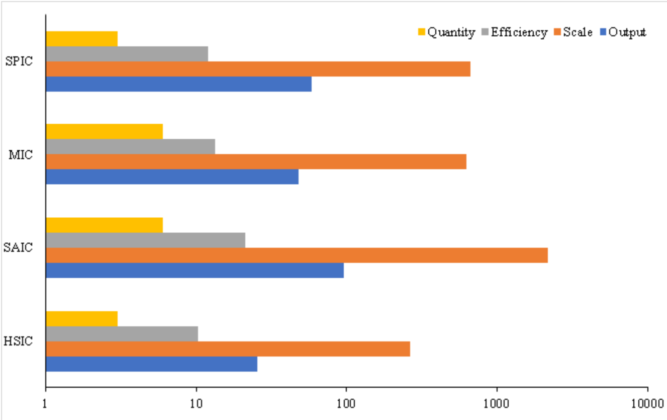


Fig. 11. Evaluation results of four cluster types in Figure.

of the industry through technological transformation, so as to improve the overall R&D strength, which will contribute to the development of the MICS. (II) For some large enterprises, cross-regional investment should be encouraged to expand the cluster scale geographically, which can not only drive the development of other regions' economy, but also expand the market scale of the enterprises themselves and promote the development of the SPICs. (III) Encouragement of cross-technology cooperation between large enterprises can realize the division of labor, so that the large enterprises can focus on the industry's core technology field and convert part of the competitive relationship into a cooperative relationship, which promotes the development of the SAICs [96].

V. Discussion and Conclusion

The development of industrial clusters is intricately linked to both market dynamics and government intervention, particularly in technology-intensive sectors such as FE, where knowledge and skilled employees play important roles. Accurate IIC and CIC are indispensable for government policymakers to devise effective cluster development strategies. Existing research on IIC has often been limited in scope, focusing solely on specific aspects or neglecting other key characteristics of industrial clusters.

This article addresses these limitations by introducing a novel approach to industrial cluster analysis using the KANNs-DBSCAN algorithm. This algorithm improves upon traditional DBSCAN by adapting to varying densities in datasets and automatically determining optimal clustering parameters based on the distance distribution characteristics of the input dataset. By applying this algorithm to cluster enterprise data according to their geographical distribution, this article effectively identified industrial clusters, providing a robust foundation for subsequent analysis.

Moreover, the CIC was enhanced. Based on Markusen's cluster research theory, this article uses enterprise patent data to classify industrial clusters for the first time. This approach is particularly suited to knowledge-intensive industries, which rely heavily on knowledge resources such as patents. By evaluating cluster types based on output, scale, efficiency, and quantity, this approach providing a novel framework for cluster classification. It also enables scholars to analysis industries, enterprises, and their characteristics in a more directly way.

Methodology employed in this article allows for monitoring the latest development of technology clusters, helping policymakers pinpoint innovation hotspots and understand cluster classification dynamics. Through a case study of the FE industry in China, our framework identified 24 clusters with growth potential, including six MICs, three HSICs, three SPICs, and six SAICs, offering valuable insights for policymakers shaping cluster development policies.

However, there are some limitations in this article. The IIC is a complex process that requires consideration of economic, political, environmental, and technical factors. Patent data alone offers a limited perspective, and integrating additional data sources is essential for comprehensive analysis. Moreover, as globalization integrates supply chains across regions and industries, patents increasingly reflect technological innovations and collaborative efforts within and across industrial clusters. Therefore, more research is necessary to understand the correlations that exist and how such attempts can be integrated to help governments and businesses better improve innovation and achieve better cluster development.

Acknowledgments

The authors would like to thank the anonymous reviewers for their constructive input on this article.

References

[1] E. Mantaeva, V. Goldenova and S. U. Kalmyk, "On the Role of Industrial Cluster in Regional Economy Development," *Vestnik Volgogradskogo gosudarstvennogo universiteta. Seriia3, Ekonomika, pravo*, vol. 144, no. 1, pp. 45-51, 2017. doi:10.15688/jvolsu3.2017.1.5.

[2] China, MIIT. (2022, Nov. 30). List of advanced manufacturing clusters in 45 countries. [Online]. Available: https://www.miit.gov.cn/jgsj/ghs/gzdt/art/2022/ art_fa5bd57e9f364b65ae48de37a319046f.html.

[3] Z. Yang, R. Sliuzas, J. Cai and H. F. L. Ottens, "Exploring spatial evolution of economic clusters: A case study of Beijing," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 19, pp. 252-265, 2012. doi:10.1016/j.jag.2012.05.017.

[4] T. Brenner, "Identification of Local Industrial Clusters in Germany," *Reg. Stud.*, vol. 40, no. 9, pp. 991-1004, 2006, doi: 10.1080/00343400601047408.

[5] M. E. Porter, *The competitive advantage of nations*,ed. New York: Free Press, 1990.

[6] J. V. Henderson, Z. Shalizi and A. J. Venables, "Geography and development," *J. Econ. Geogr.*, vol. 1, no. 1, pp. 81-105, 2001, doi: 10.1093/jeg/1.1.81.

[7] Y. Zhou, Z. Li, Y. Liu and F. Deng, "Network Proximity and Communities in Innovation Clusters Across Knowledge, Business, and Geography: Evidence From China," *IEEE Trans. Eng. Manage.*, vol. 68, no. 5, pp. 1388-1397, 2021, doi: 10.1109/TEM.2020.3032160.

[8] M. Delgado, M. E. Porter and S. Stern, "Defining clusters of related industries," *J. Econ. Geogr.*, vol. 16, no. 1, pp. 1-38, 2016, doi: 10.1093/jeg/lbv017.

[9] C. Liu and J. C. Yang, "Decoding Patent Information Using Patent Maps," *Data Sci. J.*, vol. 7, pp. 14-22, 2008, doi: 10.2481/dsj.7.14.

[10] X. Li, M. Fan, Y. Zhou, J. Fu, F. Yuan and L. Huang, "Monitoring and forecasting the development trends of nanogenerator technology using citation analysis and text mining," *Nano Energy*, vol. 71, pp. 104636, 2020, doi: 10.1016/j.nanoen.2020.104636.

[11] A. J. Trippe, "Patinformatics: Tasks to Tools," *World Pat. Inf.*, vol. 25, no. 3, pp. 211-221, 2003, doi:10.1016/S0172-2190(03)00079-6.

[12] B. J. Fifarek and F. M. Veloso, "Offshoring and the global geography of innovation," *J. Econ. Geogr.*, vol. 10, no. 4, pp. 559-578, 2010, doi: 10.1093/jeg/lbq013.

[13] L. Cassi and A. Plunket, "Research Collaboration in Co-inventor Networks: Combining Closure, Bridging and Proximities," *Reg. Stud.*, vol. 49, no. 6, pp. 936-954, 2015, doi: 10.1080/00343404.2013.816412.

[14] E. Miguelez and R. Moreno, "Relatedness, external linkages and regional innovation in Europe," *Reg. Stud.*, vol. 52, no. 5, pp. 688-701, 2018, doi: 10.1080/00343404.2017.1360478.

[15] J. Hagedoorn and M. Cloodt, "Measuring innovative performance: is there an advantage in using multiple indicators?" *Res. Policy*, vol. 32, no. 8, pp. 1365-1379, 2003, doi: 10.1016/S0048-7333(02)00137-3.

[16] J. O. Lanjouw and M. Schankerman, "Patent Quality and Research Productivity: Measuring Innovation with Multiple Indicators," *The Economic journal (London)*, vol. 114, no. 495, pp. 441-465, 2004, doi: 10.1111/j.1468-0297.2004.00216.x.

[17] J. Alcácer and M. Zhao, "Zooming in: A practical manual for identifying geographic clusters," *Strateg. Manage. J.*, vol. 37, no. 1, pp. 10-21, 2016, doi: 10.1002/smj.2451.

[18] L. Anselin and X. Li, "Tobler's Law in a Multivariate World," *Geogr. Anal.*, vol. 52, no. 4, pp. 494-510, 2020, doi: 10.1111/gean.12237.

[19] K. E. Maskus, S. Milani and R. Neumann, "The impact of patent protection and financial development on industrial R&D," *Res. Policy*, vol. 48, no. 1, pp. Feb.

[20] Y. Li, Y. Wei, Y. Li, Z. Lei and A. Ceriani, "Connecting emerging industry and regional innovation system: Linkages, effect and paradigm in China," *Technovation*, vol. 111, Mar. 2022, Art. no. 102388.

[21] W. Liu, Y. Tao and K. Bi, "Capturing information on global knowledge flows from patent transfers: An empirical study using USPTO patents," *Res. Policy*, vol. 51, no. 5, Jun. 2022, Art. no. 104509.

[22] R. Voyer, "Knowledge-Based Industrial Clustering: International Comparisons," *Springer US*, 1998.

[23] M. Komorowski, "Identifying industry clusters: a critical analysis of the most commonly used methods," *Regional studies, regional science*, vol. 7, no. 1, pp. 92-100, 2020, doi: 10.1080/21681376.2020.1733436.

[24] A. Marshall, "Principles of economics: an introductory volume,", *London: Macmillan*, 1920.

[25] G. Ellison and E. L. Glaeser, "Geographic Concentration in U. S. Manufacturing Industries: A Dartboard Approach," *The Journal of political economy*, vol. 105, no. 5, pp. 889, 1997.

[26] M. Fritsch and H. Graf, "How sub-national conditions affect regional innovation systems: The case of the two Germanys," *Pap. Reg. Sci.*, vol. 90, no. 2, pp. 331-353, 2011, doi: 10.1111/j.1435-5957.2011.00364.x.

[27] P. Sharma, S. B. S. Nookala and A. Sharma, "India's National and Regional Innovation Systems: Challenges, Opportunities and Recommendations for Policy Makers," *Ind. Innov.*, vol. 19, no. 6, pp. 517-537, 2012, doi: 10.1080/13662716.2012.718878.

[28] E. Turkina, A. Van Assche and R. Kali, "Structure and evolution of global cluster networks: evidence from the aerospace industry," *J. Econ. Geogr.*, vol. 16, no. 6, pp. 1211-1234, 2016. doi:10.1093/jeg/lbw020.

[29] P. Krugman, "Increasing Returns and Economic Geography," *The Journal of political economy*, vol. 99, no. 3, pp. 483-499, 1991, doi: 10.1086/261763.

[30] G. T. R. Lin and C. Sun, "Driving industrial clusters to be nationally competitive," *Technol. Anal. Strateg. Manage.*, vol. 22, no. 1, pp. 81-97, 2010, doi: 10.1080/09537320903438153.

[31] W. Zhao and M. Y. Toh, "Impact of Innovative City Pilot Policy on Industrial Structure Upgrading in China," *Sustainability*, vol. 15, no. 9, 2023, doi: 10.3390/su15097377.

[32] B. A. Bittencourt, A. C. Zen, F. Prevot and V. K. Schmidt, "How to Be More Innovative in Clusters? The Influence of Geographical Agglomerations on Its Firms," *J. Knowl. Econ.*, vol. 14, no. 3, pp. 2603-2629, 2023, doi: 10.1007/s13132-022-00975-2.

[33] G. M. Spencer, T. Vinodrai, M. S. Gertler and D. A. Wolfe, "Do Clusters Make a Difference? Defining and Assessing their Economic Performance," *Reg. Stud.*, vol. 44, no. 6, pp. 697-715, 2010, doi: 10.1080/00343400903107736.

[34] S. Czamanski and L. A. D. Ablas, "Identification of Industrial Clusters and Complexes: a Comparison of Methods and Findings," *Urban studies (Edinburgh, Scotland)*, vol. 16, no. 1, pp. 61-80, 1979, doi: 10.1080/713702464.

[35] A. Lagendijk, "The dynamics of industrial clustering. International comparisons in computing and biotechnology," *Res. Policy*, vol. 30, no. 2, pp. 353-354, 2001, doi: 10.1016/S0048-7333(99)00117-1.

[36] M. E. Porter and M. P. Porter, "Location, Clusters, and the 'New' Microeconomics of Competition," *Business economics (Cleveland, Ohio)*, vol. 33, no. 1, pp. 7-13, 1998.

[37] Z. Liu, X. Chen, W. Xu, Y. Chen and X. Li, "Detecting industry clusters from the bottom up based on co-location patterns mining: A case study in Dongguan, China," *Env. Plan. B-Urban Anal. City Sci.*, vol. 48, no. 9, pp. 2827-2841, 2021, doi: 10.1177/2399808321991542.

[38] P. Maskell and A. Malmberg, "The elusive concept of localization economies: towards a knowledge-based theory of spatial clustering," *Environment and planning. A*, vol. 34, no. 3, pp. 429-449, 2002, doi: 10.1068/a3457.

[39] B. Nooteboom, "Innovation, learning and cluster dynamics," *Erim Report*, pp. 155-181, 2004.

[40] Z. W. Zhao, Z. Q. Zhao and P. Zhang, "A new method for identifying industrial clustering using the standard deviational ellipse," *Sci Rep*, vol. 13, no. 1, 2023, doi: 10.1038/s41598-023-27655-8.

[41] R. Fotso, "Evaluating the indirect effects of cluster-based innovation policies: the case of the Technological Research Institutes in France," *J. Technol. Transf.*, vol. 47, no. 4, pp. 1070-1114, 2022. doi:10.1007/s10961-021-09865-2.

[42] People's Government of Yunnan Province, (2024, Jan. 24). 53000000102/202200011, *Notice of the People's Government of Yunnan Province on Issuing Several Policy Measures to Promote the High Quality Development of Key Industrial Parks in Yunnan Province*. [Online]. Available: https://www.yn.gov.cn/zwgk/zcwj/zxwj/202201/t20220128_235420.html

[43] People's Government of Ningbo Province, (2024, Jan. 16). 11330208591592676Q/2024-27585, *Ningbo Hi-Tech Zone Management Committee Issues Opinions on Several Policies on Striving to Become a World-Class Hi-Tech Park*. [Online]. Available: http://www.nbhtz.gov.cn/art/2024/1/16/art_1229026414_1784786.html

[44] A. S. Sarah, "Austin tops list for headquarters relocations in US, report says," 2024. [Online]. Available: https://www.kxan.com/news/local/austin/austin-tops-list-for-headquarters-relocations-in-us-report-says/

[45] S. S. Moral, "Industrial Clusters and New Firm Creation in the Manufacturing Sector of Madrid's Metropolitan Region," *Reg. Stud.*, vol. 43, no. 7, pp. 949-965, 2009, doi: 10.1080/00343400801932326.

[46] S. A. Rosenfeld, "Bringing business clusters into the mainstream of economic development," *EUR. PLAN. STUD.*, vol. 5, no. 1, pp. 2-23, 1997, doi:10.1080/09654319708720381.

[47] F. Benita, S. Sarica and G. Bansal, "Testing the static and dynamic performance of statistical methods for the detection of national industrial clusters," *Pap. Reg. Sci.*, vol. 99, no. 4, pp. 1137-1157, 2020, doi: 10.1111/pirs.12517.

[48] M. Titze, M. Brachert and A. Kubis, "Actors and InteractionsIdentifying the Role of Industrial Clusters for Regional Production and Knowledge Generation Activities," *Growth Change*, vol. 45, no. 2, pp. 163-190, 2014, doi: 10.1111/grow.12047.

[49] L. Jing and M. Guo, "Weighted DBSCAN Algorithm for Discovery of Spatial Features in Industries Management, *Proc. 2021 IEEE 4th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)*, 2021, pp. 383-386.

[50] Y. Xu, Z. Zhuang, W. Li and X. Zhou, "Effective community division based on improved spectral clustering," *Neurocomputing*, vol. 279, pp. 54-62, 2018, doi: 10.1016/j.neucom.2017.06.085.

[51] C. Fang and H. Liu, "Research and Application of Improved Clustering Algorithm in Retail Customer Classification," *Symmetry-Basel*, vol. 13, no. 10, pp. 1789, 2021, doi: 10.3390/sym13101789.

[52] Z. Chen and D. Zhu, "Hierarchic Clustering Algorithm Used for Anomaly Detecting," *Procedia engineering*, vol. 15, pp. 3401-3405, 2011, doi: 10.1016/j.proeng.2011.08.637.

[53] C. Bouveyron and C. Brunet-Saumard, "Model-based clustering of high-dimensional data: A review," *Comput. Stat. Data Anal.*, vol. 71, pp. 52-78, 2014, doi: 10.1016/j.csda.2012.12.008.

[54] M. Ester, H. P. Kriegel, J. Sander and X. Xu, *A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise*, ed. Portland, Oregon, USA: The AAAI Press, 1996.

[55] M. A. Mody, C. Suess and X. Lehto, "The accommodation experiencescape: a comparative assessment of hotels and Airbnb," *Int. J. Contemp. Hosp. Manag.*, vol. 29, no. 9, pp. 2377-2404, 2017, doi: 10.1108/IJCHM-09-2016-0501.

[56] L.J. Liang, H. S. C. Choi and M. Joppe, "Understanding repurchase intention of Airbnb consumers: perceived authenticity , electronic word-of-mouth, and price sensitivity," *J. Travel Tour. Mark.*, vol. 35, no. 1, pp. 73-89, 2018, doi: 10.1080/10548408.2016.1224750.

[57] C. H. Jin, H. J. Na, M. Piao, G. Pok and K. H. Ryu, "A Novel DBSCAN-based Defect Pattern Detection and Classification Framework for Wafer Bin Map," *IEEE Trans. Semicond. Manuf.*, vol. 32, no. 3, pp. 286-292, 2019. doi:10.1109/TSM.2019.2916835.

[58] S. H. Yue, P. LI, J. D. Guo and S. G. Zhou, "A statistical information-based clustering approach in distance space," *Journal of Zhejiang University. A. Science*, vol. 6, no. 1, pp. 71-78, 2005.

[59] S. Jahirabadkar and P. Kulkarni, “Algorithm to determine Îţ-distance parameter in density based clustering,” *Expert Syst. Appl.*, vol. 41, no. 6, pp. 2939-2946, 2014, doi: 10.1016/j.eswa.2013.10.025.

[60] H. H. Yang, “Serial updating rule for blind separation derived from the method of scoring,” *IEEE Trans. Signal Process.*, vol. 47, no. 8, pp. 2279-2285, 1999, doi: 10.1109/78.774771.

[61] J. Song, Y. Guo and B. Wang, “The Parameter Configuration Method of DBSCAN Clustering Algorithm, ” *Proc.IEEE*, pp. 1062-1070, 2018.

[62] Z. Zhou, J. Wang, S. Zhu and Z. Sun, “An improved adaptive and fast AF-DBSCAN clustering algorithm,” *Zhineng Xitong Xuebao (CAAI Transactions on Intelligent Systems)*, vol. 11, no. 1, pp. 93-98, 2016.

[63] J. Kim, J. Choi, K. Yoo and A. Nasridinov, “AA-DBSCAN: an approximate adaptive DBSCAN for finding clusters with varying densities,” *The Journal of supercomputing*, vol. 75, no. 1, pp. 142-169, 2019, doi: 10.1007/s11227-018-2380-z.

[64] M. M. R. Khan, M. A. B. Siddique, R. B. Arif and M. R. Oishe, “ADBSCAN: Adaptive Density-Based Spatial Clustering of Applications with Noise for Identifying Clusters with Varying Densities,” *Proc.IEEE*, 2018, pp. 107-111.

[65] R. J. G. B. Campello, D. Moulavi and J. Sander, “Density-Based Clustering Based on Hierarchical Density Estimates,” *Berlin, Heidelberg: Springer Berlin Heidelberg*, pp. 160-172.

[66] C. H. A. Logan and S. Fotopoulou, “Unsupervised star, galaxy, QSO classification: Application of HDBSCAN”. *Astron. Astrophys.*, vol. 633, 2020. doi:10.1051/0004-6361/201936648.

[67] W. J. Li, S. Q. Yan, Y. Jiang, S. Z. Zhang and L. C. Wang, “Research on method of self-adaptive determination of DBSCAN algorithm parameters”. *Computer Engineering and Applications*, vol. 55, no. 5, pp. 1-7, 2019, DOI: 10.3778/j.issn.1002-8331.1809-0018.

[68] A. Sharma and A. Sharma, “KNN-DBSCAN: Using k-nearest neighbor information for parameter-free density based clustering,” in *2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT)*, 2018.

[69] J. Song, X. Tong, L. Wang, C. Zhao and A. V. Prishchepov, “Monitoring finer-scale population density in urban functional zones: A remote sensing data fusion approach,” *Landsc. Urban Plan.*, vol. 190, pp. 103580, 2019, doi: 10.1016/j.landurbplan.2019.05.011.

[70] P. E. Stek, “Identifying spatial technology clusters from patenting concentrations using heat map kernel density estimation,” *Scientometrics*, vol. 126, no. 2, pp. 911-930, 2021, doi: 10.1007/s11192-020-03751-8.

[71] Y. Xing and P. Brimblecombe, “Urban Park layout and exposure to traffic-derived air pollutants,” *Landsc. Urban Plan.*, vol. 194, 2020, doi: 10.1016/j.landurbplan.2019.103682.

[72] S. Chen, X. Yan and B. Yang, “Move to success? Headquarters relocation, political favoritism, and corporate performance,” *Journal of corporate finance (Amsterdam, Netherlands)*, vol. 64, p. 101698, 2020. doi:10.1016/j.jcorpfin.2020.101698.

[73] R. Crescenzi and A. Jaax, “Innovation in Russia: The Territorial Dimension,” *Econ. Geogr.*, vol. 93, no. 1, pp. 66-88, 2017. doi:10.1080/00130095.2016.1208532.

[74] R. Zhao, L. Zhao, S. Deng and N. Zheng, “Trends in patents for solar thermal utilization in China,” *Renew. Sust. Energ. Rev.*, vol. 52, pp. 852-862, 2015. doi:10.1016/j.rser.2015.07.181.

[75] Y. Zhou, X. Li, R. Lema and F. Urban, “Comparing the knowledge bases of wind turbine firms in Asia and Europe: Patent trajectories, networks, and globalisation,” *Science & public policy*, vol. 43, no. 4, pp. 476-491, 2016. doi:10.1093/scipol/scv055.

[76] A. Markusen, “Sticky Places in Slippery Space: A Typology of Industrial Districts,” *Econ. Geogr.*, vol. 72, no. 3, pp. 293-313, 1996, doi: 10.2307/144402.

[77] J. Shen, X. Hao, Z. Liang, Y. Liu, W. Wang and L. Shao, “Real-Time Superpixel Segmentation by DBSCAN Clustering Algorithm,” *IEEE Trans. Image Process.*, vol. 25, no. 12, pp. 5933-5942, 2016, doi: 10.1109/TIP.2016.2616302.

[78] M. R. Hosseini, I. Martek, E. K. Zavadskas, A. A. Aibinu, M. Arashpour and N. Chileshe, “Critical evaluation of off-site construction research: A Scientometric analysis,” *Autom. Constr.*, vol. 87, pp. 235-247, 2018, doi: 10.1016/j.autcon.2017.12.002.

[79] W. W. Hood and C. S. Wilson, “The literature of bibliometrics, scientometrics, and informetrics,” *Scientometrics*, vol. 52, no. 2, pp. 291-314, 2001, doi: 10.1023/A:1017919924342.

[80] A. Darko, A. P. C. Chan, X. Huo and D. Owusu-Manu, “A scientometric analysis and visualization of global green building research,” *Build. Environ.*, vol. 149, pp. 501-511, 2019, doi: 10.1016/j.buildenv.2018.12.059.

[81] F. M. Scherer, “The Propensity to Patent,” *Int. J. Ind. Organ.*, vol. 1, no. 1, pp. 107-128, 1983, doi: 10.1016/0167-7187(83)90026-7.

[82] J. Hervás-Oliver, J. Albors-Garrigos, S. Estelles-Miguel and C. Boronat-Moll, “Radical innovation in Marshallian industrial districts,” *Reg. Stud.*, vol. 52, no. 10, pp. 1388-1397, 2018, doi: 10.1080/00343404.2017.1390311.

[83] R. Boix and V. Galletto, “Innovation and Industrial Districts: A First Approach to the Measurement and Determinants of the I-District Effect,” *Reg. Stud.*, vol. 43, no. 9, pp. 1117-1133, 2009, doi: 10.1080/00343400801932342.

[84] China, National Bureau of Statistics. (2018, Jan. 03). Statistical classification of micro, small, medium and large enterprises. [Online]. Available: http:// www.stats.gov.cn/ sj/tjbz/gjtjbz/202302/t20230213_1902763.html.

[85] J. M. Juran and A. B. Godfrey, “Juran’s quality handbook,”, *New York: McGraw Hill*, 2000.

[86] R. Koch, *The 80/20 Principle: The Secret of Achieving More with Less,* Beijing: CITIC Publishing Group, 2014.

[87] S. Mudambi, “Branding importance in business-to-business markets,” *Ind. Mark. Manage.*, vol. 31, no. 6, pp. 525-533, 2002, doi: 10.1016/S0019-8501(02)00184-0.

[88] H. Schmitz, “Does Local Cooperation Matter Evidence from Industrial Clusters in South Asia and Lat in America,” *Oxford Development Studies*, vol. 28, no. 3, pp. 786-815, 2000.

[89] S. O. Park and A. Markusen, “Generalizing new industrial districts: a theoretical agenda and an application from a non-Western economy,” *Environment and planning. A*, vol. 27, no. 1, pp. 81-104, 1995, doi: 10.1068/a270081.

[90] P. Cooke, “Regional innovation systems: development opportunities from the ‘green turn’,” *Technol. Anal. Strateg. Manage.*, vol. 22, no. 7, pp. 831-844, 2010, doi: 10.1080/09537325.2010.511156.

[91] M. J. Aranguren and M. Larrea, “Regional Innovation Policy Processes: Linking Learning to Action,” *J. Knowl. Econ.*, vol. 2, no. 4, pp. 569-585, 2011, doi: 10.1007/s13132-011-0068-1.

[92] C. Wong, J. Sheu and K. Lee, “Assessing the quest of SMEs in pivoting for new technological ventures: comparing the patenting indexes of seven developed cities,” *Scientometrics*, vol. 128, no. 7, pp. 4029-4064, 2023, doi: 10.1007/s11192-023-04729-y.

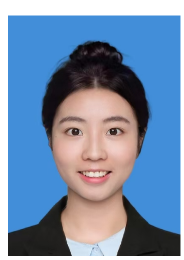
[93] N. Kshetri, “The Economics of Chip War: China’s Struggle to Develop the Semiconductor Industry,” *Computer (Long Beach, Calif.)*, vol. 56, no. 6, pp. 101-106, 2023, doi: 10.1109/MC.2023.3263267.

[94] C. Liu, Z. Peng, L. Liu and S. Li, “Innovation Networks of Science and Technology Firms: Evidence from China,” *Land*, vol. 12, no. 7, 2023, doi: 10.3390/land12071283.

[95] S. Modak, “Finding groups in data: an introduction to cluster analysis,” *J. Appl. Stat.*, 2023, doi: 10.1080/02664763.2023.2220087.

[96] R. Xiao, "Four development stages of collective intelligence," *Front. Inform. Technol. Elect. Eng.*,vol. 25, no. 7, pp. 903-916, 2024, doi: 10.1631/FITEE.2300459.

Wenguang Lin received his Ph.D. degree from the Department of Electrical and Mechanical Engineering, Xiamen University, China, in 2014. He is currently an Associate Professor with the School of Mechanical and Automotive Engineering, Xiamen University of Technology, Xiamen, China. His research interests include data-driven product design, patent knowledge mining, and complex network.



Ting Wang received her B.E. degree from Chongqing Institute of Engineering, China, in 2022. She is currently working toward the M.S. degree in the School of Mechanical and Automotive Engineering, Xiamen University of Technology, Xiamen, China. Her research interests include big data mining and machine learning applications in industry.

Zhizhen Chen received his Ph.D. degree in Finance from University of Glasgow, UK, in 2018. He is currently a Senior Lecturer of the Business School at the University of Greenwich, London, UK. His research interests include quantitative finance, banking, risk management, and machine learning application in finance.

Renbin Xiao received his Ph.D. degree in systems engineering from Huazhong University of Science and Technology, Wuhan, China, in 1993. He is currently a Professor with the School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan, China. His research interests include massive personalized customization, management systems engineering, collective intelligence, complex systems and complexity science. He has published more than 400 journal papers and more than 10 academic books.