Evaluating Global Intelligence Innovation: An Index Based on Machine Learning Methods

Abstract: This study investigates national intelligence innovation through machine learning methods. We propose a global intelligence innovation index (GIII) to evaluate the global landscape of intelligence innovation of 101 countries around the world. First, we develop a conceptual framework of national intelligence innovation based on the innovation ecosystem theory. Second, we design a machine learning method based on the unsupervised clustering-based random forest model to measure GIII. Finally, we evaluate the national intelligence innovation using GIII and provide theoretical and practical insights. The results show that global intelligence innovation development presents a convoluted situation, as high income doesn't necessarily promote intelligence innovation. Furthermore, intelligence innovation shows interesting relationships with unemployment, aging, and shares of economic sectors. GIII provides a reference to the level of intelligence innovation in various countries around the world and helps decision-makers better formulate policies to facilitate intelligence innovation development.

Keywords: Intelligence innovation; Comprehensive evaluation index system; Machine Learning

1. Introduction

The world is currently undergoing a new round of technological revolution spurred by frontier technologies such as artificial intelligence (AI), Internet of Things, cloud computing, big data, blockchain, 5G technology, etc. (Sirimanne et al., 2019). Such technological advancements have inspired profound changes in commercial operations, economic development, social activities and government practices (Wang et al., 2020). These outcomes have contributed to a newly emerged phenomenon of the "smart economy," which adopts new innovations and entrepreneurial initiatives to increase productivity and competitiveness with the overall goal of improving the quality of life of all citizens (Frank and Aznar Fernández-Montesinos, 2020). Countries around the world are seeking to capture the value created by the smart economy. For example, since 2016, the federal government of Germany has released a series of national strategy reports that outline roadmaps for several key areas in order to create a smart economy, including AI (Federal Government, 2020), blockchain (Federal Ministry for Economic Affairs and Energy, 2019), 5G (Federal Ministry of Transport and Digital Infrastructure, 2017), and data (Federal Ministry of the Interior, 2021). In the U.S., the House of Representatives passed the National AI Initiative Act of 2020, aiming to accelerate AI research and application to enhance the country's economic prosperity and national security (Thornberry, 2020).

Creating a smart economy is inseparable from intelligence innovation, which is an emerging innovation with development of the frontier technologies, such as artificial intelligence (AI), the Internet of Things, cloud computing, big data, blockchain, and 5G technology. Innovation is commonly defined as a value creation process from a commercial perspective (Schumpeter, 1983; van de Ven, 1986; Porter, 1990; Roper et al., 2008). We intend to define intelligence innovation as a value creation process for smart economy by empowering multiple aspects of society. The empowerment is achieved through both the industrialization of frontier technologies and the improvement of existing industries using frontier technologies. This concept extends the traditional perspective of commerce and incorporates multiple fields empowered by intelligence innovation, including commercial operations, economic development, social activities and government practices. Intelligence innovation works as an important indicator for evaluating the national level competitiveness and the development of smart economy. Therefore, the evaluation of the country-level performance in intelligence innovation is an important and meaningful research problem. Attempts to answer this question can provide useful insights for decision-makers to understand the worldwide intelligence innovation landscape and their own competitive position more clearly, thus empowering them to formulate more sound domestic policy.

However, existing studies on intelligence innovation evaluation either are limited to regional evaluation of intelligence innovation or focus on a single category of frontier technologies. Therefore, this paper aims to develop the Global Intelligence Innovation Index (GIII) to measure the worldwide level of intelligence innovation against the background of "smart economy" while incorporating the frontier technologies of artificial intelligence (AI), the Internet of Things, cloud computing, big data, blockchain, and 5G technology. After constructing the index system, a novel weight calculation method is proposed, and the index is tested on 101 countries. Finally, an overall generalization of the global intelligence innovation landscape and policy suggestions are presented.

This paper also introduces the machine learning methodology into the field of index construction and comprehensive evaluation, particularly in the weight calculation step. The supervised machine learning model has been increasingly useful in providing an accurate prediction of a given input by training on sample instances labelled with desirable outputs, also known as the target variable. Its function is in line with the objective of a composite index—synthesizing features of a given instance and making predictions on its performance, where the feature importance given by the model is similar to the weight of an index. Recent studies have applied supervised machine learning models for comprehensive evaluation. For example, Yan et al.(2019) adopted several machine learning models to assess water resources in China's transboundary river basins by predicting the runoff coefficient based on several influence factors and derived their importance. However, few studies have applied machine learning models for index weighting construction. In our study, the composite index, being the target variable of the dataset, is previously unknown. Thus, our dataset cannot be used directly to train a supervised machine learning model. We use an unsupervised machine learning model to find the target variable, then a supervised machine learning model is trained on the combined dataset with target variable to yield the feature importance.

More specifically, a machine learning-based method is adopted for constructing GIII, which uses k-means unsupervised clustering model with the random forest model to determine index weights. K-means has been used for clustering unlabeled data in areas including innovation assessment. Wang et al. (2015) used k-means to categorize the regional innovation systems in China. Tortorella et al. (2020) examined the effect of contingency factors on the adoption and associated barriers of Healthcare 4.0 technologies using k-means. We use the random forest model to obtain the feature importance of the k-means clustering method. The random forest model is an attractive methodology for producing an accurate classifier and providing insight regarding the discriminative ability of individual predictor variables (Archer and Kimes, 2008). Finally, the feature importance obtained by the random forest model is used as weights for aggregating GIII.

Our main contributions are summarized as follows:

- **We propose a comprehensive evaluation index to measure national intelligence innovation.** As the world moves into the smart era, measuring intelligence innovation level could provide vital insights for decision-makers. However, few studies have focused on the topic with fewer studies employing objective methods to measure the index. Based on the conceptual framework of national intelligence innovation, we construct GIII. It is a detailed-granularity index system consisting of 28 indicators, dedicated to measuring intelligence innovation.
- **We develop a machine learning-based methodology to construct GIII.** Machine learning method has become increasingly prevalent, yet few are utilized for measuring composite index. We introduce an unsupervised clustering-based random forest model to calculate the weight of indicators. Firstly, an unsupervised k-means clustering model based on principal component analysis is used to label each country. Then, a supervised random forest model is used to calculate the feature importance as the weight of each indicator. Finally, the weight is used to aggregate the GIII score.
- **We discover interesting insights related to key macroeconomic issues using GIII.** Income, unemployment, aging and the structure of economic sectors are areas of strong interest in today's macroeconomic studies. Based on the myriad data collected, we find that high income doesn't necessarily mean high intelligence innovation, but low income generally impedes intelligence innovation development. In addition, the analysis shows that unemployment is negatively connected to intelligence innovation, while for labor income share and aging, positive correlations emerge. In terms of the structure of economic sectors, GIII is positively correlated to the share of service sector, and negatively to the share of agriculture sector.

The remainder of the paper is organized as follows. In Section 2, we briefly review the related literature. Section 3 proposes the comprehensive evaluation index system of the global intelligence innovation. Section 4 describes the dataset and methodology. The index results are obtained, and further discussion is detailed in Section 5. Finally, Section 6 concludes with the research findings and future research ideas.

2. Literature review

2.1 National innovation evaluation

To provide a theoretical foundation for constructing the GIII index system, we first review the literature concerning national innovation. The earliest study is from Freeman et al. (1987) who formally proposed the concept of national innovation systems (NIS) based on Schumpeter's theory of innovation. Nelson (1993) pointed out that the goal of NIS is to promote technological innovation and improve the economic performance of a country's enterprises through institutional and organizational structural adjustment. Subsequently, numerous studies have complemented the concept of NIS both theoretically and practically. (Mowery, 1998; Cimoli, 2000; Intarakumnerd et al., 2002; Godin, 2009; Samara et al., 2012; Suseno and Standing, 2018). Dodgson et al. (2008) investigated the dynamics of NIS and examined the case of Taiwan. Hu and Hung (2014) adopted the concept of sectoral system of innovation which based on NIS to analyze the innovation performance in Taiwan's pharmaceutical industry. Yao et al. (2020) adopted a social networks lens on the NIS of China and confirmed that intercity innovation network enhanced city innovation. By drawing on the concept of Jackson (2011), NIS can be summarized as the complex relationships that are formed between actors or entities whose functional goal is to enable technology development and innovation.

In recent years, studies on evaluating national innovation have blossomed in the literature. The Global Innovation Index (GII) performed an annual assessment of national innovation capacity (Dutta et al., 2020). It comprised two subindices—the Innovation Input Subindex and Innovation Output Subindex—which were composed of 5 and 2 pillars, respectively. The overall GII score was computed using the weighted average method of 80 individual indicators. The European Innovation Scoreboard constructed an index system capturing 32 indicators to evaluate innovation among EU member states (Hollanders et al., 2021). These indices compiled information to make a comprehensive evaluation on relatively broad notion of innovation, but ignored the distinctive and multifaceted impacts of intelligence innovation.

Some other indices focused on intelligence innovation, but either were limited to regional evaluation of intelligence innovation or focused on a single category of frontier technologies. The China “Intelligence+” Social Development Index Report 2020 published by the China Academy of Information and Communication Technology (CAICT) put forward a 4-subindex system that covered production, consumption, public administration and infrastructure. The Delphi and analytic hierarchy process (AHP) methods were implemented (He et al., 2020). The level of intelligence development was measured and compared in various provinces, autonomous regions and municipalities within China, but was insufficient to provide a national-level comparison worldwide. The Artificial Intelligence Index Report 2021 constructed a comprehensive index system including indicators spanning from research, economy, education to policy, aiming to measure the vitality of the AI sector globally (Zhang et al., 2021). The report provided a comprehensive evaluation of multiple fields but focused exclusively on the impacts of AI and did not consider other intelligent technologies. Compared with existing indices, there lacks an index that can measure the worldwide national level of intelligence innovation, which provides opportunities to develop the Global Intelligence Innovation Index (GIII).

2.2 Comprehensive evaluation methods

Comprehensive evaluation methods in existing literature can be generally classified into three categories: subjective weighting methods, objective weighting methods, and combined weighting methods. The common subjective weighting methods include the analytic hierarchy process (AHP) (Lee, 2009; Chamodrakas et al., 2010; Ren et al., 2017), Delphi method (Gray et al., 2011; Abbassi et al., 2014), expert meeting method, etc. However, the subjective evaluation method is greatly influenced by human factors, which leads to strong randomness, high volatility of the evaluation results, poor comparability and low research persistence. The objective weighting method determines the weights of indicators mainly through information provided by the data itself. The common objective weighting methods include the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) (Boran et al., 2009; Büyüközkan and Çifçi, 2012), the Entropy Weight Method (EWM) (Shemshadi et al., 2011; He et al., 2018; Cavallaro et al., 2019; Sirimanne et al., 2019), and the Data Envelopment Analysis (DEA) (Surroca et al., 2016; Chen et al., 2018; Dionisio et al., 2021), etc.

To take into account the subjective judgment of the decision-maker and the objective characteristics of the evaluation object, the combined weighting method has emerged. For example, Cheng (2010) proposed the structural entropy weight method combining the expert opinion method with the fuzzy analysis method. Wang and Zhang (2011) combined the AHP and

DEA methods to jointly determine the weight vector of the comprehensive index of schemes and then obtained their correlation to reasonably evaluate them based on gray system theory.

Compared with these traditional comprehensive evaluation methods, machine learning based evaluation methods drew attention in recent years. Machine learning methods such as the k-nearest neighbor (Cover and Hart, 1967), recurrent neural network, “long short-term memory (LSTM)” (Gers et al., 2000), support vector machine (Vapnik, 2013), and decision tree (Breiman et al., 1984) have also begun to be applied in different fields. Compared with traditional statistical methods such as linear regression, machine learning methods have been proven to be more efficient and accurate in a variety of application scenarios (Jeong et al., 2016; Lamorski et al., 2017). There are few recently published studies that address the problem of evaluation using machine-learning methods. For example, Yan et al. (2019) built machine learning models (random forest, gradient boosting, and stacking) to predict the relationship between the runoff coefficient and its influencing factors in research on water resource evaluations of transboundary basins in China. However, few studies have applied machine learning models for index weighting construction where our study tries to fill the gap.

2.3 Imputation methods for missing data

Another area related to our study is the imputation methods of missing data. The phenomenon of missing data is a common, although rarely intended, problem that affects data quality that can have varying impacts on the ability of researchers to draw proper conclusions concerning the relevance of their data (D'Agostino, 2007). The existing methods for handling missing data can be mainly classified into three categories: deletion methods, weighting methods, and imputation methods.

If the missing data are missing completely at random (MCAR), the sample size is large enough, and the proportion of missing data is less than 5%, then the deletion method can be used (Deng et al., 2019). However, simply discarding missing data is not a reasonable practice, as valuable information may be lost and inferential power compromised (Enders, 2010). The weighting method is mainly used to address the problem of missing data caused by unanswered units. Its principle is to decompose the total weight of missing units into nonmissing units using the methods from the missing at random (MAR) scenario. Relevant methods include the P-S adjustment method (Politz and Simmons, 1949) and the reciprocal weighted adjustment method (Horvitz and Thompson, 1952).

The imputation method completes the dataset by computing the missing values utilizing the complex relationships between data. Several strategies inspired by statistical and machine learning methods have been developed to address this problem. Common statistical imputation methods include mean imputation, median imputation, regression imputation, and expectation-maximization algorithm imputation. While many statistical methods have been developed, many of them perform poorly in high-dimensional and large-scale data settings (D'Agostino, 2007) and cannot handle various types of data. For these reasons, machine learning methods have been introduced. For example, based on random forest algorithm, Stekhoven and Bühlmann (2012) proposed the missing forest algorithm, which can handle both categorical and continuous data and outperformed other imputation methods especially when complex interactions and non-linear relations exist in the dataset. In our study, we intend to select four classic imputation methods, two from statistical imputation methods and two from machine learning imputation methods. Then we test their performance by comparing their imputation results, in order to determine the imputation method with the best performance on our dataset.

3. Comprehensive evaluation index system of global intelligence innovation

3.1 Conceptual framework of national intelligence innovation

Demystifying the intelligence innovation process is crucial and provides a reasonable reference for the construction of Global Intelligence Innovation Index (GIII). To evaluate the intelligence innovation of each country, the priority is to answer the following question: what is the basic conceptual framework of national intelligence innovation? We investigate relevant literature and try to find the answer.

Intelligence innovation is a value creation process for smart economy by empowering multiple aspects of society. As Roper et al.(2008) suggested, the innovation value chain transforms knowledge into business value via a structured and complex process. As to the value chain of intelligence innovation, it comprises two main channels: the industrialization of frontier technologies and the improvement of existing industries using frontier technologies. Frontier technologies include artificial intelligence, the Internet of Things, big data, cloud computing, blockchain and 5G technology, which are able to take full advantage of data from collection to utilization. Intelligence innovation empowers changes in commerce, economy, society and government, thus leading to the creation of a smart economy (Frank and Aznar Fernández-Montesinos, 2020) with more innovation, newer entrepreneurial initiatives, higher productivity and enhanced competitiveness, eventually lifting the living standard of all citizens and creating great value.

As regards the framework for national intelligence innovation, we notice the perspective of the traditional input-output model and innovation ecosystem theory. Many frameworks of national innovation take the view of the traditional input-output model bounded by input and output. Guan and Chen (2012) considered the national innovation process to be a two-stage system comprised of “an upstream knowledge production process” and a “downstream knowledge commercialization process”. Zabala-Iturriagagoitia et al. (2021) used the DEA method to measure the productivity level of national innovation systems in Europe through various input and output variables. The U.S. Department of Commerce's Bureau of Technology (2007) suggested that innovation generated by individual technology projects from individual firms and industry sectors to the national and global levels is an extension of the traditional linear industrial chain of innovation, which forms the prototype of an entire ecosystem together with the macro environment in which they reside. Jackson (2011) defined an innovation ecosystem as “a system in which there are complex relationships between innovation actors or entities to enable technology development and innovation.” Except the coordination and interaction of relevant innovation entities, an innovation ecosystem emphasized their interdependent relationship with the external environment (Gomes et al., 2018).

Hence, based on the traditional input-output model and combined with innovation ecosystem theory, we propose three important components of national intelligence innovation, i.e., the impetus of intelligence innovation, resource input of intelligence innovation, performance output of intelligence innovation. In a multi-entity perspective, the framework of national intelligence innovation needs to incorporate the input-output innovation process with impetus from external environment like government policy environment, macroeconomic environment, etc.

Based on previous literature (Oh et al., 2016; Dedehayir et al., 2018; Granstrand and Holgersson, 2020), we identify the major participants involved in national intelligence innovation, including: (1) universities and (2) research institutions, or the research economy at the input side of input-output model; (3) enterprises, or the commercial economy at the output side of input-output model; and (4) governments, acting as the impetus of intelligence innovation. We propose that national intelligence innovation is the interaction between innovation participants to achieve economic and social benefits, from invisible knowledge generation and diffusion to application and commercialization.

With the identification of four major participants and three important components, we propose a conceptual framework of national intelligence innovation, as shown in Fig. 1. The impetus of national intelligence innovation plays a supporting role, which will guide and accelerate the process of intelligence innovation. Complex and dynamic innovation activities will be carried out with the help of resource inputs and intelligence innovation participants (mainly including universities, academic institutions, enterprises and governments) to improve intelligence innovation performance. Intelligence innovation performance will, in turn, promote the process of knowledge generation and diffusion, which is manifested as an "intelligence innovation ecological environment" that incorporates all participants and promotes the balance of the three ecological communities of research, development and application to realize sustainable value creation and innovation output. This conceptual framework provides a theoretical foundation for constructing the index system of Global Intelligence Innovation.

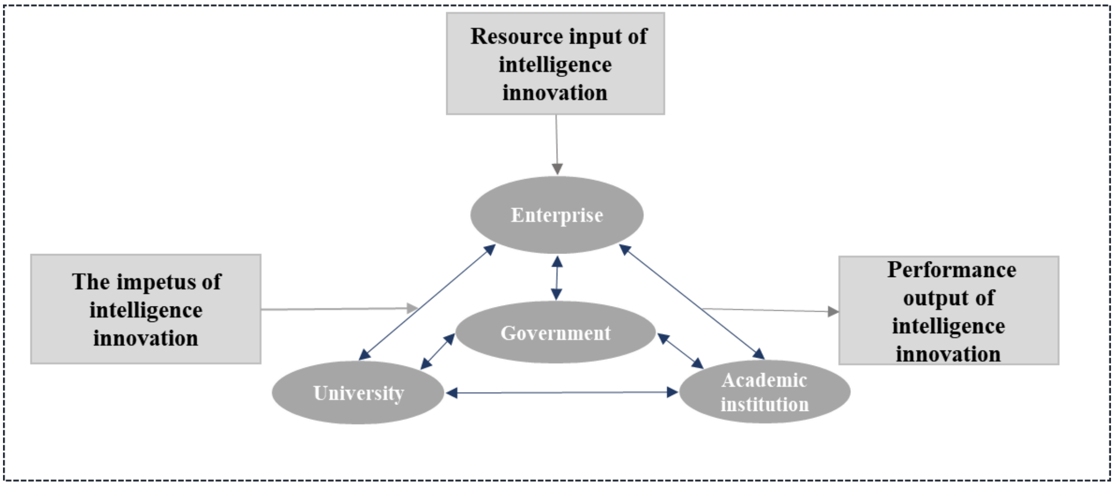

Fig. 1 The conceptual framework of national intelligence innovation

3.2 Comprehensive evaluation index system of global intelligence innovation

To directly reflect the development of national intelligence innovation worldwide, it is critical to quantitatively assess the level of intelligence innovation development in different countries. Although there is little research that evaluates global intelligence innovation using a composite index, relevant research on the national innovation evaluation index can provide some references for our study. The Global Innovation Index (GII) is an index system made up of the two subindices of innovation input and innovation output (Dutta et al., 2020). Among them, innovation input is mainly used to measure the driving factors of innovation activities and includes five indicators: institutions, human capital and research, infrastructure, market sophistication and business sophistication. Innovation output refers to the results of innovation activities, including knowledge and technology outputs, as well as creative outputs. The European Innovation Scoreboard uses 12 innovation dimensions, capturing a total of 32 indicators to assess the research and innovation performance of EU member states. All indicators lie in the four broad categories: framework conditions, investments, innovation activities and impacts (Hollanders et al., 2021).

Prior studies have mostly adopted the input-output model to construct innovation-related index systems. However, intelligence innovation which needs ecological collaboration emphasizes the interdependent relationship with the external environment (Gomes et al., 2018). That is to say, not only input and output, but also impetus from environment should be taken into account when evaluating national intelligence innovation. Consequently, three second-grade indicators of global intelligence innovation are constructed based on the conceptual framework of national intelligence innovation proposed in Section 3.1. They are the impetus, resource input and performance output of intelligence innovation.

Then we seek third-grade indicators and fourth-grade indicators following the principles of objectivity, comparability and availability. Finally, the comprehensive evaluation index system of global intelligence innovation with 28 indicators is constructed, as shown in Table 1. The index system is named as "Global Intelligence Innovation Index" (GIII).

Table 1

Comprehensive evaluation index system for global intelligence innovation

First-grade indicator	Second-grade indicator	Third-grade indicator	Fourth-grade indicator
Global intelligence innovation index	1. The impetus of intelligence innovation	1.1 Policy environment	1.1.1 The number of intelligence innovation policies
			1.1.2 The number of national science and technology awards
		1.2 Economic environment	1.2.1 GDP per capita (Thousands of dollars)
			1.2.2 The average growth rate of GDP in the last three years (%)
			1.2.3 Annual salary level of intelligence technology positions (dollars)
		1.3 Innovation environment	1.3.1 Global innovation cities of top 500
			1.3.2 Cost of business start-up procedures of GNI per capita (%)
			1.3.3 The number of job openings in the intelligence industry
	2. Resource input of intelligence innovation	2.1 Resources available for scientific research	2.1.1 The number of computer science majors ranked in the top 600 globally
			2.1.2 The number of researchers (per million inhabitants, FTL)
			2.1.3 The number of national laboratories and research centers for intelligence innovation technology
			2.1.4 The number of large-scale events held under the theme of intelligence innovation
		2.2 Financial support	2.2.1 R&D investment of government
			2.2.2 R&D investment of universities
			2.2.3 R&D investment of high-tech enterprises
		2.3 Technical foundation support	2.3.1 5G map rollout
			2.3.2 The number of supercomputers
			2.3.3 The number of secure internet servers (per million people)

		2.3.4 ICT development index
3. Performance output of intelligence innovation	3.1 Knowledge output	3.1.1 The number of papers published in the field of intelligence
		3.1.2 The number of patents related to intelligence technology
		3.1.3 Export of high-tech products (dollar)
	3.2 Industry output	3.2.1 The number of unicorns
		3.2.2 The number of firms in the intelligence industry
		3.2.3 Proportion of the added value of medium and high-tech industries in the added value of manufacturing industries (%)
	3.3 Government output	3.3.1 Government AI readiness
		3.3.2 E-government
		3.3.3 Effectiveness of government

4 Data and methodology

4.1 Data

4.1.1 Data sources and descriptive statistics

Under the GIII framework, this paper studies the top 101 countries (including some data from Hong Kong that is merged into China) in the 2019 Global Innovation Index (GII) report, which can be found in Appendix A. The statistical information is acquired from various sources, mainly including international organizations, official websites, databases, and research reports, such as:

- International organizations: World Bank, OECD
- Official websites: Wikipedia, Google, Glassdoor, Payscale, Salaryexpert, Indeed, National Bureau of Statistics, QS TOP Universities, SPEEDTEST, TOP500
- Databases: Crunchbase, Lens, UIS (The UNESCO Institute for Statistics), UNeGovKB (UN E-Government Knowledgebase), Web of Science
- Research reports: Innovation Cities Index 2019, ICT Development Index 2017, AI Watch: TES Analysis of AI Worldwide Ecosystem in 2009-2018, Government AI Readiness Index 2019.

Detailed data sources for each indicator are summarized in Table 2.

Table 2

Data sources

Indicator Code	Data Source	Indicator Code	Data Source
1.1.1	https://oecd.ai/ google	2.2.3	http://data.uis.unesco.org/ https://stat.unido.org/ https://ec.europa.eu/eurostat/web/science-technology-innovation/
1.1.2	https://wikipedia/en/wiki/Category google	2.3.1	https://www.speedtest.net/ookla-5g-map
1.2.1	https://www.kylc.com/stats	2.3.2	https://www.top500.org/statistics/
1.2.2	https://www.kylc.com/stats	2.3.3	https://data.worldbank.org/indicator/IT.NET.SECR
1.2.3	https://www.salaryexpert.com http://www.salaryexplorer.com/ https://www.glassdoor.com https://www.payscale.com	2.3.4	https://www.itu.int/net4/ITU-D/idi/2017/index.html

1.3.1	https://www.innovation-cities.com/index-2019-global-city-rankings/	3.1.1	https://www.webofscience.com
1.3.2	https://data.worldbank.org/indicator/IC.REG.COST.PC.ZS	3.1.2	https://www.lens.org
1.3.3	https://www.linkedin.com https://www.51job.com	3.1.3	https://data.worldbank.org/indicator/TX.VAL.TECH.CD
2.1.1	https://www.topuniversities.com/universities crunchbase	3.2.1	https://wikipedia./en/wiki/List_of_unicorn_startup_companies
2.1.2	https://data.worldbank.org/indicator/SP.POP.SCIE.RD.P6 http://data.uis.unesco.org/	3.2.2	https://www.crunchbase.com
2.1.3	https://publications.jrc.ec.europa.eu/repository/handle/JRC120106	3.2.3	https://data.worldbank.org/indicator/NV.MNF.TECH.ZS.UN
2.1.4	https://www.crunchbase.com	3.3.1	https://www.oxfordinsights.com/ai-readiness2019
2.2.1	http://data.uis.unesco.org/	3.3.2	https://publicadministration.un.org/egovkb
2.2.2	http://data.uis.unesco.org/ http://www.stats.gov.cn/tjsj/zxfb/202008/t20200827_1786198.html	3.3.3	http://info.worldbank.org/governance/wgi/

The descriptive statistics are performed on the original dataset to obtain the statistical summary of each indicator, as shown in Table 3. Some of the indicators contain missing data, which is summarized in Table 4.

Table 3

Descriptive statistic

Code of Indicator	Sample Size	Mean	Std.	Code of Indicator	Sample Size	Mean	Std.
1.1.1	101	5.17	9.91	2.2.3	69	16028065	61005284
1.1.2	101	9.78	34.33	2.3.1	100	149.85	804.45
1.2.1	101	2.17	2.25	2.3.2	101	4.82	25.38
1.2.2	101	3.16	1.97	2.3.3	101	748849.64	4075492.7
1.2.3	101	43226.23	35676.22	2.3.4	100	6.34	1.66
1.3.1	101	4.85	12.84	3.1.1	101	315.71	869.13
1.3.2	101	7.58	11.04	3.1.2	101	2291.55	14613.91
1.3.3	101	2779.72	12154.54	3.1.3	100	21115926001.59	108968844667.17
2.1.1	101	13.49	24.47	3.2.1	101	3.71	18.17
2.1.2	101	3754.32	1594.7	3.2.2	101	1571.1	5873.73
2.1.3	101	26.22	113.58	3.2.3	100	29.98	16.89
2.1.4	101	31.3	156.54	3.3.1	99	5.84	1.57
2.2.1	91	3396658	11569263	3.3.2	101	0.75	0.13
2.2.2	88	24389977	118688500	3.3.3	101	65.16	21.65

4.1.2 Missing value imputation

There are 2828 data points in the original dataset, of which 61 are missing (approximately 2.16% of the original dataset); the proportion of missing values in each indicator is shown in Table 4. This paper compares four common imputation methods for missing values on the original dataset, including statistical imputation methods (mean, median imputation) and machine-learning imputation methods (missForest, K-nearest neighbors (KNN) imputation). The results prove that KNN imputation outperforms all other methods on the original dataset. The process is detailed as follows.

Table 4

Missing values

Code of Indicator	2.2.1	2.2.2	2.2.3	2.3.1	2.3.4	3.1.3	3.2.3	3.3.1
Number of Missing Values in Original Dataset	10	13	32	1	1	1	1	2
Percentage of Missing Values in Original Dataset	9.80%	31.37%	12.75%	0.98%	0.98%	0.98%	0.98%	1.96%
Number of Missing Values in Test Dataset	6	20	8	1	1	1	1	1

First, all countries without missing values (65 instances, 1820 sample points) are selected from the original dataset as the original test dataset. Second, the test dataset is generated by randomly replacing a certain percentage of data points in each indicator of the original test dataset with missing values in accordance with the percentage of missing values in the corresponding indicator of the original dataset, as shown in Table 4. Third, four imputation methods—namely, mean imputation, median imputation, missForest and KNN imputation—are implemented using the ForImp, missForest (Stekhoven and Bühlmann, 2012), and VIM (Kowarik and Templ, 2016) packages on the test dataset in R. Finally, the imputation performance of missing values is measured using the normalized root mean squared error (NRMSE) (Oba et al., 2003) in Formula (1):

$$NRMSE = \sqrt{\frac{mean\left(\left(X^{true} - X^{imp}\right)^2\right)}{var\left(X^{true}\right)}} \tag{1}$$

where X^{true} is the value in the original test dataset with corresponding positions to missing values in the test dataset and X^{imp} is the imputed value with corresponding positions to missing values in the test dataset. To reduce the randomness when generating the test dataset, 20 test datasets are generated and imputed in each test. The imputation performance is evaluated by the mean of the 20 NRMSE values. The following NRMSE values all refer to this mean value.

KNN and missForest imputation require the tuning of hyperparameters. To find the optimal value, the imputation performance of each hyperparameter combination is measured using the NRMSE value and then compared (Ratolojanahary et al., 2019). The candidate values of the hyperparameters are shown in Table 5.

Table 5

Candidate values of hyperparameters

Imputation method	Hyperparameter	Candidate Value
KNN	Metric	Euclidean, Manhattan, Gower
	k	[1,15]
	weighted in distance calculation	True, False
	weighted in aggregation	True, False
missForest	ntree	50, 100, 500
	mtry	*sqrt(p), p/3*
	nodesize	[1,6]

Hyperparameters needed to tune for KNN imputation include the metric to be used for calculating the distances, the number of nearest neighbors *k*, whether the distance calculation is weighted using random forest regression, and whether the aggregation process is weighted using the distances *k* nearest neighbors. The experimental results are shown in Fig. 2. The metric used to calculate the distances does not affect the imputation performance significantly; thus, the commonly used Euclidean distance is chosen. For hyperparameters weighted in distance calculation, weighting makes imputation performance considerably less stable; thus, distance calculation is not weighted. For hyperparameters weighted in aggregation, the hyperparameter does not affect imputation performance significantly; thus, aggregation is not weighted to reduce processing time. For hyperparameter *k*, the optimal value 6 is chosen.

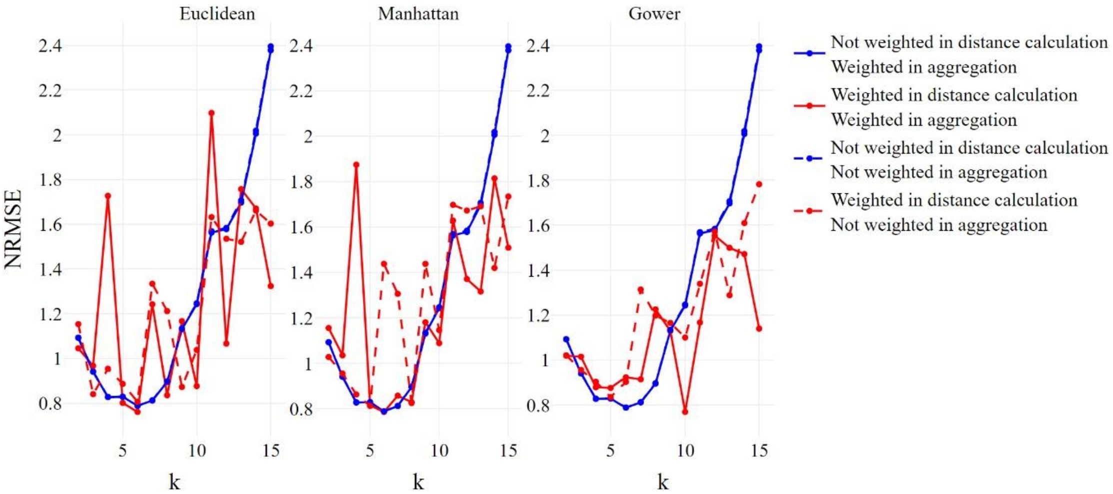

Fig. 2 Hyperparameter results of KNN

Hyperparameters needed to tune for missForest imputation include the number of trees (ntree), the number of indicators sampled at each split (mtry) with the candidate values *sqrt(p)* and *p/3* (and *p* denoting the number of indicators) referring to Probst et al. (2019), and the minimum number of samples in leaf nodes (nodesize). The test results are shown in Fig. 3. The optimal hyperparameter combination of ntree 100, mtry *sqrt(p)*, and nodesize 2 are chosen.

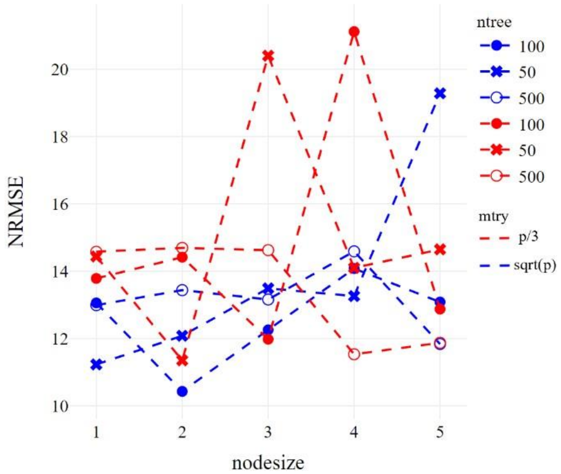

Fig. 3 Hyperparameter results of missForest

The performance of the four imputation methods is shown in Fig. 4; the KNN method outperforms the other methods. The final hyperparameter combination chosen for KNN are metric Euclidean distance, *k*=6, not weighted in distance calculation, and not weighted in aggregation.

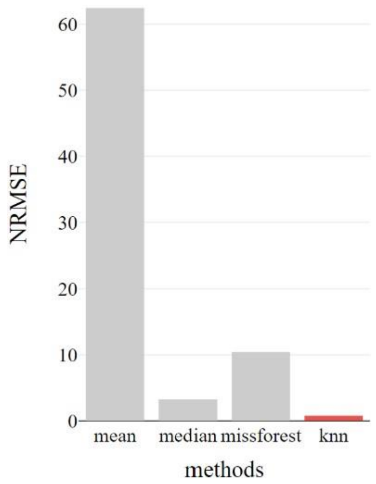

Fig. 4 Performance comparison of four imputation methods

4.2 Methodology of index construction

This study introduces machine learning methods into the process of index construction, particularly in the step of weight calculation. A random forest model based on the k-means clustering algorithm combined with principal component analysis is proposed. The following models were completed using Scikit-learn (Pedregosa et al., 2011). After data preprocessing and transformation, PCA is performed to boost the performance of the subsequent clustering algorithm. Then, k-means clustering algorithm is used to label each country. Combining the label produced by k-means and the unstandardized dataset without outlier, random forest algorithm is used to produce the weights for each indicator. Finally, the standardized dataset with outliers and the weights are combined to produce the final GIII results. The main procedures can be seen in Fig. 5, and the details are as follows.

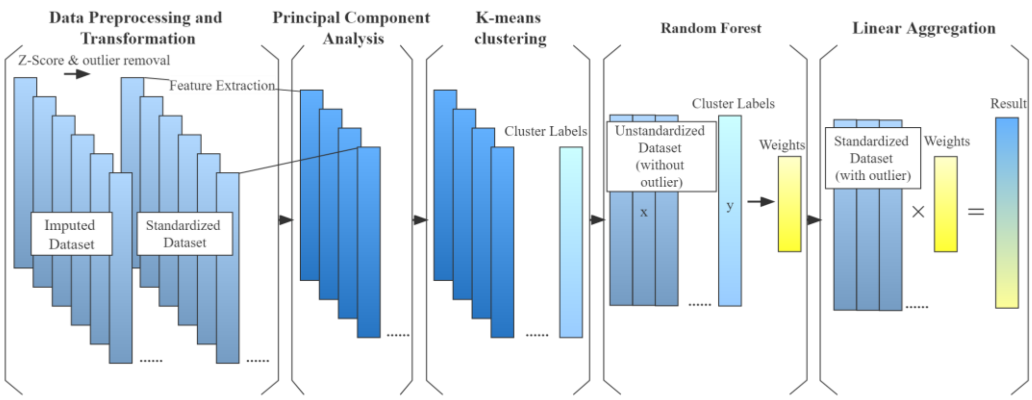

Fig. 5 Main procedures of the unsupervised clustering-based random forest model

4.2.1 Data preprocessing and transformation

The standardization of each indicator is generally performed prior to PCA to obtain indicators with similar standard deviations or scales and thus avoid biasing toward variables with significantly larger standard deviations or scales (Gewers et al.). In cluster analysis, it is also necessary to standardize each indicator to avoid the influence of indicator scale on clustering results (Kaufman and Rousseeuw, 2009). Z-score is used to standardize the original dataset.

Assume that the imputed dataset $\{X_i\}_{i=1}^{N}$ involves N countries. X_i denotes the sample of ith country, which has a total of M indicators, $X_i = (x_{i1}, x_{i2}, \dots, x_{ij}, \dots, x_{iM})'$. x_{ij} denotes the value of the ith country's jth indicator $(1 \leq i \leq N, 1 \leq j \leq M)$. z_{ij} the standardized value of x_{ij}, is calculated using the below equations:

$$z_{ij} = \frac{x_{ij} - \bar{X}_j}{S_j} \tag{2}$$

$$\bar{X}_j = \frac{1}{N} \sum_{i=1}^{N} x_{ij} \tag{3}$$

$$S_j = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (X_{ij} - \bar{X}_j)^2} \tag{4}$$

where $\bar{X}_j$ denotes the average value and S_j denotes the standard deviation of the jth indicator in N countries. Since the k-means algorithm is sensitive to outliers, how to deal with outliers is a key issue. In our paper, outliers in the imputed dataset are manually selected and removed to improve k-means performance by visualizing the Z-score standardized imputed dataset through downscaling it into two dimensions using principal component analysis (PCA), as shown in Fig. 6. China and the United States are the obvious outliers in this dataset and were thus removed. Z-score standardization is performed a second time on the imputed dataset without China and the United States. Henceforth, "standardized dataset" refers to this dataset.

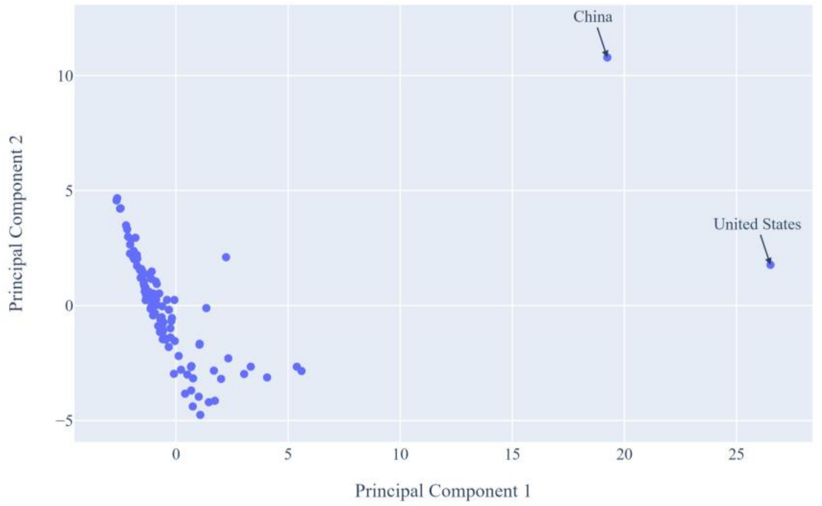

Fig. 6 Outlier detection using PCA

4.2.2 Unsupervised clustering combined with principal component analysis

1) Evaluating clustering tendency

To verify the reliability of the clustering analysis, the Hopkins statistic H (Banerjee and Davé, 2004) is used to evaluate the clustering tendency of the standardized dataset, which ranges from 0 to 1. The value is much closer to 1 when the dataset shows a stronger clustering tendency. The Hopkins statistic H for our standardized dataset is 0.82, which indicates that the dataset is suitable for clustering analysis.

Hopkins statistic H

Input: Dataset D

Procedure:

1: Randomly draw n data points p_i from D and calculate the Euclidean distance w_i between each data point p_i and its nearest point.

2: Generate n data points q_i which uniformly distribute in the space of D, and calculate the Euclidean distance u_i between each data point q_i and its nearest point.

3: Calculate the following equation:

$$H = \frac{\sum_{i=1}^{n} u_i}{\sum_{i=1}^{n} u_i + \sum_{i=1}^{n} w_i}$$

Output: Hopkins statistic H.

2) Principal component analysis

Principal component analysis (Abdi and Williams, 2010) is a common multivariate statistical analysis method used for feature extraction and dimensionality reduction. PCA transforms the original variables into a new set of linearly uncorrelated variables using linear orthogonal transformation. Based on the assumption that high-dimensional features are correlated with each other and contain redundant information in most cases, PCA is often used in machine learning and is an important step in data preprocessing by which important information from the data can be extracted and computational complexity can be reduced by dimension reduction.

The number of principal components is determined using the automatic dimensionality selection model based on the Bayesian model proposed by Minka (Minka, 2000). The 28 indicators of the standardized dataset are condensed into 20 dimensions, and the noise reduction of the dataset is expected to improve the performance of k-means clustering.

3) K-means clustering

K-means is an unsupervised clustering algorithm that uses distance to evaluate the similarity between samples. The smaller the distance between two samples is, the greater their similarity is. The core idea of the algorithm is to divide the dataset into k clusters while making the sample points within the clusters as close as possible and the distance between the clusters as large as possible.

The most important hyperparameter determined in the k-means algorithm is the number of Clusters K. To tune K, the average value of the silhouette score (Rousseeuw, 1987), the Calinski–Harabasz index (Calinski and Harabasz, 1974), and the Davies–Bouldin index (Davies and Bouldin, 1979) are used to evaluate the performance of the k-means clustering results. A series of k-means algorithms with different settings of K, ranging from 2 to 20, is performed. The final experimental results prove that the clustering result has the best performance when the hyperparameter K is set to 5. Fig. 7, Fig. 8, and Fig. 9 show the clustering results visualized in different clustering dimensions by PCA.

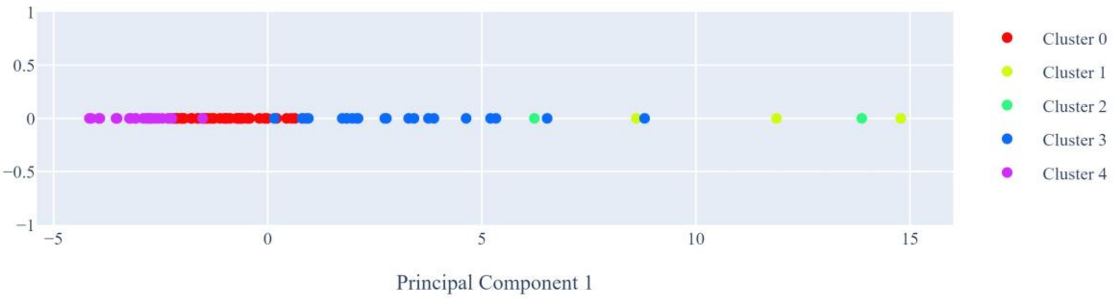

Fig. 7 One-dimensional clustering result

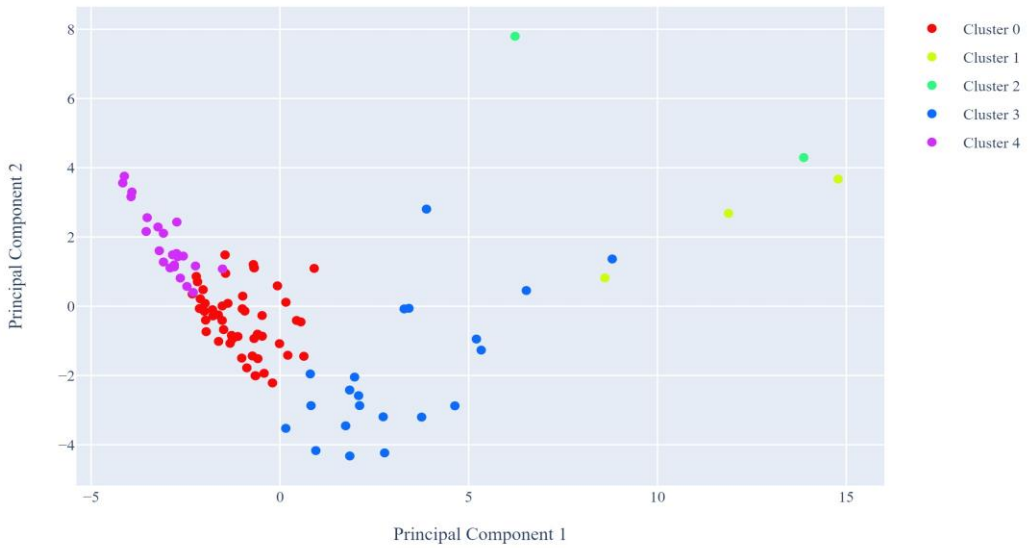

Fig. 8 Two-dimensional clustering result

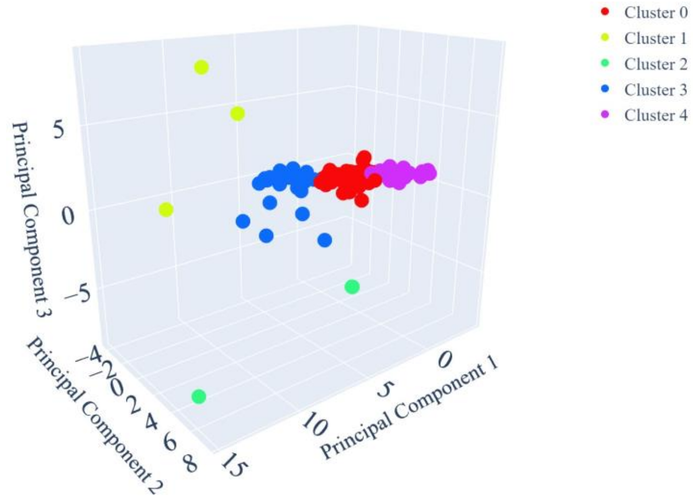

Fig. 9 Three-dimensional clustering result

4.2.3 Random forest for weighting calculation

Random forest (Breiman, 2001) is a type of supervised machine learning algorithm which belongs to the category of ensemble learning. Random forest uses the bootstrap aggregating method to generate multiple mutually independent training sets and classification and regression trees (CART) for solving classification or regression problems. The result is determined using majority voting or average probabilistic prediction. The core idea of ensemble learning is to improve performance by combining results from multiple independent algorithms, thus outperforming any single algorithm.

Given the training set $Q = \{X_i, y_i\}_{i=1}^{N}$, which contains a total of N samples, $\{X_i\}_{i=1}^{N}$ is the feature variables with X_i representing the ith sample in the unstandardized dataset with the United States and China excluded as outliers. $\{y_i | y_i \in \{0,1, \dots, K\}\}_{i=1}^{N}$ denotes the target variable, which is a categorical variable having K possible values (K classes). y_i corresponds to the result of the k-means model, with K set to 5.

The bootstrap aggregating method selects a random sample with replacement repeatedly for B times, generating B sub-training sets with the bth sub-training set $Q_{N_b} = \{X_i, y_i\}_{i=1}^{N_b}$ containing N_b samples (approximately 2/3 of the total number of samples N). Then, the bth CART, f_b, is trained on Q_{N_b} with no pruning performed. The results of B CARTs, $\{f_b(\mathrm{X})\}_{b=1}^{B}$, are combined through majority voting in which either the class with the highest number of votes from B CARTs or the average probabilistic prediction in which the class with the highest average probability prediction of B CARTs is selected.

The bootstrap aggregating method leaves around 1/3 of the training set not sampled in each sub-training set, which is called out-of-bag (OOB) data. OOB data can be used for internal error estimation to obtain the OOB error of each CART. The general error estimation of a random forest is represented by the average OOB error of all CARTs, which approximates the error obtained by cross-validation on the condition that the number of trees in a random forest is large enough (Breiman, 2001).

For bth CART, f_b in the random forest model, it is trained by recursively partitioning the root node – the sub-training set Q_{N_b}. Every node is split into two smaller subsets. The recursion is completed when the subset at a node has all the same values of the target variable. When splitting, a number m ($m \leq$ total number of features M) is specified so that at each node t, m variables are selected at random as candidate variables. To reach the terminating case of the recursion, the ideal split would make the two child nodes purer. The Gini impurity criterion is used to determine the best variable and the threshold.

Let the data at node t be represented by $Q_{N_{b,t}} = \{X_i, y_i\}_{i=1}^{N_{b,t}}$. For each candidate split $\theta = (v, t_m)$ consisting of a feature v and threshold t_m, partition the data into two subsets:

$$Q_{N_{b,t}}^{left}(\theta) = \{(x_{iv}, y_i) | x_{iv} \leq t_m\}_{i=1}^{N_{b,t}^{left}} \tag{5}$$

$$Q_{N_{b,t}}^{right}(\theta) = \{(x_{iv}, y_i) | x_{iv} > t_m\}_{i=1}^{N_{b,t}^{right}} \tag{6}$$

The Gini criterion can be calculated as:

$$G\left(Q_{N_{b,t}}, \theta\right) = \frac{N_{b,t}^{left}}{N_{b,t}} Gini\left(Q_{N_{b,t}}^{left}(\theta)\right) + \frac{N_{b,t}^{right}}{N_{b,t}} Gini(Q_{N_{b,t}}^{right}(\theta)) \tag{7}$$

$$Gini(Q)\text{=}1\text{-}\sum_{k=1}^{K}\left[p\left(\frac{k}{Q}\right)\right]^2 \tag{8}$$

where $Gini(Q)$ is used to measure the purity of the node; the larger the Gini value is, the higher the impurity. $p\left(\frac{k}{Q}\right)$ denotes the fraction of the samples labeled with class k in data Q. The candidate split that minimize the impurity is selected:

$$\theta^* = \arg\min_{\theta} G\left(Q_{N_{b,t}}, \theta\right) \tag{9}$$

For each split, the effect of this split can be measured based on the amount of decrease in impurity – Gini decrease (GD). Assume that the split at node t of CARTb is $\theta_{b,t}^*$:

$$GD(\theta_{b,t}^*) = Gini\left(Q_{N_{b,t}}\right) - G\left(Q_{N_{b,t}}, \theta_{b,t}^*\right) \tag{10}$$

Summing up the Gini decrease for each individual variable over all CARTs provides a reference for variable importance that is often very consistent with the permutation importance measure (Breiman, 2001). Generally, the mean Gini decrease (MGD) for each individual variable over all CARTs is frequently used as an estimate regarding the importance of variables:

$$W_j = MGD_j = \frac{\sum_{b=1}^{B}\sum_{t\in T_b, v=j} GD(\theta_{b,t}^*)}{\sum_{j=1}^{M}\sum_{b=1}^{B}\sum_{t\in T_b, v=j} GD(\theta_{b,t}^*))} \tag{11}$$

where W_j and MGD_j are the jth variable's weight and the MGD value, respectively. For the bth CART, it contains a set of T_b nodes.

As standardization has no obvious effect on the random forest classification result, this paper uses the unstandardized dataset without China and the United States for training random forest, in which clustering labels of each country under k-means clustering are set as target variables. The main hyperparameter settings of the random forest classification algorithm are as follows: number of trees, 500; minimum number of samples for node split, 2; minimum number of samples in leaf node, 1; and number of indicators used in each split is the square root of the total number of indicators. To minimize the influence of randomness, we perform the random forest classification algorithm 50 times, and the final weight of each indicator is the average value of the weight results for 50 times. The final weight results for each indicator are shown in Fig. 10.

The accuracy of the random forest classification algorithm is evaluated using the OOB error, which is the average OOB error from 50 random forest classification algorithms. The value of the OOB error is 0.85172, which indicates that random forest classifies the labels obtained by k-means clustering accurately.

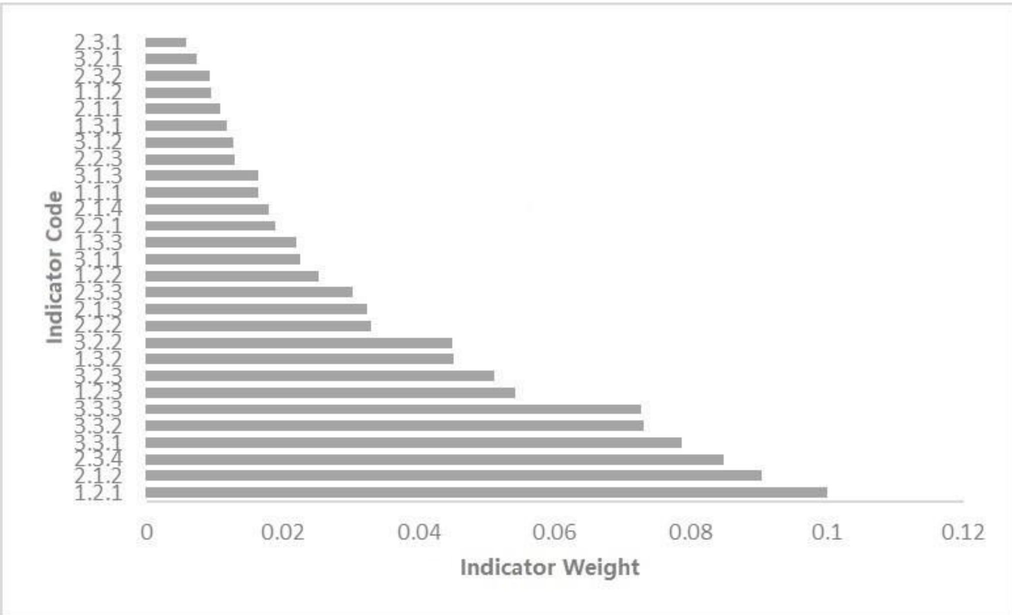

Fig. 10 Weighting results of indicators

After obtaining the indicators' weights, the final index of each country is calculated using the linear aggregation method:

$$S_i = \sum_{j=1}^{M} W_j * Z_{ij} \tag{12}$$

where S_i represents the final score of ith country, W_j represents the weight of jth indicator, and Z_{ij} represents the standardized data. The Z-score is used to standardize the original dataset including China and the United States, and negative indicators are transformed using negation.

5. Results and discussion

5.1 Results

The overall ranking of the Global Intelligence Innovation Index (GIII) is shown in Table 6. Among the 101 countries, the United States maintains considerable advantages compared to the rest of the world, showing an imbalanced development of global intelligence innovation.

Among the top 20 countries, most countries are developed countries except China. When compared with GII (Global Innovation Index), we find that the top-ranked countries in the GII are still among the top-ranked countries in GIII. In contrast, most developing countries rank relatively lower.

Regarding the three second-grade indicators, we investigate the results of the top five countries for each indicator. According to the overall ranking of the second-grade indicators, the top five countries in impetus for intelligence innovation are the United States, Luxembourg, Switzerland, Korea, and Ireland; the top five countries in resource input of intelligence innovation are the United States, China, Korea, Japan, and Germany; the top five countries in performance output of intelligence innovation are the United States, China, Singapore, England, and Germany. The United States holds first place in all second-grade indicators, and China performs well in intelligence innovation input and intelligence innovation output but falls slightly behind in impetus for intelligence innovation.

Table 6

Global Intelligence Innovation Index

Ranking	Country	GIII	Ranking	Country	GIII
1	United States	2.9241	52	Bahrain	-0.1394
2	China	1.4924	53	Brazil	-0.1627
3	Japan	1.1342	54	Kuwait	-0.1638
4	Germany	1.122	55	Costa Rica	-0.2063
5	England	1.0596	56	Montenegro	-0.2127
6	Singapore	1.055	57	Georgia	-0.214
7	Denmark	1.0512	58	Viet Nam	-0.2158
8	Switzerland	0.9921	59	Kazakhstan	-0.2206
9	Norway	0.9461	60	Serbia	-0.2277
10	South Korea	0.9324	61	Oman	-0.234
11	Sweden	0.9033	62	Brunei Darussalam	-0.2365
12	France	0.8534	63	Argentina	-0.255
13	Luxembourg	0.847	64	Colombia	-0.2966
14	Australia	0.824	65	Ukraine	-0.3107
15	Finland	0.7811	66	Belarus	-0.3227
16	Netherlands	0.7676	67	South Africa	-0.3236
17	Ireland	0.7674	68	Mauritius	-0.3243
18	Israel	0.7102	69	North Macedonia	-0.3297
19	Canada	0.697	70	Philippines	-0.3349
20	Austria	0.6772	71	Iran	-0.3596
21	Iceland	0.6417	72	Trinidad and Tobago	-0.3615
22	New Zealand	0.5721	73	Armenia	-0.3891
23	Belgium	0.517	74	Indonesia	-0.4035
24	Spain	0.4234	75	Panama	-0.409
25	United Arab Emirates	0.3842	76	Republic of Moldova	-0.412
26	Italy	0.3755	77	Jamaica	-0.427
27	Estonia	0.3743	78	Tunisia	-0.4397
28	Malta	0.2968	79	Azerbaijan	-0.4429
29	Czechia	0.2917	80	Peru	-0.4666
30	Slovenia	0.254	81	Morocco	-0.5156
31	Portugal	0.2172	82	Albania	-0.5543
32	Qatar	0.1855	83	Jordan	-0.5697
33	Cyprus	0.1767	84	Mongolia	-0.5749

34	Hungary	0.1448	85	Dominican Republic	-0.5801
35	Poland	0.1445	86	Bosnia and Herzegovina	-0.6386
36	Malaysia	0.1366	87	Botswana	-0.6566
37	Slovakia	0.1328	88	Ecuador	-0.7136
38	Lithuania	0.1282	89	Sri Lanka	-0.7177
39	Latvia	0.0932	90	Kyrgyzstan	-0.7471
40	Russian Federation	0.0818	91	Kenya	-0.7541
41	Chile	0.015	92	Egypt	-0.777
42	Greece	-0.0103	93	Rwanda	-0.7919
43	India	-0.0351	94	Lebanon	-0.8319
44	Bulgaria	-0.0464	95	Namibia	-0.8364
45	Croatia	-0.0665	96	Senegal	-0.8906
46	Saudi Arabia	-0.0916	97	Paraguay	-0.8991
47	Thailand	-0.0957	98	Tajikistan	-1.0242
48	Romania	-0.1073	99	Uganda	-1.0893
49	Mexico	-0.1156	100	Cambodia	-1.1274
50	Uruguay	-0.1216	101	United Republic of Tanzania	-1.1728
51	Turkey	-0.1295			

5.2 Analysis and discussion

5.2.1 Global Intelligence Innovation Index distribution

Fig. 11 shows the overall distribution situation of the GIII. It can be concluded that the United States has considerable advantages over the rest of the world and maintain a nearly monopolistic position in terms of intelligence innovation.

Countries with a value greater than 0 are considered countries with above average global intelligence innovation, and countries with a value less than 0 refer to countries with below average global intelligence innovation. We find that there is a large gap in the overall level of global intelligence innovation among countries around the world, but 40.59% of countries reach the global average level of intelligence innovation.

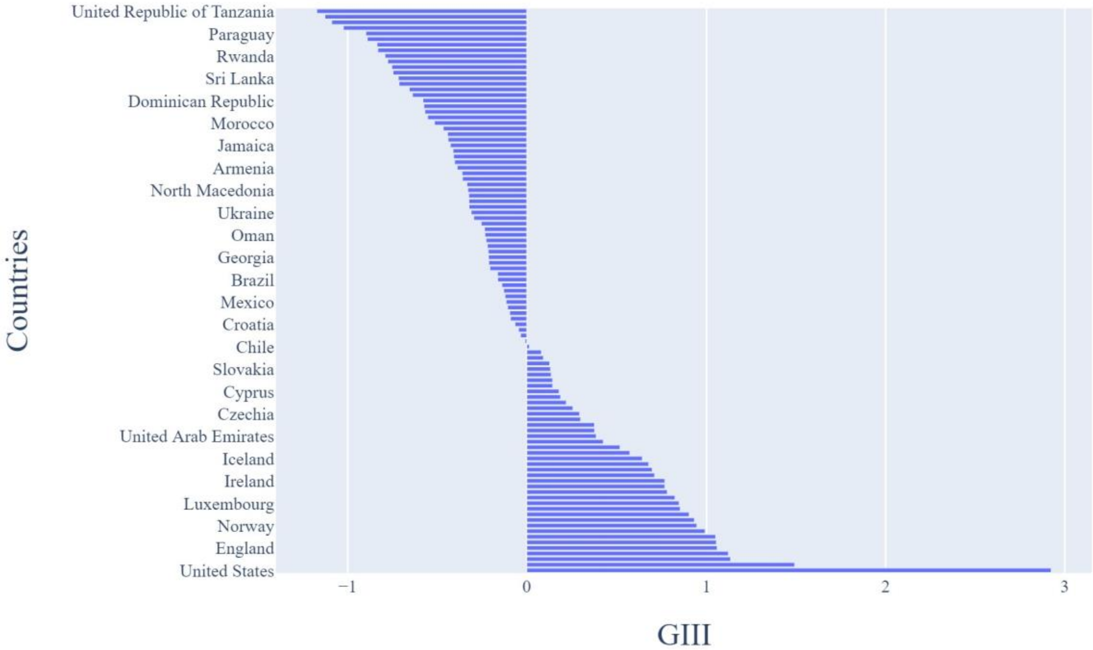

Fig. 11 Global Intelligence Innovation Index distribution

5.2.2 Correlations with macroeconomic indicators

Nowadays, national income, unemployment, aging, and the structure of economic sector are the most significant macroeconomic issues concerned by policymakers around the world. In this section, using GIII, we examine the correlation between intelligence innovation and several macroeconomic indicators to provide useful insights on these issues.

Countries' innovative abilities generally increase with their level of economic development. In terms of intelligence innovation, positive correlation relationship exists as well, with GIII score increasing with GDP per capita, as shown in Fig.12. However, by looking into the data more closely, interesting trends emerge and cannot be overgeneralized into a simple positive correlation.

To deeply examine the relationship between intelligence innovation and income, we first classify all the countries into high-income (GDP per capita above $12695), upper-middle-income (GDP per capita between $4096 and $12695), lower-middle-income (GDP per capita between $1046 and $4096) and low-income (GDP per capita below $1046), according to the criteria set by the World Bank. Then, all countries are further categorized into class A (leading type, GIII score above 0.4), class B (pursuing type, GIII score between -0.4 and 0.4), and class C (backward type, GIII score below -0.4), based on their GIII score, as shown in Table 7.

From Table 7, not all high-income countries fall into Class A, with a considerable number of countries falls into Class B, and even into Class C. Similarly, a large number of upper and lower-middle-income countries manage to perform well and are classified into Class B, with China as the exception to fall into Class A. However, all low-income countries and the majority of lower-middle-income countries fall behind and are categorized as Class C.

It can be concluded that the variance of countries' GIII scores increases with GDP per capita. The trend further manifests in Fig. 12, which shows the distribution of countries in the different income groups based on their GIII score. In other words, high-income countries do not necessarily have higher intelligence innovation performance, but a poorer economic foundation is detrimental to the development of intelligence innovation. Thus, for lower-income countries, it is preferable to focus on economic development, while for higher-income countries, other factors beyond GDP per capita play a more important role in developing intelligence innovation.

Table 7

Grading evaluation of Global Intelligence Innovation Index

Income Level	**GIII Level**	**List of countries**
High-Income	A (leading type)	Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Iceland, Ireland, Israel, Japan, Luxembourg, Netherlands, New Zealand, Norway, Singapore, South Korea, Spain, Sweden, Switzerland, United Kingdom, United States
	B (Pursuant type)	Bahrain, Brunei Darussalam, Chile, Croatia, Cyprus, Czechia, Estonia, Greece, Hungary, Italy, Kuwait, Latvia, Lithuania, Malta, Oman, Poland, Portugal, Qatar, Romania, Saudi Arabia, Slovakia, Slovenia, Trinidad and Tobago, United Arab Emirates, Uruguay
	C (Backward type)	Panama
Upper-Middle-	A (Leading type)	China

Income	B (Pursuant type)	Argentina, Armenia, Belarus, Brazil, Bulgaria, Colombia, Costa Rica, Georgia, Iran (Islamic Republic of Iran), Kazakhstan, Malaysia, Mauritius, Mexico, Montenegro, North Macedonia, Russian Federation, Serbia, South Africa, Thailand, Turkey
	C (Backward type)	Albania, Azerbaijan, Bosnia and Herzegovina, Botswana, Dominican Republic, Ecuador, Indonesia, Jamaica, Jordan, Lebanon, Mongolia, Namibia, Paraguay, Peru, Republic of Moldova
Lower-Middle-Income	B (Pursuant type)	India, Philippines, Ukraine, Viet Nam
	C (Backward type)	Cambodia, Egypt, Kenya, Kyrgyzstan, Morocco, Senegal, Sri Lanka, Tunisia, United Republic of Tanzania
Low-Income	C (Backward type)	Rwanda, Tajikistan, Uganda

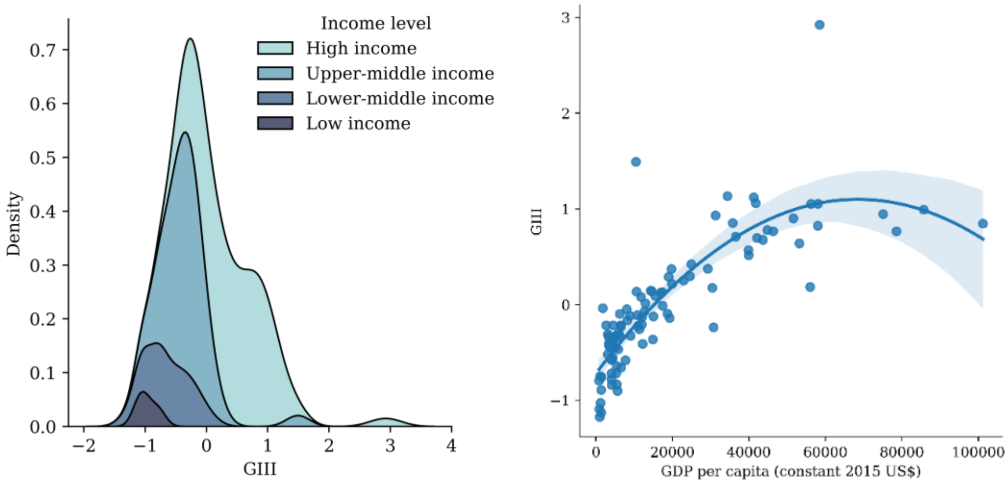

Fig. 12 Relationship between income and GIII[1].

Unemployment is another major concern for policymakers. In Fig. 13, we find that the level of intelligence innovation is slightly negatively correlated with the unemployment rate. In addition, the labor income share of GDP, the compensation to employees, increases with higher GIII score. Contrary to the usual concern that intelligence innovation and automation in general would increase unemployment and render labor factor redundant in production, the result suggests the positive impact of intelligence innovation on employment. Our findings are consistent with recent literature on the effect of automation on employment and factor shares. Acemoglu and Restrepo (2018)'s macroeconomic model shows that the two forces – the force of automation reducing employment and the force of automation creating new tasks – can reach long-term equilibrium and achieve stable balanced growth. Empirical evidence from patents also showed that advances in automation technology have a positive influence on employment (Mann and Püttmann, 2021).

The relationship between GIII and labor income share also shows another interesting insight. In general, the country's labor income share is negatively correlated to GDP per capita, as capital plays an increasingly important role in economic development when the economy grows.

[1] The data of GDP per capita was measured in 2020 and provided by the World Bank.

However, in terms of intelligence innovation, the opposite trend emerges. The labor income share is positively correlated to GIII score, implying a more significant role that labor plays in developing intelligence innovation. It suggests that countries should attach more importance to labor, especially human capital such as scientists and academics. The support for labor to develop labor capital is the key to promoting intelligence innovation.

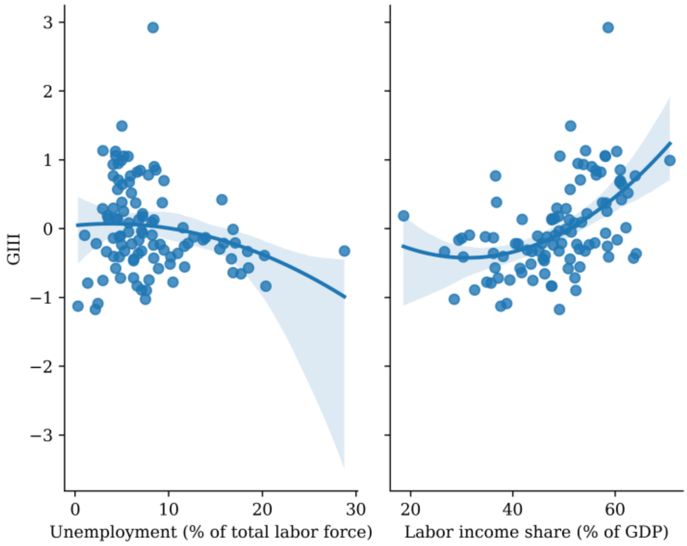

Fig. 13 Relationship between unemployment, labor income share and GIII[2].

In Fig. 14, we analyze the relationship between intelligence innovation and population aging, which shows a positive correlation. The result suggests the impact of endogenous response of technology – countries experiencing more severe aging problems turn to automation to compensate for the labor shortage. This finding is in line with recent literature, showing that intelligence innovation, especially the adoption of automation technology, is receiving a powerful boost from demographic changes (Acemoglu and Restrepo, 2022). It points out a possible solution of emphasizing intelligence innovation development to alleviate problems caused by the aging of the population.

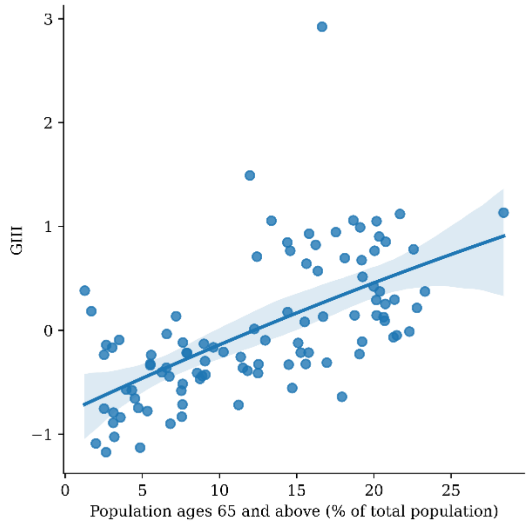

Fig. 14 Relationship between aging and GIII[3].

[2] The data of unemployment rate and labor income share was measured in 2020 and 2017 respectively, both provided by the World Bank.

The structure of economic sector is another matter worth investigating. As reported in Fig.15, a country's performance of innovation intelligence is positively linked with the share of the service sector in its GDP, slightly negatively linked with the share of the industry sector, and strongly negatively linked with the share of the agriculture sector. One possible interpretation of these relationships is that intelligence innovation promotes development in the service sector more significantly compared to the industry and agriculture sector. Another interpretation is that the development of the service sector has a greater impact on boosting the country's intelligence innovation compared to the industry and agriculture sector.

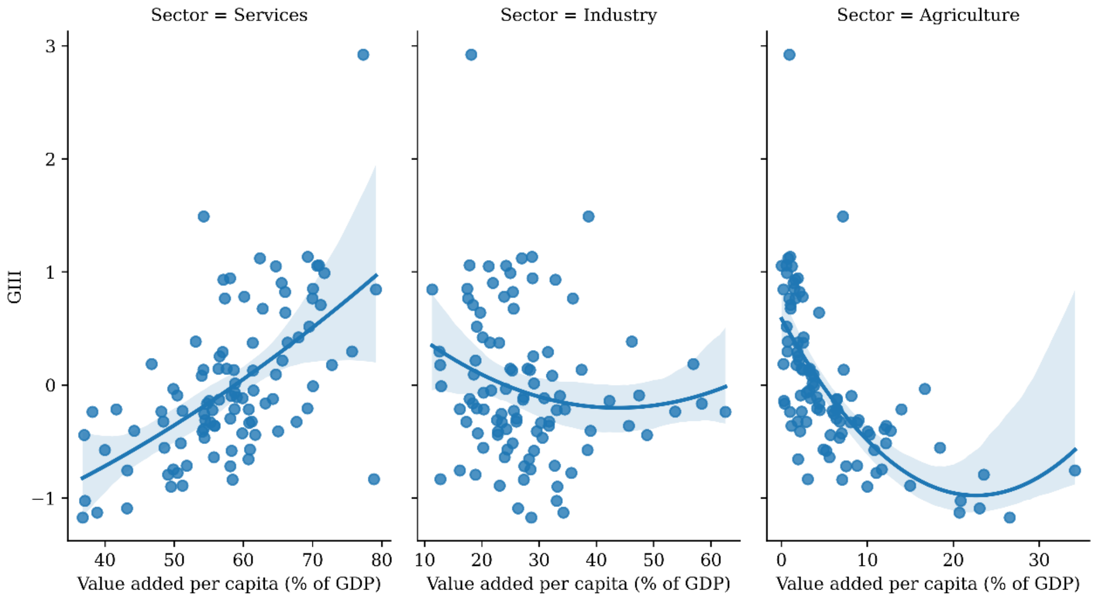

Fig. 15 Relationship between share of the economic sector and GIII[4].

6. Conclusion

Intelligence innovation plays crucial role in promoting national competitiveness and development. Meanwhile, the evaluation of global national level intelligence innovation is not only academically interesting but also practically substantive. This study aims to develop the global intelligence innovation index to evaluate the level of each country's intelligence innovation development and provide insights into the current situation and trends in global intelligence innovation development. In addition, this paper contributes to both the academic and political fields, as it enriches intelligence-related research literature and promotes national intelligence innovation development.

Theoretical contribution. This paper proposes the conceptual framework of national intelligence innovation with governments, enterprises, universities, and academic research institutions as the core participants based on the perspective of innovation ecosystem theory, which demystifies the complex process of intelligence innovation. Second, due to missing values in the original datasets, this paper imputes missing values by comparing several statistical methods (mean, median imputation) and machine-learning methods (KNN, missForest imputation) using normalized root mean squared error (NRMSE). The experimental results suggest that KNN yields the best performance in the original datasets, which verifies the relative advantages of machine

[3] The data of population ages 65 and above was measured in 2020 and provided by the World Bank.
[4] The data of sectoral value-added per capita was measured in 2019 and provided by the World Bank.

learning imputation. Finally, this paper proposes a machine learning methodology which uses an unsupervised clustering-based random forest model for weight calculation. The method introduces machine learning methods for index construction and solves the problem of lacking target variables to train supervised machine learning methods such as random forest.

Practical contribution. The results of this study indicates that higher income may could not lead to higher intelligence innovation performance, however, lower income is detrimental to the development of intelligence innovation. Therefore, countries should promote "stepwise" innovation policies in line with their differences in income and intelligence innovation development. It is preferable for countries of lower income to focus on laying down a solid economic foundation. For countries of higher income, factors other than GDP per capita should be emphasized, and "localized" policies in specific national contexts should be developed. For example, class A countries should maintain their leading edge and focus their resources on building their core competitiveness, while class B countries should identify and make up for their shortcomings, and class C countries should start from the basics and vigorously develop their economy and increase investment in education and infrastructure, gradually moving upwards through intelligence development paths. Furthermore, the labor factor share shows a positive relationship with GIII, implying its significance in supporting intelligence innovation. As a result, countries should provide more incentives for labor, especially scientists and academics, in order to promote intelligence innovation. Using GIII, the relationships between intelligence innovation and several other crucial macroeconomic indicators are investigated. In terms of unemployment, the unemployment rate decreases and labor income share increases with GIII, thus the possible downside of unemployment brought by intelligence innovation should not be overstated. For aging, countries with a higher percentage of aging population tend to develop and implement intelligence innovation more extensively, which elicits a possible way to ameliorate the economic burden of the aging population. Intelligence innovation also exhibits a stronger relationship with the development of service sector, suggesting the transformative power of intelligence innovation on a country's economic structure.

Our analysis also opens up several opportunities for future studies. First, the interaction of intelligence innovation activities between countries, such as "open innovation" (Lee et al., 2020), or sustainable development of intelligence innovation, can be considered in the construction of an intelligence innovation index system in the future. Second, as missing data are a commonly faced problem in index construction, more imputation methods could be examined and compared on similar datasets concerning socioeconomic indicators. Third, other methods of integrating machine learning algorithms into index construction can be further explored. In addition, other subjective factors could be included in the weighting method. For example, the result of the expert evaluation method can be integrated into the random forest model to make the integrated index more universal.

References

Abbassi, M., Ashrafi, M., Sharifi Tashnizi, E., 2014. Selecting balanced portfolios of R&D projects with interdependencies: A Cross-Entropy based methodology. Technovation 34, 54–63.

Abdi, H., Williams, L.J., 2010. Principal component analysis. WIREs Comp Stat 2, 433–459.

Acemoglu, D., Restrepo, P., 2018. The Race between Man and Machine: Implications of

Technology for Growth, Factor Shares, and Employment. American Economic Review 108, 1488–1542.

Acemoglu, D., Restrepo, P., 2022. Demographics and Automation. The Review of Economic Studies 89, 1–44.

Archer, K.J., Kimes, R.V., 2008. Empirical characterization of random forest variable importance measures. Computational Statistics & Data Analysis 52, 2249–2260.

Banerjee, A., Davé, R.N., 2004. Validating clusters using the Hopkins statistic, in: 2004 IEEE International Conference on Fuzzy Systems (IEEE Cat. No.04CH37542), 149-153.

Boran, F.E., Genç, S., Kurt, M., Akay, D., 2009. A multi-criteria intuitionistic fuzzy group decision making for supplier selection with TOPSIS method. Expert Systems with Applications 36, 11363–11368.

Breiman, L., 2001. Random forests. Machine learning 45, 5–32.

Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A., 1984. Classification and regression trees. CRC press.

Büyüközkan, G., Çifçi, G., 2012. A novel hybrid MCDM approach based on fuzzy DEMATEL, fuzzy ANP and fuzzy TOPSIS to evaluate green suppliers. Expert Systems with Applications 39, 3000–3011.

Calinski, T., Harabasz, J., 1974. A dendrite method for cluster analysis. Comm. in Stats. - Theory & Methods 3, 1–27.

Cavallaro, F., Zavadskas, E.K., Streimikiene, D., Mardani, A., 2019. Assessment of concentrated solar power (CSP) technologies based on a modified intuitionistic fuzzy topsis and trigonometric entropy weights. Technological Forecasting and Social Change 140, 258–270.

Chamodrakas, I., Batis, D., Martakos, D., 2010. Supplier selection in electronic marketplaces using satisficing and fuzzy AHP. Expert Systems with Applications 37, 490–498.

Chen, X., Liu, Z., Zhu, Q., 2018. Performance evaluation of China's high-tech innovation process: Analysis based on the innovation value chain. Technovation 74-75, 42–53.

Cheng, Q., 2010. Structure entropy weight method to confirm the weight of evaluating index. Systems Engineering-theory & Practice 30, 1225–1228.

Cimoli, M., 2000. Developing Innovation Systems: Mexico in a global context. Psychology Press.

D'Agostino, R.B., 2007. Overview of missing data techniques. Methods in molecular biology (Clifton, N.J.) 404, 339–352.

Davies, D.L., Bouldin, D.W., 1979. A cluster separation measure. IEEE transactions on pattern analysis and machine intelligence, 224–227.

Dedehayir, O., Mäkinen, S.J., Roland Ortt, J., 2018. Roles during innovation ecosystem genesis: A literature review. Technological Forecasting and Social Change 136, 18–29.

Deng, J., Dan, B., He, D., Tang, R., 2019. Processing Method of Missing Data and Its Developing Tendency. Statistics & Decision 35, 28–34.

Dionisio, E.A., Inácio Júnior, E., Fischer, B.B., 2021. Country-level efficiency and the index of dynamic entrepreneurship: Contributions from an efficiency approach. Technological Forecasting and Social Change 162, 120406.

Dodgson, M., Mathews, J., Kastelle, T., Hu, M.-C., 2008. The evolving nature of Taiwan's national innovation system: The case of biotechnology innovation networks. Research Policy 37, 430–445.

Dutta, S.D., Lanvin, B., Wunsch-Vincent, S., 2020. Global Innovation Index 2020: Who Will

Finance Innovation. World Intellectual Property Organization, 448 pp. https://www.wipo.int/edocs/pubdocs/en/wipo_pub_gii_2020.pdf. Accessed 23 August 2021.
Enders, C.K., 2010. Applied missing data analysis. Guilford press.
Federal Government, 2020. Artificial Intelligence Strategy of the Federal Government, 35 pp. https://www.ki-strategie-deutschland.de/files/downloads/201201_Fortschreibung_KI-Strategie.pdf. Accessed 29 August 2021.
Federal Ministry for Economic Affairs and Energy, 2019. Blockchain Strategy of the Federal Government, 24 pp. https://www.bmwi.de/Redaktion/EN/Publikationen/Digitale-Welt/blockchain-strategy.pdf?__blob=publicationFile&v=3. Accessed 29 August 2021.
Federal Ministry of the Interior, 2021. Open-Data-Strategie of the Federal Government, 32 pp. https://www.bmi.bund.de/SharedDocs/downloads/DE/publikationen/themen/moderne-verwaltung/open-data-strategie-der-bundesregierung.pdf?__blob=publicationFile&v=3. Accessed 29 August 2021.
Federal Ministry of Transport and Digital Infrastructure, 2017. 5G Strategy for Germany, 28 pp. https://www.bmvi.de/SharedDocs/EN/publications/5g-strategy-for-germany.pdf?__blob=publicationFile. Accessed 29 August 2021.
Frank, E., Aznar Fernández-Montesinos, G., 2020. Smart City = Smart Citizen = Smart Economy? In: Cornetta, G., Touhafi, A., Muntean, G.-M. (Eds.) Social, legal, and ethical implications of IoT, cloud, and edge computing technologies. Information Science Reference an imprint of IGI Global, Hershey PA, pp. 161–180.
Freeman, R., Freeman, C., Freeman, S., 1987. Technology, policy, and economic performance: lessons from Japan. Burns & Oates.
Gers, F.A., Schmidhuber, J., Cummins, F., 2000. Learning to forget: Continual prediction with LSTM. Neural computation 12, 2451–2471.
Gewers, F.L., Ferreira, G.R., Arruda, H.F. de, Silva, F.N., Comin, C.H., Amancio, D.R., Da Costa, L.F. Principal Component Analysis: A Natural Approach to Data Exploration.
Godin, B., 2009. National Innovation System. Science, Technology, & Human Values 34, 476–501.
Gomes, L.A.d.V., Facin, A.L.F., Salerno, M.S., Ikenami, R.K., 2018. Unpacking the innovation ecosystem construct: Evolution, gaps and trends. Technological Forecasting and Social Change 136, 30–48.
Granstrand, O., Holgersson, M., 2020. Innovation ecosystems: A conceptual review and a new definition. Technovation 90-91, 102098.
Gray, J.V., Roth, A.V., Leiblein, M.J., 2011. Quality risk in offshore manufacturing: Evidence from the pharmaceutical industry. Journal of Operations Management 29, 737–752.
Guan, J., Chen, K., 2012. Modeling the relative efficiency of national innovation systems. Research Policy 41, 102–115.
He, G., Song, M., Liu, G., Gao, Y., 2020. China "Intelligence+" Social Development Index Report (2020). China Academy of Information and Communication Technology, 51 pp. https://www.isc.org.cn/zxzx/xhdt/%E3%80%8A%E4%B8%AD%E5%9B%BD%E2%80%9C%E6%99%BA%E8%83%BD+%E2%80%9D%E7%A4%BE%E4%BC%9A%E5%8F%91%E5%B1%95%E6%8C%87%E6%95%B0%E6%8A%A5%E5%91%8A%EF%BC%882020%EF%BC%89%E3%80%8B.pdf. Accessed 23 August 2021.
He, Y.X., Jiao, Z., Yang, J., 2018. Comprehensive evaluation of global clean energy development

index based on the improved entropy method. Ecological Indicators 88, 305–321.
Hollanders, H., Es-Sadki, N., Rantcheva, A., 2021. European Innovation Scoreboard 2021. European Commission, 95 pp. https://ec.europa.eu/docsroom/documents/46013/attachments/1/translations/en/renditions/native. Accessed 24 August 2021.
Horvitz, D.G., Thompson, D.J., 1952. A generalization of sampling without replacement from a finite universe. Journal of the American Statistical Association 47, 663–685.
Hu, M.-C., Hung, S.-C., 2014. Taiwan's pharmaceuticals: A failure of the sectoral system of innovation? Technological Forecasting and Social Change 88, 162–176.
Intarakumnerd, P., Chairatana, P., Tangchitpiboon, T., 2002. National innovation system in less successful developing countries: the case of Thailand. Research Policy 31, 1445–1457.
Jackson, D.J., 2011. What is an innovation ecosystem. National Science Foundation 1, 1–13.
Jeong, J.H., Resop, J.P., Mueller, N.D., Fleisher, D.H., Yun, K., Butler, E.E., Timlin, D.J., Shim, K.-M., Gerber, J.S., Reddy, V.R., Kim, S.-H., 2016. Random Forests for Global and Regional Crop Yield Predictions. PloS one 11, e0156571.
Kaufman, L., Rousseeuw, P.J., 2009. Finding groups in data: an introduction to cluster analysis. John Wiley & Sons.
Kowarik, A., Templ, M., 2016. Imputation with the R Package VIM. Journal of Statistical Software 74.
Lamorski, K., Šimůnek, J., Sławiński, C., Lamorska, J., 2017. An estimation of the main wetting branch of the soil water retention curve based on its main drying branch using the machine learning method. Water Resour. Res. 53, 1539–1552.
Lee, A.H., 2009. A fuzzy supplier selection model with the consideration of benefits, opportunities, costs and risks. Expert Systems with Applications 36, 2879–2893.
Lee, S., Lee, H., Lee, C., 2020. Open innovation at the national level: Towards a global innovation system. Technological Forecasting and Social Change 151, 119842.
Mann, K., Püttmann, L., 2021. Benign Effects of Automation: New Evidence from Patent Texts. The Review of Economics and Statistics, 1–45.
Minka, T., 2000. Automatic choice of dimensionality for PCA. Advances in neural information processing systems 13, 598–604.
Mowery, D.C., 1998. The changing structure of the US national innovation system: implications for international conflict and cooperation in R&D policy. Research Policy 27, 639–654.
Nelson, R.R., 1993. National innovation systems: a comparative analysis. Oxford University Press on Demand.
Oba, S., Sato, M., Takemasa, I., Monden, M., Matsubara, K., Ishii, S., 2003. A Bayesian missing value estimation method for gene expression profile data. Bioinformatics 19, 2088–2096.
Oh, D.-S., Phillips, F., Park, S., Lee, E., 2016. Innovation ecosystems: A critical examination. Technovation 54, 1–6.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., 2011. Scikit-learn: Machine learning in Python. the Journal of machine Learning research 12, 2825–2830.
Politz, A., Simmons, W., 1949. An Attempt to Get the “Not at Homes” into the Sample Without Callbacks. Journal of the American Statistical Association 44, 9–16.
Probst, P., Wright, M.N., Boulesteix, A.-L., 2019. Hyperparameters and tuning strategies for

random forest. WIREs Data Mining Knowl Discov 9.
Ratolojanahary, R., Houé Ngouna, R., Medjaher, K., Junca-Bourié, J., Dauriac, F., Sebilo, M., 2019. Model selection to improve multiple imputation for handling high rate missingness in a water quality dataset. Expert Systems with Applications 131, 299–307.
Ren, J., Liang, H., Chan, F.T., 2017. Urban sewage sludge, sustainability, and transition for Eco-City: Multi-criteria sustainability assessment of technologies based on best-worst method. Technological Forecasting and Social Change 116, 29–39.
Roper, S., Du, J., Love, J.H., 2008. Modelling the innovation value chain. Research Policy 37, 961–977.
Rousseeuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics 20, 53–65.
Samara, E., Georgiadis, P., Bakouros, I., 2012. The impact of innovation policies on the performance of national innovation systems: A system dynamics analysis. Technovation 32, 624–638.
Shemshadi, A., Shirazi, H., Toreihi, M., Tarokh, M.J., 2011. A fuzzy VIKOR method for supplier selection based on entropy measure for objective weighting. Expert Systems with Applications 38, 12160–12167.
Sirimanne, S.N., Fredriksson, T., Garces, P.F., Gil, S.F., Jones, C., Kidane, M.J., Korka, D., van Giffen, T., 2019. Digital Economy Report 2019: Value Creation and Capture: Implications for Developing Countries. United Nations Conference on Trade and Development, 194 pp. https://unctad.org/system/files/official-document/der2019_en.pdf. Accessed 22 August 2021.
Stekhoven, D.J., Bühlmann, P., 2012. MissForest--non-parametric missing value imputation for mixed-type data. Bioinformatics (Oxford, England) 28, 112–118.
Surroca, J., Prior, D., Tribó Giné, J.A., 2016. Using panel data dea to measure CEOs' focus of attention: An application to the study of cognitive group membership and performance. Strat. Mgmt. J. 37, 370–388.
Suseno, Y., Standing, C., 2018. The Systems Perspective of National Innovation Ecosystems. Syst. Res 35, 282–307.
Thornberry, W.M., 2020. National Artificial Intelligence Initiative Act of 2020. U.S. House of Representative, 2214 pp. https://www.congress.gov/116/crpt/hrpt617/CRPT-116hrpt617.pdf#page=1210. Accessed 22 August 2021.
Tortorella, G.L., Fogliatto, F.S., Espôsto, K.F., Vergara, A.M.C., Vassolo, R., Mendoza, D.T., Narayanamurthy, G., 2020. Effects of contingencies on healthcare 4.0 technologies adoption and barriers in emerging economies. Technological Forecasting and Social Change 156, 120048.
Vapnik, V., 2013. The nature of statistical learning theory. Springer science & business media.
Wang, L., Luo, G., Sari, A., Shao, X.-F., 2020. What nurtures fourth industrial revolution? An investigation of economic and social determinants of technological innovation in advanced economies. Technological Forecasting and Social Change 161, 120305.
Wang, X., Zhang, Y., 2011. Non-uniform grey relational method based on AHP and DEA. Systems Engineering-theory & Practice 31, 1221–1229.
Wang, Y., Sutherland, D., Ning, L., Pan, X., 2015. The evolving nature of China's regional innovation systems: Insights from an exploration–exploitation approach. Technological Forecasting and Social Change 100, 140–152.

Yan, J., Jia, S., Lv, A., Zhu, W., 2019. Water Resources Assessment of China's Transboundary River Basins Using a Machine Learning Approach. Water Resour. Res. 55, 632–655.

Yao, L., Li, J., Li, J., 2020. Urban innovation and intercity patent collaboration: A network analysis of China's national innovation system. Technological Forecasting and Social Change 160, 120185.

Zabala-Iturriagagoitia, J.M., Aparicio, J., Ortiz, L., Carayannis, E.G., Grigoroudis, E., 2021. The productivity of national innovation systems in Europe: Catching up or falling behind? Technovation 102, 102215.

Zhang, D., Mishra, S., Brynjolfsson, E., Etchemendy, J., Ganguli, D., Grosz, B., Lyons, T., Manyika, J., Niebles, J.C., Sellitto, M., Shoham, Y., Clark, J., Perrault, R., 2021. Artificial Intelligence Index Report 2021. Stanford Institute for Human-Centered Artificial Intelligence, 222 pp. https://aiindex.stanford.edu/wp-content/uploads/2021/03/2021-AI-Index-Report_Master.pdf. Accessed 23 August 2021.