Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage:
https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/index
ISSN: 2462-1943

Application of Big Data in The Prevention of Work-Related Crimes in Early-Stage Construction Engineering

Lei Hao[1], Khoo Terh Jing[1,*], Zhang Ruirui[1], Ha Chin Yee[1], Shi Yangle[1], Chen Siyao[1]

[1] Department of Construction Project Management, School of Housing, Building, and Planning, Universiti Sains Malaysia, 11800 Minden, Penang, Malaysia

PAPER INFO

Paper history:
Received 13 August 2023
Received in revised form 9 October 2023
Accepted 12 October 2023
Available online 25 October 2023

Keywords:
Work-related crimes; The early stage construction; Big data technology; Construction project; Crime prevention strategy; Quantitative research

ABSTRACT

Preventing work-related crimes in the early stages of construction projects is a crucial way to ensure the safety, efficiency, and integrity of construction projects. Big data technology offers great potential to enhance crime prevention strategies through advanced data analytics and predictive models. The early stage of construction projects includes bidding, land acquisition, demolition, approval, and procurement. Work-related crimes in the early stage of construction projects mainly involve corruption, bribery crimes, and dereliction of work-related crimes. This paper discussed the application of big data in the prevention of work-related crimes in the early stage of construction projects. First of all, this paper explored the characteristics of work-related crimes in the early stage of construction projects, and the patterns and types of crimes; secondly, identified the role of big data technology in the prevention of work-related crimes in the early stage of construction projects; finally, based on the survey results and analysis, this paper proposed to use work-related crime prevention strategy of big data technology in the early stage of construction projects. The research used quantitative research methods and simple random sampling to select various construction projects in Zhejiang, Beijing, Shandong, Guangdong, Henan, Jiangsu, Hebei, Hubei, Fujian, and Liaoning, which are the ten regions with the highest occurrence of work-related crimes in China. The target samples of this research included construction workers, construction companies, relevant stakeholders, and professionals closely related to the construction industry. An online questionnaire was administered to the participants for data collection. The collected data were analyzed using descriptive analysis techniques in the Statistical Package for the Social Sciences (SPSS) software. The result of this research demonstrated the importance of big data technology in the prevention of work-related crimes in the early stages of construction projects.

1. Introduction

At present, the year-on-year growth rate of China's construction industry is almost higher than the growth rate of the country's GDP. The status of pillar industries is stable. The construction industry has a positive role in stimulating economic growth. 121020.72 billion yuan, an increase of

* *Corresponding author.*
E-mail address: terhjing@usm.my

https://doi.org/10.37934/araset.32.2.457474

3.0% over the previous year. The added value of the construction industry of the whole society was 8.33831 billion yuan, an increase of 5.5% over the previous year, and the growth rate was 2.5 percentage points higher than China's GDP [1]. The capital investment in various key projects and urban infrastructure construction in China is gradually increasing. The huge investment of the state, while promoting infrastructure construction, promoting development, and improving people's livelihood, also provides opportunities for some criminals to use their power for personal gain [2]. As shown in Figure 1:

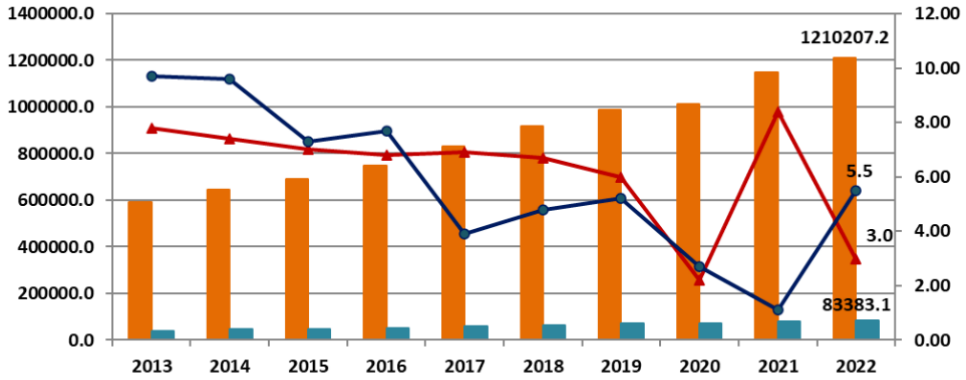

Fig. 1. 2013-2022 China's GDP, the construction industry added value and growth rate [1]

According to the statistics of the incidence rate and average loss of work-related crimes in the world, it is found that the field of engineering construction is a high-incidence field of work-related crimes in the country, and the losses caused by work-related crimes are huge [3,4]. In the era of big data, work-related crimes tend to be intelligent, concealed, and diversified. Traditional investigation methods make it difficult to deal with complex cases, complicated information, and diverse evidence. Big data, is a new type of resource with super penetration and influence, Effective use of big data technology can play a good role in preventing work-related crimes [5]. In addition, as the development of informatization enters a new stage, information technology and big data analysis methods can also help discipline inspection and supervision agencies find clues and fix evidence, insert technological "wings" into discipline inspection and supervision work, and improve work efficiency [6]. The vicious impact of frequent work-related crimes has put forward inevitable requirements for the systematization and standardization of work-related crime prevention, especially for engineering construction projects that are high-incidence work-related crime areas, the task of forming a work-related crime pre-prevention system is more urgent and necessary [7]. Therefore, strengthening prevention from the source, establishing a whole-project, whole-process, and full-coverage work-related crime prevention system, and ensuring high-quality engineering is a realistic and urgent problem facing the field of construction engineering [8].

2. Literature Review

In a broad sense, work-related crimes have rich connotations. For example, the "crime of embezzlement and bribery" stipulated in Chapter 8 of China's *Criminal Law (1997)* and the "crime of dereliction of work-related" in Chapter 9 all belong to the category of work-related crimes. The staff of government agencies, state-owned companies, enterprises, public institutions, and people's organizations authorized by the government take advantage of their power and convenience in supervising, managing, operating, and handling public and private property to engage in corruption, bribery, malpractice for personal gain, abuse of power, dereliction of work-related, and violations of

citizens' integrity. Democratic rights, undermining the rules and regulations established by the state for public affairs, and crimes that should be punished under criminal law [9]. Work-related crimes in a narrow sense can be understood as the relevant personnel engaged in public affairs according to law, in the process of performing work tasks, using the convenience or influence of their positions to carry out illegal activities, and should be punished according to the criminal law. personnel, personnel appointed by state-owned units, and other relevant personnel engaged in official duties under the law [10]. According to the statistics of the regions where the captured work-related crimes occurred, the overall distribution trend is that the number of work-related crimes in coastal areas is relatively high, and the ten provinces with relatively high occurrences are Zhejiang, Beijing, Shandong, Guangdong, Henan, Jiangsu, Hebei, Hubei, Fujian, Liaoning; the average age of the persons involved in the case is 46.9 years old, of which 20.8% are over 51 years old, 52% are 41-50 years old, 22.1% are 31-40 years old, and 30 years old or less 5.2%, high-level cadres have serious crimes on work-related, which is close to half of the total number of people who have been investigated [11].

Using big data thinking to determine the high-incidence links of work-related crimes in the field of construction engineering, search keywords "2019-2023", "construction industry", and "criminal cases" in the Alpha intelligent case database, and the results show that there are 431 entries from Chinese senior People's Courts The cases of rulings issued by intermediate people's courts, basic people's courts and special people's courts not only have detailed information but also cover work-related crimes in all links, which are very representative. Among them, the bidding process is the largest The highest proportion accounts for about 40% of the total number of cases, followed by land acquisition and demolition, which accounts for about 25% of the total number of cases, about 15% for approval, about 10% for procurement, and about 5% for completion acceptance, engineering construction accounted for about 5% [12]. Therefore, bidding links, land requisition, and demolition links, approval links, and procurement links are high-incidence links of work-related crimes in the field of construction engineering. As shown in Figure 2:

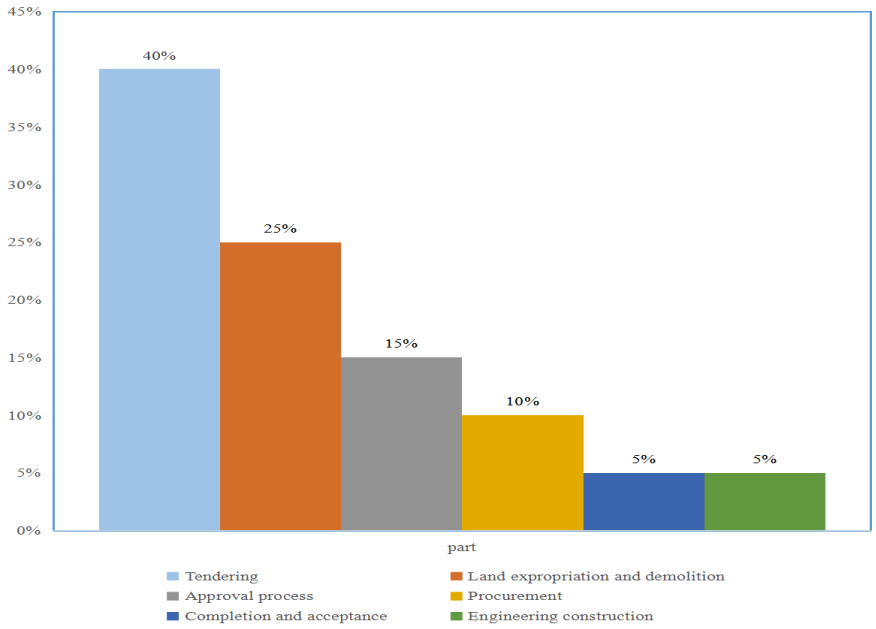

Fig. 2. Proportion of work-related crimes in various links in the field of construction engineering (Source drawn by this paper)

Due to the huge capital investment in the field of engineering construction, the huge profit margins, the many departments and links involved, and the intricate inter-interest relationship, it is

an area prone to high-incidence crimes such as corruption, bribery, and dereliction of work-related [13]. The project process can be divided into the pre-planning stage, project implementation stage, and completion acceptance stage, including project approval, planning site selection, bidding, material and equipment selection, construction, supervision, quality inspection, completion acceptance, project payment, capital operation, etc. Each link has a huge amount of capital flow, and its existence is extremely attractive. These characteristics make the engineering construction industry a field where work-related crimes are prone to and frequently occur [14]. Through the research and analysis of a series of engineering project cases, the risk of integrity may exist in various stakeholders, stages, and links of major projects, while the main risks are concentrated in non-public operation links, weak management links, power positions, and funds. The asset management part is the early stage of the project [8]. Combining international and Chinese scholars' research on the high-incidence links of work-related crimes in the field of construction engineering, to make this paper have a wider reference value, this paper focuses on the research on work-related crimes in the early stage of construction engineering, which is also the stage with the highest incidence of work-related crimes in the field of construction engineering in the world. As mentioned in the research scope of the first chapter of this paper, the early stages of construction projects studied in this paper include planning approval, land acquisition, and demolition, bidding, design and design changes, equipment procurement, that is, all links before construction, construction the middle stage and completion acceptance stage are out of the research scope of this paper. At this stage, there are many difficulties in the procuratorate's special prevention, such as single prevention means, unfamiliarity with the construction project operation process and norms, inadequate implementation of measures, etc., and putting forward corresponding countermeasures, starting from the perspective of treating both symptoms and root causes, to construct a scientific and reasonable engineering construction field the long-term prevention mechanism of work-related crime [13].

The use of big data analysis technology to fight crime is a development trend, and there are already successful examples in the world of using big data to effectively fight crime. For example, the Los Angeles Police in the United States cooperated with researchers to predict earthquakes based on variants of earthquake prediction algorithms and crime data The probability of crime occurring, accurate to within 500 square feet, has seen a 33% and 21% drop in the distribution of larceny and violent crime in Los Angeles areas where the algorithm is used [15]. With the rapid development of modern information technology represented by big data, people's production, life, and way of thinking are undergoing a major change, while the application of data to judicial work, especially the investigation of work-related crimes in the field of construction engineering, is still in the development stage [16]. Occupational crimes in the construction field are frequent, and various methods have been tried for prevention and governance, but the results are not satisfactory. The use of big data technology is expected to reduce the interference of human factors in the construction field and reduce the occurrence of corruption [17]. The characteristics of big data have been studied in various publications. McKinsey, a world-leading global management consulting firm founded by James Oscar McKinsey, a professor of accounting at the University of Chicago, defines the characteristics of big data as "those that exceed the size of a typical database. Datasets capable of being captured, stored, managed and analyzed by software tools" [18]. Gartner, an American information technology research and consulting company, defines the characteristics of big data as 3Vs: Volume, Velocity, and Variety [19]. By increasing accuracy, these 3 Vs gradually expand to 5Vs as the five characteristics of big data, namely Volume, Velocity, Variety, Veracity, Value: Volume poses the most direct challenge to the traditional IT structure, from the ability to process large amounts of information The benefits obtained in big data analysis are the main attraction of big data analysis; Velocity refers to the speed at which new data is generated and the speed at which data is

moved, and relational databases process, store and analyze data at an increasing speed; Unstructured data greatly increases the complexity of storing and analyzing big data; Veracity means that the quality of the captured data may vary greatly, and the accuracy of the analyzed data depends on the accuracy of the source data; Value is the most important in big data On the one hand, while the potential value of big data is huge, it is useless unless it can be converted into value, implementing IT infrastructure systems to store big data becomes very expensive, and businesses will need a return on investment [20]. The five characteristics of big data are summarized in Figure 3 below:

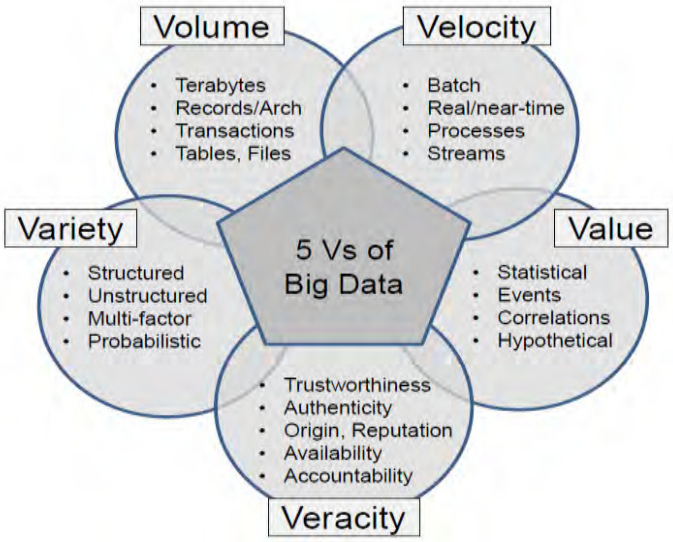

Fig. 3. 5Vs of Big Data[20]

With the continuous development of science and technology in China, information technology has become more and more widely used at all levels of society. For example, electronic documents, electronic payment, Weibo, WeChat, etc. are all derivatives of information technology, and more and more cases of work-related crimes are involved Among them, people's understanding of traditional witness and material evidence has been changed in the past. People are more looking for relevant information collection means, and big data technology is one of the important means of evidence collection today. Big data technology has real-time and richness. It can automatically track and find criminal evidence to a large extent, and it can also promote the reform of the policing model and the research of social hotspots [21]. At present, discipline inspection and supervision agencies are paying more and more attention to the combination of big data concepts and prevention work, but the traditional thinking pattern has not been completely broken through. The big data concept guides the transformation of the prevention work model and firmly establishes the concept of data priority, data analysis, and data security idea [22]. With the application of big data technology more and more widely and in more and more fields, in the financial transaction industry, the application of big data algorithms to transaction decisions is becoming more and more popular. In terms of meeting customer service needs and business process optimization, big data analysis and prediction capabilities of technology have also become irreplaceable. Compared with the wide application of big data technology in the economic, medical, and social fields, the improvement of security and law enforcement, especially the judicial field, is only in the initial stage of big data technology. In the judicial field, more certainty and objectivity are required, rather than analysis and prediction capabilities, so the application of big data technology is not widespread. At present, the application of big data technology thinking is mainly reflected in two tasks in the work of work-related crime prevention the first is the inquiry of bribery files: the procuratorial organs have classified, organized,

entered, stored, and established a database through the bribery cases and other information made by the people's courts. By accepting inquiries from social units and individuals, it can be used in cases where bribery may be more The method of setting up a "firewall" between large units and individuals and units and individuals that may accept bribes is to curb bribery crimes and then promote the construction of a social integrity system; the second is to punish and prevent annual reports: Conduct a comprehensive summary of the work-related crime cases investigated and dealt with, and summarize the occurrence characteristics, cause rules, and development trends of work-related crime cases within the year through analysis, statistics, induction, etc., and propose preventive countermeasures. Through layers of data statistics, research, and judgment, Summarize and achieve a high degree of attention to the areas of prone and frequent work-related crimes, and prevent the recurrence of work-related crimes [23]. Big data is the trend of the times. From small data to massive data to big numbers, and even to infinite data, every stage of technological innovation is intertwined with infinitely moving value guidance and continuous research investment, which will also bring constant beating Anxiety and threat factors. The development of big data science and technology is like a double-edged sword. While bringing opportunities to crime prevention, it is also accompanied by many challenges.

Based on the application of big data in various industries in China and the development trend of work-related crimes in the field of construction engineering, consult the authoritative Chinese literature website HowNet, and use the keywords "big data application", "big data crime prevention" and "big data work-related crime prevention" After the search of "big data prevention of work-related crimes in construction projects", it was found that there were 28,641 papers on the application of big data, 60 papers on big data crime prevention, 12 papers on big data work-related crime prevention, and even more papers on big data prevention of work-related crimes in construction projects. In particular, big data for the prevention of work-related crimes in the early stages of construction projects is almost still in a blank state, which once again verifies the writing value and theoretical significance of the topic of this paper.

The Research Framework is shown in Figure 4. Sources are shown in Table 1.

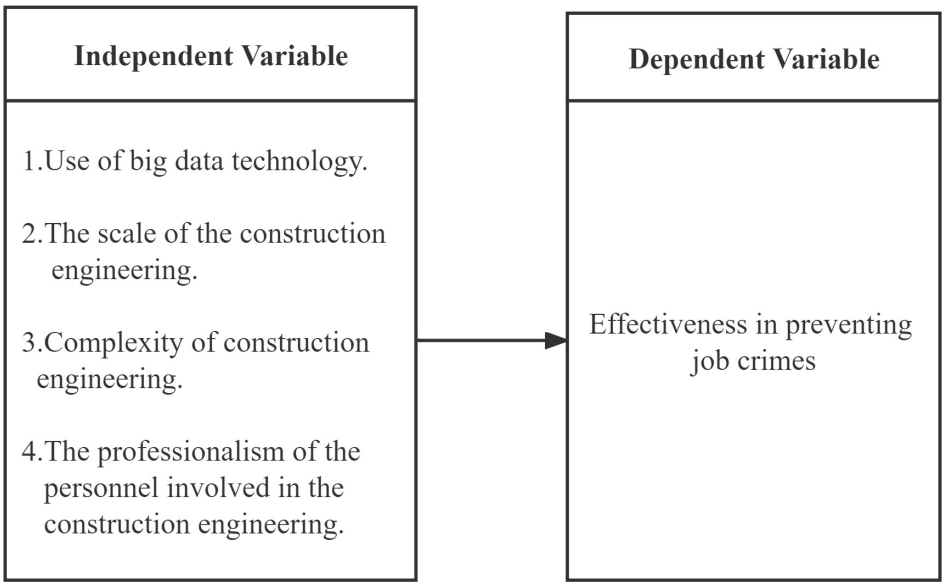

Fig. 4. Research Framework (Source drawn by this paper)

Table 1
Summary of dependent variable sources
(Source drawn by this paper)

Item	Source
1. Use of big data technology	[21,22,24,25,26]
2. The scale of construction engineering	[27]
3. Complexity of construction engineering	[28]
4. The professionalism of the personnel involved in construction engineering	[29] [30]

3. Methodology

3.1 Research Approach

The research method chosen in this paper is quantitative analysis. The reason for deciding to use quantitative research is that it is necessary to systematically and objectively explore the role of big data technology in the prevention of occupational crimes in the early stage of engineering projects. By using quantitative analysis, the relationship between variables can be quantified. To measure the impact of big data technology on the prevention of work-related crimes in the early stages of construction projects, and extend the findings to a larger population [31]. To investigate the project-based organization of bidding activities in China's construction industry and the causal complexity of corruption, also using quantitative analysis, a total of 400 Chinese construction industry participants were randomly recruited to complete the survey [32]. The rationale for choosing a quantitative assay for this research stemmed from its objectivity, precision, statistical analysis power, generalizability, advantages of large sample size, measurement and operationalization advantages, testability of hypotheses, time and cost efficiency, and ability to facilitate comparisons. Capability and benchmarking, these factors align with the research objectives and enhance the rigor and reliability of the research results.

3.2 Research Design

Overall, this paper has the following phases: The first phase is a review of existing research, reviewing existing research related to big data technology, occupational crime, and their intersection in the construction industry, and identifying key theories, concepts, and research in this field blank, formulate a conceptual framework to guide the research; the second stage is the research objectives and research questions, clarify the research objectives and research questions based on the review of existing research, ensure that the objectives and questions are consistent with the focus of the research, and address the findings in the literature the third stage is the method and sampling design, determine the quantitative analysis, select simple random sampling, determine the research area as a Chinese city, and establish the sampling frame, use appropriate statistical considerations to determine the sample size; the fourth stage is data collection, Develop a questionnaire tool, collect relevant data from selected construction projects, pre-test the questionnaire to ensure clarity, reliability and validity, conduct questionnaire surveys on a sample of confirmed construction projects, collect the use of big data technology, the occurrence of occupational crimes , project characteristics and other relevant factors; the fifth stage is data analysis, cleaning and organizing the collected data,

performing inferential statistics to summarize the characteristics and variables of the sample, conducting quantitative analysis, analyzing the relationship between variables, and evaluating big data technology Impact on the prevention of occupational crime in the early stages of construction engineering; the sixth stage is results and findings, presenting the results of the data analysis in a clear and organized manner, explaining the findings in relation to the research objectives and research questions, discussing the implications of the results and their implications for the construction industry understands the role of big data technology in the prevention of occupational crime; the seventh stage is discussion and conclusion, analyzing the research results in the context of existing research, discussing the strengths, limitations and implications of the research, for the construction industry stakeholders Provide recommendations for future research and practical impact, and summarize the main findings and conclusions of the research. The overall research flow chart is as Figure 5:

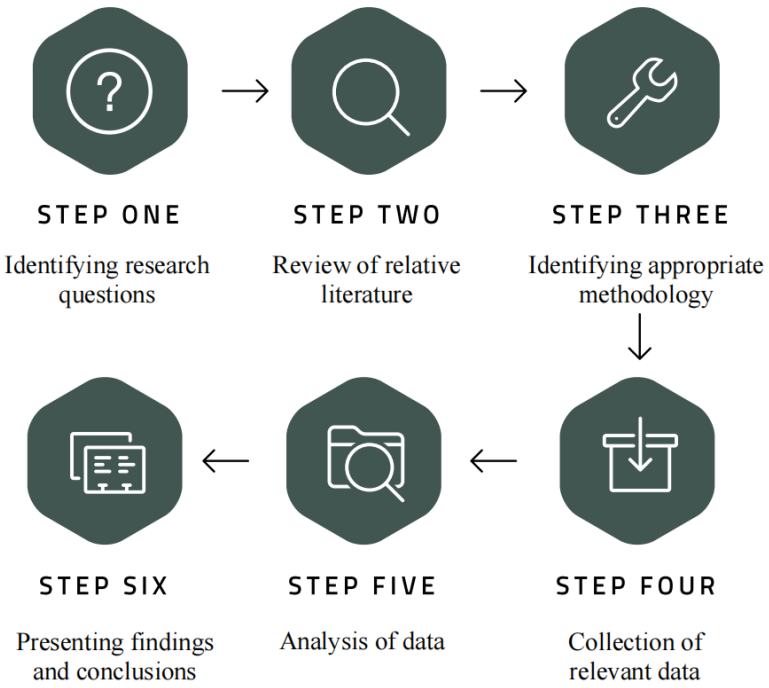

Fig. 5. Overall research flowchart [33]

3.3 Research Area

The research field of this paper focuses on the prevention of work-related crimes in the early stages of construction projects. The relevant laws for work-related crimes are based on Chinese criminal law. At the same time, considering the importance and prevalence of work-related crimes at this stage, this research is carried out in Chinese cities. Choose China the reason why cities are used as research areas, as explained in the first and second chapters, is that with the rapid development of China's economy, the construction industry is a pillar industry of China's national economy, and there are various construction projects in Chinese cities. They vary in size, complexity, and geographical location. As in Chapter 2, the statistics are based on the regions where the captured work-related crimes occurred. The overall distribution trend is that the number of work-related crimes in coastal areas is relatively high, and the number of cases is relatively high The ten provinces in China are Zhejiang, Beijing, Shandong, Guangdong, Henan, Jiangsu, Hebei, Hubei, Fujian, and Liaoning. The research area of this paper mainly focuses on these 10 provinces. In addition, the

research field covers many aspects, including Identifying and exploring different types of work-related crimes that are common in the early stage of construction projects, including bribery, corruption, fraud, illegal bidding, collusion, and other crimes.

3.4 Data Collection

The main method of data collection in this research methodology is the questionnaire. The questionnaire has been carefully designed to ensure the clarity, validity, and reliability of the data collected, as well as the consistency with the research objectives. The structure of the questionnaire consists of multiple choice The structured question composition of the response scale, including multiple-choice questions and predefined answer options, covers various aspects related to the research objectives, utilization of big data collection techniques, types of occupational crimes, preventive measures implemented and project characteristics, etc. Data on various aspects, a multiple-choice response scale allows respondents to choose the option that best represents their opinion or degree of agreement. The sampling method chosen for this research is simple random sampling. Simple random sampling ensures that each element in the target population has an equal opportunity to be selected as a sample. By randomly selecting individuals from different construction projects and companies, the sample can better reflect the diversity and characteristics of the overall population, avoiding any bias or partiality in the selection process, and each potential participant has an equal chance of being included in the sample, regardless of their characteristics or affiliation [34]. The simple random sampling flowchart in this study is shown in Figure 6.

The target sample of this paper includes a wide range of construction projects in Zhejiang, Beijing, Shandong, Guangdong, Henan, Jiangsu, Hebei, Hubei, Fujian, and Liaoning provinces in China, which can include various types of buildings, such as residential buildings, commercial complexes, foundation facility development, industrial facilities, and public infrastructure. These projects vary in scale, scope, complexity, and stakeholders involved. Relevant personnel in the specific project link of land acquisition and demolition, approval, bidding, and procurement determined according to the research objectives, to further determine the sample population as construction workers, construction companies, relevant stakeholders, and other professionals: including professionals who work closely with the construction industry or have professional knowledge. Such as architects, engineers, urban planners, safety inspectors, and consultants.

A sampling frame is the basis for selecting individuals as a sample in a simple random sampling process, each individual in the sampling frame is assigned a unique numerical code to facilitate the random selection process [35]. The authors used a random number generator to randomly select individuals from the sampling frame and invite them to participate in the questionnaire. The sampling frame was constructed by gathering information from various sources, including construction firms, trade associations, professional networks, and relevant government agencies. The research's sampling frame is a comprehensive list of persons involved in the early stages of construction projects in specific regions of Zhejiang, Beijing, Shandong, Guangdong, Henan, Jiangsu, Hebei, Hubei, Fujian, and Liaoning. Incorporating ethical considerations, the sampling frame includes the following information for each individual but does not involve personal privacy, examples are as shown in Table 2.

Fig. 6. Simple random sampling flowchart (Source drawn by this paper)

Table 2
Sample sampling frame table (Source drawn by this paper)

Serial Number	Gender	Age	Education Level	Position	Working Years	Province Where the Project is Located	Consent Status
1	Female	35	Master	Architect	10	Beijing	Consented
2	Male	43	Undergraduate	Project Manager	21	Jiangsu	Consented
.....	.....	.....		.....	.....	.....	.....

In the academic literature, the sample size is widely discussed. Choosing the correct sample size remains a great challenge for researchers. Larger sample sizes lead to ethical issues, time consumption, and financial waste, while smaller sample sizes of the research will affect the validity of the research [36]. Similarly, authors such as Collis and Hussey argue that sample size is based on analytical methods such as further analysis of confirmatory factor analysis, structural equation modeling, stochastic modeling with latent variables, structural path analysis, and multiple regression analysis that must be treated with caution [37]. In the case of analytical method studies, some studies have shown that by obtaining samples of more than 300 subjects, the estimated statistic derived from the sample may be the same as the true value in the expected population [38]. The sample size of this research was chosen according to the most important rules of thumb cited in academic research results, and the confidence level was used as a measure of the sample size, choosing a random sample to ensure unbiased conduct. This research was chosen for the calculation of simple random The formula for sampling sample size is the "sample size calculation formula", which is a standard formula used in statistics and research methods to determine the appropriate sample size needed to estimate a population parameter with a desired level of confidence and precision.

The formula is: N = (Z^2 * P * (1 - P)) / (E^2)

N: represents the sample size.

Z: This is the Z value associated with the selected confidence level, the Z value corresponds to the level of confidence or certainty required in the estimate. Combining the purpose of this research and the research time constraints, a 95% confidence level is chosen, and the Z value is 1.96.

P: Estimated proportion of the variable. A maximum variability of 0.5 is assumed.

E: This represents the desired margin of error or the maximum acceptable difference between the estimated and true proportion of the population. The margin of error defines the precision of the estimate, and this research defines E as 5%.

Calculate the sample size required to achieve the desired confidence and margin of error in estimating the proportion by substituting the values for Z, p, and E into the formula, N = (1.96^2 * 0.5 * (1 - 0.5)) / (0.05^ 2) ≈384.16, the calculated sample size is about 384.16. Since the sample size should be an integer, round it up to the nearest integer. Therefore, the recommended sample size for this research is 385. This size reflects the number of obtained responses, not the numbers distributed by questionnaires, postal and email surveys rarely have a 100% response rate in most social and managerial surveys [39]. Perhaps the most common and efficient way to ensure that the minimum sample size is met is to increase the sample size by up to 50% in the first allocation of the survey [40]. Based on the time limit of this research, the sample size was increased by 30%. Therefore, the number of questionnaires distributed in this research was set to 385+385*30%=500.5. According to the integer calculation, the total number of questionnaire samples should be 501.

Macer and Wilson [41] set the median acceptable length of online questionnaires as 15 minutes. The time required for different questionnaire structures, question types, and topics is different. Scale questions are easier to answer than open questions, so the best way is to estimate the time rather than the number of questions. For online surveys, try to keep the duration within 10 minutes, which is about 10-20 questions [42]. There should not be too many questions, and there are not many questions. The higher the willingness of the subjects to fill in, the greater the possibility of reflecting the real potential characteristics. Too long questionnaires often cause psychological boredom or fear of difficulties in the respondents, which affects the quality of the answers. Recycling reliability and validity, of course, the questions should not be too small, otherwise, it will not be able to fully cover all aspects of the research, which will affect the results of the questionnaire [43]. Based on the research questions and research purposes of this paper, out of respect for the time and energy of the participants, the questionnaire time is defined as less than 10 minutes, and the number is set to

14. The distribution of questionnaires in this research will be completed through two channels: email and the Chinese social platform WeChat.

4. Results

The data collection time is two weeks, from June 5, 2023, to June 19, 2023. Use questionnaire star software to generate a link to the questionnaire, and state in the link that this questionnaire is for academic use only, and the research of this paper purpose and mechanisms of ethical consideration. Following Roberta Sammut, Odette Griscti, Ian J. Norman [44] Efforts to Improve Response Rates to Web-Based Questionnaires, Email Advance Notification, Email Invitation, 2 Reminders, and Ease of Access Simple design followed by readers, semi-automated login improves response rates for web surveys, web surveys that take less than 10 minutes to complete are more likely to have higher response rates than longer surveys [44]. The author indicated in the link that this questionnaire survey only takes about ten minutes to complete to encourage participants to spare time to complete it. In addition, the author also thanked the participants for participating in this research. A simple random Sampling method, randomly sent to the research objects in the sampling frame by email and WeChat. In the third chapter, 501 questionnaires are planned to be sent out. As of June 19, 2023, a total of 285 questionnaires have been received, as shown in Figure 7, the response rate was 56.89%. The appropriate response rate for the questionnaire may vary depending on the context and purpose of the survey, there is no universal definition or standard response rate applicable to all types of questionnaires, however, a relatively high survey response rate is important Because it reduces sampling bias problems and increases the validity of survey-based research findings [45].

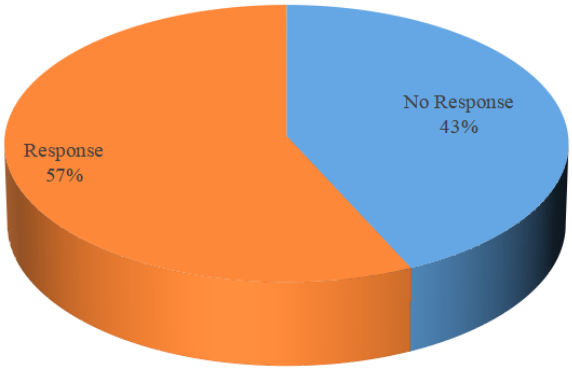

Fig. 7. Questionnaire response rate graph (Source drawn by this paper)

4.1 Descriptive Analysis

Perform descriptive statistical analysis on the single-choice questions with serial numbers 1-16, and analyze statistical data such as the mean, median, mode, standard deviation, and variance of the 11 single-choice questions. As shown in attachment 1 for the questionnaire. As Table 3 follows:

Table 3
Single-choice topic for descriptive statistical analysis table (Source drawn by this paper)

Variable	Mean	Median	Std. Deviation	Variance
1. Gender:	1.49	1	0.501	0.251
2. Age:	1.85	2	0.889	0.791
3. Education:	2.78	3	1.034	1.069
4. Years of experience in the construction industry:	1.55	1	0.678	0.459
5. The area where you are currently working is:	5.74	6	2.665	7.103
6. Is work-related crime more common in the early stages of construction projects?	3.77	4	1.092	1.193
8. In the early stage of the construction project, how often did you observe work-related crimes?	3.05	3	0.84	0.706
11. Does your company implement big data technology in the field of construction engineering?	1.65	2	0.479	0.23
13. To what extent do you think big data technology has played a role in the investigation and prevention of work-related crimes in the early stages of construction projects?	3.76	4	1.071	1.147
15. How do you evaluate the current level of understanding and adoption of big data technology in preventing work-related crimes in the construction industry?	2.11	2	0.991	0.982
16. Combining big data technology with existing security measures can strengthen the prevention of work-related crimes.	3.79	4	1.092	1.193

The results of the questionnaires show that the utilization rate of big data technology in the field of construction engineering is significantly lower, accounting for only 35.44% of the company. The status quo of the construction industry shows that in the context of preventing occupational crimes, people generally lack understanding and limited applications of big data technology. Survey data show that most construction companies have not fully adopted big data technology, which may prevent them from effectively handling professional crimes. Because only a few companies use big data analysis, the industry seems to miss the potential benefits and improvements that this technology can provide. Define multiple response sets for multiple-choice questions with serial numbers 7-19, and perform frequency analysis.

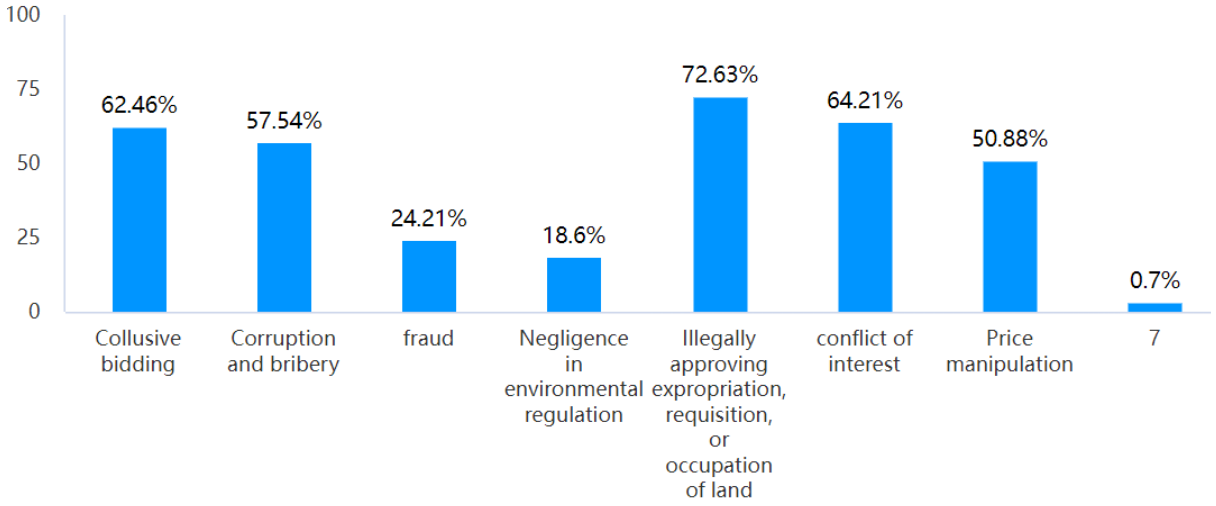

Fig. 8. Common types of work-related crimes in the early stage of construction projects (Source drawn by this paper)

As shown in Figure 8, the findings of the questionnaire reveal key insights into the types of work-related crimes that are more likely to occur in the early stages of construction work. Affected by bid rigging, corruption, and bribery were highlighted as prominent issues that could undermine fair competition and transparency in the project procurement process. In addition, illegally sanctioned expropriations, applications, and land occupations are potential threats to property rights and moral governance. The presence of conflicts of interest could compromise decision-making and lead to bias, while price manipulation was identified as a potential concern affecting project cost and financial integrity. The findings of the survey underscore the need for enhanced regulatory measures, codes of ethics, and vigilant monitoring to tackle and prevent workplace crime.

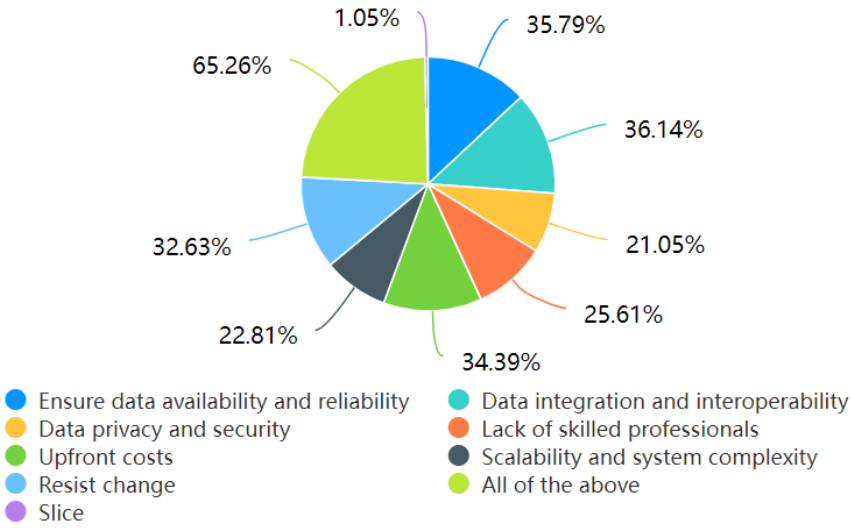

Fig. 9. The main challenges of implementing big data technology in the construction industry to prevent work-related crimes
(Source drawn by this paper)

From the survey results shown in Figure 9, the low adoption rate may be attributed to a variety of factors, including a lack of awareness, many construction companies may not be aware of the benefits and application of big data technology in fighting professional crime. Lack of this understanding may prevent them from investing in such solutions; cost and resource constraints, implementing big data technologies often requires significant financial resources and expertise that may be difficult for smaller construction businesses to afford or obtain; Resistance to change, traditional ways of thinking, and resistance to the use of new technologies may be prevalent in the construction industry, making it difficult for companies to accept big data solutions; Data security issues, some companies may have concerns about storing and managing sensitive data, fearing potential breach and cyber-attacks; and lack of regulatory push, lack of regulatory requirements or incentives to prevent career crimes in big data technologies could lead to complacency in the industry.

4.2 Prevention Strategies Using Big Data Technologies

Construction companies should establish robust data collection and management systems to efficiently capture relevant information. This includes implementing IoT devices, sensors, and other data sources to create a comprehensive data ecosystem. Emphasis is placed on integrating big data analytics into the construction process to leverage real-time insights for better decision-making.

Leverage machine learning algorithms to identify patterns and predict potential work-related crime and safety risks. Prioritize data security by employing encryption, multi-factor authentication, and access controls. Collaborate with data experts to establish industry-wide data protection best practices. Provides training programs for construction professionals to improve their data literacy and analytical skills. Encourage a data-driven culture to facilitate the seamless adoption of big data technologies.

Using big data technologies to prevent occupational crimes in the early stages of construction projects can help improve transparency, accountability, and fair practices. Here are strategies to take advantage of crime prevention by work-related:

i. Data Collection and Integration: Collect data from various sources, including project budgets, financial transactions, procurement processes, contractor information, and employee records. Integrate this data into a centralized database for analysis.
ii. Data Analysis and Anomaly Detection: Utilize big data analytics and machine learning algorithms to analyze integrated data. Look for anomalies or patterns that could indicate potential damage, such as unusually high project costs, irregular payments, or unusual procurement behavior.
iii. Financial Auditing: Implement automated financial auditing using big data analytics to monitor financial transactions and detect any suspicious activity or discrepancies.
iv. Risk Profiles: Create risk profiles for contractors, suppliers, and project stakeholders based on their historical data and behavior. This helps identify high-risk entities that may be susceptible to corrupt practices.
v. Real-time monitoring and alerting: Set up a real-time monitoring system to track financial transactions, expenses, and purchasing activity. Implement an alert mechanism to notify the appropriate authorities of any potential red flags.
vi. Transparency and Open Data: Increase transparency by making relevant project data and financial information available as much as possible. Open data initiatives can deter corruption and promote public accountability.
vii. Whistleblower Hotline: A confidential whistleblower hotline has been set up to encourage employees and stakeholders to report suspected corrupt or unethical behavior.
viii. Conflict of interest detection: Use big data analytics to identify potential conflicts of interest among project stakeholders, ensuring that those with no vested interest have undue influence on project decisions.
ix. Geospatial analysis: Leverage geospatial data to identify areas with a history of criminal activity such as corruption or fraud, which can help allocate resources and scrutiny accordingly.
x. Cooperate with Anti-Corruption Agencies: Cooperate and share relevant data with anti-corruption and other official crime agencies to support their investigative and law enforcement efforts.
xi. Data Visualization and Reporting: Use data visualization tools to present the results of data analysis in a clear and understandable format for project managers, stakeholders, and professional crime agencies such as anti-corruption.
xii. Ethics Training and Awareness: Conduct regular ethics training and awareness programs for all project stakeholders, emphasizing the importance of integrity and zero tolerance for crimes committed work-related.
xiii. Data Privacy and Security: Ensure appropriate data privacy and security measures are in place to protect sensitive information and prevent unauthorized access.

5. Conclusions

After the above research, the early stages of construction projects across China have seriously lacked big data implementation. Most construction companies lack system methods to collect and manage project data, which hinders the use of big data. The integration of big data technology and existing construction processes is still very low, resulting in missing opportunities for optimization. Personal concerns about data security and privacy are the main obstacles to adopting big data solutions in the construction industry. Interviewees acknowledge the potential benefits of big data technology, including enhancing security, preventing crimes, improving decisions, and reducing costs. This paper emphasizes the shortcomings of early big data technology in my country China. To reduce crimes related to work, improve security measures, and improve overall efficiency, the construction industry must use big data as an indispensable tool. The proposed strategic framework outlines the feasible steps that integrate big data technology into the construction process. By adopting this strategy, the construction industry can release the entire potential of big data, and it is more efficient in preventing work-related crimes in the early stages of construction engineering.

Although this research offers valuable insights and a strategic framework, certain limitations should be acknowledged: The survey was conducted in ten provinces in China and may not be representative of the entire construction industry in China. The limited sample size may affect the generalizability of the findings and may not reflect the diverse practices and challenges that exist in other regions. Survey responses can be biased because participants may provide responses based on personal perceptions or experience. Additionally, non-response bias may arise if certain construction firms or stakeholders choose not to participate in the survey. Surveys may be conducted for a limited time, limiting researchers' ability to reach larger audiences and conduct more comprehensive data analysis. The research focuses on assessing the level of big data application in the field of early construction engineering. It does not delve into the factors leading to limited adoption, nor does it address in detail the specific challenges faced by construction firms. The accuracy and reliability of survey data depend on the ability of respondents to provide accurate information. Errors or inconsistencies in the data may affect the conclusions of the research. The findings and proposed strategies are specific to the construction industry. Big data applications can vary widely across industries, and the portability of the strategy to other industries should be carefully considered. While the strategic framework presents feasible steps for big data integration, actual implementation may encounter various challenges such as financial constraints, resistance to change, and lack of technical expertise. The adoption of big data technologies has raised concerns about data privacy and security. This paper briefly discusses this issue, but a more comprehensive analysis of data protection measures and potential legal implications would be beneficial. Despite these limitations, this research provides a fundamental understanding of the current state of big data applications in early construction engineering and provides a valuable starting point for future research and practical implementation efforts. Researchers and industry practitioners should consider these limitations when interpreting the results and applying suggested strategies.

Acknowledgment

This research was funded by School of Housing, Building and Planning, Universiti Sains Malaysia.

References

[1] China Construction Industry Association. "Statistical analysis of construction industry development in 2020." (2021).

[2] L. Jian, Investigation on the Prevention of Occupational Crimes in the Field of Engineering Construction., People's Procuratorate(14) (2015) 53-55.

[3] L. Jingjing, D. Lu, L. Jue, Z. Wenxi, Research on the combat strategy of engineering job crimes based on dynamic meta-network, Journal of Railway Science and Engineering (2021).
[4] D. Jie, Engineering project management model based on the perspective of sustainable development, MANAGER'JOURNAL (2023) 96-97.
[5] W. Yanxiang, Analysis on the application of big data in the investigation of duty crimes, Modern Business Trade Industry(01) (2023) 203-206.
[6] G. Zhe, The Epistemology of Investigating Duty-Based Crimes in the Era of Big Data, Politics and Law(03) (2019) 42-54.
[7] Q. Zhong, W. Xinge, L. Chuan, Research on the "Trinity" Enterprise Occupational Crime Prevention System——Taking the Field of Engineering Construction as an Example, Legal System and Society(35) (2016) 195-196.
[8] J. GenYou, Reflections on the Prevention of Duty-Based Crimes in State-owned Enterprises in the Field of Major Projects, Prosecutors(08) (2018) 60-61.
[9] Criminal Law of the People's Republic of China, China, 1997.
[10] W. Shanshan, Exploring the Causes and Preventive Measures of Duty Crime, LEGALITY VISION (2017) 61-62.
[11] Y. Guoming, L. Biao, Y. Ya, Research on the Current Situation, Characteristics and Countermeasures of Anti-corruption Public Opinion on Duty-Based Crimes, JAC Forum (2019) 153-157.
[12] Criminal cases in the construction industry in the past five years, 2019-2023.
[13] J. Yuanyuan, Talking about how the procuratorial organs can effectively carry out the special prevention of major engineering projects, Legal System and Society, 27. (2016).
[14] Z. Cheng, Analysis of how to carry out job crime prevention work in engineering construction, How to Establish a Scientific Decision-Making Mechanism Theory Symposium - Decision Forum" Proceedings (Part 2). (2015) 126.
[15] Z. Jun, A Preliminary Study on the Transformation of the Work Path of the Initial Investigation of Job-related Crimes in the New Era——From the Perspective of Big Data Technology, Chinese Prosecutor, (24) (2015) 57-62.
[16] H. Donglin, Using big data to promote the transformation and upgrading of duty crime investigation, People's Procuratorate(14) (2016) 44-46.
[17] Z. Yaohua, Analysis on network prevention and control management of occupational crimes in the construction field, Shandong Trade Union Forum, 21(03) (2015) 76-78.
[18] Sun, Zhaohao, and Paul P. Wang. "Big Data, Analytics, and Intelligence: An Editorial Perspective." *New Math. Nat. Comput.* 13, no. 2 (2017): 75-81. https://doi.org/10.1142/S179300571702001X
[19] GARTNER, IT. "glossary: Big Data.[Sl]: Gartner Group, 2014." (2018).
[20] Anuradha, J. "A brief introduction on Big Data 5Vs characteristics and Hadoop technology." *Procedia computer science* 48 (2015): 319-324. https://doi.org/10.1016/j.procs.2015.04.188
[21] L. Chao, Analysis on Ways and Methods of Duty Crime Prevention under the Background of Data, LEGALITY VISION (2017) 124.
[22] N. Jianping, Application of Big Data and Innovation of Duty Crime Prevention, Journal of Shanxi Provincial Committee Party School of C.P.C (2018).
[23] Q. Yixin, Occupational crime prevention under big data technology, Legal System and Society (2016).
[24] L. Lijuan, Discussion on the application of big data in the prevention of job crimes, 2018 2nd National Prosecutors Reading Essay Competition Award-Winning Essays (2018).
[25] W. Yanxiang, Analysis on the application of big data in the investigation of job crimes, Modern Business Trade Industry (2023).
[26] Reja, Varun Kumar, and Koshy Varghese. "Impact of 5G technology on IoT applications in construction project management." In *ISARC. Proc. Int. Symp. Autom. Robot. Constr*, vol. 36, pp. 209-217. 2019. https://doi.org/10.22260/ISARC2019/0029
[27] W. Jianguo, M. JuNian, Thoughts on Promoting the Construction of Corruption Prevention and Supervision Mechanism in the Field of Engineering Construction with the Rule of Law, Forum on Criminology (Volume III) (2016) 485-490.
[28] S.K. Sears, G.A. Sears, R.H. Clough, J.L. Rounds, R.O. Segner, Construction project management, John Wiley & Sons 2015.
[29] W. Shihuai, Discussion on the management strategy of construction project, Chinese and foreign entrepreneurs (2019) 103.
[30] Mok, Ka Yan, Geoffrey Qiping Shen, and Jing Yang. "Stakeholder management studies in mega construction projects: A review and future directions." *International journal of project management* 33, no. 2 (2015): 446-457. https://doi.org/10.1016/j.ijproman.2014.08.007
[31] Sun, Heyuan, and Zhiyuan Liu. "Risk Prevention And Control And Construction Path Of Criminal Compliance Of Construction Engineering Corporates." *Asian Law & Public Policy Review* 7 (2022): 54-80. https://doi.org/10.55662/ALPPR.2022.702

[32] Wang, Zhixiu, Junying Liu, and Xinya Guan. "Investigating the causal complexity of corruption in project-based organizations: a scenario analysis of bidding activity in the Chinese construction sector." *Engineering, Construction and Architectural Management* 29, no. 10 (2022): 3893-3916. https://doi.org/10.1108/ECAM-10-2020-0841
[33] Creswell, John W., and J. David Creswell. *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications, 2017.
[34] Igwenagu, Chinelo. *Fundamentals of research methodology and data collection*. LAP Lambert Academic Publishing, 2016.
[35] Mishra, Shanti Bhushan, and Shashi Alok. "Handbook of research methodology." (2022).
[36] Chander, N. Gopi. "Sample size estimation." *The Journal of the Indian Prosthodontic Society* 17, no. 3 (2017): 217. https://doi.org/10.4103/jips.jips_169_17
[37] Rahi, Samar. "Research design and methods: A systematic review of research paradigms, sampling issues and instruments development." *International Journal of Economics & Management Sciences* 6, no. 2 (2017): 1-5.
[38] Bujang, Mohamad Adam, Nadiah Sa'at, A. Reena Joys, and Mariana Mohamad Ali. "An audit of the statistics and the comparison with the parameter in the population." In *AIP Conference Proceedings*, vol. 1682, no. 1. AIP Publishing, 2015. https://doi.org/10.1063/1.4932510
[39] Taherdoost, Hamed. "Sampling methods in research methodology; how to choose a sampling technique for research." *How to choose a sampling technique for research (April 10, 2016)* (2016). https://doi.org/10.2139/ssrn.3205035
[40] Taherdoost, Hamed. "Determining sample size; how to calculate survey sample size." *International Journal of Economics and Management Systems* 2 (2017).
[41] Schönbrodt, Felix D., and Marco Perugini. "At what sample size do correlations stabilize?." *Journal of Research in Personality* 47, no. 5 (2013): 609-612. https://doi.org/10.1016/j.jrp.2013.05.009
[42] Revilla, Melanie, and Carlos Ochoa. "Ideal and maximum length for a web survey." *International Journal of Market Research* 59, no. 5 (2017).
[43] Design, Questionnaire. "How to Plan, Structure and Write Survey Material for Effective Market Research (Market Research in Practice Series)(Paperback) by Ian Brace; 289 pages."
[44] Sammut, Roberta, Odette Griscti, and Ian J. Norman. "Strategies to improve response rates to web surveys: a literature review." *International Journal of Nursing Studies* 123 (2021): 104058. https://doi.org/10.1016/j.ijnurstu.2021.104058
[45] Pedersen, Mogens Jin, and Christian Videbæk Nielsen. "Improving survey response rates in online panels: Effects of low-cost incentives and cost-free text appeal interventions." *Social Science Computer Review* 34, no. 2 (2016): 229-243. https://doi.org/10.1177/0894439314563916