Integrating Traditional Two-way Fixed-effects Models with KAN and GCN-Enhanced Deep Learning Models: Exploring the Impact of Digital Inclusive Finance on High-Quality Development of Enterprise

Dedai Wei[a, +], Zimo Wang[b, +], Hanfu Kang[c, +], Xinye Sha[d, *], Yiran Xie[e], Anqi Dai[f], Kaichen Ouyang[g, *]

(Dedai Wei, Zimo Wang, and Hanfu Kang contributed equally to this work.)

[a]College of Economics, Shenyang University, Shenyang 110000, China

wdd31212@163.com

[b]Department of Philosophy, Nankai University, Tianjin 300381, China

miemie9999@sina.com

[c]Department of Mathematics, Kings College London, WC2R 2LS, London, United Kingdom

k21018922@kcl.ac.uk

[d]Graduate School of Arts and Sciences, Columbia University, NY 10027, United States of America

xs2399@columbia.edu

[e]School of Humanities and Social Sciences, University of Science and Technology of China, Hefei 230026, China

yr_x@mail.ustc.edu.cn

[f]School of Finance,Dongbei University of Finance and Economics, Dalian 116000,China

dannqiii@163.com

[g]Department of Mathematics, University of Science and Technology of China, Hefei 230026, China

oykc@mail.ustc.edu.cn

*Corresponding author: Xinye Sha(xs2399@columbia.edu),
Kaichen Ouyang(oykc@mail.ustc.edu.cn)

Abstract: In the context of the ongoing global economic transformation, high-quality enterprise development (HQED), particularly through enhancing Total Factor Productivity (TFPLP), is crucial for economic growth. Digital Inclusive Finance (DIF), as a key financial innovation, optimizes resource allocation, reduces financing costs, and increases information transparency, thereby improving financing channels for enterprises and playing an important role in HQED. However, traditional financial models, which often rely on linear assumptions, struggle to capture the complex nonlinear relationships between DIF and TFP. Deep learning, with its capacity for nonlinear modeling, offers a promising solution to this limitation. This study first uses the traditional economic double fixed-effects model to perform causal inference on the relationship between DIF (Digital Inclusive Finance) and TFPLP (Total Factor Productivity), in order to explore the preliminary structural connection between the two. Further, this paper considers corporate internal control levels, financing constraints, and corporate innovation as potential mediating variables, conducting in-depth analysis to investigate the role these factors play in the relationship between DIF and TFPLP. Through the linear analysis results provided by traditional models, we initially confirm the association between DIF and TFPLP. However, the limitations of linear models fail to fully reveal the deeper nonlinear relationship between the two. Therefore, this paper further introduces advanced deep learning methods by incorporating KAN (Kolmogorov-Arnold Network) and GCN (Graph Convolutional Network) into four classic deep

learning time series models (Transformer, LSTM, BiLSTM, and GRU), resulting in improved models such as KAN-Transformer, KAN-LSTM, KAN-BiLSTM, KAN-GRU, and GCN-Transformer, GCN-LSTM, GCN-BiLSTM, GCN-GRU. By introducing KAN and GCN, this study significantly enhances the deep learning models' ability to automatically learn features, demonstrating higher prediction accuracy and performance when capturing the complex patterns and nonlinear relationships between DIF and TFPLP. Compared to traditional methods, these deep learning models can more comprehensively capture the latent nonlinear features in the data, improving prediction accuracy and revealing deeper economic relationships. Ultimately, this study not only combines the causal inference and individual control abilities of the fixed-effects model in traditional economics but also cleverly integrates the high predictive power of deep learning models, establishing a solid bridge between economics and computer science, providing a new perspective and methodology.

Keywords: Digital Inclusive Finance; Total Factor Productivity; Deep Learning; Kolmogorov–Arnold Neural Network; Graph Convolutional Network; High-Quality Enterprise Development.

1. Introduction

In the current global economic context, high-quality development of enterprises (HQED) has become one of the core goals for driving economic growth. This is especially true in China, where the economy is shifting from fast growth to development focused on quality[1]. Improving the internal efficiency and innovation of businesses has become an important issue. Total Factor Productivity (TFPLP) is a key measure of a business's internal efficiency and innovation[2], TFPLP is calculated using LP methods, based on the to estimate the marginal output of production factors. It shows the level of output under specific technological and managerial conditions[3]. In this process, (DIF) has become an important financial innovation[4]. It is changing the traditional financial system, improving how resources are shared, and giving businesses better ways to get funding. By using technologies like the internet, big data, and artificial intelligence, DIF lowers financing costs, improves information transparency, and makes financial services more inclusive. So, DIF is becoming more important for HQED[5].

Some research has looked at the effect of DIF on enterprises, but majority concentrate on scale growth and the impact of traditional financial models on business growth. There is little research on how digital finance helps transform businesses by affecting resource allocation, technological innovation, and management improvement[6]. Traditional economic studies usually use linear models, which can't capture the complex, nonlinear relationships between financial variables and business performance. As a result, these methods struggle to handle modern financial data, especially when dealing with large, multidimensional data. Artificial intelligence (AI), as a broad field of technology, aims to solve complex problems by simulating human cognitive processes[7-9]. With the improvement of computational power and the surge of big data, deep learning, a key branch of AI, has made significant advancements. Deep learning primarily relies on neural network models, particularly deep neural networks, which simulate the structure and function of the human brain's neurons, enabling the model to learn more complex feature representations from vast amounts of data[10-12]. Recently, deep learning technologies, like neural networks and time series models (such as Transformer[13, 14], Long Short-Term Memory (LSTM), Bidirectional LSTM (BiLSTM), and Gated Recurrent Units (GRU)), have become useful tools for solving nonlinear problems in finance[15]. Among these, Transformer, which uses a self-attention mechanism, is commonly used for modeling sequence data. By using self-attention, Transformer are capable of efficiently capturing long-range dependencies in sequences, making it good at handling complex nonlinear relationships in time series data. LSTM networks are an improved type of recurrent neural network that can handle sequence data with long-term dependencies. LSTM solves the gradient vanishing problem in traditional RNNs by adding memory units[16]. BiLSTM[17] expands LSTM by processing sequence data in both forward and backward directions, helping it capture more detailed contextual information. GRU[18]，a type of LSTM, reduces computational complexity by using an update gate mechanism, while still handling long-term dependencies in sequence data well. In contrast to traditional approaches, the deep learning models are better at capturing complex nonlinear relationships and finding hidden patterns in data[19], making them useful tools for analyzing the impact of DIF on the HQED.

This research choses companies listed in China from 2011 to 2022 as the research sample. The data can be obtained in supporting material. By analyzing data from this period, the study investigates the relationship between DIF and TFPLP of enterprises. To guarantee the precision and robustness of the results, this research incorporates multiple control variables when analyzing the impact of DIF. These variables reflect the basic characteristics and operational conditions of the enterprises, which may also influence their TFP. Furthermore, the study performs heterogeneity tests, examining the distinctions between state-controlled and non-state-controlled enterprises, enterprises in different life cycle stages, and technology-based enterprises, to guarantee the universality and reliability of the conclusions. Additionally, to get rid of potential interference from the COVID-19 pandemic[20], robustness checks are conducted to account for its impact on enterprise operations. These control variables and heterogeneity tests help us gain a more thorough understanding of the mechanisms through which digital inclusive finance affects the HQDE. However, as the volume of data increases and relationships become more complex, traditional linear models are clearly insufficient to understand the potential complex nonlinear relationships between DIF and enterprise total factor productivity. Therefore, after completing the traditional empirical analysis, this study further combines deep learning technologies, particularly neural network models, to reveal the deeper nonlinear relationships between DIF and enterprise high-quality development. Specifically, this study introduces the Kolmogorov–Arnold Neural Network (KAN) model and combines it with traditional time series models (such as Transformer, LSTM, BiLSTM, and GRU). KAN is built upon the Kolmogorov–Arnold representation theorem, which states that any multivariable continuous function can be represented as a combination of certain univariate functions[21]. This network architecture, with fewer hidden layers and nonlinear activation functions, effectively approximates complex functions[22]. Compared to traditional Multi-Layer Perceptron (MLP), the KAN model offers better expression capability and accuracy. It can solve high-dimensional function mapping problems with fewer computational resources, improving both efficiency and accuracy, especially with data that has complex structures. In this study, the KAN model replaces the MLP layer in the output layer of traditional models, greatly improving prediction accuracy and performance. By including DIF and eight control variables, the study predicts enterprise total factor productivity (TFPLP) and evaluates the model's prediction performance using R^2, RMSE, MSE, and MAE. The results show that the KAN model greatly improves prediction accuracy and performance. Besides the KAN model, this study also conducts an experiment with the Graph Convolutional Network (GCN). GCN is a deep learning model designed for graph-structured data, which applies convolution operations to extract features from the graph[23, 24]. GCN can capture complex relationships between nodes, and when processing enterprise data with complicated relationships, it can effectively identify interactions between different features, providing richer feature representations for future predictions. In this study, GCN is combined with the deep learning models mentioned earlier to fully use GCN's ability to capture spatial relationships in data. By constructing graph-structured data, GCN identifies spatial relationships within the data and then inputs the extracted spatial features into time series models, such as Transformer, LSTM, BiLSTM, and GRU for temporal modeling. This allows the study to reveal the complex relationships in time as well. By combining spatial and temporal features, this study further improves the prediction effectiveness of the models. After comparative analysis, the improved models show significantly better prediction accuracy and reliability than the original models. The main innovations of this study are as follows:

1) This paper systematically explores how DIF promotes high-quality enterprise development through intermediary variables such as innovation, financing constraints, and internal control, filling a gap in the existing literature.

2) The paper innovatively combines the KAN with traditional time series models (such as Transformer, LSTM, BiLSTM, and GRU), replacing the fully connected layer (MLP layer) in traditional models. This significantly improves the model's predictive accuracy and reveals the nonlinear relationship between DIF and enterprise TFPLP.

3) The paper also combines GCN with deep learning models, utilizing GCN to understand complex dependencies in the data and integrating it with time series modeling, which further enhances the model's predictive performance.

4) This paper includes the output of TFPLP as part of the input features in GCN. This allows the model to better learn the complex nonlinear relationships between features, improving prediction accuracy.

In conclusion, by combining traditional empirical analysis with advanced deep learning technologies, this study comprehensively explores how DIF affects the high-quality development of enterprises, especially by revealing the nonlinear relationship between DIF and enterprise TFP. The introduction of innovative models addresses the shortcomings of traditional empirical analysis in capturing complex nonlinear relationships, offering a new methodology for financial research and providing important decision-making references for policymakers and business managers. The two experimental models are illustrated in **Fig. 1**.

The remainder of this paper is structured as follows: **Section 2** provides a comprehensive review of existing research on DIF and high-quality development, after which **Section 3** presents the theoretical hypothesis analysis. Next, **Section 4** reviews the research design of empirical analysis of basic regression in detail. **Section 5** focuses on experimental results and **Section 6** presents analysis of heterogeneity test. Finally, **Section 7** concludes the research with a key finding's summary, current limitations and future research recommendations.

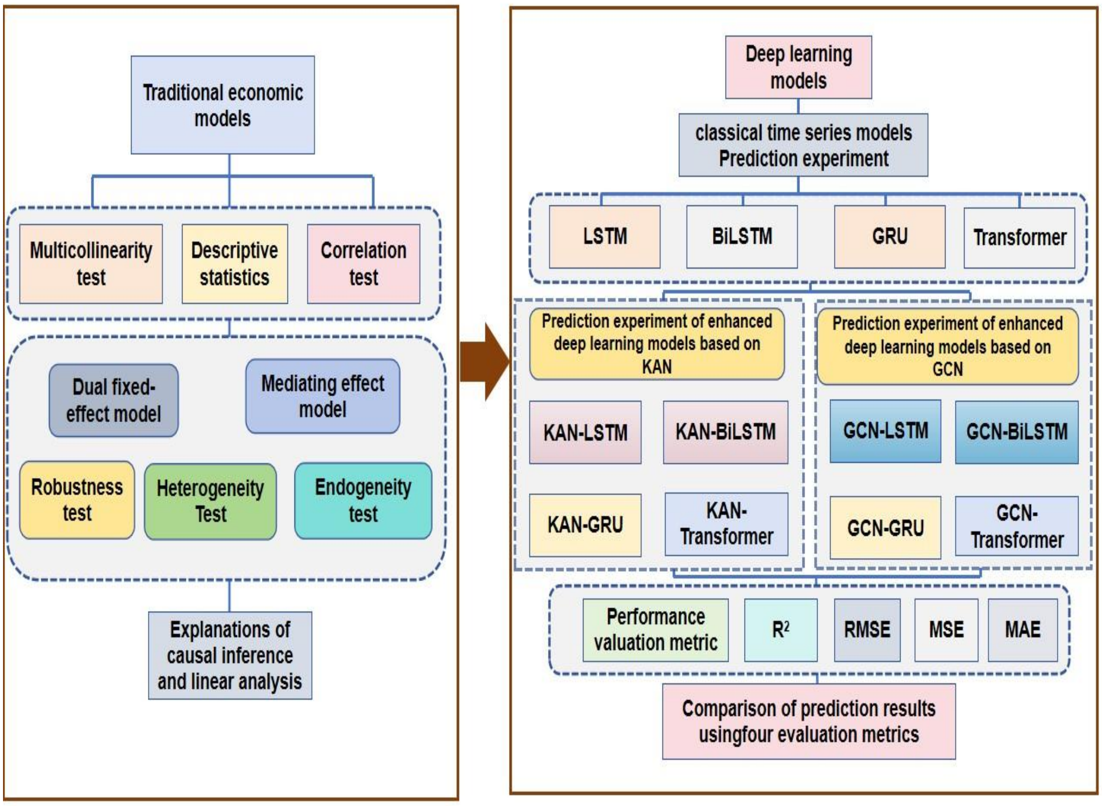


Fig. 1. Structural overview of traditional economic and deep learning models.

2. Literature review

2.1. Theoretical Framework and Interrelationship Between Digital Inclusive Finance and Total Factor Productivity

DIF refers to the use of digital technologies to provide more convenient, widespread, and low-cost financial services, particularly addressing the financing needs of small and medium-sized enterprises, low-income groups, and remote areas[5]. DIF promotes economic growth and social equity by enhancing the effectiveness of financial resource distribution and improving the accessibility of financial services. TFPLP is an essential indicator used to assess the efficiency of resource utilization in an economy or firm, reflecting the implication of technological progress and management innovation on economic performance[25]. Through this method, TFP effectively measures the overall output efficiency generated by inputs like capital and labor under different time frames and conditions. The effect of DIF on TFP is mainly shown in its ability to improve financing channels, lower capital acquisition costs, boost innovation, and increase production efficiency, which raises the TFP of businesses. Especially in areas with uneven resource distribution, DIF can reduce financing barriers, encourage technological innovation, and improve production processes, leading to better resource use and higher TFP.

2.2. Application and Limitations of Traditional Economic Models

Traditional models rest on the description of the relationship between economic variables based on assumptions. Usual models include linear regression, fixed effects, and random effects models. Using these models, with simplified assumptions, different economic factors' relations can be depicted with mathematical methods. Specially, these models turn out to be helpful while analyzing causality and trends. Traditional models interpret easily and the computation is simple; thus, it helps the policymakers. These models contain a simple structure, hence handling nonlinear relationships is difficult; especially for large-scale data, such relationships may turn out to be complex, thereby missing the detection of important patterns. The DFE model is widely used in panel data analyses[26]. The model is one of the common ones used in analysis related to panel data. This is done to control for individual and time effects so that one may analyze how policy changes or any other factor change the economic variables. Having both individual and time fixed effects, the model will be in a position to control for differences arising due to individual characteristics and time factors, hence reducing bias and increasing accuracy in conclusions.

2.3. Advantages and Challenges of Deep Learning in Economic Fields

Deep learning, a key part of machine learning, has made great progress in many fields. Built on artificial neural networks and multi-layer nonlinear processing, deep learning can learn complex patterns and features from data. In economics, deep learning is being used more and more, especially for big data analysis, predictive modeling, and modeling nonlinear relationships in economic phenomena, showing great potential and benefits[27].

Deep learning is good for dealing with complicated nonlinear relationships[28]. In most economic models, relationships are assumed linear, while the real-world economy usually has some nonlinear features; for example, fluctuation in financial markets is affected by many factors, as are corporate efficiencies and consumer behavior[29]. Deep learning can automatically learn such nonlinear features through its multilayer structure and make much more accurate predictions compared to other traditional models[30]. Deep learning also performs well in big data environments. With the increase in available data in economics, traditional methods often cannot handle large datasets. Deep learning can handle massive volumes of information and analyze them for patterns and regularities. A good example is financial analysis, where deep learning finds hidden patterns in huge volumes of market data and makes quite accurate predictions. Besides, deep learning models are flexible. Unlike traditional models, deep learning does not depend on pre-set assumptions but tunes model parameters by training the data to improve its prediction capabilities[31]. This makes deep learning more adaptive in changing economic scenarios, where it can sustain changes in different economic cycles and market conditions. Speaking generally, deep learning, with strong data processing and nonlinear modeling and adaptability, may bring huge benefits in economic perspectives. Particularly, it far outperforms traditional methods when it comes to dealing with complex problems and vast volumes of information.

2.4. Complementary Advantages of Combining Traditional Economic Models with Deep Learning Models

Traditional models, especially the DFE model, provide excellent interpretability and strong theoretical guidance to analyze economic issues. The DFE model controls for time and individual effects, thus enabling it to give efficient estimates of causal relationships in a robust manner when panel data is dealt with. However, most traditional models assume linear relations between variables. These are problematic for nonlinear, complex, and diverse data found in big, complicated datasets. In such cases, the capability of fitting and also predicting, by the model, may be limited.

Deep learning models, however, excel at handling nonlinear relationships and recognizing complex patterns. By using multi-layer neural networks, deep learning can uncover hidden patterns from large datasets, capture intricate nonlinear relationships, and does not rely on the assumptions that traditional models depend on. And deep learning models are particularly good at managing high-dimensional, large-scale economic data, often outperforming traditional models in prediction accuracy and generalization[32].

Combining traditional economic models with deep learning models approaches leverages the strengths of both. Traditional models offer a clear theoretical framework and causal reasoning for analysis, while deep learning models improve the ability to understand and analyze complex economic phenomena.

3. Theoretical hypothesis analysis

This section introduces four hypotheses, along with detailed descriptions for each.

3. 1. Direct pathways of DIF in promoting high-quality enterprise development

DIF has become a key factor in promoting HQDE by directly affecting several internal mechanisms within businesses[5]. First, DIF helps traditional industries become more productive and transform faster, making the real economy more sustainable over the long term. Second, DIF improves the efficiency of production processes, leading to better quality results[33]. This boost in productivity directly supports overall business growth, enhances competitive advantage, and maximizes corporate performance, all of which contribute to high-quality development. From this analysis, we derive the first hypothesis of this study.

Hypothesis 1: The development of DIF promotes HQDE

3. 2. Indirect paths of DIF affecting the HQDE

This section looks at how DIF affects three key factors: alleviating corporate financing constraints, enhancing internal control, and boosting innovation capacity. We will look at how these impacts indirectly promote HQDE.

3. 2. 1. Addressing financing constraints in enterprises through DIF

Financing is a major issue for Chinese businesses, mainly due to problems in the financial market and banks' unfair lending policies[34]. DIF helps solve this problem by playing a crucial role in expanding access to credit, especially for vulnerable businesses, providing them with better opportunities to secure long-term loans and supporting their ongoing development[34].

As a result, easing financing constraints through DIF indirectly improves the overall development quality of businesses. Based on this analysis, we derive the following hypothesis:

Hypothesis 2: Digital inclusive finance alleviates financing constraints to boost the HQDE.

3. 2. 2. Enhancing internal control level of enterprises through DIF

DIF is an effective way to help businesses improve their internal control mechanisms by using digital tools to make data more transparent and monitor processes in real time. These technologies allow businesses to identify mistakes, automate tasks, and ensure compliance with rules[35]. This reduces the risk of errors, fraud, and rule violations. Based on this analysis, we derive the following hypothesis:

Hypothesis 3: Digital inclusive finance enhances internal control level to boost the HQDE.

3. 2. 3. Facilitating innovation of enterprises through DIF

Innovation is a crucial factor for sustaining long-term enterprise success and maintaining a competitive edge in the market. By granting enterprises access to digital financial services, DIF enables them to invest more effectively in research and development (R&D), improving products and services[36]. Based on this analysis, we derive the following hypothesis:

Hypothesis 4: DIF facilitates innovation of enterprises to boost the HQDE.

4. Empirical strategies of empirical analysis and data sources of basic regression

The experimental design of this paper concentrates on exploring the impact of DIF on high-quality enterprise development, employing rigorous modeling and diverse empirical analysis methods. The traditional economic research design's overall framework is presented in **Fig. 2**.

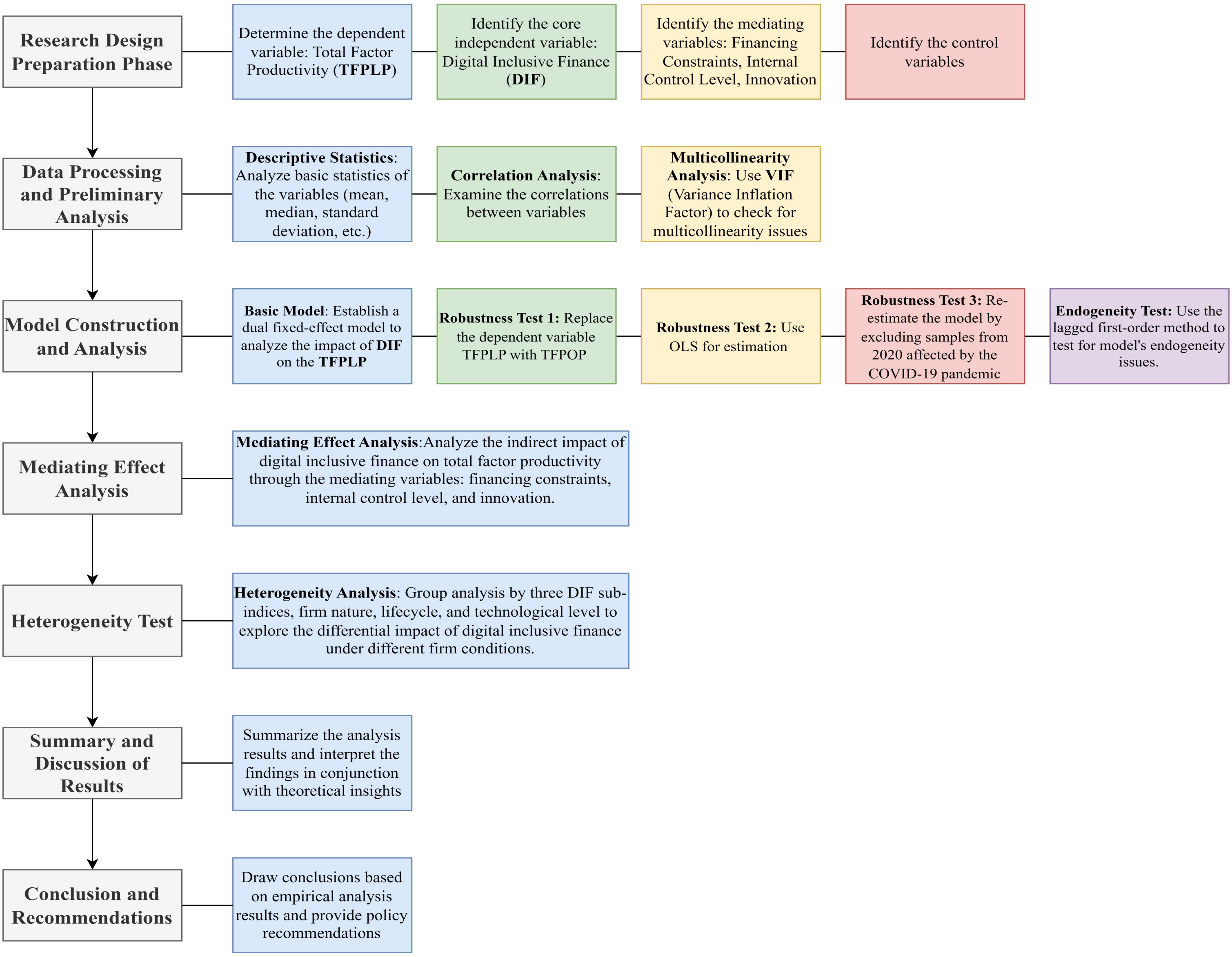


Fig. 2. The overall framework of research design.

Overall, the experimental design emphasizes the rigor of model construction and employs a variety of analytical methods to comprehensively examine the impact pathways and mechanisms of DIF on high-quality enterprise development.

4. 1. Index selection and data description

In this study, 22,291 enterprise annual observations for the period of 2011-2022 were selected as the analyzed sample, covering a wide range of industries and regions to guarantee the representativeness and breadth of the research outcomes. The sample data were strictly screened and processed. Firstly, financial enterprises were removed from the sample to avoid the interference of financial industry specificity on the research results. Secondly, ST and *ST enterprises, which often face financial difficulties or operational abnormalities that may bias the analysis results, are removed from the sample as well. Finally, enterprises with more missing values in the data were removed to make sure the completeness and reliability of the data. In order to further minimize the impact of extreme values on the results of the study, the continuous variables have been reduced-tailed at the upper and lower 1% quartiles. All company data are sourced from CSMAR database.

Table 1. The result of descriptive statistics for the variables.

Variable	Symbol	N	Mean	p50	S. D.	Min	Max
High-Quality Enterprise Development	TFPLP	22291	8.337	8.228	1.019	6.198	11.14
Digital Inclusive Finance	DIF	22291	303.3	318.1	94.9	60.58	459
Innovation	Innovation	22291	2.832	2.996	1.768	0	9.61
Financing Constraint	FC	22291	-3.835	-3.838	0.282	-5.69	0
Internal Control Level	ICL	22291	642.3	661.3	111.2	0	945
Company Size	SIZE	22291	22.28	22.08	1.288	20.03	26.33
Leverage Ratio	LEV	22291	0.414	0.403	0.2	0.057	0.887
Ownership Concentration	TOP	22291	33.93	31.8	14.62	8.43	73.7
Business Growth	Growth	22291	0.166	0.105	0.377	-0.512	2.275
Ownership Concentration	CFO	22291	0.05	0.048	0.067	-0.137	0.244
Company Age	AGE	22291	2.94	2.996	0.323	1.946	3.526
Board Size	BSIZE	22291	2.115	2.197	0.195	1.609	2.639
Management Shareholding	MSR	22291	14.23	1.135	19.69	0	67.45

Note：(1) The LP method is calculated in TFP, according to the Cobb-Douglas function, converted into the following equation for calculation: $lnY_{it} = \alpha \, ln \, Lit + \beta \, ln \, Kit + uit$. In this equation, the output Y stands for the primary business income of the listed company/the added value of the listed company, Labor L stands for the number of employees within the listed company, Capital K stands for the net value of fixed assets and cash agent for purchasing goods and accepting labor payment for intermediate product input M. (2) For the calculation method of DIF, please refer to the research on Index Compilation 2,191 listed companies. The DIF index is based on the Peking University Digital Inclusive Finance Index while the other accounting and financial indicators are based on the China Stock Market and Accounting Research Database and annual financial reports. The definitions of important variables and descriptive statistics are listed in **Table 1**.

4. 2. The measurement of variable

4. 2. 1. Explained variable: High-quality development of enterprises.

In this study, total factor productivity is used as a key indicator for measuring HQDE. TFPLP, as the explanatory variable, reflects the efficiency and overall production capability of an enterprise in relation to the input of production factors. Specifically, Total Factor Productivity represents the actual level of output that an enterprise can achieve, considering the input of capital, labor, and other production factors. By analyzing changes in TFP, we can evaluate the performance of enterprises in improving production efficiency, innovation capability, and resource allocation, thereby indirectly measuring the progress in high-quality development. It is worth noting that the Chinese government has stressed that the key to high-quality development is to improve TFP and achieve innovation-driven development. For businesses and enterprises, total factor productivity is not only an indicator to measure the overall resource allocation of enterprises, but also an important indicator to reflect the technological innovation level of enterprises. Therefore, on the basis of keeping consistent with the methods followed by most scholars, this paper chooses total factor productivity as an indicator to measure HQDE. For our benchmark specification, we use Levinsohn and Petrin's (2003) LP approach to assess enterprise development quality versus TFP. In a follow-up analysis, we used Wooldridge's (2009) one-step OP-based estimation method to recalculate TFP to demonstrate the robustness of our preliminary results.

4. 2. 2. Explanatory variable: Digital inclusive finance

In this study, DIF is used as an explanatory variable to explore its impact on financial development and economic performance. DIF refers to financial services provided through digital technologies[37], aiming to make financial services more accessible, convenient, and affordable. We consider DIF as an explanatory variable because it may influence changes in other economic or financial indicators. To comprehensively analyze the role of DIF, we decompose it into three key sub-indexes:

1) **Depth:** This refers to the extent of digital inclusive financial services' penetration among different social groups[38].
2) **Breadth**: This pertains to the coverage of digital inclusive financial services, including the types of services provided and the diversity of financial products.
3) **Degree of Digitization**: This indicates the level of digitalization of financial services, i.e., whether services are provided fully or partially through digital channels[39].

4. 2. 3. Control variables

In the study, control variables are used to isolate the effect of the primary explanatory variable to ensure that the observed results are attributable to the explanatory variable and not confounded by other factors. Here's an explanation of these eight control variables:

1) **Firm Size (SIZE)**: Represented by the natural logarithm of total assets. Firm size often influences financial and market performance. Larger firms may have more resources, stronger market positions, and greater economic stability. Controlling firm size helps to avoid confounding effects on the results[40].

2) **Financial Leverage (LEV)**: The ratio of total debt to total assets. Financial leverage reflects the level of debt in the firm's capital structure[41]. Higher financial leverage can increase financial risk and affect investment decisions and operational performance. Controlling financial leverage helps to understand the impact of capital structure on firm performance.

3) **Cash Flow Level (CFO)**: The ratio of net cash flow from operating activities to total assets. Cash flow level reflects the firm's cash generation capacity[42]. A higher cash flow level typically indicates sufficient cash to support operations and investment activities. Controlling cash flow level helps to distinguish the effects of cash flow from other financial factors.

4) **Firm Growth (GROWTH)**: The growth rate of operating revenue. Firm growth measures the expansion capability and market opportunities of the firm. Growth can impact financial conditions and market performance[43]. Controlling growth helps to understand the influence of expansion speed on overall performance.

5) **Firm Age (AGE)**: The time difference between the firm's founding year and the observation year. Firm age can affect market experience, stability, and reputation. Older firms may have more market experience and stability, so controlling for firm age helps to exclude its impact on the results.

6) **Ownership Concentration (TOP)**: The proportion of shares held by the largest shareholder. Ownership concentration reflects the shareholder structure of the firm. Firms with high ownership concentration might face different governance structures and decision-making processes, so controlling this variable helps understand its influence on performance[44].

7) **Board Size (BSIZE)**: The natural logarithm of the number of board members. Board size can affect corporate governance and decision-making efficiency. Controlling board size helps isolate the impact of governance structure on firm performance.

8) **Management Shareholding (MSR)**: The proportion of shares held by the management team. Management shareholding can align the interests of management with those of shareholders and affect strategic decisions. Controlling management shareholding helps to account for its effect on firm performance[45].

4. 2. 4. Mediating variables

Intermediary variables are used to explore the mechanisms or pathways through which the explanatory variable affects the explanatory variable. Here's an explanation of Innovation, Financing Constraints (FC), and Internal Control Level (ICL) as intermediary variables:

1) **Innovation**: Innovation typically refers to new developments and improvements in products, services, or processes within a firm[46]. As an intermediary variable, innovation can help explain how the explanatory variable influences the explanatory variable through driving technological advancements, product development, or business model improvements. For example, DIF may indirectly affect financial performance or market competitiveness by enhancing a firm's innovation capabilities.

2) **Financing Constraints (FC)**: Financing constraints refer to the difficulties and

limitations a firm faces in obtaining external funding[47]. As an intermediary variable, financing constraints can reveal how the explanatory variable influences the explanatory variable through its impact on the firm's ability to obtain and conditions of financing. For instance, firm size or financial leverage might affect financing constraints, which in turn influences the firm's innovation capacity and growth potential[48].

3) **Internal Control Level (ICL)**: The internal control level refers to the effectiveness of measures in place for financial reporting, operations, and compliance within a firm. As a intermediary variable, the internal control level can help explain how the explanatory variable impacts financial stability and operational efficiency through improving internal control. For example, effective internal control can reduce operational risks and financial irregularities, enhancing overall performance and decision-making quality.

4. 3. Reference model specification

4. 3. 1. Dual fixed-effect model

In order to more intuitively analyze the impact of DIF on enterprise high-quality development, the following regression model is established based on existing literature:

$$TFPLP_{i,t} = \alpha_0 + \alpha_1 DIF_{i,t} + \alpha control_{i,t} + \delta_i + \rho_t + \varepsilon_{i,t} \tag{1}$$

where i represents the individual, t represents the year, enterprise high-quality development (TFPLP) is the explanatory variable, digital financial inclusion (DIF), breadth (BRE), depth (DEP), digitization degree (DIG) is the core explanatory variable, and $control$ is the relevant control variable:enterprise scale (SIZE), financial leverage (LEV), cash flow level (CFO), Enterprise Growth (Growth), enterprise age (AGE), ownership concentration (TOP), Board size (BSIZE),management shareholding (MSR). δ_i represents individual fixed effects, ρ_t expressed fixed time effect, $\varepsilon_{i,t}$ is the random disturbance term.

4. 3. 2. Mediating effect model

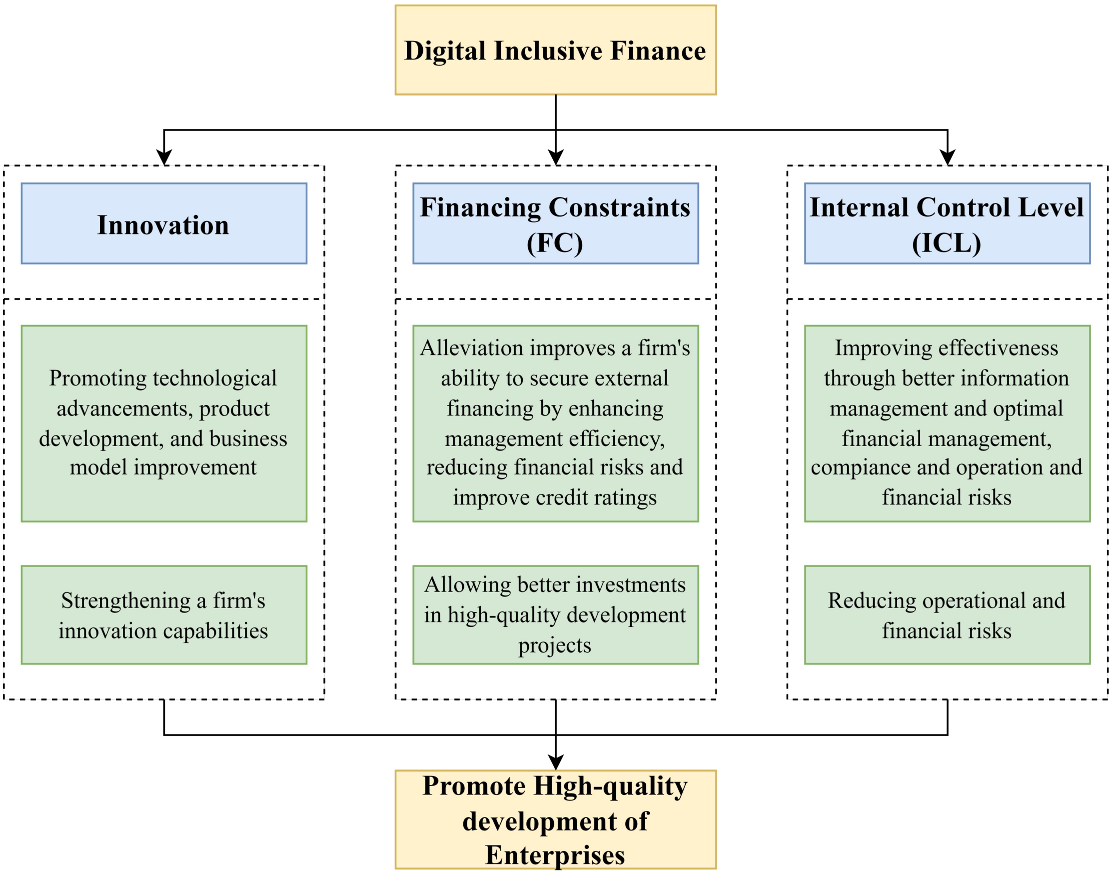


Fig. 3. The overview of mediating effect model.

To explore possible intermediary variables in the transmission mechanism of DIF on enterprise high-quality development, this section adopts the Stepwise Regression for Mediation Effect (Baron & Kenny's mediation test) model and takes financing constraint (FC)， Innovation and Internal Control Level (ICL) as the intermediary variable $M_{i,t}$ for empirical analysis and the mediating effect mechanism overview is presented in **Fig. 3**.

The three-step method is a classic approach to testing mediation effects, typically divided into three steps as follows:

Step 1: Total effect test: First, test whether the total effect of the independent variable (DIF) on the dependent variable (TFPLP) is significant. Establish the regression model:

$$TFPLP_{i,t} = \gamma_0 + \gamma_1 DIF_{i,t} + \gamma control_{i,t} + \delta_i + \rho_t + \varepsilon_{i,t} \tag{1}$$

If the regression coefficient c is significant, it demonstrates that the independent variable has a significant impact on the dependent variable, allowing for further mediation effect testing.

Step 2: Mediator Variable Effect Test: Test whether the independent variable (DIF) has a significant effect on the mediator variable (M). Establish the regression model:

$$M_{i,t} = \beta_0 + \beta_1 DIF_{i,t} + \beta Control_{i,t} + \delta_i + \rho_t + \varepsilon_{i,t} \tag{2}$$

If the regression coefficient a is significant, it indicates that the independent variable significantly affects the mediator variable, suggesting that the independent variable may influence the dependent variable through the mediator.

Step 3: Mediation effect and direct effect test: Test the impact of the mediator variable (M) on the dependent variable (TFPLP), while controlling for the independent variable (DIF). Establish the regression model:

$$TFPLP_{i,t} = \gamma_0 + \gamma_1 DIF_{i,t} + \theta M_{i,t} + \gamma control_{i,t} + \delta_i + \rho_t + \varepsilon_{i,t} \tag{3}$$

where the above 3 equations test whether there is a conduction relationship between explanatory variable, intermediary variable and explanatory variable, and whether the intermediary variable is an important influence path. $M_{i,t}$ is an intermediary variable, First, the intermediary variable is not included in Equation (1). Suppose that coefficient α_1 indicates that the impact of DIF on enterprise high-quality development (TFPLP) is significant. In that case, regression is carried out based n Equation (2) to evaluate the effect of DIF on $M_{i,t}$ and further bring the intermediary variables into the regression framework to form Equation (3). If these variables are significant, it implies that a mediating effect is evident. If γ_1 is significant, it implies that $M_{i,t}$ has a partial mediating effect. If γ_1 is not significant, it implies that $M_{i,t}$ has a complete mediating effect.

4. 3. 3. Estimation methods for Endogeneity concerns

In analyzing the impact of DIF on TFPLP, there may be a reciprocal causal relationship between the two, leading to endogeneity issues. Additionally, using a fixed effects model for panel data regression may result in bias due to endogeneity. To estimate the model, the lagged value of the core explanatory variable is selected as an instrument to reduce estimation bias coming from the empirical regression. We use the lagged value ($DIF_{i,t-1}$) of the core explanatory variable ($DIF_{i,t}$). The process is shown in **Fig. 4**. This is because the lagged variable is generally uncorrelated with the current random disturbance term but may be correlated with the current explanatory variable, meeting the requirements for an instrument.

First stage regression: Use the instrumental variable $DIF_{i,t-1}$ to perform a regression on and obtain the predicted value $D\hat{I}F_{i,t}$, The model can be structured as follows:

$$D\hat{I}F_{i,t} = \lambda_0 + \lambda_1 DIF_{i,t-1} + \lambda_2 \mathrm{control}_{i,t} + \delta_i + \rho_t + \varepsilon_{i,t} \tag{4}$$

where, $D\hat{I}F_{i,t}$ is the predicted value obtained from the first stage regression.

Second stage regression: In the second stage, substitute the predicted value $D\hat{I}F_{i,t}$ obtained from the first stage regression into the original baseline regression model:

$$TFPLP_{i,t} = \phi_0 + \phi_1 D\hat{I}F_{i,t} + \phi control_{i,t} + \delta_i + \rho_t + \varepsilon_{i,t} \tag{5}$$

where, by using $D\hat{I}F_{i,t}$ instead of the original $DIF_{i,t}$, the model effectively addresses the endogeneity issue.

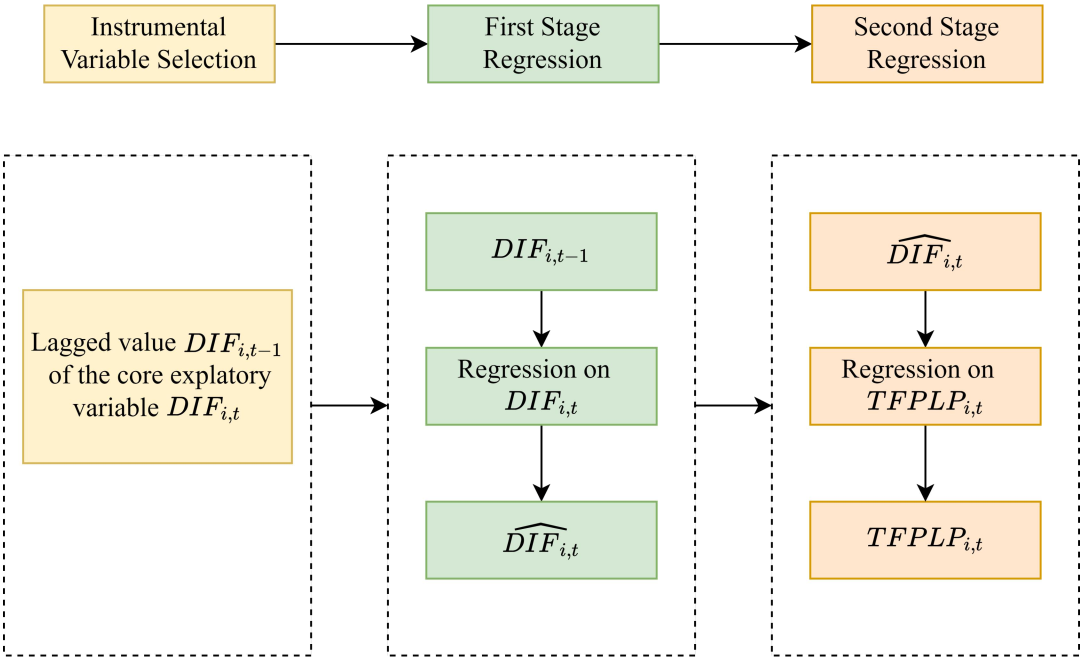


Fig. 4. The overview of endogeneity estimation.

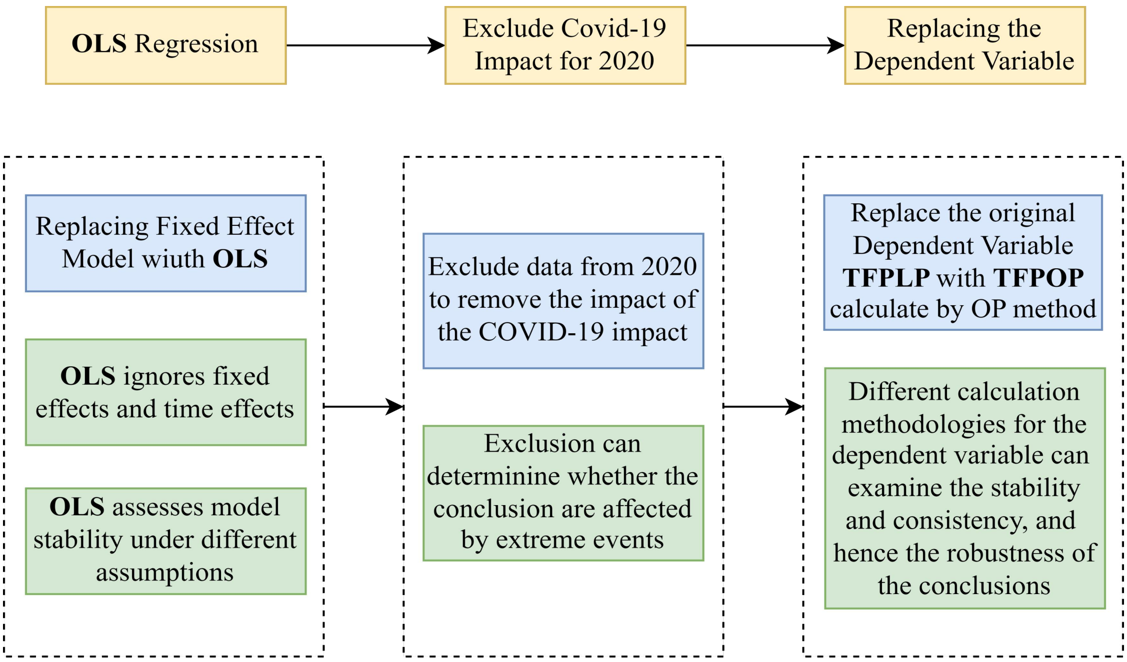


Fig. 5. The process of robustness estimation.

4. 3. 4. Estimation methods for robustness concerns

To assess the robustness of the empirical results from the previously constructed fixed effects regression model Eq (1) and make sure the rigor of the study's outcomes, this paper opts to modify the regression model to an Ordinary Least Squares (OLS) regression, which is applied to verify the validity of the original model.

$$TFPLP_{i,t} = \beta_0 + \beta_1 DIF_{i,t} + b\,\text{control}_{i,t} + \varepsilon_{i,t} \tag{6}$$

where, individual fixed effects δ_i and fixed time effect $\rho_t$are not included.

Additionally, this study further validates the model's robustness by excluding the impact of the 2020 COVID-19 pandemic from the original fixed effects model and by replacing the dependent variable (TFPLP) with a variable calculated using the OP method (TFPOP) and the process of robustness estimation illustrated in **Fig. 5**.

5. Empirical results and analysis.

5. 1. Multicollinearity analysis

Table 2. The result of VIF test.

Variable	VIF	1/VIF
SIZE	1.60	0.626817
LEV	1.49	0.672060
MSR	1.24	0.805599
AGE	1.10	0.905590
Growth	1.01	0.985874
CFO	1.07	0.930484
TOP	1.07	0.937819
BSIZE	1.08	0.923215
Mean VIF	1.21	0.848432

Table 2 presents the Variance Inflation Factor (VIF) and its reciprocal (1/VIF) for various variables in the regression model, used to assess the degree of multicollinearity among the explanatory variables. In general, a VIF value less than 10 indicates that multicollinearity is not a serious problem. From **Table 2**, we can see that the VIF values for SIZE, LEV, MSR, AGE, Growth, CFO, TOP, and BSIZE are all well below 10, with the highest VIF being 1.60 for SIZE. The corresponding 1/VIF values are all above 0.1, further confirming that multicollinearity is not a significant issue in this model. Specifically, SIZE has a VIF of 1.60 and 1/VIF of 0.626817, indicating almost no multicollinearity. LEV has a VIF of 1.49 and 1/VIF of 0.672060, indicating a low level of multicollinearity. The VIF values for MSR, AGE, Growth, CFO, TOP, and BSIZE range from 1.01 to 1.24, with corresponding 1/VIF values between 0.805599 and 0.985874, indicating almost no multicollinearity. The mean VIF value is 1.21, with a corresponding mean 1/VIF of 0.848432, further supporting the conclusion that multicollinearity is not an issue in this dataset.

5. 2. Correlation analysis

Table 3-1. Correlation analysis result.

Variables	TFPLP	DIF	Innovation	FC	ICL	SIZE	LEV
TFPLP	1						
DIF	0.367***	1					
Innovation	0.362***	0.225***	1				
SA	-0.003	-0.372***	0.078***	1			

ICL	0.208***	0.093***	0.137***	0.105***	1		
SIZE	0.813***	0.141***	0.379***	0.025***	0.129***	1	
LEV	0.487***	-0.020***	0.115***	-0.045***	-0.093***	0.516***	1

Table 3-2. Correlation analysis result.

Variables	TOP	Growth	CFO	AGE	BSIZE	MSR
TOP	1					
Growth	-0.00600	1				
CFO	0.094***	0.043***	1			
AGE	-0.094***	-0.058***	0.013*	1		
BSIZE	0.021***	-0.018***	0.032***	0.047***	1	
MSR	-0.082***	0.062***	0.011*	-0.233***	-0.204***	1

Note: "***", "**" and "*" respectively indicate that the indicators are significant at the level of 1%, 5% and 10%.

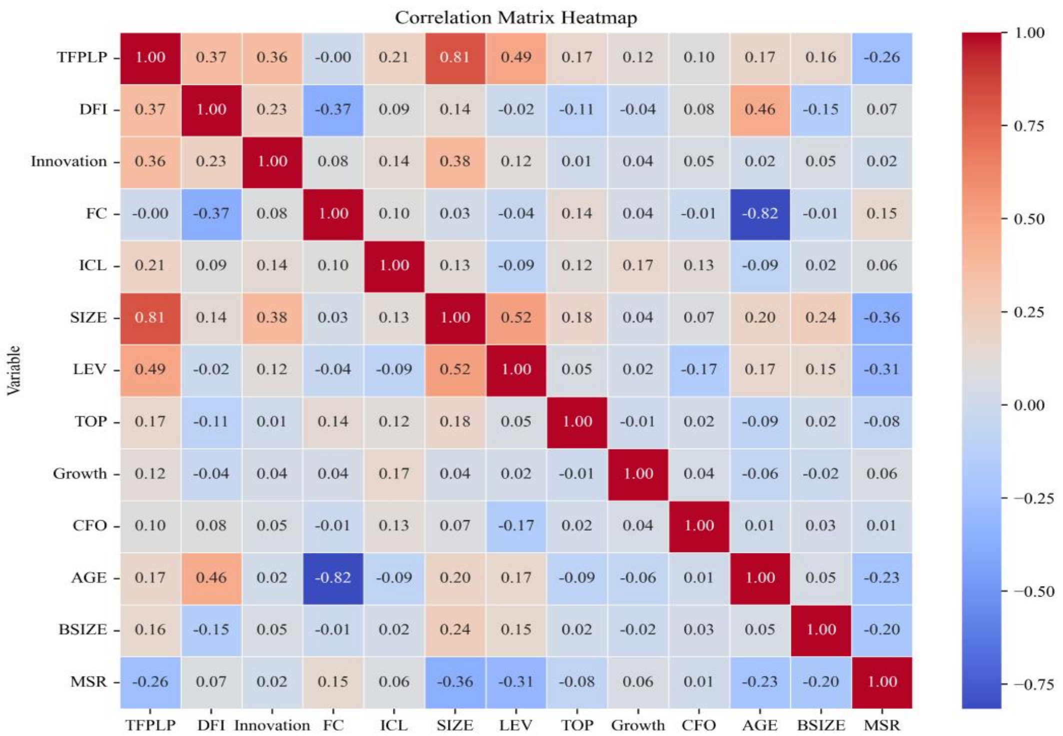


Fig. 6. The correlation matrix heatmap.

Table 3 and Fig. 6 reveal several strong relationships among the variables, such as a positive correlation between TFPLP and FC (0.81) and between SIZE and BSIZE (0.83), indicating that these pairs move in tandem. Conversely, there are strong negative correlations, such as between Innovation and FC (-0.82) and between LEV and SIZE (-0.69), suggesting that as one variable increases, the other tends to decrease. Most other variables, such as Growth, TOP, and CFO, show weak correlations, indicating limited linear relationships among them. Overall, the matrix highlights key variable interactions that may warrant further investigation.

Overall, the analysis shows robust relationships between the key variables and TFPLP, emphasizing the significant role of firm size, financial indicators, and internal control in driving productivity. The significant positive and negative correlations highlight the complex interplay of these factors in determining firm performance.

5. 3. Baseline regression analysis

5. 3. 1. Baseline regression analysis for DIF

This section analyzes the impact of DIF on the high-quality development of enterprises by comparing the results before and after adding control variables.

Table 4. Regression results for testing the role of DIF.

Variables	(1)	(2)
	TFPLP	TFPLP
DIF	0.00204***	0.00076***
	(5.16)	(2.59)
SIZE		0.55284***
		(97.53)
LEV		0.17039***
		(7.84)
TOP		0.00144***
		(3.85)
GROWTH		0.24830***
		(47.37)
CFO		0.76915***
		(21.54)
AGE		-0.06413
		(-1.37)
BSIZE		0.09417***
		(4.83)
MSR		-0.00084***
		(-2.92)
_cons	7.59965***	-4.45569***
	(230.21)	(-26.58)
N	22291	22291
R^2	0.389	0.669
Year	Yes	Yes
FE	Yes	Yes

Note: Values in brackets are *t* values. "***", "**" and "*" respectively indicate that the indicators are significant at the level of 1%, 5% and 10%.

From **Table 4**, it can be seen that in Model 1, DIF has a significant positive impact on high-quality development of enterprises (TEPLP), with a coefficient of 0.00204 and a t-value of 5.16, significant at the 1% level (***). This indicates that the improvement of DIF significantly promotes the high-quality development of enterprises. Model 1 does not include other control variables and only demonstrates the direct impact of DIF on the high-quality development of enterprises.

In Model 2, with the introduction of multiple control variables, the impact of DIF) on high-quality development of enterprises (TEPLP) remains significant, with a coefficient of 0.00076 and a t-value of 2.59, significant at the 1% level (). The introduced control variables include firm size (SIZE, coefficient of 0.55284, t-value of 97.53), financial leverage (LEV, coefficient of 0.17039, t-value of 7.84***), firm growth (GROWTH, coefficient of 0.24830, t-value of 47.37***), and cash flow level (CFO, coefficient of 0.76915, t-value of 21.54***), all of which significantly affect the high-quality development of enterprises. Other control variables including ownership concentration (TOP), firm age (AGE), board size (BSIZE), and manager's shareholding ratio (MSR) are also considered in Model 2, but their impacts are relatively small or insignificant.

Comparing Model 1 and Model 2, the impact of DIF is significant in both models, but the coefficient decreases from 0.00204 to 0.00076, indicating that the influence of DIF weakens after introducing control variables, yet it remains positively significant. Control variables such as firm size (SIZE), financial leverage (LEV), firm growth (GROWTH), and cash flow level (CFO) show high significance in Model 2, demonstrating their crucial impact on the high-quality development of enterprises. After introducing control variables, the explanatory power of Model 2 has a significant improvement, with R^2 increasing from 0.389 to 0.669, indicating that Model 2 can more comprehensively explain the factors influencing the high-quality development of enterprises. Overall, the results are more robust after introducing control variables, further confirming the key role of DIF in promoting high-quality development of enterprises, while also highlighting the importance of firm size, financial leverage, growth rate, and cash flow level.

5. 4. Robustness tests

Table 5. Robustness check and dealing with endogeneity concerns.

Variables	**One-period lag**	**TFPOP**	**OLS**	**FE (eliminating 2020)**
	(1)	(2)	(3)	(4)
DIF	0.00500***	0.00184***	0.00496***	0.00167***
	(42.57)	(21.29)	(44.93)	(18.82)
SIZE	0.53460***	0.40688***	0.53692***	0.54686***
	(153.38)	(69.10)	(164.21)	(90.40)
LEV	0.82998***	0.02591	0.82260***	0.17498***
	(38.67)	(1.15)	(41.16)	(7.52)
TOP	0.00367***	0.00154***	0.00356***	0.00198***
	(14.71)	(3.95)	(15.41)	(4.92)
GROWTH	0.26422***	0.25229***	0.25791***	0.24036***
	(27.62)	(46.69)	(29.28)	(42.85)
CFO	0.70871***	0.79227***	0.70633***	0.78191***
	(12.94)	(21.35)	(13.95)	(19.99)
AGE	-0.45444***	-0.04566	-0.43500***	-0.10466**

	(-33.08)	(-1.03)	(-35.10)	(-2.29)
BSIZE	0.15784***	0.07601***	0.14531***	0.12555***
	(8.34)	(3.73)	(8.22)	(5.92)
MSR	-0.00202***	-0.00046	-0.00145***	-0.00071**
	(-9.63)	(-1.54)	(-7.72)	(-2.26)
_cons	-4.73377***	-3.07447***	-4.19920***	-4.53032***
	(-47.32)	(-19.53)	(-52.49)	(-28.02)
N	19747	22291	22291	19972
R^2	0.770	0.582	0.774	0.663
Year	Yes	Yes	Yes	Yes
FE	Yes	Yes	No	Yes

The regression results of Model 1 in **Table 5** after applying the first-order lagging show that the coefficients of the year variables are all significant, indicating that these years have a significant impact on the high-quality development of enterprises (TFPLP). The regression coefficient for DIF remains positive and significant, suggesting that even after controlling for the effects of lagged years, DIF still has a positive impact on high-quality enterprise development, which supports the positive role of DIF in business growth. Additionally, control variables such as SIZE, LEV, and GROWTH remain significant in the model after applying the first-order lagging, implying that their influence is not diminished by the lagging process. Overall, the model's R^2 value is relatively high (0.77), indicating strong explanatory power, and the lagging process has not significantly reduced the model's fit. This further demonstrates the effectiveness of first-order lagging in controlling endogeneity and improving the accuracy of the model's estimates.

As shown in **Table 5**, Model 2 changes the explanatory variable from TFPLP to TFPOP. The results of this model indicate that the coefficient for DIF remains significant and positive, which means that DIF has a positive impact not only on high-quality enterprise development (TFPLP) but also on total factor productivity (TFPOP) as an alternative measure. This further demonstrates that the positive effect of DIF on enterprise development is robust and does not depend on a specific explanatory variable.

Model 3 uses Ordinary Least Squares (OLS) for regression, and the results show that most variables have significant coefficients and consistent directions. This indicates that, even without considering fixed effects and time effects, the relationship between the explanatory variables and the dependent variable is robust. The R-squared value of the OLS model is 0.774, suggesting strong explanatory power, with most variables having a significant and positive impact on the outcome. However, since the OLS model does not account for individual and time fixed effects, there might be omitted variable bias, which could affect the robustness of the results. Model 4 is a fixed effects model that excludes the impact of the 2020 COVID-19 pandemic. It shows coefficients with similar directions to those in Model 3, although some coefficient values differ. After excluding the 2020 data, the R-squared value slightly decreases (to 0.663) but remains high, indicating that the model still has strong explanatory power. The fixed effects model, by controlling for individual and time effects, effectively reduces potential biases. Although some coefficients are smaller compared to Model 3, their significance remains high, indicating that the model results are still robust after controlling for more potential influencing factors. Excluding the

2020 data also helps test the impact of the pandemic on the results, enhancing the credibility of the findings.

5. 5. Mechanism verification

This section further assesses the transmission path between DIF and HQDE from the perspectives of enterprise FC, innovation and ICL.

5. 5. 1. The mediating effect in financing constraint (FC)

In the study, introducing Financing Constraint (FC) as an intermediary variable is both necessary and beneficial. By incorporating FC, the study can better understand how DIF promotes high-quality enterprise development by alleviating financing constraints. This mediating effect helps uncover the internal mechanisms through which DIF influences enterprise development, providing a more comprehensive explanatory framework.

We can see that, from **Table 6**, DIF has a significant direct effect on high-quality enterprise development (coefficient 0.00076, $p < 0.01$), indicating that DIF effectively enhances enterprise quality. In the second step, DIF has a significant negative impact on financial constraint (FC) (coefficient -0.00098, $p < 0.01$), suggesting that DIF reduces financial constraints for enterprises. In the third step, after introducing FC into the model, DIF's positive effect on high-quality development remains significant (coefficient 0.00162, $p < 0.01$), while FC also has a significant negative impact on high-quality development. This indicates that FC plays a partial mediating role in the relationship between DIF and high-quality development. In other words, DIF not only directly promotes enterprise high-quality development but also indirectly contributes by reducing financial constraints. This dual impact shows that DIF influences enterprise development through a complex network of pathways, rather than merely through simple capital injection. Therefore, the mediating effect of DIF in reducing financing constraints reveals its fundamental role in enhancing resource allocation efficiency, which is crucial for improving enterprise competitiveness and long-term growth.

Table 6. The results of mediating effect in FC.

	Mediating effect in Financial Constraint		
Variable	TFPLP (1)	FC (2)	TFPLP (3)
DIF	0.00076***	-0.00098***	0.00162***
	(2.59)	(-34.10)	(18.69)
SIZE	0.55284***	-0.00403**	0.54985***
	(97.53)	(-2.07)	(96.20)
LEV	0.17039***	-0.00667	0.17730***
	(7.84)	(-0.90)	(8.12)
TOP	0.00144***	0.00047***	0.00178***
	(3.85)	(3.60)	(4.70)
GROWTH	0.24830***	-0.00512***	0.24889***
	(47.37)	(-2.86)	(47.45)
CFO	0.76915***	0.01738	0.75883***
	(21.54)	(1.42)	(21.07)
AGE	-0.06413	-0.16984***	-0.09558**

	(-1.37)	(-11.54)	(-2.21)
BSIZE	0.09417***	-0.00504	0.11758***
	(4.83)	(-0.75)	(5.94)
MSR	-0.00084***	0.00075***	-0.00083***
	(-2.92)	(7.52)	(-2.84)
FC			-0.00130**
			(-2.06)
_cons	-4.45569***	-2.96246***	-4.57006***
	(-26.58)	(-56.86)	(-27.59)
N	22291	22291	22291
R^2	0.669	0.591	0.658
Year	Yes	Yes	Yes
FE	Yes	Yes	Yes

5. 5. 2. Innovation

By including innovation as an intermediary variable, the study can explore how DIF indirectly promotes high-quality development by fostering corporate innovation. Analyzing this mediating effect helps reveal the role of DIF in the innovation process, providing a deeper understanding of how financial inclusion policies influence corporate innovation behaviors and high-quality development, thereby offering theoretical support for policy formulation.

From Model 1 of **Table 7**, This table applies the three-step method to verify the mediating effect of DIF on high-quality enterprise development. First, Column 1 shows that the direct effect of DIF on high-quality development (TFPLP) is significant, with a regression coefficient of 0.00076 and a t-value of 2.59, indicating that DIF has a major beneficial impact on advancing high-quality development. Second, Column 2 demonstrates that DIF also has a strong impact on innovation, with a regression coefficient of 0.00326 and a t-value of 12.70, suggesting that DIF significantly promotes innovation within enterprises. Finally, Column 3 further reveals that innovation serves as a partial mediating role between DIF and high-quality development. After controlling for innovation, the indirect effect of DIF on high-quality development through innovation remains significant, with a regression coefficient of 0.00164 and a t-value of 19.42. This means that DIF not only directly promotes high-quality development but also indirectly enhances it by fostering innovation. Therefore, the mediating effect of innovation between DIF and high-quality development is confirmed, indicating that DIF contributes to enterprise development through both direct and innovation-driven pathways By fostering innovation, DIF gives companies greater flexibility and creativity, allowing them to not only survive in the current market but also excel in future competition. Therefore, the mediating effect of DIF in promoting innovation strengthens its core role in the strategic development of enterprises.

Table 7. The results of mediating effect in Innovation.

	Mediating effect in Innovation		
	TFPLP (1)	Innovation (2)	TFPLP (3)
DIF	0.00076***	0.00326***	0.00164***
	(2.59)	(12.70)	(19.42)
SIZE	0.55284***	0.48280***	0.55296***
	(97.53)	(27.67)	(94.83)

LEV	0.17039***	-0.22562***	0.17583***
	(7.84)	(-3.38)	(8.05)
TOP	0.00144***	0.00332***	0.00181***
	(3.85)	(2.87)	(4.76)
GROWTH	0.24830***	0.04145***	0.24915***
	(47.37)	(2.59)	(47.51)
CFO	0.76915***	0.10137	0.75951***
	(21.54)	(0.92)	(21.10)
AGE	-0.06413	-0.39967***	-0.09839**
	(-1.37)	(-3.03)	(-2.28)
BSIZE	0.09417***	0.14644**	0.11852***
	(4.83)	(2.42)	(5.99)
MSR	-0.00084***	0.00380***	-0.00080***
	(-2.92)	(4.27)	(-2.75)
Innovation			0.00646***
			(2.68)
_cons	-4.45569***	-8.13379***	-4.62646***
	(-26.58)	(-17.43)	(-30.03)
N	22291	22291	22291
R^2	0.669	0.173	0.658
Year	Yes	Yes	Yes
FE	Yes	Yes	Yes

5. 5. 3. Internal Control Level (ICL)

Table 8 illustrates the mediating role of internal control in the impact of different variables on corporate performance. First, DIF (institutional ownership) has a direct effect on corporate performance (TFPLP) with a coefficient of 0.00076 in the first column, which is significant at the 1% level (***), indicating a positive effect of institutional ownership on corporate performance. The second column shows that FI significantly influences the internal control level (ICL) with a high coefficient of 0.44758, also significant at the 1% level, suggesting that institutional ownership greatly enhances internal control within a company. In the third column, the effect of DIF on corporate performance slightly increases to 0.00077, indicating that although internal control as a mediating variable reduces some of the direct effect, its positive impact remains significant. Moreover, the direct impact of SIZE (company size) on corporate performance is 0.55284, and after considering the mediating role of internal control, the coefficient changes from 17.82246 to 0.55006, highlighting the significant moderating role of internal control in large companies. Thus, the data suggests that strengthening internal control helps optimize resource allocation and management efficiency, thereby improving corporate performance. Thus, the mediating effect of DIF in enhancing internal control further solidifies its strategic role as a pillar of comprehensive enterprise development.

Table 8. The results of mediating effect in ICL.

Mediating effect in Internal Control Level			
	TFPLP (1)	Internal Control Level (2)	TFPLP (3)

DIF	0.00076***	0.44758***	0.00077***
	(2.59)	(18.91)	(2.63)
SIZE	0.55284***	17.82246***	0.55006***
	(97.53)	(25.42)	(97.43)
LEV	0.17039***	-8.8e+01***	0.20292***
	(7.84)	(-20.63)	(9.32)
TOP	0.00144***	0.65283***	0.00114***
	(3.85)	(13.17)	(3.05)
GROWTH	0.24830***	45.90371***	0.23615***
	(47.37)	(24.30)	(44.56)
CFO	0.76915***	98.92659***	0.74914***
	(21.54)	(9.11)	(21.06)
AGE	-0.06413	-4.0e+01***	-0.05964
	(-1.37)	(-14.96)	(-1.28)
BSIZE	0.09417***	18.41006***	0.08839***
	(4.83)	(4.86)	(4.55)
MSR	-0.00084***	0.25115***	-0.00089***
	(-2.92)	(6.25)	(-3.09)
ICL			0.00027***
			(13.16)
_cons	-4.45569***	2.8e+02***	-4.56793***
	(-26.58)	(16.08)	(-27.34)
N	22291	22291	22291
R^2	0.669	0.127	0.673
Year	Yes	Yes	Yes
FE	Yes	Yes	Yes

6. Further discussion: heterogeneity analysis

This section will examine heterogeneity from the perspectives of enterprise life cycle, region, enterprise ownership and whether it is a technology enterprise.

6. 1. Heterogeneity test analysis for DIF sub-index

This section analyzes the impact of DIF depth, breadth, and digitalization on the HQDE. Through regression analysis of these dimensions, we explore their specific roles in enhancing enterprise development, incorporating control variables into the analysis. This analysis helps confirm the influence of various aspects of DIF while also revealing the roles of other important factors in HQDE.

Table 9. Heterogeneity: Digital inclusive finance sub-index.

Variables	(1)	(2)	(3)
	TFPLP	TFPLP	TFPLP
BRE	0.00162***		
	-19.25		

DEP		0.00108***	
		-19.3	
DIG			-0.00005
			(-0.38)
SIZE	0.54934***	0.55494***	0.55322***
	-96.07	-97.65	-97.59
LEV	0.18006***	0.16475***	0.17315***
	-8.24	-7.56	-7.97
TOP	0.00169***	0.00178***	0.00143***
	-4.45	-4.68	-3.81
Growth	0.25049***	0.24222***	0.24808***
	-47.72	-46.24	-47.3
CFO	0.75989***	0.77110***	0.76861***
	-21.1	-21.41	-21.53
AGE	-0.13209***	0.16605***	-0.057
	(-2.94)	-5.31	(-1.22)
BSIZE	0.11749***	0.10386***	0.09422***
	-5.94	-5.26	-4.83
MSR	-0.00088***	-0.00068**	-0.00087***
	(-3.01)	(-2.35)	(-3.03)
_cons	-4.42510***	-5.26343***	-4.42510***
	(-27.97)	(-40.26)	(-26.46)
N	22291	22291	22291
R^2	0.658	0.658	0.669
Year	Yes	Yes	Yes
FE	Yes	Yes	Yes

Note: Values in brackets are *t* values. "***", "**" and "*" respectively indicate that the indicators are significant at the level of 1%, 5% and 10%.

From the results shown in **Table 9**, The first sub-index, BRE, shows a significant positive effect on TFPLP, with a coefficient of 0.00162, indicating that broader digital financial inclusion is positively associated with higher-quality firm development. The negative t-value (-19.25) suggests a robust result, significant at the 1% level, highlighting the importance of financial inclusiveness breadth in fostering firm performance.

For the second sub-index, DEP, the coefficient of 0.00108 also shows a positive and significant impact on TFPLP, indicating that greater depth in digital financial inclusion enhances high-quality firm development. The corresponding t-value (-19.3) confirms the statistical significance at the 5% level, reinforcing the finding that a deeper penetration of digital financial services correlates with improved firm outcomes.

The third sub-index, DIG, representing the degree of digitalization, shows an insignificant effect on TFPLP, with a coefficient close to zero (-0.00005) and a t-value of -0.38. This result implies that the degree of digitalization alone may not significantly impact high-quality firm development. The lack of significance suggests that while the breadth and depth of digital financial inclusion are important, the mere presence of digitalization may not be sufficient to influence firm performance outcomes.

Table 10. Heterogeneity: ownership, life circles of enterprises and technical level.

Variables	Enterprise property rights		Life circles of enterprises			Technical level	
	State-owned enterprises	Private enterprises	Growth	Maturity	Decline	High	Low
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
DIF	-0.0002	0.00131***	0.00147***	0.00135***	0.00164	0.00141***	0.00068
	(-0.45)	-3.3	-2.88	-10.89	-1.34	-13.95	-1.41
SIZE	0.54363***	0.54218***	0.53698***	0.60551***	0.60372***	0.49410***	0.59772***
	-51.41	-76.57	-56.29	-64	-14.83	-70.02	-60.11
LEV	-0.05308	0.25763***	0.17394***	0.08748**	0.31153***	0.22631***	0.03378
	(-1.31)	-9.76	-4.79	-2.4	-2.62	-8.67	-0.89
TOP	0.00327***	0.00079	0.00091	0.00071	0.00087	0.00003	0.00484***
	-5.14	-1.6	-1.43	-1.22	-0.45	-0.07	-7.26
GROWTH	0.24623***	0.24716***	0.21840***	0.26213***	0.30854***	0.24164***	0.25150***
	-26.87	-39.02	-26.69	-27.53	-12.97	-37.09	-31
CFO	0.78223***	0.75727***	0.67125***	0.96185***	0.36368**	0.73889***	0.73120***
	-12.68	-17.39	-9.62	-14.45	-2.38	-17.08	-12.9
AGE	-0.12749	-0.04983	-0.00629	-0.11303*	-0.11108	0.08513*	-0.23939***
	(-1.40)	(-0.88)	(-0.08)	(-1.80)	(-0.43)	-1.68	(-3.02)
BSIZE	0.05758*	0.10148***	0.06624*	0.07959***	0.19103**	0.13058***	0.04644
	-1.69	-4.16	-1.93	-2.66	-2.39	-5.66	-1.41
MSR	0.00161	-0.0005	-0.00099**	0.00052	-0.00276	-0.00033	0.00007
	-0.68	(-1.64)	(-2.00)	-1.2	(-1.57)	(-1.03)	-0.12
_cons	-3.86543***	-4.35279***	-4.19802***	-5.55164***	-5.66525***	-3.80063***	-4.93546***
	(-11.41)	(-21.63)	(-14.69)	(-23.50)	(-5.51)	(-21.35)	(-16.25)
N	7510	14781	9040	8648	4603	13608	8683
R^2	0.625	0.684	0.704	0.726	0.551	0.684	0.608
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

6. 2. Heterogeneity test analysis for state-owned and private enterprises

State-controlled and non-state- controlled enterprises differ significantly in terms of resource access, policy support, and market orientation, making it essential to conduct a heterogeneity test between these two types of enterprises.

According to the analysis of columns (1)-(2) in **Table 10**, the impact of DIF on state-controlled and non-state- controlled enterprises shows significant differences. Non-state-owned enterprises are more sensitive to DIF, with an impact coefficient of 0.00131 and a t-value of 3.30, indicating that DIF significantly promotes their high-quality development. In contrast, state-controlled enterprises have an impact coefficient of -0.00020 and a t-value of -0.45, showing that the impact of DIF on them is not significant. Non-state-controlled enterprises face greater financing pressure in market competition and can better leverage DIF to acquire resources and drive growth, whereas state-owned enterprises, with stable financing channels and government support, experience a relatively small marginal effect from DIF.

6. 3. Heterogeneity test analysis for different life circles of enterprises

Enterprises face different challenges and development needs at different stages of their lifecycle (growth, maturity, decline), making it necessary to conduct a heterogeneity test across these stages[49].

The data of columns (3)-(5) in **Table 10** reveals that DIF's impact on high-quality development varies across different stages of the corporate lifecycle. Growing companies (coefficient 0.00147, t-value 2.88) and mature companies (coefficient 0.00135, t-value 10.89) are significantly promoted by DIF, likely because these companies are at critical stages of expansion or market consolidation and thus have a greater need for external funds and resources. In contrast, declining companies have an impact coefficient of 0.00164 and a t-value of 1.34, which is not statistically significant, indicating that DIF's effectiveness in aiding company transformation or recovery is limited.

6. 4. Heterogeneity test analysis for high and low technical enterprises

Conducting a heterogeneity test between high-tech and low-tech enterprises is necessary because DIF may have different impacts on companies with varying levels of technological sophistication[50]. High-tech enterprises, which often rely more on innovation and R&D, may more easily obtain the funding and resources they need for innovation through digital finance, thereby promoting high-quality development[51]. In contrast, low-tech enterprises may face more technological bottlenecks, and even with financial support, their ability to translate that into high-quality development might be limited. This test helps to clarify the differential impact of digital finance on enterprises at various technology levels.

According to the analysis of columns (6)-(7) in **Table 10**, DIF has a significant and positive impact on high-tech enterprises, with an impact coefficient of 0.00141 and a t-value of 13.95, suggesting that DIF effectively supports innovation and market expansion in these companies. For low-tech enterprises, however, the impact coefficient is 0.00068 and the t-value is 1.41, indicating that DIF does not significantly promote their high-quality development. This could be because high-tech enterprises have greater innovation needs and market potential, enabling them to significantly enhance their competitiveness with DIF support, whereas low-tech enterprises, with limited overall development potential, see relatively minimal benefits from DIF even with financial support.

6. 5 Summary of traditional economic empirical analysis

In the above experiment, we analyzed the linear relationship between DIF and TFPLP using traditional econometric methods like the dual fixed-effects model, mediation effect model, and heterogeneity robustness checks. We found that DIF has a positive effect on TFPLP. This shows that changes in DIF can help improve TFPLP in traditional economic models. But traditional models have limits in explaining complex nonlinear relationships. To explore the possible nonlinear links between DIF and TFPLP, we will use deep learning time series models. Deep learning methods are better at finding nonlinear patterns in data, giving us clearer predictions for understanding the complex relationship between the two. Therefore, we will use deep learning models in the next experiments to look deeper into this relationship.

7. The application of deep learning models in prediction of TFPLP

In this section, based on the double fixed-effects model used in performing the linear regression analysis of DIF and TFPLP, their deeper nonlinear relationship is explored with four classic time series models: LSTM, BiLSTM, GRU, and Transformer, for prediction. The KAN and GCN methods are employed to improve the accuracy of the four models for the prediction and explanation of nonlinear relationships between DIF and TFPLP. It begins the chapter with a review of four classic models and presents the experiments. Basic concepts are then introduced regarding KAN and GCN on how they enhance the models. Four evaluation metrics are used for analysis, containing R^2, RMSE, MSE, MAE, to demonstrate the improvements in predictive performance for enhanced models and further probe into the complicated nonlinear relationship between DIF and TFPLP.

7. 1 The overview of 4 classical time series models

This section provides an introduction to four classic time series models: LSTM, BiLSTM, GRU, and Transformer, highlighting their key features and differences in handling sequential data.

7. 1. 1 LSTM (Long Short-Term Memory)

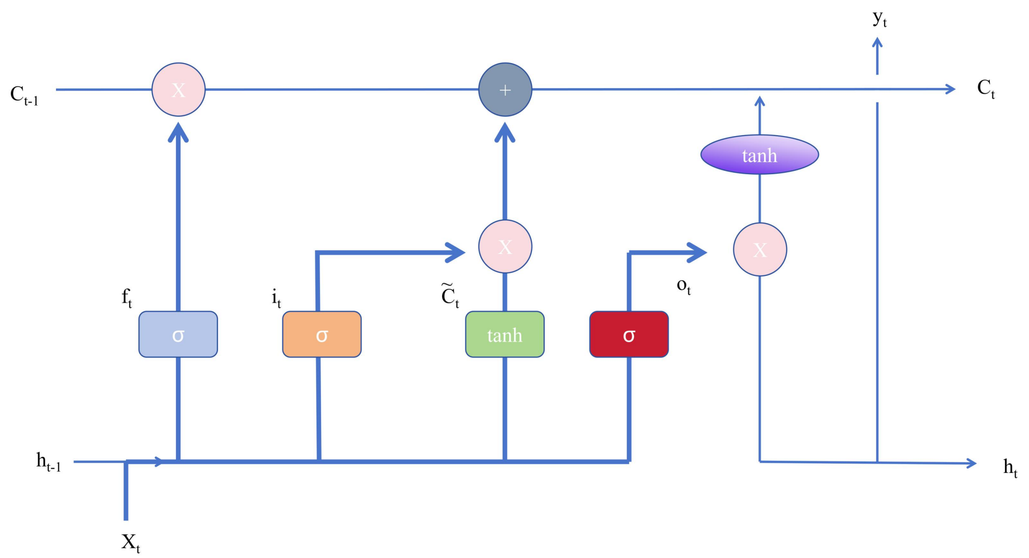


Fig. 7. The architecture of LSTM model.

Improved variants of the RNNs include the LSTM, as presented in **Fig. 7**, which was specifically designed to handle the issues of vanishing and exploding gradients experienced by traditional RNNs when operating on data with longer sequences. It uses three different gates for controlling the information flow in a way that the network can learn and remember long-term dependencies. Three gates in LSTM are:

Forget Gate: It has to decide which information from previous time step memory has to be discarded.

Input Gate: It determines the portions of current that are to stay in the memory.

Output Gate: The gate is used to decide what information in memory will flow to the next time step hidden state.

Because the state of memory cells carries information, the LSTM can keep it across time steps. In a traditional RNN, LSTM can avoid the problem of vanishing gradients by the balanced work of the forget gate, input gate, and output gate[16].

7. 1. 2 BILSTM (Bidirectional Long Short-Term Memory)

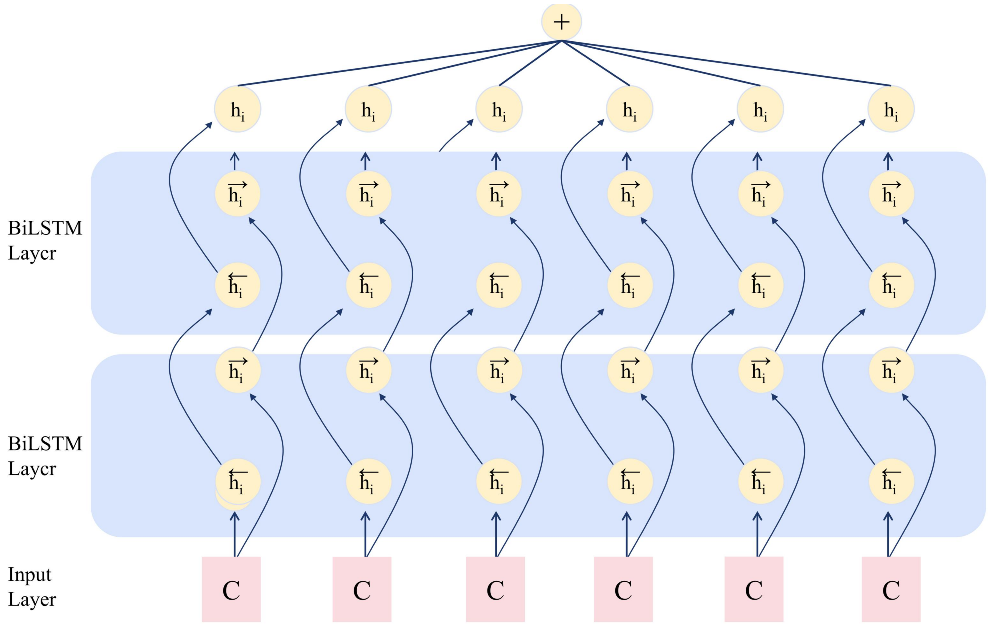


Fig. 8. The architecture of BiLSTM model.

BiLSTM is an extension of the standard LSTM that enhances its capacity to capture both past and future context within sequence data. Its architecture is presented in **Fig. 8**, unlike a standard LSTM, which processes data in a single direction (from the beginning to the end of a sequence), BiLSTM processes data in both forward and backward directions[17].

The bidirectional LSTM consists of two components: one that processes the sequence from start to end (forward LSTM) and another that processes it from end to start (backward LSTM). This approach enables the network to access both past and future context for each time step, making it particularly valuable for tasks where understanding both prior and subsequent information is crucial.

7. 1. 3 GRU (Gated Recurrent Unit)

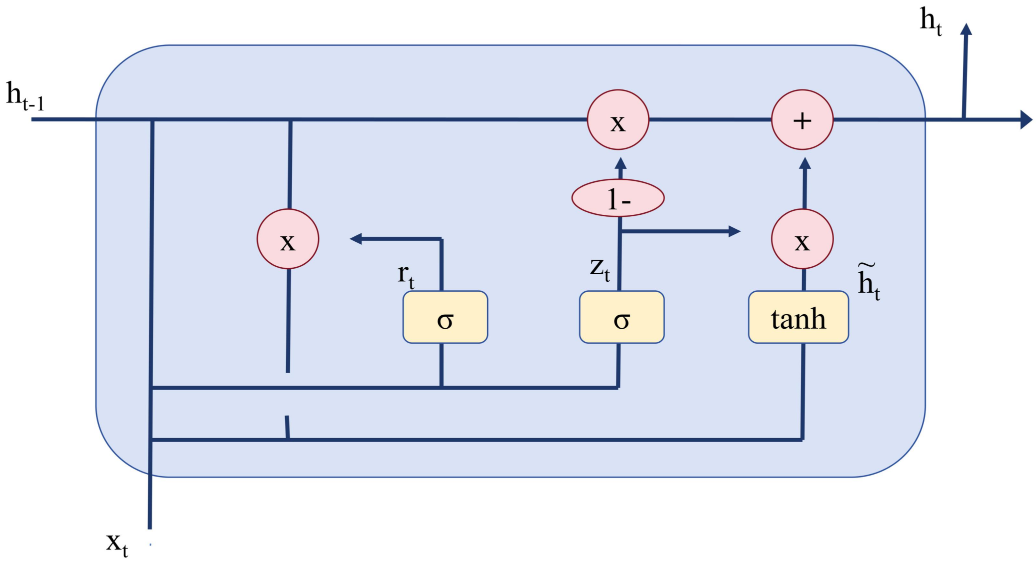


Fig. 9. The architecture of GRU model.

GRU is a variant of the standard LSTM, and it is designed to streamline the architecture while preserving the ability to capture long-term dependencies in sequence data[18]. Like LSTM, as shown in **Fig. 9**, GRU is a type of recurrent neural network (RNN), but it uses fewer gates and has a simpler structure, which makes it computationally more efficient. GRU operates with two primary gates:

Update Gate: The update gate regulates the proportion of the previous memory (hidden state) to retain and determines how much of the new information should be incorporated[52].

Reset Gate: The reset gate controls the extent to which the previous memory should be forgotten. It determines the amount of information to discard from the previous time step when calculating the current hidden state.

In contrast to LSTM, which has three gates (input, forget, and output), the GRU combines the forget and input gates into a single update gate, which makes it more computationally efficient.

7. 1. 4 Transformer

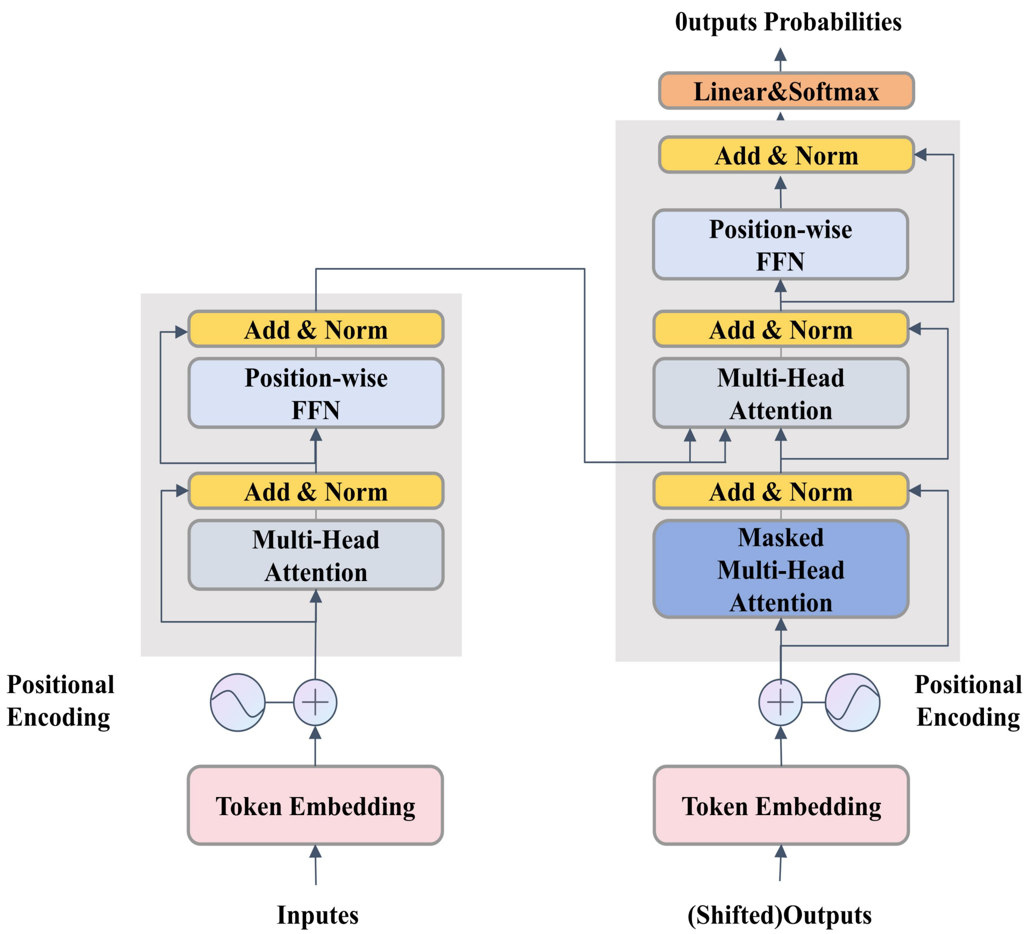


Fig. 10. The architecture of Transformer model.

Transformer is a deep learning model architecture (shown in **Fig. 10**) that relies entirely on a self-attention mechanism, bypassing the need for traditional recurrent or convolutional layers. Originally introduced in the paper "Attention is All You Need," it has since become the foundation for many state-of-the-art models in natural language processing (NLP) and other sequence-based tasks.

The key feature of the Transformer is its self-attention mechanism[53], enabling the model to evaluate the importance of various elements in the input sequence in relation to one another, regardless of their positions[54]. This mechanism allows the Transformer to model the dependencies in data long-range without the drawback of step-by-step processing present in RNNs or LSTM. The result is that the Transformer processes data in parallel and is faster compared to RNN-based models when training[55]. Moreover, since it does not rely on sequential processing of data, the Transformer can handle long-range dependencies much more transparently than what is usually hard to achieve by RNNs and LSTM.

7. 2 Experimental methods and data

In this experiment, we aim to improve time series prediction performance by enhancing the KAN (Kolmogorov–Arnold Network) and GCN (Graph Convolutional Network) models. The experiment involves two series of improved models: KAN Improved Model Series and GCN Improved Model Series, each of which enhances traditional time series models (LSTM, Transformer, BiLSTM, GRU) with specific processing methods. Below, we describe each series in detail.

This study further processes the experimental data (**Table 1**) by selecting corporate annual observation data from the complete period of 2011 to 2022, covering a wide range of industries and regions to ensure the representativeness and breadth of the research results.

7. 3 Performance valuation metrics

To comprehensively assess the effectiveness of models on time series data, this section employs three commonly used metrics to compare and analyze the effectiveness of each model. These metrics are R^2 (coefficient of determination), RMSE (root mean square error), and MSE (mean square error). By integrating these three metrics, the evaluation provides a well-rounded assessment of each model's performance in time series forecasting tasks, ensuring the objectivity and reliability of the results.

- **R^2:** R^2 represents the proportion of variance in the dependent variable that is explained by the independent variables. Its value ranges from 0 to 1, with higher values indicating that the model captures a larger portion of the variance, reflecting stronger predictive power.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2} \tag{133}$$

where y_i stands for the actual value, $\hat{y}_i$ stands for the predicted value, $\overline{y}$ stands for the average value of the actual values, and n denotes the sample size.

- **RMSE:** The RMSE on the test set assesses the model's ability to generalize to new data. A strong model should not only perform well on training data but, more importantly, maintain consistent performance on previously unseen data.

$$RMSE_{test} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Y_{test,i} - Y_{pred,test,i})^2} \tag{134}$$

where N stands for the quantity of samples within the test set, and the formula uses $Y_{train,\,i}$ and $Y_{pred,\ train,\,i}$ to calculate the mean value of squared discrepancies between the actual and predicted values for each sample in the test set, evaluating the model's capability of generalization.

- **MSE:** MSE is the average of the squared differences between predicted and actual values, reflecting the overall prediction error in the model. Like RMSE, MSE is sensitive to large errors, making it an effective metric for identifying significant prediction deviations. Smaller MSE values suggest better prediction accuracy.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{135}$$

where y_i represents the actual value, $\hat{y}_i$ represents the predicted value, and *n* denotes the sample size.

- **MAE:** MAE calculates the average magnitude of errors in predictions, disregarding their direction. It is the mean of the absolute differences between predicted and actual values across the test sample, offering a simple interpretation of the average error size.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \tag{136}$$

where y_i represents the actual value, $\hat{y}_i$ represents the predicted value, and *n* denotes the sample size.

7. 4 The analysis of results in 4 classical time series models

Table 11 The comparison results of 4 classical time series models.

Models	The results in training sets				The results in test sets			
	RMSE	R^2	MSE	MAE	RMSE	R^2	MSE	MAE
Transformer	0.0055	0.9876	0.0055	0.0491	**0.4638**	**0.8525**	**0.2151**	**0.3447**
BiLSTM	0.0044	0.9901	0.0044	0.0337	0.6685	0.6935	0.4469	0.4985
GRU	**0.0038**	**0.9914**	**0.0038**	**0.0264**	0.6942	0.6695	0.4820	0.5069
LSTM	0.0053	0.9880	0.0053	0.0422	0.6389	0.7201	0.4082	0.4743

From **Table 11**, it is clear that the four classical time series models show differences in performance on the training and test sets. First, BiLSTM performs the best on the training set, with the lowest RMSE (0.0044), MSE (0.0337), and MAE (0.0337), and the highest R^2 value (0.9901), indicating its strongest fitting ability on the training data. However, when it comes to the test set, Transformer shows the best performance, with the lowest RMSE (0.4638), MSE (0.2151), and MAE (0.3447), as well as the highest R^2 value (0.8525), indicating the strongest generalization ability on the test data. In contrast, although BiLSTM performs best on the training set, its performance on the test set is slightly weaker, with RMSE (0.6685), MAE (0.4985), and R^2 value (0.8525) lower than Transformer. GRU and LSTM perform relatively poorly on the test set, with GRU showing the highest RMSE (0.6942) and MAE (0.5069), indicating its lower accuracy in handling the test data. Overall, while BiLSTM performs best on the training set, Transformer outperforms all models on the test set, demonstrating the strongest generalization ability.

7. 5 The Enhanced Deep Learning Framework Based on KAN

To better understand the application of KAN in enhancing the four previously mentioned models, this chapter first introduces the KAN model and compares it with the traditional MLP model in neural networks. Then, it presents the mechanisms and methods of the four modified models individually.

7. 5. 1 The KAN and MLP models

- Kolmogorov–Arnold Neural Network (KAN) is built upon the Kolmogorov-Arnold representation theorem, which decomposes a multivariable continuous function into a superposition and composition of univariate functions, enabling the approximation of any continuous function. The theorem states that any continuous function $f(x_1, x_2, \ldots, x_n)$ can be represented as:

$$f(x) = \sum_{q=1}^{2n+1} \Phi_q \left(\sum_{p=1}^{n} \phi_{q,p}(x_p) \right) \tag{7}$$

where $\phi_{q,p}$ are nonlinear transformation functions applied to the input x_p, and Φ_q are nonlinear functions for final aggregation. In KAN, this decomposition can replace the fully connected layers (MLP), with the implementation process as follows:

1) **First Layer** (Transformation Layer): Each input variable x_p undergoes an independent nonlinear transformation through the univariate functions $\phi_{q,p}$ (xp).

2) **Second Layer** (Aggregation Layer): The outputs from the first layer are summed and passed through nonlinear functions Φ_q to construct the final function representation.

As presented in **Fig. 11**, learnable activation function of KAN applied to the edges: The activation function is dynamically learnable and acts on the weights of the edges, while each node performs a simple summation operation. Furthermore, the KAN structure, from **Fig. 12**, is closer to the implementation of mathematical theory, using learnable activation functions. The activation functions apply to the non-linear mapping part of the deep structure, while the linear transformation part is fixed.

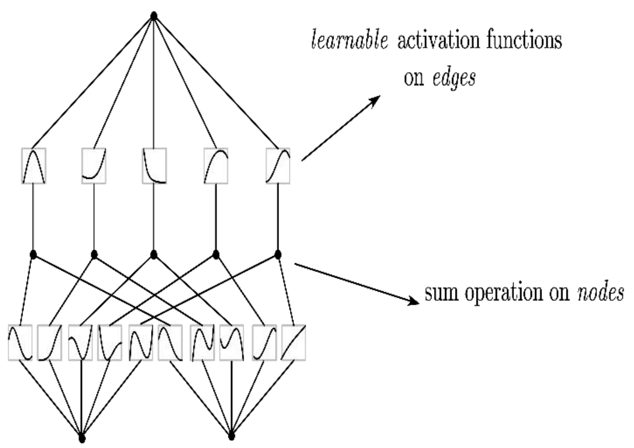


Fig. 11. The shallow architecture of KAN.

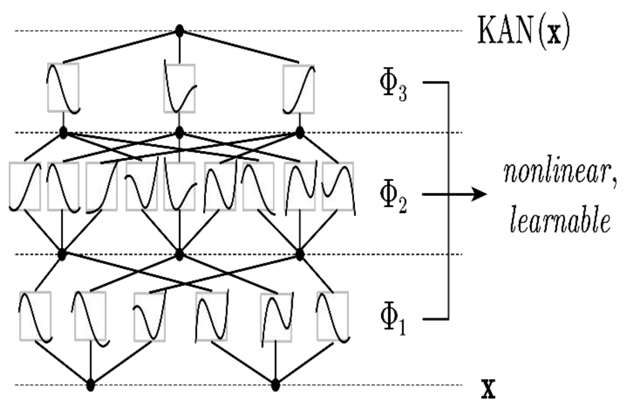


Fig. 12. The deep architecture of KAN.

- Multi-Layer Perceptron (MLP) is based on the Universal Approximation Theorem. As shown in **Fig. 13** the activation function fixed at the node: Each node (neuron) in the network has a fixed activation function (such as ReLU, Sigmoid, etc.), while the edges (connection weights) are learnable parameters and we can see that from **Fig. 14**, MLP consists of alternating linear transformations (weight matrices) and fixed non-linear transformations (activation functions), with parameters at each layer being learnable. This ultimately enables complex feature extraction and mapping.

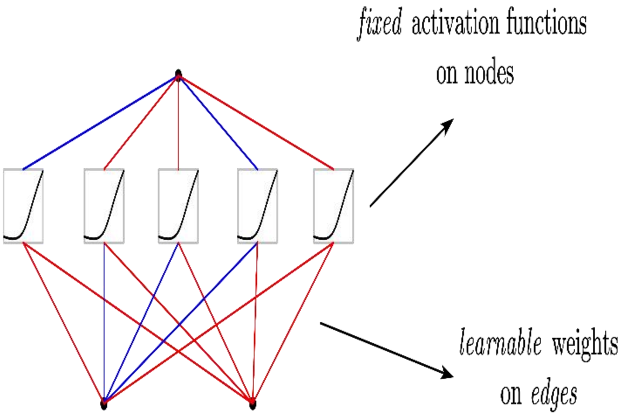


Fig. 13. The shallow architecture of MLP.

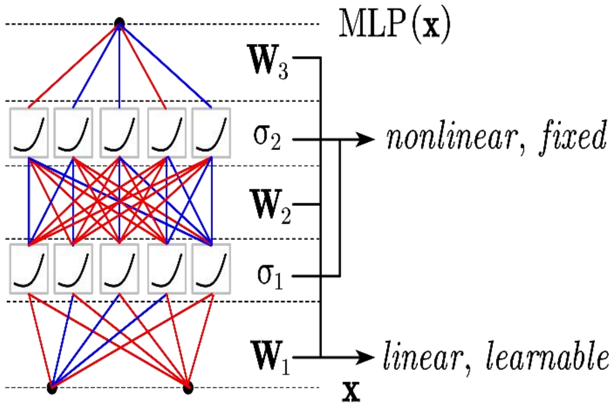


Fig. 14. The deep architecture of MLP.

This KAN structure avoids the redundant fully connected operations of traditional MLPs, improves representation efficiency, reduces parameter count, and still retains the ability to approximate any continuous function.

7. 5. 2 The experimental methods of the enhanced deep learning models based on KAN

In the KAN improved model series, we focus on replacing the fully connected layer in traditional time series models with KAN, leveraging its unique structure to improve model expressiveness. The structure is as follows:

- **Data Preprocessing:** Data Cleaning and Denoising: First, we clean the enterprise time series data, which involves handling missing values (e.g., using interpolation, mean imputation, etc.), removing outliers (e.g., using Z-scores or boxplots to identify and remove abnormal data), and standardizing/normalizing the data (scaling the data to a specific range, which helps with faster training in deep learning models).
- **Model Design:** In traditional time series models (LSTM, Transformer, BiLSTM, GRU), a fully connected layer (FC) is typically used for prediction at the final stage. In this experiment, we propose replacing the fully connected layer with KAN (Kolmogorov–Arnold Network).
- **Train-Test Split:** The dataset is split into training and testing sets with a ratio of 8:2, where 80% of the data will be used for training and the remaining 20% will be used for testing. This guarantees the model is trained on a sufficiently large portion of the data while keeping a portion aside for evaluation to assess the model’s generalization capability.
- **Objective and Hypothesis**: The goal is to enhance the model's capability of capturing nonlinear relationships and complex dependencies by substituting the fully connected layer with KAN, thereby enhancing prediction accuracy in time series forecasting. The hypothesis is that KAN can handle the complexity of time series data more effectively, especially in high-dimensional, multivariate scenarios.

7. 5. 3 The overview of the enhanced deep learning models based on KAN

This section introduces the mechanisms of the four improved models. **Fig. 15** clearly illustrates the experimental process in which KAN replaces the fully connected layer in each of the four models, ultimately generating the final output.

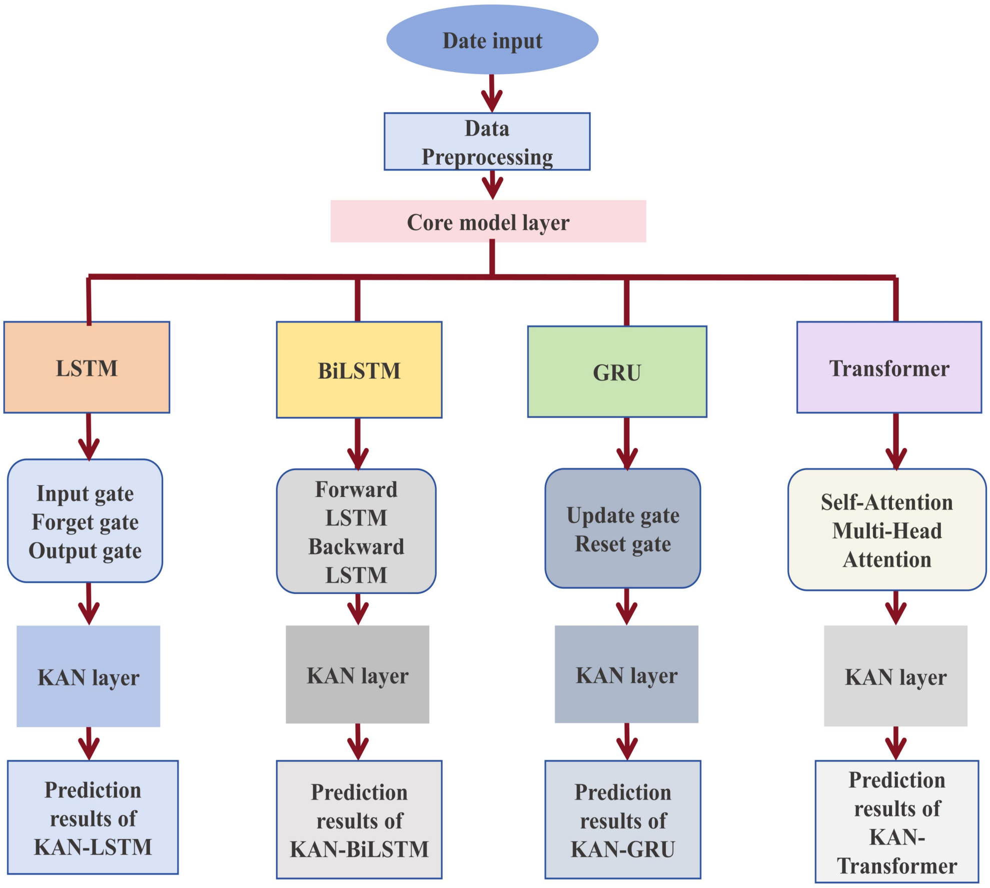


Fig. 15. The experimental flowchart of improved KAN series models.

- **KAN-LSTM model:** KAN-LSTM is a combined neural network model that unites LSTM with KAN, joining their powers to process data with time series dependence in an intricate manner. The model was based on the concept that LSTM might capture the dynamic relationships of time series data, while KAN replaces a traditional fully connected layer with nonlinear transformations that enhance the representation of the model.
- **KAN-BiLSTM model:** The KAN-BiLSTM model combines the merits of BiLSTM and KAN to enhance the ability of the model in handling complex sequential data through bidirectional temporal modeling and strong nonlinear mapping. BiLSTM captures the bidirectional dependencies in time series, while KAN replaces the traditional fully connected layer with nonlinear transformations, improving the predictive power of the model.
- **KAN-GRU model:** The KAN-GRU model represents the combination of strengths of GRU and KAN by improving the capability of prediction over sequential data based on efficient temporal modeling and powerful nonlinear mapping. GRU extracts temporal dependencies, while KAN, which replaces the fully connected layer in the traditional setting, enhances predictiveness through nonlinear mapping.
- **KAN-Transformer model:** The Transformer and KAN model combine to improve the model's performance through effective operations on complex sequential data and tasks. Transformer captures long-range dependencies because of its self-attention mechanisms, while KAN provides nonlinear transformations over feature representations to enhance its predictive

power.

7. 5. 4 The comparison analysis of results in enhanced deep learning models based on KAN

Table 12 The comparison results of performance in 4 improved KAN time series models.

Models	The results in training sets				The results in test sets			
	RMSE	R^2	MSE	MAE	RMSE	R^2	MSE	MAE
Transformer	0.0055	0.9876	0.0055	0.0491	0.4638	0.8525	0.2151	0.3447
KAN-Transformer	0.0044	0.99	0.0044	0.0386	**0.4497**	**0.8613**	**0.2023**	**0.3382**
BiLSTM	0.0044	0.9901	0.0044	0.0337	0.6685	0.6935	0.4469	0.4985
KAN-BiLSTM	**0.0037**	**0.9917**	**0.0037**	**0.0255**	0.5345	0.8041	0.2857	0.3924
GRU	0.0038	0.9914	0.0038	0.0264	0.6942	0.6695	0.482	0.5069
KAN-GRU	0.0046	0.9899	0.0046	0.041	0.4598	0.855	0.2114	0.3441
LSTM	0.0053	0.988	0.0053	0.0422	0.6389	0.7201	0.4082	0.4743
KAN-LSTM	0.004	0.991	0.004	0.0292	0.5987	0.7542	0.3584	0.4436

From **Table 12 and Fig. 16**, it can be seen that KAN-BiLSTM performs better than BiLSTM on the training set, though the difference is minimal. Both models show similar RMSE, MSE, and MAE, with KAN-BiLSTM having a slightly higher R^2 value (0.9917) compared to BiLSTM (0.9901), indicating a small improvement. However, on the test set, KAN-Transformer performs the best, with the lowest RMSE (0.4497), MSE (0.2023), and MAE (0.3382), as well as the highest R^2 value (0.8613), demonstrating the best generalization ability. KAN-BiLSTM still outperforms BiLSTM on the test set, with RMSE (0.5345), MAE (0.3924), and R^2 value of 0.8613. Although it does not outperform KAN-Transformer, it still performs exceptionally well compared to other models. In contrast, KAN-GRU and KAN-LSTM perform relatively worse on the test set, especially in RMSE and MSE, and fall behind KAN-Transformer and KAN-BiLSTM. Overall, models with KAN integration generally show improved performance on the test set, with KAN-Transformer achieving the best performance, and **Fig. 17** shows the comparison results of true values and predictions.

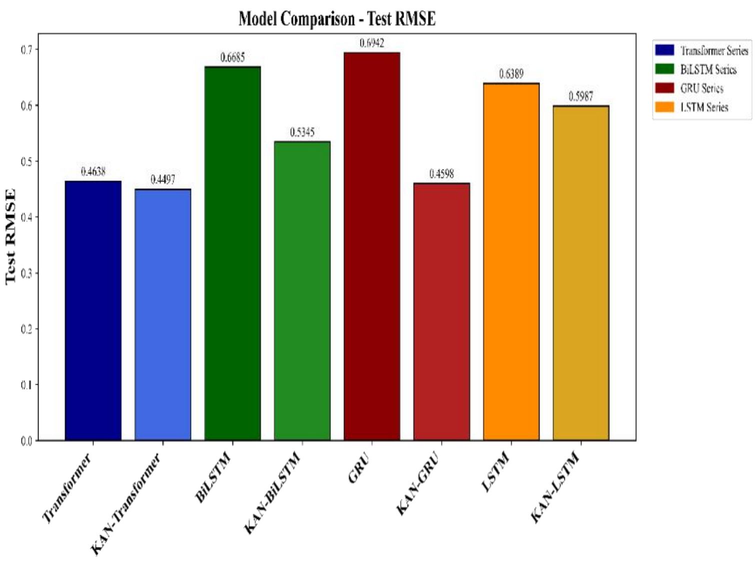


Fig. 16-a. The result of RMSE in KAN series models.

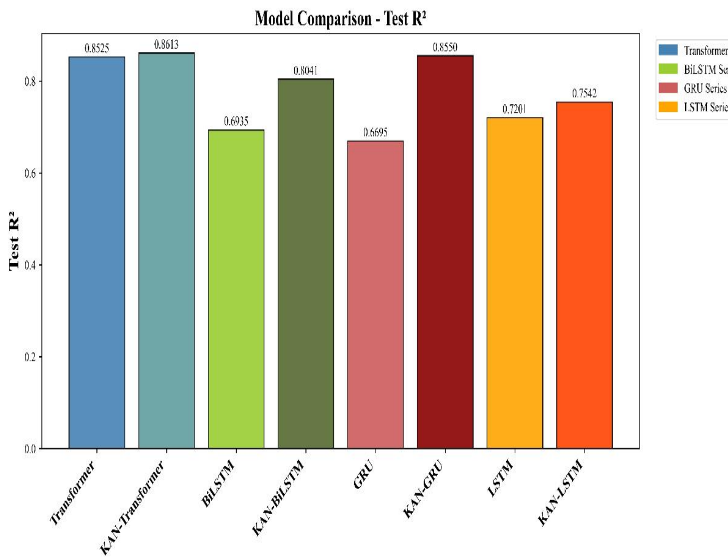


Fig. 16-b. The result of R^2 in KAN series models.

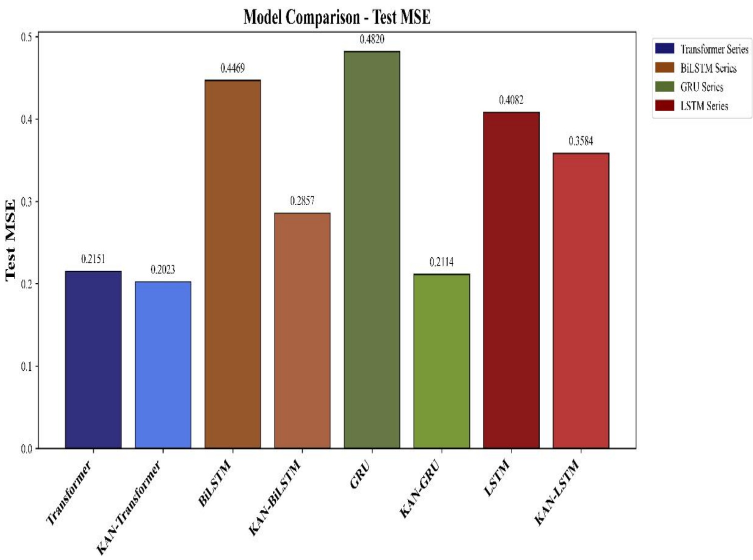


Fig. 16-c. The result of MSE in KAN series models.

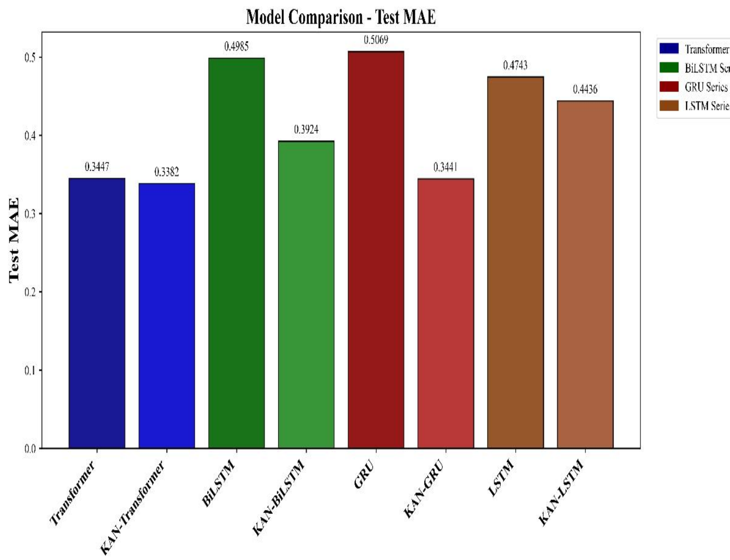
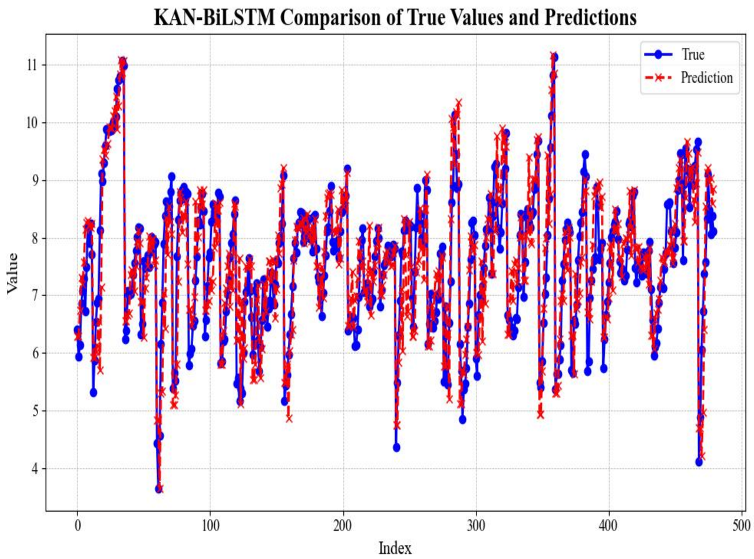
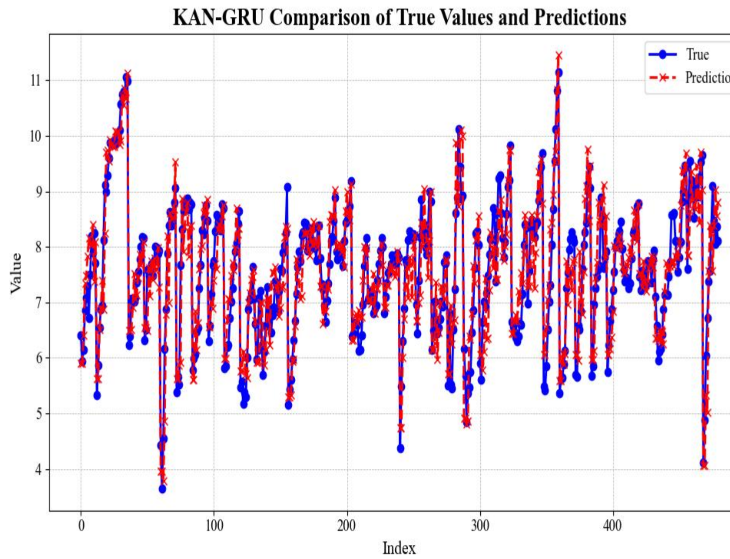


Fig. 16-d. The result of MAE in KAN series models.





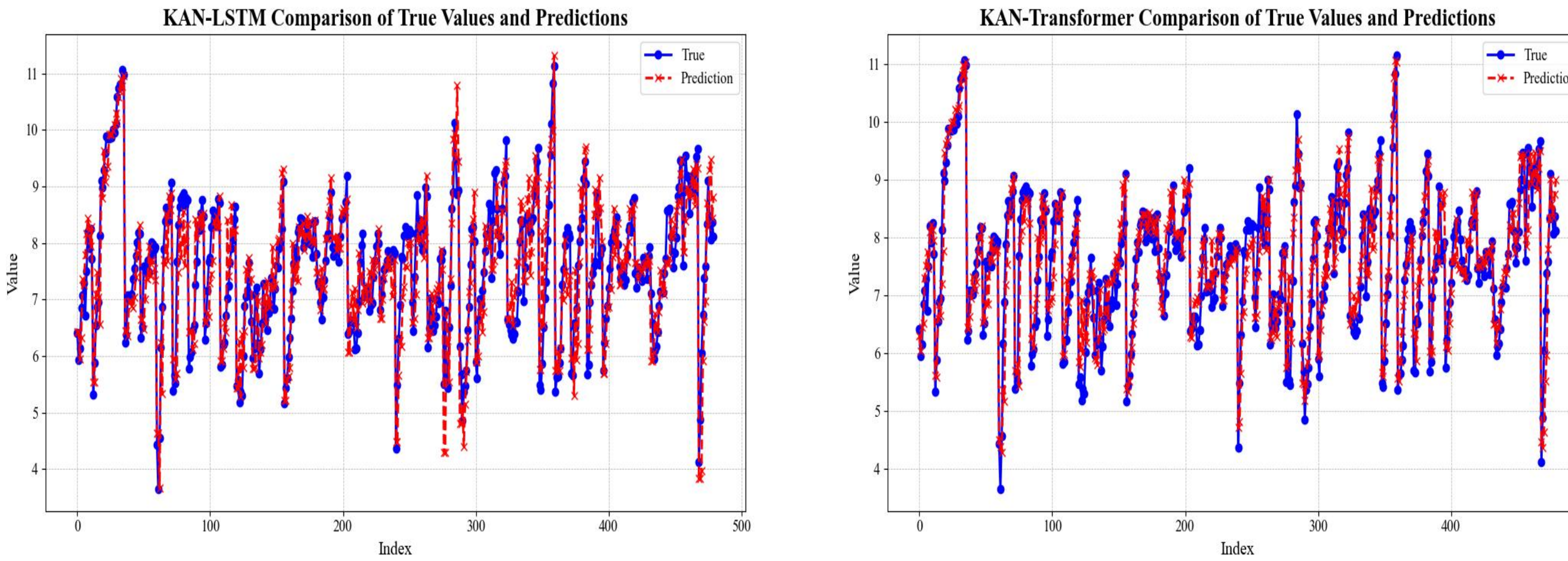


Fig. 17. The comparison results of true values and predictions in 4 KAN series models.

7. 6 The enhanced deep learning framework based on GCN

In addition to using KAN to improve the four models mentioned, we also employ Graph Convolutional Networks (GCN) to enhance these models. This section will first introduce the basic concept of GCN, followed by a detailed explanation of the mechanisms of the four GCN-based modified models.

7. 6. 1 GCN model

Graph Convolutional Networks (GCN), the architecture is provided in **Fig. 18**, are one of deep learning models designed for graph-structured data. In the context of this study, we leverage GCN to model multi-enterprise digital financial data (DIF), where nodes represent variables such as DIF, eight control variables (including Company Size, Leverage Ratio, Ownership Concentration, Business Growth, Company Age, Board Size, and Management Shareholding), and the target variable TFPLP. Edges in this graph represent the relationships between these variables, capturing both statistical correlations and domain-specific prior knowledge. The primary target of GCN in this context is to capture the representation of each variable (node) by aggregating information from its related variables (neighboring nodes) in the graph. In GCN, the feature information of each node is updated by not only considering its own value (e.g., a variable's time-series data at a specific time step) but also by propagating the features from neighboring nodes that share relationships with it. This propagation allows GCN to explicitly capture the dependencies between input variables (DIF and control variables) and their combined impact on the output variable TFPLP.

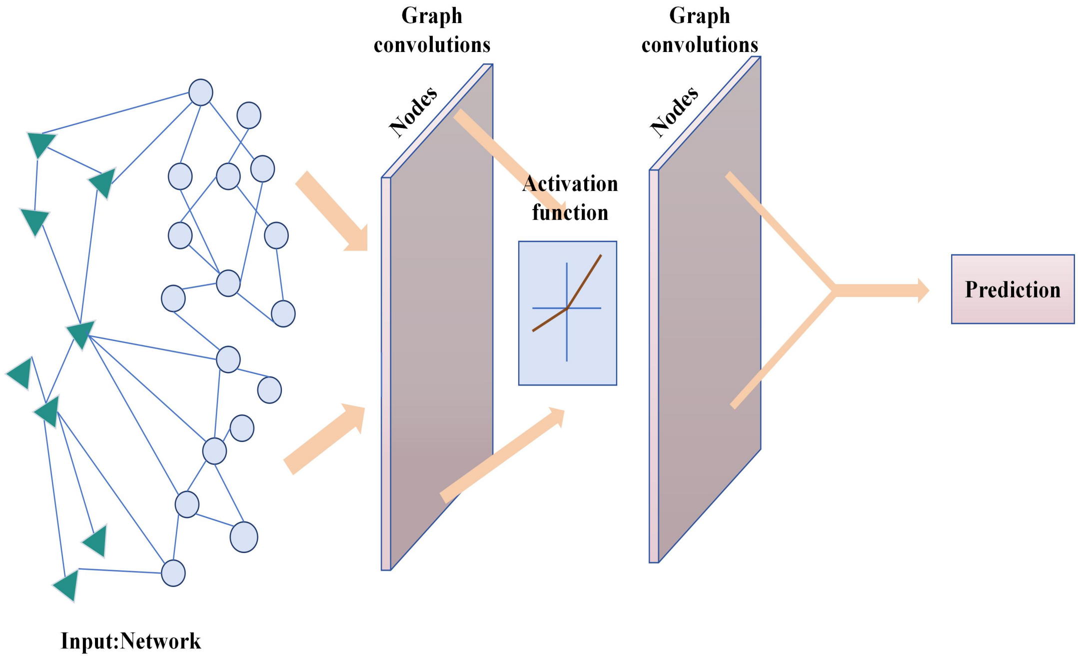


Fig. 18. The architecture of GCN model.

This information propagation is done through convolution operations, similar to those in traditional Convolutional Neural Networks (CNNs). However, instead of working with grid-structured data like images, GCN works on nodes and their connections in a graph. The main idea of GCN is to update the feature representation of each node by applying a convolution operation using the adjacency matrix, which shows the relationships between nodes. This adjacency matrix can include both statistical relationships (like correlation or mutual information) and prior knowledge about variable dependencies, making GCN well-suited for modeling the complex interactions in digital financial data. Assume the graph has N nodes, and each node has a feature vector of dimension D. The node feature matrix is $X \in R$^{$N \times D$} and the adjacency matrix of the graph is $A \in R$^{$N \times N$}. The update formula for GCN is as follows:

$$H^{(l+1)} = \sigma\left(\hat{A}H^{(l)}W^{(l)}\right) \tag{8}$$

where, $H^{(l)}$ represents the node feature matrix at the *l-th* layer, and $H^{(l)} = X$ is the input feature matrix of the nodes. $\hat{A} = D^{-1/2}, A\ D^{-1/2}$ is the normalized adjacency matrix, where A is the original adjacency matrix, and D is the degree matrix. $W^{(l)}$ is the weight matrix at the *l-th* layer. σ is the activation function, typically ReLU or another nonlinear function.

With this formula, GCN updates the node features by aggregating information from neighboring nodes via the normalized adjacency matrix Â. The output of each layer serves as the input to the next layer, allowing the model to aggregate information from progressively larger neighborhoods of each node.

7. 6. 2 The experimental methods of the enhanced deep learning models based on GCN

In the GCN improved model series, we propose combining Graph Convolutional Networks (GCN) with traditional time series models, using graph-structured data to extract features that further improves the model's prediction capability and the process is provided in **Fig. 19**.

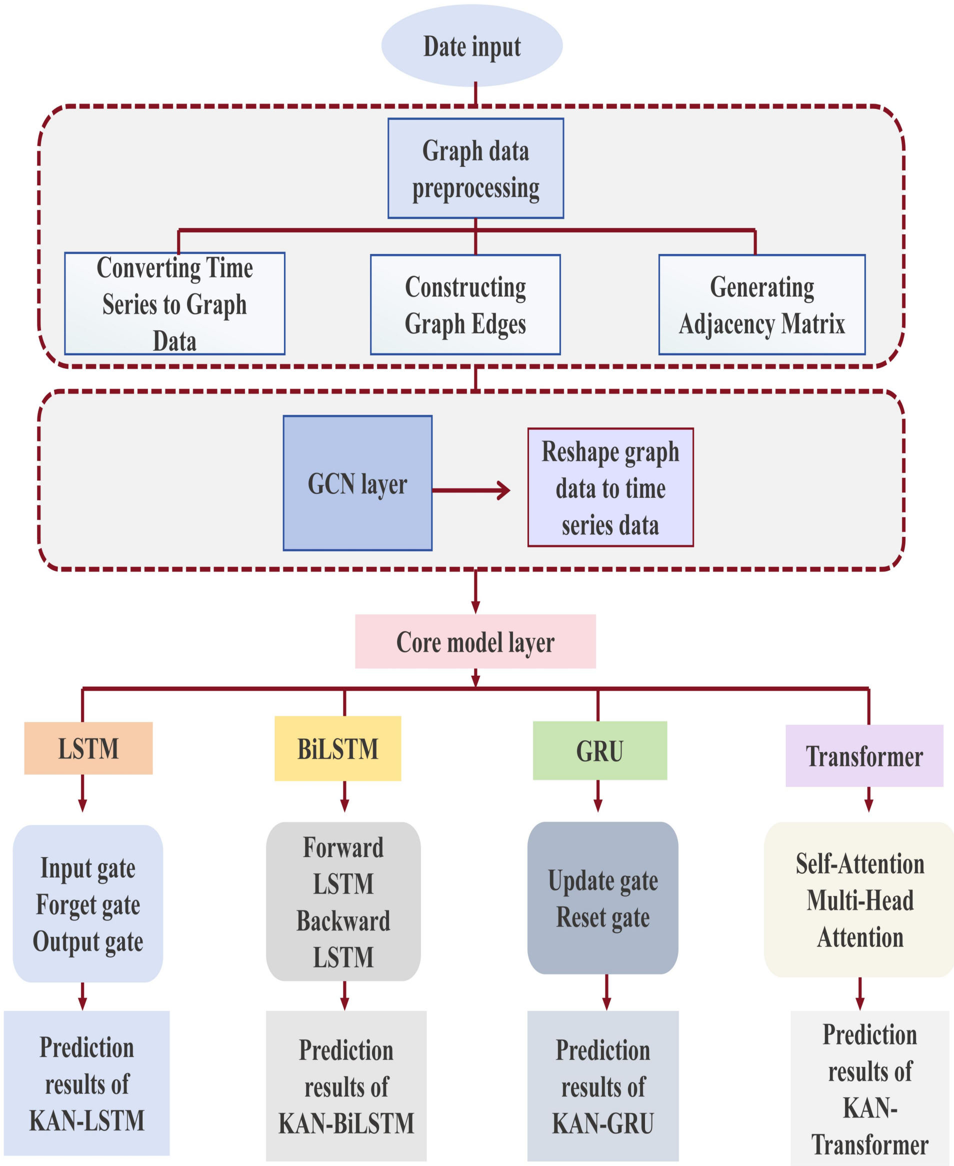


Fig. 19. The experimental flowchart of improved GCN series models.

- **Data Preprocessing:**

1) **Converting Time Series to Graph Data:** The enterprise time series data is first converted into graph data. Specifically, input variables (DIF and the other eight variables) as well as output variables (TFPLP) are treated as nodes in the graph.

2) **Constructing Graph Edges:** The correlation between variables (Pearson correlation) is used to build edges between the nodes. These edges reflect the interdependencies between different variables in the time series data.

3) **Generating Adjacency Matrix:** Based on the constructed edges, an adjacency matrix is created to represent the connections between nodes in the graph. This adjacency matrix serves as one of the inputs for the GCN to learn the relationships between nodes.

- **Model Design:**

1) Construct graph: Define nodes and edges based on time series data and variable correlations.

2) Graph convolution: Perform graph convolution to extract local and global structural information.

3) Feature extraction: GCN generates enriched node features, incorporating both individual and relational data.

4) Model integration: After data processing and nonlinear transformation, we pass GCN features into traditional time series models (LSTM, Transformer, BiLSTM, GRU) for further temporal processing.

5) Enhanced prediction: Leverage both graph and temporal information to improve model performance.

Train-Test Split: In the GCN improved model series, the dataset will be split into training and testing sets with a 9:1 ratio. This means most of the data will be used for training to ensure the model learns the complex relationships within the graph structure, while only a small portion will be reserved for testing and evaluation.

Objective and Hypothesis: The goal is to leverage the graph structure, combined with time series data's temporal dependencies, to improve the model's ability to capture both the relationships between variables and temporal patterns, thus enhancing prediction accuracy. The hypothesis is that the graph structure effectively captures complex dependencies between variables, and GCN can optimize this relationship modeling, leading to improved performance of traditional time series models.

7. 6. 3 The overview of the enhanced deep learning models based on GCN

- **GCN-LSTM**

The model begins with the input layer where graph-structured data that has been pre-processed is fed. The adjacency matrix captures the relationships and is used to extract features via graph convolutions in the GCN layer. The output features are passed through a Fully Connected Layer (FC) with Rectified Linear Unit (ReLU) activation to produce more general features for temporal modeling. The reshaped features are fed to the LSTM layer, which learns short and long term dependencies in the data. Finally, the output of the LSTM layer is passed to a Fully Connected Layer which would generate the final prediction.

- **GCN-BiLSTM**

The GCN-BiLSTM model is your time-series and graph structure time series processing, combining the data process strengths of GCN and BiLSTM. The first part of the model is a GCN layer that extracts relationships between nodes based on the graph's adjacency matrix, and determine high-dimensional feature representations for each node. Lastly, the input to the BiLSTM layer is obtained from the GCN layer output that goes through a Fully Connected Layer (Dense Layer), which uses nonlinear activation functions (like ReLU) to transform the features. As a result, the BiLSTM layer processes the transformed features, capturing both forward and backward temporal dependencies in the data and offering a comprehensive context for the temporal relationship of the features. Last, the output of BiLSTM layer goes through one more Fully Connected Layer, and generates the final result..

- **GCN-GRU**

The GCN-GRU model integrates the strengths of GCN and GRU, combining GCN's ability to extract features from graph-structured data with GRU's efficiency in modeling time-series data. Initially, the input data passes through the GCN layer, where the adjacency matrix captures relationships between nodes, creating feature representations. These features are then processed through a Fully Connected Layer for transformation before being fed into the GRU layer. In the GRU layer, the gating mechanism (update and reset gates) controls information flow, capturing both long-term and short-term dependencies in the data. Finally, the output from the GRU layer goes through another Fully Connected Layer to produce the final prediction. .

- **GCN-Transformer**

The GCN-Transformer model combines GCN's feature extraction from graph-structured data with the Transformer's ability to model global temporal dependencies. First, GCN extracts node features using graph convolution and the adjacency matrix, capturing both local and global relationships. These features are then transformed through a fully connected layer before being passed to the Transformer. The Transformer's self-attention mechanism models long-range temporal dependencies by calculating attention weights across time steps, enabling efficient handling of long sequences. Multi-head self-attention and feedforward networks further enhance feature extraction. The final output is generated through a fully connected layer, making the model well-suited for tasks like time-series forecasting and traffic flow prediction.

7. 6. 4 The comparison analysis of results in 4 improved GCN time series models

Table 13 The comparison results of performance in 4 improved GCN time series models.

Models	The results in training sets				The results in test sets			
	RMSE	R^2	MSE	MAE	RMSE	R^2	MSE	MAE
Transformer	0.0055	0.9876	0.0055	0.0491	0.4638	0.8525	0.2151	0.3447
GCN-Transformer	0.1598	0.9435	0.0255	0.1042	0.4304	0.8729	0.1852	0.3178
BiLSTM	0.0044	0.9901	0.0044	0.0337	0.6685	0.6935	0.4469	0.4985
GCN-BiLSTM	0.2307	0.8821	0.0532	0.1558	**0.4167**	**0.8809**	**0.1736**	**0.3037**
GRU	**0.0038**	**0.9914**	**0.0038**	**0.0264**	0.6942	0.6695	0.482	0.5069
GCN-GRU	0.2147	0.8979	0.0461	0.1461	0.4361	0.8696	0.1902	0.3208
LSTM	0.0053	0.988	0.0053	0.0422	0.6389	0.7201	0.4082	0.4743
GCN-LSTM	0.1805	0.9278	0.0326	0.1075	0.4354	0.87	0.1896	0.3189

From **Table 13** and **Fig. 20**, it can be observed that the GCN improvement did not show significant effects on the training set. On the training set, the performance of GCN-BiLSTM is similar to that of BiLSTM, with only minor differences in RMSE (0.2307 vs. 0.0044), MSE (0.1558 vs. 0.0337), and MAE (0.0292 vs. 0.0337), and the R^2 value (0.8821 vs. 0.9901) also shows little difference. This indicates that the introduction of GCN did not significantly improve the model's fit on the training data. While GCN contributes to feature extraction in the graph structure, it did not notably enhance the model's fitting capability on the training set. However, on the test set, the performance of the GCN-improved model shows a clear improvement. GCN-BiLSTM performs the best on the test set with RMSE (0.4167), MSE (0.1736), and MAE (0.3037), and its R^2 value is 0.8806, significantly higher than other models. This shows that GCN greatly improved the model's generalization ability and predictive accuracy. Compared to BiLSTM, GCN-BiLSTM shows significant improvements on the test set, especially in R^2 and error metrics, demonstrating the effectiveness of GCN in enhancing the model's generalization. In addition to GCN-BiLSTM, GCN-Transformer also shows improvements. GCN-Transformer performs better than Transformer on the test set, with RMSE (0.4304), MSE (0.2023), and MAE (0.3178), and its R² value (0.8729) also increased, though it is still lower than GCN-BiLSTM, but higher than Transformer. However, GCN-GRU and GCN-LSTM perform relatively poorly on the test set. Although their performance improved compared to the non-improved GRU and LSTM, they still fall short of GCN-BiLSTM and GCN-Transformer.

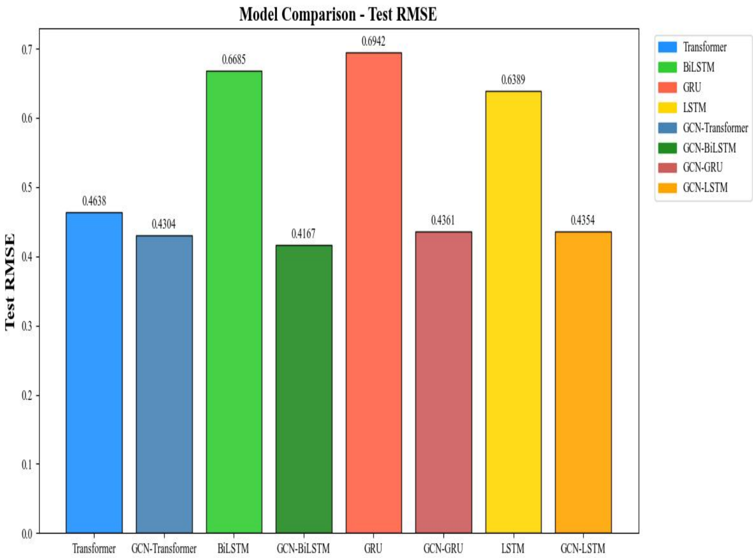


Fig. 20-a. The result of RMSE in GCN series models.

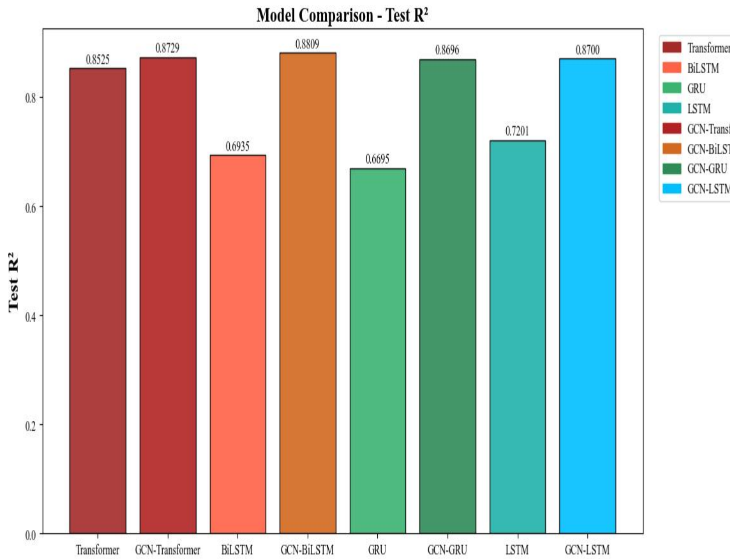


Fig. 20-b. The result of R^2 in GCN series models.

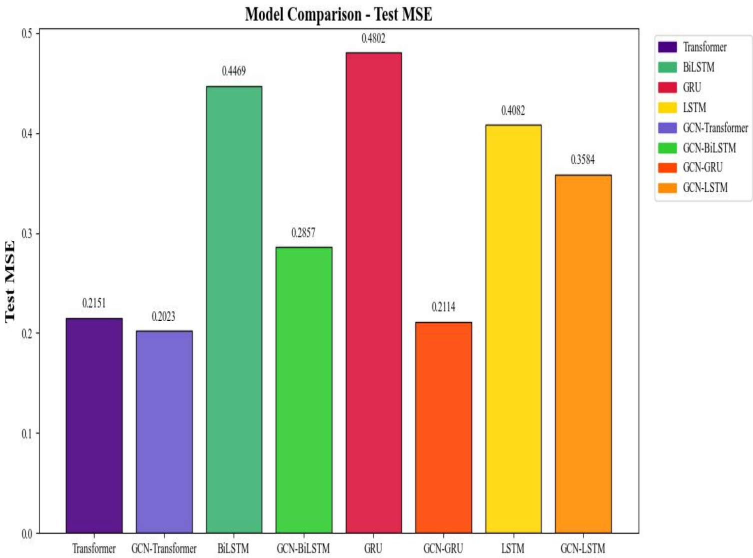


Fig. 20-c. The result of MSE in GCN series models.

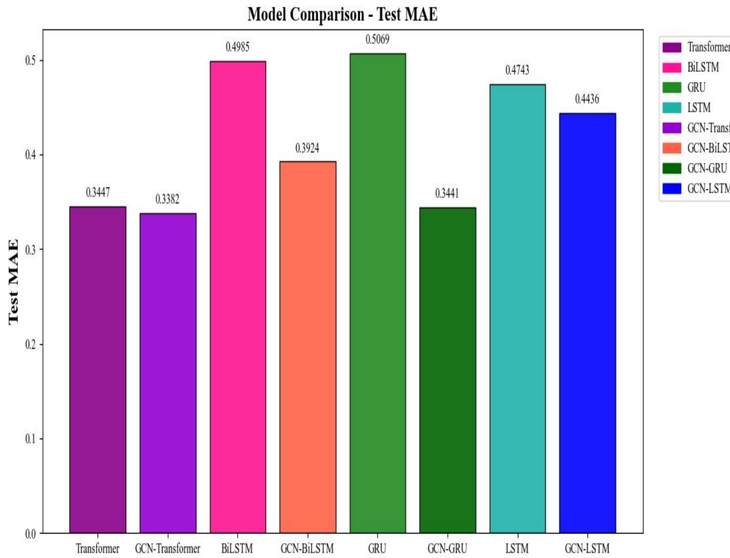


Fig. 20-d. The result of MAE in GCN series models.

7. 7 Comprehensive analysis of comparison results in KAN and GCN time series models

Table 14 Comprehensive analysis of comparison results in improved KAN and GCN series models.

Models	The results in training sets				The results in test sets			
	RMSE	R^2	MSE	MAE	RMSE	R^2	MSE	MAE
Transformer	0.0055	0.9876	0.0055	0.0491	0.4638	0.8525	0.2151	0.3447
KAN-Transformer	**0.0044**	**0.99**	**0.0044**	**0.0386**	0.4497	0.8613	0.2023	0.3382

GCN-Transformer	0.1598	0.9435	0.0255	0.1042	**0.4304**	**0.8729**	**0.1852**	**0.3178**
BiLSTM	0.0044	0.9901	0.0044	0.0337	0.6685	0.6935	0.4469	0.4985
KAN-BiLSTM	**0.0037**	**0.9917**	**0.0037**	**0.0255**	0.5345	0.8041	0.2857	0.3924
GCN-BiLSTM	0.2307	0.8821	0.0532	0.1558	**0.4167**	**0.8809**	**0.1736**	**0.3037**
GRU	**0.0038**	**0.9914**	**0.0038**	**0.0264**	0.6942	0.6695	0.482	0.5069
KAN-GRU	0.0046	0.9899	0.0046	0.041	0.4598	0.855	0.2114	0.3441
GCN-GRU	0.2147	0.8979	0.0461	0.1461	**0.4361**	**0.8696**	**0.1902**	**0.3208**
LSTM	0.0053	0.988	0.0053	0.0422	0.6389	0.7201	0.4082	0.4743
KAN-LSTM	**0.004**	**0.991**	**0.004**	**0.0292**	0.5987	0.7542	0.3584	0.4436
GCN-LSTM	0.1805	0.9278	0.0326	0.1075	**0.4354**	**0.87**	**0.1896**	**0.3189**

From the **Table 14**, it can be seen that both KAN and GCN have improved the model's performance, with GCN showing more significant improvements, especially on the test set. First, the performance on the training set is generally similar, with both KAN and GCN showing some reduction in error. However, the key difference is seen in the test set performance, where the improvement due to GCN is especially noticeable in the GCN-BiLSTM model. On the test set, GCN-BiLSTM shows the best performance, with the lowest RMSE (0.4167), MSE (0.1736), and MAE (0.3037), as well as the highest R^2 value (0.8809), significantly outperforming other models. This indicates that the inclusion of GCN greatly enhanced the model's generalization ability, making GCN-BiLSTM the best-performing model on the test set. KAN-BiLSTM also shows some improvement, with RMSE (0.5345), MSE (0.2857), and MAE (0.3924) lower than BiLSTM, and R^2 (0.8613) increased, but still falls short of GCN-BiLSTM. Among other models, Transformer and GCN-Transformer perform relatively well on the test set, with GCN-Transformer showing the best results compared to Transformer, with lower RMSE (0.4304), MSE (0.2023), and MAE (0.3178), and a higher R^2 (0.8729), indicating that GCN also brings positive effects to Transformer. In contrast, GRU and LSTM models perform poorly on the test set. Although GCN-GRU and GCN-LSTM show some improvement compared to their non-improved versions, their RMSE, MSE, and MAE are still high, and the R^2 values are lower, indicating that the improvement effect for these models is limited.

8. Conclusion

This study delves into the impact of DIF on TFPLP to reveal how DIF drives high-quality development in enterprises. First, the traditional DFE model was used to analyze the linear relationship between DIF and TFPLP. Subsequently, the Kolmogorov–Arnold Neural Network (KAN) and Graph Convolutional Network (GCN) were integrated to improve four classical time-series models (such as Transformer, LSTM, BiLSTM, and GRU), successfully uncovering the nonlinear relationship between DIF and enterprise productivity. This series of analyses provides essential empirical evidence for understanding the deep role that DIF plays in fostering innovation, improving resource allocation efficiency, and enhancing managerial effectiveness within enterprises.

The key contributions of this study are as follows: First, the study systematically explores how DIF drives high-quality development in enterprises through intermediary variables such as innovation, financing constraints, and internal controls, filling a gap in the existing literature. Second, the study innovatively combines Kolmogorov–Arnold Neural Networks (KAN) with traditional deep learning time-series models (such as Transformer, LSTM, BiLSTM, and GRU), significantly improving prediction performance and revealing the nonlinear relationship between DIF and enterprise TFPLP. Lastly, by incorporating Graph Convolutional Networks (GCN), this research effectively captures spatial features from the data and combines them with the temporal dependencies of time-series models. This integration of spatial and temporal features has further enhanced the predictive capability of the model and opened up new perspectives for the application of deep learning in the financial field.

While this research provides an innovative analytical framework and methodology for understanding the relationship between Digital Inclusive Finance and high-quality enterprise development, there are a few avenues for further research. First, with the ongoing development of the DIF field, future studies could expand the sample scope to include enterprises from different regions and industries, thereby validating the generalizability of the findings. Second, further optimization of deep learning models, especially in the integration of multidimensional data and model parameter tuning, could enhance prediction accuracy and generalization capacity. Additionally, with advancements in technology, exploring more flexible deep learning architectures, including adaptive neural networks and reinforcement learning, may better address increasingly complex and real-time data needs.

In summary, this study integrates deep learning techniques with traditional economic empirical models to uncover the deep role of Digital Inclusive Finance in promoting high-quality enterprise development. As digitalization and intelligent processes continue to advance, DIF will provide more opportunities for enterprise innovation and growth. The methods and approaches proposed in this study offer valuable insights for future research and practice in related fields, providing decision-makers and business managers with valuable guidance.

Data availability

All data sources and experimental results are provided within the manuscript or supplementary information files. For any further inquiries, please contact the corresponding author.

Author contributions

D.W. and K.O. conceptualized the study; Z.W., H.K., and K.O. curated the data; H.K., K.O., and Z.W. performed formal analysis; D.W., K.O., and H.K. developed the methodology; K.O. and X.S. managed project administration; D.W., Z.W., and H.K. provided resources; D.W., Z.W., and H.K. developed the software; K.O. and X.S. supervised the work; D.W. wrote the original draft; D.W., Z.W., and X.S. reviewed and edited the manuscript. All authors reviewed and approved the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to K.O or X.S..

Reference

[1] S. Wang, Y. Zhou, J. Guo, and J. Mao, “Did high speed rail accelerate the development of tourism economy? – Empirical analysis from Northeast China,” *Transport Policy,* vol. 143, pp. 25-35, 2023/11/01/, 2023.

[2] L. Dai, J. Zhang, and S. Luo, “Effective R&D capital and total factor productivity: Evidence using spatial panel data models,” *Technological Forecasting and Social Change,* vol. 183, pp. 121886, 2022/10/01/, 2022.

[3] M. G. Tsionas, “Convex non-parametric least squares, causal structures and productivity,” *European Journal of Operational Research,* vol. 303, no. 1, pp. 370-387, 2022/11/16/, 2022.

[4] K. Ma, “Digital inclusive finance and corporate green technology innovation,” *Finance Research Letters,* vol. 55, pp. 104015, 2023/07/01/, 2023.

[5] B. Guo, Y. Feng, and J. Lin, “Digital inclusive finance and digital transformation of enterprises,” *Finance Research Letters,* vol. 57, pp. 104270, 2023/11/01/, 2023.

[6] A. Strimovskaya, and S. Barykin, “A multidimensional approach to the resource allocation problem (RAP) through the prism of industrial information integration (III),” *Journal of Industrial Information Integration,* vol. 34, pp. 100473, 2023/08/01/, 2023.

[7] K. Ouyang, S. Fu, Y. Chen, Q. Cai, A. A. Heidari, and H. Chen, “Escape: an optimization method based on crowd evacuation behaviors,” *Artificial Intelligence Review,* vol. 58, no. 1, pp. 19, 2024.

[8] H. Wang, T. Fu, Y. Du, W. Gao, K. Huang, Z. Liu, P. Chandak, S. Liu, P. Van Katwyk, and A. Deac, “Scientific discovery in the age of artificial intelligence,” *Nature,* vol. 620, no. 7972, pp. 47-60, 2023.

[9] L. Jiao, X. Song, C. You, X. Liu, L. Li, P. Chen, X. Tang, Z. Feng, F. Liu, and Y. Guo, “AI meets physics: a comprehensive survey,” *Artificial Intelligence Review,* vol. 57, no. 9, pp. 256, 2024.

[10] H. Shui, X. Sha, B. Chen, and J. Wu, "Stock weighted average price prediction based on feature engineering and Lightgbm model." pp. 336-340.

[11] F. Liu, S. Guo, Q. Xing, X. Sha, Y. Chen, Y. Jin, Q. Zheng, and C. Yu, "Application of an ANN and LSTM-based Ensemble Model for Stock Market Prediction." pp. 390-395.

[12] E. Egrioglu, and E. Bas, “A new deep neural network for forecasting: Deep dendritic artificial neural network,” *Artificial Intelligence Review,* vol. 57, no. 7, pp. 171, 2024.

[13] A. Vaswani, “Attention is all you need,” *Advances in Neural Information Processing Systems*, 2017.

[14] D. Wei, Z. Wang, M. Qiu, J. Yu, J. Yu, Y. Jin, X. Sha, and K. OuYang, “Multiple Objectives Escaping Bird Search Optimization and Its application in Stock Market Prediction Based on Transformer Model,” 2024.

[15] X. Wang, L. Zhang, Y. Liu, C. Zhao, and K. Wang, “Solving task scheduling problems in cloud manufacturing via attention mechanism and deep reinforcement learning,” *Journal of Manufacturing Systems,* vol. 65, pp. 452-468, 2022/10/01/, 2022.

[16] S. Hochreiter, “Long Short-term Memory,” *Neural Computation MIT-Press*, 1997.

[17] A. Graves, and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM networks." pp. 2047-2052.

[18] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y.

Bengio, “Learning phrase representations using RNN encoder-decoder for statistical machine translation,” *arXiv preprint arXiv:1406.1078*, 2014.

[19] Y. Matsuo, Y. LeCun, M. Sahani, D. Precup, D. Silver, M. Sugiyama, E. Uchibe, and J. Morimoto, “Deep learning, reinforcement learning, and world models,” *Neural Networks,* vol. 152, pp. 267-275, 2022/08/01/, 2022.

[20] A. Kramer, and K. Z. Kramer, “The potential impact of the Covid-19 pandemic on occupational status, work from home, and occupational mobility,” *Journal of Vocational Behavior,* vol. 119, pp. 103442, 2020/06/01/, 2020.

[21] Y. Gao, Z. Hu, W.-A. Chen, M. Liu, and Y. Ruan, “A revolutionary neural network architecture with interpretability and flexibility based on Kolmogorov–Arnold for solar radiation and temperature forecasting,” *Applied Energy,* vol. 378, pp. 124844, 2025/01/15/, 2025.

[22] Y. Wang, J. Sun, J. Bai, C. Anitescu, M. S. Eshaghi, X. Zhuang, T. Rabczuk, and Y. Liu, “Kolmogorov–Arnold-Informed neural network: A physics-informed deep learning framework for solving forward and inverse problems based on Kolmogorov–Arnold Networks,” *Computer Methods in Applied Mechanics and Engineering,* vol. 433, pp. 117518, 2025/01/01/, 2025.

[23] T. N. Kipf, and M. Welling, “Semi-supervised classification with graph convolutional networks,” *arXiv preprint arXiv:1609.02907*, 2016.

[24] K. Ouyang, and S. Fu, “Graph Neural Networks Are Evolutionary Algorithms,” *arXiv preprint arXiv:2412.17629*, 2024.

[25] F. Di Maio, P. C. Rem, K. Baldé, and M. Polder, “Measuring resource efficiency and circular economy: A market value approach,” *Resources, Conservation and Recycling,* vol. 122, pp. 163-171, 2017/07/01/, 2017.

[26] K. A. Whalen, “Three circular business models that extend product value and their contribution to resource efficiency,” *Journal of Cleaner Production,* vol. 226, pp. 1128-1137, 2019/07/20/, 2019.

[27] M. S. Suliman, and H. Farzaneh, “Synthesizing the market clearing mechanism based on the national power grid using hybrid of deep learning and econometric models: Evidence from the Japan Electric Power Exchange (JEPX) market,” *Journal of Cleaner Production,* vol. 411, pp. 137353, 2023/07/20/, 2023.

[28] I. Gabur, D. P. Simioniuc, R. J. Snowdon, and D. Cristea, “Machine learning applied to the search for nonlinear features in breeding populations,” *Frontiers in Artificial Intelligence,* vol. 5, pp. 876578, 2022.

[29] N. Nargis, A. K. M. G. Hussain, S. Asare, Z. Xue, A. Majmundar, P. Bandi, F. Islami, K. R. Yabroff, and A. Jemal, “Economic loss attributable to cigarette smoking in the USA: an economic modelling study,” *The Lancet Public Health,* vol. 7, no. 10, pp. e834-e843, 2022/10/01/, 2022.

[30] N. Shlezinger, Y. C. Eldar, and S. P. Boyd, “Model-based deep learning: On the intersection of deep learning and optimization,” *IEEE Access,* vol. 10, pp. 115384-115398, 2022.

[31] L. Maliar, S. Maliar, and P. Winant, “Deep learning for solving dynamic economic models,” *Journal of Monetary Economics,* vol. 122, pp. 76-101, 2021/09/01/, 2021.

[32] J. Zhang, Z. Li, X. Song, and H. Ning, “Deep Tobit networks: A novel machine learning approach to microeconometrics,” *Neural Networks,* vol. 144, pp. 279-296, 2021/12/01/, 2021.

[33] P. J. Woodfield, Y. M. Ooi, and K. Husted, “Commercialisation patterns of scientific knowledge in traditional low- and medium-tech industries,” *Technological Forecasting and Social Change,* vol. 189, pp. 122349, 2023/04/01/, 2023.

[34] C. Li, Y. Wang, Z. Zhou, Z. Wang, and A. Mardani, “Digital finance and enterprise financing constraints: Structural characteristics and mechanism identification,” *Journal of Business Research,* vol. 165, pp. 114074, 2023/10/01/, 2023.

[35] Q. Qian, X. Chao, and H. Feng, “Internal or external control? How to respond to credit risk contagion in complex enterprises network,” *International Review of Financial Analysis,* vol. 87, pp. 102604, 2023/05/01/, 2023.

[36] D. Schepis, S. Purchase, and B. Butler, “Facilitating open innovation processes through network orchestration mechanisms,” *Industrial Marketing Management,* vol. 93, pp. 270-280, 2021/02/01/, 2021.

[37] A. F. Aysan, G. Gozgor, and Z. Nanaeva, “Technological perspectives of Metaverse for financial service providers,” *Technological Forecasting and Social Change,* vol. 202, pp. 123323, 2024/05/01/, 2024.

[38] R. Ricci, D. Battaglia, and P. Neirotti, “External knowledge search, opportunity recognition and industry 4.0 adoption in SMEs,” *International Journal of Production Economics,* vol. 240, pp. 108234, 2021/10/01/, 2021.

[39] L. Oliveira, A. Fleury, and M. T. Fleury, “Digital power: Value chain upgrading in an age of digitization,” *International Business Review,* vol. 30, no. 6, pp. 101850, 2021/12/01/, 2021.

[40] W. Hadid, and M. Hamdan, “Firm size and cost system sophistication: The role of firm age,” *The British Accounting Review,* vol. 54, no. 2, pp. 101037, 2022/03/01/, 2022.

[41] S. A. Basha, H. Bennasr, and M. Goaied, “Financial literacy, financial development, and leverage of small firms,” *International Review of Financial Analysis,* vol. 86, pp. 102510, 2023/03/01/, 2023.

[42] Z. Chen, and T. W. Archibald, “Maximizing the survival probability in a cash flow inventory problem with a joint service level constraint,” *International Journal of Production Economics,* vol. 270, pp. 109191, 2024/04/01/, 2024.

[43] C.-Y. Lee, “Geographical clustering and firm growth: Differential growth performance among clustered firms,” *Research Policy,* vol. 47, no. 6, pp. 1173-1184, 2018/07/01/, 2018.

[44] O. Farooq, and I. Zarouali, “Financial centers and ownership concentration: When is ownership concentration value relevant? Evidence from an emerging market,” *Research in International Business and Finance,* vol. 38, pp. 236-245, 2016/09/01/, 2016.

[45] D. Manry, and K. Nathan, “Greenmail premia, board composition and management shareholdings,” *Journal of Corporate Finance,* vol. 5, no. 4, pp. 369-382, 1999/12/01/, 1999.

[46] M. Nourallah, P. Öhman, and S. Hamati, “Financial technology and financial capability: Study of the European Union,” *Global Finance Journal,* vol. 62, pp. 101008, 2024/09/01/, 2024.

[47] J. Abdin, A. Sharma, R. Trivedi, and C. Wang, “Financing constraints, intellectual property rights protection and incremental innovation: Evidence from transition economy firms,” *Technological Forecasting and Social Change,* vol. 198, pp. 122982, 2024/01/01/, 2024.

[48] J. Ni, X. Cao, W. Zhou, and J. Li, “Customer concentration and financing constraints,” *Journal of Corporate Finance,* vol. 82, pp. 102432, 2023/10/01/, 2023.

[49] N. J. Cruz-Hinojosa, and J. A. Gutiérrez-de-Mesa, “Literature review of the situation research faces in the application of ITIL in Small and Medium Enterprises,” *Computer Standards &*

Interfaces, vol. 48, pp. 124-138, 2016/11/01/, 2016.

[50] P. He, S. Zhang, and C. He, “Impacts of logistics resource sharing on B2C E-commerce companies and customers,” *Electronic Commerce Research and Applications,* vol. 34, pp. 100820, 2019/03/01/, 2019.

[51] S. Liu, “The influence of financial asset allocation on the innovation and investment of high-tech enterprises,” *Finance Research Letters,* vol. 62, pp. 105138, 2024/04/01/, 2024.

[52] Z. Niu, G. Zhong, G. Yue, L.-N. Wang, H. Yu, X. Ling, and J. Dong, “Recurrent attention unit: A new gated recurrent unit for long-term memory of important parts in sequential data,” *Neurocomputing,* vol. 517, pp. 1-9, 2023/01/14/, 2023.

[53] A. Khan, and B. Lee, “DeepGene Transformer: Transformer for the gene expression-based classification of cancer subtypes,” *Expert Systems with Applications,* vol. 226, pp. 120047, 2023/09/15/, 2023.

[54] B. Kelenyi, V. Domsa, and L. Tamas, “SAM-Net: Self-Attention based Feature Matching with Spatial Transformers and Knowledge Distillation,” *Expert Systems with Applications,* vol. 242, pp. 122804, 2024/05/15/, 2024.

[55] Y. Qiao, K. Hashimoto, A. Eriguchi, H. Wang, D. Wang, Y. Tsuruoka, and K. Taura, “Parallelizing and optimizing neural Encoder–Decoder models without padding on multi-core architecture,” *Future Generation Computer Systems,* vol. 108, pp. 1206-1213, 2020/07/01/, 2020.