Analysis of Driving Factors for Green Economic Development and Innovation Capability Based on Deep Learning

Hongyuan Wang[1], Yujia Wu[2], Jirou Ma[3] and Yihang Lei[1]
[1] School of Management Science and Engineering, Central University of Finance and Economics, Beijing, China
[2] School of Business and Management, Shanghai Dianji University, Shanghai, China
[3] School of Economics and Management, University of Science and Technology Beijing, Beijing, China

Abstract—This study employs deep learning methods to analyze the key driving factors influencing the proportion of green patent applications across 30 Chinese provinces from 2010 to 2023. By constructing a multilayer perceptron (MLP) model and integrating SHAP value analysis, the marginal contributions of factors such as economic foundations, innovation investment, digitalization levels, and environmental governance to green patent applications were quantitatively evaluated. The results indicate that chemical oxygen demand (COD) emissions and sulfur dioxide (SO_2) emission intensity are the primary barriers to green technological innovation, whereas R&D investment, new product development expenditures, and digital transformation provide substantial support for green technologies. Furthermore, optimizing the employment structure of the tertiary sector and increasing the average years of education per capita are shown to play a significant role in driving the growth of green patent applications. Finally, this paper proposes policy recommendations, including strengthening pollution control, optimizing economic structures, accelerating the development of green service industries, and enhancing human capital, to provide theoretical support and practical insights for achieving green economic transformation.

Index Terms—green patents; deep learning; environmental governance; innovation investment; digital transformation

I. INTRODUCTION

1.1 Research Background

Green economic development and green technological innovation have become central issues in global sustainable development. Under the collective efforts of the international community, the concept of green development, with a primary focus on reducing carbon emissions and improving environmental quality, is gradually emerging as a key theme in national policy agendas. In particular, within the framework of the Paris Agreement, countries have committed to achieving carbon neutrality, presenting unprecedented challenges and transformation demands for traditional economic models. As the world's second-largest economy and largest carbon emitter, China is striving to transition from a high-pollution, high-energy-consumption development model to a green, low-carbon, and sustainable economic model by promoting green technological innovation and optimizing its economic structure.

Green patent applications are widely regarded as a critical indicator of green technological innovation. They not only reflect a nation’s or region’s research and development capabilities in green technologies but also indicate the market’s responsiveness to the demand for environmental technologies. In the ongoing transition toward a green economy, identifying the driving factors behind green patent applications is essential for formulating targeted policies and promoting the development of green technologies. However, the factors influencing green patent applications are complex and multifaceted, encompassing aspects such as economic foundations, innovation investment, digitalization levels, and environmental governance. This complexity poses significant challenges to understanding the mechanisms that drive green patent applications.

1.2 Research Questions

Although numerous studies have explored the relationships between green patent applications and factors such as economic development, technological advancement, and environmental governance, most of these studies rely on linear regression models or other traditional statistical methods. These approaches are often insufficient to fully capture the complex, nonlinear relationships among driving factors. Moreover, regional disparities in economic development levels, innovation capabilities, and policy implementation intensities further complicate the interactions among variables. Hence, this study seeks to address the following core questions:

1. Which factors have a significant impact on the proportion of green patent applications?
2. What are the marginal contributions of each influencing factor to green patent applications?
3. In the process of green economic transition, which intervention measures should be prioritized in different domains (e.g., digitalization levels, environmental governance, R&D investment)?

This work was supported by the Jianlong Innovation and Entrepreneurship Fund under Grant No. BKZZJH202506.Corresponding author: Hongyuan Wang (e-mail: why126@email.cufe.edu.cn).

To address these questions, this study adopts deep learning methods, particularly the multilayer perceptron (MLP), to better capture the nonlinear relationships between driving factors and the proportion of green patent applications. Additionally, variable importance analysis based on the SHAP (Shapley Additive Explanations) framework is employed to quantify the impact of each variable on green patent applications.

1.3 Research Significance

The study of the driving factors behind green patent applications holds both academic and practical policy significance. Academically, this research introduces deep learning techniques as a methodological innovation to the field of green economy studies. Deep learning excels in capturing the nonlinear relationships within high-dimensional data, addressing the limitations of traditional statistical models. This allows for a more comprehensive understanding of the complex mechanisms driving green patent applications.

Practically, this study provides scientific evidence for governments and enterprises to formulate green economic development policies by analyzing the marginal contributions of various factors. For instance, the research may identify "corporate digitalization levels" and "environmental pollution control" as critical drivers of green patent applications, which can assist governments in optimizing policy priorities and encouraging enterprises to increase investments in digital technologies and green innovation. Furthermore, analyzing variables with low importance can help optimize resource allocation and avoid policy blind spots.

II. Literature Review

2.1 Green Economy and Innovation Driving Factors

The core objective of a green economy is to achieve a harmonious balance between economic growth and environmental protection. Its driving factors can be categorized into key domains such as economic foundations, innovation investment, digitalization levels, and environmental governance.

Economic Foundations:

Economic foundations provide critical support for green economic development. Studies have shown that the level of economic development determines the capacity for green technology research, development, and application. For instance, regions with high Gross Domestic Product (GDP) can offer financial support for green technological innovation and facilitate the commercialization of green products (Zhang et al., 2021). Additionally, the importance of economic structure in green development has also been highlighted. The growth of the service sector (tertiary industry) not only reduces the proportion of high-pollution industries but also promotes the expansion of knowledge-intensive industries (Vertakova & Plotnikov, 2017).

Innovation Investment:

Technological innovation is regarded as a core driving force for green economic development. Enterprises' investment in research and development (R&D) can significantly enhance green technological innovation capabilities, which is often reflected in the number of green patent applications (Karuppiah et al., 2022). Moreover, patent approvals and the diffusion of technologies play a crucial role in driving green economic growth (Zhang et al., 2021).

Digitalization Levels:

Digital transformation has injected new momentum into green economic development. The digital economy plays a vital role in improving energy efficiency and reducing resource waste (Wang, 2024). Furthermore, digital technologies such as the Internet of Things (IoT), cloud computing, and artificial intelligence (AI) can optimize production processes and support enterprises in achieving green transformation (Xiong, 2023).

Environmental Governance:

Environmental governance is considered a critical pillar of green economic development. Effective pollution control measures, such as reducing chemical oxygen demand (COD) and sulfur dioxide (SO_2) emissions, are closely associated with the increase in green technologies (Makhosheva et al., 2024). Additionally, the improvement of policy and regulatory frameworks plays a pivotal role in advancing the green economy (Heshmati, 2018).

2.2 Limitations of Traditional Regression Analysis Methods

In studies on green economic development, traditional regression analysis methods (e.g., linear regression or fixed-effects models) are widely used to evaluate the relationships between variables. However, these traditional approaches exhibit significant limitations when dealing with complex systems:

Neglect of Nonlinear Relationships

Traditional regression methods assume linear relationships between variables, making it difficult to capture the complex nonlinear characteristics often present in reality. For example, the driving factors of green economic development (e.g., digitalization levels and the number of green patent applications) may exhibit diminishing marginal effects or asymmetric growth trends. In recent years, more flexible nonlinear modeling techniques, such as quantile regression and distributed regression, have been introduced to better capture the complex relationships between variables (Huang et al., 2017). Additionally, distributed regression allows for model extensions to cover the entire data distribution, making it more suitable for capturing the nonlinear characteristics of green economic systems (Klein, 2023).

Limited Capacity for High-Dimensional Data Processing

As the dimensionality of green economy-related data increases, traditional regression methods demonstrate significant limitations in handling high-dimensional data, especially when the number of variables far exceeds the sample size. Modern machine learning methods, such as random forests and XGBoost, have been shown to perform better in high-dimensional data scenarios. These methods not only excel in identifying important variables but also enhance the predictive accuracy of models (Dastile et al., 2020).

Neglect of Dynamic Effects

Traditional models often simplify dynamic effects, for instance, treating time variables as fixed or random effects while overlooking the temporal variation characteristics of variables at different time points. Mixed-effects regression models offer a potential solution by allowing the coexistence of linear and nonlinear dynamic effects, enabling a more accurate representation of how the driving factors of green economic development evolve over time (Lohse et al., 2020).

Impact of Data Quality on Model Performance

Studies have shown that both traditional regression methods and modern machine learning approaches may experience significant declines in predictive performance when faced with low-quality data. Issues such as substantial noise or missing values in the dataset can severely impair the model's predictive capabilities. To address this challenge, researchers recommend integrating data preprocessing techniques, such as regularization methods, to mitigate the impact of noise on model performance (Christodoulou et al., 2019).

2.3 Applications of Deep Learning in Economic Research

In recent years, deep learning has emerged as a powerful nonlinear modeling tool and has gradually become an important method in economic research. Its core advantage lies in its ability to capture complex relationships between variables. The multilayer perceptron (MLP), as a foundational model in deep learning, leverages its multilayer neural network structure to automatically extract features from high-dimensional data and effectively handle nonlinear relationships. This capability has made it increasingly valuable for applications in fields such as green economy and innovation research.

The advantages of deep learning are primarily reflected in the following aspects. First, MLP exhibits significant strengths in capturing nonlinear relationships between variables. By incorporating nonlinear activation functions (e.g., ReLU) in the hidden layers, it can model complex relationships that traditional linear models fail to recognize. This feature is particularly critical for studying the proportion of green patent applications, as the interactions and nonlinear characteristics among influencing factors are prominent. Second, the multilayer perceptron can handle high-dimensional data and automatically select important features through the weight-updating process, effectively ignoring redundant variables. This capability enhances the model’s predictive accuracy and robustness. Furthermore, deep learning demonstrates strong generalization ability. By introducing regularization techniques, such as Dropout, it effectively prevents overfitting, enabling the model to perform better on test data.

In the field of green economy research, the application of deep learning has provided critical support for quantifying and predicting complex relationships. For instance, some studies have utilized deep learning models to predict the dynamic relationship between energy efficiency and carbon emissions, thereby optimizing the formulation of regional carbon reduction policies. These models excel at accurately capturing pollution emission patterns across different time periods, addressing the limitations of traditional statistical models. Similarly, the application of deep learning in innovation research has also gained considerable attention. By analyzing the relationship between innovation investment and patent output, studies have found that deep learning can effectively quantify the nonlinear impact of R&D funding on the number of patents. Moreover, in studies on regional innovation capacity, neural network models that integrate geographic, economic, and technological variables have provided more scientifically grounded insights for government decision-making.

In the fields of digitalization and environmental economics, the application of deep learning has also demonstrated strong adaptability and effectiveness. For example, studies based on artificial neural networks have revealed that the level of corporate digitalization significantly improves resource efficiency and reduces environmental pollution. Through interpretability analysis of deep learning models, researchers have further uncovered the synergistic effects between digital technologies and green innovation. These successful cases highlight that deep learning offers a novel methodological support for economic research. It not only enhances predictive accuracy but also helps researchers uncover hidden economic mechanisms, providing deeper insights into complex interactions within the economy.

2.4 Research Gaps and Contributions

Despite the increasing application of deep learning in economic research, there are still significant gaps in its use within the field of green economy. First, a large body of existing studies predominantly relies on linear or semi-parametric models, failing to fully leverage the capacity of deep learning to capture nonlinear relationships. Green economic development involves multidimensional driving factors, such as economic foundations, innovation investment, digitalization levels, and environmental governance, which exhibit complex interactions. Traditional linear models are inadequate in uncovering these intricate mechanisms. Moreover, existing research has paid limited attention to the exploration of interactions between variables. For instance, the synergistic effects between corporate digitalization levels and R&D investment remain insufficiently validated.

Another significant research gap lies in the insufficient quantification of the marginal contributions of variables. Traditional studies often focus on the impact of a single factor on green patent applications, neglecting the relative importance of multiple variables. This limitation results in a lack of clear priority references for policymaking, thereby reducing the efficiency of policy implementation.

In addition, many studies employ overly simplistic approaches to model the dynamic characteristics of time series data, often relying on fixed-effects or random-effects models. Such methods are inadequate for fully capturing the dynamic evolution of green economic development processes.

This study addresses the aforementioned gaps through several innovations. First, it employs a deep learning model based on a multilayer perceptron (MLP) to effectively capture the complex nonlinear relationships influencing the proportion of green patent applications, offering a novel technical approach to green economy research. Second, by incorporating SHAP-based variable importance analysis, the study quantifies the marginal contributions of different driving factors, providing a clear quantitative basis for policymaking in green technological innovation.

Additionally, the study pays special attention to interactions between variables, revealing the synergistic effects of digitalization levels and innovation investment on green patent applications through model results.

These innovations not only address the limitations of existing literature but also provide valuable insights for academic research and policy practices in the field of green economy development.

Compared with other popular nonlinear models such as random forest and XGBoost, the multilayer perceptron (MLP) model used in this study offers greater flexibility in learning deep hierarchical representations. While tree-based models are effective in handling categorical variables and feature interactions, they often rely on heuristic splitting rules that may miss subtle nonlinear structures. In contrast, MLPs can capture continuous nonlinear patterns through dense connections and activation functions, making them particularly suitable for high-dimensional, interdependent socioeconomic data. Furthermore, the integration of SHAP values with MLP enhances interpretability, bridging the gap between model accuracy and policy relevance—an aspect that is often limited in deep learning applications.

III. Data and Model Selection

3.1 Data Sources

The data used in this study consist of panel data from 31 provincial-level administrative regions in China, spanning the period from 2010 to 2023. These data cover multiple dimensions, including economic foundations, innovation investment, digitalization levels, and environmental governance, providing a comprehensive sample for analyzing the driving factors of green economic development.

The relatively long time span of the dataset allows the study to capture the dynamic changes in the proportion of green patent applications across regions, as well as to uncover the regional disparities in green innovation capabilities.

In this study, the proportion of green patent applications is selected as the core dependent variable, calculated as the ratio of green patent applications to total patent applications. This metric directly reflects the share of green technological innovation within overall technological innovation, serving as a key indicator for measuring green economic development and innovation capability.

To comprehensively analyze the driving factors of the green patent application proportion, the study focuses on four primary dimensions of independent variables: economic foundations, innovation investment, digitalization levels, and environmental factors.

Variables representing economic foundations include GDP, the employment share of the tertiary industry, average years of education per capita, and the intensity of educational expenditure. Together, these variables reflect regional differences in economic strength and human capital foundations.

Innovation investment is measured through variables such as the full-time equivalent of R&D personnel in large industrial enterprises, the proportion of new product development expenditure to GDP, the number of innovative enterprises per 100 people, and the ratio of patent grants to the total population. These variables capture the level of resource input and output capacity in innovation activities within each region.

Digitalization levels, as an emerging factor driving green economic development in recent years, are represented by indicators such as corporate digitalization levels, the digital economy index, and the density of robot installations. These metrics reflect regional differences in technology adoption and digital transformation.

Environmental factors include the proportion of environmental protection expenditure in fiscal spending, forest coverage rate, the ratio of chemical oxygen demand (COD) emissions to GDP, and the ratio of sulfur dioxide (SO_2) emissions to GDP. These variables emphasize the government's environmental governance efforts and the intensity of pollutant emissions.

The comprehensive selection of the dependent variable and independent variables provides a solid data foundation for an in-depth analysis of the driving mechanisms behind the proportion of green patent applications.

The data were primarily obtained from the National Bureau of Statistics of China, including the *China Statistical Yearbook* and *China Statistical Yearbook on Science and Technology* (2010–2023). All data were manually collected from official websites and regional statistical bulletins to ensure accuracy and consistency.

3.2 Data Preprocessing

Before initiating the analysis, rigorous data preprocessing was conducted to ensure the scientific validity and accuracy of the model. Given the significant differences in the scales and value ranges of the variables, all independent variables were standardized using Z-score normalization to prevent imbalanced influences on model weights. The standardization formula is:

$$Z = \frac{x - \mu}{\sigma} \tag{1}$$

where x represents the original value, μ is the mean of the variable, and σ is the standard deviation of the variable. Through this process, all variables were transformed into a standard normal distribution with a mean of 0 and a standard deviation of 1. This not only improved the convergence speed of the model but also minimized biases caused by large value ranges among the variables.

After standardization, the dataset was randomly divided into a training set, validation set, and test set, with proportions of 70%, 15%, and 15%, respectively. The training set was used for parameter optimization in the deep learning model by adjusting the weights of the neural network based on the input-output relationships in the samples. The validation set was employed to tune the model's hyperparameters, such as learning rate and the number of network layers, ensuring the optimal structure and training performance of the model. Finally, the test set was reserved to evaluate the model's generalization performance, primarily assessing its predictive capability on unseen data.

The independent partitioning of the training, validation, and test datasets ensured the reliability of the model and provided a robust framework for assessing its performance. In addition, during the data preprocessing stage, the integrity of the variables was thoroughly examined to identify and address any potential missing or outlier values. An initial assessment indicated that the data quality across all samples was high, with all variables demonstrating logical values and no apparent outliers or missing data. This ensured a strong data foundation for subsequent model training and analysis.

3.3 Deep Learning Model

To analyze the driving factors influencing the proportion of green patent applications, this study constructed a deep learning regression model based on a multilayer perceptron (MLP). The model architecture consists of an input layer, multiple hidden layers, and an output layer, leveraging nonlinear activation functions and optimization techniques to achieve precise data fitting.

The input layer is designed to include all independent variables, which are fed into the model after Z-score standardization to ensure a balanced weight distribution among variables with different scales. This standardized input allows the model to effectively process and integrate information from diverse factors while maintaining consistency across variables.

The design of the hidden layers forms the core of the model. This study adopts a three-layer hidden structure with 128, 64, and 32 neurons in the first, second, and third layers, respectively. Each hidden layer uses the ReLU (Rectified Linear Unit) activation function, mathematically defined as:

$$\mathrm{ReLU}(x) = \max(0, x) \tag{2}$$

The nonlinear characteristics of the ReLU activation function enhance the model's ability to learn complex relationships while effectively avoiding the vanishing gradient problem.

To prevent overfitting, Dropout regularization is applied after each hidden layer, randomly dropping a portion of the neurons' activation values during training. The dropout rate is set to 30%, which reduces the network's dependency on specific neurons and improves its generalization ability. The output layer of the model consists of a single linear neuron responsible for generating predictions of the proportion of green patent applications. This layer maps the complex features learned by the hidden layers into a continuous regression output.

In terms of optimization, the model uses the Adam (Adaptive Moment Estimation) optimizer, which combines the advantages of momentum and adaptive learning rates to accelerate convergence during training. The learning rate is set to **0.0005**, ensuring that the model updates its weights stably and avoids getting stuck in local optima.

The loss function selected is Mean Squared Error (MSE), defined as:

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \tag{3}$$

where y_i represents the true value, $\hat{y}_i$ is the predicted value, and N is the number of samples.

The use of MSE provides a clear and interpretable metric for quantifying the deviation between the predicted and true values. It also serves as the objective for the optimization process, guiding the model to minimize prediction errors and improve accuracy.

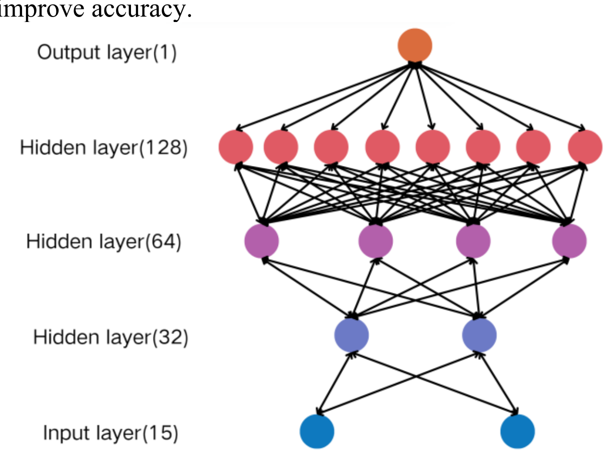

Fig. 1. Framework of the Deep Learning Model: A visualization of the neural network structure from the input layer to the output layer, including input variables, the number of neurons in the hidden layers, activation function types, and the prediction target of the output layer.

To improve reproducibility and clarity, we provide a detailed summary of the deep learning model configuration based on our implementation in MATLAB. The model consists of a feature input layer followed by three fully connected hidden layers containing 128, 64, and 32 neurons, respectively. Each hidden layer uses the ReLU activation function and is followed by a dropout layer with a dropout rate of 0.3 to reduce overfitting.

The output layer includes a single neuron with a linear activation function, optimized using a regression loss. The training process adopts the Adam optimizer with an initial learning rate of 0.0005 and a mini-batch size of 32. The model is trained for a maximum of 200 epochs. Early stopping is implemented through validation monitoring on a separate 15% validation set.

These hyperparameters (e.g., number of neurons, dropout rate, learning rate) were selected based on empirical tuning and best validation performance during experimentation. The final configuration balances model complexity and generalization performance, ensuring robustness and interpretability.

3.4 Model Evaluation Metrics

To comprehensively evaluate the performance of the deep learning model, this study employs Mean Squared Error (MSE) and the Coefficient of Determination (R^2) as the primary evaluation metrics.

In the test set, a lower MSE value indicates that the model is able to effectively capture the fluctuation characteristics of the proportion of green patent applications.

The coefficient of determination (R^2) is used to measure the model's ability to explain the variability in the data, with values ranging from 0 to 1. The formula for R^2 is:

$$R^2 = 1 - \frac{\mathrm{SS_{res}}}{\mathrm{SS_{tot}}} \quad (4)$$

Where:

$\mathrm{SS_{res}} = \sum_{i=1}^{N}(y_i - \hat{y}_i)^2$: Residual Sum of Squares, representing the unexplained variance.

$\mathrm{SS_{tot}} = \sum_{i=1}^{N}(y_i - \bar{y})^2$: Total Sum of Squares.

When the R^2 value approaches 1, it indicates that the model has a high degree of fit to the data. Conversely, when R^2 approaches 0, it suggests that the model has limited predictive power.

In this study, by calculating MSE and R^2 values on the test set, the model's predictive accuracy and generalization performance were validated, demonstrating its robustness and applicability to capturing the driving factors behind green patent applications.

3.5 Variable Importance Analysis

To further quantify the marginal contribution of each variable to the proportion of green patent applications, this study employs a SHAP (Shapley Additive Explanations)-based analysis method. SHAP values are derived from cooperative game theory and are used to calculate the marginal contribution of each variable to the model's output, revealing the importance of each variable in the model's predictions.

The specific steps of the variable importance analysis are as follows:

1. Replace the values of a specific variable with its mean while keeping all other variables unchanged.
2. Perform a new model prediction based on the modified dataset and record the change in the prediction.
3. Compare the original prediction with the new prediction to calculate the marginal contribution of the variable.

The specific formula for calculating variable importance is:

$$\text{Importance}_j = \frac{1}{N}\sum_{i=1}^{N}\left|\hat{y}_i^{\text{original}} - \hat{y}_i^{\text{modified}}\right| \quad (5)$$

where:

$\hat{y}_i^{\text{original}}$ is the original predicted value with all variables intact.

$\hat{y}_i^{\text{modified}}$ is the predicted value after replacing the j variable with its mean.

N is the total number of samples.

To facilitate comparisons, the importance scores of all variables are normalized to a range between 0 and 1 using the formula:

$$\text{Normalized Importance}_j = \frac{\text{Importance}_j}{\sum_{k=1}^{M}\ \text{Importance}_k} \quad (6)$$

where M is the total number of variables.

The higher the normalized importance score, the greater the variable's contribution to the prediction of the proportion of green patent applications.

Using this analysis method, the study successfully quantified the marginal contributions of each independent variable and plotted a ranking of variable importance. This provides robust scientific evidence for promoting green economic development and enhancing innovation capacity.

IV. Empirical Results

4.1 Model Performance Evaluation

This study utilized a multilayer perceptron (MLP) deep learning model to analyze and predict the driving factors of the proportion of green patent applications. The model demonstrated satisfactory performance. On the test set, the evaluation results showed a Mean Squared Error (MSE) of 0.2439 and a Coefficient of Determination (R^2) of 0.7529. These results indicate that the model effectively captures the complex nonlinear relationships between the proportion of green patent applications and its multidimensional driving factors. Furthermore, the high predictive accuracy of the model highlights its robustness and reliability in quantifying and interpreting the key influences on green patent applications. What's more, Over the course of 1800 iterations, the model gradually stabilized, demonstrating good convergence.

The R^2 value represents the proportion of the variance in the target variable that the model can explain. In fields such as social sciences and economics, where data often contain high noise levels and complex interactions, an R^2 value exceeding 0.75 on the test set is considered relatively high. This indicates that the model is not only capable of capturing the effects of key driving factors in green economic development but also demonstrates strong generalization ability.

The consistency of the model's performance on the validation set and test set further confirms that the model does not suffer from overfitting, maintaining its predictive power across unseen data. This highlights the reliability and robustness of the deep learning model in analyzing and predicting the dynamics of green patent applications.

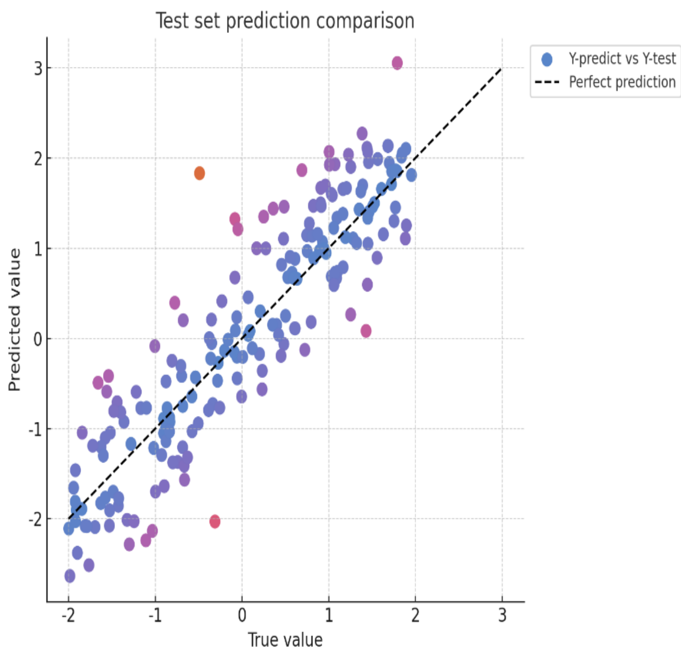

Fig. 2. Predicted vs. Actual Values on Test Set: Model Performance Validation

From the scatter plot results (see the figure above), a clear positive correlation is observed between the model's predicted values and the true values. The majority of the points are distributed near the y = x line, indicating that the model can accurately predict the proportion of green patent applications.

While some points deviate from the trend line, possibly reflecting specific regional factors or noise in the data, the overall fit is highly satisfactory. This provides a robust foundation for the subsequent variable importance analysis, ensuring the reliability of insights derived from the model's predictions.

4.2 Variable Importance Analysis

To further investigate the driving mechanisms behind the proportion of green patent applications, this study employed a SHAP-based analysis using the mean replacement method to quantify the importance of each variable. This method approximates SHAP values by replacing the selected variable with its mean while keeping all other variables constant. Although it is a simplified interpretation of SHAP, it effectively reveals the marginal impact of each variable on the model’s output. The results reveal significant differences in the impact of variables across different dimensions. The specific importance scores are as follows:

Rank	Variable Name	Importance Score
1	Chemical oxygen demand (COD) emissions/GDP	0.1337
2	SO_2 emissions/GDP	0.1259
3	Employment share of the tertiary industry	0.1246
4	New product development expenditure/GDP	0.0852
5	Average years of education per capita	0.0792
6	Number of innovative enterprises per 100 people	0.0654
7	Forest coverage rate	0.0601
8	Digital economy index	0.0583
9	Corporate digitalization level	0.0509
10	Number of granted patents/population	0.0473
11	Intensity of educational expenditure	0.0454
12	GDP	0.0423
13	Robot installation density	0.0342
14	Full-time equivalent R&D personnel in large industrial enterprises	0.0303
15	Environmental protection expenditure/general fiscal expenditure	0.0173

Fig. 3. Feature Importance Derived from SHAP Approximation: Ranking of Driving Factors for Green Patent Applications

4.3 In-Depth Analysis of Key Variables

Based on the SHAP value-based variable importance analysis, this study comprehensively evaluated the key factors influencing the proportion of green patent applications. The results highlight that environmental governance and economic foundations are the two most critical dimensions driving green technological innovation.

Specifically:

Environmental governance: Among the environmental factors, COD emissions/GDP and SO_2 emissions/GDP received the highest importance scores. This underscores the critical role of pollution reduction in stimulating green technological innovation.

Economic foundations: Within this dimension, the employment share of the tertiary industry and average years of education per capita emerged as significant contributors, indicating the importance of industrial structure optimization and human capital enhancement in fostering green innovation.

The following sections will discuss the mechanisms through which these variables influence green patent applications and propose corresponding policy recommendations to further promote green economic development and innovation.

1. The Key Driving Role of Environmental Governance

Environmental governance serves as a direct driver of green technological innovation, particularly in high-pollution industries. The results reveal that **COD emissions/GDP** and **SO_2 emissions/GDP** have the highest SHAP values among all variables, at 0.1337 and 0.1259, respectively. This indicates that the intensity of pollutant emissions has a profound impact on green technological innovation.

These findings suggest that stringent pollution control and reduction efforts not only address environmental concerns but also create strong incentives for the development and adoption of green technologies, as industries strive to meet regulatory requirements and improve their environmental performance.

COD Emissions/GDP is a key indicator of water pollution severity, with higher values often associated with significant industrial pollution. Such pollution pressures local governments and enterprises to collaboratively seek technological solutions to reduce emission levels. For example, the development of water treatment technologies not only alleviates environmental stress but also directly drives the growth of green patent applications.

In regions with high industrial concentration, the carrying capacity of water resources often becomes a critical constraint on economic development. As a result, the demand for COD reduction has, in practice, become a significant driver of green technology research and development, encouraging innovation to address these pressing environmental challenges.

Similarly, the high SHAP value of SO_2 emissions/GDP reflects the indirect role of industrial pollutants in driving green technological innovation. Sulfur dioxide, a major byproduct of industrial fuel combustion and power generation, not only degrades air quality but also poses significant threats to human health and ecosystems.

In recent years, governments have implemented measures such as carbon trading policies and clean energy subsidies to gradually reduce reliance on traditional energy sources, thereby fostering the development of green energy and low-emission technologies. However, disparities in policy enforcement across regions remain a challenge. In high-emission areas, the lack of motivation to invest in green technologies continues to hinder progress, suggesting the need for more targeted and equitable policy interventions.

2. The Indirect Support Role of Economic Foundations

Economic foundations provide stable resources and favorable conditions for promoting green technological innovation, with their marginal contribution to the proportion of green patent applications ranking just behind environmental governance. Among these factors, the employment share of the tertiary industry (SHAP value: 0.1246) and average years of education per capita (SHAP value: 0.0792) stand out as key variables. These metrics respectively reflect the roles of economic structure optimization and human capital accumulation in supporting green economic development.

The significant impact of the employment share of the tertiary industry highlights the critical role of economic structural transformation from industrial to service sectors in driving green technological innovation. The service industry, particularly knowledge-intensive sectors such as green finance, environmental consulting, and technology services, not only consumes fewer resources but also actively promotes the development and application of green technologies.

As the scale of the service industry expands, innovation resources become more concentrated, significantly enhancing innovation efficiency. In major urban areas, the growth of the tertiary sector provides a stable market demand for green technologies, further facilitating their commercialization and large-scale adoption.

At the same time, the importance of average years of education per capita reflects the foundational role of human capital in green technological innovation. Higher education levels not only enhance the workforce's capacity to absorb advanced technologies but also foster a greater pool of talent with innovative mindsets.

In regional economic development, areas with higher education levels tend to exhibit stronger green innovation capabilities. This trend is particularly evident in provinces with a higher proportion of green patent applications, where a well-educated workforce serves as a key driver of green technological advancement.

V. **Policy Recommendations**

5.1Strengthen Pollution Governance and Drive Green Technological Innovation Through Incentive Mechanisms

Pollution governance is a critical driver of green technological innovation. SHAP value analysis in this study reveals that COD emissions/GDP (0.1337) and SO_2 emissions/GDP (0.1259) have significant impacts on green patent applications. The intensity of pollutant emissions strongly influences green innovation, emphasizing the need for effective governance.

The core of pollution control lies in applying external policy tools to exert pressure on enterprises, compelling them to improve technologies and processes. Simultaneously, economic and policy incentives should be provided to stimulate green innovation. This transmission mechanism can be summarized as the "Pressure-Incentive-Innovation" model.

(1).Pressure Pathway in Policy Transmission Mechanisms

The pressure pathway directly motivates enterprises to improve their pollution control behaviors by increasing the cost of emissions and enforcing stricter environmental regulations. Studies have shown that pollution charging systems and carbon trading policies are effective economic instruments that significantly enhance enterprises' environmental investments and their motivation for green technology research and development (Ellerman et al., 2007). Under a carbon trading mechanism, enterprises are compelled to develop more efficient technologies to reduce emission costs. This external pressure fosters endogenous innovation incentives. For instance, the study by Ma & Chang (2023) demonstrates that carbon trading markets significantly promote green patent applications across 75 developing countries, particularly in high-emission industries and regions. Such mechanisms highlight the role of regulatory and economic pressures in driving enterprises to prioritize green technological innovation, ultimately contributing to sustainable economic transformation.

In addition, the strict enforcement of environmental regulations is another critical pathway for applying pressure. For instance, local governments can effectively curb excessive emissions by implementing real-time emission monitoring and conducting regular environmental assessments.

A notable example is Jiangsu Province, where an online monitoring platform has been employed to strengthen real-time management of COD emissions, driving the rapid development of water treatment technologies (Wang et al., 2023). This demonstrates that environmental regulation not only reduces pollutant emissions but also creates new opportunities for green technological innovation.

By ensuring rigorous oversight, environmental governance can simultaneously promote compliance and stimulate technological progress, reinforcing the role of regulation as both a constraint and a catalyst for innovation.

(2). Incentive Pathway in Policy Transmission Mechanisms

While applying pressure, policies also need to guide enterprises towards green technological innovation through diverse incentive mechanisms. This incentive pathway primarily involves financial subsidies, tax incentives, and green finance support.

Jiang et al. (2022) found that green finance can significantly enhance enterprise investment in green technology R&D by alleviating financing constraints. Specifically, green credit and green bonds provide low-cost financing while imposing environmental requirements on the use of funds, thereby promoting green innovation. Financial subsidies, as a traditional incentive mechanism, continue to play an important role in green technological innovation. For example, the central government in China has established multiple green technology special funds in recent years to support the development of low-emission technologies in high-pollution industries. Research indicates that these specialized funds play a crucial role in technological breakthroughs and the commercialization of innovations, with a particularly notable impact on small and medium-sized enterprises (SMEs) (Yu et al., 2021).

These policy incentives work together to reduce financial barriers and create favorable conditions for enterprises to invest in green technologies, accelerating the transition to a more sustainable and innovative economy.

(3). Innovation Pathway in Policy Transmission Mechanisms

The combined effects of policy pressure and incentives stimulate innovation demand within enterprises, driving the rapid development of green technologies. For instance, the integration of carbon trading policies and green finance can create a virtuous innovation ecosystem. Zhao & Xin (2021) suggest that green finance support can alleviate financing constraints, improve enterprises' ability to cope with carbon trading costs, and simultaneously promote the diffusion of green technologies. This mechanism accelerates the market application of green technologies by lowering the costs of technological development and market entry barriers. Additionally, the innovation pathway is reflected in cross-industry collaboration and international technology transfer. For example, large enterprises in developed regions can assist underdeveloped areas by transferring technologies, enabling low-cost pollution control and the introduction of green technologies. Wang & Zhao (2020) indicate that the establishment of regional green technology platforms not only enhances overall pollution control efficiency but also provides innovation resources and market support to small and medium-sized enterprises (SMEs).

These innovation-driven pathways, facilitated by both domestic and international collaboration, ensure that green technologies can be rapidly developed, deployed, and scaled, creating an environment conducive to sustainable technological progress and innovation.

(4). Practical Case Analysis and Optimization Suggestions

Globally, the European Union Emissions Trading System (EU ETS) is one of the successful cases. According to Ellerman et al. (2007), the EU ETS has effectively driven large-scale technological innovation in the industrial sector through strict emission quota allocation and dynamic price regulation. In contrast, China's carbon trading market is still in its early stages, with significant regional imbalances in its development (Zhao & Xin, 2021).

Therefore, optimizing the carbon trading mechanism should be a key focus of policy adjustments.

Specific Recommendations:

1. **Gradually increase emission standards** and reduce free emission allowances to encourage enterprises to actively engage in market trading. This would incentivize companies to adopt greener technologies and reduce their carbon footprints.
2. **Establish a dedicated green technology fund** to support the research and development of low-carbon technologies and their deployment in high-pollution industries. This targeted financial support can help drive innovation and accelerate the adoption of sustainable practices.
3. **Improve the carbon trading data monitoring system** to ensure market transparency, which would enhance the efficiency of decision-making for enterprises. Clear and accessible market information is essential for businesses to make informed investments in green technologies.

5.2 Optimizing Economic Structure to Promote Green Services and Human Capital Development

Optimizing the economic structure is an important means to achieve sustainable development of the green economy. The development of the green service industry and the accumulation of human capital complement each other, collectively forming the key support for green technological innovation. From the SHAP value analysis in this study, it can be seen that the employment share of the tertiary industry and average years of education per capita have significant marginal contributions to the proportion of green patent applications, indicating that economic structure optimization not only directly improves resource allocation efficiency but also indirectly stimulates the demand for technological innovation.

1. Promoting the Specialization of Green Services and Regional Collaboration

The expansion of the green service industry is an important path for achieving economic transformation. As a low-resource consumption and high-knowledge intensity sector, it can significantly enhance the sustainability of the economic system. Research shows that industries such as green finance, environmental consulting, and ecological technology services play a key role in providing financial support and technological solutions (Ma, 2022). Particularly in developing countries, the green service industry has significantly promoted green technological innovation by directing capital flows into the low-carbon economy sector. To promote the scaling and specialization of the green service industry, policies need to strike a balance between market mechanisms and fiscal support. First, the relevant policy framework for green finance should be optimized. For instance, expanding the green bond market and raising bond approval standards can guide funds toward efficient green technology projects (Tufail et al., 2024). Second, strengthening regional collaboration can effectively improve the resource allocation efficiency of the green service industry. For example, while the eastern part of China has abundant green finance resources, service coverage in the central and western regions is relatively low. Therefore, it is recommended to establish cross-regional green service centers to facilitate the free flow of technology and capital between regions, thereby narrowing the gap in green economic development across regions.

Practical examples show that Jiangsu Province, by creating a regional green technology evaluation platform, has promoted the development of the green service industry and effectively facilitated the local economy's green transformation (Song et al., 2022). This experience can serve as a model for nationwide promotion, providing policy support and technological guidance for the green service industry in underdeveloped regions.

2. Strengthening Human Capital Development and Cultivating Green Economy Innovation Talent

The accumulation of human capital is a core driving force behind green technological innovation. The improvement in education levels not only significantly enhances the adaptability of the labor market but also provides high-quality intellectual support for the development of the green service industry. Ma (2022) suggests that the optimal allocation of educational resources and the increase in average years of education play an important role in promoting green patent applications.

To further enhance the support provided by human capital to the green economy, policies should focus on the following areas:

First, strengthen the integration of green economy-related courses into the education system. For example, Germany's dual education system offers a model where schools and enterprises jointly train professionals, providing a large pool of skilled talent for green technological innovation. China can draw on this experience by incorporating courses in environmental economics and sustainable energy management into higher education curricula while accelerating the reform of vocational education institutions.

Second, enhance collaboration between universities and enterprises, by establishing joint laboratories and industry-academia-research cooperation projects to facilitate the rapid development and application of green technologies. Song et al. (2022) highlight that this collaborative model not only improves the conversion rate of university research results but also provides enterprises with innovative solutions.

Moreover, the government should provide policy incentives to guide highly skilled green technology talent to flow into underdeveloped regions. For example, under the Belt and Road Initiative, regional educational cooperation and international exchange can bring advanced educational resources and technical training to the central and western regions, thereby enhancing the green innovation capacity of regional economies.

3. Building a Collaborative Policy Transmission Mechanism

The development of the green service industry and human capital construction needs to achieve synergy in policy. Specifically, a "Green Economic Development Center" could be established to integrate green services and educational resources. This center would not only provide market analysis, technical support, and capital matchmaking for the green service industry but also serve as a hub for human capital training, ensuring a seamless connection between the education system and market demand.

This mechanism would effectively enhance the efficiency of policy transmission, creating a virtuous cycle of "service industry expansion — talent supply — technological innovation." By aligning the needs of the market with the capabilities of the workforce, such a collaborative framework would drive the sustainable growth of both green services and green innovation, fostering a more dynamic and innovative green economy.

Reference

Zhang, J., Lu, G., Skitmore, M. *et al.* A critical review of the current research mainstreams and the influencing factors of green total factor productivity. *Environ Sci Pollut Res* **28**, 35392–35405 (2021). https://doi.org/10.1007/s11356-021-14467-4

Vertakova, Y., & Plotnikov, V. (2017). Problems of sustainable development worldwide and public policies for green economy. *Economic Annals-XXI*, 166, 4–10. https://doi.org/10.21003/ea.V166-01

Karuppiah, K., Sankaranarayanan, B., Ali, S., Alarjani, A., & Mohamed, A. (2022). Causality analytics among key factors for green economy practices: Implications for sustainable development goals. *Frontiers in Environmental Science*, 10. https://doi.org/10.3389/fenvs.2022.933657

Guo, F., Yang, C., & Feng, S. (2024). The Impact of Digital Economy on Green Technology Innovation and Its Mechanism: Evidence from 274 Cities in China. *Emerging Markets Finance and Trade*, *60*(9), 1971–1985. https://doi.org/10.1080/1540496X.2023.2297925

Xiong, Z. (2023). Literature Review on Digital Economy Enabling Green Development. *Frontiers in Business Economics and Management*.11(2):192-194. 10.54097/fbem.v11i2.12588

Makhosheva, S., Kandrokova, M., Yandieva, L., & Uyanaeva, K. (2024). Research on the institutional environment of the green economy: foreign and domestic experience.*BIO Web of Conferences*.84. 10.1051/bioconf/20248401019

Heshmati, A. (2018). An Empirical Survey of the Ramifications of a Green Economy. *Socially Responsible Investment eJournal*. https://doi.org/10.1504/IJGE.2018.092359
Huang, Q., Zhang, H., Chen, J., & He, M. (2017). Quantile regression models and their applications: A review. *Journal of Biometrics & Biostatistics*.08(03) 10.4172/2155-6180.1000354
Klein, N. (2023). Distributional regression for data analysis. *Annual Review of Statistics and Its Application.* https://doi.org/10.1146/annurev-statistics-040722-053607
Dastile, X., Çelik, T., & Potsane, M. (2020). Statistical and machine learning models in credit scoring: A systematic literature survey. *Applied Soft Computing.* https://doi.org/10.1016/j.asoc.2020.106263
Lohse, K., Shen, J., & Kozlowski, A. (2020). Modeling longitudinal outcomes: A contrast of two methods. *Journal of Motor Learning and Development.* 10.1123/jmld.2019-0007
Christodoulou, E., Ma, J., Collins, G., et al. (2019). A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *Journal of Clinical Epidemiology.* 10.1016/j.jclinepi.2019.02.004
Ellerman, A.D., and B. Buchner (2007): The European Union Emissions Trading Scheme: Origins, Allocation, and 2005 Results. *Review of Environmental Economics and Policy*, I: 1 (April) http://dx.doi.org/10.1093/reep/rem003
Jiang, S., Liu, X., Liu, Z., Shi, H., & Xu, H. (2022). Does green finance promote enterprises' green technology innovation in China? *Frontiers in Environmental Science*, 10, 981013. https://doi.org/10.3389/fenvs.2022.981013
Ma, J., & Chang, C.-P. (2023). The role of green finance in green innovation: Global perspective from 75 developing countries. *Emerging Markets Finance and Trade*, 59, 3109–3128.https://doi.org/10.1080/1540496X.2023.2210720
Wang, Y., Liu, J., Yang, X., Shi, M., & Ran, R. (2023). The mechanism of green finance's impact on enterprises' sustainable green innovation. *Green Finance*. https://doi.org/10.3934/gf.2023018
Zhao, Y., & Xin, L. (2021). Research on green innovation countermeasures of supporting the circular economy to green finance under big data. *Journal of Enterprise Information Management*, 35(6), 1305–1322. https://doi.org/10.1108/JEIM-01-2021-0039
Yu, C., Wu, X., Zhang, D., Chen, S., & Zhao, J. (2021). Demand for green finance: Resolving financing constraints on green innovation in China. *Energy Policy*, 153, 112255. https://doi.org/10.1016/j.enpol.2021.112255
Ma, W. (2022). Exploring the role of educational human capital and green finance in total-factor energy efficiency in the context of sustainable development. *Sustainability*, 15(1), 429.https://doi.org/10.3390/su15010429
Tufail, M., Song, L., & Khan, Z. (2024). Green finance and green growth nexus: Evaluating the role of globalization and human capital. *Journal of Applied Economics*, 26(1), 1–20. https://doi.org/10.1080/15140326.2024.2309437
Song, J., Zhou, H., Gao, Y., & Guan, Y. (2022). Digital inclusive finance, human capital and inclusive green development—Evidence from China. *Sustainability*, 14(16), 9922. https://doi.org/10.3390/su14169922