Article

Analysis and Prediction of New Energy Vehicle Development in China Based on Regression and Time Series Models

Fangyan Ma [1,*] and Xue Chen [1]

[1] Hainan Vocational University of Science and Technology, Hainan, 571126, China
* Correspondence: Fangyan Ma; Hainan Vocational University of Science and Technology, Hainan, 571126, China

Abstract: The rapid expansion of new energy vehicles (NEVs) in China represents a vital element of the global transition toward sustainable transportation and carbon neutrality. This study develops a comprehensive modeling framework to quantitatively examine the main influencing factors and predict the future trajectory of China's NEV market. Grey relational analysis is first employed to determine that the density of public charging infrastructure and government subsidy policies exhibit the strongest correlation with NEV sales. Subsequently, an ARIMA (0,1,1) time series model is constructed and validated to forecast annual NEV sales in China for the period 2024-2033, revealing a consistent upward trend. To explore the global implications, a Vector Autoregressive (VAR) model is established, indicating a significant yet intricate relationship among the rise of NEVs, the traditional fuel vehicle industry, global oil prices, and associated R&D investments. Furthermore, a population competition model incorporating policy resistance factors is simulated to evaluate the potential impact of international trade barriers. The results show that although external resistance may decelerate growth, it cannot reverse the overall expansion of China's NEV industry. This research offers a solid theoretical foundation and data-driven insights for policymakers and industry stakeholders, emphasizing the necessity of continuous infrastructure investment and strategic policy support.

Keywords: new energy vehicles; ARIMA model; grey relational analysis; multiple linear regression; population competition model; policy impact

Published: 19 November 2025

1. Introduction

The global automotive industry is experiencing a profound and accelerating transformation driven by the dual imperatives of mitigating climate change and ensuring long-term energy security [1]. In this context, the development of new energy vehicles (NEVs) has emerged as a strategic pathway for achieving sustainable mobility and reducing carbon emissions. As the world's largest automobile market and one of the foremost producers of NEVs, China occupies a pivotal position in this transition [2]. Over the past decade, the Chinese government has introduced a comprehensive portfolio of supportive measures, including purchase subsidies, tax incentives, preferential licensing policies, and mandates for the construction of extensive charging infrastructure networks. These efforts have collectively stimulated technological innovation, industrial investment, and consumer adoption, propelling China's NEV sector into a period of rapid and sustained expansion.

According to data from the National Bureau of Statistics, China's NEV sales have increased from only a few thousand units in the early 2010s to several million units per year in recent years, establishing the country as a global leader in electric mobility [3]. This surge reflects not only strong domestic demand but also the continuous advancement of battery technology, improvements in energy efficiency, and the growing competitiveness of Chinese automakers in the international market. Moreover, the NEV industry has generated substantial spillover effects, driving growth in related sectors such as lithium-ion battery manufacturing, renewable energy integration, and smart grid development.

Despite these achievements, several challenges remain unresolved. The rapid evolution of the NEV industry has introduced complex interdependencies among policy design, market behavior, and technological innovation. Understanding the quantitative relationships among these factors is essential for maintaining sustainable growth and avoiding policy or investment inefficiencies [4]. Therefore, a systematic and data-driven approach is required to capture the dynamics of the NEV market and to forecast its future trajectory under different economic and policy scenarios.

To address this research gap, the present study proposes an integrated mathematical modeling framework designed to analyze, predict, and interpret the evolution of China's NEV market. The objectives of this study are fourfold:

(1) to identify and rank the principal factors influencing NEV adoption and diffusion in China;

(2) to develop a robust and reliable predictive model for medium-term market growth;

(3) to assess the global economic and industrial implications of China's NEV expansion, particularly in relation to conventional vehicle manufacturing, oil markets, and technological investment flows; and

(4) to simulate the potential consequences of international policy responses-such as trade barriers or subsidy adjustments-on the sustainable development of China's NEV industry.

The remainder of this paper is organized as follows. Section 2 outlines the methodological framework, including the application of multiple linear regression, grey relational analysis, ARIMA, VAR, and population competition models. Section 3 presents an in-depth analysis of the results, covering factor identification, sales forecasting, global impact assessment, and policy simulation. Finally, Section 4 concludes the study with a summary of key findings, theoretical implications, and recommendations for policy and future research directions [5].

2. Methodology and Model Construction

2.1. Data Preprocessing and Variable Selection

The empirical foundation of this study rests on the use of high-quality, multi-source data to ensure both accuracy and analytical robustness. Annual time-series data spanning the period from 1989 to 2023 were systematically collected from multiple authoritative sources, including the National Bureau of Statistics of China, the China Association of Automobile Manufacturers, and other comprehensive industry reports. This extensive dataset provides a solid empirical basis for modeling the development dynamics of China's new energy vehicle (NEV) sector [6].

To capture the complex and multidimensional nature of NEV development, the variables were categorized into five analytical dimensions, reflecting policy, infrastructure, economic, technological, and social factors. This classification enables a more holistic assessment of the interrelated drivers influencing NEV market expansion. The specific variable design is as follows:

1) **Policy (X_1):** Quantified by the intensity and continuity of national and local subsidy policies, including purchase incentives and tax exemptions, which reflect the degree of governmental intervention and support for NEV promotion.
2) **Infrastructure (X_2):** Represented by the number of publicly available charging piles, a crucial indicator of the accessibility and convenience of NEV use. The expansion of charging networks directly affects consumer confidence and adoption rates.

3) **Economy (X_3):** Comprised of the retail prices of gasoline and electricity, which together characterize the relative operating cost advantages of NEVs compared to traditional internal combustion engine vehicles.
4) **Technology (X_4):** Measured by the average cost of power batteries, serving as a proxy for technological progress and manufacturing efficiency. Declining battery costs are a major driver of market competitiveness and affordability.
5) **Society (X_5):** Indirectly represented by national carbon emission levels and survey-based indices of public environmental awareness. These variables collectively reflect the societal pressure and consumer attitudes that encourage sustainable mobility choices.

The dependent variable, **Y**, represents the overall development level of NEVs, proxied by the annual sales volume. This indicator effectively captures both market demand and production capacity over time, making it suitable for dynamic modeling and forecasting.

A rigorous data preprocessing procedure was conducted prior to model construction to ensure data reliability and consistency. The process included:

1) **Missing Data Imputation:** Missing values were supplemented using linear interpolation and trend estimation techniques to maintain the continuity of time-series data.
2) **Outlier Detection and Correction:** Abnormal observations were identified through statistical diagnostics such as z-scores and residual analysis, and corrected where necessary to prevent model distortion.
3) **Normalization:** All variables were standardized to eliminate dimensional disparities, enabling direct comparison and ensuring stable numerical behavior in subsequent model estimation.

This systematic data preparation ensures that the final dataset is both statistically sound and theoretically representative, thereby laying a reliable foundation for the construction of regression, grey relational, and time-series forecasting models introduced in the following sections.

2.2. Multiple Linear Regression and Grey Relational Analysis Model

To initially capture the relationship between the independent variables and NEV sales, a multiple linear regression model was established:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \varepsilon \tag{1}$$

Where β_0 is the constant term, β_1 to β_5 are the regression coefficients for each factor, and ε is the random error term.

However, given the potential for complex, non-linear relationships and the need to evaluate the relative importance of each factor, grey relational analysis (GRA) was subsequently employed. This method is particularly effective for analyzing systems with limited data and uncertain information. The analysis procedure is as follows:

1.Determine the Reference and Comparison Sequences: The NEV sales volume is set as the reference sequence X_0, while the influencing factors X_1 to X_5 are the comparison sequences.

2.Data Normalization: The raw data is normalized to eliminate scale effects.

3.Calculate the Grey Relational Coefficient: The coefficient between the reference sequence and each comparison sequence at point k is calculated using the formula:

$$\gamma(x_0(k), x_i(k)) = \frac{\Delta_{min} + \rho\Delta_{max}}{\Delta_{ik} + \rho\Delta_{max}} \tag{2}$$

The comprehensive evaluation result B was obtained through the fuzzy transformation B=A*R.

where Δ_{ik} is the absolute difference at point k, Δ_{min} and Δ_{max} max are the global minimum and maximum differences, and ϱis the discrimination coefficient, set to 0.5.

Compute the Grey Relational Grade (GRG): The average of the relational coefficients for each sequence is computed, yielding the GRG γ_{0i}.A higher GRG indicates a stronger influence of that factor on NEV sales (As shown in Table 1).

Table 1. Grey-weighted correlation.

Considerations	X_1	X_2	X_3	X_4	X_5
Relatedness	0.8081	0.8045	0.7599	0.7484	0.6196

Accordingly, we can also find the greatest correlation between public charging piles and subsidy policies.

Figure 1 shows the correlation of variables x1 through x7 over a 35-year period. Where the solid line represents x1 and the dashed lines represent x2 through x7. The chart illustrates the relative changes and trends in these variables over time.

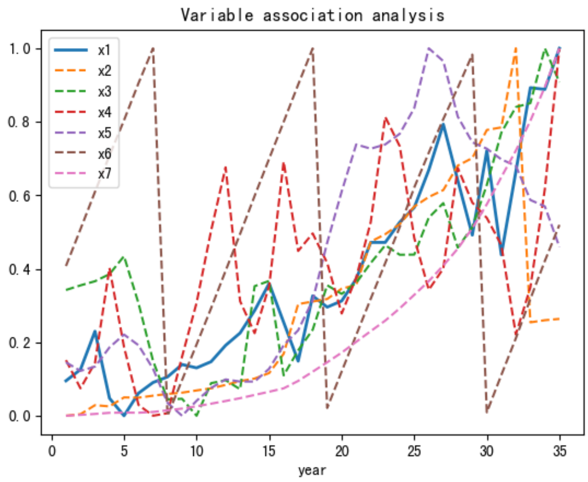

Figure 1. Variable association analysis.

According to the above gray correlation algorithm, it can be found that public charging piles and subsidy policy are the factors with the greatest correlation in the development of new energy electric vehicles in China, which can be judged as the most important influencing factors, so from the data, the government needs to formulate a policy to increase the support of new energy vehicles, In particular, subsidies should be increased to subsidize the purchase of new energy vehicles, as well as more charging piles to make charging more convenient.

2.3. ARIMA Model for NEV Sales Forecasting

To predict the future trajectory of NEV sales in China, an Autoregressive Integrated Moving Average (ARIMA) model was adopted. The ARIMA(p,d,q) model is a cornerstone of time series analysis, where p, d, and q represent the orders of autoregression, integration (differencing), and moving average, respectively. The modeling process involved the following steps:

1.Stationarity Test: The Augmented Dickey-Fuller (ADF) test was applied to the original sales data series. The null hypothesis of non-stationarity could not be rejected (p-value = 0.974), confirming the series was non-stationary.

2.Differencing: First-order differencing was performed to stabilize the series. The ADF test on the differenced series rejected the null hypothesis (p-value = 0.000), confirming stationarity. This established the parameter d=1.

3.Model Identification: The Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots of the stationary series were analyzed to identify potential p and q values.

4.Model Selection and Validation: Several ARIMA models were estimated, and the ARIMA (0,1,1) model was selected as optimal based on the lowest Akaike Information Criterion (AIC = 234.308) and Bayesian Information Criterion (BIC = 238.887). The Ljung-Box Q-test statistics for the residuals were all non-significant ($p > 0.1$), indicating that the residuals were white noise and the model was well-specified. The model's goodness-of-fit, represented by R^2, was 0.843. (As shown in Table 2)

Table 2. ADF Inspection Form.

Variant	Difference stage	T	P	AIC	Threshold Value		
					1%	5%	10%
EV	**0**	**0.238**	**0.974**	172.072	-3.646	-2.954	-2.616
	1	-7.451	0.000	164.206	-3.646	-2.954	-2.616
	2	-4.157	0.001	152.45	-3.77	-3.005	-2.643

Figures 2 and Figure 3 the ACF plot shows trailing, and the PACF plot shows truncation after lag 1, supporting an MA (1) component.

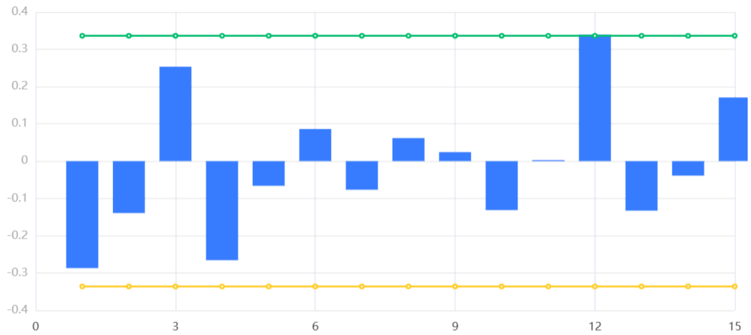

Figure 2. Autocorrelogram of Final Differential Data (ACF).

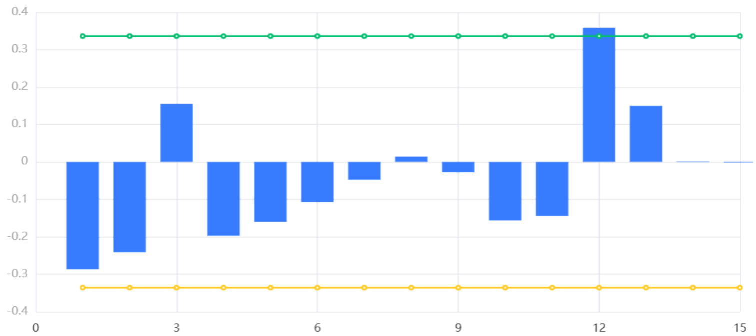

Figure 3. Partial Autocorrelation Chart for Final Differential Data (PACF).

The final ARIMA (0,1,1) model is expressed as:

$$y(t) = 1.648 - 0.412 \cdot \varepsilon(t-1) \quad (3)$$

where y(t) is the differenced series of NEV sales, and $\varepsilon(t-1)$ is the previous error term.

2.4. Vector Autoregressive (VAR) Model for Global Impact Analysis

To analyze the dynamic impact of China's NEV development on the global traditional energy vehicle industry, a Vector Autoregressive (VAR) model was constructed. The model incorporates four key variables: global sales of new energy vehicles (Y_1), global sales of conventional energy vehicles (Y_2), changes in global oil prices (Y_3), and the amount of R&D in conventional energy technologies (Y_4).

A Pearson correlation analysis was first conducted to screen for significant relationships. The results showed a strong positive correlation between NEV sales and traditional energy R&D (r = 0.932, p = 0.068), and a strong negative correlation between conventional vehicle sales and oil prices (r = -0.849).

The general form of the VAR model with lag p is:

$$Y_t = \varphi_0 + \varphi_1 Y_{t-1} + \cdots + \varphi_p Y_{t-p} + BT_t + \quad (4)$$

where Y_t is a vector of the endogenous variables, φ_0 is a constant vector, $\varphi_1 \cdots \varphi_p$ are coefficient matrices, X_t represents exogenous variables, and ε_t is a vector of error terms. After estimation, the following key relationship was derived from the model:

$$y_1 = -3.1011y_2 + 0.0002y_3 + 0.0061y_4 + 0.0173 \quad (5)$$

This suggests that an increase in global conventional vehicle sales is associated with a decrease in NEV sales, while increases in oil prices and traditional R&D spending are associated with an increase in NEV sales, indicating complex market and technological substitution effects.

Figure 4 visually represents the Pearson correlation coefficients between the global variables, with color intensity indicating the strength and direction of the relationship.

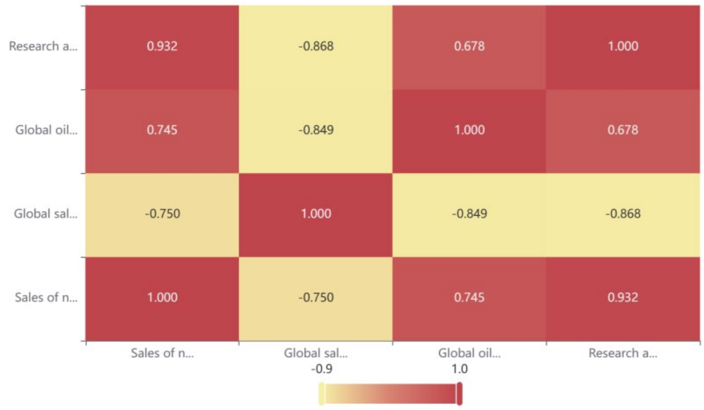

Figure 4. Correlation Coefficient Heat Map of Global Variables.

2.5. Population Competition Model with Policy Resistance

To simulate the impact of international trade policies and resistance on the competition between NEVs and traditional fuel vehicles, a population competition model was developed, inspired by ecological Lotka-Volterra models. The model is defined by the following system of differential equations:

$$\frac{dx}{dt} = r_1 x(1 - \frac{x}{n_1} - \frac{y}{n_2}) \quad (6)$$

$$\frac{dy}{dt} = r_2 y(1 - \frac{x}{n_1} - \Phi\frac{x}{n_2}) \quad (7)$$

Here r_1,r_2 are intrinsic growth rates of the two vehicle populations. Φ is the "boycott strength" coefficient, representing the cost increase or market access restriction for NEVs due to foreign policies (e.g., tariffs, subsidy exclusions). A higher Φ implies stronger resistance.

This model was solved numerically in MATLAB using the delay differential equation solver (dde23) to simulate the evolution of vehicle ownership over time under different scenarios of resistance strength Φ.

3. Results and Discussion

3.1. Key Influencing Factors of NEV Development in China

The results of the grey relational analysis, as summarized in Table 3, provide a clear hierarchy of factors driving NEV adoption in China. The number of public charging piles (GRG=0.8081) and subsidy policies (GRG=0.8045) are identified as the two most critical factors. This underscores the fundamental importance of mitigating "range anxiety" through widespread infrastructure and making NEVs financially accessible to consumers.

The strong showing of social/environmental awareness (GRG=0.7731) indicates a growing consumer consciousness that aligns with national sustainability goals.

Table 3. ARIMA Model Forecast for China's NEV Sales (2024-2033).

Projected value	
Number of steps	**Projected results**
2024	65.14232942
2025	66.79036838
2026	68.43840734
2027	70.0864463
2028	71.73448526
2029	73.38252422
2030	75.03056318
2031	76.67860214
2032	78.32664109
2033	79.97468005

3.2. NEV Sales Forecast for the Next Decade

The ARIMA (0,1,1) model demonstrated excellent fit and forecasting capability. The forecast for annual NEV sales in China from 2024 to 2033 shows a consistent and steady growth, increasing from approximately 65.14 million units in 2024 to nearly 80 million units by 2033 (As shown in Table 3).

Figure 5 shows the original data (blue), the model's fitted values (green), and the forecasted values (yellow) for the next 10 years.

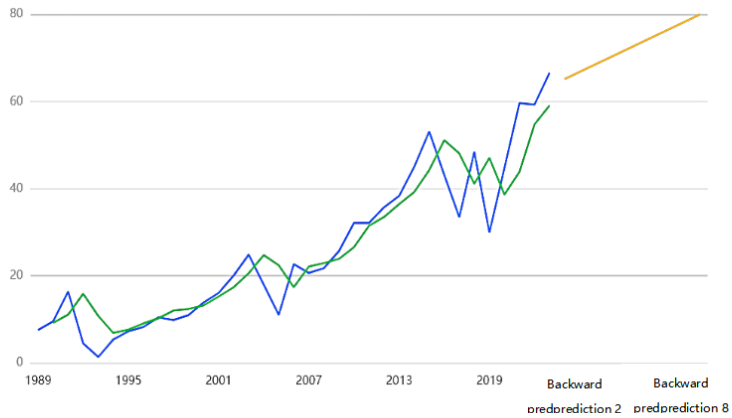

Figure 5. Time Series Chart of NEV Sales.

This prediction suggests that the Chinese NEV market is far from saturation and is expected to maintain its growth momentum, reinforcing the country's leading position in the global electric vehicle landscape.

3.3. Global Impact and Inter-Industry Dynamics

The VAR model results reveal a nuanced relationship between the NEV sector and the traditional automotive ecosystem. The negative coefficient between global NEV sales and conventional vehicle sales (-3.1011) confirms a substitution effect, signaling a gradual market share shift. However, the positive correlation between NEV sales and both oil prices (0.0002) and traditional R&D spending (0.0061) is more complex. It suggests that high oil prices, while making EVs more attractive, also reflect a strong global economy with high overall energy demand. Similarly, increased R&D in traditional technologies might be a competitive response from incumbent automakers, but it occurs within a broader technological landscape that also benefits from spillovers to EV-related technologies.

3.4. Simulation of International Policy Resistance

The population competition model was simulated for different values of the boycott strength coefficient Φ.The results, visualized in a multi-panel graph, clearly demonstrate that as Φ increases, the growth curve of NEV ownership $(y(t))$ becomes flatter, and the time required for NEVs to reach a certain market penetration level lengthens.

Figure 6 show the change in vehicle ownership over time for fuel vehicles (green) and NEVs (blue) under different boycott coefficients.

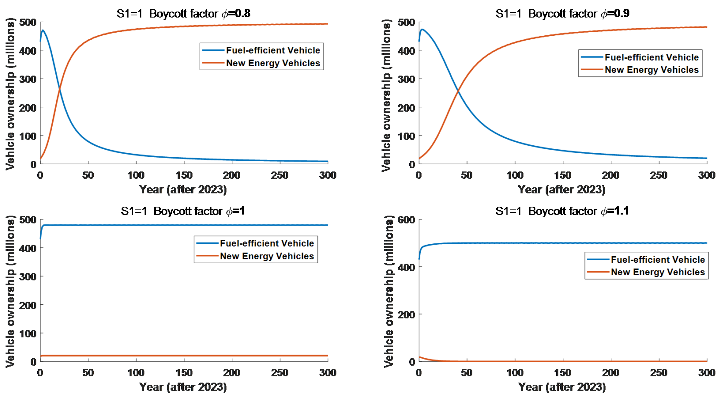

Figure 6. Trends in New Energy Vehicle Development Under Different Resistance Levels.

This simulation leads to a critical insight: while protectionist policies in foreign markets can act as a headwind and slow down the international expansion of China's NEV industry, they are unlikely to halt its progress altogether. The inherent growth drivers-domestic policy support, infrastructure development, and technological advancement-appear robust enough to sustain the industry's development, even in a more hostile international trade environment.

4. Conclusion

This study presents a multi-faceted quantitative analysis of the development of new energy vehicles in China. By integrating grey relational analysis, ARIMA forecasting, VAR modeling, and population competition simulation, we provide a comprehensive and robust examination of the sector's dynamics.

The main conclusions are as follows:

1.Infrastructure and subsidies are paramount. The development of public charging networks and financial incentives are the most powerful levers for promoting NEV adoption in China.

2.Sustained growth is expected. The Chinese NEV market is projected to experience steady growth over the next decade, with annual sales potentially approaching 80 million units by 2033.

3.The global impact is significant but complex. The rise of NEVs is directly challenging the traditional vehicle market, but it remains intertwined with global energy prices and technological trends.

4.External resistance is a manageable risk. International policy barriers can dampen growth but are unlikely to reverse the overall positive trajectory of China's NEV industry, given its strong domestic fundamentals.

Policy Implications:

Based on these findings, we recommend that policymakers in China:

1.Sustain and optimize fiscal support policies while gradually shifting from direct purchase subsidies to support for R&D and infrastructure.

2.Accelerate the nationwide deployment of high-quality, intelligent charging infrastructure to fully eradicate range anxiety.

3.Foster international cooperation and dialogue to mitigate trade frictions and explore collaborative standards and projects.

This study presents a comprehensive, data-driven framework for a project aimed at combating online IWT. The use of a fuzzy comprehensive evaluation model provides an objective methodology for selecting TRAFFIC as the ideal implementing partner, moving beyond subjective choice. The strong Pearson correlation between online and overall IWT validates the project's strategic focus. The linear regression model offers a quantifiable means to predict and track the project's impact, fostering accountability.

References

1. T. C. Mills, "Applied time series analysis: A practical guide to modeling and forecasting," *Academic press*, 2019.
2. T. Yang, X. Zhao, Q. Sun, Y. Zhang, and J. Xie, "Elucidating the anti-inflammatory activity of platycodins in lung inflammation through pulmonary distribution dynamics and grey relational analysis of cytokines," *Journal of Ethnopharmacology*, vol. 323, p. 117706, 2024. doi: 10.1016/j.jep.2024.117706
3. M. Zhou, and W. Dong, "Grey correlation analysis of macro-and micro-scale properties of aeolian sand concrete under the salt freezing effect," In *Structures*, December, 2023, p. 105551.
4. S. Hong, and W. R. Reed, "Metaanalysis and partial correlation coefficients: A matter of weights," *Research Synthesis Methods*, vol. 15, no. 2, pp. 303-312, 2024. doi: 10.1002/jrsm.1697
5. Y. Chen, Y. Peng, and L. Shen, "A Model for Predicting Salaries in Big Data Roles: An Integration of Random Forest and Adaboost-KNN Models," In *Proceedings of the 2024 4th International Conference on Artificial Intelligence, Big Data and Algorithms*, June, 2024, pp. 568-574. doi: 10.1145/3690407.3690504
6. K. Trabelsi, Z. Saif, M. W. Driller, M. V. Vitiello, and H. Jahrami, "Evaluating the reliability of the athlete sleep behavior questionnaire (ASBQ): a meta-analysis of Cronbach's alpha and intraclass correlation coefficient," *BMC Sports Science, Medicine and Rehabilitation*, vol. 16, no. 1, p. 1, 2024. doi: 10.1186/s13102-023-00787-0