World Electric Vehicle Journal

MDPI

Article

Digital Transformation and Green Innovation in New Energy Vehicle Manufacturing Firms: Evidence from China

Xing Xin [1], Yu Chen [2] and Jiang Hu [2,3,*]

[1] School of Modern Finance, Jiaxing Nanhu University, Jiaxing 314001, China; xinxing898@163.com
[2] School of Automotive Business, Hubei University of Automotive Technology, Shiyan 442000, China; cy792775855@163.com
[3] School of International Education, Hubei University of Automotive Technology, Shiyan 442000, China
* Correspondence: 20150007@huat.edu.cn

Abstract

Digital transformation is increasingly reshaping firms' innovation processes. Meanwhile, new energy vehicle (NEV) manufacturing firms face growing challenges in reconciling rapid technological advancement with resource-intensive production and stringent environmental requirements. This study examines whether digital transformation is associated with corporate green innovation and identifies its potential transmission channels. Using a balanced panel of 85 Chinese A-share-listed NEV manufacturing firms from 2014 to 2023, we employ a double machine learning framework to estimate the relationship between digital transformation and green innovation. The results reveal a significant positive association between digital transformation and green innovation. This finding remains robust across alternative cross-fitting configurations, machine learning algorithms, winsorization thresholds, and sample restrictions, with supplementary support from instrumental-variable estimation and propensity score matching. Mechanism tests suggest that R&D personnel input, R&D funding input, and human resource level represent potential transmission channels. Heterogeneity analysis further reveals positive coefficient estimates across different firm and industry categories. Extended analyses reveal positive associations with green invention patents, green utility-model patents, total patent applications, and non-green patent applications. These findings highlight the importance of integrating digital capabilities with innovation resources and workforce development to support green technological innovation in NEV manufacturing.

Keywords: digital transformation; green innovation; new energy vehicle manufacturing firms; double machine learning; human resource level

Academic Editor: Joeri Van Mierlo

Received: 9 October 2026
Revised: 16 November 2026
Accepted: 18 November 2026
Published: 19 November 2026

Copyright: © 2026 by the authors. Published by MDPI on behalf of the World Electric Vehicle Association. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Against the backdrop of global climate action and carbon-neutrality goals, the green transition of transportation and manufacturing has become an important component of sustainable development. Green innovation combines technological progress with environmental improvement and enables firms to reconcile competitiveness with environmental responsibility (Liu et al., 2023 [1]; Liang and Sun, 2024 [2]). This issue is especially salient in the new energy vehicle (NEV) industry. Although NEVs can reduce emissions during vehicle use, the production of power batteries, critical materials, electric drive systems, automotive electronics, and complete vehicles remains resource- and energy-

intensive. Moreover, advances in electrification, intelligence, and connectivity have increased the technological interdependence among hardware, software, materials, and manufacturing processes. Green innovation is therefore central not only to cleaner production, but also to the safety, performance, and lifecycle sustainability of NEV products.

Digital transformation offers NEV manufacturers a potential means of responding to these intertwined technological and environmental demands. For incumbent firms, it involves strategic renewal and the reconfiguration of business models and organizational capabilities rather than the isolated adoption of digital tools (Simsek et al., 2024 [3]). Nevertheless, digitalization does not automatically generate environmental benefits. Digital technologies may improve operational performance while also creating additional resource consumption and environmental or social costs; their sustainability value depends on how effectively they are integrated with green practices and organizational objectives (Cardinali and De Giovanni, 2022 [4]). Whether digital transformation is translated into stronger green innovation in NEV manufacturing therefore remains an empirical question.

Existing firm-level studies generally report a positive relationship between digital transformation and green innovation, but most rely on broad samples of listed companies or manufacturing firms (Tao et al., 2024 [5]; Xu et al., 2024 [6]). Recent reviews further indicate that this literature remains geographically concentrated and that industry-specific settings require greater attention (Song et al., 2025 [7]). This limitation matters for NEV manufacturing because its innovation process combines complex technological interdependence, intensive R&D, extensive data requirements, and stringent environmental constraints. Consequently, evidence drawn from general corporate samples may not fully explain how digital capabilities are converted into green innovation in this industry, through which internal channels this conversion occurs, or whether this relationship varies across different firm and industry contexts.

To address this gap, this study uses a balanced panel of 85 Chinese A-share-listed NEV manufacturing firms from 2014 to 2023, comprising 850 firm-year observations. Green innovation is measured using the combined number of green invention patent applications and green utility-model patent applications, while digital transformation is measured using the CSMAR corporate digital transformation index. The empirical analysis is centered on a double machine learning framework and evaluates the robustness of the main findings by varying cross-fitting schemes, replacing machine learning algorithms, treating extreme values, and employing instrumental-variable estimation. Building on this empirical framework, the analysis examines R&D personnel input, R&D funding input, and human resource level as three potential transmission channels. The study further conducts heterogeneity analyses across high-tech and non-high-tech firms, firms operating in high- and low-carbon industries, and industries with different levels of digitalization to assess whether the estimated relationship varies across firm and industry contexts. It also distinguishes between green invention and green utility-model patent applications and extends the analysis to total and non-green patent applications to assess whether the innovation-enhancing effects of digital transformation extend beyond green technological innovation and whether the expansion of green innovation crowds out innovation in non-green technological domains.

The empirical results show that digital transformation is consistently associated with a higher level of green innovation among NEV manufacturing firms, and this relationship remains robust across alternative machine learning specifications, sample treatments, and endogeneity controls. The mechanism analysis distinguishes three potential internal channels: the allocation of employees to specialized R&D activities, the intensity of financial resources devoted to R&D, and the educational composition of the overall workforce. The subgroup results further show that the positive relationship between digital transformation and green innovation remains evident among high-tech firms, firms operating in

low-carbon industries, and firms operating in industries with lower levels of digitalization. Further analysis shows that digital transformation is positively associated with green invention patent applications, green utility-model patent applications, total patent applications, and non-green patent applications. These findings indicate that its innovation-enhancing effects encompass both substantive and incremental forms of green technological development and do not appear to crowd out innovation in non-green technological fields.

This study contributes to the literature in three respects. First, it moves beyond the broad settings of listed firms and general manufacturing enterprises by focusing specifically on the NEV manufacturing industry, where multiple technologies are highly coupled, R&D activities are particularly complex, and green manufacturing constraints are relatively stringent. This industry-specific setting helps reveal how the green innovation effects of digital transformation operate within a complex technological system. Second, the study distinguishes between two complementary dimensions of human resources. R&D personnel input captures the functional allocation of employees directly engaged in innovation, whereas human resource level captures the educational quality of the firm's overall workforce through the proportion of employees holding a bachelor's degree or above. Together with R&D funding input, these mechanisms explain how digital capabilities may support green innovation through specialized innovation labor, financial resource formation, and workforce skill upgrading. Third, the study applies double machine learning to accommodate potentially complex nonlinear relationships and supplements the analysis with instrumental-variable identification. It also investigates heterogeneity related to firms' technological characteristics, industry energy intensity, and industry-level digitalization, together with green invention, green utility-model, total, and non-green patent outcomes, thereby providing a more detailed characterization of the boundary conditions and breadth of the innovation-enhancing effects associated with digital transformation.

2. Literature Review and Research Hypotheses

2.1. Digital Transformation and Green Innovation

Digital transformation goes beyond the adoption of isolated digital technologies; it entails organization-wide changes in strategy, organizational structures, business processes, value creation, and capabilities enabled by combinations of digital technologies (Hanelt et al., 2021 [8]; Vial, 2019 [9]; Verhoef et al., 2021 [10]). These changes alter how firms process information and allocate resources. In particular, digital technologies reduce the costs of storing, computing, transmitting, searching, tracking, and verifying information, thereby providing a new infrastructure for corporate innovation (Goldfarb and Tucker, 2019 [11]).

One pathway from digital transformation to green innovation concerns information processing and knowledge integration. Digitally collected data on energy use, emissions, technological development, and market demand allow firms to identify environmental constraints and technological opportunities more rapidly. Digital platforms also connect R&D, production, procurement, and marketing activities and broaden access to external knowledge. Information technology capabilities can expand the breadth and depth of firms' knowledge bases (Wei et al., 2022 [12]), while the innovation value of big data depends on the variety and velocity of information rather than on data volume alone (Ghasemaghaei and Calic, 2020 [13]). Greater source variety can further strengthen knowledge search and dynamic adaptation (Pedota, 2023 [14]). Digital transformation can therefore support green innovation by improving the identification, combination, and reuse of dispersed technological knowledge.

A second pathway concerns the reallocation and coordination of innovation factors. Data-factor development can facilitate information sharing and improve the allocation of

resources devoted to green innovation (Liu et al., 2026 [15]). In manufacturing, Industry 4.0 technologies may strengthen green innovation through open innovation (Mubarak et al., 2021 [16]), while digital supply-chain platforms can convert information-processing capabilities into improved environmental performance (Li et al., 2020 [17]). Firm-level evidence also indicates that digital transformation can expand firms' resource and knowledge bases and promote green innovation (He et al., 2024 [18]). These findings suggest that the effect arises not from technological deployment alone, but from the reorganization of information, knowledge, and innovation resources.

This logic is particularly relevant to NEV manufacturing, where power batteries, electric drive systems, automotive electronics, lightweight materials, thermal management, and intelligent technologies are closely interdependent. Research on electric-vehicle battery management illustrates the increasing importance of digital analytical capabilities: machine-learning methods are increasingly used to estimate lithium-ion battery state of charge and state of health (Poh et al., 2022 [19]), while advanced data-driven methods are being developed for automotive battery health prognostics (Poh, 2025 [20]). These engineering studies do not directly establish a firm-level green innovation effect, but they demonstrate that data, algorithms, and computing capabilities are becoming embedded in core NEV technologies. At the organizational level, digital technologies and digital organization can also reshape green innovation networks and strengthen collaboration among NEV firms (Li et al., 2023 [21]). However, systematic firm-level evidence remains limited on whether overall digital transformation promotes green innovation and which internal resources and organizational conditions support that relationship. Accordingly, the following hypothesis is proposed:

H1. *Digital transformation significantly promotes green innovation in NEV manufacturing firms.*

2.2. The Transmission Channel of R&D Personnel Input

Green innovation is knowledge-intensive and requires sustained professional coordination. In NEV manufacturing, technological development spans power batteries, electric drive systems, automotive electronics, new materials, and software. Firms therefore require R&D personnel from different professional backgrounds who can absorb external knowledge, use digital tools, and combine technological components across disciplinary and functional boundaries.

Digital transformation changes the capability structure of corporate R&D. As data analytics, simulation, intelligent manufacturing, and software development become embedded in innovation processes, firms need a larger functional allocation of employees who can directly undertake R&D tasks and combine domain expertise with digital tools. The innovation benefits of big-data applications also depend on supportive human-resource practices and management commitment (El-Kassar and Singh, 2019 [22]). Thus, the informational advantages created by digitalization can enter technological development only when firms allocate sufficient personnel to specialized R&D activities.

Increasing the share of R&D personnel can broaden technological search, improve knowledge absorption, and strengthen the cross-disciplinary coordination required for complex green technologies. R&D personnel input therefore provides a human-capital channel through which digital capabilities can be converted into green innovation. Accordingly, the following hypothesis is proposed:

H2. *Digital transformation promotes green innovation in NEV manufacturing firms by increasing R&D personnel input.*

2.3. The Transmission Channel of R&D Funding Input

Green technology development also requires sustained financial support. NEV-related R&D involves investment in equipment, software, experimental testing, engineering validation, and repeated design iteration. Because development cycles are long and technological uncertainty is high, firms must maintain stable funding throughout the innovation process.

Digital transformation can improve the allocation of internal innovation resources by making project information more visible and R&D decisions more traceable. It may also reduce information asymmetry between firms and external capital providers, thereby improving access to innovation finance. Existing evidence identifies R&D investment as a transmission channel between digital transformation and green innovation (Yang and Xu, 2025 [23]), while research on digitalized green supply chains similarly indicates that financing conditions influence the conversion of digital capabilities into environmental improvements (Liao et al., 2024 [24]). Digital transformation may therefore affect both firms' willingness to allocate funds to R&D and their capacity to sustain that investment.

When financial resources can be continuously directed toward technological experimentation, design optimization, and engineering validation, firms are better able to sustain green technology projects through lengthy development cycles. R&D funding input therefore constitutes the material channel through which digital capabilities support green technological outcomes. Accordingly, the following hypothesis is proposed:

H3. *Digital transformation promotes green innovation in NEV manufacturing firms by increasing R&D funding input.*

2.4. The Transmission Channel of Human Resource Level

Green innovation requires not only specialized R&D personnel but also a high-quality workforce capable of absorbing technological knowledge and supporting cross-functional collaboration. In NEV manufacturing, green technology development involves the integration of complex technological and environmental information across R&D, production, procurement, quality control, and data management. Unlike R&D personnel input, which reflects the functional allocation of employees directly engaged in research and development, human resource level reflects the educational quality and knowledge foundation of the firm's overall workforce.

Digital transformation can reshape firms' workforce composition by increasing their demand for highly educated and skilled employees. As big data, artificial intelligence, simulation systems, and intelligent manufacturing become embedded in production and business processes, firms require employees with stronger analytical skills, learning capabilities, and interdisciplinary knowledge. Previous research emphasizes that workforce skill development is essential for the effective implementation of organizational digital transformation (Sousa and Rocha, 2019 [25]). The integration of big-data technologies with human resource practices can also strengthen firms' sustainable capabilities (Singh and El-Kassar, 2019 [26]). Firms undergoing digital transformation therefore have greater incentives to recruit and retain highly educated employees and to reallocate their workforce toward knowledge- and skill-intensive positions. Consistent with this reasoning, firm-level evidence shows that digital transformation can improve human capital, with the proportion of employees holding a bachelor's degree or above used as an indicator of human capital level (Wang and Zhong, 2024 [27]).

A higher human resource level can strengthen firms' capacity to absorb digital technologies, interpret environmental information, and coordinate knowledge across organizational functions. Highly educated employees are generally better equipped to identify environmental problems, understand complex technological information, and combine digital knowledge with product design, engineering validation, and process

improvement. Existing research further shows that green human capital plays an important role in translating human resource practices into green innovation (Song et al., 2021 [28]). Human resource level therefore constitutes a workforce-quality foundation through which digital capabilities can be converted into green technological outcomes. Accordingly, the following hypothesis is proposed:

H4. *Digital transformation promotes green innovation in NEV manufacturing firms by improving their human resource level.*

Based on the preceding theoretical analysis, this study develops a conceptual framework to explain how digital transformation influences green innovation in NEV manufacturing firms. Digital transformation is hypothesized to promote green innovation directly by enhancing information processing, knowledge integration, and innovation resource allocation. In addition, three distinct but complementary transmission channels are proposed: R&D personnel input, R&D funding input, and human resource level. These channels reflect the roles of specialized R&D personnel allocation, financial investment in innovation, and the educational quality of the overall workforce in converting digital capabilities into green innovation outcomes. Figure 1 illustrates the proposed conceptual framework and the hypothesized relationships (H1–H4).

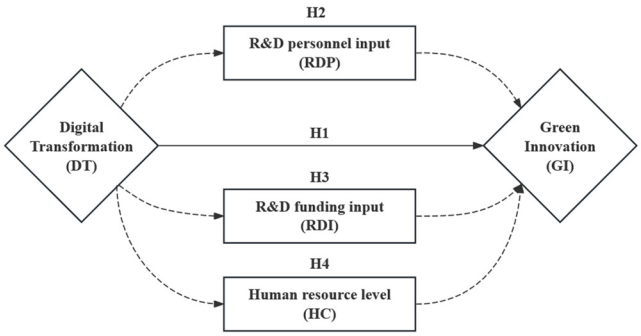


Figure 1. Conceptual Framework and Hypothesized Relationships.

3. Research Design

3.1. Data and Sample

This study selects Chinese A-share-listed firms engaged in new energy vehicle (NEV) manufacturing and related supply-chain activities as the research sample. In addition to NEV manufacturers, the sample also covers listed firms involved in the production of power batteries and battery materials, electric motors and drive systems, automotive electronics and electrical components, thermal-management systems, transmission and chassis components, lightweight structural parts, tires, and automotive glass. Sample firms are selected based on their principal business revenue composition and industry classifications, with a focus on companies whose principal business activities are related to NEV manufacturing and its supporting supply chain. Data on corporate digital transformation, financial characteristics, corporate governance, and R&D activities are obtained from the CSMAR database. Green patent application data are sourced from the CNRDS and the CNIPA.

The sample period spans 2014–2023. The period begins in 2014 because China's NEV industry entered a stage of rapid expansion during this time, accompanied by continuous technological development and increasingly close linkages between vehicle manufacturers and core component suppliers. Meanwhile, the key variables used in this study, including corporate digital transformation, green patent applications, financial

characteristics, and corporate governance, can be consistently matched from 2014 onward. The sample period ends in 2023 to ensure the completeness and comparability of core variables across different databases, thereby facilitating reliable data matching and the construction of a balanced panel dataset. Accordingly, 2023 is set as the final year of the sample period.

Following common research practice, the initial sample was processed as follows. First, firms designated as ST or *ST were excluded, leaving only firms with normal listing status. Second, firm-year observations with missing values for the core variables were removed. Third, linear interpolation was applied to a small number of intermittent missing values in non-core continuous control variables. We further retained only firms with continuous observations throughout the 2014–2023 period, resulting in a balanced panel of 85 NEV manufacturing firms and 850 firm-year observations. All empirical analyses were conducted using Stata 18.0.

3.2. Variables and Measurement

The empirical analysis involves green innovation, digital transformation, mechanism variables, control variables, and variables used in the extended analyses. Their definitions and measurements are presented in Table 1.

Table 1. Variable Definitions.

Variable Type	Variable	Symbol	Measurement
Dependent variable	Green innovation	GI	ln(1 + green invention patent applications + green utility-model patent applications)
Core explanatory variable	Digital transformation	DT	CSMAR Corporate Digital Transformation Index
Alternative explanatory variable	Adjusted digital transformation	DT_adj	CSMAR index excluding digitalization outcomes, with the remaining weights renormalized
Mechanism variable	R&D personnel input	RDP	Number of R&D personnel/total number of employees
Mechanism variable	R&D funding input	RDI	R&D expenditure/operating revenue
Mechanism variable	Human resource level	HC	Share of Employees with a Bachelor's Degree or Above
Control variable	Firm size	Size	ln(total assets)
Control variable	Leverage	Lev	Total liabilities/total assets
Control variable	Tobin's Q	TobinQ	Firm-level Tobin's Q
Control variable	Profitability	ROA	Return on total assets
Control variable	Growth	Growth	Operating revenue growth rate
Control variable	Chair–general manager duality	Dual	1 if the board chair and general manager positions are held by the same person; 0 otherwise
Extended-analysis variable	Green invention patent applications	GIP	ln(1 + green invention patent applications)
Extended-analysis variable	Green utility model patent applications	GUMP	ln(1 + green utility model patent applications)
Extended-analysis variable	Total patent applications	TPA	ln(1 + total patent applications)
Extended-analysis variable	Non-green patent applications	NGPA	ln(1 + non-green patent applications)

Corporate green innovation (GI) is measured using the annual number of green patent applications. Green patents provide an observable indicator of firms' technological innovation activities related to environmental protection, energy conservation, and resource efficiency (Li et al., 2023 [21]). Previous studies have distinguished between green invention patents and green utility-model patents to capture different forms of green technological innovation (Yang and Xu, 2025 [23]). Compared with patent grants, patent applications more closely reflect the timing of innovation activities and are less affected by examination delays. Accordingly, GI is calculated as the natural logarithm of one plus the sum of green invention patent applications and green utility-model patent applications. This measure captures firms' green patenting activities across different levels of technological complexity.

The core explanatory variable, digital transformation (DT), is measured using the Corporate Digital Transformation Index provided by the CSMAR database. The index evaluation system comprises five dimensions at the listed-company level—strategic guidance (34.72%), technology-driven development (16.20%), organizational empowerment (9.69%), digitalization outcomes (27.13%), and digital application (8.84%)—as well as one macro-level dimension, environmental support (3.42%). Together, these dimensions capture firms' digital transformation from the perspectives of digital strategy, technology adoption, organizational change, transformation outcomes, and the external environment. This study directly uses the composite index provided by the database; a higher index value indicates a higher level of corporate digital transformation. The detailed dimensions and their measurement components are presented in Table 2.

Table 2. Dimensions and Measurement Components of the Corporate Digital Transformation Index.

First-Level Dimension	Weight	Key Measurement Components
Strategic guidance	34.72%	Digital-related positions established at the management level; digital innovation orientation; and strategic emphasis on digitalization.
Technology-driven development	16.20%	Artificial intelligence, big data, cloud computing, blockchain, and other digital technologies.
Organizational empowerment	9.69%	Investment in digital capital and human resources; and the development of digital infrastructure and innovation bases.
Environmental support	3.42%	The industry innovation environment, the R&D environment, and regional digital infrastructure.
Digitalization outcomes	27.13%	Digital patents, standards, academic papers, qualifications, and other related innovation outcomes.
Digital application	8.84%	The application of digital technologies in technological innovation, process innovation, and business innovation.

As shown in Table 2, the digitalization outcomes dimension of the CSMAR index includes digital patents and other innovation-related outputs, which may overlap with the patent-based measure of green innovation. To assess the robustness of the empirical findings to this potential measurement overlap, an adjusted digital transformation index (DT_adj) is constructed by excluding the digitalization outcomes dimension, which accounts for 27.13% of the original index. The weights of the remaining five dimensions are proportionally renormalized to 47.65% for strategic guidance, 22.23% for technology-driven development, 13.30% for organizational empowerment, 4.69% for environmental support, and 12.13% for digital application. The adjusted index is subsequently employed as an alternative explanatory variable to examine whether the positive relationship between digital transformation and green innovation persists after excluding innovation-related outcomes from the original index.

Three mechanism variables are considered to examine the potential roles of R&D personnel input, R&D funding input, and human resource level. R&D personnel input (RDP) is measured as the proportion of R&D personnel in total employment. This indicator reflects the allocation of specialized human resources to firms' research and development activities. Existing research identifies human capital accumulation and the development of specialized capabilities as important factors linking digital transformation to corporate green innovation (Wang and Zhong, 2024 [27]). Accordingly, the share of R&D personnel is used to capture firms' innovation-related human resource input.

R&D funding input (RDI) is measured as R&D expenditure divided by operating revenue, reflecting the intensity of financial resources allocated to research and development relative to firms' operating scale. R&D investment provides essential financial support for technological development, experimentation, and innovation activities. Previous studies have identified R&D investment as an important factor in the relationship between digital transformation and corporate green innovation (Xu et al., 2024 [6]; Yang and Xu, 2025 [23]).

Human resource level (HC) is measured as the proportion of employees holding a bachelor's degree or higher in total employment, reflecting the educational composition of firms' workforce. Employees with higher educational attainment may possess stronger capabilities for learning, knowledge integration, and the application of advanced technologies, which are particularly relevant to the technological complexity of NEV manufacturing. Unlike RDP, which focuses on personnel directly engaged in R&D activities, HC captures the broader educational structure of firms' human resources. This distinction allows the analysis to examine the potential role of workforce human capital separately from specialized R&D personnel input.

The control variables include firm size (Size), leverage (Lev), Tobin's Q (TobinQ), profitability (ROA), growth (Growth), and chair–general manager duality (Dual). Size is measured as the natural logarithm of total assets, while Lev is calculated as total liabilities divided by total assets. TobinQ captures firms' market valuation and growth opportunities, and ROA reflects profitability relative to total assets. Growth is measured by the operating revenue growth rate. Dual is a binary variable equal to one when the board chair and general manager positions are held by the same individual and zero otherwise. These definitions are consistent with Table 1.

The extended analyses employ four alternative innovation outcomes: green invention patent applications (GIP), green utility-model patent applications (GUMP), total patent applications (TPA), and non-green patent applications (NGPA). Existing studies indicate that different patent categories capture distinct forms of technological innovation (Yang and Xu, 2025 [23]). GIP is measured as the natural logarithm of one plus the number of green invention patent applications, reflecting relatively substantive green technological advances. GUMP is measured as the natural logarithm of one plus green utility-model patent applications, capturing incremental and application-oriented green technological improvements. TPA is measured as the natural logarithm of one plus total patent applications, representing firms' overall patenting activities. NGPA is measured as the natural logarithm of one plus non-green patent applications and is introduced to examine innovation activities outside green technological fields. These measures allow the analysis to distinguish between different forms of green innovation and assess whether the relationship between digital transformation and innovation extends to broader technological domains (Liu et al., 2023 [1]).

3.3. Empirical Strategy

To examine the effect of digital transformation on green innovation in NEV manufacturing firms, this study employs a double machine learning (DML) framework as its

primary estimation approach. Unlike conventional regression models that require the functional relationships between explanatory variables and outcomes to be specified in advance, DML allows these relationships to be estimated flexibly through machine learning. This feature is particularly relevant in the present study, as the relationships between firm characteristics, digital transformation, and green innovation may involve nonlinearities and complex interactions. By combining machine learning with Neyman orthogonalization and cross-fitting, DML helps reduce regularization bias and overfitting in estimating the parameter of interest (Chernozhukov et al., 2018 [29]). The partially linear DML model is specified as follows:

$$GI_{it} = \theta DT_{it} + g(W_{it}) + \varepsilon_{it} \quad (1)$$

$$DT_{it} = m(W_{it}) + \nu_{it} \quad (2)$$

where GI_{it} denotes the green innovation level of firm i in year t, and DT_{it} represents its digital transformation level. W_{it} is a vector of observed control information, including firm size, leverage, Tobin's Q, profitability, revenue growth, chair–general manager duality, and year indicators. The functions g(·) and m(·) represent unknown nuisance functions, while ε_{it} and ν_{it} are error terms. The parameter θ is the coefficient of interest, capturing the relationship between digital transformation and green innovation after accounting for the observed control information.

Under this framework, DML estimates the conditional expectations of green innovation and digital transformation given W_{it}, respectively, and obtains their corresponding residuals. The target parameter is then estimated using the relationship between these residualized variables. This partialling-out procedure, together with Neyman orthogonalization, reduces the sensitivity of the estimated coefficient to errors in the machine learning estimation of the nuisance functions (Chernozhukov et al., 2018 [29]). Random forest is employed to estimate the nuisance functions because it can capture nonlinear relationships and interactions without imposing a prespecified functional form (Breiman, 2001 [30]). The estimation is implemented using the Stata ddml framework (Ahrens et al., 2024 [31]), with pystacked used to implement the random forest learner (Ahrens et al., 2023 [32]). For the computational implementation, the empirical analysis is conducted in Stata/MP 18.0 using ddml (version 1.4.4) and pystacked (version 0.7.5), in conjunction with Python 3.12.4 and scikit-learn 1.6.1. The baseline random forest estimator employs the default hyperparameter settings of scikit-learn, including 100 trees, the squared-error splitting criterion, unrestricted maximum tree depth, a minimum of two observations required for node splitting, and a minimum of one observation per leaf node. All available predictors are considered at each split (max_features = 1.0), with bootstrap sampling enabled. No additional hyperparameter optimization is performed. The neural network and gradient boosting estimators employed in the robustness checks are likewise implemented using their respective default hyperparameter settings. Detailed hyperparameter configurations of the three machine-learning algorithms are provided in Supplementary Table S1 in the accompanying Supplementary Materials, while the Stata estimation code is available through the GitHub repository specified in the Data Availability Statement.

To account for the panel structure of the data, this study adopts firm-level five-fold cross-fitting using the fcluster(id) option. All observations belonging to the same firm are assigned to the same fold. In each iteration, four folds are used for training and the remaining fold for out-of-sample prediction, ensuring that observations from the same firm do not appear simultaneously in the training and validation samples. The Stata random-number generator is initialized with a fixed seed of 42. As no separate Python seed is specified, pystacked's default pyseed(-1) setting is used, under which the Python seed is drawn from Stata's random-number generator. For statistical inference, standard errors

are clustered at the firm level using the cluster(id) option to account for heteroskedasticity and within-firm dependence across annual observations. Firm-level cross-fitting prevents information leakage across training and validation samples, whereas firm-clustered standard errors account for within-firm dependence in statistical inference.

3.4. Further Empirical Design

Following the baseline DML estimation, the empirical analysis is extended in three directions: robustness and endogeneity checks, mechanism analysis, and heterogeneity and innovation-outcome analyses. The first examines the stability of the baseline results through robustness tests and instrumental-variable estimation. The second investigates the roles of R&D personnel, R&D funding input, and human resource level. The third explores whether the effects vary across firms and whether they extend to alternative innovation outcomes.

Although DML flexibly accommodates observed firm characteristics and their potentially nonlinear relationships, reverse causality, omitted variables, and other unobserved factors may still affect the core estimates. To further examine potential endogeneity, this study constructs an interaction-based instrumental variable. Specifically, the instrumental variable (IV_station_LDT) is constructed by interacting the historical presence of Ming-dynasty postal stations with the firm's own one-period-lagged digital transformation index:

$$IV_{it} = Station_i^{Ming} \times DT_{i,t-1} \quad (3)$$

where $Station_i^{Ming}$ is a binary indicator equal to one if at least one Ming-dynasty postal station is recorded in the city where firm i is registered, and zero otherwise. The historical station indicator is matched to firms according to their registered cities, while $DT_{i,t-1}$ denotes the digital transformation index of the same firm lagged by one year.

The selection of this instrument is grounded in previous empirical studies that employ historical postal infrastructure to address potential endogeneity. Historical postal networks may capture persistent regional differences in transportation and information connectivity, which are potentially relevant to contemporary digital development (Xu et al., 2024 [33]). Building on this rationale, existing studies have constructed interaction-based instruments by combining historical postal station characteristics with year indicators or time-varying economic variables (Guo et al., 2023 [34]; Zhang et al., 2024 [35]; Yang et al., 2025 [36]). These studies provide empirical and methodological support for incorporating historical postal infrastructure into instrumental-variable specifications, while the present study adapts this approach to firm-level digital transformation dynamics.

Building on this construction, the relevance of the instrument rests on the relationship between firms' historical location characteristics and their digital transformation trajectories. The lagged digital transformation index captures each firm's prior digitalization level, while the historical postal-station indicator distinguishes firms according to their registered locations. Their interaction is expected to be associated with current digital transformation by capturing differences in firms' digitalization trajectories across locations with different historical transportation and information-transmission conditions. Regarding the exclusion restriction, the historical postal-station indicator predates contemporary corporate innovation, while firm fixed effects absorb its time-invariant standalone effect and other time-invariant firm characteristics. Year fixed effects account for common temporal shocks, and firm-level controls further capture observable differences across firms. These design features reduce the likelihood that the instrument is related to current green innovation through time-invariant location characteristics or common macroeconomic changes. Taken together, the IV estimates are interpreted as complementary evidence consistent with the baseline association.

In addition to the instrumental-variable approach, this study employs propensity score matching (PSM) to address potential sample-selection bias associated with observable firm characteristics. Firms are classified into treatment and control groups according to whether their digital transformation index in 2015 is above the sample median. Propensity scores are estimated using a logit model based on firm-level characteristics in 2014. A 1:2 nearest-neighbor matching procedure with a caliper of 0.05 is then implemented under the common-support condition. The matched firms are subsequently retained for panel regression analysis over the period 2015–2023.

The mechanism analysis focuses on three potential channels through which digital transformation may be associated with corporate green innovation: R&D personnel input (RDP), R&D funding input (RDI), and human resource level (HC). These channels are examined within a three-step DML framework, focusing on the relationship between digital transformation and each mechanism variable, as well as the association between the mechanism variable and green innovation conditional on digital transformation. The corresponding empirical specifications are as follows:

$$M_{it}^{(k)} = \alpha_k DT_{it} + f_k(W_{it}) + u_{it}^{(k)} \quad (4)$$

$$GI_{it} = \theta_k DT_{it} + \beta_k M_{it}^{(k)} + h_k(W_{it}) + \varepsilon_{it}^{(k)} \quad (5)$$

where $M_{it}^{(k)}$ denotes the mechanism variable, with k representing RDP, RDI, or HC; W_{it} represents the vector of firm-level control variables and year indicators; and $f_k(\cdot)$ and $h_k(\cdot)$ are unknown functions capturing the effects of these covariates. In Equation (4), α_k captures the relationship between digital transformation and the corresponding mechanism variable. In Equation (5), β_k captures the association between the mechanism variable and green innovation conditional on digital transformation, while θ_k represents the corresponding coefficient of digital transformation in the joint specification.

To assess the sensitivity of the baseline estimates to model specifications and sample treatments, a series of robustness checks is conducted from two perspectives. Regarding estimation methods, the number of cross-fitting folds is adjusted from the baseline five-fold setting to three and eight folds, while neural networks and gradient boosting are employed as alternatives to random forests. These adjustments examine the sensitivity of the estimates to different sample-splitting schemes and machine learning algorithms. Regarding sample treatments, continuous variables are winsorized at the 1% and 5% levels, and the model is re-estimated after excluding firms located in municipalities directly under the central government and observations from the COVID-19 pandemic years (2020–2022). These procedures help assess the influence of extreme observations, regional sample composition, and pandemic-related shocks on the baseline findings.

To examine the heterogeneity in the relationship between digital transformation and corporate green innovation, this study conducts analyses along three dimensions: technological characteristics, industry energy intensity, and industry-level digitalization. Technological characteristics are classified according to the National Bureau of Statistics' High-tech Industry Classification (Manufacturing) (2013) and High-tech Industry Classification (Services) (2013), with firms assigned to high-tech and non-high-tech groups accordingly. Industry energy intensity is classified based on the National Bureau of Statistics' statistical definition of the six major energy-intensive industries, distinguishing between high-carbon and low-carbon firms. Industry-level digitalization is divided using the annual median, with firms above the median in each year assigned to the high-digitalization group and the remaining firms assigned to the low-digitalization group.

To further characterize the innovation outcomes associated with digital transformation, this study extends the analysis along two dimensions: patent types and technological fields. Within green innovation, green invention patent applications (GIP) and

green utility-model patent applications (GUMP) are examined separately to distinguish between different forms of green technological innovation. Green invention patents generally involve greater technological complexity and more substantive technological advances, whereas utility-model patents tend to reflect application-oriented and incremental improvements. This distinction helps identify whether the relationship between digital transformation and green innovation varies across different levels of technological sophistication. The analysis is further extended to total patent applications (TPA) and non-green patent applications (NGPA), capturing firms' overall patenting activities and innovation outside green technological fields, respectively. This broader examination provides additional insight into whether the innovation-related association of digital transformation is confined to green technological activities or extends to other areas of corporate innovation.

4. Results

4.1. Descriptive Statistics

Table 3 reports the descriptive statistics for 85 Chinese A-share-listed new energy vehicle (NEV) manufacturing firms over the period 2014–2023, comprising 850 firm-year observations. Green innovation (GI) exhibits a mean of 1.7990 and a standard deviation of 1.5645, with values ranging from 0 to 5.3181, suggesting substantial variation in green patenting activity across firm-year observations. Digital transformation (DT) also displays considerable dispersion, with a mean of 41.5273, a standard deviation of 9.2987, and a range of 23.4600–68.3067. These distributional differences indicate that the sampled firms are characterized by varying levels of digitalization and green innovation. Among the control variables, the average leverage ratio is 0.4768, while the mean values of Tobin's Q and return on assets (ROA) are 1.8625 and 0.0412, respectively. Revenue growth exhibits a relatively wide distribution, ranging from −0.5612 to 4.5190, reflecting substantial variation in firm-level growth performance. The mean of CEO duality is 0.2894, indicating that the chairperson and chief executive officer positions are held by the same individual in approximately 28.94% of firm-year observations.

Table 3. Descriptive Statistics of Main Variables.

Variables	N	Mean	SD	Min	Max
GI	850	1.7990	1.5645	0.0000	5.3181
DT	850	41.5273	9.2987	23.4600	68.3067
Size	850	23.0612	1.2230	20.3292	27.4677
Lev	850	0.4768	0.1654	0.0698	0.9375
TobinQ	850	1.8625	1.0857	0.7966	10.3605
ROA	850	0.0412	0.0470	−0.2574	0.2129
Growth	850	0.1650	0.3157	−0.5612	4.5190
Dual	850	0.2894	0.4538	0.0000	1.0000

4.2. Baseline Results

Table 4 presents the baseline DML estimate in Column (1), the two-way fixed-effects (TWFE) estimate in Column (2), and the DML estimate using the adjusted digital transformation index DT_adj in Column (3). The baseline DML model employs random forests with five-fold firm-level cross-fitting and controls for observed firm characteristics and year effects. Standard errors are clustered at the firm level. As shown in Column (1), the coefficient of digital transformation (DT) is positive and statistically significant at the 1% level ($\beta = 0.0393$), indicating a positive relationship between digital transformation and green innovation among new energy vehicle (NEV) manufacturing firms. Given that

green innovation is measured as the natural logarithm of one plus green patent applications, a one-standard-deviation increase in DT, equivalent to 9.2987 index points, is associated with an approximately 44.1% increase in one plus green patent applications. This result indicates that the estimated relationship is economically meaningful and provides empirical support for H1.

Column (2) reports the TWFE estimate after controlling for firm and year fixed effects. The coefficient of digital transformation remains positive and statistically significant at the 5% level ($\beta = 0.0177$, SE = 0.0084). This estimate is smaller than the baseline DML coefficient of 0.0393 because the two models rely on different sources of variation. The TWFE coefficient is identified from changes within the same firm over time after absorbing time-invariant firm characteristics, whereas the DML estimate is obtained by flexibly adjusting for observed firm-level controls and year effects without absorbing firm fixed effects. Despite the difference in magnitude, both estimates indicate a statistically significant positive relationship between digital transformation and green innovation.

Column (3) further evaluates the robustness of the findings using an alternative measure of digital transformation. As described in Section 3.2, the adjusted index (DT_adj) excludes the digitalization outcomes dimension of the original CSMAR index to address potential overlap with the patent-based measure of green innovation. The DML estimate remains positive and statistically significant at the 1% level ($\beta = 0.0276$, SE = 0.0067). This result indicates that the positive association persists after excluding innovation-related outputs from the digital transformation measure. Taken together, the TWFE and adjusted-index estimates provide complementary evidence supporting the robustness of the positive relationship across alternative model specifications and measurement approaches.

Table 4. Baseline and Alternative-Measure Estimation Results.

	(1) **DML**	**(2)** **TWFE**	**(3)** **DML**
	GI	GI	GI
DT	0.0393 *** (0.0083)	0.0177 ** (0.0084)	—
DT_adj	—	—	0.0276 *** (0.0067)
Firm fixed effects	No	Yes	No
Control variables	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
Firm-clustered SE	Yes	Yes	Yes
Observations	850	850	850

Note: Firm-clustered standard errors are reported in parentheses. ** and *** indicate significance at the 5% and 1% levels, respectively.

4.3. Robustness Checks

Table 5 presents the robustness checks under alternative estimation settings. Regarding cross-fitting configurations, changing the number of folds from five to three and eight yields estimated coefficients of 0.0418 and 0.0417, respectively, both significant at the 1% level. Regarding machine learning algorithms, replacing random forests with neural networks and gradient boosting produces coefficients of 0.0539 and 0.0379, respectively, both of which are also significantly positive at the 1% level. These results indicate that the positive relationship between digital transformation and corporate green innovation does not depend on a particular cross-fitting configuration or machine learning algorithm.

The analysis further examines the sensitivity of the estimates to data treatment and sample composition. Winsorizing continuous variables at the 1% and 5% levels yields

coefficients of 0.0394 and 0.0418, respectively, both significant at the 1% level. After excluding firms located in municipalities directly under the central government, the estimated coefficient is 0.0400. Excluding observations from the COVID-19 pandemic years (2020–2022) produces a coefficient of 0.0373. Both estimates remain statistically significant at the 1% level, based on 750 and 595 firm-year observations, respectively. Overall, the positive relationship between digital transformation and corporate green innovation persists under alternative data treatments and different geographical and temporal sample restrictions.

Table 5. Robustness Checks.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Changing sample-splitting ratios		Replacing machine learning algorithms		Winsorization		Excluding municipalities	Excluding COVID-19 Years
	K-folds = 3	K-folds = 8	Neural network	Gradient boosting	1% winsorization	5% winsorization	K-folds = 5	K-folds = 5
	GI	GI	GI	GI	GI	GI	GI	GI
DT	0.0418 *** (0.0087)	0.0417 *** (0.0089)	0.0539 *** (0.0118)	0.0379 *** (0.0083)	0.0394 *** (0.0084)	0.0418 *** (0.0088)	0.0400 *** (0.0091)	0.0373 *** (0.0086)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm-clustered SE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	850	850	850	850	850	850	750	595

Note: Firm-clustered standard errors are reported in parentheses. *** indicates significance at the 1% level.

4.4. Endogeneity Analysis

Table 6 reports the instrumental-variable estimation results obtained using two-stage least squares (2SLS), with firm and year fixed effects and standard errors clustered at the firm level. In the first stage, the coefficient of IV_station_LDT is 0.2992 (SE = 0.0451) and statistically significant at the 1% level, indicating a positive association between the instrument and firms' digital transformation. The Kleibergen–Paap rk LM statistic is 23.104 ($p < 0.001$), rejecting the null hypothesis of underidentification. The Kleibergen–Paap rk Wald F statistic is 44.050, alleviating concerns regarding weak identification. These diagnostics assess the relevance and strength of the instrument; however, the exclusion restriction remains unverified. In the second stage, the estimated coefficient of digital transformation is 0.0689 (SE = 0.0285) and statistically significant at the 5% level ($p = 0.018$). The positive coefficient is directionally consistent with the baseline DML result and is interpreted as complementary evidence consistent with the positive association between digital transformation and corporate green innovation.

To address potential sample selection concerns, this study employs propensity score matching (PSM). Firms are divided into treatment and control groups according to the median digital transformation index in 2015. Based on firm-level characteristics in 2014, a Logit model is used to estimate propensity scores, followed by 1:2 nearest-neighbor matching with a caliper of 0.05 and a common support restriction. Table 7 presents the PSM results. Before matching, the coefficient of digital transformation is 0.0182 ($p = 0.035$), based on 765 firm-year observations. After matching, 73 firms and 657 firm-year observations are retained, accounting for 85.88% of the original sample. The coefficient remains

significantly positive at 0.0191 (p = 0.043), suggesting that the positive relationship between digital transformation and green innovation persists after matching.

Table 6. Instrumental Variable Estimates.

	(1)	(2)
	DT	GI
IV_station_LDT	0.2992 *** (0.0451)	—
DT	—	0.0689 ** (0.0285)
KP rk LM statistic	23.104	—
KP rk LM p-value	0.0000	—
KP rk Wald F statistic	44.050	
Control variables	Yes	Yes
Year fixed effects	Yes	Yes
Firm fixed effects	Yes	Yes
Firm-clustered SE	Yes	Yes
Observations	850	850

Note: Firm-clustered standard errors are reported in parentheses. ** and *** indicate significance at the 5% and 1% levels, respectively.

Table 7. Propensity Score Matching Results.

	(1)	(2)
	GI	GI
DT	0.0182 ** (0.0085)	0.0191 ** (0.0093)
Control variables	Yes	Yes
Year fixed effects	Yes	Yes
Firm fixed effects	Yes	Yes
Firm-clustered SE	Yes	Yes
PSM	No	Yes
Observations	765	657
Number of firms	85	73

Note: Firm-clustered standard errors are reported in parentheses. ** indicates significance at the 5% level.

4.5. Mechanism Tests

Table 8 reports the results of the mechanism analysis. For R&D personnel input (RDP), the coefficient of digital transformation is significantly positive ($\beta = 0.1815$, $p < 0.01$). When RDP and digital transformation are jointly included in the green innovation equation, RDP exhibits a positive coefficient of 0.0287 ($p < 0.01$), while the coefficient of digital transformation decreases from 0.0393 to 0.0341. A similar pattern emerges for R&D funding input (RDI). Digital transformation is positively associated with RDI ($\beta = 0.0794$, $p < 0.01$), and RDI is positively related to green innovation after accounting for digital transformation ($\beta = 0.1152$, $p < 0.01$). The corresponding coefficient of digital transformation declines to 0.0299.

Beyond R&D-related factors, this study further examines whether digital transformation is associated with changes in firms' human resource level (HC), measured by the proportion of employees holding a bachelor's degree or higher. The results show that digital transformation is positively associated with this proportion ($\beta = 0.0020$, $p = 0.075$), with

the coefficient significant at the 10% level. In the joint specification, HC exhibits a positive and statistically significant association with green innovation (β = 3.0742, $p < 0.01$), while the coefficient of digital transformation remains positive and statistically significant (β = 0.0324).

Taken together, these findings are consistent with the proposed roles of R&D personnel input, R&D funding input, and human resource level as three transmission channels through which digital transformation is associated with green innovation, thereby providing empirical support for H2–H4.

Table 8. Mechanism Analysis.

	Baseline	**RDP**		**RDI**		**HC**	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	GI	RDP	GI	RDI	GI	HC	GI
DT	0.0393 *** (0.0083)	0.1815 *** (0.0697)	0.0341 *** (0.0083)	0.0794 *** (0.0183)	0.0299 *** (0.0075)	0.0020 * (0.0011)	0.0324 *** (0.0078)
RDP	—	—	0.0287 *** (0.0102)	—	—	—	—
RDI	—	—	—	—	0.1152 *** (0.0341)	—	—
HC	—	—	—	—	—	—	3.0742 *** (0.7834)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm-clustered SE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	850	850	850	850	850	850	850

Note: Firm-clustered standard errors are reported in parentheses. * and *** indicate significance at the 10% and 1% levels, respectively.

4.6. Heterogeneity Analysis

Table 9 reports the subsample estimates across firms' technological attributes, industry carbon-emission attributes, and industry digitalization levels. Regarding technological attributes, the coefficient of digital transformation is positive and statistically significant for high-tech firms (β = 0.0367, $p < 0.01$), while the corresponding estimate for non-high-tech firms is 0.0189 and statistically insignificant ($p = 0.392$). Regarding industry carbon-emission attributes, the coefficient is positive and statistically significant among firms in low-carbon industries (β = 0.0363, $p < 0.01$), whereas the estimate for firms in high-carbon industries is 0.0021 and statistically insignificant ($p = 0.923$). Regarding industry digitalization, the estimated coefficients are 0.0257 for firms in highly digitalized industries ($p = 0.054$) and 0.0359 for firms in industries with lower levels of digitalization ($p = 0.005$).

Taken together, the estimated coefficients are positive across all subsamples, although their magnitudes and statistical precision vary. Formal cross-group tests do not reject the equality of the corresponding coefficients at conventional significance levels. The interaction coefficients and corresponding Wald-test *p*-values are provided in Supplementary Table S2 in the accompanying Supplementary Materials. The results therefore indicate that the positive relationship between digital transformation and green innovation is directionally consistent across the examined firm and industry contexts.

Table 9. Heterogeneity Analysis.

	Technical Attribute		**Carbon Emissions**		**Industry Digitalization**	
	(1)	(2)	(3)	(4)	(5)	(6)

	High-tech	Non-high-tech	High-carbon	Low-carbon	High digitalization	Low digitalization
	GI	GI	GI	GI	GI	GI
DT	0.0367 ***	0.0189	0.0021	0.0363 ***	0.0257 *	0.0359 ***
	(0.0099)	(0.0221)	(0.0218)	(0.0096)	(0.0134)	(0.0128)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Firm-clustered SE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	721	129	136	714	426	424

Note: Firm-clustered standard errors are reported in parentheses. * and *** indicate significance at the 10% and 1% levels, respectively.

4.7. Extended Analysis of Innovation Effects

To distinguish between different forms and levels of green innovation, this study examines green invention and utility model patent applications separately. Green invention patents generally reflect more substantive technological advances and higher levels of technological novelty, whereas utility model patents primarily capture incremental improvements and application-oriented technological developments. As reported in Table 10, the estimated coefficients of digital transformation are 0.0358 for green invention patent applications (GIP) and 0.0275 for green utility model patent applications (GUMP), both statistically significant at the 1% level. These results indicate that digital transformation is positively associated with both more substantive green technological innovation and incremental, application-oriented green innovation, suggesting that its innovation-enhancing effects extend across different forms of green technological development.

The analysis is further extended to firms' overall patenting activities. When total patent applications (TPA) are used as the dependent variable, the coefficient of digital transformation remains positive and statistically significant at the 1% level, with an estimated value of 0.0242. This finding indicates that the innovation-enhancing effects of digital transformation are not confined to green technological innovation but are also reflected in firms' aggregate patent output. Thus, the contribution of digital transformation to innovation appears to extend beyond a particular technological domain.

A further question is whether the expansion of green innovation comes at the expense of innovation activities in other technological fields. To examine this possibility, non-green patent applications (NGPA) are considered separately. The estimated coefficient of digital transformation is 0.0221 and is statistically significant at the 1% level. This positive association, together with the increases in green invention and utility model patent applications, suggests that digital transformation promotes patent output across both green and non-green technological domains. Thus, no evidence of a crowding-out effect on non-green patent output is observed.

Taken together, the results show that the innovation-enhancing effects of digital transformation extend beyond green innovation alone. Its positive association with different types of green patents, total patent applications, and non-green patent applications points to a broader expansion of firms' innovation output. This finding provides a more comprehensive understanding of the role of digital transformation in shaping firms' innovation activities.

Table 10. Extended Analysis of Innovation Outcomes.

	(1)	**(2)**	**(3)**	**(4)**
	GIP	GUMP	TPA	NGPA
DT	0.0358 ***	0.0275 ***	0.0242 ***	0.0221 ***
	(0.0075)	(0.0065)	(0.0064)	(0.0063)

Controls	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Firm-clustered SE	Yes	Yes	Yes	Yes
Observations	850	850	850	850

Note: Firm-clustered standard errors are reported in parentheses. *** indicates significance at the 1% level.Figure 2 synthesizes the empirical findings presented in Tables 4–10, integrating the baseline results, robustness and endogeneity tests, mechanism analysis, heterogeneity analysis, and extended innovation outcomes into a unified analytical framework. Collectively, these findings provide empirical evidence on the relationship between digital transformation and corporate green innovation across alternative model specifications, potential transmission channels, and innovation outcomes.

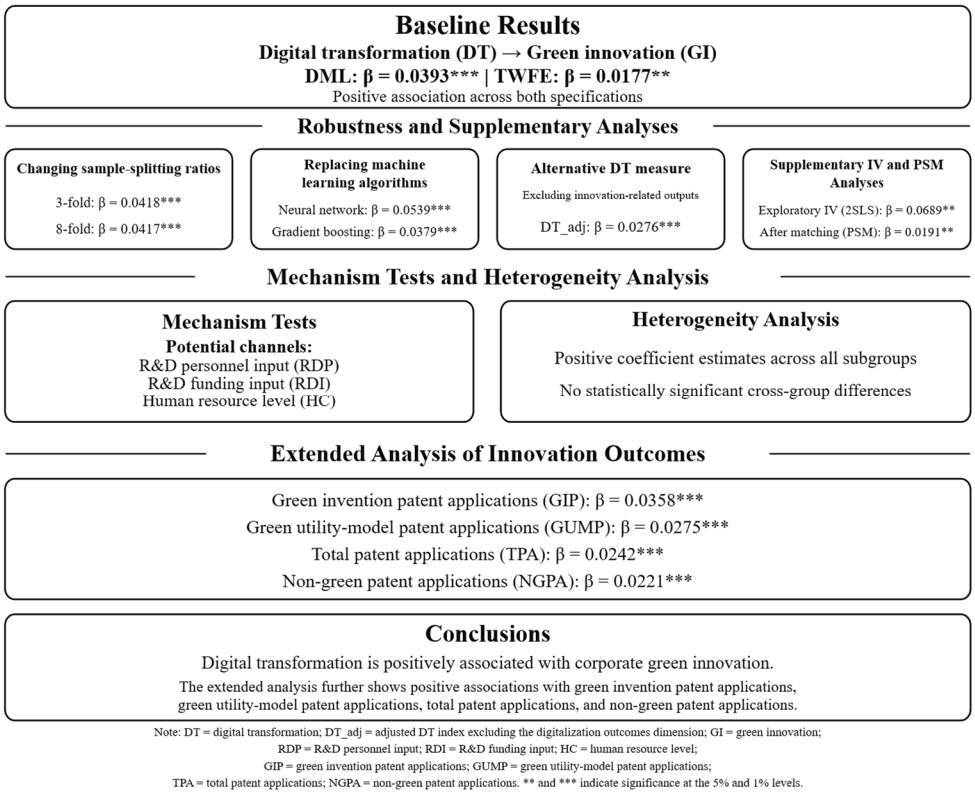


Figure 2. Summary of Empirical Findings.

5. Discussion

The positive relationship between digital transformation and green innovation found in this study is consistent with evidence from general samples of listed and manufacturing firms (Tao et al., 2024 [5]; Xu et al., 2024 [6]). It also accords with the broader "twin transition" perspective, according to which digital and environmental innovation are increasingly interconnected (Aklilu et al., 2025 [37]). The contribution of this study is therefore not to claim for the first time that digitalization is related to green development. Rather, it provides firm-level evidence on how this relationship is manifested in NEV manufacturing, examines the internal resources that may serve as potential transmission channels, and assesses its consistency across different firm and industry contexts. This distinction is important because policy-level recognition of the green potential of digital technologies does not by itself explain how firms convert digital capabilities into observable green technological outputs.

Compared with general manufacturing, NEV production represents a particularly interdependent technological system. Power batteries and battery materials, electric drive

systems, automotive electronics, thermal management, lightweight components, software, and intelligent control must be developed and validated jointly. Changes in one module may affect vehicle safety, energy efficiency, product performance, and the design of other modules. Digital transformation may therefore be associated with green innovation not only because it supports production automation and reduces information costs, but also because digital systems can create data continuity among design, simulation, testing, manufacturing, quality control, and market feedback. The results complement prior NEV research on digital innovation networks (Li et al., 2023 [21]) by providing firm-level evidence of a positive association between digital transformation and firms' internal green patent output. A potential industry-specific explanation is that, in a tightly coupled technological system, the environmental value of digitalization may partly reflect its capacity to coordinate multidisciplinary knowledge and repeated engineering validation.

The three mechanism results provide exploratory evidence regarding the internal resource foundations that may underlie the observed relationship. R&D personnel input captures the functional allocation of employees directly engaged in technological development, while R&D funding input represents the financial support devoted to equipment, software, experimentation, and repeated validation. Human resource level, measured through the educational composition of the overall workforce, reflects a broader absorptive foundation extending beyond the R&D department. This distinction is especially relevant to NEV firms because digital information must be understood and applied not only by researchers but also by employees in production, procurement, quality control, data management, and other functions. The observed associations therefore suggest that specialized innovation labor, sustained financial commitment, and broad workforce capability may constitute complementary channels rather than interchangeable inputs. Digital technologies may generate green innovation value when firms possess the personnel, funding, and organization-wide knowledge base needed to convert data into technological and process improvements.

The subgroup analysis provides additional contextual insight into how digital transformation may operate across different firm and industry environments. Statistically significant positive estimates are observed among high-tech firms, firms operating in low-carbon industries, and firms operating in industries with lower levels of digitalization, although formal cross-group tests do not identify significant differences between the corresponding coefficients. These dimensions may nevertheless shape how firms apply digital technologies to innovation activities. In high-tech firms, accumulated technological knowledge and absorptive capacity may facilitate the integration of digital tools and data into R&D processes. In carbon-intensive production settings, digital investment may initially be directed toward environmental monitoring, traceability, process control, and regulatory compliance before being reflected in patent output. Industry digitalization may also influence the form of digital deployment: firms in less digitalized industries may focus on basic digital infrastructure and process integration, whereas firms in highly digitalized industries may emphasize more advanced and complex applications.

The patent-based results provide a further NEV-specific insight. Digital transformation is positively associated with both green invention patents and green utility-model patents, suggesting that it supports relatively substantive technological advances as well as incremental, application-oriented improvements. In NEV manufacturing, these forms of innovation are likely to be complementary: advances in batteries, electric drive systems, or thermal management often require corresponding modifications to equipment, production processes, and component applications. Moreover, total and non-green patent applications also increase with digital transformation. The evidence therefore does not suggest that firms expand green patenting simply by diverting a fixed stock of innovative resources away from other technological fields. Instead, digitalization appears to broaden

knowledge search, technological recombination, and the overall scope of firms' innovation activities, although patent counts cannot establish the commercial or environmental value of every invention.

Recent developments in artificial intelligence illustrate how these mechanisms may operate in practice. AI-assisted design and simulation can help NEV firms process testing data, identify design anomalies, and shorten iterative engineering cycles. In battery management, data-driven digital twins have been used to improve the estimation and interpretation of battery state of charge and state of health (Njoku et al., 2024 [38]). Intelligent digital-twin architectures have also been proposed for the diagnosis and prognostic health management of electric-vehicle powertrains (El Hadraoui et al., 2024 [39]), while research on battery-oriented digital twins further highlights their potential for performance, lifecycle, and safety management (Rajesh et al., 2024 [40]). Such applications can support predictive maintenance, reduce avoidable component replacement and downtime, and connect operational data with subsequent product and process improvement. These examples are consistent with the workforce, R&D, and knowledge-integration capabilities identified in this study, but they should be treated as technological illustrations rather than applications separately identified by the empirical analysis.

The positive average relationship should nevertheless not be interpreted as implying that greater digitalization is invariably beneficial. Digital infrastructure, sensors, computing equipment, AI training, and data storage consume electricity and materials and may generate electronic waste; recent OECD analysis similarly emphasizes that the twin transition involves both environmental opportunities and environmental costs (Aklilu et al., 2025 [37]). At the firm level, digital projects may create cost overruns, legacy-system incompatibilities, duplicated platforms, and greater organizational complexity when investment exceeds managerial and absorptive capacity. NEV firms also face cybersecurity, intellectual-property, and data-governance risks because battery, vehicle-operation, R&D, and supply-chain data may be commercially sensitive and safety-critical. AI applications introduce additional concerns regarding model opacity, biased training data, model drift, and erroneous predictions. These risks provide a plausible explanation for why the benefits of digital transformation are not uniform across firms. Firms should therefore evaluate digitalization using innovation, environmental, security, and lifecycle criteria rather than the number of digital tools deployed, and should combine staged investment with cybersecurity governance, human oversight, and continuous capability development.

6. Conclusions and Implications

6.1. Conclusions

Using a balanced panel of 85 Chinese A-share-listed NEV manufacturing firms from 2014 to 2023, this study applies double machine learning to examine the relationship between digital transformation and green innovation. The positive relationship persists when the digitalization outcomes dimension is excluded from the digital transformation index and remains robust to changes in the number of cross-fitting folds and machine-learning algorithms, the treatment of extreme observations, and modifications to the sample scope. The instrumental-variable and propensity-score-matching analyses yield directionally consistent results and are interpreted as complementary evidence supporting the baseline association.

The mechanism analysis provides exploratory evidence consistent with R&D personnel input, R&D funding input, and human resource level as three potential transmission channels linking digital transformation to green innovation. Specifically, digital transformation is positively associated with each channel variable, while each channel variable is positively associated with green innovation when digital transformation is included in the corresponding specification. Together, these results are consistent with the proposed

transmission pathways. The subgroup estimates are positive across the examined firm and industry categories, although their magnitude and statistical precision vary. Formal cross-group tests do not reject coefficient equality at conventional significance levels, indicating that the relationship is directionally consistent across the examined contexts. The extended analysis further identifies positive associations between digital transformation and green invention patents, green utility-model patents, total patent applications, and non-green patent applications. Taken together, the findings suggest that digital transformation is associated with both substantive and incremental forms of green technological development, as well as a broader expansion of firms' patenting activities, with no observable evidence of crowding out innovation in non-green technological fields.

6.2. Implications

Theoretically, the study contextualizes the digital transformation–green innovation relationship within the technologically interdependent setting of NEV manufacturing. Its industry-specific contribution lies in providing exploratory evidence that three complementary resource dimensions—specialized R&D labor, sustained R&D funding, and the educational quality of the overall workforce—may serve as potential channels linking digital capabilities to green innovation. The analysis further considers this relationship across firms with different technological attributes, industry carbon-emission attributes, and industry digitalization levels, while the extended results show positive associations across both green and non-green innovation outcomes.

For managers, digital transformation should be integrated with R&D and workforce planning rather than treated as an isolated technology-purchasing program. NEV firms should connect data across design, testing, production, quality control, procurement, and after-sales feedback; maintain stable funding for experimentation and validation; and combine specialist recruitment with broader employee training. AI and digital-twin applications should be introduced through staged projects with clear performance, environmental, explainability, and cybersecurity requirements. Firms in high-carbon or less technologically capable settings may first need to strengthen data quality, workforce skills, and cross-functional coordination before undertaking complex digital investments.

For policymakers, support should extend beyond subsidies for digital equipment. Public digital infrastructure, industrial Internet platforms, shared testing facilities, workforce development, and standards for secure data exchange can reduce adoption costs and improve the conversion of digital information into green technologies. Differentiated policies are also needed: high-carbon firms may require digital support to be coordinated with cleaner-process retrofits, while firms in less digitalized industries may benefit from accessible platforms and technical services that enable catch-up. Evaluation systems should account for innovation outcomes, energy and material use, and digital environmental footprints.

Several limitations define the scope of these implications. Chinese listed NEV firms are generally larger and better resourced than unlisted firms and small component suppliers, and China's industrial policy, digital infrastructure, environmental regulation, capital markets, and supply-chain structure may differ from those of other countries. The analytical framework may therefore be transferable, but the estimated relationships should not be assumed to generalize automatically across institutional environments, firm sizes, energy systems, or market structures. Patent applications also do not fully capture commercialization, environmental performance, or economic value, and the study does not directly identify individual AI or digital-twin applications. Future research could use cross-country and SME samples, direct measures of digital technologies, patent quality and commercialization indicators, lifecycle emissions, and formal cross-group coefficient tests to evaluate the external validity and technological specificity of the findings.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/wevj17110607/s1, Table S1: Hyperparameter Configurations of Machine-Learning Learners; Table S2: Formal Tests of Cross-Group Coefficient Differences.

Author Contributions: Conceptualization, X.X. and J.H.; methodology, Y.C. and X.X.; software, Y.C.; validation, X.X., Y.C. and J.H.; formal analysis, X.X. and Y.C.; investigation, X.X. and J.H.; resources, X.X. and J.H.; data curation, X.X. and Y.C.; writing—original draft preparation, X.X.; writing—review and editing, Y.C. and J.H.; visualization, Y.C.; supervision, J.H.; project administration, J.H. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Research on the Leapfrog Development of Zhejiang's "415X" Industrial Clusters under the "Core–Collaboration" Spatial Organization (award number: Y202661738).

Data Availability Statement: The processed dataset and Stata estimation code used in this study are publicly available in a GitHub repository. The processed dataset is available at https://github.com/tksgk6h5hb-wq/Green-innovation/blob/main/11.12.xlsx (accessed on 12 November 2026), and the Stata estimation code is available at https://github.com/tksgk6h5hb-wq/Green-innovation/blob/main/11.15.do (accessed on 15 November 2026). Additional information regarding the data processing procedures can be obtained from the corresponding author upon reasonable request.

Acknowledgments: The authors gratefully acknowledge the support received during the preparation of this study and thank all authors for their contributions to the completion of this manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Liu, X.; Liu, F.; Ren, X. Firms' digitalization in manufacturing and the structure and direction of green innovation. *J. Environ. Manag.* **2023**, *335*, 117525. https://doi.org/10.1016/j.jenvman.2023.117525.
2. Liang, P.; Sun, X. Does digital transformation promote the green innovation of China's listed companies? *Environ. Dev. Sustain.* **2024**, *26*, 22199–22235. https://doi.org/10.1007/s10668-024-05059-6.
3. Simsek, Z.; Heavey, C.; König, A.; Stam, W. Leading digital transformation in incumbent firms: A strategic entrepreneurship framing. *Strateg. Entrep. J.* **2024**, *18*, 91–102. https://doi.org/10.1002/sej.1499.
4. Cardinali, P.G.; De Giovanni, P. Responsible digitalization through digital technologies and green practices. *Corp. Soc. Responsib. Environ. Manag.* **2022**, *29*, 984–995. https://doi.org/10.1002/csr.2249.
5. Tao, A.; Wang, C.; Zhang, S.; Kuai, P. Does enterprise digital transformation contribute to green innovation? Micro-level evidence from China. *J. Environ. Manag.* **2024**, *370*, 122609. https://doi.org/10.1016/j.jenvman.2024.122609.
6. Xu, C.; Sun, G.; Kong, T. The impact of digital transformation on enterprise green innovation. *Int. Rev. Econ. Finance* **2024**, *90*, 1–12. https://doi.org/10.1016/j.iref.2023.11.001.
7. Song, C.; Teh, S.Y.; Alnoor, A.; Shao, M. Trends of the digital transformation and green innovation using PRISMA and bibliometric analysis review. *Discov. Sustain.* **2025**, *6*, 1010. https://doi.org/10.1007/s43621-025-01855-w.
8. Hanelt, A.; Bohnsack, R.; Marz, D.; Antunes Marante, C. A systematic review of the literature on digital transformation: Insights and implications for strategy and organizational change. *J. Manag. Stud.* **2021**, *58*, 1159–1197. https://doi.org/10.1111/joms.12639.
9. Vial, G. Understanding digital transformation: A review and a research agenda. *J. Strateg. Inf. Syst.* **2019**, *28*, 118–144. https://doi.org/10.1016/j.jsis.2019.01.003.
10. Verhoef, P.C.; Broekhuizen, T.; Bart, Y.; Bhattacharya, A.; Dong, J.Q.; Fabian, N.; Haenlein, M. Digital transformation: A multidisciplinary reflection and research agenda. *J. Bus. Res.* **2021**, *122*, 889–901. https://doi.org/10.1016/j.jbusres.2019.09.022.
11. Goldfarb, A.; Tucker, C. Digital economics. *J. Econ. Lit.* **2019**, *57*, 3–43. https://doi.org/10.1257/jel.20171452.
12. Wei, S.; Xu, D.; Liu, H. The effects of information technology capability and knowledge base on digital innovation: The moderating role of institutional environments. *Eur. J. Innov. Manag.* **2022**, *25*, 720–740. https://doi.org/10.1108/EJIM-08-2020-0324.
13. Ghasemaghaei, M.; Calic, G. Assessing the impact of big data on firm innovation performance: Big data is not always better data. *J. Bus. Res.* **2020**, *108*, 147–162. https://doi.org/10.1016/j.jbusres.2019.09.062.

14. Pedota, M. Big data and dynamic capabilities in the digital revolution: The hidden role of source variety. *Res. Policy* **2023**, *52*, 104812. https://doi.org/10.1016/j.respol.2023.104812.
15. Liu, X.; Wang, M.; Wang, T. Enterprise green innovation: The role of data factor inputs. *Int. Rev. Econ. Finance* **2026**, *105*, 104769. https://doi.org/10.1016/j.iref.2025.104769.
16. Mubarak, M.F.; Tiwari, S.; Petraite, M.; Mubarik, M.; Raja Mohd Rasi, R.Z. How Industry 4.0 technologies and open innovation can improve green innovation performance? *Manag. Environ. Qual. Int. J.* **2021**, *32*, 1007–1022. https://doi.org/10.1108/MEQ-11-2020-0266.
17. Li, Y.; Dai, J.; Cui, L. The impact of digital technologies on economic and environmental performance in the context of industry 4.0: A moderated mediation model. *Int. J. Prod. Econ.* **2020**, *229*, 107777. https://doi.org/10.1016/j.ijpe.2020.107777.
18. He, Q.; Ribeiro-Navarrete, S.; Botella-Carrubi, D. A matter of motivation: The impact of enterprise digital transformation on green innovation. *Rev. Manag. Sci.* **2024**, *18*, 1489–1518. https://doi.org/10.1007/s11846-023-00665-6.
19. Poh, W.Q.T.; Xu, Y.; Tan, R.T.P. A review of machine learning applications for Li-ion battery state estimation in electric vehicles. In *2022 IEEE PES Innovative Smart Grid Technologies–Asia (ISGT Asia)*; IEEE: Piscataway, NJ, USA, 2022; pp. 265–269. https://doi.org/10.1109/ISGTAsia54193.2022.10003481.
20. Poh, W.Q.T. Advanced Data-Driven Methods for Automotive Battery Health Prognostics. Doctoral Dissertation, School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, 2025. https://doi.org/10.32657/10356/199630.
21. Li, H.Y.; Liu, Q.; Ye, H.Z. Digital development influencing mechanism on green innovation performance: A perspective of green innovation network. *IEEE Access* **2023**, *11*, 22490–22504. https://doi.org/10.1109/ACCESS.2023.3252912.
22. El-Kassar, A.N.; Singh, S.K. Green innovation and organizational performance: The influence of big data and the moderating role of management commitment and HR practices. *Technol. Forecast. Soc. Change* **2019**, *144*, 483–498. https://doi.org/10.1016/j.techfore.2017.12.016.
23. Yang, T.; Xu, B. Green innovation in manufacturing enterprises and digital transformation. *Econ. Anal. Policy* **2025**, *85*, 571 – 578. https://doi.org/10.1016/j.eap.2024.12.024.
24. Liao, F.; Hu, Y.; Chen, M.; Xu, S. Digital transformation and corporate green supply chain efficiency: Evidence from China. *Econ. Anal. Policy* **2024**, *81*, 195–207. https://doi.org/10.1016/j.eap.2023.11.033.
25. Sousa, M.J.; Rocha, Á. Digital learning: Developing skills for digital transformation of organizations. *Future Gener. Comput. Syst.* **2019**, *91*, 327–334. https://doi.org/10.1016/j.future.2018.08.048.
26. Singh, S.K.; El-Kassar, A.N. Role of big data analytics in developing sustainable capabilities. *J. Clean. Prod.* **2019**, *213*, 1264–1273. https://doi.org/10.1016/j.jclepro.2018.12.199.
27. Wang, X.; Zhong, X. Digital transformation and green innovation: Firm-level evidence from China. *Front. Environ. Sci.* **2024**, *12*, 1389255. https://doi.org/10.3389/fenvs.2024.1389255.
28. Song, W.; Yu, H.; Xu, H. Effects of green human resource management and managerial environmental concern on green innovation. *Eur. J. Innov. Manag.* **2021**, *24*, 951–967. https://doi.org/10.1108/EJIM-11-2019-0315.
29. Chernozhukov, V.; Chetverikov, D.; Demirer, M.; Duflo, E.; Hansen, C.; Newey, W.; Robins, J. Double/debiased machine learning for treatment and structural parameters. *Econom. J.* **2018**, *21*, C1 – C68. https://doi.org/10.1111/ectj.12097.
30. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. https://doi.org/10.1023/A:1010933404324.
31. Ahrens, A.; Hansen, C.B.; Schaffer, M.E.; Wiemann, T. ddml: Double/debiased machine learning in Stata. *Stata J.* **2024**, *24*, 3–45. https://doi.org/10.48550/arXiv.2301.09397.
32. Ahrens, A.; Hansen, C.B.; Schaffer, M.E. pystacked: Stacking generalization and machine learning in Stata. *Stata J.* **2023**, *23*, 909–931. https://doi.org/10.1177/1536867X231212426.
33. Xu, J.; Wang, C.; Yin, X.; Wang, W. Digital economy and rural household resilience: Evidence from China. *Agric. Econ.* **2024**, *70*, 244–263. https://doi.org/10.17221/317/2023-AGRICECON.
34. Guo, P.; Hu, X.; Zhao, S.; Li, M. The growth impact of infrastructure capital investment: The role of regional innovation capacity—Evidence from China. *Econ. Res.-Ekon. Istraž.* **2023**, *36*, 2142632. https://doi.org/10.1080/1331677X.2022.2142632.
35. Zhang, X.; Tang, T.; Mo, E. Can urban e-commerce transformation improve economic resilience? A quasi-natural experiment from China. *PLoS ONE* **2024**, *19*, e0304014. https://doi.org/10.1371/journal.pone.0304014.
36. Yang, Y.; Wang, F.; Tang, B. The impact and mechanism of e-commerce on rural agriculture: Evidence from China. *NJAS Impact Agric. Life Sci.* **2025**, *97*, 2523380. https://doi.org/10.1080/27685241.2025.2523380.
37. Aklilu, A.Z.; Dussaux, D.; Verrier, A. Quantifying digital innovation for the twin transition: Historical trends, current landscape and links with environmental innovation. In *OECD Environment Working Papers*; OECD Publishing: Paris, France, 2025; p. 261. https://doi.org/10.1787/8cc4cff0-en.

38. Njoku, J.N.; Nwakanma, C.I.; Kim, D.-S. Explainable data-driven digital twins for predicting battery states in electric vehicles. *IEEE Access* **2024**, *12*, 83480–83501. https://doi.org/10.1109/ACCESS.2024.3413075.
39. El Hadraoui, H.; Ouahabi, N.; El Bazi, N.; Laayati, O.; Zegrari, M.; Chebak, A. Toward an intelligent diagnosis and prognostic health management system for autonomous electric vehicle powertrains: A novel distributed intelligent digital twin-based architecture. *IEEE Access* **2024**, *12*, 110729–110761. https://doi.org/10.1109/ACCESS.2024.3441517.
40. Rajesh, P.K.; Soundarya, T.; Jithin, K.V. Driving sustainability—The role of digital twin in enhancing battery performance for electric vehicles. *J. Power Sources* **2024**, *604*, 234464. https://doi.org/10.1016/j.jpowsour.2024.234464.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.