Article

Identifying the Policy Direction of National R&D Programs Based on Data Envelopment Analysis and Diversity Index Approach

Myoungjae Choi [1], Ohjin Kwon [2], Dongkyu Won [3] and Wooseok Jang [3],*

[1] Korea Institute of Science and Technology Information, University of Science and Technology, Daejeon 34141, Korea; mj1234@kisti.re.kr

[2] Department of Open Data Convergence Research, Korea Institute of Science and Technology Information, Busan 48059, Korea; dbajin@kisti.re.kr

[3] Division of Data analysis, Korea Institute of Science and Technology Information, Seoul 02456, Korea; dkwon@kisti.re.kr

* Correspondence: wsjang90@kisti.re.kr; Tel.: +82-2-3299-608

Abstract: The Korean government has been continuously conducting diverse national R&D programs to discover new growth engines. The Republic of Korea is one of the countries with the largest investment in national R&D, but its efficiency was relatively low. In response, this study established a framework to identify the characteristics and direction of outstanding R&D programs. In this study, the performance of the R&D programs was identified in the sub-program unit. The efficiency of the national R&D program was analyzed using the data envelopment analysis model through the outputs of the national R&D programs such as papers and patents. However, patent and paper output would take time to be realized. Therefore, this study also calculated the diversity index of R&D programs to identify their potential expected performance. This study applied the suggested framework in the electric vehicle fields, which is one of the core growth engines of South Korea. A list of outstanding programs was identified from the National Institute of Science and Technology Information (NTIS) data. Additionally, this study also discovered the main technology areas and their current issues of outstanding and -new R&D programs. These results could contribute to suggesting the policy direction to conduct high-performance national R&D programs.

Keywords: research development program; national research efficiency; data envelopment analysis; diversity index; data analysis

1. Introduction

Currently, the South Korean government is advocating "innovative growth" in an attempt to alleviate the imbalances in the economic structure that is too geared toward particular manufacturing sectors, while preparing for a decrease in working-age populations [1]. Innovative growth refers to a growth model for increasing economic productivity as well as bringing innovation to socio-economic systems through recent technologies that drive growth. In fact, the global economy is changing according to the flow of the knowledge economy based on the Fourth Industrial Revolution and not the industrialization model. In other words, the concept of innovative growth involves developing innovative technologies to solve new problems and encompasses mid-to-long-term development and technological industrialization processes lasting more than five years [2]. To this end, the government has specified not only seven key infrastructure areas such as artificial intelligence and the Internet of Things of the Fourth Industrial Revolution, but also three areas that are defined as DNA (data, 5G network, artificial intelligence) to be innovative infrastructure areas. According to the information from 2020 MSIT (Ministry

of Science and Information communications technology), these areas would impact early changes in other industrial sectors.

In this process, the South Korean government began discussing the efficiency of R&D (Research and Development) investments. South Korea's share of R&D investment is actually among the highest, yet its efficiency is relatively low [3]. In fact, with regard to the scale of R&D investment, Korea's R&D investment in 2022 was expected to be 29.8 trillion won, the largest scale in terms of government R&D investment relative to GDP and the second largest budget in total R&D investment. However, the problem with the transparency and efficiency in the expenditure of the national R&D budget has consistently been raised [4]. The ranking in scientific and technological competitiveness such as technology transfer rates and science infrastructure has hardly improved for the past several years. According to KOSTAT (Statistics Korea), the technology transfer rate declined by about 5 percentage points in 2019 compared to 2015. Because of the limited national budget, the Korean government has found it the most important issue in terms of its R&D investment to look for the fields for high efficiency. Moreover, the scale of South Korea's R&D investment is steadily increasing due to continuous national projects such as the Digital New Deal and the Green New Deal Project. It is expected that such efforts can lead to a breakthrough growth in national competitiveness in terms of science and technology if the efficiency in R&D budget allocation can be improved.

To maximize the efficiency of R&D investment, it is necessary to invest a large budget in areas that have the potential to yield a high ROI (return on investment). In addition, it is also needed to boost investment efficiency while minimizing inefficiency by following systematic budgeting [5]. Here, the analysis was carried out using a data envelopment analysis method. The DEA (data envelopment analysis) approach, which can clearly define the efficiency, was selected as an analysis method to select several excellent programs. DEA has DMUs (decision-making units) with multiple inputs and multiple outputs. This model is a non-parametric approach to selecting DMUs with a maximal efficiency that does not require several criteria in advance. For this reason, it is used when several variables with clear inputs and outputs need to be analyzed at the same time, and it is a method that is widely utilized for efficiency analysis for public R&D programs. In other words, ways to allocate budgets proactively to areas that are likely to maximize performance in the future should be identified, and strategies that ensure greater output relative to budget inputs should be devised [6]. However, existing research has only performed DEA at the level of government-funded research institutes. As a result, it was difficult to know which national R&D programs offered more efficiency [7].

The present study addressed this issue by conducting DEA based on patent and paper analysis, which was not commonly carried out with the existing institutional analysis units, but a subprogram unit in a national R&D program. In doing so, the study explored how programs can be more efficient. A more specific goal is to perform an efficiency analysis through the use of NTIS (National Science and Technology Information Service) data in the electric vehicle segment. This aims to increase efficiency in national R&D programs as well as to serve as a useful model for establishing science and technology policies and distributing budgets in the short or long term [8]. Moreover, we intend to overcome the limitations of existing DEAs using additional analysis conducted after DEA.

Specifically, by using NTIS program data, this study conducts in-depth research not previously available to more efficiently allocate limited R&D funds [9]. An analysis of NTIS data reveals it includes several programs related to national research. These details comprise, for instance, the total amount of funding, educational backgrounds of researchers, the length of the program, and the number of team members involved. In addition, under the performance information, the number of published papers and the registration status of patents can be found. Several studies have been conducted to examine the efficiency of public institutions by setting public institutions and government-funded laboratories as DMUs, but few studies have used DEA on a program in the unit of national

research subprogram managed by public institutions and government-funded laboratories [10].

Therefore, we conducted a highly utilized efficiency analysis for each DEA result value by setting the DMU for NTIS programs in program units rather than institutional units. In previous studies, such analysis involved the evaluation of institutions in terms of their efficiency and inefficiency. It is believed that the characteristics of public institutions vary greatly, an analysis of program efficiency at an institution-level poses limitations. This is because each research institution tends to specialize in a certain area of research. Hence, it could be decided in advance which public institutions will perform better based on the type of program a country is interested in developing. By contrast, our research seeks to understand which types of R&D investment programs with certain characteristics tend to succeed in each field, not by comparing across fields. Our study, therefore, focused on a specific field for an efficiency analysis by program unit. By doing so, we determined what program would be evaluated well in a particular technical field and why [11].

This study conducts a DEA approach to calculate the performance efficiency by national R&D programs. DEA is a nonparametric method based on linear programming that measures efficiency through multiple input and output variables. As a form of analysis without specifying a function, it derives a frontier with a maximal efficiency value of 1 and measures inefficiency by calculating the distance between that and the frontier [12]. This allows us to determine which factors require benchmarking and to perform a direct analysis using a production possibility set. DEA analyses were conducted for papers using citation information, impact factor values, and the number of papers, and for patents using reference and citation counts of patent and family information, and the number of cases, resulting in a list of outstanding R&D programs [13].

Existing studies on R&D performance (efficiency) analysis have primarily been based on quantitative data obtained through DEA. A number of patents and papers, as well as impacts factor and citations, were the major factors. However, quantitative performance is not evident until a year or more has passed since investment or research was conducted. To overcome these limitations, our study also included a qualitative analysis in which a diversity index that reflects the characteristics of the field was used. The concept of a diversity index consists largely of variety, balance, and disparity, and these characteristics can be used to define diversity in a field (program) [14]. Indeed, the R&D program can be easily improved or transformed with constant exchanges of information with a high degree of diversity [15]. It improves the flexibility and resilience of technologies as well as utilization methods by expanding the pool of resources. Consequently, programs with high diversity stimulate innovation and lead to improved productivity [16]. In some bibliometric studies, variety or balance was used to measure a predetermined "category" to define the level of development of the technology [17]. Similarly, by producing diversity indicators on the application plans for the various technologies, this study identifies technologies with a wide scope of usability and high development potential.

Furthermore, existing studies have traditionally used the excellence of a research group or program as a criterion for evaluation. However, in the present study, the topics were derived for each program to draw implications from the program obtained. Natural language processing was used to derive similar topics out of the vast amount of NTIS program data based on keywords or abstracts included in the data. Therefore, the implications of a particular area will be presented in a more concise manner. The research procedures and methods for this study are outlined in Section 2.

2. Methods

This study mainly consists of three modules. The entire process of this study is illustrated in Figure 1 below, and each module is discussed in detail starting from Section 2.1.
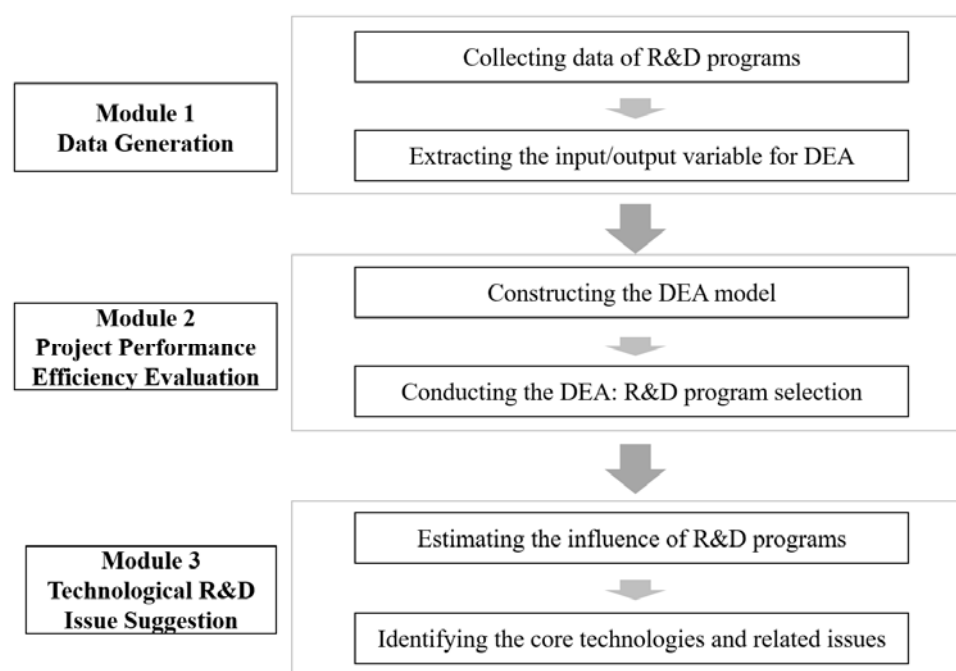
Figure 1. Research framework of DEA model in technological R&D programs.

2.1. Data Generation

The first step in conducting an R&D performance analysis is data generation. We collected information about national R&D programs from NTIS, which is a government-sponsored information portal operated by KISTI (Korea Institute of Science and Technology Information, South Korea). NTIS provides information on national R&D programs, including plans, tasks, manpower, and outcomes. Moreover, the portal provides comprehensive information about all national R&D programs. Launched in 2008, this portal delivers automatic recommendations for customized keyword information, systematic and automated management tools for tracking research status, detailed information on R&D by country and customized package information, and, recently, information on the latest research trends [18]. Users can access information about national R&D programs, such as information on programs related to national R&D programs, achievements (papers, patents), and equipment for research facilities. With its user-centered design, and by making close to 150 million national R&D-related records available, NTIS offers a wealth of information, including about public research institutions, researchers, and programs [19].

For this study, sub-programs within NTIS related to our research interests were searched. A total of two types of information were collected from the sub-program search. The first includes a description of the program, including its title, summary, and keywords. In addition, data on participants and R&D budgets were collected for use as input in the DEA. The second type is that associated with achievement generated by the program. As part of the achievement information, we collected patent and paper information derived from the program. The collected information was used in the DEA's output and qualitative analysis. An overview of the information collected is presented in Table 1.

Table 1. Gathered information from NTIS data.

Information Part		Gathered Information
Program description		Title, abstract, keyword, the number of researchers, R&D budget
Program output	Patent	Title, application/granted, assignee, IPC code, contribution
	Paper	Title, keyword, journal, contribution

In the next step, the input and output values for the DEA were derived from the collected information. To select input variables and output variables, previous studies similar to this study were investigated. As a result, it came to our attention that in many previous studies, the information on the R&D budget and workforce were set as input variables. In addition, output variables were mainly set based on the number of paper or patent publications. In addition, according to previous research, it was generally set at around five years, and R&D efficiency and trends were analyzed for the corresponding period. This study also diagnosed the efficiency of the program based on the five-year data. (Table 2)

Table 2. Previous research for DEA.

Research	Unit	Input Variable	Output Variable
Lee and Park et al. (2005) [20]	Nation	R&D expenditure, researchers	Patent, paper
Kocher et al. (2006) [21]	Nation	R&D expenditure, researchers	Paper
Coccia et al. (2007) [22]	Institution	researchers, public fund	Publication
Wang and Huang et al. (2007) [23]	Nations	R&D stocks, researchers	Patent, paper
Cullman et al. (2009) [24]	Nations	R&D expenditure, researchers	Patent
Zhao et al. (2015) [25]	Province	Researchers, R&D product, development costs	Patent, number of new products
Park and Shin (2018) [9]	Institution	Researchers, R&D fund	Patent, paper

Furthermore, DEA's output variables were selected as those frequently utilized in paper and patent analysis [20–27]. Considering the purpose of government R&D, development capability and the scope and coverage of R&D were selected as important criteria in this study. In addition, as novelty is a crucial factor in patents, additional variables were set to identify the degree of information exchange. The feature of selected output variables was summarized as Table 3.

Table 3. Feature of selected output variables.

Feature	Type	Variables
Development Capability	Paper	Citation, impact factor
	Patent	Forward citation
Scope and Coverage	Paper	Impact factor (indirect)
	Patent	Family size
Information Exchange	Paper	-
	Patent	Backward citation
Intensity	Paper	Frequency (weighted)
	Patent	Frequency (weighted)

Based on the previous research, the DEA input used the R&D budget and the number of participating personnel based on previous research. This information was gathered through the program description [15]. Moreover, the level of contribution by program was utilized to reflect the contribution of subprograms searched based on their achievements. In setting variables, both quantitative and qualitative factors were considered, which was

one of the points that distinguished this study from previous studies. Input and output variables constructed in such a way are reported in Table 4.

Table 4. Input and output variables for DEA.

Information Part	Variable	Description
Input variable	R&D budget	Government R&D budget allocation for strategic investments in government R&D programs
	Workforce	The number of researchers was classified based on the academic background of the researchers
Output variable	Patent	
	Frequency (weighted by contribution)	The number of applied patents during the R&D program
	Number of forward citations	The number of other patents cited a specific patent
	Number of backward citations	The number of reference patents that are cited by a specific patent
	Number of family patents	The number of patents for similar publications in different countries
	Paper	
	Frequency (weighted by contribution)	The number of papers applied during the R&D program
	Impact factor	The yearly mean number of citations of articles published in the last two years in a given journal
	Number of citations	The number of times that a publication has been cited by other publications

In this process, the frequency was calculated by the sum of weighted contributions for each publication. Contribution was the assigned values by the principal investigator depending on the degree of association of R&D programs. Workforce was calculated by the number of Bachelor's, Master's, and doctoral researchers in R&D programs, so we collected the information of the number of Bachelor's, Master's, and doctoral researchers. Workforce was calculated by weighting Ph.D., Master's, and Bachelor's personnel at 1, 0.5, and 0.2, respectively.

2.2. Program Performance Efficiency Evaluation

This study utilized the BCC (Banker, Charnes, and Cooper) model, which is designed as a linear function based on constant returns to scale [27]. VRS (variable returns to scale) assumes based on the variable with increasing production scale, thus acting as a hybrid between a DRS (decreasing return to scale) model and an IRS (increasing returns to scale) method [28]. Indeed, the BCC model closely resembles the curve of the rise and decline in technical life expectancy. An economics perspective finds that when new technologies or products emerge, there is a brief rise to a certain point followed by a gradual decline [29].

The CCR (Charnes, Cooper, and Rhodes) model is the most common form of the DEA and it assumes that the ratio of the weighted sum of outputs to each weighted sum of

input is not greater than 1 for the DMU. Furthermore, the model was started on the assumption that the sum of the weights of the input variable and output variable was greater than zero. The following is an equation that shows how the DEA is analyzed and the efficiencies obtained. First, in analyzing DMUs of Z, when DMU$_z$(z =1,2,…, z) is set as an input variable of Z (x_n(n = 1,2,…, N)) to derive an output variable of M, the efficiency of the DMU of K of the kth observation value is shown in Equation (1). DMU represents a program unit.

In NTIS data, it is a word that represents the business group. X and Y values represent each variable and lambda represents a weight value. Moreover, the S value can be understood as a dummy variable value that has no direct influence. In the BCC model, it is premised on VRS, and if on an input basis, it can be shown as follows:

$$\theta^{k^*} = \min_{\theta, \lambda, s^-, s^+} \left\{ \theta^* - \varepsilon \left(\sum_{n=1}^{N} s^-_n + \sum_{m=1}^{M} s^+_m \right) \right\}$$

s. t.

$$\theta^k x^k_n \geq \sum_{z=1}^{Z} x^z_n \lambda^z (n = 1,2, \dots, N)$$

$$y^k_m \leq \sum_{z=1}^{Z} y^z_m \lambda^z (m = 1,2, \dots, N)$$

$$\sum_{z=1}^{Z} \lambda^z = 1$$

$$\lambda^z, s^-, s^+, x^k, y^k \geq 0$$

(1)

Papers and patents are created in different processes in the R&D stage. According to technology readiness level (TRL), papers could mainly be produced based on the results of the TRL stages 1 to 3, which are basic research stages and small experimental stages [30,31]. In contrast, patents are mainly published in the technology concretization and prototype development stage after the TRL Stage 3. Thus, it was possible to separate and analyze the publications to diagnose at which stage of technology development started to appear, and at which stage capabilities are mainly invested. Based on these analyses, it could be anticipated to predict the TRL of the technologies and their expected commercialization period.

In this study, the two BCC models presented in Equation (1) were constructed: Model I and Model II. In Model I, the R&D budget and the number of researchers in Table 2 were utilized as input. Three variables in the paper, (1) frequency (weighted by contribution), (2) impact factor, and (3) the number of citations, were utilized as outputs. This study defined the efficiency calculated in Model I as paper efficiency. In Model II, input was the same as shown in Model I. Four variables in patent, (1) frequency (weighted by contribution), (2) the number of backward citations, (3) the number of forward citations, and (4) the number of family patents was utilized as outputs. The efficiency calculated in Model II was defined as the patent efficiency.

Our portfolio analysis, which was illustrated in Figure 2, was based on a performance efficiency analysis in terms of patents and papers for R&D programs. Each R&D program was allocated the quadrant by calculating paper efficiency and patent efficiency. For the programs placed on the right side of the *y*-axis, their paper efficiency was max (1). Similarly, for the programs placed above the *x*-axis, their patent efficiency is 1. Therefore, for the program assigned to Quadrant I, both the patent and patent efficiencies of the R&D program were 1. As for the program placed in Quadrant III, both the paper and patent efficiencies of the R&D program were less than 1 [32]. In this study, the program with max paper (patent) efficiency could be assumed that they have the superior performance of papers (patents) compared to the inputs of budget and workforce. Therefore, an analysis of the potential performance of the R&D programs in Quadrant 1 was performed. Those

programs in Quadrants II and IV were also selected and analyzed if meaningful implications are anticipated [33].
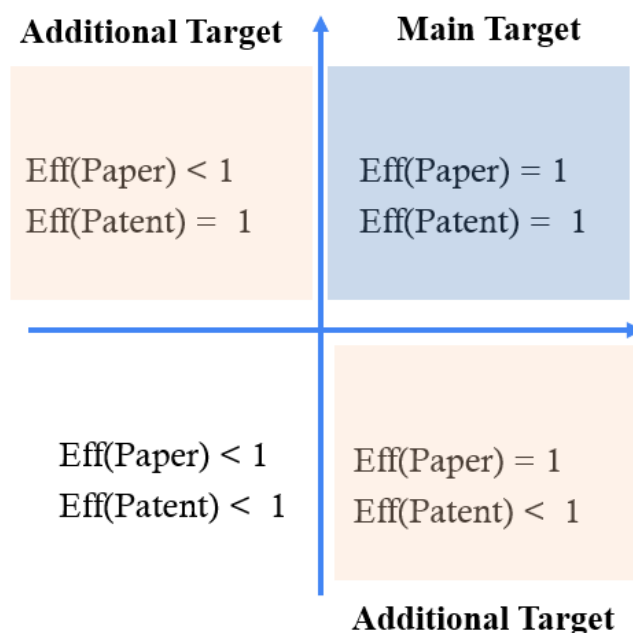

Figure 2. Quadrant of patent/paper performance efficiency.

2.3. Technological R&D Issue Suggestion

The R&D program performance efficiency analysis in Section 2.2 takes only the outcomes published before the analysis into account. In other words, these programs are evaluated solely on their outcomes up to this point. However, outcomes such as patents and papers take time to be realized. It was also considered that outcomes need to be analyzed in conjunction with the impact they have.

Hence, it is required to estimate and evaluate the potential impact of R&D programs. To do so, we introduced the diversity index, which indirectly measures the ripple effects expected of the program in the future. To utilize the index, patents and papers were categorized. We identified the characteristics of patents and papers as categories, and then analyzed the relevant areas to which these categories belong. With regard to patent potential performance, the IPC code was assigned to a category; regarding published papers, to which SCIMago classification the journal containing the paper belonged was determined [34].

First, a cosine dissimilarity measurement was performed to determine if derived performance was heterogeneous. This is calculated as Equation (2) below, where a_i is defined by the number of papers (patents) belonging to a certain category. This index increases as the number of papers shared between the two categories decreases. In the case of a high index, it indicates that the major effects of the two programs occur in different areas, so knowledge can be created across disparate fields, which is interpreted as impacting heterogeneous fields [35].

$$1 - \frac{\sum a_i \cdot b_i}{\sqrt{\sum a_i^2}\sqrt{\sum b_i^2}} \tag{2}$$

Second, the Shannon–Weaver index, an indicator of category balance and diversity, was used to analyze the diversity of derived performance [36]. This is calculated using Equation (3) below and p_i represents the proportion of the paper's (patent) that pertains to a particular category. This index increases when (1) the number of categories increases or (2) the proportions of categories are uniform. As a result, if the Shannon–Weaver index

is high, it can be concluded that these R&D programs and their outcomes would influence various areas in the future [37].

$$\sum p_i \ln\left(\frac{1}{p_i}\right) \tag{3}$$

Using the two aforementioned indexes, we selected several R&D programs that are considered to have high influence, and the program and sub-program, which is a sub-unit, were analyzed. Based on expert opinion, we analyzed (1) what technologies are primarily related to each R&D program and (2) what issues need to be addressed. As part of the analysis, we conducted a survey based on expert opinion about what technologies the subprograms within the programs are associated with [38]. Moreover, the program's descriptions on keywords and summary statements were used to identify any existing issues. With them, we attempted to identify technologies that could deliver both high potential performance and be influential and provided implications relating to those issues.

3. Results and Discussion

To analyze this, electric vehicles, which include automobiles that use batteries and fuel cells as power sources, were selected as a study area. This field is one of the intensive investment sectors that was developed in response to climate change around the world [39]. The government has designated the sector as one of its national growth engines and is investing a large number of resources to ensure its competitiveness [40]. As a result, it is a key field for Korea's future as well as significant investments in the world being made, making it suitable for the analytical purposes of this study [41].

3.1. Data Generation

We explored the electric vehicle-related programs on NTIS. Based on the search formula for five-year programs conducted during the period of 2014 to 2018, the first program list was compiled. The search formula was created by combining words related to electric vehicles, such as eco-friendly vehicle, battery-electric car, and hydrogen-fueled vehicle. Sub-programs were selected as the analysis target when sub-programs contained keywords in title, summary, or keywords fields. Approximately 2000 subprograms were collected this way [42].

In the next step, from the list of first-searched programs, a program with a strong connection to core technologies in the field of electric vehicles was selected. Through surveys of electric vehicle sector experts, a list of highly relevant programs was compiled. Hence, the second list of 800 subprograms was developed. A fair number of the sub-programs (approximately 410) were organized by the Ministry of Trade, Industry, and Energy. We also identified other subprograms administrated by the Ministries of Science and ICT (information and communications technology), SMEs and Startups, and Education.

Based on the final list of subprograms, input and output data were generated for DEA. For this purpose, we gathered outcomes as well as program information for 800 subprograms from NTIS. This process involved selecting a program that has published numerous papers and patents. A total of 17 programs (P01–P17) and their 415 subprograms were selected for final analysis.

The descriptive statistics of selected 17 programs were as shown in Table 5. For calculating the corresponding variable values, patent and paper information not available on NTIS was obtained from the patent (USPTO, WIPS) and paper (SCOPUS) databases. The descriptive statistics in Table 5 were calculated as the cumulative sum of the results generated by the R&D programs from 2014 to 2018. For example, if ten papers in a program were found as outcomes, the statistics in Table 4 were equivalent to the sum of the IF value of ten papers. In this study, the cumulative sum of input variables and output variables was calculated as each program, and the descriptive statistics of n = 17 were presented. Thus, in this study, Model I analyzed the paper efficiency of 17 data (R&D programs) through five variables (2 input variables and 3 output variables). Model II also analyzed

the patent efficiency of 17 data (R&D programs) through 6 variables (2 input and 4 output variables).

Table 5. Descriptive statistics for 17 R&D programs.

	Variable	# of Data	Average	Standard Deviation
Input	Workforce	17	273.48	289.36
	R&D budget (100 million/won)	17	145.55	292.34
Output (paper)	Frequency	17	22.16	23.75
	Citation	17	870.89	913.81
	Impact factor	17	205.29	215.64
Output (Patent)	Frequency	17	69.69	144.08
	Backward citation	17	383.41	820.10
	Forward citation	17	61.84	127.61
	Family patent	17	91.41	233.76

3.2. Program Performance Efficiency Evaluation

Based on the data generated in Section 3.1, DEA was performed. In this section, the BCC (Banker, Charnes, and Cooper) model was utilized. Since the scale of selected R&D programs was different, the influence of scale should be eliminated. Contrary to the CCR (Charnes, Cooper, and Rhodes) model, the BCC model, which was assumed "variable return to scale (VRS) efficiency" was suitable for this study as it considered both scale efficiency and pure technical efficiency [29]

To identify the performance of R&D programs, two DEA models were constructed to calculate the paper and patent efficiency mentioned in Section 2.2. Patents reflected the characteristics of technology development, and papers could reflect the characteristics of research. Therefore, to analyze the performance of the research and the technology development stage, respectively, two independent models (Model I with paper variables as output and Model II with patent attributes as output) were constructed.

In this study, the decision-making unit (DMU) was selected 17 R&D programs in Section 3.1. These DMUs formulated paper and patent outputs based on the inputs of the workforce and budget. We assumed that the level of the R&D programs' performance is proportional to the quantity and quality of outcomes (papers and patents), so the cumulative sum of outcomes in program units was utilized. In addition, output variables were defined through the analysis of variables such as citations of papers and patents, which are representative of each program's outcomes. Therefore, this study calculated the paper efficiency through the publication of papers and patents in units of programs, not individual paper units.

To determine the most efficient program in DEA, there must be a positive correlation between the input and output data. Correlation analysis was therefore conducted. The results are reported in Table 6.

Table 6. Correlation analysis of input-output variables.

	Input Variable / Output Variable	Workforce	R&D Budget
Paper	Frequency	0.206	0.255
	Citation	0.038	0.092

Patent	Impact factor	0.416 *	0.479 **
	Frequency	0.956 ***	0.983 ***
	Backward citation	0.951 ***	0.972 ***
	Forward citation	0.936 ***	0.970 ***
	Family patent	0.958 ***	0.980 ***

*: $p < 0.1$ **: $p < 0.05$, ***: $p < 0.01$ (two-tailed test).

The analysis revealed that there were positive correlations across all variables. Among the variables of paper analysis, a weak positive correlation was detected for some variables. In the case of patent, all output variables showed p-values below 0.01, indicating a strong positive correlation. In other words, the patent had a relatively high correlation between input and output in comparison to paper. Because these R&D programs mainly focused on steps involving technology concretization and prototype development (after TRL stage 3). Therefore, it was appropriate to define and analyze all of these variables as input and output variables in the DEA model.

Accordingly, the efficiency of 17 programs was analyzed through DEA. The efficiency of a program was measured in values between 0 and 1, with 1 representing the highest efficiency. A described in Section 2.2, the efficiency was assessed by separating output variables of the paper and patent and by establishing two DEA models. The results are summarized in Table 7 below. The R&D program name was not opened, however, Section 3.3 outlined key technologies and issues for each program.

Table 7. Result of DEA: 17 R&D programs.

Program	Paper		Patent		Quadrant
	CRS	VRS	CRS	VRS	
P01	1	**1**	1	**1**	**I**
P02	0.007	0.083	0.137	0.569	III
P03	0.274	**1**	0.631	**1**	**I**
P04	0.112	0.546	0.247	0.498	III
P05	1	**1**	1	**1**	**I**
P06	0.174	0.717	0.226	0.227	III
P07	0.009	0.443	0.197	**1**	II
P08	0.008	0.115	0.091	0.303	III
P09	0.005	0.072	0.098	0.410	III
P10	0.152	0.339	1	**1**	I
P11	0.004	0.180	0.092	0.483	III
P12	0.011	0.091	0.127	0.386	III
P13	0.173	0.900	1	**1**	II
P14	0.005	0.744	0.207	**1**	II
P15	0.136	0.537	0.624	0.839	III
P16	0.045	0.106	0.880	0.980	III
P17	0.679	**1**	1	**1**	**I**

The efficiency of DEA was calculated using the VRS rather than CRS value. The VRS assumption was applied to the DEA efficiency calculation since the nature of national research programs makes R&D input changeable and it is preferable to maximize efficiency.

Table 7 shows that four programs had performance efficiency in terms of paper analysis equal to 1 (P01, P03, P05, and P17). Further, for programs with performance efficiency not equal to 1, the efficiency tended to be lower than that of patents, thus, the deviation of

performance efficiency in terms of paper analysis between R&D programs was relatively high. In fact, there were some programs in which the performance efficiency in terms of paper analysis was low since these programs were conducted after a prototype development for electric vehicle technology. Therefore, patent efficiency was more considered important in this study. As shown in Table 7, eight programs with maximum performance efficiency were found for patent analysis. Among them, four (P01, P03, P05, and P17) showed maximum performance efficiency for paper analysis as well. Therefore, four programs were placed in Quadrant I, four in Quadrant II, and seven in Quadrant III.

Typically, programs located in Quadrant III had performance efficiency in terms of paper below average. In particular, the correlation analysis between the efficiency of patent and paper revealed an R-square value of 0.570. In this case, the corresponding test statistic had significance at a level of 0.05 because the p-value was less than 0.05. Thus, it was demonstrated that the performance efficiency of patent and paper had a fairly high correlation. To further analyze this, raw data were assessed. As a result, the performance efficiency in terms of the paper analysis was low for programs involving heavy manpower and research costs, but performance efficiency in cases related to patents was quite high. It can be explained by the fact that many of the large-scale programs are programs of the Ministry of Trade, Industry, and Energy, which conducts research on technologies. Thus, our study identified four programs where singularities existed, P03, P07, P14, and P17. In this regard, sub-programs belonging to these four programs were examined according to the technology they are associated with. First, technology has been categorized as a major investment development area being introduced by the Industrial Technology R&D Strategy, which was published by the Korea Ministry of Trade, Industry, and Energy. Within the technology, except for the common part technology, five technology groups were made: driving and power system (T01), conditioning and thermal management system (T02), battery and electric charging system (T03), fuel cell system (T04), and hydrogen charging system (T05) the TRL stage 5.

Accordingly, we concluded that the patent performance is more important given the nature of the Ministry of Trade, Industry, and Energy's R&D programs. We, therefore, analyzed a total of eight programs in Section 3.3, including four in Quadrant I and four in Quadrant II.

3.3. Technological R&D Issue Suggestion

This section calculated the diversity index for eight programs selected in Section 3.2. From the content of the previous study, it was concluded that the greater the diversity and heterogeneity, the more likely the program's outcomes will be utilized in a variety of places in the future or have the potential to improve. As indicated in Section 2.3, the diversity index for each program was calculated by assigning the outcomes (papers and patents) to categories. These results could be shown as in Table 8.

Table 8. Result of diversity index: selected eight R&D programs.

Quadrant	Program	Paper		Patent	
		Shannon–Weaver	Cosine Distance	Shannon–Weaver	Cosine Distance
Quadrant I	P01	1.763	0.650	1.321	0.162
	P03	2.659	0.740	1.900	0.575
	P05	1.512	0.643	-	0.542
	P17	2.591	0.641	2.008	0.452
Quadrant II	P07	1.931	0.660	2.134	0.554
	P10	2.322	0.527	1.081	0.111
	P13	2.091	0.696	2.174	0.486
	P14	2.699	0.632	2.784	0.578

Table 8 is organized by dividing into the diversity of paper and patent. In the case of paper analysis, Shannon–Weaver values were high in P03, P14, and P17. In the basic research programs P03 and P17, various topics were explored. It was particularly noteworthy that in P03, a large number of papers were in the pure sciences such as physics and chemistry, rather than only disciplines that are related to engineering such as mechanics or information technology. As a result, fields were more evenly distributed. In addition, P14, which deals with technology across electric vehicles, carried out research related to key technologies related to electric vehicles and platforms linking to these technologies. Therefore, engineering fields such as energy and fuel, transportation science, mechanical engineering, and chemical engineering accounted for a high proportion.

According to cosine similarity, each program had approximately similar values. The value of P03 was found to be relatively high. The reason is that in the basic research phase, there are numerous fields of academic study that can be applied to electric vehicles. Not only physics and information technology is needed to produce electric vehicles, but also convergence technologies among many fields. Thus, many papers belonging to multidisciplinary fields were found, so their heterogeneity was relatively evident since these papers were more rooted in pure science such as physics, chemistry, and material science than the other programs.

In the case of the patent, Shannon–Weaver values were not calculated for P03 since all derived patents were assigned the same IPC (International Patent Classification) code. Shannon–Weaver index in patents was calculated to be the highest in P14 for similar reasons as with papers. In fact, many of the IPC codes were allocated to b60*, which represents ordinary vehicles, because the many patents were related to vehicle platforms and their components. Several programs have remarkable topics of patents. For instance, the case of P07, in contrast to other programs, had a relatively high heterogeneity due to the many patents related to materials such as C01* and C02*.

Thus, our study identified four programs where singularities existed, P03, P07, P14, and P17. According to opinions from experts, all of the selected R&D programs were considered as important, which could play a validation role in this study. They insisted that the selected four programs have the core promoted technologies. These technologies contributed to improving the competitiveness of electric vehicles, for instance, consumer acceptance such as operating mileage and charging time of vehicles. In this regard, subprograms belonging to these four programs were examined according to the technology they are associated with. First, technology has been categorized as a major investment development area being introduced by the Industrial Technology R&D Strategy, which was published by the Korea Ministry of Trade, Industry, and Energy. Based on the analysis results and related expert opinion, this study determined five technology areas as target areas except for the common part technology: driving and power system (T01), conditioning and thermal management system (T02), battery and electric charging system (T03), fuel cell system (T04), and hydrogen charging system (T05).

The five main technologies were divided in this manner so that potential issues could be identified from the program. Through analyzing the keywords and summaries of each program, a number of issues were derived. The extracted keywords and their related issues are listed in Table 9.

Table 9. Result of keyword analysis: selected four R&D programs.

Field	Program	Issue (Keyword)
T01	P07	Driving motor (heavy rare earth, magnet, coercivity)
	P14	Driving power distribution (power distribution control, powertrain, electromagnetic transmission)
T02	P03	Material (heat-exchanger, phase change heat transfer)
		Driving heat management (driving performance, battery thermal management)

	P03	Energy density (electroactive material, anode material, separator)
T03	P14	Battery management system (cylindrical cells, high safety, high voltage battery)
T04	P07	Hydrogen storage system (solid hydrogen storage system, metal hydride)
	P17	Fuel cell (DME (dimethyl ether) hydrogen reformer, micro-sensor) Catalyst (reformer catalyst)
T05	P14	Hydrogen supply (hydraulic piston type compressor, tube trailer) Hydrogen storage (balance of tank, regulator, receptacle)

In the driving and power system (T01) field, issues related to driving motors, power transmission devices, and power conversion devices were identified. Regarding P07, the main issue concerned driving motors. Specifically, the research was focused on replacing magnetic materials included in motors with materials other than rare earth elements. Regarding motors, it mainly addresses the issue of using eco-friendly materials rather than improving technical specifications. The reason is that rare earth elements are responsible for pollution of the environment, while at the same time, their supply and demand are affected by the political issues of China. Next, P14's issues were related to platforms and devices that transmit power. Technologies of devices and systems for controlling power transmission devices were also included. In P14, a number of issues were identified for platforms that were integrated with power generators (motors) or power conversion devices, as opposed simply to dealing with power transmission devices. It could be said that electric vehicle technologies should be characterized by environmental protection and efficient energy usage. Thus, it would be important to develop novel eco-friendly components and their efficient platforms rather than the existing traditional platforms based on ESG management.

For the conditioning and thermal management system (T02) field, the issue of thermal management was mainly identified. There was information on topics such as heat exchangers used in automobiles or lightening and solidification of materials. Additionally, heat accumulation during driving was an issue, particularly regarding the area of battery and power generation. Due to the fact that internal and external temperatures directly affect the mileage and other characteristics of electric vehicles, maintaining the constant temperature of components was a major challenge for the program, as compared with internal combustion engine vehicles. One of the important issues for batteries is thermal management to maintain the driving performance of vehicles regardless of external temperature. Diverse components and related composites were anticipated to improve the range of use, which would be critical commercial factors for electric vehicles.

For battery and electric charging system (T03) field, issues related to batteries were primarily identified. A major issue identified in P03 was battery energy capacity. Various studies have been conducted on improving the efficiency of cathodes and electrolytes. Generally, this kind of research involved introducing alternative substances such as sulfur or suggesting new substances. Compare with other vehicles, electric vehicles tended to move short and mid-distances, but it is important to increase mileage in vehicles to improve the users' convenience. In this way, this program aimed to increase consumer acceptance of electric vehicles. In P14, issues related to battery packs were identified. The focus of the research was large to improve not only safety against thermal problems but also energy efficiency through the integration of battery packs and power conversion devices. To develop a safer and more efficient battery, many researchers attempt to alter the characteristics or layout of the battery pack. Future major areas of battery R&D programs would increase mileage and develop high-performance batteries with both 1) novel anode and cathode materials or 2) a novel process of configuration optimization.

In terms of fuel cell system (T04), the major issues were hydrogen storage and stacks of fuel cells in vehicles. P07 identified the main issue related to hydrogen storage devices.

Such a device is designed to store high-temperature, high-pressure hydrogen, and to use it to power a vehicle. The research was conducted to develop storage materials that could accommodate hydrogen at high temperatures (100 °C). The objective was to increase the driving distance of hydrogen fuel cell vehicles to improve the efficiency of hydrogen. In P17, the issue of the fuel cell stack was identified. The main task was to develop control technology in real-time using microsensors and increase the hydrogen production rate of fuel cells. In addition, catalysts were developed to increase the efficiency of operation. These developments mainly responded to the issue of improving life expectancy for fuel cells and reducing the impact of the external environment on the performance, and future influential R&D direction would be continued.

In the hydrogen charging system (TO5) field, hydrogen supply and storage have been identified as major issues. As for the supply, research was conducted on increasing the capacity of trailers that carry hydrogen or improving the efficiency of hydrogen compression. In the case of storage, research was conducted on external hydrogen storage devices that maintained high pressure and prevented hydrogen exposure. By doing this, the program was challenged by the issue of safely transporting hydrogen to the charging station and storing it safely. The future direction of R&D is that hydrogen charging technology is being developed in various ways. For instance, beyond high-pressure hydrogen charging technologies, many researchers attempt to utilize metal hydride to store the hydrogen by combining a metal element with hydrogen affinity and a transition metal between crystal lattices.

4. Conclusions

Through the NTIS database, which contains national R&D information of South Korea, we conducted DEA using paper and patent information. This process involved a performance analysis by subprogram unit. Furthermore, the performance efficiency of each R&D program was calculated by utilizing various indexes based on two sources of the flow of knowledge, papers and patents. The study also identified several R&D programs with outstanding paper and patent performance based on the results. Additionally, it included a post-analysis of potential outcomes. Based on the diversity index, this process was constructed of a list of programs that were expected to produce significant accomplishments. To eventually establish the implications, a number of issues in the current technology area were derived from the abstracts and keywords for each sub-program. We applied the proposed framework to the electric vehicle sector, an industry that Korea considers a key growth engine. We examined 17 R&D programs pertaining to electric vehicles. Following both quantitative and qualitative analysis, we then selected four programs with high performance and potential for future development, thereby deriving various hot issues and implications related to electric vehicles.

This study offers two research contributions. First, an integrated framework was presented to analyze the outcome of the R&D program. In many DEA approaches, input-output relationships were analyzed using only quantitative variables. However, there are many unrealized outcomes of R&D. Current quantitative variables such as frequency and citation merely could reflect issues directly, so it is necessary to conduct additional research to arrive at a meaningful conclusion. For this reason, we also conducted qualitative analysis by actively utilizing textual data such as summary and keywords as well as additional information such as the category of papers/patents. In this process, we evaluated the influence of the program in terms of qualitative data. As part of assessing the impact or potential performance of the program, we identified technical issues based on the sub-program's content registered in NTIS, which allowed us to make a more informed conclusion. In fact, based on the case of electric vehicles, which is Korea's key R&D investment field, we produced implications concerning five different technology types, from driving and power systems to hydrogen charging systems.

Second, we even considered potential performance through the diversity index. In the previous DEA approach, the output consisted only of completed performance, such

as patents and papers. However, the characteristics of R&D cause a time-lag in performance. Because of this, it is difficult to evaluate R&D programs accurately. There are various studies that address this issue of time-lag, but it is difficult to say whether outcomes will happen based on past trends. The R&D performance must also consider impacts in the future, not just those of the present. Different areas of research have encountered different timings in terms of realizing, so we sought to identify potential performance not yet realized. We used the diversity index for our research as it allows for identifying areas that could have a significant influence on knowledge exchange. As a result, we concluded that this index could contribute to constructing a useful indicator in similar future developmental studies.

However, this research has some limitations. First, the data themselves have limitations. Due to the limited number of NTIS target programs (programs), and to the fact that the search terms were limited to electric vehicles, it was difficult to collect data concentrated in a specific year or to obtain voluminous data continuously. Since the electric vehicle field has only recently received attention, it has been difficult to obtain past data and analyze relevant programs. The period of selected R&D programs was rarely matched: the start and end years of each R&D program. For this reason, it was hard to conduct the dynamic analysis including time-series analysis. The time-lag problem is also an important factor to influence the quantitative statistics of R&D outcome analysis. Therefore, we conducted a single-period approach with the CCR model. As a result, future studies will validate efficiency and usability by extending the analyzable time range and performing analyses on a wider variety of program and sub-program sets.

In this process, there is a need to establish a systematic validation process. In this study, the appropriateness of the selected programs was judged based on expert opinions after the programs selected by the suggested framework. In addition, all of these programs tended to increase the amount and efficiency of performance, and thus the corresponding framework attained reasonable results. It needed to supplement a more systematic validation process as an area to be improved in our future studies. Future studies will validate the efficiency and usability by extending the analyzable time range and performing an analysis on a wider variety of program and sub-program sets. In particular, programs that efficiently generated results by setting the period in several units will be individually selected. Additionally, both quantitative methods such as decision tree and qualitative methods such as keyword analysis will all be used to define the characteristics of the excellent R&D program. These results will be used as a basis for systematizing the validation process.

Another problem exists with relying on partial performance derived from papers and patents. Thus, future studies will utilize wider indicators, such as the number of cases related to technology transfer or mergers and acquisitions. Patents and papers are important output indicators, but in the context of national R&D, technical and professional performance should also be measured, not just research performance. We intend to conduct an advanced analysis in the future by referring to the number of technology transfers as well as several other examples. Additionally, we plan to conduct further analysis reflecting the performance still to be realized and research related to this topic. As part of the preparation for this research plan, various analyzable indicators will be developed based on NTIS data to be used in future research.

References

1. Kim, S.S.; Choi, Y.S. The innovative platform programme in South Korea: Economic policies in innovation-driven growth. *Foresight STI Gov.* **2019**, *13*, 13–22. https://doi.org/10.17323/2500-2597.2019.3.13.22.
2. Kim, G.-P.; Kang, G.S.; Kim, J.; Oh, T.; Lee, H.J.; Son, W. Innovative Growth Strategy in the US, Europe and Japan. *KIEP* **2020**, *10*, 1–7. https://hdl.handle.net/11540/12153.
3. Lee, J.-I. An econometric model for evaluating success/failure of an R&D project: The case of Korea. *Int. J. Inf. Technol. Manag.* **2011**, *10*, 69–79. https://doi.org/10.1504/IJITM.2011.037763.
4. Hwang, B.Y.; Park, S.H.; Kim, D.C. Efficiency of Public R&D Management Agency and Its Improvement toward Open Innovation. *J. Open Innov.* **2021**, *7*, 200. https://doi.org/10.3390/joitmc7030200.
5. Xu, P.; Zhang, M.; Gui, M. How R&D Financial Subsidies, Regional R&D Input, and Intellectual Property Protection Affect the Sustainable Patent Output of SMEs: Evidence from China. *Sustainability* **2020**, *12*, 1207. https://doi.org/10.3390/su12031207.
6. Pishdar, M.; Shakib, M.D.; Antucheviciene, J.; Vilkonis, A. Interval Type-2 Fuzzy Super SBM Network DEA for Assessing Sustainability Performance of Third-Party Logistics Service Providers Considering Circular Economy Strategies in the Era of Industry 4.0. *Sustainability* **2021**, *13*, 6497. https://doi.org/10.3390/su13116497.
7. Drew, J.; Kortt, M.; Dollery, B. What determines efficiency in local government? A DEA analysis of NSW local government. *Econ. Pap.* **2015**, *34*, 243–256. https://doi.org/10.1111/1759-3441.12118.
8. Xia, Q.; Jin, M.; Wu, H.; Yang, C. A DEA-based decision framework to determine the subsidy rate of emission reduction for local government. *J. Clean Prod.* **2018**, *202*, 846–852. https://doi.org/10.1016/j.jclepro.2018.08.171.
9. Park, J.H.; Shin, K. Efficiency of government-sponsored R&D projects: A metafrontier DEA approach. *Sustainability* **2018**, *10*, 2316. https://doi.org/10.3390/su10072316.
10. Luna, D.E.; Gil-Garcia, J.R.; Luna-Reyes, L.F.; Sandoval-Almazan, R.; Duarte-Valle, A. Improving the performance assessment of government web portals: A proposal using data envelopment analysis (DEA). *Inf. Polity* **2013**, *18*, 169–187. https://doi.org/10.5555/2657015.2657021.
11. Andersen, P.; Petersen, N.C. A procedure for ranking efficient units in data envelopment analysis. *Manag. Sci.* **1993**, *39*, 1261–1264.
12. Banker, R.D. Estimating most productive scale size using data envelopment analysis. *Eur. J. Oper. Res.* **1984**, *17*, 35–44. https://doi.org/10.1016/0377-2217(84)90006-7.
13. Gattoufi, S.; Oral, M.; Reisman, A. Data envelopment analysis literature: A bibliography update (1951–2001). *J. Socio-Econ. Plan. Sci.* **2004**, *38*, 159–229. https://doi.org/10.1016/S0038-0121(03)00023-5.
14. Rao, C.R. Diversity and dissimilarity coefficients: A unified approach. *Theor. Popul. Biol.* **1982**, *21*, 24–43. https://doi.org/10.1016/0040-5809(82)90004-1.
15. Van Rijnsoever, F.J.; van Den Berg, J.; Koch, J.; Hekkert, M.P. Smart innovation policy: How network position and project composition affect the diversity of an emerging technology. *Res. Policy* **2015**, *44*, 1094–1107. https://doi.org/10.1016/j.respol.2014.12.004.
16. Stirling, A. A general framework for analysing diversity in science, technology and society. *J. R. Soc. Interface* **2007**, *4*, 707–719. https://doi.org/10.1098/rsif.2007.0213.
17. Bordons, M.; Bravo, C.; Barrigón, S. Time-tracking of the research profile of a drug using bibliometric tools. *J. Am. Soc. Inf. Sci.* **2004**, *55*, 445–461. https://doi.org/10.1002/asi.10397.
18. Jeong, J.Y.; Kang, I.; Choi, K.S.; Lee, B.-H. Network analysis on green technology in national research and development projects in Korea. *Sustainability* **2018**, *10*, 1043. https://doi.org/10.3390/su10041043.
19. Lee, J.Y.; Kumari, R.; Jeong, J.Y.; Kim, T.-H.; Lee, B.-H. Knowledge Discovering on Graphene Green Technology by Text Mining in National R&D Projects in South Korea. *Sustainability* **2020**, *12*, 9857. https://doi.org/10.3390/su12239857.
20. Lee, H.Y.; Park, Y.T. An international comparison of R&D efficiency: DEA approach. *Asian J. Technol. Innov.* **2005**, *13*, 207–222. https://doi.org/10.1080/19761597.2005.9668614.

21. Kocher, M.G.; Luptacik, M.; Sutter, M. Measuring productivity of research in economics: A cross-country study using DEA. *Socio-Econ. Plan. Sci.* **2006**, *40*, 314–332.

22. Cooper, W.W.; Seiford, L.M.; Tone, K. *Data Envelopment Analysis: A Comprehensive Text with Models, Applications References and DEA-Slover Software*; Springer: Berlin/Heidelberg, Germany, 2007.

23. Wang, E.C.; Huang, W. Relative efficiency of R&D activities: A cross-country study accounting for environmental factors in the DEA approach. *Res. Policy* **2007**, *36*, 260–273.

24. Cullman, A.; Schmidt-Ehmcke, J.; Zloczysti, P. Innovation, R&D efficiency and the impact of the regulatory environment: A two-stage semi-parametric DEA approach. Econstor 2009; DIW Discussion No.883. Available online: https://www.econstor.eu/bitstream/10419/29763/1/599519827.pdf (accessed on 23 August 2021)

25. Zhao, H. The Empirical Study on R&D Efficiency of High-Tech Industry in China. *Int. J. Trade Econ. Financ.* **2015**, *6*, 49.

26. Zhou, Y.; Dong, F.; Liu, Y.; Li, Z.; Du, J.; Zhang, L. Forecasting emerging technologies using data augmentation and deep learning. *Scientometrics* **2020**, *123*, 1–29. https://doi.org/10.1007/s11192-020-03351-6.

27. Toloo, M.; Nalchigar, S. A new integrated DEA model for finding most BCC-efficient DMU. *Appl. Math. Model.* **2009**, *33*, 597–604. https://doi.org/10.1016/j.apm.2008.02.001.

28. Butler, T.W.; Li, L. The utility of returns to scale in DEA programming: An analysis of Michigan rural hospitals. *Eur. J. Oper. Res.* **2005**, *161*, 469–477. https://doi.org/10.1016/j.ejor.2003.09.019.

29. Banker, R.D.; Charnes, A.; Cooper, W.W. Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Manag. Sci.* **1984**, *30*, 1078–1092. https://doi.org/10.1287/mnsc.30.9.1078.

30. Mankins, J.C. Technology readiness levels. *White Pap.* **1995**, *1995*, 6.

31. Jou, G.T.; Yuan, B.J. Utilizing a novel approach at the fuzzy front-end of new product development: A case study in a flexible fabric supercapacitor. *Sustainability* **2016**, *8*, 740. https://doi.org/10.3390/su8080740.

32. Fan, J.-L.; Zhang, X.; Zhang, J.; Peng, S. Efficiency evaluation of CO_2 utilization technologies in China: A super-efficiency DEA analysis based on expert survey. *J. CO_2 Util.* **2015**, *11*, 54–62. https://doi.org/10.1016/j.jcou.2015.01.004.

33. Rusydiana, A.S.; Sanrego, Y.D. Measuring the performance of Islamic banking in Indonesia: An application of Maslahah-efficiency quadrant (MEQ). *J. Islamic Monet. Econ. Financ.* **2018**, *3*, 79–98. https://doi.org/10.21098/jimf.v3i0.909.

34. Falagas, M.E.; Kouranos, V.D.; Arencibia-Jorge, R.; Karageorgopoulos, D.E. Comparison of SCImago journal rank indicator with journal impact factor. *FASEB J.* **2008**, *22*, 2623–2628. https://doi.org/10.1096/fj.08-107938.

35. Li, B.; Han, L. Distance Weighted Cosine Similarity Measure for Text Classification. In Proceedings of the International Conference on Intelligent Data Engineering and Automated Learning, Hefei, China, 20–23 October 2013; pp. 611–618.

36. Spellerberg, I.F.; Fedor, P.J. A tribute to Claude Shannon (1916–2001) and a plea for more rigorous use of species richness, species diversity and the 'Shannon–Wiener'Index. *Glob. Ecol. Biogeogr.* **2003**, *12*, 177–179. https://doi.org/10.1046/j.1466-822X.2003.00015.x.

37. Al-Fedaghi, S. A conceptual foundation for the Shannon-Weaver model of communication. *Int. J. Soft Comput.* **2012**, *7*, 12–19. https://doi.org/10.3923/ijscomp.2012.12.19.

38. Magnussen, S.; Boyle, T. Estimating sample size for inference about the Shannon-Weaver and the Simpson indices of species diversity. *For. Ecol. Manag.* **1995**, *78*, 71–84. https://doi.org/10.1016/0378-1127(95)03596-1.

39. Choi, H.; Koo, Y. Effectiveness of battery electric vehicle promotion on particulate matter emissions reduction. *Transp. Res. Part D Transp. Environ.* **2021**, *93*, 102758. https://doi.org/10.1016/j.trd.2021.102758.

40. Kwon, L.-N.; Park, J.-H.; Moon, Y.-H.; Lee, B.; Shin, Y.; Kim, Y.-K. Weak signal detecting of industry convergence using information of products and services of global listed companies-focusing on growth engine industry in South Korea. *J. Open Innov. Technol. Mark. Complex.* **2018**, *4*, 10. https://doi.org/10.1186/s40852-018-0083-6.

41. Ki, J. A Comparative Analysis of Autonomous Vehicle Policies among Korea, Japan, and France. 2020. Available online: https://hal.archives-ouvertes.fr/hal-02562482/document (accessed on 12 September 2021)

42. Yuan, K.; Song, Y.; Shao, Y.; Sun, C.; Wu, Z. A charging strategy with the price stimulus considering the queue of charging station and EV fast charging demand. *Energy Procedia* **2018**, *145*, 400–405. https://doi.org/10.1016/j.egypro.2018.04.046.