



## Article

# An Exploration of Prediction Performance Based on Projection Pursuit Regression in Conjunction with Data Envelopment Analysis: A Comparison with Artificial Neural Networks and Support Vector Regression

Xiaohong Yu <sup>1,2</sup>  and Wengao Lou <sup>3,\*</sup> 

<sup>1</sup> College of Humanities and Law, Shanghai Business School, Shanghai 200235, China; yuxiaohong@nuaa.edu.cn

<sup>2</sup> College of Economics and Management, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

<sup>3</sup> School of Information Management, Shanghai Lixin University of Accounting and Finance, Shanghai 201209, China

\* Correspondence: louwengao@lixin.edu.cn

**Abstract:** Data envelopment analysis (DEA) is a leading approach in performance analysis and discovering newer benchmarks, and the traditional DEA models cannot forecast the future efficiency of decision-making units (DMUs). Machine learning, such as the artificial neural networks (ANNs), support vector machine/regression (SVM/SVR), projection pursuit regression (PPR), etc., have been viewed as beneficial for managers in predicting system behaviors. PPR is especially suitable for small and non-normal distribution samples, the usual cases in DEA analysis. This paper integrates DEA and PPR to cover the shortcomings we faced while using DEA and DEA-BPNN, DEA-SVR, etc. This study explores the advantages of combining these complementary methods into an integrated performance measurement and prediction model. Firstly, the DEA approach is used to evaluate and rank the efficiency of DMUs. Secondly, we establish two DEA-PPR combined models to describe the DEA efficiency scores (also called the production function) and the DEA-efficient frontier function. The first combined model's input variables are input-output indicators in the DEA model, and the output variable is the DEA efficiency. In the second model, its input variables are input or output indicators in the DEA model, and the output variable is the optimal input indicator for input-oriented DEA or the output indicator for output-oriented DEA. We conducted positive research on two examples with actual data and virtual small, medium-sized, and large samples. Compared with the DEA-BPNN and DEA-SVR models, the results show that the DEA-PPR combined model has more vital global optimization ability, better convergence, higher accuracy, and a simple topology. The DEA-PPR model can obtain robust results for both small and large cases. The DEA-BPNN and DEA-SVR models cannot obtain robust results for small and medium-sized samples due to overfitting. For large samples, the DEA-PPR model outperforms DEA-BPNN, DEA-SVR, etc. The DEA-PPR combined model possesses better suitability, applicability, and reliability than the DEA-BPNN model, the DEA-SVR model, etc.

**Keywords:** projection pursuit regression (PPR); data envelopment analysis (DEA); artificial neural networks (ANNs); support vector machine/regression (SVM/SVR); efficiency measure; decision-making units (DMUs); combined model

**MSC:** 68T07; 90B50; 68V99



**Citation:** Yu, X.; Lou, W. An Exploration of Prediction Performance Based on Projection Pursuit Regression in Conjunction with Data Envelopment Analysis: A Comparison with Artificial Neural Networks and Support Vector Regression. *Mathematics* **2023**, *11*, 4775. <https://doi.org/10.3390/math11234775>

Academic Editors: Marcio Basilio, Valdecy Pereira and Marcos Dos Santos

Received: 1 November 2023

Revised: 17 November 2023

Accepted: 22 November 2023

Published: 26 November 2023



**Copyright:** © 2023 by the authors. 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 (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The measurement of the efficiency or performance of homogeneous production or services, namely decision-making units (DMUs), has long been a hotspot or subject of

interest for researchers in many fields. Since Charnes et al. [1] proposed the radial-based data envelopment analysis (called DEA, DEA-CCR, or CCR) under input-oriented cases with a constant return-to-scale (CRS), DEA is widely and successfully used to measure the performance or efficiency of DMUs [2,3]. DEA is a nonparametric basis for evaluating the multiple input–output efficiency of different DMUs. To better meet various situations and conditions, a variety of DEA models have been proposed successively [4]. Distinct calculation principles, a simple structure, fewer input–output indicators, low-cost data collection, and easy application characterize them.

Meanwhile, the DEA model also has some shortcomings. For example, DEA is very sensitive to data noise and cannot be used for prediction [5]. When new DMUs are added, they must be remodeled, and the efficiency of the original DMUs is changed. At the same time, the shortcomings of these DEA models happen to be the advantage of another nonparametric modeling approach—the artificial neural network (ANN) [6]. Furthermore, DEA models treat DMUs as a black box; inputs enter and outputs exit without considering the intervening steps [7]. Therefore, Athanassopoulos and Curram [8] proposed to combine the DEA and BPNN models for efficiency analysis and conducted empirical research on the operating efficiency of the Cobb–Douglas production function and bank; the results show that the DEA model has better performance in simulating production functions, and the BPNN model is equivalent to DEA in the efficiency ranking of DMUs; they conducted in-depth research on how to combine DEA and BPNN better to give full play to their respective advantages. Since then, extensive theoretical and applied research has been conducted on the combination of ANN and DEA [4,9–16]. Additionally, other machine learning methods (such as SVR and SVM) are also combined with DEA in applied and comparative studies [4,11,17–20]. Some scholars have optimal experimental research on the integrated model of RSM (response surface methodology) and ANN-DEA. So far, extensive and in-depth research has been conducted on combined models such as ANN-DEA, achieving good results. However, there are still areas to be improved and perfected. First, for small and medium-sized samples in DEA, the input–output data and the evaluation results (efficiency scores) usually do not obey normal distribution, and it is not easy to guarantee the generalization ability and practical value of ANN or SVR combined models. Although some studies divide the sample data into training and test subsets [11,13,21], all studies do not use validation subsets to monitor the training process to avoid over-training. In training the ANN-DEA combined model, we cannot judge whether over-training has occurred, so it is not easy to guarantee its generalization ability and practical value. Second, to establish SVR (SVM) and ANN, the principles of determining input and output variables in different studies are often inconsistent, resulting in unconvincing modeling results. Some scholars [11,13,21–23] took the input–output indicators of the DEA model as the input variables of the ANN model. The DEA model evaluated the efficiency scores as the output variables of the ANN for modeling, but did not discuss how to reasonably determine the number of hidden-layer neurons and avoid over-training. Farahmand et al. [24] and Zhu et al. [11] used the input–output indicators of the DEA model as the input variables of SVR and the efficiency value evaluated via DEA as the output variables of SVR. Through empirical research, Zhu et al. [11] believed that the performance of the ANN-DEA model was better than that of SVR-DEA. Zhong et al. [13] argued that the performance of the ANN-DEA model was better than that of SVR-DEA, random forest (RF-) DEA, gradient boosting regressor (GBR-) DEA, and other models. Bose [15] established an input-oriented CCR-DEA model, took values uniformly within the output-indicator range of DEA-efficient samples, generated a certain number of modeling virtual samples, and used the output indicators of DEA as the input variables of the ANN model and the input indicators of the DEA model as the output variables of the ANN model; without a test subset, the DEA-efficient frontier ANN model was established.

It can be seen from those mentioned above that although the academic community has carried out extensive research on the combined prediction model of ANN-DEA (including SVR-, RF-, etc.), there are still many problems to be further researched, especially on the

efficiency of small and medium-sized samples that do not obey normal distribution; a better model should be established with a better generalization ability and robustness. On the other hand, projection pursuit regression (PPR) is particularly applied to modeling with nonlinear, small, and medium-sized samples that do not obey normal distribution [25–28]. Therefore, this paper introduces PPR into efficiency research to establish a DEA-PPR combined model to obtain more reliable and robust results. It expands not only the efficiency evaluation method but also the application field of the PPR model.

This study fulfills a practical need and improves benchmarking and decision-making processes by exploring an innovative performance measurement and prediction framework using DEA-PPR. The proposed combined model utilizes DEA as a preprocessor, and the subsequent PPR model conducts prediction tasks for the best performance output for each DMU. This paper provides insight into two examples to build the DEA-PPR combined models and compare the performance of different combined models such as DEA-BPNN, DEA-SVR, etc.

In summary, the primary purpose and motivations for this research paper are fivefold:

- (1) To present a new DEA-PPR combined model for performance measurement and prediction, thus bridging the research gap through methodological advancement;
- (2) To provide empirical support for the proposed model using two datasets through streamlining sequential processes of DEA measurement and DEA-PPR prediction;
- (3) To provide a way of thinking to improve managerial efficiency and enhance administrative flexibility in selecting actionable options from theoretical and practically feasible alternatives and potential progress monitoring via the DEA-PPR combined model;
- (4) To discuss the advantages and disadvantages of machine learning, such as BPNN, SVR, PPR, RF, etc.;
- (5) To put forward the basic principles and some matters needing attention for building machine learning, such as BPNN, PPR, SVR, etc.

The rest of this paper is organized as follows: Section 2 is a review of the literature on DEA combined models and PPR; Section 3 is about CCR-DEA and PPR modeling principles; Section 4 is on the empirical research on the production function of combined models of DEA and the DEA-efficient frontier function; Section 5 is the results and discussion; and Section 6 describes the limitations and future research fields.

## 2. Literature Review

### 2.1. DEA and Its Combined Models

DEA, proposed in 1978 [1], provides an effective tool for measuring the efficiency of the homogeneous DMUs of non-profit (e.g., schools, hospitals, and local authorities) and for-profit (e.g., banks, public houses, corporates, listed companies) organizations or persons (e.g., scholars, employees). Since then, many new DEA models have been introduced, such as BCC, SBM, super-efficiency (SE-) DEA, SE-SBM, (SE-)SBM with undesirable output, dynamic DEA, and the window analysis technique [2–4]. DEA models with multiple input and output indicators have been widely used to assess DMUs' efficiency.

On the other hand, the DEA models used for performance assessment have inherent disadvantages. First, DEA models are easily affected by data noise. Second, when new DMUs are added, we must re-establish the DEA, and the efficiency of all the original DMUs must be changed. It is not easy to compare the efficiency of different datasets. To overcome the above disadvantages, Athanassopoulos and Curram [8] thought that DEA and artificial neural networks (ANNs) were nonparametric methods, in that no assumption is made, and the inputs and outputs are used to describe an operational process. They started work to combine DEA with BPNN and studied the known Cobb–Douglas production function with two inputs and one output (simulated data). A series of 16 datasets were artificially generated and tested using the ANN-DEA combined models. The number of nodes in the input and output layers of the ANN model was equal to the DEA model. The results showed that the DEA performed more satisfactorily as a tool for estimating empirical production functions. Furthermore, they researched 250 commercial bank branches, using four inputs

and three outputs. Take the network (topology) structure of the BPNN model as 4-10-3 (that is, the input layer, hidden layer, and output layer have four, ten, and three neurons, respectively, and the number of network connection weights is  $5 \times 10 + 11 = 61$ ); the input-layer and output-layer neurons of the BPNN model are the same as the input–output indicators of the DEA model; that is, the BPNN model is used to simulate the operation of the bank branch. Randomly select a 20% (50) proportion of test (validation) subset data; it is believed that the more validation data, the better the model’s generalization ability is. The error changes in the validation data are used to monitor the training process. When the error reaches the minimum and does not decrease with further training, we take the network connection weights with the minimum error of the validation data (called the early-stopping method), eliminate the influence of over-training as much as possible, and improve the model’s generalization ability and practical value. The results show that the ANN is instead a tool for obtaining relative rankings of DMUs based on their predicted outcome. Since then, scholars have extensively researched other DEA combined models. (These previous studies added meaningful value to the existing literature, as summarized in Table 1).

**Table 1.** The detailed information on the articles establishing a combined DEA model with machine learning such as BPNN, RF, SVR, etc.

References	DEA Model	Topology Employed in a Combined Model	Combined Model	Number of Samples (Validation, Test)	To Obey the Rule of Thumb	Model for Efficiency Score (ES) and Efficient Frontier (EF)
Wu et al. [5]	CCR (2, 3)	5-10-1, 5-4-1	two BPNNs	142 (84)	no	ES
Athanassopoulos et al. [8]	CCR (2, 1) *, DEA (4, 3)	2-3-1 **, 4-10-3	BPNN	250 (50)	yes	ES, DEA > BPNN; efficiency rank, BPNN > DEA
Na et al. [9]	CCR (5, 2)	5-3-2	BPNN	13N, N	no	ES
Ma et al. [10]	CCR (6, 1)	4-4-3	BPNN	38 (5)	no	ES
Zhu et al. [11]	CCR (2, 2)	4-?-1	BPNN, GANN, SVR, ISVR	948 (48)	yes	ES, GANN > BPNN > ISVM > SVM
Olanrewaju [12]	CCR (1, 1)	1-5-1	BPNN	8	no	ES
Zhong et al. [13]	SESBM (3, 1)	3-?-1	15 MLs such as CART, CIT, Bagging, RF, BPNN, etc.	710 (3:1)	no	EF, BPNN > ET > RF > GBR
Anouze et al. [14]	CCR (5, 4), and 15 environmental variables	9-5-1	15 MLs such as CART, CIT, Bagging, RF, BPNN, etc.	151 (2:1)	no	The bagging and RF are better than BPNN, CIT, etc.
Bose et al. [15]	CCR (2, 2)	2-3-2-2, 5-4-3-2	Two BPNNs	12, 99	no	EF
Kwon [16]	two DEAs, CCRs (4, 1)	5-7-1, 5-8-1, 5-8-1, 5-7-1; 4-3-1, 4-2-1, 4-4-1, 4-5-1	eight BPNNs	56 (17)	no	ES, EF
Hong et al. [17]	CCR (4, 4)	/	SOM <sup>c</sup>	50	/	/
Yang et al. [20]	SBM (5, 5)	trial-and-error	SVM <sup>c</sup>	creating at most 500 instances, 10-fold CV	/	ES
Saeidi et al. [22]	CCR (4, 2)	6-?-1	BPNN	26	no	ES and EF
Kwon et al. [23]	CCR (3, 3), CCR (3, 1)	15-9-1, 6-30-1, 15-3-1, 15-3-1	four BPNNs	181 (37, 36) <sup>#</sup>	no	ES, EF
Ren et al. [29]	CCR (5, 5)	10-21-1	BPNN	5N, N	no	ES
Tsolas et al. [30]	eight DEAs, CCR (3, 2)	3-3-1, 6-?-1	two BPNNs	160 (4:1)	no	ES, EF

Table 1. Cont.

References	DEA Model	Topology Employed in a Combined Model	Combined Model	Number of Samples (Validation, Test)	To Obey the Rule of Thumb	Model for Efficiency Score (ES) and Efficient Frontier (EF)
Zhang et al. [31]	SESBM (3, 2)	5-10-1; 5-10-20-1; 5-10-10-10-1; 5-10-10-20-1; 5-10-20-30-1	11 MLs such as BPNN, SVR, etc.	420 (30), 5-fold CV	no	ES, BPNN is the optimal
Fallahpour et al. [32]	CCR (3, 3)	trial-and-error	SVM	48 (12)	/	ES
Yazdanparast et al. [33]	Z-DEA (1, 17)	17-?-1	BPNN	150 (45)	no	EF
Sreekumar et al. [34]	CCR (3, 8), BCC (3, 8)	11-?-1	GRNN	49	/	ES
Kao et al. [35]	CCR (10, 2)	12-?-1	two SVMs <sup>c</sup>	91	/	ES
Barros et al. [36]	PCA-CCR (2, 2)	4-20-1	BPNN	50	no	ES
Sanei et al. [37]	SBM (3, 3)	5-6-1	three BPNNs	155 (46)	no	EF
Liu et al. [38]	BBC (4, 4)	7-8-4	BPNN	120 (20)	no	ES

Notes: \*: CCR(2,1) denotes the used CCR-DEA model with two inputs and one output. \*\*: the “2-3-1” BPNN model denotes the used BPNN model with two inputs, three neurons in the hidden layer, and one output; the “?” means that the article does not specify the number of neurons in the hidden layer. c: the “c” denotes the model used for classification. #: 181 (37, 36) denotes that the numbers of the total, validation, and test subsets are 181, 37, and 36, respectively; 250 (50) denotes the numbers of the total and test subsets are 250, and 50, respectively; 160 (4:1) denotes that the total data are divided into training and test subsets according to a 4:1 ratio.

It can be seen from Table 1 that there are more than 10 DEA combined models, including ANN (mainly BPNN and RBFNN models), SVR/SVM, RF, the bagging model, etc. There are three main functions (purposes) of establishing a DEA combined model. The first purpose is to reveal the function relationships between the input–output indicators of the DEA model; that is, the input and output indicators of the ANN model are exactly or partially the same as the DEA model [8,15]. The second purpose is to reveal the working principle of the DEA model [11,14,23,29]; that is, the efficiency score calculated via the DEA model is used as the output value of the machine learning model (ML) (e.g., BPNN, RF, SVR, etc.); input–output indicators of the DEA model are used as the inputs of the ML; to take all DMUs as modeling samples; and to divide the samples into training and test datasets (also called validation datasets in some studies, but they play the role of a test dataset, which is only used to evaluate the performance of the model instead of monitoring the training process to avoid over-training). This combined model is a subsystem of the DEA model, which can be used to calculate the value of input–output efficiency. Still, it cannot be used to simultaneously calculate input excesses and output shortages (or shortfalls). The third purpose is to construct the DEA-efficient frontier function of DEA obtained by establishing the DEA model, that is, to use the DEA-efficient (valid) DMUs as the modeling samples of the ML [9,15,16,23]. Since there are a relatively small number of valid DMUs, some scholars generate a certain number of virtual DMUs by uniformly taking values within the range of the indicator value of valid DMUs. Some scholars established a BPNN model for small samples. Therefore, building a combined DEA model is not a substitute for the DEA model but can augment rather than replace the DEA model. Na et al. [9] aimed to determine the total staff employed in government organizations in China; collected statistical data from 27 provinces, municipalities, and autonomous regions; set five input indicators (such as population) and two output indicators (administrative expenditure and the number of officially employed staff); and established the CCR and CGS-DEA models. There were 10 and 13 provinces that were valid for DEA. Then, according to the 10 and 13 valid DEA samples (all as training samples; no test and validation samples), the input layer had five neurons (that is, the indicators of the DEA model), and the output layer had two neurons (the output indicators of the DEA model); a DEA-efficient frontier BPNN model



was established with a network structure of 5-3-2 and 26 connection weights, which is significantly larger than the number of training samples. It does not meet the requirements for BPNN modeling, as the number of training samples must be more than the number of network connection weights. Ren et al. [29] established a DEA model for a total of four DMUs (A, B, C, and D), with two input indicators and three output indicators; they generated uniformly distributed samples at five levels (which were all training samples; no test and validation samples), used a genetic algorithm to build a BPNN model with a 10-21-1 network topology, and constructed an efficient frontier of DEA with 253 network connection weights, which is 50 times the number of training samples; it does not meet the basic requirements for building a BPNN model. Na et al. [9] and Ren et al. [29] did not use validation samples to monitor the training process and could not judge whether over-training occurred in the training process. Because the training samples were less than the number of connection weights [9], and the number of training samples was even less than the number of hidden-layer neurons [29], over-training must occur during the training process. Therefore, it is impossible for BPNN models established in this way to have a generalization ability and practical value. In [15], the first case was to establish a DEA model based on the collected data of doctors and nurses (two input indicators) and the number of outpatients and inpatients (two output indicators) from 12 hospitals, and obtain three valid DEA samples (A, B, and D); then, they generated 125 virtual DEA-efficient DMUs using linear combinations of real DEA-efficient DMUs, A, B, and D. The researchers established a BPNN model to calculate the optimal numbers of doctors and nurses according to the number of inpatients and outpatients; that is, the output indicators of the DEA model were used as the input variables of the BPNN model, and the input indicator of the DEA-efficient DMUs was used as the output variable of BPNN. The second example studied in [15] is based on the borrowing data of 99 Indian microfinance institutions that had been collected; there were five input indicators (such as assets, etc.) and two output indicators (the number of borrowers and borrowings); according to the sample data of the six DEA-efficient DMUs, a total of 729 virtual DEA-valid samples were generated. The authors used the DEA model inputs as BPNN inputs and the DEA outputs as the BPNN outputs, established the BPNN model, and tried to obtain the optimal levels of inputs under a certain number of borrowers and borrowings—assets, offices, personnel, and expenses. For the two examples in Ref. [15], the authors did not explain whether the samples were divided into training, test, and validation subsets, nor how to determine a reasonable number of hidden-layer neurons and avoid over-training. Therefore, the validity and reliability of the results of Ref. [15] require further investigation. Tsolas et al. [30] studied the operational efficiency of 160 bank branches in Greece based on operational data; there were three inputs, namely personnel expenses, rents and depreciation, and operational expenses, and two outputs, namely net interest income and non-interest income (fee and trading income). The researchers established input-oriented DEA models, mainly including radial CCR, BCC, NIRS, NDRS, and non-radial FDH. Russell and others obtained the input–output efficiency score, and divided the efficiency score into four levels of 1–4; the authors took the input–output indicators of the DEA model as the inputs of the BPNN model (six in total) and the efficiency score level as the output (one), and randomly divided the 160 samples into training and test subsets at a ratio of 8:2. They established a BPNN model with certain results; however, since the test subset was not used to monitor the training process in real time, it was impossible to determine whether over-training occurred or not during the training process. Know et al. [23] employed ten years of longitudinal data from Japanese electronic manufacturing firms from 2003 to 2012, and each firm-year was treated as an individual DMU; the researchers eliminated non-normal data and obtained 1419 samples for modeling. There were three inputs for the DEA model (namely employees, total assets, and operating expenses) and two outputs (namely revenue and market value); they established output-oriented CCR and BCC models to produce input and output projections for each inefficient DMU to become DEA-efficient. Then, two BPNN models were established; the inputs of the BPNN1 model were the input–output indicators

(five) of the DEA model, and the output variable was the efficiency score of the DEA model, revealing the production function of the DEA model. When building the BPNN1 model, the samples were randomly partitioned into training and test datasets at a 7:3 ratio, without a validation dataset; the BPNN1 model had a higher prediction accuracy, with only six DMUs beyond 10% error and with a maximum error of 17%. However, the input and output variables of the BPNN2 model were the input–output indicators of the DEA-valid DMUs, revealing the frontier function of the DEA model; the BPNN2 split 1419 DMUs into 772 training and 647 test samples, the established BPNN model demonstrated a high prediction accuracy of less than 10% error for 95% of the DMUs, but Ref. [23] had no validation samples to monitor the training process in real time, and it is impossible to determine whether over-training occurred or not during the training process. Similarly, Kwon (2014) also established two types of BPNN models that reveal the production function and frontier function of the DEA model based on the input–output indicators of the DEA model and their modeling results. The author split the data into training and test datasets at a ratio of 7:3; 54 samples were used to train the network to learn underlying patterns, and the remaining 24 samples were used to test the network. A total of five BPNN models were established, with network topologies of 8-11-1, 8-6-1, 8-4-1, 8-18-1, and 8-14-1, and connection weights of 111, 61, 41, 181, and 141, respectively; except for the third model, the numbers of network connection weights were larger than the number of training samples, and so they do not meet the requirements for building a BPNN model. A further review of the research on combining the DEA model with neural networks such as BPNN, RBFNN, PNN (probabilistic neural network), and SOM to establish DEA-efficient frontier functions or DEA production functions for efficiency prediction, efficiency classification, the screening of training data, and data processing refers to Kwon [16]. The combined DEA models with ANN are also used for risk management, feature selection, etc. In recent years, ML, such as SVR/SVM, RF, bagging, RSM, lightGBM, etc., has been introduced in combination with various DEA models, such as SBM, SESBM with undesirable output, and window analysis [4,11,22,31]. After removing outliers, Zhong et al. [13] obtained 710 sample data points from 95 rural commercial banks from 2011 to 2018 in Guangdong Province, China. The input indicators of the SESBM model were the number of employees, fixed assets, and intermediate business expenses, and the output indicator was the intermediate business income. The dataset was randomly split into training and test subsets with a proportion of 3:1. The authors used SESBM to measure the efficiency of DMUs in the training subset and then adjusted all the training datasets to the efficiency frontier based on projection values. The DEA model's input and output indicators were the machine learning model's input and output variables, respectively, and 15 machine learning models, such as BPNN and SVR, were established. Four performance metrics were used to evaluate the different ML algorithms, namely the mean square error (MSE), the mean absolute error (MAE), the root mean square error (RMSE), and the coefficient of correlation  $R^2$ . After comparing 15 MLs, the results show that the top four MLs for efficiency evaluation performance are BPNN, the extra trees regressor (ETR), RF, and the gradient boosting regressor (GBR). SVR is the worst, and BPNN is the best. Zhong et al. [13] also established combined models between SESBM with undesirable output and SVR, BPNN, etc.; the BPNN model is better than the other models.

To sum up, many scholars have conducted in-depth research on DEA combined models and achieved good results, which have played an essential role in management and other fields. However, the following problems still exist:

- (1) The model's generalization ability and prediction accuracy are difficult to guarantee without a validation dataset. To establish MLs such as BPNN, except for [8], no other authors divided the samples into training, validation, and test datasets with similar properties. No study discusses how, according to the error changes of the validation dataset, to use the early-stopping or regularization methods to prevent over-training, and use the trial-and-error approach to determine a reasonable number of hidden-layer neurons to ensure the generalization ability and practical value [6,39–43].

Although, most studies clearly state (and some studies do not) to randomly divide the samples into training and test datasets (in some studies, the validation dataset is a test dataset) with similar properties according to the ratios of 8:2, 7:3, or 3:1 (or according to some algorithm). There are also studies stating that over-training should be avoided. Still, since no validation dataset is used to monitor the training process, it is impossible to judge whether over-training has occurred. When the number of training samples is less than three times the number of connection weights due to the extensive network topology, over-training is easy to occur during training. In the case of over-training, even if the error of the training dataset is minimal, and even if the RMSE of the test dataset is casually small, the established model has no generalization ability and practical value. Scholars should pay more attention to this problem.

- (2) It is difficult to establish reliable and effective DEA-ML combined models for DEA modeling problems, usually with only small and medium-sized samples or frontier modeling with only a few DEA-valid samples. Most DEA efficiency modeling uses small or medium-sized samples. For modeling DEA-efficient frontier functions, small samples are usually used. To establish the DEA-BPNN combined model, most of the literature does not meet the basic requirement—the number of training samples must be more than the number of connection weights. Ref. [42] put forward a rule of thumb: you should aim to have at least five times as many cases (training samples) as connection weights in the network, and preferably ten times as many for establishing a reliable and effective BPNN model. According to the rule of thumb, you can determine the reasonable number of neurons in the hidden layer through trial-and-error. To establish DEA combined with SVR, RF, etc., you should avoid over-training and overfitting to carefully determine the model's parameters with small and medium-sized samples.
- (3) It is not easy to judge which ML model is better than others. Among the ML models currently used for DEA combined modeling, some studies believe that BPNN performs better [10,12,30], while others believe that SVR and other models are better [13]. Which model has better performance is also a question worthy of study.

It can be seen that there are still several tricky problems that need to be improved in the establishment of reliable and effective DEA-combined MLs with a good generalization ability and production precision, particularly for frontier surface functions in the case of small and medium-sized samples.

## 2.2. Projection Pursuit Regression (PPR) Model

PPR is consistent with the BPNN model regarding the nonlinear approximation ability, especially for the nonlinear modeling of small and medium-sized samples that do not obey normal distribution [25–28,44,45]. The constraint condition of the PPR model is that the sum of the squares of the weights of all independent variables equals one. The modeling usually starts from linear to quadratic; cubic; polynomial ridge function (PRF); and one to two, three, etc., PRFs. So, there is a low possibility of over-training or overfitting during modeling [26,27,44,46]. It has been widely used in small and medium-sized samples such as agricultural engineering, water conservancy, demography, and earthquakes [26,44,46]. However, there are no studies on DEA-PPR combined modeling. Therefore, in this paper, the PPR model is introduced into efficiency evaluation research for the first time to construct the DEA-PPR combined model, hoping to improve the generalization ability, robustness, and reliability, mainly to solve the efficiency production function and frontier function of DEA combined modeling with small and medium-sized samples, and expand the DEA combined model and the new method of input–output efficiency research, as well as the application field of the PPR model.

The following are the innovations and contributions of this paper:

- (1) We are the first to establish the DEA-PPR combined model to effectively and reliably solve the problem of input–output efficiency of small and medium-sized samples that



- do not obey the normal distribution, to overcome the disadvantages of MLs, such as BPNN, SVR, etc., that can only be applied to large samples.
- (2) We proposed the modeling principles and steps for establishing a BPNN model with good generalization ability. Through empirical research, the reliable and effective DEA-BPNN, DEA-SVR, and DEA-PPR combined models were found to verify the effectiveness of each other. The DEA-PPR model has a relatively better generalization ability and prediction accuracy among them.
  - (3) We established the DEA-PPR combined model to simulate the production function by setting the efficiency score to one, adopting the optimization technology to obtain the frontier function directly, and realizing the unification of the production and frontier surface functions.
  - (4) We established the DEA-PPR combined model of the frontier function of the DEA by generating virtual samples according to the valid DEA samples. According to the input-oriented DEA-PPR combined model, the optimal input quantity can be obtained; on the contrary, if the output-oriented DEA-PPR combined model is used, the optimal output quantity can be obtained, providing a decision-making basis and technical paths for DMUs to organize production, strengthen management, improve efficiency, and reduce costs.

### 3. Methodology and Principle

In this paper, for the prediction research of input–output efficiency, we establish combined models of DEA with BPNN, PPR, and SVR to improve the generalization ability, robustness, and practical value.

This paper briefly introduces some typical models to evaluate input–output efficiency.

#### 3.1. DEA-CCR Model

Charnes et al. [1] first proposed the DEA-CCR model. There are  $n$  DMUs, and each DMU has  $m$  inputs and  $q$  outputs. We set  $x_i$  and  $y_r$  to represent the inputs and outputs of the DMU,  $n \times m$  is the input matrix, and  $n \times q$  is the output matrix. For each DMU, we can obtain the ratio  $\frac{\sum_{r=1}^q u_r y_{rk}}{\sum_{i=1}^m v_i x_{ik}}$  of all its outputs to inputs (where  $u_r$  is the output weight and  $v_i$  the input weight); then, the problem of the CCR model is converted to the problem of selecting the best weights. We take the input orientation as an example and obtain the specific planning equation, as in Equation (1):

$$\begin{aligned}
 \max \theta_k &= \frac{\sum_{r=1}^q u_r y_{rk}}{\sum_{i=1}^m v_i x_{ik}} \\
 \text{s.t. } &\frac{\sum_{r=1}^q u_r y_{rj}}{\sum_{i=1}^m v_i x_{ij}} \leq 1 \\
 &u \geq 0, v \geq 0 \\
 &(i = 1, 2, \dots, m; r = 1, 2, \dots, q; k, j = 1, 2, \dots, n)
 \end{aligned} \tag{1}$$

The goal of the CCR model is to maximize the efficiency value of the  $r$ th DMU under the restriction that the other DMUs' efficiency score is less than or equal to 1. We can solve the problem by transforming it into linear programming. Thus, by constructing an effective frontier, all DMUs either fall within the effective frontier (CCR-effective, CCR-efficient, efficient, and valid) or outside the effective frontier (CCR-ineffective, CCR-efficient, invalid, and inefficient).

In practice, we would like to replace Equation (1) with the corresponding linear programming pairwise model and obtain Equation (2).

$$\begin{aligned} \min & \theta \\ \text{s.t.} & \sum_{j=1}^n \lambda_j x_{ij} \leq \theta x_{ik} \\ & \sum_{j=1}^n \lambda_j y_{rj} \geq y_{rk} \\ & (i = 1, 2, \dots, m; r = 1, 2, \dots, q; k, j = 1, 2, \dots, n; \lambda \geq 0) \end{aligned} \quad (2)$$

In Equation (2),  $\lambda$  is the coefficient of the DMU, and the optimal solution ( $\theta$ ) of Equation (2) represents the efficiency score. The range of the  $\theta$  is (0,1]. For a specific DMU, the DMU is efficient when and only when  $\theta = 1$ , and if  $\theta < 1$ , the DMU is inefficient.

The CCR model is developed based on the constant return-to-scale (CRS) assumption. The DEA-BCC (or BCC) model, on the other hand, is based on a variable return-to-scale (VRS). The BCC and CCR models differ only in that the former, but not the latter, includes the convexity condition of  $\sum_{j=1}^n \lambda_j = 1, \lambda_j \geq 0, \forall j$  in its constraints. Thus, as might be expected, they share properties in common and exhibit differences. They also share properties with the corresponding additive models [2,3,47].

Using the CCR and BCC models helps to determine the DMU's overall technical and scale efficiencies, and whether the data exhibit a VRS.

Furthermore, the CCR and BCC models fail to distinguish the DMUs with the highest efficiency scores of "1"; therefore, to overcome this problem, the super-efficiency CCR or BCC models were proposed in Ref. [2].

The CCR, BCC, and super-efficiency models are radial models and cannot fully consider the effect of slackness on efficiency. The slack-based-measured (SBM) DEA model (or SBM) and the super-efficiency SBM (called SE-SBM) were proposed to solve this problem [2]. Based on the slackness measurement, SBM is a non-radial method suitable for measuring efficiency when the input and output vary non-proportionally. SE-SBM is a model that combines super efficiency and the SBM.

In the super-efficiency evaluation method, the efficient DMUs are removed from the set, and the efficiency of the DMUs is re-evaluated; thus, the original non-efficient DMUs remain unchanged, and the original efficient DMUs can be greater than 1, then they can be compared [2,13].

According to Cooper et al. [3], the DEA models have a rule of thumb—the minimum number of DMUs should be equal to or greater than  $\max\{q \times m, 3 \times (q + m)\}$ .

### 3.2. Machine Learning

Machine learning (ML) is a multi-disciplinary subject involving many disciplines, such as probability theory, statistics, approximation theory, convex analysis, and algorithm complexity theory. After decades of continuous development, ML is a well-known method that "uses algorithms to parse data, learn from it, and then make decisions or predictions about something unknown in the world" [10,40,42]. An important limitation of ML, such as BPNN, RF, SVR, etc., is that of over-training or overfitting. An ML trained to generalize well on new data (called the generalization ability) will produce a correct input–output mapping, even when the input differs slightly from the examples used to prepare the ML. However, when an ML model learns too many input–output examples, the ML model may memorize training examples that are not true of the underlying function that is to be modeled. Such a phenomenon is referred to as over-training or overfitting. During the training process or optimizing the parameters, the error of the training subset continuously decreases, even almost to zero. The error of the validation subset first decreases and then rises again. This is a sure sign that over-training or overfitting is occurring, and you should stop training once deterioration in the error of the validation subset is observed [40,42]. When an ML model is over-trained, it loses the ability to generalize between similar input–output patterns [6,40,42]. The most effective approach to ensure the generalization ability

is to reserve data to cross-validate the ML's performance. First, the available dataset is randomly partitioned into training, validation, and test subsets. Second, the training subset is used to train the ML model, the validation subset is used to validate the ML model, and the generalization performance of the established ML model is measured on the test subset. Third, if the validation and test errors are close, the trained ML model will likely generalize well; otherwise, we must re-train the ML model [40,42]. If there is no validation or test subset, judging whether over-training has occurred is challenging. BPNN and SVR are the most widely used MLs, and the PPR is not used to combine with the DEA model. We briefly introduce the principles of the BPNN and PPR models as follows.

### 3.2.1. BPNN Model

The BPNN model, proposed by Rumelhart et al. [6,39,40,42,43,46], is currently the most widely used. The BPNN model consists of an input layer, a hidden layer (usually 1, but 2–3 layers are optional), and an output layer. Each layer consists of multiple neurons. It simulates the human processing process of external input information.

Corresponding to the BPNN with one hidden layer, the output of the  $j$ th output node of the  $p$ th training sample is

$$O_{pj} = f_2 \left[ \sum_{h=1}^H \omega_{jh} f_1 \left( \sum_{i=1}^n \omega_{hi} x_{pi} + \theta_h \right) + \theta_j \right] \quad (3)$$

In Equation (3),  $\omega_{hi}$  and  $\omega_{jh}$  are the connection weights between the input layer and the hidden layer and between the hidden layer and the output layer;  $\theta_j$  and  $\theta_h$  are the thresholds of the hidden layer and the input layer; and  $f_2(\cdot)$  and  $f_1(\cdot)$  are the transfer functions of the output and hidden layers, respectively.  $n$  and  $H$  are the number of neurons in the input and hidden layers, respectively.

According to the principle of least squares, the objective function is obtained as

$$Q(\omega) = \min \left\{ \sum_{p=1}^n E_p^2 \right\} = \min \left\{ \sum_{p=1}^n \sum_{j=1}^m (t_{pj} - O_{pj})^2 \right\} \quad (4)$$

In Equation (4),  $t_{pj}$  is the expected output value;  $m$  is the number of neurons in the output layer.

According to the existence theorem proposed by Hornik et al. [48], as long as sufficient neurons are in the hidden layer, a BPNN can always make the error of the training samples as small as possible (even close to 0). However, if the hidden neurons are too many, the BPNN will likely remember the pattern of the training samples and fail to generalize the training samples. Therefore, the network topology must be as compact as possible (the number of hidden layers and neurons should be as small as possible). Second, the collected data must be split into training, validation, and test subsets with similar properties. The validation subset is used to monitor the training process. At the beginning of the training process, the error of training and validation subsets monotonically decrease. The error of the validation subset reduces to a minimum and then rises as the training continues, indicating that over-training occurred. The connection weights before over-training are taken. Refs. [39,42,43,46] proposed the basic principles and steps to be followed when building a BPNN model.

### 3.2.2. PPR Model

The PPR model was proposed by Friedman and Stuetzle [25]. Because its constraint is that the sum of the squares of the weights of each independent variable is equal to 1, over-training is less likely to occur. Therefore, the PPR model has better reliability and robustness and is especially suitable for small and medium-sized (SM) samples that do not obey the normal distribution. In addition, in comparison with other techniques such as Jackknife, Bootstrap, Monte Carlo, Lasso, and Robust Regression, the PPR model does not sample from the total dataset and can reveal the natural structure characteristics of the data; of course, the modeling process is also reliable and unbiased [46,49,50]. The other

techniques do sample to model and some bias exists. Furthermore, the distribution of the efficiency from DEA and the economies is very non-normal. Methods such as Jackknife, Lasso et al. are mainly suitable for normal distribution data.

Assuming the normalized data of the independent variable  $x(i, j)$ , we can obtain the sample projection value  $z(i) = \sum_{j=1}^m a(j)x(i, j)$  and establish the PPR model based on the cubic polynomial ridge function (PRF) of the sample projection value, as follows [25,26,46]:

$$\begin{aligned} f[z(i)] &= c_0 + c_1 z(i) + c_2 [z(i)]^2 + c_3 [z(i)]^3 \\ &= c_0 + c_1 \sum_{j=1}^m a(j)x(i, j) + c_2 \left[ \sum_{j=1}^m a(j)x(i, j) \right]^2 + c_3 \left[ \sum_{j=1}^m a(j)x(i, j) \right]^3. \end{aligned} \quad (5)$$

In Equation (5),  $c_0 \sim c_3$  is the coefficient of the cubic PRF, and  $m$  is the number of independent variables. We use the least squares method and take the minimum of the sum of squares fitting errors as the objective function:

$$Q(a, C) = \min \sum_{i=1}^n \{y(i) - f[z(i)]\}^2 \quad (6)$$

Based on the parasitism–predation algorithm [51], we obtain the optimal global solution of Equation (6), the coefficient ( $c_0 \sim c_3$ ) of the cubic PRF of Equation (5), and the optimal projection vector coefficient  $a(1), a(2), \dots, a(m)$  as well.

Due to the length limitation of this article, more detailed information on MLs can be found in Refs. [52,53].

#### 4. Empirical Researches

Whether the number of DMUs is small or large, the number of DEA-efficient DMUs is small. To establish the combined model of the DEA-efficient frontier function, we must generate sufficient virtual DEA-efficient samples.

##### 4.1. Empirical Illustration Using Hospital Data

We use the data for 12 hospitals as an example from Cooper et al. [3] that depict the CCR efficiencies in Refs. [3,6]. The samples' data are shown in Table 2.

The primary purposes of establishing the two DEA-PPR combined models are as follows:

- (1) To analyze the relationship between the efficiency and the DEA input–output indicators, judge the importance of the inputs and outputs, and predict the efficiency of new hospital data;
- (2) To study how to improve the hospitals' managerial efficiency, and to provide the lower bounds for the inputs (the numbers of doctors and nurses) for each inefficient DMU to produce or service its current level of outputs (the numbers of inpatients and outpatients).

According to Refs. [3,6], we also run a input-oriented CCR under CRS, applying the MaxDEA 9.0 software [2]. Table 2 shows the DEA efficiencies of the twelve DMUs. The efficiency scores of the three DEA-efficient DMUs (A, B, and D) equal “1”.

As aforementioned, we established the DEA-CCR model and obtained the efficiency of each DMU, determined whether one DMU is efficient or inefficient, constructed the efficient frontier of DEA, calculated the inputs' excesses of each decision unit, etc. However, the optimal solution is not single for the DEA model, is used to establish discontinuously changing frontiers and production functions (models for calculating input–output efficiency), and cannot predict the efficiency values of other (new) DMUs and their optimal input quantities (for input-oriented and specific outputs, the output-oriented is the optimal input). So, we should establish combined models such as DEA-PPR to overcome the disadvantages of DEA. This study will select the DEA-ANN, DEA-PPR, and DEA-SVR combined models and compare their performance.

**Table 2.** The DMUs' data and their DEA efficiency, as well as the predicted results.

DMUs	Original Data				ES	Efficient/ Inefficient	PPR Model			
	I1 *	I2	O1	O2			PPR-1	PPR-2	PPR	ESN
A	20	151	100	90	1	Efficient	0.996	0.018	1.014	1
B	19	131	150	50	1	Efficient	0.974	0.017	0.992	1
C	25	160	160	55	0.883	Inefficient	0.902	−0.012	0.891	0.874
D	27	168	180	72	1	Efficient	0.995	−0.008	0.987	0.941
E	22	158	94	66	0.764	Inefficient	0.787	−0.011	0.777	0.748
F	55	255	230	90	0.835	Inefficient	0.807	0.023	0.831	0.791
G	33	235	220	88	0.902	Inefficient	0.930	−0.002	0.928	0.902
H	31	206	152	80	0.796	Inefficient	0.806	−0.024	0.782	0.752
I	30	244	190	100	0.960	Inefficient	0.897	0.022	0.919	0.960
J	50	268	250	100	0.871	Inefficient	0.905	−0.028	0.877	0.819
K	53	306	260	147	0.955	Inefficient	1.002	−0.035	0.968	0.873
L	38	284	250	120	0.958	Inefficient	0.979	0.008	0.984	0.958
B-C **	20.9	141.2	160	55	1	Efficient	0.980	0.012	0.993	1(M) #
B-E	16.8	120.6	94	66	1	Efficient	0.982	0.015	0.997	0.963(N)
B-F	33.8	212.9	230	90	1	Efficient	1.002	−0.012	0.990	
B-G	29.8	208.6	220	88	1	Efficient	1.001	0.013	1.013	
B-H	24.7	164	152	80	1	Efficient	0.996	0.001	0.997	
B-I	28.8	207.1	190	100	1	Efficient	1.002	0.014	1.016	
B-J	37.5	233.3	250	100	1	Efficient	1.002	−0.018	0.985	
B-K	43.3	292.3	260	147	1	Efficient	0.984	−0.012	0.972	
B-L	36.4	259.5	250	120	1	Efficient	1.000	0.011	1.011	

Notes: \*: I1 and I2 denote two inputs, the doctor and nurse, respectively; O1 and O2 denote two outpatient and inpatient outputs, respectively; ES denotes the efficiency score. \*\*: B-C, B-E, . . . , and K-L represent the benchmark of DMU C, E, . . . , and K, respectively. #: the letters M and N in the bracket denote the new DMUs, and the column of ESN is the efficiency scores after adding new DMUs.

#### 4.1.1. To Establish the DEA-PPR Combined Model of the DEA Production Function

According to the CCR model mentioned above, we obtain the efficiencies of 12 DMUs. Establishing a robust and reliable combined model is challenging due to the small samples. Therefore, we establish a PPR model with PRF by taking the efficiencies of 12 DMUs as the values of the output variable of PPR and the input–output indicators as the input variables of the PPR model. Because the CCR model describes a complex input–output relationship, the PPR model with a linear PRF cannot satisfy the accuracy. In this paper, we tried to establish a PPR model with a quadratic PRF. We input the above data into our compiled program based on the PPA [26,51] and obtain the optimal global solution, the first-dimension PPR model, and the DMUs' efficiency scores (ES),

$$ES_{PPR-1} = 1.00007 + 0.06054z_1 - 0.36947z_1^2 \quad (7)$$

where,  $z_1 = -0.05135I_1 - 0.71259I_2 + 0.43926O_1 + 0.54464O_2$ . The ES is shown in the  $ES_{PPR-1}$  column of Table 2. Since the coefficient of the quadratic term of the Formula (7) is less than 0, Formula (7) has a maximum value. When  $z_1 = 0.081928$ , the maximum value of  $ES_{PPR-1}$  is 1.00751.

It can be seen that, for the PPR model of the input-oriented DEA model, the two outputs are positive to the efficiency, and the two inputs are negative to the efficiency. If we take appropriate measures to reduce the input values of one DMU (or increase



the output values), the  $z_1$  will increase, and the ES rise; otherwise, the ES will decrease. The relationship between the ES and the input variables is entirely consistent with the production function of the DEA model. It can be seen from  $z_1$  that the impact of I2 (nurse) on ES is more significant than that of I1 (doctor). Therefore, reducing the number of nurses is more conducive to improving the ES of these hospitals, while the two output indicators are not much different.

According to the PPR model (7), the mean absolute error (MAE) is calculated as  $MAE = 0.018$ , the mean absolute percentage error as  $MAPE = 2.2\%$ , and the root mean squared error as  $RMSE = 0.024$ . That is to say, Equation (7) has revealed the production function of the DEA model well, but significant errors still exist in very few DMUs. For example, the ES of DEA-inefficient DMU K (1.002) is greater than DEA-efficient DMU B, which cannot accurately reflect the ES of the CCR. Then, we further build a PPR model with the second quadratic PRF to describe the ES:

$$ES_{PPR-2} = 0.00310 + 0.07887z_2 + 0.03985z_2^2 \quad (8)$$

where  $z_2 = -0.87938I_1 + 0.31175I_2 + 0.31072O_1 + 0.18248O_2$ . The ES in Equation (8) is shown in the  $ES_{PPR-2}$  column of Table 2. Since the coefficient of the quadratic term is greater than 0, Equation (8) has a minimum value. When  $z_2 = -0.98959$ , the minimum value  $ES_{PPR-2}$  is  $-0.03592$ . Thus, the PPR model with two quadratic PRFs reveals the production function of the DEA model as follows:

$$ES_{PPR} = ES_{PPR-1} + ES_{PPR-2} \\ = 1.00381 + 0.06054z_1 - 0.36947z_1^2 + 0.07887z_2 + 0.03985z_2^2 \quad (9)$$

The ES calculated according to Equation (9) is shown in the  $ES_{PPR}$  column of Table 2. Therefore, for the DEA-PPR combined model of the original DEA model, if we take appropriate measures to reduce the input (or increase the output), the  $z_1$  or  $z_2$  increase, and the ES will rise; otherwise, the ES will decrease. Because the  $z_2$  of all DMUs are greater than  $-0.98959$ , the larger the  $z_2$  is, the higher the ES of the PPR model is. The relationship between the ES of the PPR model and the input variables is precisely the same as the production function of the original DEA model. Meanwhile, according to the combined DEA-PPR model in Equation (9) with two PRFs, the impact of I2 (nurse) on the hospital's ES is reduced because its coefficient in  $z_2$  is greater than 0. In contrast, the impact of I1 (doctor) is significantly enhanced because the coefficient of  $z_2$  reaches 0.87938, which is significantly larger than that of I1. Therefore, the impacts of the number of nurses and doctors on the hospital's ES are reduced, and the effects of the two output indicators are not much different, which is consistent with the actual operation of the hospital.

The linear relationship between the ES calculated with the DEA-PPR combined model and the ES of the original DEA model is  $ES_{PPR} = 0.0303 + 0.968ES_{DEA}$ , and the correlation coefficient is  $R = 0.9737$ , indicating that the ES values calculated according to the two models are significantly correlated at the 0.01 level. For the DEA-PPR combined model,  $MAE = 0.014$ ,  $MAPE = 1.5\%$ , and  $RMSE = 0.017$ . The performance metrics, such as MAE, MAPE, etc., are better than that of Equation (7). The ES of DMU K from Equation (9) equals 0.984 and is still more significant than that of benchmark DMU L (0.972), but smaller than that of other DEA-efficient and benchmark DMUs. Therefore, the performance metrics and the distribution of the ES of DMUs show that the established DEA-PPR combined model has a high prediction accuracy, reveals the production function characteristics of the CCR, and suggests good practical value.

We may suppose that, through investigation or other methods, we have obtained the numbers of the outpatients, inpatients, doctors, and nurses of two more hospitals, as shown in Tables 2 and 3, named M and N. If we rebuild the CCR, the ES of the 14 DMUs is shown in Table 2 (in the ESN column). Except for DMUs A, B, G, I, and L, the other DMUs' ES has changed. Furthermore, DMU D changes from being DEA-efficient to inefficient. So, we cannot rank the DMUs objectively when adding new DMUs.

**Table 3.** Comparison of the optimal number of doctors and nurses based on PPR and BPNN combined models.

DMUs	I <sub>1</sub>	I <sub>2</sub>	O <sub>1</sub>	O <sub>2</sub>	D <sub>PPR</sub>	N <sub>PPR</sub>	D <sub>B</sub>	N <sub>B</sub>	D <sub>PPR-27</sub>	N <sub>PPR-27</sub>	D <sub>PPR-8</sub>	N <sub>PPR-8</sub>	D <sub>NN-729</sub>	N <sub>NN-729</sub>
A	20	151	100	90	20.04	150.95	20.00	150.19	20.02	150.98	20.00	151.00	20.26	149.17
B	19	131	150	50	20.08	129.91	19.01	131.31	20.07	129.92	20.04	129.95	20.19	129.94
D	27	168	180	72	25.57	169.45	26.93	167.17	25.55	169.47	25.52	169.50	25.29	168.20
C	25	160	160	55	21.63	140.44	21.06	140.16	21.62	140.45	21.59	140.48	21.47	139.30
E	22	158	94	66	16.61	120.82	14.68	122.47	16.60	120.83	16.58	120.86	17.97	122.48
F	55	255	230	90	32.51	214.18	33.84	194.51	32.48	214.20	32.45	214.23	32.56	214.94
G	33	235	220	88	31.31	207.04	32.97	191.43	31.29	207.07	31.26	207.10	31.33	207.61
H	31	206	152	80	23.84	164.90	24.86	163.26	23.82	164.92	23.79	164.95	23.58	163.41
I	30	244	190	100	29.86	206.06	30.86	186.89	29.84	206.09	29.81	206.12	29.85	206.51
J	50	268	250	100	35.61	235.25	35.42	200.64	35.59	235.27	35.55	235.31	35.69	236.11
K	53	306	260	147	42.15	293.42	33.42	196.15	42.12	293.46	42.09	293.48	42.05	292.72
L	38	284	250	120	37.99	257.93	34.96	200.38	37.96	257.96	37.93	257.99	38.03	258.24
M	25	150	170	79	25.45	272.52	/	/	22.43	172.54	25.40	172.57	25.18	171.38
N	30	300	90	130	23.85	191.45	/	/	23.81	191.49	23.79	191.51	23.73	191.02

According to the above-established DEA-PPR combined model (9), we can quickly obtain the ES of DMU M and N to be 0.991 and 0.249, respectively.

For the above problem, since there are only 12 samples and four input–output indicators, even if we adopt the leave-one-out cross-validation method, the number of neurons in the hidden layer is one (there are seven connection weights); it does not meet the rule of thumb for building a BPNN model—the number of training samples should be more than five times the number of connection weights, so it is impossible to construct a DEA-BPNN, as well as DEA-SVR and DEA-RF models, with good generalization ability.

#### 4.1.2. To Establish the DEA-PPR Combined Model of the DEA-Efficient Frontier Function

Because there is usually a relatively small number of DEA-efficient DMUs, we cannot directly build a DEA-BPNN combined model based on the efficient DMUs only. Ref. [15] constructed 125 virtual efficient DMUs based on three DEA-efficient DMUs, obtained 128 samples, and established a DEA-BPNN combined model. However, since Ref. [5] did not specify whether to divide the samples into training, validation, and test subsets with similar properties, nor how to prevent over-training (or overfitting, over-learning), the generalization ability and robustness of the established DEA-BPNN combined model is questionable and not explicit.

Based on Equation (9), the DEA-PPR combined model of the DEA production function, if we set the ES to be 1, we can obtain the frontier function model of the DEA-efficient DMUs. If I<sub>2</sub> (I<sub>1</sub>), O<sub>1</sub>, and O<sub>2</sub> are assumed, I<sub>1</sub> (I<sub>2</sub>) can be obtained, and the rest can be obtained similarly. Of course, with the virtual samples of Ref. [15], we can also establish the DEA-PPR combined model of the DEA-efficient frontier functions. Referring to [15], we divided the closed interval [0,1] into five equal subintervals, each of the five values (0.2, 0.4, 0.6, 0.8, 1.0) for DEA-efficient A, B, and D, respectively. We generated  $5^3 = 125$  combinations

of DEA-efficient DMUs in all. We have a matrix of the above proportions,  $P$ , consisting of 125 rows and three columns, and matrix  $S$  of three DEA-efficient DMUs, as follows:

$$P = \begin{bmatrix} 0.2 & 0.4 & 0.6 & 0.8 & 1.0 & 0.2 & 0.4 & \dots & 1.0 & 1.0 & 1.0 & 1.0 \\ 0.2 & 0.2 & 0.2 & 0.2 & 0.2 & 0.4 & 0.4 & \dots & 1.0 & 1.0 & 1.0 & 1.0 \\ 0.2 & 0.2 & 0.2 & 0.2 & 0.2 & 0.2 & 0.2 & \dots & 0.4 & 0.6 & 0.8 & 1.0 \end{bmatrix}^T$$

$$S = \begin{bmatrix} 20 & 151 & 100 & 90 \\ 19 & 131 & 150 & 50 \\ 27 & 168 & 180 & 72 \end{bmatrix}$$

Thus, we form the matrix  $V = P_{125 \times 3} S_{3 \times 4}$  and generate 125 ( $5^3$ ) virtual linear combinations of the DEA-efficient DMUs. Along with the matrix representing the DEA-efficient DMUs taken singly, we thus have  $125 + 3 = 128$  linear combinations.

We take the above 128 samples as modeling samples and the nine benchmark DMUs as test samples (or solve the optimal input of these DMUs), input the data into the PPA-based program, and establish the DEA-PPR combined model with one quadratic PRF of the DEA-efficient frontier function. For the input-oriented CCR model, referring to [14], we, respectively, establish, under the given output conditions, the combined DEA-PPR models for the optimal number of doctors and nurses ( $D_{PPR}$ ,  $N_{PPR}$ ) (represented by the letters  $D$  and  $N$ , respectively), as follows:

$$D_{PPR} = 38.6034 + 8.2998z_D - 0.00461z_D^2 \quad (10)$$

$$N_{PPR} = 263.057 + 54.931z_N + 0.0048z_N^2 \quad (11)$$

where  $z_D = 0.85137O_1 + 0.52457O_2$ ,  $z_N = 0.65370O_1 + 0.75676O_2$ ;  $O_1, O_2$  has been centered on having a mean zero and scaled to have a standard deviation of 1. From the formulas of  $z_D$  and  $z_N$ , it can be seen that the number of outpatients significantly impacts the optimal number of doctors. In contrast, the number of inpatients significantly impacts the optimal number of nurses, but the difference between the two indicators is insignificant.

#### 4.1.3. To Determine the Optimal Number of Doctors and Nurses

According to Equations (10) and (11), the frontier function of the DEA-PPR combined model, we can determine the optimal numbers of doctors and nurses. The optimal numbers of doctors and nurses for nine inefficient DMUs (verification samples) are shown in the  $D_{PPR}$  and  $N_{PPR}$  columns of Table 3.

#### 4.1.4. To Compare the Performance of the Different Models

Based on the optimal numbers of doctors and nurses obtained from the DEA combined models, the performance metrics of nine verification samples, such as the MAE, MAPE, RMSE, correlation coefficient ( $R$ ), maximum absolute error ( $E_{A-max}$ ), and maximum relative (percentage) error ( $E_{R-max}$ ) of the DEA-PPR combined model, were obtained, as shown in the  $D_{PPR}$  and  $N_{PPR}$  rows of Table 4. The performance metrics of the DEA-BPNN model established in Ref. [15] are shown in the  $D_B$  and  $N_B$  rows of Table 4.

From the performance metrics of different models shown in Table 4, it can be seen that the generalization ability of the DEA-PPR combined model with one quadratic PRF of the DEA-efficient frontier function is significantly better than that of the DEA-BPNN combined model in Ref. [15]. For the optimal number of doctors and nurses, the DEA-BPNN combined model has a relatively poor generalization ability, with the maximum relative errors reaching 29.6% and 49.0%, respectively, which shows that over-training occurs when the BPNN model is established (but Ref. [15] does not explicitly state whether they use the validation or test subsets to monitor the training process and how to avoid over-training).

**Table 4.** Comparison of performance between different models with different virtual combinations.

	MAE	MAPE (%)	RMSE	R	$E_{A-max}$	$E_{R-max}$ (%)
$D_B$	2.346	8.0	3.695	0.898	9.880	29.6
$D_{PPR}$	1.068	5.3	1.190	0.989	1.589	7.9
$D_{SVR}$	1.083	5.4	1.193	0.957	1.827	9.1
$D_{NN}$	1.163	4.2	1.281	0.983	2.262	6.8
$D_{PPR-27}$	1.068	3.7	1.188	0.989	1.882	5.6
$D_{PPR-8}$	1.065	3.7	1.182	0.989	1.820	5.3
$D_{PPR-729}$	1.070	3.66	1.192	0.989	1.894	5.8
$D_{NN-729}$	1.209	4.39	1.283	0.988	1.807	7.0
$D_{SVR-729}$	1.621	6.47	1.736	0.983	3.163	17.0
$N_B$	27.48	14.1	40.62	0.895	96.15	49.0
$N_{PPR}$	1.081	0.7	1.202	1	1.946	1.3
$N_{SVR}$	1.081	0.7	1.212	0.999	1.999	1.3
$N_{NN}$	1.111	0.6	1.253	0.999	2.379	1.0
$N_{PPR-27}$	1.078	0.6	1.199	1	1.939	0.9
$N_{PPR-8}$	1.075	0.6	1.193	1	1.672	0.9
$N_{PPR-729}$	1.083	0.56	1.204	1	1.952	0.9
$N_{NN-729}$	1.298	0.75	1.504	1	2.807	1.6
$N_{SVR-729}$	2.648	1.65	3.230	0.999	6.629	4.4

#### 4.1.5. To Analyze the Robustness and Reliability of the DEA-PPR Combined Models

To further verify the robustness and reliability of the PPR model for small samples, we once divided the closed interval  $[0,1]$  into three and two equal subintervals, respectively, and generated virtual DEA-efficient samples. The former three values are 0.33, 0.66, and 1, and the latter two values are 0.5 and 1. We can generate  $3^3 = 27$  and  $2^3 = 8$  virtual samples, a total of 30 and 11 samples. We take the nine DEA-inefficient DMUs as the test subset and establish the DEA-PPR combined model with one quadratic PRF of the DEA-efficient frontier function based on the 27 and 8 virtual samples (respectively, represented with the subscripts 27 and 8), as follows:

$$D_{PPR-27} = 38.666 + 9.0173z_{D-27} - 0.0196z_{D-27}^2 \quad (12)$$

$$N_{PPR-27} = 263.554 + 59.576z_{N-27} + 0.0203z_{N-27}^2 \quad (13)$$

where  $z_{D-27} = 0.8515O_1 + 0.5243O_2$ ,  $z_{N-27} = 0.6543O_1 + 0.7562O_2$ ;  $O_1, O_2$  has been normalized to have a mean zero and scaled to have a variance of 1, the same as follows:

$$D_{PPR-8} = 36.366 + 9.9389z_{D-8} - 0.0743z_{D-8}^2 \quad (14)$$

$$N_{PPR-8} = 247.515 + 65.574z_{N-8} + 0.0489z_{N-8}^2 \quad (15)$$

where  $z_{D-8} = 0.8497O_1 + 0.5274O_2$ ,  $z_{N-8} = 0.6526O_1 + 0.7577O_2$ .

Since the virtual samples and quantities used to build DEA-PPR combined models differ, the normalized values ( $O_1, O_2$ ) are also slightly different.

It can be seen from Table 4 that the performance metrics, such as the MAE, MAPE, and RMSE, of models with different virtual samples have a good consistency. Even if only eight virtual samples are generated, the established DEA-PPR combined models have a high fitting accuracy and generalization ability, indicating that the DEA-PPR combined models have good robustness and reliability.

#### 4.1.6. To Establish the DEA-SVR Combined Model of the DEA-Efficient Frontier Function

Theoretically, an SVR model can minimize the structural risk, so it has a good nonlinear approximation ability, and has also been widely used in nonlinear data modeling, such as constructing DEA frontier functions [11,54,55]. However, in practice, the optimal values of the penalty factor and the width coefficient of the radial basis kernel function must be obtained via the multi-fold cross-validation method [54,55]; otherwise, there is a high possibility of overfitting during optimization. The performance metrics of the SVR model are directly related to whether the values of parameters such as the penalty factor and width coefficient are reasonable, so the model does not have strong stability [11]. Due to the length limitation, we only built the DEA-SVR combined model for the case of 125 virtual samples (taking nine inefficient DMUs and two new DMUs, M and N, to be test samples). If the value ranges of the penalty factor and the width coefficient are both  $[10^{-2}, 10^5]$ , we obtained the optimal solutions of the two parameters, 0.019 and  $10^5$  (the width coefficient is equal to the boundary value, which is not reasonable). The mean square error of the training samples is already less than  $10^{-7}$ , but the error of the above nine benchmark verification samples is vast; the MAE of the optimal number of nurses is 21.01, the MAPE is 13.91%, the RMSE is 39.07, the  $E_{A-max}$  is 112.65, and the  $E_{R-max}$  is 74.63; clearly, overfitting occurred during the optimization process. Suppose the value ranges of the penalty factor and the width coefficient are adjusted to  $[10^{-2}, 10^3]$ . In that case, we obtain the optimal solutions of two parameters, 586.7 and  $10^3$  (the width coefficient is still equal to the boundary value, which is not reasonable), and the mean square error of the training sample is 0.36; the MAE, MAPE, RMSE,  $E_{A-max}$ ,  $E_{R-max}$ , and other data on the optimal number of nurses in the nine benchmark verification samples of the DEA-SVR combined model are shown in the  $N_{SVR}$  row of Table 4. Similarly, when the DEA-SVR combined model of the optimal number of doctors is established when the value ranges of the penalty factor and the width coefficient are both  $[10^{-2}, 10^5]$ , the optimization process has overfitting; when the value ranges are adjusted to  $[10^{-2}, 10^3]$ , the optimal solutions are 484.0 and  $10^3$ , respectively; the MAE, MAPE, RMSE,  $E_{A-max}$ ,  $E_{R-max}$  and other data of the established DEA-SVR combined model are shown in the  $D_{SVR}$  row of Table 4. It can be seen that the DEA-SVR combined model has poorer performance than the DEA-PPR combined model.

It can be seen from the above modeling process that if the value of the penalty factor is too small, there is a high possibility of overfitting. Unfortunately, there is no commercial software for the SVR model that can, while optimizing the parameters, monitor the error changes of the test samples to prevent overfitting. It brings great difficulties and challenges to establish a DEA-SVR combined model.

#### 4.1.7. To Establish the DEA-BPNN Combined Model of the DEA-Efficient Frontier Function

Although Ref. [15] generated 125 virtual samples and established a DEA-BPNN combined model of the frontier function, the errors of the nine verification samples were large (see Table 4), among which, for the relative errors of the optimal numbers of doctors and nurses in benchmark K, are as high as 22.8% and 32.9%, respectively. The relative error of the optimal number of nurses in benchmark L is also 22.8%, indicating that over-training may have occurred in the training process. So, we try to rebuild the DEA-BPNN combined model. Although many scholars have used the BPNN model to construct the frontier function and production function of the CCR model, few follow the basic principles and steps of establishing a reliable and robust BPNN model. According to the existence theorem proposed by Hornik et al. [48], as long as the hidden layer has sufficient neurons, the error of the training subset of a BPNN can reach as small a value as possible (even close to 0). However, the established BPNN model may have no generalization ability. Therefore, the core of building a BPNN model is to prevent over-training during the training process. We establish a DEA-BPNN combined model with good reliability, robustness, and generalization ability, and follow the basic principles and steps as follows:



- (1) We randomly divide the samples into training, validation, and test subsets with similar properties according to the ratio of 2:1:1. (The number of the validation and test subsets should account for at least 15%, respectively);
- (2) We use the trial-and-error method and make the BPNN topology as compact as possible (usually one hidden layer and the number of neurons in the hidden layer is as small as possible). The ratio of the number of training subsets to the number of connection weights must be greater than one and should be more than five, preferably ten, according to the rule of thumb;
- (3) We use the training subset to adjust the connection weights to reduce the sum of squares error (SSE) of the training subset and the validation subset to monitor the training process. Along with the training process, the SSE of the training subset gradually decreases, and the SSE of the validation subset first falls to a specific minimal value, and then begins to rise again, which is a sure sign that over-training is occurring. We stop training (called the early-stopping method). To take the network weights before the SSE begins to rise, we establish the BPNN model;
- (4) We use the test subset to measure the prediction ability and performance of the BPNN model. If the SSE of the test subset is reasonably close to or slightly larger (generally less than 1.3 times) than the SSE of the training and validation subsets, the established BPNN model has a good generalization ability, reliability, robustness, and prediction ability, as well as practical value. Otherwise, we should restart the process from (3) until the BPNN model has good generalization and prediction abilities.

The above principles and steps must be wholly followed; otherwise, the generalization ability and practical value of the established BPNN cannot be guaranteed.

We randomly selected 68, 30, and 30 training, validation, and test samples from the above 128 samples. The three subsets had the same properties (the mean and standard deviation were almost the same; if they were inconsistent, we should resample). According to [15], the input layer had two neurons, outpatient and inpatient, and the output layer had one output, the optimal number of nurses or doctors. We established two BPNN models to predict them, respectively. According to the rule of thumb, the number of training samples must reach more than five times the number of connection weights, and the number of hidden-layer neurons cannot exceed three. Using the STATISTICA Neural Networks (SNNs) software [42], we used the logistic transformation function for the hidden layer and output layer; considering that the model must have a certain degree of extrapolation needs, the input and output data were linearly transformed into the range of [0.2, 0.8]. We used the quasi-Newton optimization algorithm and monitored the training process with the validation subset. We compared and studied the performance of the models with 1, 2, and 3 hidden neurons. Under the premise of following the above principles and no over-training, the model with one neuron in the hidden layer had a high enough performance; if the number of hidden-layer neurons was two or three, there is a high possibility of over-training, and the performance was not improved significantly. We established the DEA-BPNN combined models for the optimal numbers of nurses and doctors, respectively, and the performance metrics of nine benchmark verification samples are shown in the  $N_{NN}$  and  $D_{NN}$  rows of Table 4.

The performance metrics, such as the MAE, MAPE, RMSE,  $R$ ,  $E_{A-max}$ , and  $E_{R-max}$  of the training, validation, and test subsets of the DEA-BPNN combined models for the optimal numbers of doctors and nurses, are shown in Table 5.

**Table 5.** Comparison of performance metrics of three subsets of DEA-BPNN combined models.

Model	Training Subset				Validation Subset				Test Subset			
	MAE	MAPE	$E_{A-max}$	$E_{R-max}$	MAE	MAPE	$E_{A-max}$	$E_{R-max}$	MAE	MAPE	$E_{A-max}$	$E_{R-max}$
$D_{NN}$	0.496	1.52	1.597	11.68	0.513	1.49	1.515	4.35	0.527	1.33	1.822	6.75
$N_{NN}$	2.170	2.06	3.245	8.60	2.417	2.17	6.13	9.47	2.649	2.66	10.06	10.29
$D_{NN-729}$	0.429	1.13	2.250	14.00	0.429	1.31	3.28	24.9	0.430	1.21	2.02	13.26
$N_{NN-729}$	0.833	0.37	15.75	15.40	0.787	0.30	9.32	2.13	0.880	0.39	8.55	6.37

It can be seen from Table 5 that the performance metrics of the training, validation, and test subsets are relatively close, indicating that the established BPNN models have a good generalization ability and reliability.

In cases where only 27 and 8 virtual samples are generated, even if we use the leave-one-out cross-validation method, it is impossible to establish DEA-BPNN and DEA-SVR combined models with generalization ability and reliability due to the small samples.

#### 4.1.8. To Establish the Combined Models of the DEA-Efficient Frontier Function for Large Samples

To verify the suitability of the DEA-PPR combined model for large samples, we generated 729 ( $9^3$ ) samples taking each of the nine values (0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1) for efficient DMUs A, B, and D, respectively. The minimal values of the samples were equal to that of 125 samples. We randomly divided the 731 samples into 371, 180, and 180 training, validation, and test datasets according to the ratio of 2:1:1. We established the DEA-PPR combined models for doctors and nurses, and the performance metrics of the nine benchmark verification samples are shown in the  $D_{PPR-729}$  and  $N_{PPR-729}$  rows of Table 4. By obeying the principles and steps, we established the DEA-BPNN combined models, and the performance metrics are shown in the  $D_{NN-729}$  and  $N_{NN-729}$  rows of Table 4. The performance metrics of the three subsets are shown in the  $D_{NN-729}$  and  $N_{NN-729}$  rows of Table 5. The optimal number of doctors and nurses is shown in the  $D_{NN-729}$  and  $N_{NN-729}$  columns of Table 3. We set the search range of parameters to be  $[10^{-2}, 10^3]$ , optimized the SVR models, and obtained the optimal values to be (7.762, 0.626) and (7.762, 3.206), respectively, which show that overfitting did not occur. The performance metrics of the nine benchmark verification samples are shown in the  $D_{SVR-729}$  and  $N_{SVR-729}$  rows of Table 4.

From the performance metrics of different models in Table 4, it can be seen that, for generating 125 and 729 virtual samples, respectively, the DEA-PPR, DEA-BPNN, and DEA-SVR combined models we established without over-training or overfitting have quite a high prediction accuracy and generalization ability. However, the DEA-PPR combined model slightly outperforms the DEA-BPNN and DEA-SVR for medium-sized samples, and the DEA-PPR combined model outperforms the DEA-BPNN and DEA-SVR for large samples. Furthermore, the DEA-BPNN combined model has a better generalization ability than the DEA-SVR combined models under medium-sized and large samples. More importantly, due to the strongly nonlinear approximation ability of the BPNN and SVR models, there is a high possibility of over-training and overfitting during training or optimizing the parameters within unreasonable search ranges. Cross-validation methods with validation subsets must be used, which require a cumbersome modeling process and cannot guarantee the generalization and prediction abilities of the model without following the principles and steps. Because the PPR model with a quadratic PRF is characterized by a relatively simple structure and moderate nonlinear approximation ability, it is especially suitable for modeling problems that are not very complex, such as the frontier surface, the production function, etc., of the DEA model. Whether for 729, 125, 27, or 8 virtual samples, the established DEA-PPR combined models have quite a good generalization ability and prediction ability. The PPR model is especially suitable for small and medium-sized samples that do not obey normal distribution. Furthermore, the DEA model is also ideal for analysis

and modeling to research the input–output efficiency of small and medium-sized samples. Therefore, for the problem of small and medium-sized samples, the DEA-PPR combined model has unique advantages, such as good generalization, robustness, and reliability.

#### 4.2. Empirical Illustration Using China’s Provincial Carbon Dioxide Emission Quotas

As the world’s top CO<sub>2</sub> emitter (accounting for 26.5% of the world’s total emissions) and the second largest economy, China’s achievement of peak CO<sub>2</sub> emissions and carbon neutrality (called the “dual carbon” target) will play a pivotal role in reducing the world’s CO<sub>2</sub> emissions. Therefore, reasonably allocating China’s provincial carbon emission quotas (CPCEQs) plays a vital role in promoting high-quality development and striving to achieve a “dual carbon” target. In this study, firstly, it is necessary to judge and evaluate the advantages and disadvantages of different allocation methods. Secondly, it is required to study the carbon emission efficiency of different periods (e.g., comparing efficiency between 2019 and 2030), and so on. DEA is the most used model to measure the efficiency of carbon emissions. It is a pity we cannot use the DEA model to compare the efficiency change between different periods and the advantages and disadvantages of other allocating methods.

The primary purposes of establishing the two DEA-PPR combined models are as follows:

- (1) To build a relationship between the carbon emissions efficiency and the DEA inputs and outputs, judge the importance of the inputs and outputs, and predict the efficiency of new carbon-allocating methods and the quotas in 2030;
- (2) To study ways to improve China’s provincial carbon emissions efficiency and help to implement the “dual carbon” target, and provide lower bounds for the inputs (the carbon emissions) for each inefficient DMU to produce its current level of outputs (the GDP and population).

##### 4.2.1. Data Resource

Referring to Gomes et al. [56], we take the gross domestic production (GDP) and the population of provinces as the output indicators and the actual carbon emissions as the input indicator. The data of the output indicators are from the *China Statistical Yearbook* (2020) (<http://www.stats.gov.cn/sj/ndsj> (accessed on 8 July 2023)), and the input data are from Chinese carbon emission accounts & datasets (<https://www.ceads.net> (accessed on 8 July 2023)). The data are shown in Table 6.

**Table 6.** The input–output data and ES of the DEA model and the predicted ES with the DEA-PPR combined model.

Provinces	CO <sub>2</sub> (Mt)	Population (10 <sup>4</sup> Persons)	GDP (CNY 100 Million)	ES	Quotas	ES-q	Quotas-y	ES-y	Quotas-2030	ES-2030	SVR
Beijing	89.2	2154	35,371	1	89.2	0.6906	426.5	0.6516	84.3	0.9578	0.9877
Tianjin	158.5	1562	14,104	0.417	66.1	0.7636	179.2	0.6503	157.7	0.6094	0.4351
Hebei	914.2	7592	35,105	0.3133	286.4	0.3407	453.0	0.8803	910.0	0.4506	0.3410
Shanxi	564.9	3729	17,027	0.2562	144.7	0.5744	207.5	0.8229	727.0	0.3222	0.2874
Inner Mongolia	794.3	2540	17,213	0.1280	101.7	0.6701	547.6	0.2707	712.2	0.3484	0.1726
Liaoning	533.4	4352	24,909	0.3146	167.8	0.5214	361.3	0.6240	504.0	0.2491	0.3401
Jilin	203.7	2691	11,727	0.5245	106.8	0.6702	218.6	0.5558	213.5	0.5236	0.5353
Heilongjiang	278.2	3751	13,613	0.5231	145.5	0.5779	290.6	0.5776	262.9	0.4580	0.5302
Shanghai	192.9	2428	38,155	0.5581	107.7	0.6301	374.8	0.7978	173.0	0.7542	0.5641
Jiangsu	804.6	8070	99,632	0.6410	515.8	1.2084	787.1	0.9723	760.2	0.8652	0.6389
Zhejiang	381.4	5850	62,352	0.7201	274.7	0.4306	546.3	0.8834	379.7	0.8558	0.7108

Table 6. Cont.

Provinces	CO <sub>2</sub> (Mt)	Population (10 <sup>4</sup> Persons)	GDP (CNY 100 Million)	ES	Quotas	ES-q	Quotas-y	ES-y	Quotas-2030	ES-2030	SVR
Anhui	408.1	6366	37,114	0.5923	241.7	0.4090	366.2	0.9428	427.8	0.4555	0.5950
Fujian	278.1	3973	42,395	0.5922	164.7	0.5287	352.6	0.9392	285.6	0.6615	0.5945
Jiangxi	242.3	4666	24,758	0.7393	179.1	0.4996	276.8	0.8492	267.5	0.6407	0.7402
Shandong	937.1	10,070	71,068	0.4823	452.0	0.4691	824.4	0.7480	885.4	0.5838	0.4947
Henan	460.6	9640	54,259	0.9059	417.3	0.3018	528.1	1	458.5	0.8236	0.8966
Hubei	354.8	5927	45,828	0.6636	235.4	0.4297	417.2	0.8935	371.9	0.5774	0.6600
Hunan	310.6	6918	39,752	0.8430	261.9	0.3874	385.3	0.9731	325.7	0.7527	0.8359
Guangdong	569.1	11,521	107,671	1	569.1	1.2367	826.4	1	510.3	1.4683	0.9877
Guangxi	246.7	4960	21,237	0.7689	189.7	0.4799	232.6	1	258.7	0.7177	0.7617
Hainan	43.1	945	5309	1	43.1	0.8542	74.5	0.6900	67.8	0.8221	0.9860
Chongqing	156.3	3124	23,606	0.7891	123.3	0.6135	238.1	0.8199	155.5	0.8515	0.7736
Sichuan	315.2	8375	46,616	1	315.2	0.3491	572.2	0.7915	313.7	0.9902	0.9877
Guizhou	261.1	3623	16,769	0.5394	140.9	0.5831	165.6	1	288.3	0.5035	0.5414
Yunnan	186.0	4858	23,224	1	186.0	0.4869	367.0	0.6421	216.1	0.7992	0.9881
Shaanxi	296.3	3876	25,793	0.5085	150.7	0.5553	308.7	0.7174	294.9	0.5161	0.5184
Gansu	164.5	2647	8718	0.6397	105.2	0.6820	108.8	1	172.4	0.6670	0.6323
Qinghai	51.8	608	2966	0.8322	43.1	0.8170	118.2	0.3213	48.9	0.8353	0.8346
Ningxia	212.4	695	3748	0.2028	43.1	0.8257	38.0	1	222.7	0.3959	0.2404
Xinjiang	455.3	2523	13,597	0.2214	100.8	0.6800	271.3	0.4695	408.2	0.2431	0.2574
Tianjin <sup>b</sup>		1562	14,104	1	60.7	0.9783					0.6401
Hebei <sup>b</sup>		7592	42,258	1	285.7	0.9669					0.7387
Shanxi <sup>b</sup>		3729	20,756	1	140.3	0.9747					0.6924
Inner Mongolia <sup>b</sup>		2540	17,213	1	96.7	0.9757					0.6219
Liaoning <sup>b</sup>		4352	24,909	1	164.0	0.9673					0.8048
Jilin <sup>b</sup>		2691	14,978	1	101.3	0.9869					0.6233
Heilongjiang <sup>b</sup>		3751	20,878	1	141.2	0.9745					0.6931
Shanghai <sup>b</sup>		2515	38,155	1	107.7	0.9107					0.8277
Jiangsu <sup>b</sup>		10,479	99,632	1	515.8	0.9770					0.622
Zhejiang <sup>b</sup>		5850	62,352	1	274.7	0.8479					0.6197
Average	/	/	/	0.6239	/	0.6089	/	0.7828	/	0.6567	/

Notes: <sup>b</sup> denotes the benchmark of the provinces of DEA-efficient DMUs. The quotas are obtained according to Refs. [56,57] using the ZSG-DEA model, and its ES is ES-q; we obtain the quotas (denoted in “Quotas-y” column of Table 6) according to an allocation system consisting of multiple variables. The quotas—2030 are obtained according to the GDP and carbon emission intensity of the “Outline” and “Plan”.

We establish the input-oriented BCC model to measure the ES of the Chinese provincial carbon emissions in 2019, referring to Refs. [56,57] and the ES shown in Table 6. According to Refs. [56,57], we built the ZSG-DEA to allocate the provincial quotas and obtain the quotas, as shown in the “quotas” column of Table 6. If we rebuild the BCC model using the quotas and make the ES values all equal to 1, we cannot judge whether the ES will rise or decline. Meanwhile, we can allocate the provincial quotas by establishing an allocation system consisting of multiple variables [46] and obtain the quotas denoted in the “Quotas-y” column of Table 6. But how do we judge the ZSG-DEA’s quotas and the quotas of the allocation system? We can obtain the provincial quotas in 2030 according to the GDP and carbon emission intensity in the Action Plan for Carbon Peaking before 2030 (namely “Plan”), as well as the 14th Five-Year Plan for Economic and Social Development and

the Outline of the Vision for 2035 (namely “Outline”) of provinces in China. We cannot compare and evaluate the ES of different allocation methods and periods solely through the DEA model because adding more DMUs changes the ES of the original DMUs. So, we must establish combined models to do so.

#### 4.2.2. To Build the Combined Model Characterizing the DEA Production Function

##### (1) To build a DEA-PPR combined model

Referring to Section 4.1, we take the ES as the output variable and the three input–output indicators as the input variables, and build the DEA-PPR combined model with one quadratic PRF, and the 30 training and ten verification samples as shown in Table 6 (noted “b”), as follows:

$$ES_{PPR-1} = 0.7099 - 0.5571z_1 + 0.1613z_1^2 \quad (16)$$

where  $z_1 = 0.8622I_1 - 0.4605O_1 - 0.2110O_2$ ,  $I_1$  denotes CO<sub>2</sub> (Mt),  $O_1$  denotes the population, and  $O_2$  denotes the GDP, as shown in Table 6. The performance metrics of the training samples, such as the RMSE, AAE, MAPE, and  $E_{A-max}$ , are 0.1129, 0.0834, 17.48%, and 0.3305, respectively. The performance metrics of the verification samples are 0.0737, 0.0615, 6.15%, and 0.1172, respectively. The performance of the DEA-PPR model is not good enough, and we rebuild the DEA-PPR model with the second quadratic PRF as follows:

$$ES_{PPR-2} = -0.0557 + 0.0727z_2 + 0.0876z_2^2 \quad (17)$$

where  $z_2 = -0.7195I_1 + 0.4261O_1 - 0.5484O_2$ . The performance metrics of the training samples, such as the RMSE, AAE, MAPE, and  $E_{A-max}$ , are 0.0871, 0.0621, 14.83%, and 0.2448, respectively. The performance metrics of the verification sample are 0.0552, 0.0501, 5.00%, and 0.0819, respectively. Although we can build the third PRF to improve the performance, it is shown that the performance of the DEA-PPR model is good enough to describe the DEA production function. That is to say, the DEA-PPR combined model can approximately characterize the DEA production function as follows:

$$ES_{PPR} = ES_{PPR-1} + ES_{PPR-2} \quad (18)$$

The calculated ES of the provincial quotas, quotas-y, and quotas—2030 are all shown in the “ES-q”, “ES-y”, and “ES-2030” columns of Table 6.

According to the allocation system, the average ES of the quotas is greater than that of the actual carbon emissions, and shows that the allocated quotas are reasonable and improve the efficiency of carbon emissions. The average ES in 2030 is less than the actual carbon emissions in 2019 and shows that the carbon emissions efficiency in 2030 is lower than in 2019. To raise the carbon emissions efficiency in 2030, we must take measurements to reduce carbon emissions. If we can reduce carbon emissions by 5% in 2030, the average ES calculated with the DEA-PPR combined model is 0.6446, greater than the average ES in 2019.

The average ES of the quotas from the ZSG-DEA model is lower than the average ES of the actual carbon emissions, and shows that the ZSG-DEA model for carbon emission quotas can equalize the carbon emission efficiency of provinces, which does reduce carbon emission efficiency.

##### (2) To build the DEA-SVR and DEA-BPNN combined models

There are too few samples to build DEA-BPNN and DEA-RF combined models. We build an SVR model according to 4.6. We set the parameters’ search range to be [0.01, 10] and obtain the optimal solution; the penalty factor equals 0.0258, and the width coefficient equals 10. The performance metrics of the training samples, such as the RMSE, AAE, and MAPE, are 0.0155, 0.0123, and 3.15, respectively. The performance metrics of the verification sample are 0.3205, 0.3116, and 31.16, respectively. Overfitting occurred during the optimization process.



#### 4.2.3. To Build the Combined Models Characterizing the DEA Frontier Function

We cannot build the combined models of the DEA-efficient frontier function using the five DMUs only. Referring to Section 4.2, we also generate  $2^5$  or  $3^5$  virtual DEA-efficient DMUs by dividing the value interval  $[0, 1]$  into two  $(0.5, 1)$  or three  $(0.3333, 0.6667, 1)$  equal parts in a linear combination with the data of five DEA-efficient samples. We generated 32 and 243 virtual samples, respectively.

##### (1) To build a DEA-PPR combined model

We take the DEA model's input indicator as the PPR model's output variable and the two output indicators as the input variables. We take the ten samples in Table 6, denoted as "b", as the verification samples.

We build the PPR model with the first quadratic PRF as follows:

$$C_{PPR-32} = 677.96 + 272.23z_{32} + 0.9597z_{32}^2 \quad (19)$$

$$C_{PPR-243} = 765.99 + 172.26z_{243} + 0.4784z_{243}^2 \quad (20)$$

where  $z_{32} = 0.8784O_1 + 0.4779O_2$  and  $z_{243} = 0.8792O_1 + 0.4765O_2$ . The performance metrics of the training and validation subsets, such as the RMSE, AAE, MAPE, and  $E_{A-max}$ , are shown in Table 7, respectively. The performance metrics show that the DEA-PPR combined model has good performance, generalization ability, and robustness for 32 or 243 virtual samples. The DEA-PPR combined model with one quadratic PRF already has a high enough prediction accuracy.

**Table 7.** Performance comparison of the PPR, BPNN, and SVR models in different virtual combinations.

Model	Training Subset					Validation Subset				
	MAE	MAPE (%)	RMSE	$E_{A-max}$	$E_{R-max}$ (%)	MAE	MAPE (%)	RMSE	$E_{A-max}$	$E_{R-max}$ (%)
$C_{PPR-32}$	12.30	3.29	15.53	43.20	36.41	8.37	5.05	14.66	35.00	27.37
$C_{SVR-32}$	10.92	5.03	15.00	45.78	106.3	42.54	27.43	53.41	132.2	49.06
$C_{PPR-243}$	12.08	1.89	14.69	45.33	41.81	9.52	5.73	14.86	37.51	22.82
$C_{SVR-243}$	11.40	2.10	14.67	51.29	119.1	36.45	24.33	44.44	106.2	44.77
$C_{BPNN-243}$	11.74	1.60	14.51	52.78	9.27	22.77	14.03	30.83	69.65	52.94
$C_{SVR-243}$	7.39	1.88	10.54	71.50	166.0	34.35	28.80	37.13	59.44	97.98

##### (2) To build the DEA-SVR and DEA-BPNN combined models

We build an SVR model according to Sections 4.1.6 and 4.1.7. When the parameters' search range is  $[0.01, 8.7096]$ , the RMSE equals the RMSE of the DEA-PPR combined model. The optimal solution is that the penalty factor equals 0.2041, and the width coefficient equals 8.7096; that is, the optimal width coefficient equals the upper bound value. The performance metrics of the training and validation subsets are shown in Table 7.

It can be seen that the DEA-SVR combined model has almost the same performance for the training subset as DEA-PPR but has inferior performance for the validation subset compared to the DEA-PPR combined model. Overfitting occurred during the optimization process.

Furthermore, the main disadvantage of the SVR model is that its performance is directly dependent on the search range of the parameters. If we change the upper boundary limit, the performance will change, too. If we set the upper boundary limit to 100, the SVR has a poorer generalization ability.

##### (3) To build the DEA-BPNN and DEA-RF combined models

The 32 virtual samples are too few to build DEA-BPNN and DEA-RF combined models. The 243 virtual samples are barely suitable for building the DEA-BPNN and DEA-RF combined models.

We build a DEA-BPNN combined model. We randomly split the 248 (243 + 5) samples into training, validation (especially for the BPNN and RF model), and test subsets in the proportion of 6:2:2 [31,42,45] and obtain 148, 50, and 50 training, validation, and test samples. According to the rule of thumb, we should preferably have ten times as many training samples as the number of the connection weights. We determine the number of hidden neurons to be three; the connection weights are 13. The transfer functions in the hidden and output layers are Sigmoid applying the SNN [42]. We train the BPNN, monitor the error change in the validation subset, and build the BPNN model without over-training. The performance metrics of the DEA-BPNN combined model of the training and validation subset are shown in Table 7. The performance metrics of the validation and test subsets are 11.41, 1.61%, 14.64, 47.81, and 111.00, and 11.71, 3.78%, 13.83, 25.79, and 3.85, respectively. The performance metrics show that the built BPNN model has a good generation ability but is still poorer than the DEA-PPR combined model. According to the built model, we can conclude that GDP significantly impacts efficiency more than the population.

We build a DEA-RF combined model. We randomly split the 248 (243 + 5) samples into training and validation subsets in the proportion of 1:1 [46]. We take the default value of the parameters of the DPS software [46], the number of trees being 300, and build the DEA-RF combined model. The performance metrics of the DEA-RF combined model of the training and validation subsets are shown in Table 7. The validation subset's performance metrics are greater than the training subset's, indicating that overfitting has occurred in the optimization. The built DEA-RF combined model has a poor generalization ability.

The comparison of the performance metrics of different combined models shows that the DEA-PPR model has the best performance and generalization ability, robustness, and prediction accuracy, whether for large samples or small and medium-sized samples. The DEA-BPNN model without over-training has the second-best performance for medium-sized and large samples. The DEA-SVR combined model has a poorer generalization ability than the DEA-PPR and DEA-BPNN combined models, and overfitting occurs in optimizing the parameters under medium-sized samples. The DEA-RF combined model has the poorest generalization ability of the four combined models for large samples. We cannot build DEA-BPNN, DEA-SVR, and DEA-RF combined models for small samples.

## 5. Results and Concluding Remarks

### 5.1. The PPR and DEA Models Have Similarities in Frontier Morphology and Theoretical Consistency

According to the DEA modeling principle, the frontier surface of the input-oriented (or output-oriented) DEA is the top surface of a convex polyhedron that is convex (or concave) to the coordinate origin. The PPR model based on the quadratic PRFs comprises multiple (at least one) quadratic surfaces, which are convex polyhedra. It can be seen that the convex polyhedron of the PPR model and the top surface of the frontier surface of the DEA model have a similar morphology and consistent theory. Therefore, there is a low possibility of overfitting in building the DEA-PPR combined model, which thus has a good generalization ability and robustness. The widely used BPNN, RF, and SVR models with very complex curved surfaces have a low morphological similarity and theoretical consistency with the DEA model.

### 5.2. The Characterization Ability of the DEA-PPR Combined Model to the DEA Production Function

The main disadvantage of DEA is the static analysis of the existing DMUs. The production function constructed via DEA is discontinuous, so the newly added DMUs cannot be analyzed. If the DEA model is re-established with the newly added DMUs, the efficiency and ranking of the original DMUs will be changed, leading to a lack of comparability. For example, after adding new DMUs (M and N) and rebuilding the DEA model, the ES of the seven original DMUs changed, and DMU D changed from being DEA-efficient to inefficient.

To overcome its disadvantages, scholars have established combined models, such as DEA-BPNN, DEA-SVR, DEA-RF, etc., based on the input–output indicators and the ES of the DEA model that have played a good role. However, most combined models, such as BPNN, RF, etc., are mainly suitable for large samples and not suitable for small and medium-sized samples. Therefore, under the cases of small and medium-sized samples, we are the first to propose an innovative idea of establishing a DEA-PPR combined model and its modeling steps. For the two actual examples, we established reliable and effective combined models, describing the production function and the frontier surface function of the CCR, respectively. Especially in the cases of small and medium-sized samples that do not obey the normal distribution, it is challenging to establish a reliable DEA-BPNN, DEA-SVR, or DEA-RF combined model for the production or frontier surface functions of the CCR with generalization ability and prediction accuracy.

It can be seen from Tables 2 and 6 that the established DEA-PPR combined model can characterize the production function of the DEA model, has sufficient fitting accuracy, and predicts the ES of new DMUs. We input the data of the new DMUs (M and N) into the DEA-PPR combined model (Equation (10)), and obtained their ES values of 0.991 and 0.249, respectively.

Therefore, the DEA-PPR combined model can exploit the advantages of both nonparametric models and overcome their shortcomings.

### 5.3. The Characterization Ability of the DEA-PPR Combined Model to DEA-Efficient Frontier Function

We establish the DEA model to obtain the DEA-efficient frontier function consisting of a series of line segments of the polyline (approximately). There are only a few DEA-efficient DMUs, and we must fill the gap between the broken line and the actual situation. Of course, if new DMUs are added, it is impossible to analyze whether it is on the frontier surface or to calculate its optimal input amount. Therefore, according to the DEA-efficient DMUs, some scholars have divided each indicator into equal parts to generate virtual DEA-efficient DMUs. According to different equivalent fractions, we can establish combined models such as DEA-PPR, DEA-BPNN, DEA-SVR, etc., to describe the DEA-efficient frontier function. To establish the BPNN, RF, and SVR combined models, we must randomly divide the samples into training, validation, and test subsets with similar properties. We use cross-validation to train or optimize the parameters to avoid over-training.

We usually establish combined models for large samples, such as DEA-PPR, DEA-BPNN, DEA-SVR, etc., with a good generalization ability; otherwise, it is difficult to guarantee the generalization ability. The DEA-PPR combined model has a better generalization ability than the DEA-BPNN model, the DEA-SVR model, etc.

The PPR model is especially suitable for small and medium-sized samples of nonlinear and non-normal distribution. Our empirical research generates medium-sized and medium-sized virtual DMUs with a linear combination of DEA-efficient DMUs. We established reliable DEA-PPR combined models with a good generalization ability. Differently to the BPNN and SVR combined models, according to the DEA-PPR combined model, we can easily judge the importance and ranking of the indicators according to the coefficients of different inputs (or outputs). The DEA-PPR combined model outperforms the DEA-BPNN and DEA-SVR combined models.

In short, the DEA-PPR combined model outperforms DEA-BPNN, DEA-SVR, etc., for small, medium-sized, and large samples, and should be recommended for use regarding generalization ability, robustness, and accuracy.

According to the DEA-PPR combined model, we can obtain the optimal inputs (outputs) for new DMUs and take measures to improve efficiency and management.

## 6. Limitations and Future Research

This study mainly focuses on overcoming the weaknesses of the DEA model and the shortcomings of the DEA-BPNN and DEA-SVR combined models not being well applied

to small and medium-sized samples. We introduce the PPR model and first establish a DEA-PPR combined model with a good generalization ability. However, there are some limitations. Future research should use more examples to verify the generalization ability and accuracy between DEA-BPNN, DEA-SVR, and DEA-PPR for establishing production and frontier functions. Second, we take the DEA-CCR model as an example, and in the next step, we should study other DEA models to verify the applicability of the DEA-PPR combined model. This paper is the first to focus on the DEA-PPR combined model, which requires more empirical research to confirm its universality and reliability.

**Author Contributions:** Conceptualization, W.L. and X.Y.; methodology, X.Y.; software, W.L.; validation, X.Y.; formal analysis, W.L.; investigation, X.Y.; resources, X.Y.; data curation, X.Y.; writing—original draft preparation, W.L. and X.Y.; visualization, X.Y.; supervision, W.L.; project administration, X.Y.; funding acquisition, X.Y. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research is supported by Shanghai Business School Plateau Discipline Project of Business Administration, No. SHGDD-GSGL-2023-02-02.

**Data Availability Statement:** The data are available on request.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Charnes, A.; Cooper, W.W.; Rhodes, E. Measuring the efficiency of decision making units. *Eur. J. Oper. Res.* **1978**, *2*, 429–444. [CrossRef]
- Cheng, G. *Data Envelopment Analysis: Methods and MaxDEA Software*; Intellectual Property Press: Beijing, China, 2014. Available online: <http://www.maxdea.cn/> (accessed on 12 June 2023).
- Cooper, W.; Seiford, L.; Tone, K. *Data Envelopment Analysis—A Comprehensive Text with Models, Applications, References and DEA-Solver Software*; Kluwer Academic Publishers: Boston, MA, USA, 2007.
- Panwar, A.; Olfati, M.; Pant, M.; Snasel, V. A review on the 40 years of existence of data envelopment analysis models: Historic development and current trends. *Arch. Comput. Methods Eng.* **2022**, *29*, 5397–5426. [CrossRef] [PubMed]
- Wu, D.; Yang, Z.; Liang, L. Using dea-neural network approach to evaluate branch efficiency of a large Canadian bank. *Expert Syst. Appl.* **2006**, *31*, 108–115. [CrossRef]
- Bishop, C. *Neural Networks for Pattern Recognition*; Oxford University Press: Oxford, UK, 1995.
- Lewis, H.F.; Sexton, T.R. Network DEA: Efficiency analysis of organizations with complex internal structure. *Comput. Oper. Res.* **2004**, *31*, 1365–1410. [CrossRef]
- Athanassopoulos, A.D.; Curram, S.P. A comparison of data envelopment analysis and artificial neural networks as tool for assessing the efficiency of decision making units. *J. Oper. Res. Soc.* **1996**, *47*, 1000–1016. [CrossRef]
- Na, J.; Zhou, Z.; Zhou, H. A decision-analysis approach to determine the total staff employ in government organization—Establishment of local government organizations. *Chin. J. Manag. Sci.* **1997**, *5*, 7–17.
- Ma, C.; Wu, C.; Zhang, S.; Yang, Y.; Li, H.; Han, Y. Decision making method for variable-rate fertilization based on data envelopment analysis and artificial neural network. *Trans. CSAE* **2004**, *20*, 152–155.
- Zhu, N.; Zhu, C.; Emrouznejad, A. A combined machine learning algorithms and DEA method for measuring and predicting the efficiency of Chinese manufacturing listed companies. *J. Manag. Sci. Eng.* **2021**, *6*, 435–448. [CrossRef]
- Olanrewaju, O. Integrated index decomposition analysis-artificial neural network-data envelopment analysis (IDA-ANN-DEA)—Implementation guide. *Energy Effic.* **2021**, *14*, 71–78. [CrossRef]
- Zhong, K.; Wang, Y.; Pei, J. Super efficiency SBM-DEA and neural network for performance evaluation. *Inf. Process. Manag.* **2021**, *58*, 102728. [CrossRef]
- Anouze, A.L.M.; Bou-Hamad, I. Data envelopment analysis and data mining to efficiency estimation and evaluation. *Int. J. Islam. Middle East. Financ. Manag.* **2019**, *12*, 169–190. [CrossRef]
- Bose, A.; Patel, G.N. “NeuralDEA”—A framework using neural network to re-evaluate DEA benchmarks. *OPSearch* **2015**, *52*, 18–41. [CrossRef]
- Kwon, H. Exploring the predictive potential of artificial neural networks in conjunction with DEA in railroad performance modeling. *Int. J. Prod. Econ.* **2017**, *183*, 159–170. [CrossRef]
- Hong, H.K.; Ha, S.H.; Shin, C.K.; Park, S.C.; Kim, S.H. Evaluating the efficiency of system integration projects using data envelopment analysis (DEA) and machine learning. *Expert Syst. Appl.* **1999**, *16*, 283–296. [CrossRef]
- Tsappes, G.; Papathanasiou, J.; Georgiou, A.C. An exploratory DEA and machine learning framework for the evaluation and analysis of sustainability composite indicators in the EU. *Mathematics* **2022**, *10*, 2277. [CrossRef]
- Mirmozaffari, M.; Shadkam, E.; Khalili, S.M.; Kabirifar, K.; Yazdani, R.; Gashteroodkhani, T.A. A novel artificial intelligent approach: Comparison of machine learning tools and algorithms based on optimization DEA Malmquist productivity index for eco-efficiency evaluation. *Int. J. Energy Sect. Manag.* **2021**, *25*, 523–550. [CrossRef]



20. Yang, X.; Dimitrov, S. Data envelopment analysis may obfuscate corporate financial data: Using support vector machine and data envelopment analysis to predict corporate failure for nonmanufacturing firms. *INFOR Inf. Syst. Oper. Res.* **2017**, *55*, 295–311. [\[CrossRef\]](#)
21. Emrouznejad, A.; Shale, E. A combined neural network and DEA for measuring efficiency of large scale datasets. *Comput. Ind. Eng.* **2009**, *56*, 249–254. [\[CrossRef\]](#)
22. Saeidi, S.; Jouybanpour, P.; Mirvakilli, A.; Iranshahi, D.; Klemeš, J.J. A comparative study between modified data envelopment analysis and response surface methodology for optimisation of heterogeneous biodiesel production from waste cooking palm oil. *J. Clean. Prod.* **2016**, *136*, 23–30. [\[CrossRef\]](#)
23. Kwon, H.; Lee, J.; Roh, J. Best performance modeling using complementary DEA-ANN approach—Application to Japanese electronics manufacturing firms. *Benchmarking Int. J.* **2016**, *23*, 704–721. [\[CrossRef\]](#)
24. Farahmand, M.; Desa, M.; Nilashi, M. A combined data envelopment analysis and support vector regression for efficiency evaluation of large decision making units. *Int. J. Eng. Technol.* **2014**, *6*, 2310–2321.
25. Friedman, J.; Stuetzle, W. Projection pursuit regression. *J. Am. Stat. Assoc.* **1981**, *76*, 817–823. [\[CrossRef\]](#)
26. Lou, W. *The Projection Pursuit Theory Based on Swarm Intelligence Optimization Algorithms—New Developments, Applications, and Software*; Fudan University Press: Shanghai, China, 2021.
27. Hwang, T.; Lay, S.; Maechler, M. Regression modeling in back-propagation and projection pursuit learning. *IEEE Trans. Neural Netw.* **1994**, *5*, 342–353. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Zhan, H.R.; Zhang, M.K.; Xia, Y.C. Ensemble projection pursuit for general nonparametric. *arXiv* **2022**, arXiv:2210.14467.
29. Ren, H.; Ma, X.R.; Li, H.B. Improvement of input evaluation for giant projects based on GA-BP neural network. *Syst. Eng.—Theory Pract.* **2015**, *35*, 1474–1481.
30. Tsolas, I.E.; Charles, V.; Gherman, T. Supporting better practice benchmarking: A DEA-ANN approach to bank branch performance assessment. *Expert Syst. Appl.* **2020**, *160*, 113599. [\[CrossRef\]](#)
31. Zhang, Z.; Xiao, Y.; Niu, H. DEA and Machine Learning for Performance Prediction. *Mathematics* **2022**, *10*, 1776. [\[CrossRef\]](#)
32. Fallahpour, A.; Kazemi, N.; Molani, M.; Nayyeri, S.; Ehsani, M. An Intelligence-Based Model for Supplier Selection Integrating Data Envelopment Analysis and Support Vector Machine. *Iran. J. Manag. Stud.* **2018**, *11*, 209–241.
33. Yazdanparast, R.; Tavakkoli-Moghaddam, R.; Heidari, R.; Aliabadi, L. A hybrid Z-number data envelopment analysis and neural network for assessment of supply chain resilience: A case study. *Cent. Eur. J. Oper. Res.* **2021**, *29*, 611–631. [\[CrossRef\]](#)
34. Sreekumar, S.; Mahapatra, S. Performance modeling of Indian business schools: A DEA-neural network approach. *Benchmarking Int. J.* **2011**, *18*, 221–239. [\[CrossRef\]](#)
35. Kao, H.; Huang, C.; Chen, J. Classification using DEA and SVM approaches: The empirical study of higher education. *Information* **2013**, *16*, 7801–7810.
36. Barros, C.; Wanke, P. Insurance companies in Mozambique: A two-stage DEA and neural networks on efficiency and capacity slacks. *Appl. Econ.* **2014**, *46*, 3591–3600.
37. Sanei, R.; Hosseinzadeh lotfi, F.; Fallah, M.; Sobhani, F. An estimation of an acceptable efficiency frontier having an optimum resource management approach, with a combination of the DEA-ANN-GA technique (A case study of branches of an insurance company). *Mathematics* **2022**, *10*, 4503. [\[CrossRef\]](#)
38. Liu, Q.; Shang, J.; Wang, J.; Niu, W.; Qiao, W. Evaluation and prediction of the safety management efficiency of coal enterprises based on a DEA-BP neural network. *Resour. Policy* **2023**, *83*, 103611. [\[CrossRef\]](#)
39. Lou, W.G. Evaluation and prediction of soil quality based on artificial neural network in the Sanjiang Plain. *Chin. J. Manag. Sci.* **2002**, *10*, 79–83.
40. Haykin, S. *Neural Networks and Learning Machines*; China Machine Press: Beijing, China, 2009.
41. Zhang, G.; Patuwo, E.; Hu, M. Forecasting with artificial neural networks: The state of the art. *Int. J. Forecast.* **1998**, *14*, 35–62. [\[CrossRef\]](#)
42. StatSoft, Inc. *STATISTICA Neural Networks*; StatSoft, Inc.: Tulsa, OK, USA, 2011.
43. Zhong, C.; Lou, W.G.; Wang, C. Neural Network-Based Modeling for Risk Evaluation and Early Warning for Large-Scale Sports Events. *Mathematics* **2022**, *10*, 3228. [\[CrossRef\]](#)
44. Singhee, A. SiLVR: Projection Pursuit for Response Surface Modeling. In *Machine Learning in VLSI Computer-Aided Design*; Elfadel, I., Boning, D., Li, X., Eds.; Springer Nature: Cham, Switzerland, 2019.
45. Hall, P. On projection pursuit regression. *Ann. Stat.* **1989**, *17*, 573–588. [\[CrossRef\]](#)
46. Yu, X.H.; Xu, H.Y.; Lou, W.G.; Xu, X.; Shi, V. Examining energy eco-efficiency in China's logistics industry. *Int. J. Prod. Econ.* **2023**, *258*, 108797. [\[CrossRef\]](#)
47. 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. [\[CrossRef\]](#)
48. Hornik, K.; Stinchcombe, M.; White, H. Multilayer feedforward networks are universal approximators. *Neural Netw.* **1989**, *2*, 359–366. [\[CrossRef\]](#)
49. Chen, C.; Tuo, R. Projection Pursuit Gaussian Process Regression. *IIEE Trans.* **2023**, *55*, 901–911. [\[CrossRef\]](#)
50. Tao, H.; Tao, J.; Li, Q.; Aihemaiti, M.; Jiang, Y.; Yang, W.; Wei, J. Average relative flow of single-wing labyrinth drip irrigation tape based on projection pursuit regression. *Sci. Rep.* **2022**, *12*, 8543. [\[CrossRef\]](#) [\[PubMed\]](#)



51. Mohamed, A.A.A.; Hassan, S.A.; Hemeida, A.M.; Alkhalaf, S.; Mahmoud, M.; Eldin, A.M.B. Parasitism–Predation algorithm (PPA): A novel approach for feature selection. *Ain Shams Eng. J.* **2020**, *11*, 293–308. [[CrossRef](#)]
52. Athey, S.; Imbens, G. Machine learning methods that economists should know about. *Annu. Rev. Econ.* **2019**, *11*, 685–725. [[CrossRef](#)]
53. Marsland, S. *Machine Learning: An Algorithmic Perspective*; CRC Press: Boca Raton, FL, USA; Taylor & Francis Group: Abingdon, UK, 2015.
54. Valero-Carreras, D.; Aparicio, J.; Guerrero, N.G. Support vector frontiers: A new approach for estimating production functions through support vector machines. *Omega* **2021**, *104*, 102490. [[CrossRef](#)]
55. Gomes, E.G.; Lins, M.P.E. Modelling undesirable outputs with zero sum gains data envelopment analysis models. *J. Oper. Res. Soc.* **2008**, *59*, 616–623. [[CrossRef](#)]
56. Zhou, X.; Niu, A.Y.; Lin, C.X. Optimizing carbon emission forecast for modelling China’s 2030 provincial carbon emission quota allocation. *J. Environ. Manag.* **2023**, *325*, 116523. [[CrossRef](#)]
57. Tang, Q.; Zhang, C. Data processing system (DPS) software with experimental design, statistical analysis and data mining developed for use in entomological research. *Insect Sci.* **2013**, *20*, 254–260. [[CrossRef](#)]

**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.