Production system efficiency optimization through application of a hybrid artificial intelligence solution

Joao Henrique Cavalcanti, Tibor Kovacs, Andrea Ko & Károly Pocsarovszky

Published online: 14 Sep 2023.

Submit your article to this journal ↗

Article views: 133

View related articles ↗

View Crossmark data ↗

Taylor & Francis
Taylor & Francis Group

Production system efficiency optimization through application of a hybrid artificial intelligence solution

Joao Henrique Cavalcanti [ID], Tibor Kovacs, Andrea Ko and Károly Pocsarovszky

Department of Information Systems, Corvinus University of Budapest, Budapest, Hungary

ABSTRACT

Industry 4.0 seeks waste reduction via the optimization of production systems integrating technology and process. In addition to evaluating existing methods and technologies, academia also develops new ones. This research proposes a new hybrid artificial intelligence (AI) solution for production system efficiency optimization that combines data envelopment analysis (DEA), machine learning (ML)-based simulation and genetic algorithms (GAs) using real-world sensor data from a thermoelectric power plant. In the proposed method, DEA is employed to identify the production system's efficient frontier, which is used to build an ML model that predicts production efficiency through simulation. A genetic algorithm is then utilized to propose those settings that result in optimized production efficiency. Although the possibility of combining DEA-ML and ML-GA has been discussed in the literature, no research was found that combines these three methods for production efficiency optimization. The proposed solution was tested and validated using real-world data. The benefits of the hybrid AI solution were measured by comparing its predicted efficiency with the efficiencies achieved by running production with conventional control-loops based control systems. The results show that considerable efficiency improvement can be achieved using the proposed hybrid AI solution.

1. Introduction

Energy is one of the most important commodities in the world. Many products or services utilize energy either directly or indirectly; therefore, energy generation efficiency has a substantial impact on global economies. Lower energy generation efficiency and the subsequent increase in energy costs result in increased prices of multiple products/services, generating inflation, loss of purchasing power, and in extreme cases, even economic recession. Increasing energy generation efficiency contributes to sustainable economic development. Consequently, academia constantly proposes new methods, applying scientific knowledge to improve production efficiencies and focusing on the reduction of energy production costs.

Lean manufacturing uses a collection of tools and techniques to eliminate waste in production systems as much as possible, and consequently, to reduce production costs. As science and technology advance, more cutting-edge technologies are used to support lean production systems, leading to the fourth industrial revolution or Industry 4.0. Industry 4.0 utilizes a wide range of technological tools, such as artificial intelligence (AI), robotics, the Internet of Things, and cloud computing, aimed at cost/waste reductions and product enhancements. Aimed at extending lean manufacturing and Industry 4.0 capabilities, this research has the goal of developing a hybrid AI solution to optimize production efficiency to be tested with the real-world data of a thermoelectric power plant. The hybrid AI solution combines data envelopment analysis (DEA), machine learning (ML) based simulation and genetic algorithms (GA). DEA is used to identify the production system's efficient frontier to be applied in the second step of building an ML model that predicts production efficiency through simulation. The GA proposes the corresponding optimal settings of inputs and process control parameters. The proposed hybrid AI solution aims to reduce waste and increase production efficiency, subsequently contributing to lower energy production costs and creating a positive impact on the

environment and society. The research applies a novel approach and demonstrates through a real-world practical example how such a hybrid AI solution could increase production efficiency, which can also be applied in other fields. The AI model was tested in two test cases, and they are referred as model 1 and model 2.

The manuscript is organized as follows: Section 1 provides a brief overview of the current state of the art of optimization problems for energy generation production systems using technology and artificial intelligence. Section 2 describes the methods employed in the hybrid AI solution, followed by Section 3, which discusses the specificities of the dataset and the variables. Section 4 describes the implementation steps of the model (data preparation, DEA, ML model creation, and GA), followed by Section 5, where the results are presented. Section 6 discusses the results of the final model by comparing it with the current way of working that uses conventional control-loops based process control, highlighting the achievable efficiency increments and the implementation boundaries. Section 7 concludes the article with benefits, limitations, and potential future steps of the research.

1.1. Literature review

This proposed AI solution uses DEA, ML and GA. DEA a widely utilized, well-established nonparametric method for estimating the relative efficiencies from observed, historical data by using linear programming to identify the distance of each decision-making unit (DMU) to the frontier, composed of a set of the most efficient DMUs (Ahmed et al. 2019). Moreover, ML is a multidisciplinary subject that combines knowledge of probability theory, statistics, approximation theory, convex analysis, and algorithm complexity theory. ML specializes in how computers simulate or implement human learning behaviours to acquire new knowledge or skills and reorganize existing knowledge structures to continuously improve their performance (Zhu, Zhu, and Emrouznejad 2020). GA is the last component of the proposed hybrid AI solution. GA is a well-known metaheuristic optimization algorithm inspired by the biological evolution process. The GA mimics the Darwinian theory of survival in nature by selecting the best set of variables (Bataineh, Kaur, and Jalali 2022).

In the literature review, Scopus database was examined as one of the main search engines for analysing relevant academic articles. Data collection was performed in August 2022, focusing on the previous 10 years of articles. The present study used the following keywords to search for the relevant research outputs: 'production efficiency', 'DEA' and 'machine learning'. The authors did not obtain any paper containing all three keywords. When the keywords were limited to two of the three previously mentioned, the following output was obtained:

The combination of 'production efficiency' and 'DEA' in the keyword search generated 331 papers, displaying an increasing trend in the number of documents published per year. Environmental science was the main subject domain area, followed by the agricultural domain. The business and management as well as the engineering subjects were also relevant. The top 10 articles (based on the number of citations) include econometrics-related papers on various industries (Ding et al. 2020; Han et al. 2021; Sağlam 2018; Wang and Feng 2021) and environmental problems, such as pollution treatment, waste disposal (Sun, Guo, and Wang 2019), environmental pollution and land surface damage (Ying, Chiu, and Lin 2019), are also discussed. From these articles, six detail Chinese environmental challenges.

The combination of 'production efficiency' and 'machine learning' in the keyword search produced 70 papers, also displaying an increasing trend over the years. The main subject areas were engineering and computer science. In the top 10 articles (based on the number of citations), the most popular topic was the application of deep neural networks in industry and agriculture (Zheng et al. 2019; Liu et al. 2020; Lee et al. 2019; He et al. 2021).

For 'DEA' and 'machine learning', keyword searches yielded 88 papers, again with an increasing trend in documents published per year. The main subject areas were computer science and engineering. In the top 10 articles (based on the number of citations), the most popular subject is related to the industry and financial sector. Salehi et al., (2020) combined DEA and the MLP approach to investigate and improve the adaptive capacity of resilient systems in a petrochemical plant. Mirmozaffari et al., (2020) suggest the combination of DEA and innovative clustering algorithms for DMU efficiency comparison of 24 cement companies. Tayal et al., (2020) apply DEA and

K-mean clustering to evaluate the total energy consumption and CO_2 emissions for an efficient sustainable facility layout. Li et al., (2017) extend the cross-sectional DEA models to time-varying Malmquist DEA to evaluate multiple decision-making units (DMUs or companies). Saberi et al., (2013) combine artificial neural networks (ANNs) and DEA in credit scoring. Zhong et al., (2021) propose a combination of ML algorithms and DEA to manage the problem of statistical noise in data. The researchers compared 15 ML algorithms, and the BPNN (backpropagation neural network) had the best performance in the DEA.

While the number of papers increased significantly in the last 4 years (doubled or produced even higher growth), research on production efficiency using a combination of DEA and ML is lacking. Key subject areas of the SCOPUS keyword search can be found in table 1.

Table 1. Key subject areas of the SCOPUS keyword search (N. B. one document may belong to multiple subject areas).

	production efficiency and DEA	production efficiency and ML	DEA and ML
Engineering	67	32	33
Environmental Science	109	7	8
Agricultural and Biological Sciences	88	13	2
Computer Science	31	27	39
Business, Management and Accounting	71	6	16
Social Sciences	55	4	10
Economics, Econometrics and Finance	63	-	4
Energy	51	6	10
Mathematics	31	3	21
Decision Sciences	29	3	17
Chemical Engineering	6	10	3
Total documents	331	70	88

To examine the identified articles more closely, Hafeez et al., (2020) used deep learning to forecast electric load and heuristics optimized ML model tuning. The model offers guidance for energy production systems by giving maximum capabilities of energy generation, therefore avoiding waste of energy that cannot be stored or dispatched to the grid. Similarly, Wen et al., (2019) and Zhou et al., (2020) aimed to predict photovoltaic power output by using ML methods and a GA to improve the prediction accuracy of the model, hence enabling the grid administrator to prepare the grid for energy loads and avoiding the waste of energy that has no grid or storage to be dispatched to. Merei et al., (2013) optimized the GA by changing the component sizes and model settings of solar/wind/diesel energy production systems with

different battery technologies. The results indicated that optimization is possible and economical.

Furthermore, Król and Ocłoń (2018) measured costs and energy efficiency for heat and energy generation by evaluating different approaches for combined heat and power (CHP) plants in Poland. In this kind of power plant, steam that would otherwise be wasted is used for heating purposes. In this case study, the cost efficiency of the CHP plant is improved if natural gas engines are used. Han et al., (2021) proposed a method for selecting input variables by applying DEA to measure the energy efficiency of a chemical process. Moreover, the approach of Xu et al., (2019) aimed at reusing heat in a solid-state thermoelectric generator (TEG), converting the waste heat to electricity using the Seebeck phenomenon. ML models have been widely utilized in the modelling, designing, and predicting for energy systems. Ten major ML models are frequently employed in energy systems: ANN, MLP, ELM, SVM, WNN, ANFIS, decision trees, deep learning, ensembles, and advanced hybrid ML models (Mosavi et al. 2019).

Combining ML with GA and DEA with ML has been widely discussed in the literature. DEA and ML were selected to measure and predict efficiency in manufacturing, finances and supplier selection (L. Liu, Huang, and Zhan 2019; Cheng et al. 2017; Hong, Leem, and Kim 2019; Salehi, Veitch, and Musharraf 2020). Furthermore, a GA was integrated into ML models to identify optimal features/hyperparameters (Badnjević et al. 2019; Di Noia et al. 2020; Ko et al. 2017). ML was utilized within the GA to explore advantages in accelerating searches. In the case of having a complex fitness function with high calculation costs, using an ML model as a fitness function could reduce the computing time.

Although combining two of the three methods (DEA, ML, GA) has been discussed and used in diverse research, combining DEA, ML, and GA has never been implemented, according to the literature review. Hence, this gap in the literature leads to a new field of application to be explored in this manuscript.

Moreover, combining two methods (DEA-ML, DEA-GA, or ML-GA) would result in very different capabilities. As mentioned, each combination can be used for different purposes, but none of them could be used effectively for production system optimization. It is worth mentioning that while GA can be used for production system optimization, its implementation

in non-linear systems is problematic where input and output relations are complex and cannot be easily described mathematically, leading to incorrectly optimized systems.

As proposed in this research, combining DEA-ML can address this problem by using AI to model the system mathematically based on historic data. It is worth mentioning that the proposed approach of using DEA-ML to predict efficiencies of production systems based on inputs, control parameters, and outputs was not found in previous studies.

Furthermore, since optimization algorithms rely on objective functions that describe input–output relations, their usage is not possible in complex systems that cannot be easily described mathematically as state-space equations. Hence, control loops are used in such cases for the controllable sub-systems, being the main competition to the proposed hybrid AI model. No other AI or optimization model was found in the literature for further comparisons.

The AI solution applied in this research is aimed at providing optimal input and process control settings for complex processes, in our real-world case for a thermoelectric power plant. Feedback loop control models, such as the proportional-integral-derivative (PID) controller model, are widely used for controlling and optimizing production systems. However, they may not be sufficient for controlling more complex processes, especially if they are non-linear systems or contain significant delay components, resulting in suboptimal settings. Modern control theory heavily relies on the state space representation, where a control system is described by a set of inputs, outputs, and state variables connected by a set of differential equations (Liu and Barabási 2016). The state space model can be described as the state and output equations, as shown in Equation 1.

$$\begin{cases} \dot{x}(t) = f(t, x(t), u(t); \theta) \\ y(t) = h(t, x(t), u(t), \theta) \end{cases} \quad (1)$$

Where the state vector $x(t) \in \mathbb{R}^N$ represents the internal state of the system at time t, the input vector $u(t) \in \mathbb{R}^R$ captures the known input signals, and the output vector $y(t) \in \mathbb{R}^R$ captures the set of experimentally measured variables. The functions $f(\cdot)$ and $h(\cdot)$ describe the behaviour of the complex system, and

Θ collects the system's parameters. Our method is aimed at approximating the function $h(\cdot)$, where the output $y(t)$ is estimated using data envelopment analysis. Feedback control loops, such as PID controllers that continuously calculate an error and apply a corrective action to minimize that error, will be used (Chia 2018; Moshayedi, Shuvam Roy, and Liao 2019), at lower levels of the system, controlling simple processes. The aim of the proposed hybrid AI system, therefore, is to improve the overall control process of the complex system by considering efficiencies, reducing arbitrariness, and consequently raising production efficiency. The comparison of achieved efficiency between the conventional system and the proposed AI model was used to validate the new approach.

2. Materials and methods

The objective of this study is to propose a hybrid AI solution that is capable of raising the production efficiency of a complex system, in our real-world case for a thermoelectric power plant that is controlled by conventional control loops. Data were collected from sensors of the plant, pre-processed, and a hybrid AI solution composed of DEA-ML-GA was applied to it, recommending the optimal production settings for a certain desired output. The performance of the proposed new hybrid AI solution was measured by comparing the efficiency achieved in the past with the predicted efficiency expected from applying the hybrid AI solution's recommended settings.

2.1. Dea

DEA can be performed by two different models, the DEA-CCR model (Charnes, Cooper, and Rhodes 1978) and DEA-BCC model (Banker, Charnes, and Wager Cooper 1984), where DEA-BCC should mostly be applied when DMUs are measured in heterogeneous conditions and scaling. For this research, the DEA-CCR model was selected considering the homogeneity of the DMU conditions and scaling. Moreover, DEA can be input- or output-oriented; for this research, the input-oriented DEA-CCR model was chosen to minimize the consumption of production inputs, and efficiency was calculated in that sense. CCR model can be described by Equation 2 (Mustafa, Ullah Khan, and Mustafa 2021)

$$\theta* = min\theta$$
$$\text{s.t.} \sum_{q=1}^{n} x_{pq}\lambda_q \leq \theta x_{po} p = 1, 2, 3, \ldots, k$$
$$\sum_{q=1}^{n} y_{rq}\lambda_q \geq y_{ro} r = 1, 2, 3, \ldots, l \tag{2}$$
$$\lambda_q \geq 0 r = 1, 2, 3, \ldots, l$$

Where x_{po}, y_{ro}, are the pth input and rth output of the DMU under study;

λ_q stands for those variables which indicate the effect of prominent factors which of efficient DMUs are used as reference by other DMUs to compare their efficiency with.

$\theta*$, is the comparative technical efficiency of DMU.

By comparing the efficiency of a specific unit with the performance of a group of similar DMUs, it is possible to create a rank of most efficient DMUs which can be later investigated, and the efficiency of inefficient DMUs can be increased by copying the behaviour of the most efficient ones. pyDEA (Raith, Rouse, and Seiford 2019), a software package developed in Python for conducting data envelopment analysis (DEA), was employed to apply the DEA-CCR model.

2.2. Machine learning

This research proposes the use of elastic net (LN method), gradient boosting (RT method) and support vector machine (RT method), which are well-known, successfully implemented ML methods. They are widely used to predict continuous data based on continuous dependent variables (Laref et al. 2019; Mokhtari, Navidi, and Mooney 2020; Touzani, Granderson, and Fernandes 2018).

Moreover, Robust Scaler, Min-Max Scaler, Standard Scaler, Max-Abs Scaler, Robust Scaler, Quantile Transformer (Normal), Quantile Transformer (Uniform), and Power Transformer, widely utilized pre-processing methods were selected as potential candidates for the AI pipeline (Singh and Singh 2022). Grid Search Cross Validation was done by dividing the dataset into multiple folds and training the model on each fold while using the remaining folds for validation based on some scoring metric. The employment of R-squared is chosen as one of the standard statistical measures to evaluate regression analyses in any scientific area, and this study complies with it (Chicco, Warrens, and Jurman 2021). Therefore, the performance of the models was evaluated based on R-squared, and the best performing combination of ML and pre-processing methods

were selected. Gradient boosting proved to be the best scoring model for each of the ML models trained.

Gradient boosting is a machine learning algorithm that has gained popularity in recent years due to its ability to handle complex datasets and to produce accurate predictions. It is an ensemble method that combines multiple weak learners to create a strong learner. The algorithm iteratively adds new trees to the model, with each new tree correcting the errors of the previous tree. It calculates the negative gradient of the loss function with respect to the predicted values and uses this as the target variable for the next tree. The final prediction is a weighted sum of the predictions of all the trees, where each tree is given a weight proportional to its contribution to the overall accuracy of the model. Gradient boosting can handle complex datasets and produce accurate predictions (Zhang et al. 2019).

Scikit-learn, a free ML software library for the Python programming language, was used to create the ML pipelines, including ML models and pre-processing methods.

2.3. Genetic algorithm

The basic elements of the GA are chromosome representation, fitness selection, and biological-inspired operators. Chromosomes are composed of a set of different values, which are referred to as genes, and represent one possible solution for the optimization problem. The fitness function is applied to assign a score for all the chromosomes in the population. The biological-inspired operators are selection, mutation, and crossover (Katoch, Singh Chauhan, and Kumar 2021).

In selection, the chromosomes are selected based on their score calculated by the fitness function; there are different selection methods, such as the rank-based selection technique, roulette wheel, and tournament selections. Since the rank-based selection technique led to better performance, it was applied to the hybrid AI model. In this strategy, individuals in the population are ranked first based on their fitness values, which reflect how well they solve the problem at hand. The ranking process provides a way to differentiate the individuals based on their relative fitness, allowing for the creation of a probability distribution that takes the ranking into account. The probability distribution is then used to select a set of individuals for reproduction, with the fittest individuals having a higher probability of being selected. However, this method also allows for

the selection of lower-ranked individuals, providing a chance for diversity in the population and avoiding premature convergence to sub-optimal solutions. This helps to balance the exploration and exploitation of the solution space, leading to a more thorough search for optimal solutions (Hussain, Shad Muhammad, and Nauman Sajid 2017; Orong, Sison, and Medina 2018).

The crossover operator, which is a random locus that changes the sub-sequences between two chromosomes to create offspring, is chosen. Mutation randomly flips chromosomes based on a probability rate, low mutations could lead to local optimum and very high mutations could lead to random non-optimal results, hence different mutations were applied, and results were checked to assert that there was convergence towards a global optimum (Lawal et al. 2021).

2.4. Z-score method

Outliers were removed from the dataset because they have the potential to distort the calculations and to lead to incorrect conclusions. Outliers are data points that differ significantly from the rest of the data, and they can arise from measurement errors, missing data or natural variability. Measurement error is a common problem in industrial IoT systems where a single event can lead to loss of data from a number of sensors (Y. Liu et al. 2020). Removing outliers from a dataset can improve the accuracy and reliability of statistical analysis and predictions. Outlier removal was done using z-score method. The Z-score is a statistical measure that shows how far a data point is from the mean of a dataset. It is calculated by subtracting the mean from a data point and dividing the result by the standard deviation. A Z-score of 0 indicates that the data point is exactly at the mean of the dataset, while higher Z-scores indicate data points that are further away from the mean. In statistics, a commonly used threshold for identifying outliers is a Z-score of ±3. This means that any data point that falls outside of 3 standard deviations from the mean was considered an outlier.

3. Database and production system variables

3.1. Production system variables

Two hybrid AI models were built and tested to evaluate the benefits of process optimisation. Four different types of sensor data are distinguished in the dataset containing information about the production system: *input parameters* are the main sources that feed the production system, *process control parameters* are settings used to adjust production, *state variables* are measurements resulting from the production system, and *output parameters* are the main result of the production process (Figure 1). The input efficiency of the production system can be defined as the ratio of inputs to outputs.

In our real-world example, the thermoelectric power plant, there are three inputs used for combustion, for both models: wood infeed A (t/h), wood infeed B (t/h), and oil infeed (kg/h). Wood A and Wood B receive the same raw material, but each feeds the furnace from different points. The process control parameters that are used to adjust the performance of the combustion process in our dataset are the fluid bed primary air pressure (bar), burner primary air pressure (bar), burner secondary pressure (bar), and burner tertiary air pressure (bar). State variables that describe the secondary outputs of the system are listed as follows: steam pressure (bar), steam temperature (°C), furnace temperature A (°C), furnace temperature B (°C), pre-ECO flue gas temperature
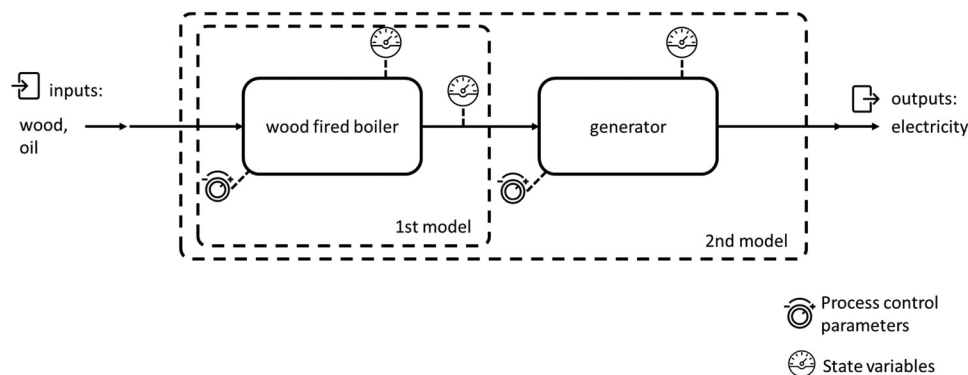

Figure 1. Conceptual model of the production system.

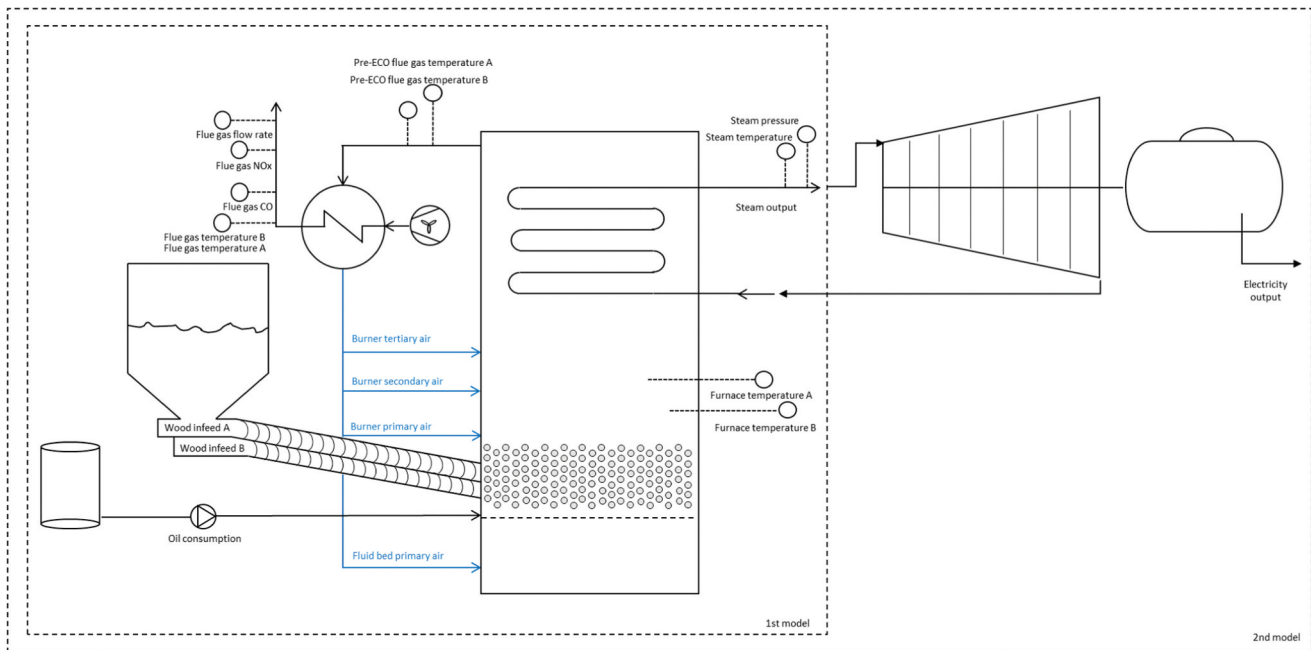**Figure 2.** Schematic diagram of the thermoelectric plant.

A (°C), pre-ECO flue gas temperature B (°C), flue gas temperature A (°C), flue gas temperature B (°C), flue gas NOx conc. (ppm), flue gas CO conc. (ppm), and flue gas flow rate (m3/h). Temperatures measured by different sensors located in different positions are alphabetically enumerated (A, B, etc.). These temperatures are consequences of the production process and are important indicators for the process, as they can act as constraints for production. Each of these indicators may have a range of desired values and running production with process indicators outside of their allowed range can cause accidents and/or environmental damage and should be avoided. The output parameter in the first model is the generated steam production (MJ/h), while the output for the second model is the generated electricity (MW/hr). The output of the first model (steam production with the corresponding steam pressure and steam temperature) could be taken as the additional process control parameter for the second model. Figure 2 illustrates the process plant of our real-world case, indicating the various sensor measurements.

3.2. Database structure and cleaning

The dataset contains values of sensor measurements (input parameters, output parameters, process control parameters, and state variables) of a thermoelectric

power plant as a regular time series dataset for every day and hour during 2020. As the data were not validated, a data cleaning process was needed as the first step of the research. Data duplications identified in the raw dataset were removed, data duplications could happen during the data collection due to errors in the information system and data warehousing. Moreover, measurement errors such as negative values or values close to zero did not reflect reality, these errors occurred when sensors were malfunctioning; hence, these values were also removed from the dataset. On the other hand, for some records in the dataset, the sum of production inputs measured by sensors (wood infeed A, wood infeed B, and oil infeed) was equal to 0, which seems to be the incorrect measurement, as it is impossible to generate electricity without using the mentioned production inputs; therefore, these cases were also removed from the dataset. Outlier removal was applied by the Z-score method for each set of sensor data types.

4. Proposed AI solution

In this research, a unique combination of DEA, GA, and ML was implemented, creating a hybrid artificial intelligence solution capable of optimizing the production efficiency of a thermoelectric power plant using manufacturing sensor data. All components of the hybrid AI solution have their own purpose: DEA estimates the

production system's efficient frontier and calculates DEA efficiency for all time periods – measured states, ML approximates the relationship between the input, output, process control parameters, and state variables and predicts production efficiency through simulation, finally, the GA proposes the optimal input and process control parameter settings corresponding to the desired output. Implementing such an AI solution can reduce production costs by optimizing resource usage and delivering economic and environmental benefits. The implementation of such a unique and new AI model can be utilized in other production plants, as a digital twin, providing decision support for the SCADA (supervisory control and data acquisition) system.

The proposed AI solution as applied to the thermoelectric power plant reads a dataset containing sensor measurements. Using domain expert knowledge, variables of the dataset were selected and classified into four data types: inputs, outputs, process control parameters, and state variables. Outliers were removed using the z-score method followed by the DEA efficiency calculation. Constant return to scale with free disposability assumptions were used for the DEA efficiency calculations. The DMUs in the DEA consist of hourly measurements of inputs and outputs. For modelling the plant's behaviour, multiple ML models were created with their respective hyperparameters and pre-processing methods to predict production efficiencies for different inputs and process control parameters. Corresponding production outputs and state variables were also predicted by the ML models created. At the same time, the prediction of production efficiency was performed also considering the previously predicted output in addition to inputs and process control parameter values. All prediction ML Models were validated, and the validations are presented in Section 5.2. Figure 3 summarizes the DEA and ML model training phases of the AI solution.

The GA component of the model searches for the feasible input and process control parameter combinations for any given output and predicts efficiencies to find the optimal one. The fitness function of the GA is the previously trained ML model, capable of predicting efficiencies. Additionally, state variables have a range of maximum and minimum desired values are considered as constraints in the GA. For each solution generated by the GA, state variables were predicted using previously trained ML models, and solutions with state variable
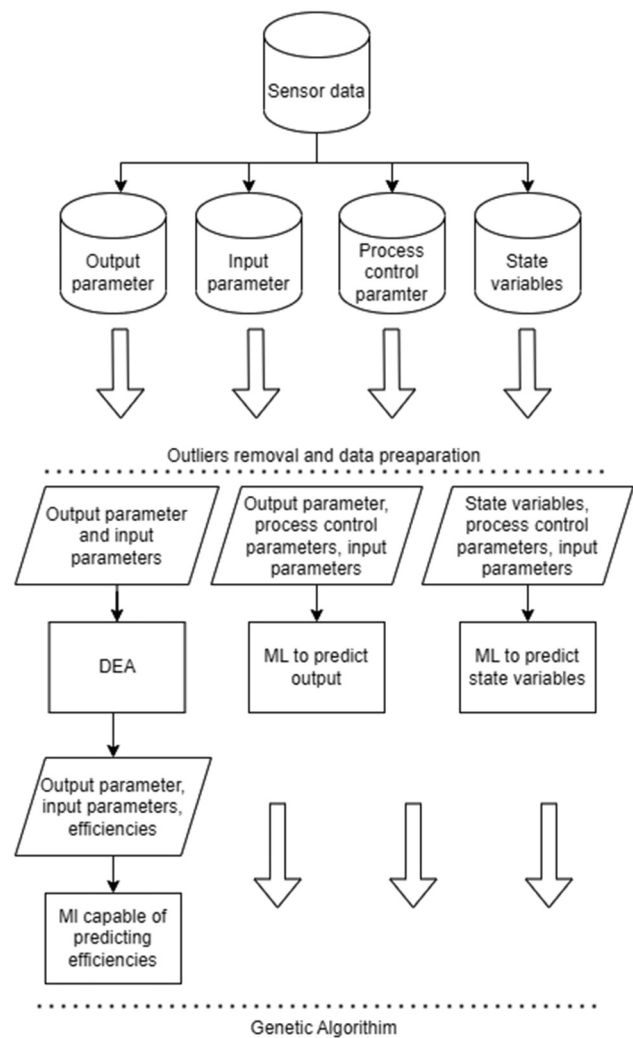

Figure 3. DEA efficiency estimation and ML training phases of the hybrid AI solution.

values beyond the desired range were considered unfeasible.

Furthermore, considering that in energy generation contracts, there are specific minimum values of energy to be generated and not achieving this minimum amount can lead to the breach of the contract, the predicted energy generation for each possible solution should always be greater than the contracted energy generation plan. Hence, if the GA solution predicts an energy generation slightly smaller than the desired/contracted output, this solution is considered infeasible. In contrast, if the predicted energy is greater than the desired/contracted output, the solution is considered feasible, and the excess energy production is considered waste and penalizes the efficiency of the solution. Figure 4 shows how the GA engine is programmed.
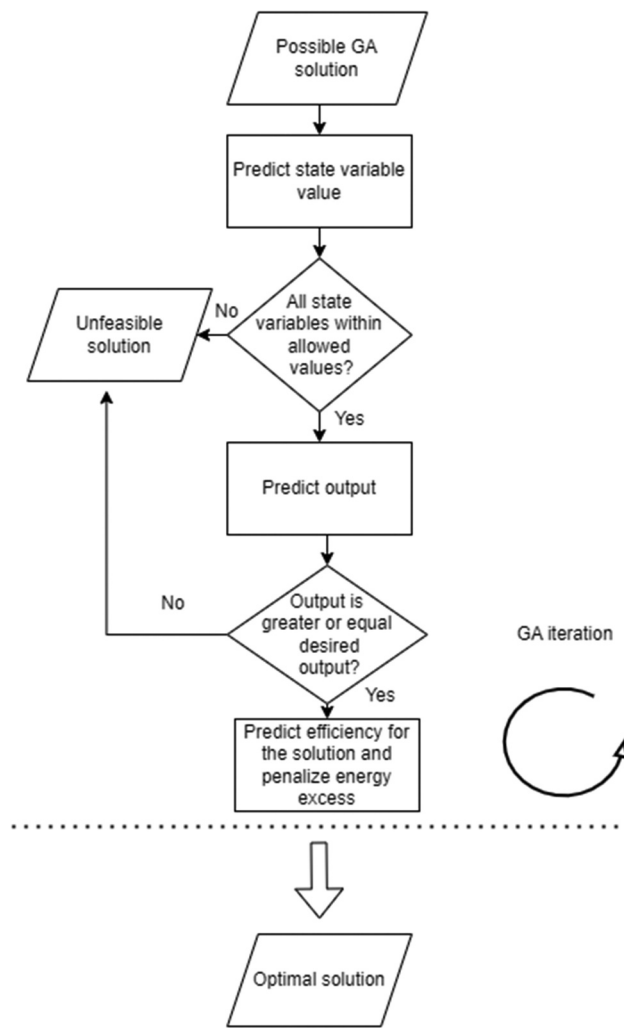
Figure 4. The Genetic algorithm optimisation phase of the hybrid AI solution.

Exceeding energy generation is penalized according to Equation 2, where efficiency is equal to 1 minus the division of planned output by the predicted output.

$$e = 1 - (Planned/Predicted) \tag{3}$$

The combination of the three main components of the model, DEA, ML and GA, works as a powerful AI solution capable of predicting the relative production efficiencies and searching different settings and configurations to select the optimal one (Cavalcanti, Kovács, and Kő 2022).

5. Results

5.1. DEA

For this research, DEA efficiency was calculated considering the inputs of the production system wood infeed A, wood infeed B, and oil infeed and the output the steam production for model 1 and electricity production for model 2. DMUs are measurements of inputs and output at a certain time, measured for each hour of the day. The result of DEA was appended to the raw dataset; hence, the dataset is extended to contain a new column with the DEA efficiency calculated for each record.

Since DEA only calculates relative efficiency, the best efficiency achieved may not mean the best possible efficiency, but the best efficiency identified in the historical dataset. Figure 5 shows the histogram of DEA relative efficiency scores throughout year 2020. The goal of the AI solution proposed in this research is to reduce efficiency loss and push it to as much as possible to 100%. It is worth to note that the distribution of efficiencies for model 2, with electric output is bimodal. The distribution with lower efficiencies was for the winter period. Further modelling may be required to separate the two distinct states and develop separate ML models.

The measure of the correlation coefficient provides information about the closeness of two variables; high/low correlations indicate high/low impact of the x variable on the y variable, whether causal or not (Senthilnathan 2019). To obtain a better overview of how different sensor data can positively or negatively correlate with each other and with the DEA efficiency, correlations between them were obtained. Figure 6 shows the correlation coefficients between selected inputs, the output, control parameters, state variables, and DEA efficiency for model 1 and 2, respectively. It is possible to determine which parameters have the most impact on each other and on the relative efficiency score by obtaining the highest positive and negative correlation values in the table. To improve the visualization, a red–blue colour scale was applied: blue means a higher positive correlation in contrast with red, representing high negative correlations.

The inputs wood and oil infeed have the strongest negative correlation with efficiency for both models; hence, a larger number of inputs resulted in lower efficiency. In addition to the input parameters, the process control parameters fluid bed primary air pressure (bar), burner secondary pressure (bar), burner tertiary air pressure (bar) showed a negative correlation with efficiency, in contrast with the burner primary air pressure (bar), which
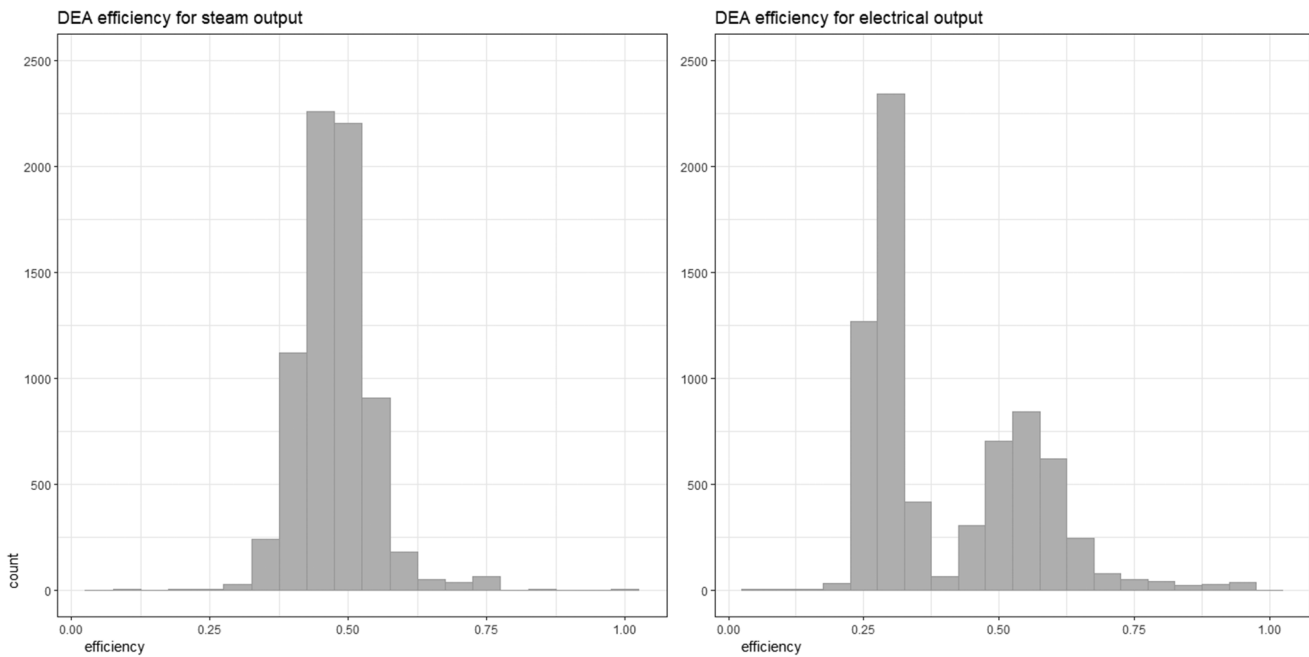
Figure 5. DEA relative efficiency distributions of the two models for year 2020.
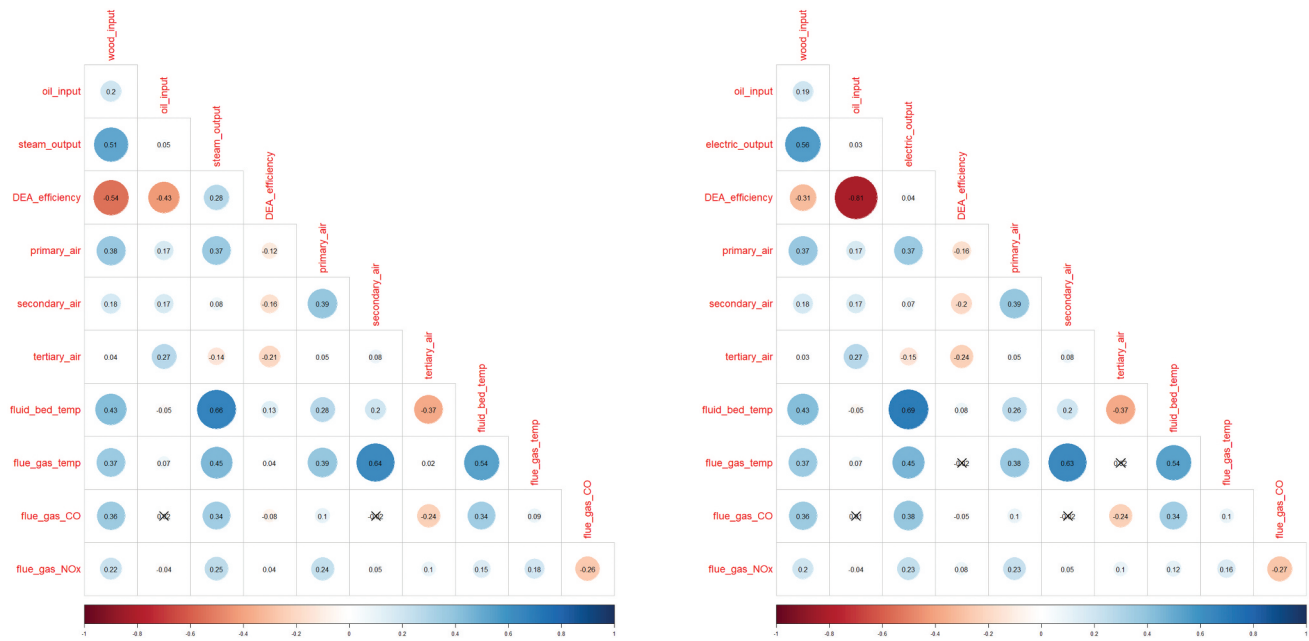
Figure 6. Correlation table of selected input, output, process control and state variables of the two models.

showed a slightly positive correlation with efficiency. The state variable flue gas temperatures showed a considerable positive correlation with efficiency, indicating that the efficiencies are higher when these gas temperatures are increased. Another interesting conclusion from the correlation table is that while the steam production is positively correlated with efficiency the correlation between electric output and DEA efficiency is almost insignificant. However, as the production system has output limitations, increasing production output deliberately is not possible.

To better understand and illustrate how certain inputs, process control parameters and state variables interact, four-dimensional charts were produced including DEA efficiencies. Figure 7 illustrates how

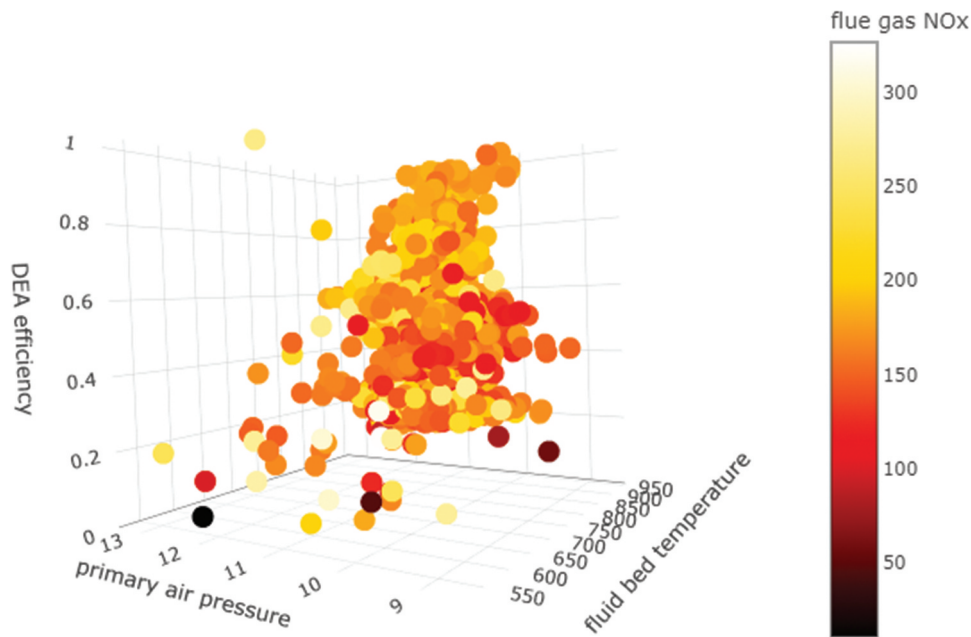**Figure 7.** Relationship between process control and state variables as well as DEA efficiency achieved.

a process control parameter: the primary fluid bed air pressure affects DEA efficiency and the state variables: furnace fluid bed temperature and the NOx concentration of the flue gas (visualised with the colour scale ranging from dark brown through red to yellow). It looks like that the highest DEA efficiencies are achieved at a certain combination of fluid bed pressure and fluid bed temperature with a corresponding, optimal flue gas NOx concentration. Both lower and higher NOx concentrations indicate lower efficiencies. The GA phase of the hybrid AI model will search for the optimal production setup and recommend input and process control parameters within the constraints of desired output and state variables.

5.2. Machine learning models

After the DEA calculation, three kinds of ML models were created to predict production output, state variables, and production efficiency. Production output and state variables were predicted based on inputs and process control parameter values. On the other hand, the prediction of production efficiency was performed considering the previously predicted output in addition to inputs and process control parameter values.

The AI solution uses grid search to search various combinations of pre-processing methods (robust scaler, min-max scaler, standard scaler, max-abs scaler,

robust scaler, normal quantile transformer, uniform quantile transformer and power transformer), ML methods (gradient boosting, elastic net and support vector machine) and hyperparameters' possible values. Grid search is a hyperparameter optimization method that performs a complete search over a given subset of the hyperparameter space and selects the best solution based on a performance score. This subset of parameters has a well-known possible range of values for each ML model, and the process of choosing the optimal value is an empirical process; therefore, a grid search of these possible values is a common practice (Shekar and Dagnew 2019). The proposed AI solution extends the capability of grid search and uses it not only for hyperparameter optimization but also for ML and pre-processing method selection.

R-squared (R^2) was the chosen scoring method for the grid search for the optimal model. In addition, the mean squared error (MSE) was calculated to further understand the model's errors. The hyperparameter space used for the grid search varied depending on the ML model. Gradient boosting had possible hyperparameter learning rates of 0.05, 0.1, 0.15, and 0.2. Moreover, the elastic net model had possible hyperparameters alpha and L1 ratio, constituting a linear space from 0 to 1 and logarithm space from −1 to 1, respectively. The support vector machine, on the other hand, had c and gamma hyperparameter values

of 0.1 or 2. All possible combinations of models, pre-processing methods, and hyperparameters were tested, R scores were measured for each possible combination, and the best scoring models were selected, producing a set of optimal models capable of predicting steam outputs, relative efficiencies, and state variables.

Gradient boost was the method that had the best performance for each trained model; hence, predicting efficiency, steam/electricity output and state variables will use gradient boosting as the base method. Table 2 shows the learning rate, optimal pre-processing method, and R^2/MSE scores of each trained ML model. ML models for predicting state variable values can be used for both models 1 and 2, on the other hand, predicting efficiencies and outputs (steam and electricity) requires distinct ML models.

The optimal pre-processing method had some variation; on the other hand, the learning rate had, in the majority, a value of 0.2 with a few cases of 0.15 R^2 scores. Some low R^2 values were observed for flue gas NOx conc., flue gas CO conc., steam pressure, and steam temperature. However, considering that the possible range of values constraining the production is quite large compared with the MSE and that the production constraints are flexible, meaning that certain deviations from the ideal range do not cause such an impact, the models can be considered to have sufficient performance for this specific production system. Since feature selection for the models was performed using expert domain knowledge, raising the efficiency of those MLs can be achieved by adding new unknown, relevant independent variables at the cost of more calculation time. Another way to raise the performance of ML models is to increase the

possible values space for hyperparameters, such as the learning rate, or even to test different ML models. However, as previously mentioned, these ML model performances were considered sufficient for this specific case.

5.3. Genetic algorithm

The final step of the model is the application of the GA. To run the GA, the min and max constraints of each input, process control parameter and state variable were needed, to create the gene space of the solution's possible values. Therefore, the GA only generated solutions with inputs and process control parameter values within the stated ranges.

For this research, a simulation was performed for a production demand of 168 MJ/h steam output for model 1 and electricity generation of 48 MW/hr for model 2. Each generation of the GA receives a set of chromosomes, of which it contains a set of genes that in this case, have different values of production inputs and process control parameters. Moreover, the fitness function was calculated for each possible solution to optimize efficiency. First, the steam/electricity output are predicted using the previously created ML model. If the predicted steam output is lower than the desired output demanded (168 MJ/h for model 1 and 48 MW/hr for model 2), the fit value of the chromosome is 0. Otherwise, the fit value is calculated by predicting efficiency based on the chromosome's genes that contain both production inputs and process control parameters, and the predicted steam output calculated by the previously mentioned ML model. Excess energy generation that surpasses the

Table 2. ML model optimal parameters and performance.

Dependent variable	Learning rate	Pre-processing method	R2	MSE
steam pressure (bar)	0.20	Power Transformer	0.2808	6.9069
steam temperature (°C)	0.20	Quantile Transformer (Uniform)	0.3029	5.4983
furnace temperature A (°C)	0.20	Min-Max Scaler	0.6103	646.2882
furnace temperature B (°C)	0.20	Power Transformer	0.5485	449.4722
flue gas temperature A (°C)	0.20	Quantile Transformer (Normal)	0.5168	319.0069
flue gas temperature B (°C)	0.20	Power Transformer	0.8686	27.0330
flue gas NOx conc. (ppm)	0.20	Max-Abs Scaler	0.2363	473.2964
flue gas CO conc. (ppm)	0.15	Quantile Transformer (Normal)	0.2533	252467.0000
pre-ECO flue gas temperature A (°C)	0.20	Min-Max Scaler	0.6534	71.1030
pre-ECO flue gas temperature B (°C)	0.15	Power Transformer	0.6915	74.9431
flue gas flow rate (m3/h)	0.20	Quantile Transformer (Normal)	0.5446	14.9431
steam production (MJ/h)	0.20	Quantile Transformer (Normal)	0.5217	157.6350
generated electricity (MW/hr)	0.20	Quantile Transformer (Normal)	0.6616	8.1102
Predict efficiency (model 1)	0.20	Max-Abs Scaler	0.9960	0.000082
Predict efficiency (model 2)	0.20	Max-Abs Scaler	0.9981	0.000065

desired output penalizes the fitness function according to Equation 1.

For model 1, GA iterated for 1000 generations, while a similar simulation for model 2 required higher number of iterations, over 30,000 generations. A mutation rate of 10% was applied in both cases. Figure 8 shows the evolution of the fitness value over the generations of model 1 and model 2, respectively.

As demonstrated in Figure 8, for model 1, by the end of the 1000th generation, the GA reaches an efficiency value of 70%, which is equivalent to the fitness value. Moreover, the best solution was achieved before the 200th generation indicating a quick convergence. Further generations have small or no improvement in production efficiency. On the other hand, model 2 reaches an efficiency of around 93% by the end of the 10,000th generation indicating that longer time is required for convergence. Different mutation rates were used to test that the model did not converge to a local optimum. With different mutation rates, such as 10-20-30-25%, the algorithm approximately converged to the same efficiency value, reinforcing that the global optimum was potentially achieved (Hassanat et al. 2019).

After convergence, the GA selects the best chromosome containing a set of optimal genes. Table 3 shows the results for each input and process control parameter recommended by the AI solution for optimal efficiency for models 1 and 2.

According to the proposed hybrid AI, to achieve optimal efficiency and produce a steam/electricity output of 168 MJ/h and 48 MW/hr for models 1 and 2, respectively, the input parameters and process control parameters should be set as shown in Table 3. The GA also predicts the values of state variables, which is shown in Table 4 as an example for model 2.

6. Discussion

In the previous example, model 1 achieved an efficiency of 70%, while model 2 achieved an efficiency of 93% exceeding the average what was achieved in the past, using traditional, loop control-based method. Therefore, it indicates that the proposed hybrid AI method is superior to the traditional, loop control-based method. The past average efficiency for model 1 when the steam output ranged from 167 to 169 MJ/h was 45.4% and for model 2 when the electricity output ranged from 47 to 49 MW/hr the efficiency achieved was 37%, reinforcing that the hybrid AI solution can increase efficiency (by 25% for model 1 and 58% for model 2) and reduce waste, if implemented. Furthermore, the predicted state variables for the recommended input and process control settings are all within the desired ranges, satisfying the production requirements (Table 4).

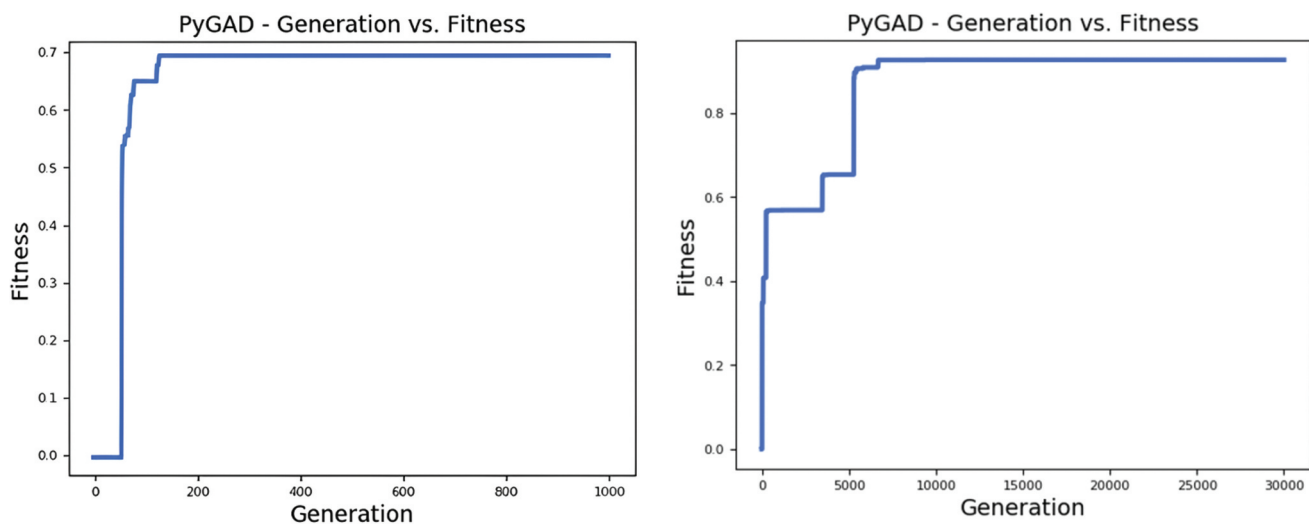Once the DEA efficiency estimation and the ML training phases are completed, optimal input and

Figure 8. Improvement of the fitness through generations of GA.

Table 3. Optimal recommendations for an output of 168 MJ/h and 48 MW/hr (model 1 and 2).

		Recommendation for optimal efficiency Model 1 for 168 MJ/h	Recommendation for optimal efficiency Model 2 for 48 MW/hr
Input	wood infeed A (t/h)	32.4	28.30
	wood infeed B (t/h)	39.6	28.56
	oil infeed (kg/h)	0.0	0.0
Process control	burner primary air pressure (bar)	0.0	0.00084
	burner secondary pressure (bar)	0.3	0.28
	burner tertiary air pressure (bar)	0.0	0.00091
	fluid bed primary air pressure (bar)	12.2	10.73

Table 4. State variable values for the optimal solution.

State variable	Value for the optimal solution (model 1)	Value for the optimal solution (model 2)	Desired Min. value	Desired Max. value
Steam pressure (bar)	96.8	97.2	95	102
Steam temperature (°C)	532.1	534.3	532	535
Furnace temperature A (°C)	824.1	799.8	None	900
Furnace temperature B (°C)	824.4	787.9	None	900
Pre-ECO flue gas temperature A (°C)	432.2	414.2	None	None
Pre-ECO flue gas temperature B (°C)	434.4	43.7	None	None
Flue gas temperature A (°C)	15.2	181.3	120	None
Flue gas temperature B (°C)	137.9	123.7	120	None
Flue gas NOx conc. (ppm)	198.2	184.2	None	200
Flue gas CO conc. (ppm)	1083.2	348.6	None	5000
Flue gas flow rate (m3/h)	66.5	62.9	None	None

process control settings can be calculated for multiple desired outputs. Therefore, GA phase can be decoupled from the DEA and ML model phases and the GA can be run multiple times to search for optimal solutions. This decoupling considerably reduces the calculation time, as the most time-consuming calculations are performed for the DEA and ML phases.

Calculation time was measured and is shown in Table 5. GA calculation time increased significantly as the number of generations went from 1000 to 30,000. Caching results for different ranges of outputs can reduce the time for the production setting recommendations, generating a table containing optimal settings for the most common outputs will eliminate the need of recalculating same results repeatedly. Another consideration about calculations time is that the time can vary depending on the system used for the calculation. For this research, a standard commercial notebook was utilized to test the model.

Figure 9 shows the efficiency that could be achieved with the hybrid AI solution in comparison with what was achieved using traditional loop control-based method, for different desired steam/electricity outputs for models 1 and 2. The dotted line shows the estimated efficiency that could be achieved using the hybrid AI solution that corresponds to the DEA efficiency frontier, while the dots represent the efficiency archived without it. The difference between the black and the grey dots is, that the grey ones represent observations, where constraints were violated for any of the state variables. Subsequently, they were excluded from the solution space. The area between the dots and the dotted line indicates how much efficiency increase could be achieved using the hybrid AI solution. It can be observed that the hybrid AI solution searches for the most efficient settings, which was achieved in the past and giving recommendations that leads to the most efficient DMUs located on the efficiency frontier. The area above the dotted line represents the waste of efficiency regardless of the usage of the hybrid AI solution. As seen in Figure 9, the efficiency improvements are significant using the hybrid AI solution. Moreover, it can be observed that there were multiple occasions when the production system operated in non-ideal conditions violating the state variable constraints, the AI solution not just

Table 5. Calculation time of the AI solution for models 1 and 2.

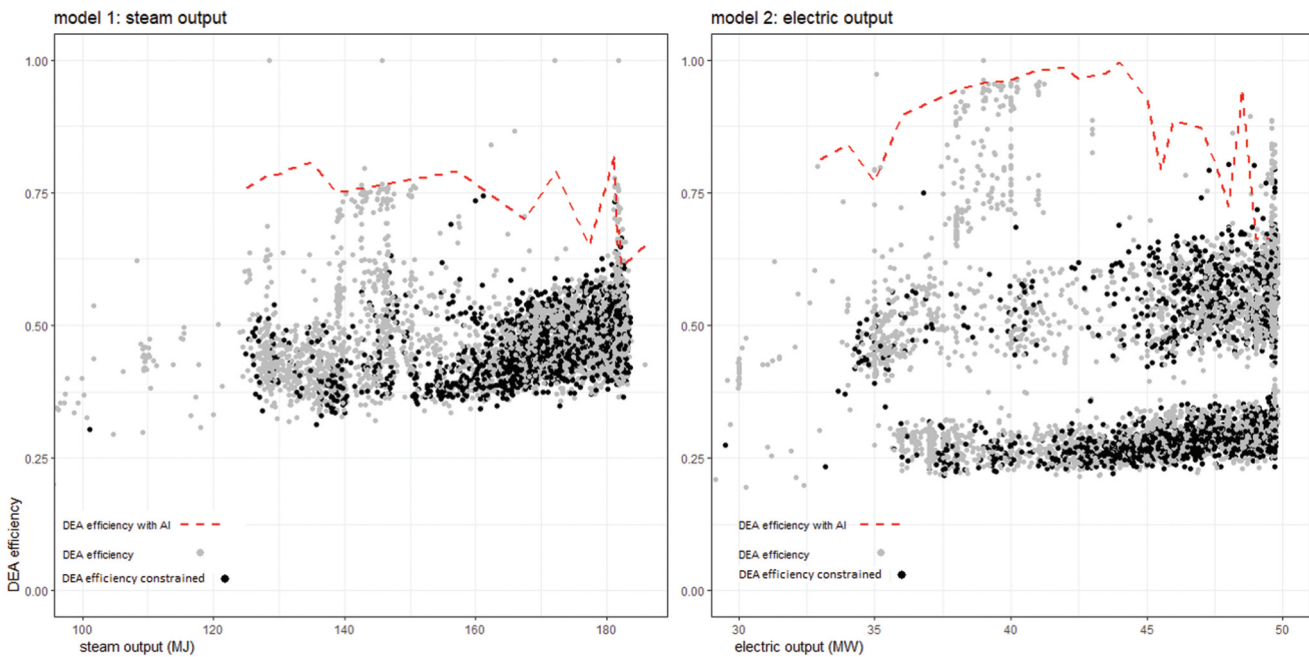| Model	DEA Calculation Time	ML Model Calculation Time	GA Calculation Time
Model 1	2906.49s	1965.89s	248.86s
Model 2	2681.01s	2925.68s	3285.34s

Figure 9. Efficiencies with and without the AI solution for model 1 and 2 (grey points are observations violating state variable constraints).

increase efficiency but guarantees that the maximum efficiency is achieved without compromising state variables constraint. Investigating the area above the dotted line and trying different approaches to reduce it could lead to further improvements in production performance. Another interesting pattern visible on Figure 9 is that with the use of the hybrid AI solution, the correlation between efficiency and steam output diminishes, therefore low steam outputs can be as efficient or even more than running production at full capacity, producing high steam outputs. This gives more flexibility to the production system, allowing management to decide how much to produce without being concerned with getting penalised with higher waste. This research therefore demonstrated the benefits of the hybrid AI system, being able to optimise input and control parameters settings, leading to more efficient production, and consequently reducing production costs and resource usage compared to the tradition loop control-based model. Considering that thermoelectric generated energy has a relatively high environmental impact, efficient usage of inputs can reduce CO_2 emissions, leading to a cleaner production system, benefiting individuals and the society.

7. Conclusion

Although combining DEA-ML and ML-GA has been widely discussed for different purposes in the literature, the combination of the three methods was not identified in the literature. Neither general AI models capable of optimizing production systems by recommending inputs and production settings were found. In this research, a hybrid AI solution was proposed combining DEA-GA-ML that is capable of learning from historic manufacturing sensor data and optimizing the production efficiency. The viability of the proposed hybrid AI solution was tested on the real-world data of a thermoelectric power plant. The comparison of the hybrid AI solution with the traditional loop control-based methods indicated, that it can improve the relative efficiency of the plant by some 25–58%, reducing production costs and resource usage, hence resulting in economic and environmental benefits.

From implementation aspects, considerations should be given to calculation times as there might be limitations for larger models with more inputs, process control and state variables as well as higher numbers of observations. These limitations might be overcome by decoupling the DEA-ML phases of the proposed hybrid AI solution from

the GA phase. Running the DEA-ML phases less frequently could significantly improve the overall calculation time, with the penalty of leading to somewhat less accurate predictions, as some of the most recent data are not included in the ML model. The benefits of having a more up-to-date ML model, that is based on the most recent data should be balanced with the cost of calculation time. Reducing the dataset to include only the most recent sensor signals could also improve the calculation time, however it would impact prediction performance. Therefore, the size of the dataset size and the DEA-ML calculation frequency should be adjusted according to the requirements and the limitations of the production system.

Since the DEA-ML engine is trained on such historic data are the result of running the plant using the traditional loop control-based method, it consists of a set of inefficient production settings. Once the hybrid AI solution is used, the plant should run on settings that are based on the recommendations of the GA. While these settings are the best solutions based on the historic data, there could be better solutions with higher efficiency, that the DEA-ML engine could not identify. One way to overcome these limitations of the hybrid AI system is the implementation design of experiments (DOE). DOE could continuously generate a wide range of experimental data containing optimal and nonoptimal production settings, that can be utilized in training the DEA-ML engine, pushing the efficiency frontier further and leading to better relative efficiencies.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

ORCID

Joao Henrique Cavalcanti http://orcid.org/0000-0001-7738-731X

References

Ahmed, S., M. Z. Hasan, M. MacLennan, F. Dorin, M. W. Ahmed, M. M. Hasan, S. M. Hasan, M. T. Islam, and J. A. M. Khan. 2019. "Measuring the Efficiency of Health Systems in Asia: A Data Envelopment Analysis." *British Medical Journal Open* 9 (3): e022155. https://doi.org/10.1136/bmjopen-2018-022155.

Badnjević, A., L. Gurbeta Pokvić, M. Hasičić, L. Bandić, Z. Mašetić, Ž. Kovačević, J. Kevrić, and L. Pecchia. 2019. "Evidence-Based Clinical Engineering: Machine Learning Algorithms for Prediction of Defibrillator Performance." *Biomedical Signal Processing and Control* 54:101629. https://doi.org/10.1016/j.bspc.2019.101629.

Banker, R. D., A. Charnes, and W. Wager Cooper. 1984. "Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis." *Management Science* 30 (9): 1078–1092. https://doi.org/10.1287/mnsc.30.9.1078.

Bataineh, A. A., D. Kaur, and S. M. J. Jalali. 2022. "Multi-Layer Perceptron Training Optimization Using Nature Inspired Computing." *Institute of Electrical and Electronics Engineers Access* 10:36963–36977. https://doi.org/10.1109/ACCESS.2022.3164669.

Cavalcanti, J. H., T. Kovács, and A. Kő. 2022. "Production System Efficiency Optimization Using Sensor Data, Machine Learning-Based Simulation and Genetic Algorithms." *Procedia CIRP* 107:528–533. https://doi.org/10.1016/j.procir.2022.05.020.

Charnes, A., W. W. Cooper, and E. Rhodes. 1978. "Measuring the Efficiency of Decision Making Units." *European Journal of Operational Research* 2 (6): 429–444. https://doi.org/10.1016/0377-2217(78)90138-8.

Cheng, Y., J. Peng, Z. Zhou, G. Xin, and W. Liu. 2017. "A Hybrid DEA-Adaboost Model in Supplier Selection for Fuzzy Variable and Multiple Objectives." *IFAC-Papersonline* 50 (1): 12255–12260. https://doi.org/10.1016/j.ifacol.2017.08.2038.

Chia, K. S. 2018. "Ziegler-Nichols Based Proportional-Integral-Derivative Controller for a Line Tracking Robot." *Indonesian Journal of Electrical Engineering and Computer Science* 9 (1): 221–226. https://doi.org/10.11591/ijeecs.v9.i1.pp221-226.

Chicco, D., M. J. Warrens, and G. Jurman. 2021. "The Coefficient of Determination R-Squared is More Informative Than SMAPE, MAE, MAPE, MSE and RMSE in Regression Analysis Evaluation." *Peer Journal Computer Science* 7:e623. https://doi.org/10.7717/peerj-cs.623.

Ding, T., W. Huaqing, J. Jia, Y. Wei, and L. Liang. 2020. "Regional Assessment of Water-Energy Nexus in China's Industrial Sector: An Interactive Meta-Frontier DEA Approach." *Journal of Cleaner Production* 244:118797. https://doi.org/10.1016/j.jclepro.2019.118797.

Di Noia, A., A. Martino, P. Montanari, and A. Rizzi. 2020. "Supervised Machine Learning Techniques and Genetic Optimization for Occupational Diseases Risk Prediction." *Soft Computing* 24 (6): 4393–4406. https://doi.org/10.1007/s00500-019-04200-2.

Hafeez, G., K. Saleem Alimgeer, and I. Khan. 2020. "Electric Load Forecasting Based on Deep Learning and Optimized by Heuristic Algorithm in Smart Grid." *Applied Energy*

269:114915. https://doi.org/10.1016/j.apenergy.2020.114915.

Han, Y., S. Liu, Z. Geng, G. Hengchang, and Q. Yixin. 2021. "Energy Analysis and Resources Optimization of Complex Chemical Processes: Evidence Based on Novel DEA Cross-Model." *Energy* 218:119508. https://doi.org/10.1016/j.energy.2020.119508.

Hassanat, A., K. Almohammadi, E. Alkafaween, E. Abunawas, A. Hammouri, and V. B. Surya Prasath. 2019. "Choosing Mutation and Crossover Ratios for Genetic Algorithms—A Review with a New Dynamic Approach." *Information* 10 (12): 390. https://doi.org/10.3390/info10120390.

He, Z., T. Shi, J. Xuan, and T. Li. 2021. "Research on Tool Wear Prediction Based on Temperature Signals and Deep Learning." *Wear* 478:203902. https://doi.org/10.1016/j.wear.2021.203902.

Hong, H.-K., B.-H. Leem, and S.-M. Kim. 2019. "Using a Hybrid Model of DEA and Decision Tree Algorithm C5. 0 to Evaluate the Efficiency of Ports." *The Journal of the Korea Contents Association* 19 (7): 99–109. https://doi.org/10.5392/JKCA.2019.19.07.099.

Hussain, A., Y. Shad Muhammad, and M. Nauman Sajid. 2017. "Performance Evaluation of Best–Worst Selection Criteria for Genetic Algorithm." *Mathematics and Computer Science* 2 (6): 89–97. https://doi.org/10.11648/j.mcs.20170206.12.

Katoch, S., S. Singh Chauhan, and V. Kumar. 2021. "A Review on Genetic Algorithm: Past, Present, and Future." *Multimedia Tools and Applications* 80 (5): 8091–8126. https://doi.org/10.1007/s11042-020-10139-6 .

Ko, T., J. Hyuk Lee, H. Cho, S. Cho, W. Lee, and M. Lee. 2017. "Machine Learning-Based Anomaly Detection via Integration of Manufacturing, Inspection and After-Sales Service Data." *Industrial Management & Data Systems* 117 (5): 927–945. https://doi.org/10.1108/IMDS-06-2016-0195.

Król, J., and P. Ocłoń. 2018. "Economic Analysis of Heat and Electricity Production in Combined Heat and Power Plant Equipped with Steam and Water Boilers and Natural Gas Engines." *Energy Conversion and Management* 176:11–29. https://doi.org/10.1016/j.enconman.2018.09.009 .

Laref, R., E. Losson, A. Sava, and M. Siadat. 2019. "On the Optimization of the Support Vector Machine Regression Hyperparameters Setting for Gas Sensors Array Applications." *Chemometrics and Intelligent Laboratory Systems* 184:22–27. https://doi.org/10.1016/j.chemolab.2018.11.011.

Lawal, A. I., G. O. Oniyide, S. Kwon, M. Onifade, E. Köken, and N. O. Ogunsola. 2021. "Prediction of Mechanical Properties of Coal from Non-Destructive Properties: A Comparative Application of MARS, ANN, and GA." *Natural Resources Research* 30 (6): 4547–4563. https://doi.org/10.1007/s11053-021-09955-w.

Lee, S. Y., B. Adhi Tama, S. Jun Moon, and S. Lee. 2019. "Steel Surface Defect Diagnostics Using Deep Convolutional Neural Network and Class Activation Map." *Applied Sciences* 9 (24): 5449. https://doi.org/10.3390/app9245449.

Li, Z., J. Crook, and G. Andreeva. 2017. "Dynamic Prediction of Financial Distress Using Malmquist DEA." *Expert Systems with Applications* 80:94–106. https://doi.org/10.1016/j.eswa.2017.03.017.

Liu, Y.-Y., and A.-L. Barabási. 2016. "Control Principles of Complex Systems." *Reviews of Modern Physics* 88 (3): 035006. https://doi.org/10.1103/RevModPhys.88.035006.

Liu, Y., T. Dillon, Y. Wenjin, W. Rahayu, and F. Mostafa. 2020. "Missing Value Imputation for Industrial IoT Sensor Data with Large Gaps." *IEEE Internet of Things Journal* 7 (8): 6855–6867. https://doi.org/10.1109/JIOT.2020.2970467.

Liu, L., Y. Huang, and X. Zhan. 2019. "The Evolution of Collective Strategies in SMEs' Innovation: A Tripartite Game Analysis and Application." *Complexity* 2019:1–15. https://doi.org/10.1155/2019/9326489.

Liu, C., L. Le Roux, C. Körner, O. Tabaste, F. Lacan, and S. Bigot. 2020. "Digital Twin-Enabled Collaborative Data Management for Metal Additive Manufacturing Systems." *Journal of Manufacturing Systems*. https://doi.org/10.1016/j.jmsy.2020.05.010.

Merei, G., C. Berger, and D. Uwe Sauer. 2013. "Optimization of an Off-Grid Hybrid PV–Wind–Diesel System with Different Battery Technologies Using Genetic Algorithm." *Solar Energy* 97:460–473. https://doi.org/10.1016/j.solener.2013.08.016.

Mirmozaffari, M., M. Yazdani, A. Boskabadi, H. Ahady Dolatsara, K. Kabirifar, and N. Amiri Golilarz. 2020. "A Novel Machine Learning Approach Combined with Optimization Models for Eco-Efficiency Evaluation." *Applied Sciences* 10 (15): 5210. https://doi.org/10.3390/app10155210.

Mokhtari, S., W. Navidi, and M. Mooney. 2020. "White-Box Regression (Elastic Net) Modeling of Earth Pressure Balance Shield Machine Advance Rate." *Automation in Construction* 115:103208. https://doi.org/10.1016/j.autcon.2020.103208.

Mosavi, A., M. Salimi, S. Faizollahzadeh Ardabili, T. Rabczuk, S. Shamshirband, and A. R. Varkonyi-Koczy. 2019. "State of the Art of Machine Learning Models in Energy Systems, a Systematic Review." *Energies* 12 (7): 1301. https://doi.org/10.3390/en12071301.

Moshayedi, A. J., A. Shuvam Roy, and L. Liao. 2019. "PID Tuning Method on AGV (Automated Guided Vehicle) Industrial Robot." *Journal of Simulation and Analysis of Novel Technologies in Mechanical Engineering* 12 (4): 53–66.

Mustafa, F. S., R. Ullah Khan, and T. Mustafa. 2021. "Technical Efficiency Comparison of Container Ports in Asian and Middle East Region Using DEA." *The Asian Journal of Shipping and Logistics* 37 (1): 12–19. https://doi.org/10.1016/j.ajsl.2020.04.004.

Orong, M. Y., A. M. Sison, and R. P. Medina. 2018. "A New Crossover Mechanism for Genetic Algorithm with Rank-Based Selection Method." In *Proceedings of 2018 5th International Conference on Business and Industrial Research: Smart Technology for Next Generation of Information, Engineering, Business and Social Science, ICBIR 2018*. https://doi.org/10.1109/ICBIR.2018.8391171.

Raith, A., P. Rouse, and L. M. Seiford. 2019. "Benchmarking Using Data Envelopment Analysis: Application to Stores of a Post and Banking Business." In *Multiple Criteria Decision*

Making and Aiding, 1–39. Springer International Publishing. https://doi.org/10.1007/978-3-319-99304-1_1.

Saberi, M., M. Sadat Mirtalaie, F. Khadeer Hussain, A. Azadeh, O. Khadeer Hussain, and B. Ashjari. 2013. "A Granular Computing-Based Approach to Credit Scoring Modeling." *Neurocomputing* 122:100–115. https://doi.org/10.1016/j.neucom.2013.05.020.

Sağlam, Ü. 2018. "A Two-Stage Performance Assessment of Utility-Scale Wind Farms in Texas Using Data Envelopment Analysis and Tobit Models." *Journal of Cleaner Production* 201:580–598. https://doi.org/10.1016/j.jclepro.2018.08.034.

Salehi, V., B. Veitch, and M. Musharraf. 2020. "Measuring and Improving Adaptive Capacity in Resilient Systems by Means of an Integrated DEA-Machine Learning Approach." *Applied Ergonomics* 82:102975. https://doi.org/10.1016/j.apergo.2019.102975 .

Senthilnathan, S. 2019. "Usefulness of Correlation Analysis." *SSRN 3416918*. https://doi.org/10.2139/ssrn.3416918.

Shekar, B. H., and G. Dagnew. 2019. "Grid Search-Based Hyperparameter Tuning and Classification of Microarray Cancer Data." In *2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP)*, 1–8. IEEE. https://doi.org/10.1109/ICACCP.2019.8882943.

Singh, D., and B. Singh. 2022. "Feature Wise Normalization: An Effective Way of Normalizing Data." *Pattern Recognition* 122:108307. https://doi.org/10.1016/j.patcog.2021.108307 .

Sun, J., L. Guo, and Z. Wang. 2019. "Technology Heterogeneity and Efficiency of China's Circular Economic Systems: A Game Meta-Frontier DEA Approach." *Resources, Conservation and Recycling* 146:337–347. https://doi.org/10.1016/j.resconrec.2019.03.046.

Tayal, A., A. Solanki, and S. Preet Singh. 2020. "Integrated Frame Work for Identifying Sustainable Manufacturing Layouts Based on Big Data, Machine Learning, Meta-Heuristic and Data Envelopment Analysis." *Sustainable Cities and Society* 62:102383. https://doi.org/10.1016/j.scs.2020.102383.

Touzani, S., J. Granderson, and S. Fernandes. 2018. "Gradient Boosting Machine for Modeling the Energy Consumption of Commercial Buildings." *Energy and Buildings* 158:1533–1543. https://doi.org/10.1016/j.enbuild.2017.11.039.

Wang, M., and C. Feng. 2021. "The Consequences of Industrial Restructuring, Regional Balanced Development, and Market-Oriented Reform for China's Carbon Dioxide Emissions: A Multi-Tier Meta-Frontier DEA-Based Decomposition Analysis." *Technological Forecasting and Social Change* 164:120507. https://doi.org/10.1016/j.techfore.2020.120507.

Wen, L., K. Zhou, S. Yang, and L. Xinhui. 2019. "Optimal Load Dispatch of Community Microgrid with Deep Learning Based Solar Power and Load Forecasting." *Energy* 171:1053–1065. https://doi.org/10.1016/j.energy.2019.01.075.

Xu, X., W. Yongjia, L. Zuo, and S. Chen. 2019. "Multimaterial Topology Optimization of Thermoelectric Generators." In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, 59186: V02AT03A064. American Society of Mechanical Engineers. https://doi.org/10.1115/DETC2019-97934.

Ying, L., Y.-H. Chiu, and T.-Y. Lin. 2019. "Coal Production Efficiency and Land Destruction in China's Coal Mining Industry." *Resources Policy* 63:101449. https://doi.org/10.1016/j.resourpol.2019.101449.

Zhang, Z., Y. Zhao, A. Canes, D. Steinberg, and O. Lyashevska. 2019. "Predictive Analytics with Gradient Boosting in Clinical Medicine." *Annals of Translational Medicine* 7 (7): 152–152. https://doi.org/10.21037/atm.2019.03.29.

Zheng, Y.-Y., J.-L. Kong, X.-B. Jin, X.-Y. Wang, S. Ting-Li, and M. Zuo. 2019. "CropDeep: The Crop Vision Dataset for Deep-Learning-Based Classification and Detection in Precision Agriculture." *Sensors* 19 (5): 1058. https://doi.org/10.3390/s19051058.

Zhong, K., Y. Wang, J. Pei, S. Tang, and Z. Han. 2021. "Super Efficiency SBM-DEA and Neural Network for Performance Evaluation." *Information Processing & Management* 58 (6): 102728. https://doi.org/10.1016/j.ipm.2021.102728.

Zhou, Y., N. Zhou, L. Gong, and M. Jiang. 2020. "Prediction of Photovoltaic Power Output Based on Similar Day Analysis, Genetic Algorithm and Extreme Learning Machine." *Energy* 204:117894. https://doi.org/10.1016/j.energy.2020.117894.

Zhu, N., C. Zhu, and A. Emrouznejad. 2020. "A Combined Machine Learning Algorithms and DEA Method for Measuring and Predicting the Efficiency of Chinese Manufacturing Listed Companies." *Journal of Management Science and Engineering* 6 (4): 435–448. https://doi.org/10.1016/j.jmse.2020.10.001.