

## Article

# Feature Selection by Binary Differential Evolution for Predicting the Energy Production of a Wind Plant

Sameer Al-Dahidi <sup>1</sup>, Piero Baraldi <sup>2,\*</sup>, Miriam Fresc <sup>2</sup>, Enrico Zio <sup>2,3</sup> and Lorenzo Montelatici <sup>4</sup>

<sup>1</sup> Department of Mechanical and Maintenance Engineering, School of Applied Technical Sciences, German Jordanian University, Amman 11180, Jordan; sameer.aldahidi@gju.edu.jo

<sup>2</sup> Energy Department, Politecnico di Milano, Via Lambruschini 4, 20156 Milan, Italy; miriam.fresc@polimi.it (M.F.); enrico.zio@polimi.it (E.Z.)

<sup>3</sup> MINES-Paris, PSL University, CRC, 06904 Sophia Antipolis, France; enrico.zio@mines-paristech.fr

<sup>4</sup> Research Development and Innovation, Edison Spa, 20121 Milan, Italy; lorenzo.montelatici@edison.it

\* Correspondence: piero.baraldi@polimi.it; Tel.: +39-02-23996345

**Abstract:** We propose a method for selecting the optimal set of weather features for wind energy prediction. This problem is tackled by developing a wrapper approach that employs binary differential evolution to search for the best feature subset, and an ensemble of artificial neural networks to predict the energy production from a wind plant. The main novelties of the approach are the use of features provided by different weather forecast providers and the use of an ensemble composed of a reduced number of models for the wrapper search. Its effectiveness is verified using weather and energy production data collected from a 34 MW real wind plant. The model is built using the selected optimal subset of weather features and allows for (i) a 1% reduction in the mean absolute error compared with a model that considers all available features and a 4.4% reduction compared with the model currently employed by the plant owners, and (ii) a reduction in the number of selected features by 85% and 50%, respectively. Reducing the number of features boosts the prediction accuracy. The implication of this finding is significant as it allows plant owners to create profitable offers in the energy market and efficiently manage their power unit commitment, maintenance scheduling, and energy storage optimization.

**Keywords:** wind energy; prediction; feature selection; binary differential evolution; artificial neural networks; ensemble



**Citation:** Al-Dahidi, S.; Baraldi, P.; Fresc, M.; Zio, E.; Montelatici, L. Feature Selection by Binary Differential Evolution for Predicting the Energy Production of a Wind Plant. *Energies* **2024**, *17*, 2424. <https://doi.org/10.3390/en17102424>

Academic Editors: Mohammadreza Aghaei and Aref Eskandari

Received: 30 January 2024

Revised: 2 May 2024

Accepted: 15 May 2024

Published: 18 May 2024



**Copyright:** © 2024 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 transition from conventional fossil-fueled power plants to renewable energy sources (RESs), such as wind and solar, could bring with it service reliability issues that must be carefully considered [1]. The aleatory and intermittent nature of RESs complicates the matching of energy production to the load demand, which is fundamental for a reliable energy supply to consumers [2]. For this reason, it is important to predict the electricity production from RES plants, which can be performed based on weather data [3]. Accurate predictions allow for the formulation of profitable offers in the energy market and the efficient management of power unit commitment, load increment and decrement decisions, maintenance scheduling, and energy storage optimization [4].

Approaches for predicting energy production can be categorized as physics-based or data-driven [5]. Given the difficulty of developing accurate physics-based models that receive, as input, the weather forecast and provide, as output, the prediction of the energy production, artificial intelligence (AI) models built by considering historical weather data and corresponding real productions have become popular [6].

The selection of the weather features to be used as AI-model inputs can significantly influence the prediction accuracy. This problem, referred to as feature selection [7], is becoming very relevant in the era of big data, given the abundance of available information

with different levels of relevance for solving the specific problem, as “*We are drowning in information and starving for knowledge*” [8]. In the case of wind energy production, several weather features made available by different weather forecast providers, including pressure, temperature, and wind speed at different altitudes in various locations near the plant area, are typically available [9].

Feature selection methods can be classified as filters, wrappers, or embedded [10,11]. Filter methods score individual features or feature subsets based on “proxy measures” of the “relevance” of the features, computed considering general characteristics of the data [11]. Wrapper methods evaluate the goodness of a subset of features as the performance of the specific prediction model, typically measured in terms of prediction accuracy [11]. In wrapper methods, a search algorithm is used as a “wrapper” around the prediction model: the search engine searches for the best solution, i.e., feature subset, among all the possible feature subsets of the  $p$  available features by evaluating the performance of the associated model. During the search for the optimal solution, the accuracy of the prediction model obtained for each candidate solution is directly used as an evaluation function to compare the different solutions selected by the search engine [12].

Filter methods are generally computationally more efficient than wrapper methods because obtaining proxy measures from data is less time-consuming than developing and evaluating the performance of prediction models. For example, a filter feature selection approach based on the relief method was applied to wind velocity prediction in [13]. However, wrapper approaches achieve greater accuracy by tailoring the feature selection to the specific prediction model employed [12,14]. In contrast, filter methods ignore the selected features’ actual effects on the prediction accuracy of the model. A review of the application of filter and wrapper feature selection methods to energy production prediction is presented in Section 2. The main limitation of wrapper methods is that the AI models typically used for energy prediction are computationally intensive to build, and, therefore, they cannot be developed with multiple subsets of features, as required.

Embedded methods perform the feature selection task directly during the development of the prediction model by computing properly defined metrics [15]. Computationally, they perform better than wrappers because they provide integration between modeling and feature selection [15]. This can be accomplished, for example, by considering a two-objective function: maximization of the goodness-of-fit and minimization of the number of variables [16]. Examples of embedded methods are least absolute shrinkage and selection operator (LASSO) and elastic net, which build a linear model of the output based on the least-squares method and shrink to zero the smallest regression coefficients [17], and various decision-tree-based algorithms, e.g., classification and regression tree (CART) [18], random forest (RF) [19], and XGBoost [20]. These methods are not considered in the context of this work as they assume linearity of the prediction model, which is not realistic in the context of wind energy production prediction.

In the present work, a novel wrapper approach for selecting the optimal set of weather features to be used for wind energy prediction is proposed. Its definition requires the following:

- (a) An algorithm that efficiently searches candidate subsets of weather features (search engine);
- (b) A prediction model;
- (c) An evaluation function that measures the accuracy of the prediction models.

With respect to (a), the binary differential evolution (BDE) algorithm [21] is employed due to its simplicity and effectiveness in exploring the decision space. Its superiority to other evolutionary algorithms (EAs) in feature selection problems has been shown [22].

With respect to (b), ensembles of artificial neural networks (ANNs) for wind energy prediction provide more accurate and robust results than the individual models of the ensemble [23]. Specific to the same dataset used in this work, the mean absolute error (MAE) of an ensemble of echo state networks (ESNs) was 7.1–9.1% lower than that of the best single baseline model [24]. Similarly, reference [23] reports improvements of 9.2%, 8.7%, and 9.2% for the MAE, root mean square error (RMSE), and weighted mean

absolute error (*WMAE*) when using an ensemble of ANNs rather than the best single baseline model. The reason is that the diverse models of the ensemble enhance overall performance by complementing each other's errors and leveraging their strengths in different zones of the learning space while also overcoming their respective limitations [25]. In practice, developing an ensemble of prediction models entails addressing two issues: (i) the definition of the base models and (ii) the aggregation of their predictions. In this work, ANNs are used as base models, and their outcomes are aggregated using the median operator, which has been shown to be more robust than other statistical indicators, such as the mean, with respect to possible outlier predictions by individual models [26]. Diversity among the base models is obtained by using a bootstrap aggregating (BAGGING) algorithm, which trains each model using a different subsample of the training set [27].

With respect to (c), the performance metric used in this work for evaluating the accuracy of the prediction model is the *WMAE* [23]. *WMAE* provides an estimate of the average prediction error normalized with respect to the actual energy production, which allows for a comparison of the prediction accuracy when the production capacities change [23].

The original contributions of this work are three-fold:

1. The development of a wrapper feature selection approach based on the novel combination of BDE and an ensemble of ANNs. Since the computational efforts needed to develop an ensemble of ANNs is proportional to the number of individual models of the ensemble, the wrapper feature selection is performed using an ensemble made of a number of ANNs smaller than that of the final prediction model;
2. The utilization of weather features obtained from various providers as potential inputs for the prediction model, which is shown to be able to significantly boost the prediction accuracy.

The effectiveness of the proposed wrapper feature selection approach is verified by considering real data from a 34 MW wind power plant. The set of weather features includes the pressure, temperature, and wind speed at different altitudes taken at various locations near the plant area and obtained from two weather forecast providers.

The remaining part of this paper is organized as follows: In Section 2, the motivation for selecting the relevant features is stated, and the available feature selection techniques for wind energy prediction are recalled. Section 3 presents the proposed BDE-based wrapper feature selection approach for wind energy prediction. Section 4 illustrates the real case study of a 34 MW wind plant. Section 5 presents the results of the application to the real case study and compares the performance of the proposed approach with that of a model that considers the whole set of available weather features and the model currently used by the wind plant owners. Some conclusions and future recommendations are given in Section 6.

## 2. The Motivation for Feature Selection

The main motivations for feature selection are as detailed in [28]: (a) irrelevant features unnecessarily increase the complexity of the prediction problem; (b) noisy features can degrade the prediction accuracy and increase the risk of data overfitting; (c) the elimination of unimportant inputs allows for a reduction in the resources needed for collecting, storing, and processing the data; and (d) the physical interpretability of the prediction can benefit from a small number of features.

Several feature selection methods have been successfully applied in different fields, such as text learning, pattern recognition, genetics, and statistics [29]. The selection or not of a feature is typically encoded in terms of a binary variable that takes the value of 1 or 0, respectively. Therefore, when  $p$  features are available, the size of the search space is  $2^p$ . Since an exhaustive search that evaluates all the possible feature subsets is commonly impractical, an efficient search engine is needed.

Both filter and wrapper methods perform a search for the optimal feature subset in the space of all possible feature combinations. For this, they require a strategy to

be defined for the search. Three sequential search strategies can be distinguished [30]: (i) the forward selection (FS) strategy starts with a model composed of just one feature and sequentially (by adding one feature at a time) selects the feature that most improves the prediction model; (ii) the backward elimination (BE) strategy starts with a model formed by all the  $p$  features and sequentially removes the feature that has the smallest impact on the model performance; (iii) a hybrid form of these greedy algorithms, called hybrid stepwise-selection or bi-directional selection, which performs both forward and backward selections at each step and selects the best option of the two [31]. However, the sequential search strategies are characterized by a major drawback; the order of parameter entry (or deletion) affects the selected model [32]. To overcome this issue, the use of EA-based approaches [33], such as genetic algorithms (GAs) [34], the BDE algorithm [21], particle swarm optimization (PSO) [35], the coral reef optimization (CRO) algorithm [36], or a combination of these techniques, have been shown to be effective even if they are computationally more demanding. In practice, the main advantages of EAs are (i) their fast convergence to a near-global optimum, (ii) their superior global searching capability in complicated search spaces, and (iii) their applicability even when gradient information is not readily achievable.

With respect to the prediction algorithm to be used within the feature selection wrapper approach for the development of the prediction model, AI-based algorithms such as ANNs, extreme learning machines (ELMs), Gaussian processes (GPs), nearest neighbor searches (NNs), support vector regression (SVR), and RF are typically used [37,38].

#### *Feature Selection for Wind Energy Predictions*

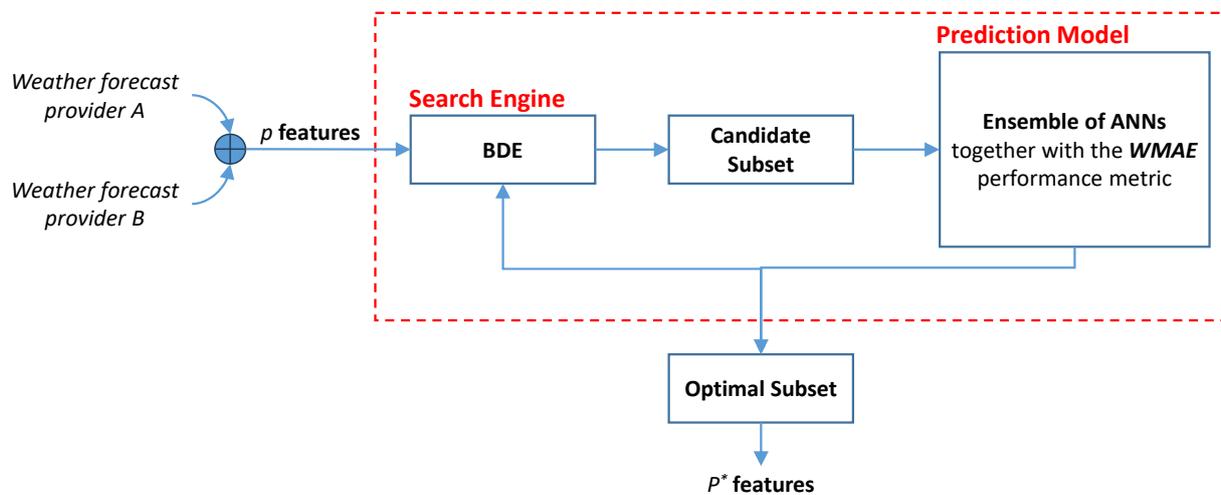
Considering the feature selection problem in the context of predicting the energy production of wind plants, Abdoos [39] proposed a hybrid approach, which combines variational mode decomposition (VMD) for the decomposition of the wind-power time series into different modes, Gram–Schmidt orthogonalization (GSO) for the elimination of redundant features, and ELMs for the prediction of the short-term wind power. Osório et al. [40] proposed a hybrid approach that combines evolutionary and adaptive techniques to forecast short-term wind power. The proposed approach integrates mutual information (MI) to select the most representative features from among the available wind power data, wavelet transform (WT) to break down the wind-power time series into components with reduced noise and an adaptive neuro-fuzzy inference system (ANFIS) to accurately estimate the wind power and whose hyperparameters are set using evolutionary PSO (EPSO). Jursa [41] proposed an approach for selecting features from among weather data obtained from a numerical weather prediction (NWP) model and measured the power data collected from various wind farms. Specifically, PSO was used as search engine and ANNs as prediction models. The work was extended in [42] using DE as search engine. Kou et al. [43] proposed an online adaptive ensemble model whose base models are multiple time-dependent warped Gaussian processes (WGPs) for the probabilistic prediction of wind power production. The input feature set and the length of time window for the historical wind speed data were dynamically selected by resorting to a sequential forward greedy search.

Differential evolution (DE) is one of the state-of-the-art methods for optimization [44,45]. The algorithm has been recently modified to improve its capability for finding the optimal solution and for reducing the computational burden in different application domains, such as for the optimization of the operational parameters of an aluminum friction–stir welding process of dissimilar materials (AA6061-T6 and AA5083-H112) [46], for the identification of parameters of photovoltaic models [47], and for the optimal positioning of flexible alternating-current transmission system controllers for reactive power management [48]. In this work, we focus on binary DE (BDE), a variation of DE specifically designed for problems with binary decision spaces. Note that despite the extensive research conducted in this field, wrapper feature selection approaches that combine a BDE algorithm as the search engine and an ensemble of ANN models as the prediction model have not yet been

developed for wind energy production prediction. In practice, employing an ensemble of ANN models is advantageous as it tends to yield more accurate predictions compared with individual models, as has been observed in various engineering applications [49,50]. Therefore, this research aims to improve the accuracy of wind energy production prediction by developing a wrapper feature selection approach combining BDE and an ensemble of models.

### 3. The Proposed Feature Selection Method

The proposed feature selection method is illustrated in Figure 1. It combines the BDE algorithm as the search engine (Section 3.1) and an ensemble of ANNs as the prediction model (Section 3.2). The weather features are collected by two different weather forecast providers, namely A and B, which predict weather features of different typologies and on different time scales.



**Figure 1.** The proposed BDE-based wrapper approach for wind energy prediction.

#### 3.1. Binary Differential Evolution (BDE) for Feature Selection

Given the relatively small number of weather forecasting features, i.e.,  $p < 100$ , which is typical of problems related to the prediction of wind energy production, we use a probabilistic search algorithm based on BDE [21,45].

BDE belongs to the family of evolutionary (or genetic) algorithms [21,45], which are optimization methods aimed at finding the global optimum of a set of real objective functions of one or more decision variables [22]. More specifically, BDE is a population-based optimization method, working iteratively through a wrapper algorithm [51].

In BDE, the search for the optimal solution is started by initializing a population of candidate solutions (artificial chromosomes,  $NP$ ) (Figure 2) [52]. New solutions are established by randomly varying existing ones through mutation (with a scaling factor denoted as  $SF$ ) and/or crossover (or recombination) (with a crossover rate denoted as  $Cr$ ) while verifying the performance of the prediction model via a *fitness* function [53]. Based on that, solutions are ranked, and those that will be maintained in the next generation are selected. The selected potential solutions are subjected to random variations, and the process will be iteratively repeated.

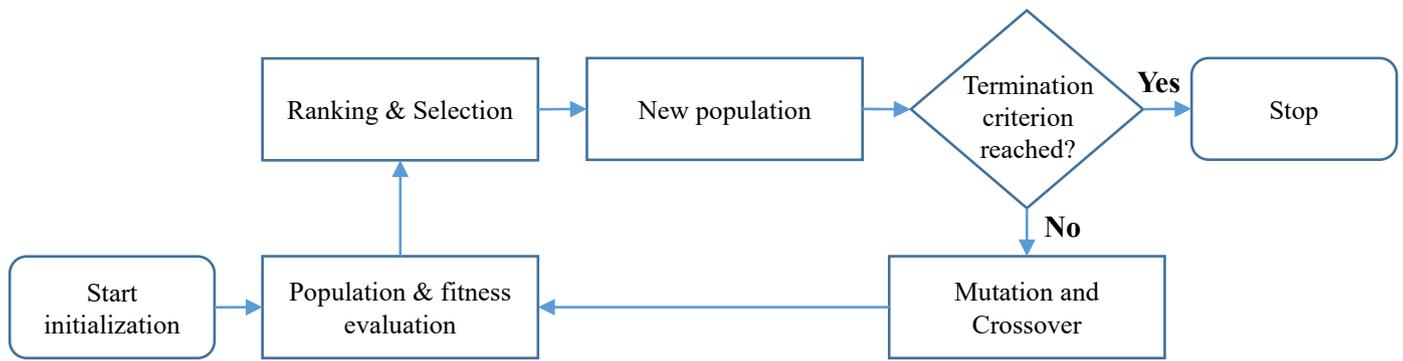


Figure 2. Flowchart of the BDE evolutionary algorithm.

Specifically, in feature selection problems, each candidate solution (an artificial chromosome) is typically represented as a vector of  $p$  binary bits/genes, which encodes the presence (1) or absence (0) of the features [54]. The BDE starts with an initial  $g$ -th population of candidate solutions. The candidate solutions are iteratively manipulated while verifying the predefined *fitness* function. The iterations continue until a predefined termination criterion is reached (e.g., a maximum number of iterations,  $G_{max}$ ) (refer to Appendix A for more details).

### 3.2. Ensemble of ANNs for Wind Energy Prediction

Ensembles of models have been used to improve the prediction accuracy and robustness of a single prediction model in various fields of application [55,56]. Particularly, in the field of wind energy prediction, the effectiveness of an ensemble of ANNs compared with individual ANN models was shown in [23].

An ensemble of models comprises multiple prediction models (called base models) whose prediction outcomes are aggregated into a final prediction outcome (Figure 3).

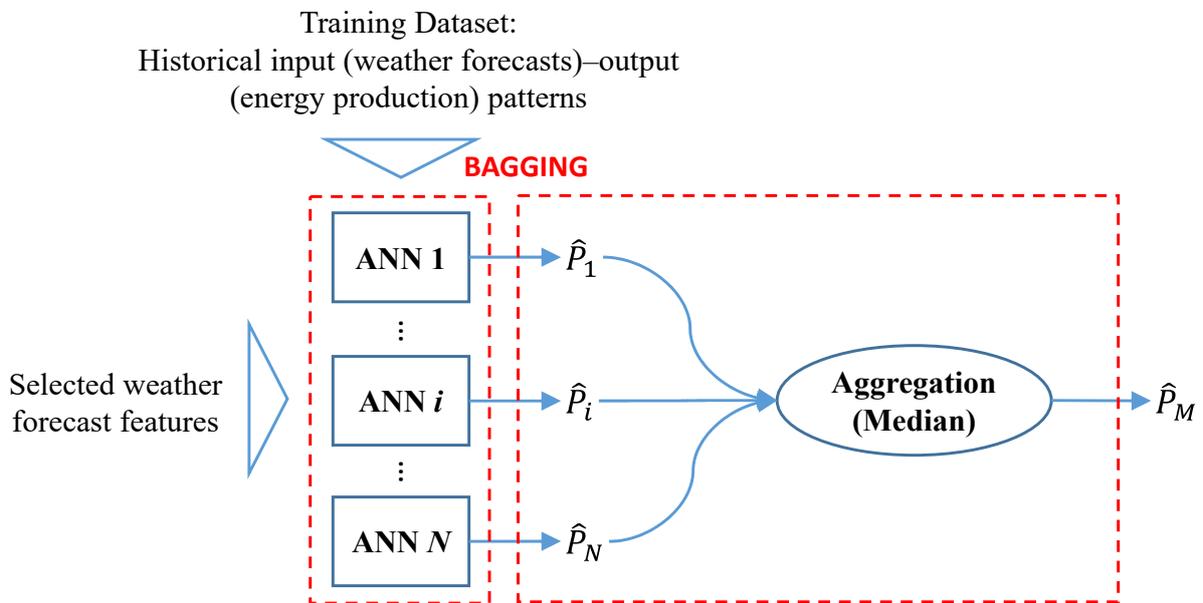


Figure 3. The ensemble of ANNs.

In practice, the development of an ensemble of prediction models requires the following [57]:

1. The generation of  $N$  diverse base models for leveraging their strengths and overcoming their drawbacks;

- The establishment of a strategy for aggregating the base models' outcomes,  $\hat{P}_i, i = 1, \dots, N$ , into a final outcome,  $\hat{P}_M$ .

In this work, feedforward ANNs are used as base models, and the diversity among the  $N$  base models is obtained by using the BAGGING technique [27,57]. In practice, the training set of each individual model was obtained by randomly sampling with replacement the number of patterns equal to that of the original training set.

To reduce the computational efforts needed to use ensembles of models, an ensemble made by a limited number of  $N_{reduced}$  ANNs is used during the BDE search. Specifically,  $N_{reduced} < N$  ANNs of the ensemble are selected so as to provide the smallest prediction error on a validation set made up of  $N_{val}$  patterns that are different from those used to train the models. Since the diversity of the models is guaranteed by the presence of  $N_{reduced}$  ANNs and the best-performing ANNs are selected, the performance of the ensemble is guaranteed while the computational burden is reduced.

The accuracy of the predictions is evaluated using the  $WMAE$  as the performance metric (Equation (1)), which corresponds to the relative prediction error [27]:

$$WMAE^m = \frac{\sum_{j=1}^{N_{test}^m} |\hat{P}^j - P^j|}{\sum_{j=1}^{N_{test}^m} P^j} \quad (1)$$

where  $WMAE^m$  is the  $WMAE$  computed considering the data corresponding to one month;  $P^j$  and  $\hat{P}^j$  are the true and predicted energy production of the  $j$ -th test pattern, respectively;  $N_{test}$  is the total number of input/output patterns of the test dataset; and  $N_{test}^m$  is the number of input/output patterns of the  $m$ -th month of the test dataset.

Given the seasonality of energy production from wind plants, the metric is computed as an average of the  $WMAE$  over 12 consecutive months (Equation (2)):

$$WMAE^{year} = \frac{\sum_{m=1}^{12} WMAE^m}{12} \quad (2)$$

The individual model outcomes are aggregated by calculating their median value to obtain the ensemble prediction [23] (Figure 3). The median operator is preferable to other statistical indicators, such as the mean, because it is more robust. This is due to the potential presence of individual models that provide predictions with significant errors on certain test patterns [27].

#### 4. Case Study

We consider the problem of selecting the best subset of weather forecast features to predict the energy production of a 34 MW wind plant [23]. The available  $p = 71$  weather features, collected from two weather forecast providers, here denoted as A and B, are hereafter described (Table 1):

**Table 1.** Weather features provided by the two weather forecast providers.

Provider	S	D	T	P	WG	RH	Height	Location	Typology
Provider A	✓	✓	✓	✓	✓	✓	10 and 100 m	4 different	Hourly
Provider B	✓	✓	✓	✓	✓	✓	10, 50, and 100 m	locations	Tri-hourly

✓ indicates the weather feature is available by the Provider.

- Twenty-four (24) weather features,  $x_k^A, k = 1, \dots, 24$ , forecasted every three hours by weather data provider A, corresponding to the wind speed (S) in the direction (D) from west to east ( $u$ ) and from north to south ( $v$ ); the temperatures (T) and pressures (P) at different heights and in different locations around the aerogenerators;
- Forty-four (44) weather features,  $x_k^B, k = 1, \dots, 44$ , forecasted every hour by weather data provider B, corresponding to the wind speed and wind gust (WG), i.e., a sudden,

brief increase in the wind, in two directions ( $u$  and  $v$  components); the temperature ( $T$ ), pressure ( $P$ ), and relative humidity ( $RH$ ) at various heights and in various locations different from those of provider A;

- Three (3) time features related to the calendar and the time of the prediction,  $x_k^{Time}$ ,  $k = 1, 2, 3$ , which are considered to account for the periodicity and seasonality of the energy production. They are the week number, the hour at which the prediction refers to, and its delay with respect to the time at which the production is predicted.

As reported in Table 1, the two weather forecast providers whose meteorological data have been used offer a large variety of weather features covering different locations and heights and referring to different time horizons of prediction. The large number of weather features ( $p = 71$ ) renders the feature selection task challenging because of the size of the search space. Furthermore, the partially redundant information content of some of the features complicates the search, which leads to the need to select those that allow the best performance to be obtained while eliminating the others. The proposed feature selection method is shown to be able to properly address these challenges by exploiting the capability of BDE to explore large feature spaces and that of a wrapper approach to select the most effective features in the case of partially redundant feature information content.

The available weather data and the corresponding hourly plant energy production refer to the period from January 2011 to December 2014. Alignment between the tri-hourly data of provider A and the hourly data of provider B was performed by considering only the tri-hourly timestamps. A forecast horizon of up to 4 days was used to train the ANN prediction models. The prediction performance was assessed for up to 1 day, which is the horizon of interest of the plant owners.

Among the available 24 weather features provided by provider A and the 3 time features, company experts selected, by trial-and-error, 19 features, which cannot be revealed for confidentiality reasons and will be referred to as “Benchmark 2” and used for comparison.

## 5. Results

Section 5.1 presents a statistical analysis of the correlation among the features, which was conducted to facilitate the interpretation of the results of the feature selection. Section 5.2 discusses the results achieved by applying the BDE algorithm, and Section 5.3 discusses the prediction performance obtained by the ensemble of ANNs.

### 5.1. Data Analysis

The correlation between the whole set of available  $p = 71$  weather features was investigated by applying the spectral clustering algorithm [58]. The aim was to identify groups of largely correlated features characterized by similar behaviors. The similarity among couples of features was evaluated by computing the pointwise difference with reference to an “approximately zero” fuzzy set defined by a bell-shaped function, which maps the pointwise difference to a similarity value. The parameter  $\sigma$  of the bell-shaped function was set to 9, in accordance with [59]. The following clusters of similar features were identified:

- Three clusters were made up of a single feature corresponding to the time (hour, delay and week of the prediction). As expected, these features have small correlations with all the others;
- A cluster consisting of 24 features corresponding to the horizontal wind speed at four different locations and two different heights provided by provider A and the horizontal wind speed and gust at four different locations and three different heights provided by provider B;
- A cluster consisting of 24 features containing the vertical wind speed at different locations and heights provided by both providers A and B;
- A cluster consisting of eight features containing the temperature measured at four different locations provided by both providers A and B;

- A cluster made up of eight features containing the pressure measured at four different locations provided by both providers A and B;
- A cluster made up of four features containing the relative humidity measured at four different locations provided by provider B.

The analysis has shown that the groups of correlated features are homogeneous from the point of view of the measured signals. In particular, groups of the  $u$  components of the wind speed, the  $v$  components of the wind speed, the temperature, the pressure, and the relative humidity are recognized. Each feature of a group is highly correlated with features of the same group and not correlated with features of other groups.

The analysis highlights the fact that the two providers provide redundant weather features. Therefore, it is expected that a reduction in the number of features to be provided as input for the prediction models may allow more accurate results to be obtained.

### 5.2. BDE Optimization for Feature Selection

The prediction model used within the BDE optimization is an ensemble of  $N_{reduced} = 10$  ANNs trained using the 2011–2012 data. The best-performing models were selected among  $N = 500$  models evaluated on  $N_{val}$  patterns of a validation dataset, which were different from those used to train the models. The choice to consider  $N = 500$  ANNs was derived by the solution currently adopted by the wind plant operator. Increasing the number of base models could improve the prediction accuracy, but up to a certain limit; beyond that limit, the performance gain becomes negligible, but the complexity of the model and computational resources associated to it would greatly increase. Undoubtedly, a good compromise solution between prediction accuracy and model complexity has been adopted by the plant operator.

With regard to the BDE search, the most critical hyperparameters affecting the robustness of the results are the number of chromosomes ( $NP$ ), the maximum number of generations ( $G_{max}$ ), the crossover rate ( $Cr$ ), and the scale factor ( $SF$ ). In this work, the values of these parameters have been set by trial-and-error considering the ranges suggested in [60,61]. The prediction performance for the 2013 data was evaluated using Equation (1). Table 2 reports the setting of the hyperparameters used in this work.

**Table 2.** BDE hyperparameters.

$NP$	$G_{max}$	$Cr$	$SF$	$Fitness$
100	1900	0.65	0.7	WMAE <sup>m</sup> Equation (1)

The set of features obtained from the BDE optimization is formed by  $p^* = 10$  features, whose detailed list is not reported here for confidentiality reasons. The selected features were forecast by both providers at various locations and at different altitudes. It is interesting to mention that the BDE selection confirms the choice made by the company expert of using only time and wind speed features for energy production prediction. Time features facilitate the identification of temporal patterns related to daily and seasonal trends in the wind behavior. Features related to wind speed are selected since the power generated by wind turbines is directly proportional to the cube of the wind speed [62]. Also, some of the wind gust features at different locations provided by provider B have been selected, since they allow short-term variations in wind speed to be anticipated.

The proposed approach has been developed in MATLAB<sup>®</sup> (version 2019) and the computational time needed on a high-speed computational cluster (with 20 nodes and 129.085 GB memory) is equal to 16 h. The computational demand is mainly due to the necessity of training an ensemble of ANNs for each chromosome of each generation. Note, however, that the feature selection is performed offline before the development of the ensemble prediction model for energy forecasting. The obtained result confirms the feasibility of using the method for predicting the energy production of the wind plant considered.

### 5.3. Prediction Performance

The final prediction model is an ensemble of  $N = 500$  ANNs that receives as input the optimal feature set identified by the proposed feature selection approach. Its performance has been computed considering two different partitions of the data in the training and test sets:

- Partition 1: data collected in the years 2011–2012 were used as the training set and data collected in the year 2013 were used as the test set to assess the prediction performance;
- Partition 2: data collected in the years 2012–2013 were used as the training set and data collected in the year 2014 were used as the test set to assess the prediction performance. Note that the verification of the performance on data for the year 2014 required the retraining of the ANNs with data taken from the previous two years. The plant owners followed this procedure to consider possible modifications of the plant behavior due to component replacement, deterioration, and maintenance activities.

The prediction performance is assessed by resorting to the mean absolute error ( $MAE$ ) in addition to the  $WMAE$  (Equation (1)). It corresponds to the average absolute error (Equation (3)):

$$MAE = \frac{\sum_{j=1}^{N_{test}} |\hat{p}^j - p^j|}{N_{test}} \quad (3)$$

where  $P^j$  and  $\hat{P}^j$  are the true and predicted energy production of the  $j$ -th test pattern, respectively, and  $N_{test}$  is the total number of input/output patterns of the test set.

It is worth mentioning that the  $WMAE$  metric (Equation (1)) differs from the  $MAE$  metric (Equation (3)) due to the presence of the total monthly production at the denominator. Therefore, if two months are characterized by the same  $MAE$  but different energy productions are considered, a larger  $WMAE$  is associated with the one with the lower production.

Figure 4 shows the  $WMAE$  (Figure 4a) and  $MAE$  (Figure 4b) performances of the  $N = 500$  ANNs ensemble obtained using as input the selected features (proposed), all the available 71 features (i.e., Benchmark 1), and the features currently used by the company (i.e., Benchmark 2).

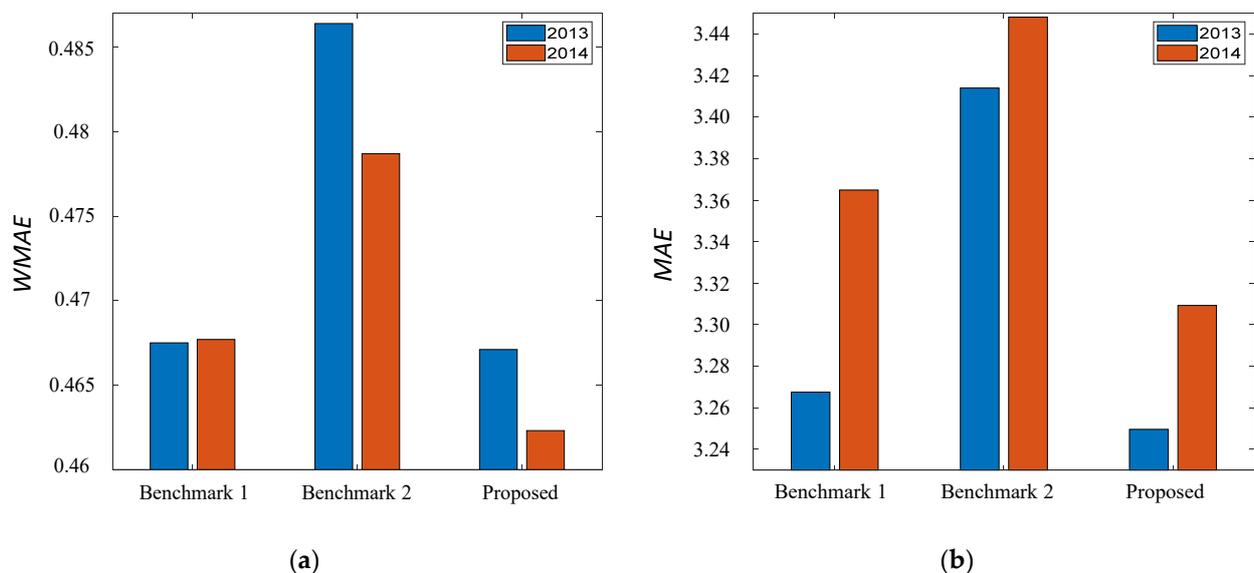


Figure 4. The (a)  $WMAE$  and (b)  $MAE$  for the test years 2013 and 2014.

The accuracy of Benchmark 2 is less satisfactory than that obtained by the other two models, which include features forecast by both providers. Overall, the 10 features selected by the proposed method allow for the development of the most accurate ANN ensemble model.

To effectively evaluate the enhancements obtained by the proposed approach with respect to the two-performance metrics, we define the performance gain ( $PG_{METRIC}$ ) associated with each performance metric (Equation (4)):

$$PG_{METRIC} = \left( \frac{METRIC_{Benchmark} - METRIC_{Proposed}}{METRIC_{Benchmark}} \right) \times 100\% \quad (4)$$

where  $METRIC_{Benchmark}$  is the performance metric obtained by considering the whole set of available weather features (Benchmark 1) or the weather features selected by the plant owners' experts (Benchmark 2), whereas  $METRIC_{Proposed}$  is the performance metric obtained using the selected weather features of the proposed approach.

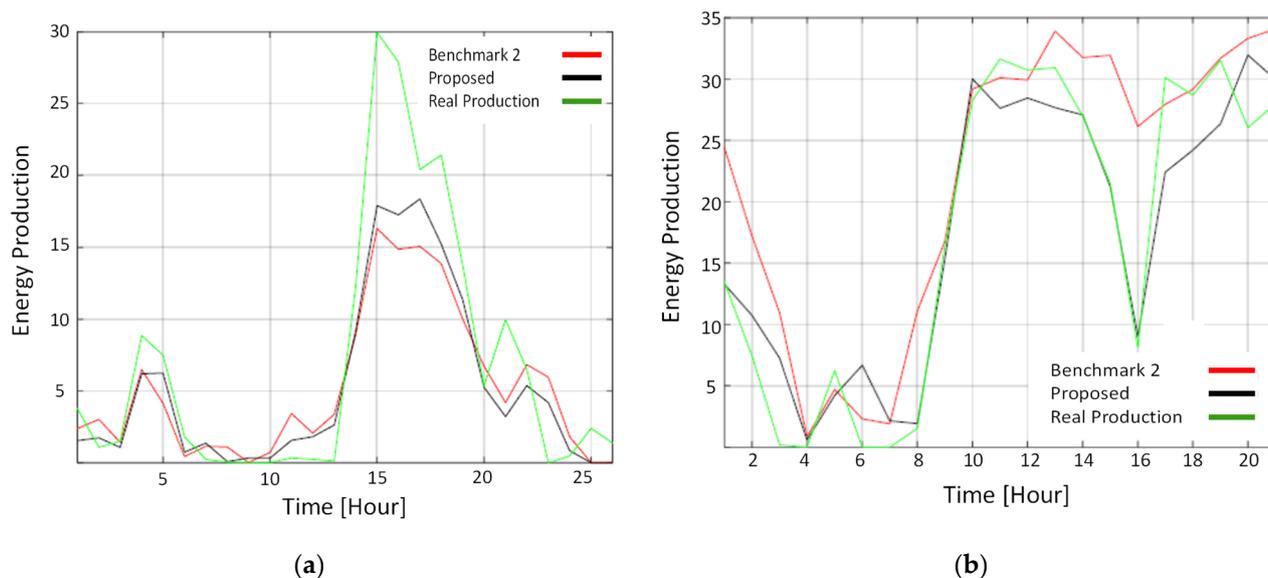
Table 3 reports the performance gains of the  $WMAE$  and  $MAE$  obtained by the proposed approach with respect to the approach that considers the whole set of 71 available weather features (Benchmark 1) and the approach that considers the 19 features selected by the plant owners' experts (Benchmark 2) for the 2013 and 2014 test sets. Positive values of the  $PG_{METRIC}$  indicate the superiority of the proposed approach to the use of the benchmarks. One can recognize the following:

- Considering the  $WMAE$ , the proposed approach outperforms Benchmark 1 by 0.06% and 1.18% for the 2013 and 2014 predictions, respectively. When considering the  $MAE$ , it performs 0.46% and 1.55% better for the 2013 and 2014 predictions, respectively. The obtained improvement in the prediction accuracy has been considered significant by the owners of the wind plants for the economic efficiency of their operation. Also, the results confirm that not all features are necessary for wind energy prediction, as some features contain redundant or irrelevant information that can negatively affect the training of the NNs. This is evident in Benchmark 1, where the use of all features causes the NNs to slightly overfit the training data, hindering their generalization to new data.
- Considering the  $WMAE$ , the proposed approach outperforms Benchmark 2 by 4.16% and 3.29% for the 2013 and 2014 predictions, respectively. When considering the  $MAE$ , it outperforms Benchmark 2 by 4.69% and 4.06% for the 2013 and 2014 predictions, respectively. This result demonstrates that the proposed wrapper approach outperforms human experts in the feature selection task.

**Table 3.** Performance gains obtained by using the proposed feature selection with respect to the two benchmarks for the 2013 and 2014 test datasets.

	With Respect to Benchmark 1 (71 F)		With Respect to Benchmark 2 (19F)	
	$PG_{WMAE}$ (%)	$PG_{MAE}$ (%)	$PG_{WMAE}$ (%)	$PG_{MAE}$ (%)
2013	0.06	0.46	4.16	4.69
2014	1.18	1.55	3.29	4.06
Mean	~0.62	~1.01	~3.73	~4.38

Figure 5 shows the actual energy production (green), the energy production predictions obtained by the approach adopted by the plant owners (red), and the proposed approach (black) of consecutive tri-hourly time steps during different days in December 2013 (Figure 5a) and December 2014 (Figure 5b). One can recognize the capability of the model to predict the minima and maxima of energy production based on the selected feature set. In contrast, the model based on the feature set selected by the plant owners failed in this task (e.g., at  $t = 16$  h in Figure 5b).



**Figure 5.** Examples of energy production predictions obtained by the model adopted by the plant owners and the proposed approach to the actual productions for the (a) 2013 and (b) 2014 data.

#### 5.4. Comparison with Other State-of-the-Art Feature Selection Techniques

Table 4 reports a list of works regarding feature selection in the context of predicting the energy production of wind plants. The performance of the feature selection methods is evaluated considering their gain in accuracy compared to the persistence forecasting method, which assumes that the wind energy production at the next time step is equal to the current energy production [40]. The gain, as defined by Equation (4), is computed considering various accuracy measures so as to facilitate the comparison across the feature selection methods applied in different case studies. For instance, the proposed wrapper feature selection approach applied to the 2013 and 2014 data achieves a performance gain of 59% and 60% when considering the *MAE* and of 50% and 51% when considering the *WMAE*, respectively, with respect to the persistence forecasting technique (i.e., when used as a benchmark in Equation (4)). The proposed approach is superior to the other wrapper approaches [41–43]. The filter approach proposed in [40], based on the use of the entropy measure, significantly outperforms all wrapper approaches in terms of the gain computed by the *NMAE* and Normalized *RMSE* (*NRMSE*). This unexpected finding [12,14] warrants further investigation, since the comparison whose results are reported in Table 4 is performed on different case studies. Future work will include directly applying the proposed feature selection method and that of [40] for the same case study.

**Table 4.** Comparison of the performance of the proposed feature selection approach with other state-of-the-art techniques in the context of wind energy prediction.

Work	Approach	Algorithms	Evaluation Function	Performance Gain (%)		
				<i>NMAE</i>	<i>NRMSE</i>	<i>MAPE</i>
Osório et al. [40]	Filter	MI-WT-EPPO-ANFIS	MI (entropy)	83%	80%	Not Available
Jursa [41]	Wrapper	PSO-ANN/NNs	<i>NBIAS</i> * and <i>NRMSE</i>	Not Available	14.5%	Not Available
Jursa and Rohrig [42]	Wrapper	PSO/DE-ANN/NNs	<i>NRMSE</i>	Not Available	10.75%	Not Available

Table 4. Cont.

Work	Approach	Algorithms	Evaluation Function	Performance Gain (%)		
				NMAE	NRMSE	MAPE
Kou et al. [43]	Wrapper	Sequential forward greedy search–OMWGP	MAPE	Not Available	Not Available	3–30% **
This work	Wrapper	BDE–Ensemble of a reduced number of ANNs	WMAE	59% and 60% ***	50% and 51% ***	60% and 58% ***

\* NBIAS: Normalized bias. \*\* Depending on the forecast horizon \*\*\* Computed for 2013 and 2014 data, respectively.

## 6. Conclusions

A feature selection method has been developed to identify the optimal set of weather variables for energy production prediction in wind plants. We have considered the case in which the prediction model is an ensemble of artificial neural networks (ANNs), which provides more satisfactory prediction accuracy than individual ANN models. The proposed feature selection method is based on a wrapper approach that uses a binary differential evolution (BDE) algorithm to search for the optimal feature subset for an ensemble of a smaller number of ANNs than the ensemble model actually used.

The proposed feature selection method has been applied to weather and energy production data collected from a 34 MW wind plant. The weather features are obtained from two weather forecast providers, whose features are different in terms of their timing and feature typology. The results show that the ensemble model developed with the selected features improves the prediction performance of the model currently used by the plant owners while using a smaller number of features than the currently adopted model.

Future work will include the comparison of the proposed method with other state-of-the-art feature selection methods for the same case study. Also, the possibility of using other data-driven techniques as prediction models will be investigated. Specifically, recurrent neural networks, such as echo state networks and long short-term memory networks, will be considered due to their proven effectiveness in dealing with stochastic time-series data. Finally, future work will consider the use of advanced evolutionary algorithms to reduce the computational burden required by fleets of wind plants and the transfer learning of the knowledge gained from the feature selection at one plant to other plants of the fleet.

**Author Contributions:** Conceptualization, S.A.-D., P.B., E.Z. and L.M.; methodology, S.A.-D., P.B., E.Z. and M.F.; software, S.A.-D., P.B. and M.F.; validation, S.A.-D., P.B., M.F., E.Z. and L.M.; formal analysis, S.A.-D., P.B. and M.F.; investigation, S.A.-D., P.B. and M.F.; resources, S.A.-D., P.B., E.Z. and L.M.; data curation, S.A.-D. and M.F.; writing—original draft preparation, S.A.-D., P.B., M.F. and E.Z.; writing—review and editing, S.A.-D., P.B., M.F., E.Z. and L.M.; visualization, S.A.-D., P.B., M.F. and E.Z.; supervision, S.A.-D. and P.B.; project administration, E.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** Restrictions apply to the availability of these data. Data were obtained from Edison Spa and are available from Lorenzo Montelatici with the permission of Edison Spa.

**Conflicts of Interest:** Author Lorenzo Montelatici was employed by the company Edison Spa. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Abbreviations

The following notations and acronyms are used in this manuscript:

AI Artificial Intelligence  
ANNs Artificial Neural Networks

ANFIS	Adaptive Neuro-Fuzzy Inference System
BE	Backward Elimination
BAGGING	Bootstrapping AGGREGATING
BDE	Binary Differential Evolution
CRO	Coral Reef Optimization
CART	Classification And Regression Tree
EPSO	Evolutionary PSO
ELMs	Extreme Learning Machines
EAs	Evolutionary Algorithms
FS	Forward Selection
GAs	Genetic Algorithms
GSO	Gram–Schmidt Orthogonalization
GBM	Gradient Boosting Machine
GPs	Gaussian Processes
LASSO	Least Absolute Shrinkage and Selection Operator
MI	Mutual Information
NWP	Numerical Weather Prediction
NSDBE	Non-Dominated Sorting Binary Differential Evolution
NNs	Nearest Neighbor search
PCA	Principal Component Analysis
PSO	Particle Swarm Optimization
RF	Random Forest
RESs	Renewable Energy Sources
SVR	Support Vector Regression
VMD	Variational Mode Decomposition
WGP	Warped GPs
WT	Wavelet Transform
WMAE	Weighted Mean Absolute Error
MAE	Mean Absolute Error
RH	Relative Humidity
$T$	Temperature
$P$	Pressure
WG	Wind Gust
S	Wind Speed
D	Wind Direction
$x_k^A$	Forecasted weather features provided by provider A, $k = 1, \dots, 24$
$x_k^B$	Forecasted weather features provided by provider B, $k = 1, \dots, 44$
$x_k^{Time}$	Time features related to the periodicity and seasonality of the weather, $k = 1, 2, 3$
$k$	Generic forecasted weather feature
$u$	Wind speed in the direction from west to east
$v$	Wind speed in the direction from north to south
$\sigma$	Bell-shaped function parameter
$N$	Number of ensemble models
$N_{reduced}$	Number of models of the reduced ensemble
$i$	Generic model of the ensemble, $i = 1, \dots, N$
$N_{val}$	Total number of input/output patterns of the validation dataset
$N_{test}$	Total number of input/output patterns of the test dataset
$j$	Generic test pattern, $j = 1, \dots, N_{test}$
$N_{test}^m$	Total number of input/output patterns in the $m$ -th month of the test dataset, $m = 1, \dots, 12$
$m$	Generic month, $m = 1, \dots, 12$
$p_j$	True energy production of the $j$ -th test pattern
$\hat{p}_j$	Predicted energy production of the $j$ -th test pattern

$\hat{P}_i$	Energy production predicted by the $i$ -th ANN model of the ensemble, $i = 1, \dots, N$
$\hat{P}_M$	Energy production predicted by the ensemble as the median of the $N$ individual models
$p$	Number of weather features
$p^*$	Optimal number of weather features
$g$	Generic generation of the BDE search, $g = 1, \dots, G_{max}$
$G_{max}$	Maximum number of generations
$b$	Generic chromosome's bit/gene, $b = 1, \dots, p$
$c$	Generic chromosome, $c = 1, \dots, NP$
$NP$	Number of chromosomes
$z_c^g, \tilde{z}_c^g$	Target $c$ -th chromosome at the $g$ -th generation and its mapped continuous version
$z_{c,b}^g, \tilde{z}_{c,b}^g$	Generic $b$ -th bit/gene of the $c$ -th chromosome at the $g$ -th generation and its mapped continuous version
$rand_{c,b}$	Random number sampled from a uniform distribution in $[0, 1]$
$\tilde{v}_c^g, v_c^g$	Donor or mutant chromosome associated with $\tilde{z}_c^g$ and its binary transform, respectively
$\tilde{v}_{c,b}^g, v_{c,b}^g$	Generic $b$ -th bit/gene of the $c$ -th donor or mutant chromosome at the $g$ -th generation and its binary transform, respectively
$r_1, r_2, r_3$	Three random integers
$OL$	Opposite learning
$x_{p,k}^g$	$OL$ parameter at each $g$ -th generation
$u_c^g$	$c$ -th trial chromosome at the $g$ -th generation
$u_{c,b}^g$	Generic $b$ -th bit/gene of the $c$ -th trial chromosome at the $g$ -th generation
$i_{rand}$	Random integer number
$Cr$	Crossover rate
$SF$	Scale factor $\in [0, 2]$
$fitness$	Fitness function used within the BDE search
$PG_{METRIC}$	Performance gain of a performance metric $METRIC$
$METRIC_{Benchmark}$	Performance metric obtained by the benchmark approach
$METRIC_{Proposed}$	Performance metric obtained by the proposed approach

## Appendix A

The detailed steps of the employed BDE algorithm are hereafter reported for completeness.

The generic  $b$ -th bit (gene),  $z_{c,b}^g$ ,  $b = 1, \dots, p$ , of the  $c$ -th chromosome (also called the target chromosome)  $z_c^g$ ,  $c = 1, \dots, NP$ , at the  $g$ -th generation,  $g = 1, \dots, G_{max}$ , is mapped into a continuous variable,  $\tilde{z}_{c,b}^g \in [0, 1]$  using the mapping operator (Equation (A1)):

$$\tilde{z}_{c,b}^g = \begin{cases} 0.5 * rand_{c,b} & \text{if } z_{c,b}^g = 0 \\ 0.5 + 0.5 * rand_{c,b} & \text{if } z_{c,b}^g = 1 \end{cases} \quad (A1)$$

where  $rand_{c,b}$  is a random number sampled from a uniform distribution in  $[0, 1]$ . Then, the generic  $c$ -th chromosome,  $\tilde{z}_c^g$ ,  $c = 1, \dots, NP$ , where  $NP$  is the number of chromosomes, will undergo the following genetic operations:

1. Mutation. Three chromosomes of the mutant population are selected by sampling three integer indices,  $r_1$ ,  $r_2$ , and  $r_3$ , from a discrete uniform distribution in  $[1, NP]$ . Then, a random vector (called a donor or mutant chromosome),  $\tilde{v}_c^g$ , is generated,  $c \neq \{r_1, r_2, r_3\}$  (Equation (A2)):

$$\tilde{v}_c^g = \tilde{z}_{r_1}^g + SF \left( \tilde{z}_{r_2}^g - \tilde{z}_{r_3}^g \right) \quad (A2)$$

where  $SF$  is a scaling factor that belongs to the interval  $[0, 2]$  [51].

Once the random vector is generated, each  $b$ -th bit/gene,  $b = 1, \dots, p$ , is scaled by applying a sigmoid function to assure that the mutation operator falls in the range of  $[0, 1]$  (Equation (A3)):

$$\tilde{v}_{c,b}^g = \frac{1}{1 + e^{-\tilde{v}_{c,b}^g}} \quad (\text{A3})$$

Finally, the inverse operator is applied to transform the  $\tilde{v}_c^g$  into the binary variables  $v_{c,b}^g$  of the donor (or mutant) chromosome  $v_c^g$  (Equation (A4)):

$$v_{c,b}^g = \begin{cases} 0 & \text{if } \tilde{v}_{c,b}^g \leq 0.5 \\ 1 & \text{if } \tilde{v}_{c,b}^g > 0.5 \end{cases} \quad (\text{A4})$$

2. Crossover (or Recombination). This step entails generating a trial chromosome,  $u_c^g$ , by exchanging the bits/genes between the target and donor chromosomes,  $z_c^g$  and  $v_c^g$ , respectively. This is achieved by resorting to the binomial crossover operator (Equation (A5)):

$$u_{c,b}^g = \begin{cases} v_{c,b}^g & \text{if } rand_{c,b} \leq Cr \text{ or } i = i_{rand} \\ z_{c,b}^g & \text{if } rand_{c,b} > Cr \text{ or } i \neq i_{rand} \end{cases} \quad (\text{A5})$$

where  $i_{rand}$  is a random integer number sampled from a uniform discrete distribution in  $[1, \dots, NP]$ , and  $Cr$  is the crossover rate, i.e., the probability that two binary vectors (i.e., solution candidates) will experience a crossover operation during the evolutionary process.

3. Opposite Learning. To introduce unexplored candidates, a swapping of the genes is sometimes performed (all the 0 bits become 1 and vice versa), depending on the value of the opposite learning (OL) parameter, which is sampled randomly in  $[0, 1]$  from uniform distributions for each chromosome of each  $g$ -th generation (Equation (A6)):

$$x_{p,k}^g = \begin{cases} 1 - x_{p,k}^g & \text{if } rand \leq OL \\ x_{p,k}^g & \text{otherwise} \end{cases} \quad (\text{A6})$$

4. Replacement. Many alternatives can be followed for the creation of the new population. Here, the non-dominated sorting binary differential evolution (NSDBE) strategy is used, as it is able to find more widespread solutions than other methods (e.g., multi-objective tabu search, vector-evaluated genetic algorithm) [63]. At the generic  $g$ -th generation, the population of  $2 * NP$  chromosomes comprising all  $u_p^g$  and  $x_p^g$  candidates is ranked using a fast, non-dominated sorting algorithm that identifies non-dominated solutions, after having evaluated the fitnesses of all the  $2 * NP$  chromosomes. For a single-objective search problem like the one at hand, the selection consists of taking the first  $NP$  chromosomes with higher fitness.

The function used to evaluate the chromosomes (fitness) is the weighted mean absolute error (WMAE), which, in the case of interest for this work, measures the wind production predictions' accuracy relative to the real total monthly production values (Equation (1) in Section 3.2). The set of features,  $p^*$ , with the smallest fitness function constitutes the inputs of the final prediction model.

## References

1. Jain, R.; Mahajan, V. Load Forecasting and Risk Assessment for Energy Market with Renewable Based Distributed Generation. *Renew. Energy Focus* **2022**, *42*, 190–205. [CrossRef]
2. Bakeer, A.; Magdy, G.; Chub, A.; Jurado, F.; Rihan, M. Optimal Ultra-Local Model Control Integrated with Load Frequency Control of Renewable Energy Sources Based Microgrids. *Energies* **2022**, *15*, 9177. [CrossRef]
3. Meenal, R.; Binu, D.; Ramya, K.C.; Michael, P.A.; Vinoth Kumar, K.; Rajasekaran, E.; Sangeetha, B. Weather Forecasting for Renewable Energy System: A Review. *Arch. Comput. Methods Eng.* **2022**, *29*, 2875–2891. [CrossRef]

4. Ponkumar, G.; Jayaprakash, S.; Kanagarathinam, K. Advanced Machine Learning Techniques for Accurate Very-Short-Term Wind Power Forecasting in Wind Energy Systems Using Historical Data Analysis. *Energies* **2023**, *16*, 5459. [[CrossRef](#)]
5. Duan, J.; Wang, P.; Ma, W.; Fang, S.; Hou, Z. A Novel Hybrid Model Based on Nonlinear Weighted Combination for Short-Term Wind Power Forecasting. *Int. J. Electr. Power Energy Syst.* **2022**, *134*, 107452. [[CrossRef](#)]
6. Abisoeye, B.O.; Sun, Y.; Zenghui, W. A Survey of Artificial Intelligence Methods for Renewable Energy Forecasting: Methodologies and Insights. *Renew. Energy Focus.* **2024**, *48*, 100529. [[CrossRef](#)]
7. Alshammari, A. Generation Forecasting Employing Deep Recurrent Neural Network with Metaheuristic Feature Selection Methodology for Renewable Energy Power Plants. *Sustain. Energy Technol. Assess.* **2023**, *55*, 102968. [[CrossRef](#)]
8. Xiao, Y.; Zou, C.; Chi, H.; Fang, R. Boosted GRU Model for Short-Term Forecasting of Wind Power with Feature-Weighted Principal Component Analysis. *Energy* **2023**, *267*, 126503. [[CrossRef](#)]
9. Xie, Y.; Li, C.; Li, M.; Liu, F.; Taukenova, M. An Overview of Deterministic and Probabilistic Forecasting Methods of Wind Energy. *iScience* **2023**, *26*, 105804. [[CrossRef](#)]
10. Senthil Kumar, P. Improved Prediction of Wind Speed Using Machine Learning. *EAI Endorsed Trans. Energy Web* **2019**, *6*, e2. [[CrossRef](#)]
11. Houndekindo, F.; Ouarda, T.B.M.J. Comparative Study of Feature Selection Methods for Wind Speed Estimation at Ungauged Locations. *Energy Convers. Manag.* **2023**, *291*, 117324. [[CrossRef](#)]
12. Salcedo-Sanz, S.; Cornejo-Bueno, L.; Prieto, L.; Paredes, D.; García-Herrera, R. Feature Selection in Machine Learning Prediction Systems for Renewable Energy Applications. *Renew. Sustain. Energy Rev.* **2018**, *90*, 728–741. [[CrossRef](#)]
13. Senthil Kumar, P.; Lopez, D. Feature Selection Used for Wind Speed Forecasting with Data Driven Approaches. *J. Eng. Sci. Technol. Rev.* **2015**, *8*, 124–127. [[CrossRef](#)]
14. Kohavi, R.; John, G.H. Wrappers for Feature Subset Selection. *Artif. Intell.* **1997**, *97*, 273–324. [[CrossRef](#)]
15. Urbanowicz, R.J.; Meeker, M.; La Cava, W.; Olson, R.S.; Moore, J.H. Relief-Based Feature Selection: Introduction and Review. *J. Biomed. Inf.* **2018**, *85*, 189–203. [[CrossRef](#)] [[PubMed](#)]
16. Bolón-Canedo, V.; Sánchez-Marroño, N.; Alonso-Betanzos, A. A Review of Feature Selection Methods on Synthetic Data. *Knowl. Inf. Syst.* **2013**, *34*, 483–519. [[CrossRef](#)]
17. Hastie, T.; Tibshirani, R.; Wainwright, M. *Statistical Learning with Sparsity: The Lasso and Generalizations*; CRC Press Taylor & Francis Group: New York, NY, USA, 2015; ISBN 9781498712170.
18. Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J. *Classification and Regression Trees*, 1st ed.; Taylor & Francis: New York, NY, USA, 1984; Volume 19, ISBN 0412048418.
19. Hapfelmeier, A.; Ulm, K. A New Variable Selection Approach Using Random Forests. *Comput. Stat. Data Anal.* **2013**, *60*, 50–69. [[CrossRef](#)]
20. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; Association for Computing Machinery: New York, NY, USA; pp. 785–794.
21. Bäck, T. *Evolutionary Algorithms in Theory and Practice: Evolution Strategies, Evolutionary Programming, Genetic Algorithms*; Oxford University Press, Inc.: New York, NY, USA, 1996; ISBN 0-19-509971-0.
22. Price, K.; Storn, R.M.; Lampinen, J.A. *Differential Evolution: A Practical Approach to Global Optimization*, 1st ed.; Springer: Berlin/Heidelberg, Germany, 2005; ISBN 3540209506.
23. Al-Dahidi, S.; Baraldi, P.; Zio, E.; Legnani, E. A Dynamic Weighting Ensemble Approach for Wind Energy Production Prediction. In Proceedings of the 2017 2nd International Conference on System Reliability and Safety—ICSRS, Milan, Italy, 20–22 December 2017; IEEE: Piscataway, NJ, USA, 2018.
24. Al-Dahidi, S.; Baraldi, P.; Nigro, E.; Zio, E.; Lorenzo, M. An Ensemble of Echo State Networks for Predicting the Energy Production of Wind Plants. In Proceedings of the 30th European Safety and Reliability Conference and the 15th Probabilistic Safety Assessment and Management Conference, Venice, Italy, 1–5 December 2020; Baraldi, P., Di Maio, F., Zio, E., Eds.; Research Publishing Services: Venice, Italy, 2020; pp. 1–8.
25. Campagner, A.; Ciucci, D.; Cabitza, F. Aggregation Models in Ensemble Learning: A Large-Scale Comparison. *Inf. Fusion.* **2023**, *90*, 241–252. [[CrossRef](#)]
26. Yang, J.; Rahardja, S.; Fränti, P. Mean-Shift Outlier Detection and Filtering. *Pattern Recognit.* **2021**, *115*, 107874. [[CrossRef](#)]
27. Breiman, L. Bagging Predictors. *Mach. Learn.* **1996**, *24*, 123–140. [[CrossRef](#)]
28. Theng, D.; Bhoyar, K.K. Feature Selection Techniques for Machine Learning: A Survey of More than Two Decades of Research. *Knowl. Inf. Syst.* **2024**, *66*, 1575–1637. [[CrossRef](#)]
29. Dhal, P.; Azad, C. A Comprehensive Survey on Feature Selection in the Various Fields of Machine Learning. *Appl. Intell.* **2022**, *52*, 4543–4581. [[CrossRef](#)]
30. Silva, L.; Bispo, B.; Teixeira, J.P. Features Selection Algorithms for Classification of Voice Signals. *Procedia Comput. Sci.* **2021**, *181*, 948–956. [[CrossRef](#)]
31. Siedlecki, W.; Sklansky, J. On Automatic Feature Selection. *Intern. J. Pattern Recognit. Artif. Intell.* **1988**, *2*, 197–220. [[CrossRef](#)]
32. Whittingham, M.J.; Stephens, P.A.; Bradbury, R.B.; Freckleton, R.P. Why Do We Still Use Stepwise Modelling in Ecology and Behaviour? *J. Anim. Ecol.* **2006**, *75*, 1182–1189. [[CrossRef](#)]

33. Slowik, A.; Kwasnicka, H. Evolutionary Algorithms and Their Applications to Engineering Problems. *Neural Comput. Appl.* **2020**, *32*, 12363–12379. [[CrossRef](#)]
34. Khan, P.W.; Byun, Y.C. Genetic Algorithm Based Optimized Feature Engineering and Hybrid Machine Learning for Effective Energy Consumption Prediction. *IEEE Access* **2020**, *8*, 196274–196286. [[CrossRef](#)]
35. El Bourakadi, D.; Yahyaouy, A.; Boumhidi, J. Improved Extreme Learning Machine with AutoEncoder and Particle Swarm Optimization for Short-Term Wind Power Prediction. *Neural Comput. Appl.* **2022**, *34*, 4643–4659. [[CrossRef](#)]
36. Pérez-Aracil, J.; Casillas-Pérez, D.; Jiménez-Fernández, S.; Prieto-Godino, L.; Salcedo-Sanz, S. A Versatile Multi-Method Ensemble for Wind Farm Layout Optimization. *J. Wind. Eng. Ind. Aerodyn.* **2022**, *225*, 104991. [[CrossRef](#)]
37. Qiao, Q.; Yunusa-Kaltungo, A.; Edwards, R.E. Feature Selection Strategy for Machine Learning Methods in Building Energy Consumption Prediction. *Energy Rep.* **2022**, *8*, 13621–13654. [[CrossRef](#)]
38. Jiang, B.; Liu, Y.; Geng, H.; Wang, Y.; Zeng, H.; Ding, J. A Holistic Feature Selection Method for Enhanced Short-Term Load Forecasting of Power System. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 1–11. [[CrossRef](#)]
39. Abdoos, A.A. A New Intelligent Method Based on Combination of VMD and ELM for Short Term Wind Power Forecasting. *Neurocomputing* **2016**, *203*, 111–120. [[CrossRef](#)]
40. Osório, G.J.; Matias, J.C.O.; Catalão, J.P.S. Short-Term Wind Power Forecasting Using Adaptive Neuro-Fuzzy Inference System Combined with Evolutionary Particle Swarm Optimization, Wavelet Transform and Mutual Information. *Renew. Energy* **2015**, *75*, 301–307. [[CrossRef](#)]
41. Jursa, R. Variable Selection for Wind Power Prediction Using Particle Swarm Optimization. In Proceedings of the GECCO 2007: Genetic and Evolutionary Computation Conference, London, UK, 7–11 July 2007.
42. Jursa, R.; Rohrig, K. Short-Term Wind Power Forecasting Using Evolutionary Algorithms for the Automated Specification of Artificial Intelligence Models. *Int. J. Forecast.* **2008**, *24*, 694–709. [[CrossRef](#)]
43. Kou, P.; Liang, D.; Gao, F.; Gao, L. Probabilistic Wind Power Forecasting with Online Model Selection and Warped Gaussian Process. *Energy Convers. Manag.* **2014**, *84*, 649–663. [[CrossRef](#)]
44. Doerr, B.; Zheng, W. Working Principles of Binary Differential Evolution. *Theor. Comput. Sci.* **2020**, *801*, 110–142. [[CrossRef](#)]
45. Gong, T.; Tuson, A.L. Differential Evolution for Binary Encoding. *Adv. Soft Comput.* **2007**, *39*, 251–262. [[CrossRef](#)] [[PubMed](#)]
46. Luesak, P.; Pitakaso, R.; Sethanan, K.; Golinska-Dawson, P.; Srichok, T.; Chokanat, P. Multi-Objective Modified Differential Evolution Methods for the Optimal Parameters of Aluminum Friction Stir Welding Processes of AA6061-T6 and AA5083-H112. *Metals* **2023**, *13*, 252. [[CrossRef](#)]
47. Yuan, S.; Ji, Y.; Chen, Y.; Liu, X.; Zhang, W. An Improved Differential Evolution for Parameter Identification of Photovoltaic Models. *Sustainability* **2023**, *15*, 13916. [[CrossRef](#)]
48. Kar, M.K.; Kumar, S.; Singh, A.K.; Panigrahi, S. Reactive Power Management by Using a Modified Differential Evolution Algorithm. *Optim. Control Appl. Methods* **2023**, *44*, 967–986. [[CrossRef](#)]
49. Al-Dahidi, S.; Ayadi, O.; Alrbai, M.; Adeeb, J. Ensemble Approach of Optimized Artificial Neural Networks for Solar Photovoltaic Power Prediction. *IEEE Access* **2019**, *7*, 81741–81758. [[CrossRef](#)]
50. Al-Dahidi, S.; Di Maio, F.; Baraldi, P.; Zio, E.; Seraoui, R. A Novel Ensemble Clustering for Operational Transients Classification with Application to a Nuclear Power Plant Turbine. *Int. J. Progn. Health Manag.* **2015**, *6*, 1–21. [[CrossRef](#)]
51. Khushaba, R.N.; Al-Ani, A.; Al-Jumaily, A. Feature Subset Selection Using Differential Evolution and a Statistical Repair Mechanism. *Expert. Syst. Appl.* **2011**, *38*, 11515–11526. [[CrossRef](#)]
52. Fogel, D.B. Practical Advantages of Evolutionary Computation. In Proceedings of the SPIE 3165, Applications of Soft Computing, San Diego, CA, USA, 13 October 1997; pp. 14–22.
53. He, X.; Zhang, Q.; Sun, N.; Dong, Y. Feature Selection with Discrete Binary Differential Evolution. In Proceedings of the 2009 International Conference on Artificial Intelligence and Computational Intelligence—AICI, Shanghai, China, 7–8 November 2009; pp. 327–330.
54. Pampara, G.; Engelbrecht, A.P.; Franken, N. Binary Differential Evolution. In Proceedings of the 2006 IEEE International Conference on Evolutionary Computation, Vancouver, BC, Canada, 16–21 July 2006; pp. 1873–1879.
55. Yan, L.; Liu, Y. An Ensemble Prediction Model for Potential Student Recommendation Using Machine Learning. *Symmetry* **2020**, *12*, 728. [[CrossRef](#)]
56. Bonissone, P.P.; Xue, F.; Subbu, R. Fast Meta-Models for Local Fusion of Multiple Predictive Models. *Appl. Soft Comput. J.* **2011**, *11*, 1529–1539. [[CrossRef](#)]
57. Polikar, R. Ensemble Based Systems in Decision Making. *Circuits Syst. Mag. IEEE* **2006**, *6*, 21–45. [[CrossRef](#)]
58. Mohar, B. Some Applications of Laplace Eigenvalues of Graphs. *Graph. Symmetry Algebr. Methods Appl.* **1997**, *497*, 225–275.
59. Angstenberger, L. *Fuzzy Pattern Recognition*; Kluwer Academic Publishers: Alphen am Rhein, The Netherlands, 2001.
60. Zielinski, K.; Weitkemper, P.; Laur, R.; Kammeyer, K.D. Parameter Study for Differential Evolution Using a Power Allocation Problem Including Interference Cancellation. In Proceedings of the 2006 IEEE Congress on Evolutionary Computation—CEC, Vancouver, BC, Canada, 16–21 July 2006.
61. Rönkkönen, J.; Kukkonen, S.; Price, K.V. Real-Parameter Optimization with Differential Evolution. In Proceedings of the 2005 IEEE Congress on Evolutionary Computation, Scotland, UK, 2–5 September 2005; Proceedings. IEEE: Piscataway, NJ, USA, 2005; Volume 1.

62. Chang, T.J.; Wu, Y.T.; Hsu, H.Y.; Liao, C.M.; Chu, C.R. Assessment of Wind Characteristics and Wind Turbine Characteristics in Taiwan. *Renew. Energy* **2003**, *28*, 851–871. [[CrossRef](#)]
63. Deb, K.; Agrawal, S.; Pratap, A.; Meyarivan, T. A Fast Elitist Non-Dominated Sorting Genetic Algorithm for Multi-Objective Optimization: NSGA-II. In Proceedings of the Parallel Problem Solving from Nature PPSN VI: 6th International Conference, Paris, France, 18–20 September 2000; Proceedings 6. Springer: Berlin/Heidelberg, Germany, 2000; Volume 1917. [[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.