Islanding detection in distributed energy resources based on gradient boosting algorithm

Resul Azizi1 | Reza Noroozian2

1 Marmara Research Centre, Energy Institute, Scientific and Technological Research Council of Turkey, Gebze, Kocaeli, Turkey
2 Department of Electrical Engineering, Faculty of Engineering, University of Zanjan, Zanjan, Iran

Abstract
This paper proposes a novel passive-based intelligent method for anti-islanding. Passive methods generally suffer from the improper tuning of threshold values for measured variables and false detection when the active or reactive power mismatch is small. Conversely, intelligence-based methods highly depend on the choice of an appropriate model, the universality of data and selected features. In addition, the risk of overfitting and underfitting for a single model due to an improper selected feature or insufficient data is always possible. This condition makes a single classification model unreliable for practical purposes. The method used in this work is a sequential ensemble of intelligence-based models called gradient boosting. The sequential form of an ensemble is substantially designed to solve the underfitting problem. In addition, by selecting the best combination of features and ensembles, algorithm overfitting can be prevented and the attribute data space can be constructed with maximum discrimination. Furthermore, this condition yields a model with a zero non-detection zone. As a result, a multi-objective optimisation method called the non-dominated sorting genetic algorithm-III is used to tune the model. The reliability and economy of this method for distribution networks with various types of inverter-based and synchronous distributed energy resources are demonstrated.

1 | INTRODUCTION

The increasing penetration and integration of distributed energy resources (DERs) and the complexities in their operating modes cause novel difficulties to emerge in modern power systems. This ever-increasing growth poses problems for the economic, safe and stable use of DERs to supply loads to new types of networks, such as microgrids. Modern DERs can operate with a power system or solely with various control modes. Hence, it is necessary to detect unintentional islanding and switch between operational or control modes on time to maintain the stability of systems. Unintentional islanding is a condition where one or more DERs remain connected and feeding one or more local loads while they are disconnected from the main power system. Some of the many threats and security problems posed by this condition are as follows.

1. Out of phase reconnection of reclosers, which can cause large transient torques in the islanded region to rotational loads, such as pumps and blowers [1].
2. Hazard to personnel and protection issues when DERs cannot provide a sufficient fault current for protection devices [1].
3. Transient over voltages due to rapid loss of load if an adequate ground source is not present in the islanded part [2].
4. Power quality problems and degradation of voltage stability due to a power mismatch or poor voltage and frequency regulation [3].

Islanding detection methods (IDMs) are generally divided into two main groups: local and remote methods. In remote methods, all switches in the local electrical power system (EPS) are monitored by a central decision-making system. Based
on their connection or disconnection, the islanded region is detected, and control signals are sent for post-fault operation of the islanded region. The remote methods have zero non-detection zones (NDZs), high reliability, degradation-free power quality, and applicability to multi-DER systems. However, these methods are complicated. In addition, vast infrastructure changes and high operational costs are indispensable. Examples of remote IDMs are introduced in [1] and [3].

In local IDMs, all measurement, detection, and decision-making are done near the utility. These methods are mainly classified into three groups: active, passive, and hybrid methods. Active methods have a very small NDZ. However, their main drawback is that they complicate and degrade the control system of the DER, leading to a longer detection time and power quality problems due to continuously injecting a perturbation to the grid. In modern power systems, DERs can operate in a post-islanded condition. For post-islanding operation of DERs, it is necessary to change between control modes, and the system is more susceptible to disturbances. Thus, these methods are not a good choice because they lead the islanded part to instability and avoid post-island operation and control of the network. Examples of active methods are introduced in [1] and [4]. On the other hand, passive IDMs are economic, and they have a fast detection speed and no effect on power quality in either grid-connected or islanded operation. However, these methods suffer from a large NDZ and are sensitive to the adjustment of threshold values [1, 3]. To overcome the drawbacks of traditional local methods, hybrid IDMs are suggested. The general procedure for islanding detection with these methods is to implement passive methods unless a suspicious condition is detected and activates active methods [5, 6]. Hybrid methods can mitigate power quality and NDZ problems, but the detection time worsens in some cases and complicates the DERs’ control systems, particularly the inverter-based ones. Nonetheless, the risk of instability for the post-islanding condition remains.

To attain a secure and economic islanding detection, adapt it to contemporary inverter-based DER technologies and avoid the aforementioned deficiencies of the traditional methods, various manipulations are performed in local methods. In [7], parameters are transformed to the space vector domain before setting the threshold values. However, the following method depends on the type of DER and is applicable for synchronous DERs. In addition, it depends on extracting the hidden characteristic of system variables using digital filters. In another study, the set point of the rate of change of frequency (ROCOF) relays is adjusted using alternative methods, such as prior knowledge of the power curve and the study of all power imbalances of DERs [8]. This method also suffers from NDZ to a great extent, and it depends on daily load curves. In [9], a harmonic tuned filter was installed at the DERs to extract a definite harmonic order of impedance, and in [10], the second harmonic voltage of the inverters was extracted for islanding detection. However, these methods depend on setting a predefined threshold for islanding detection and on the network size and topology. Methods such as deviation of the oscillation frequency are suggested in [11]. This paper uses a chaotic theory oscillator called the Helmholtz oscillator, which has great discrimination between chaotic and normal signals. The dynamic behavior of loads and the derivative of their equivalent resistance in a microgrid are used as an IDM in [12]. These methods can detect islanding with higher accuracy, but their main drawbacks are multiple settings of thresholds, which are highly affected by noise and delays in detection time.

Subsequent efforts have focused on intelligent methods and their combination with traditional active, passive, or signal processing methods [13–16]. The outstanding feature of intelligence-based data mining (IBDM) methods is that the thresholds are not set manually. The models learn to adjust themselves according to data. However, these methods intensely depend on the universality of data, measured values (features) and possible training scenarios. IBDM methods are burdened with more costs as the amount of data-gathering equipment increases, and classifiers, such as a decision tree (technically called weak learners), usually have high classification errors in a nonlinear data space. Therefore, signal processing methods are introduced in [17–20] to extract outstanding features of data and minimise the use of measurement devices. Strong classification methods, such as support vector machines (SVMs) [21] or extreme learning machines (ELMs) [22], are used to classify attribute data space more precisely. These solutions increase accuracy substantially, decrease the NDZ dramatically and shorten the detection time. However, their programming and processing are complicated and entail a high risk of overfitting due to false measured data. This situation makes these methods inapplicable for practical purposes.

As a result, this paper introduces an ensemble learning method called gradient boosting to construct a strong model from weak learners, such as a decision tree. This classifier is easy to programme, and it demonstrates the performance of a strong IBDM model, such as SVM or ELM. Moreover, sensitivity to false measurements, noise, the NDZ and overfitting have been progressively minimised using a novel, multi-objective optimisation method called the non-dominated sorting genetic algorithm-III (NSGA-III) [23]. NSGA-III also helps to detect easily and deterministically all possible models that have the following characteristics:

1. An optimal number of the features and a minimum number of measurement devices.
2. An appropriate construction of ensembles for minimizing the processing time and NDZ, and feasibility for programming in a local protection system (minimum complexity).
3. Select the best sliding time window.
4. Detect a minimum number of additional features that can minimise NDZ to a desirable level.

Total ensemble methods are advantageous primarily because they attempt to fit the model to the data step-by-step, and only a few integer parameters must be defined. They are easier to programme and demonstrate the same performance as conventional strong classifiers. The NSGA-III helps to optimally define these parameters so that the total model selects the best
feature space with minimum generalisation error and robustness to overfitting.

The remainder of this article is as follows. Section 2 introduces the topology of gradient boosting and the tuning of NSGA-III for optimal feature selection. Section 3 develops the proposed method for the power system. Section 4 describes the test system. Simulation results are presented in Section 5 and compared with those of previous IBDM methods in Section 6. Finally, in Section 7, the conclusion is drawn.

2 | NSGA-III-BASED TUNING GRADIENT BOOSTING ALGORITHM

The intelligence-based IDMIs provide a fast detection time and high accuracy. These methods are technically classified as weak (basic) learners, such as the decision tree and k-nearest neighbour methods, or as strong learners, such as the SVM and ELM. Weak learners are easy and straightforward to tune and train. Unfortunately, they classify attribute data space linearly and underfit and demonstrate poor performance for non-linearly dispersed data. Intelligence-based islanding detection with weak learners decrease the NDZs compared to traditional methods. However, because of nonlinearity and the complexity of the power system’s feature space, these methods still leave considerable NDZs, particularly in inverter-based DERs [15].

In contrast, strong learners can classify attribute space nonlinearly and more flexibly. High accuracy may be achieved by well-tuning the parameters and coefficients of the model. However, their proper tuning is controversial and not straightforward. In addition, concentrating on adjusting these parameters to increase the test data accuracy increases the risk of overfitting. Therefore, they are impractical for real system implementation.

Altogether, weak and strong classification methods only train a mono-model for classification. If this model is trained improperly because of an erroneous measurement or outlier data, the entire decision is affected adversely and fails. According to these results, intelligence-based models with a mono-classifier are not sufficiently reliable for practical purposes.

2.1 | Gradient boosting

Here, instead of using a mono-classifier, an ensemble of classifiers is implemented. In ensemble learning, instead of using a mono-classifier, a group of models participates in the training and decision-making process. The main goal is to iteratively and, if necessary, by adjusting the training speed, suppress the decision power of the models with a high prediction error and empower the correctly predicting models. In ensemble learning, the decision is taken by a majority vote of the models to minimise the error rate. These methods are generally classified into two categories: bagging and boosting. In bagging, the main goal is avoiding overfitting, and every model is trained separately by randomly selected parts of the training data. Then, the average vote of the data is used for the decision. On the other hand, in boosting methods, the main goal is to construct a strong model from weak learners. The total model helps to solve the underfitting problem of weak learners and is sufficiently strong to classify the data space nonlinearly. Moreover, it is straightforward and less complex for tuning and training. Boosting methods are trained in a sequence, and in every iteration, compensation for the mistakes of the previous models is attempted (Adaboost) or the classification error is slowly (step-by-step) decreased (gradient boosting). Adaboost [24] and gradient boosting [25] are two outstanding ensemble boosting methods. In Adaboost, all data has a weight factor, and in every iteration, the weight factor of wrongly classified data increases, while those of the others decrease. Adaboost cannot demonstrate good performance in a complex data space or in the presence of incorrectly classified data and outliers. In every iteration, the model progressively concentrates on classifying these data. In addition, it is very sensible to the selected features of attribute space, and it overfits in some feature spaces.

In gradient boosting (see Figure 1), instead of emphasizing misclassified data, the classification error for all data is defined as a function to minimise in every iteration. As stated in Table 1, first, a simple prediction model \( F_0(x) \) is defined as the probability of data belonging to a class. The goal is to minimise a differentiable loss function (the mean square error) using an optimisation method called the gradient decent algorithm (a first-order iterative optimisation algorithm for finding the minimum of a differentiable function). In every iteration (round), the distance between the prediction and class of data is considered an error rate of the model. These errors are used to calculate the gradient, which is basically the partial derivatives of the loss function \( (L(y, F) = \frac{1}{2}(y - \hat{F}(x))^2) \). This procedure obtains the residues of the model \( (y - \hat{F}(x) = -\frac{\partial L(\hat{y}, \hat{F}(x))}{\partial \hat{F}(x)}) \). It helps to find the steepness of the error function, which can be used to determine the direction for changing the model parameters to maximally reduce the error in the next training iteration (round). In the next step, a regression weak model (a regression tree) \( h_b(x) \) is fit to the residues to achieve a model with lower error. This procedure continues iteratively until the optimisation process reaches a point with minimum error, and the residues become close to zero (or until an appropriate number of rounds are completed). Therefore, the algorithm reaches the minimum of the loss function, and movement in any direction will not improve the performance.

Notwithstanding all the advantages of gradient boosting over traditional strong classifiers, it is yet a strong model, and it is greedy to learn training data in every iteration. This, in turn, can cause the model to construct complex regression trees with high depth, particularly when the models have many features. Thus, to increase the accuracy and minimise the overfitting risk of gradient boosting, in every iteration, the residue is multiplied by a shrinkage parameter \( (0 < \alpha < 1) \). This parameter causes the model to concentrate to compensate for a portion of the total error in every iteration. Therefore, misclassified data with a large effect on the error function is overwhelmed. However, a very small shrinkage parameter can decrease the convergence speed of the model, so a trade-off is necessary. The pseudocode of gradient boosting is as follows:
2.2 NSGA-III-based optimisation of gradient boosting

Selecting data space is an important part of constructing an intelligence-based model. Data space is a combination of the features (attributes) that help intelligent methods to learn and classify data. A suitable data space is where there is maximum discrimination between data of every class and an optimum number of the features to avoid complexity. In addition, the risk of overfitting is minimised in this space. Using ensemble learning methods, a strong classification model can be constructed only by training a group of basic models. These models are easy and straightforward to train and programme on local measurement devices because they only contain a group of comparison trees (regression trees). Unlike an SVM or ELM, they do not have parameters that can strictly affect the total results through small changes. Moreover, in gradient boosting, the model emphasises minimizing the entire classification error instead of attempting to classify misclassified data. Therefore, the performances of highly overfitted models are suppressed by every learning iteration (because of erroneous measurement and outliers), and their effect on the entire result is minimised.

Gradient boosting can be constructed by adding every weak learner until the error stops decreasing. In addition, in every iteration, weak learners are proposed to minimise the loss function as much as possible without any limitation on increasing the depth of the regression trees. This greedy performance for the gradient boosting can increase the risk of overfitting, specifically in a large and complex data space. By selecting the proper values of the parameters and the optimum combination of features (the best data space), it is possible to remove the overfitting probability and to remove the NDZ in islanding detection altogether. A good constructed gradient boosting with a proper data space can positively affect the performance of the model. Here, a multi-objective optimisation algorithm called the NSGA-III
TABLE 1  Gradient boosting algorithm

| Input: Take training set \((x_i, y_i)\)  |
|-------------------|--------------------------------|
| loss function \(L(y, F)\)   |
| learning rate \(\alpha\) (0<\(\alpha<1\)) |

1: Create an initial simple model to classify data
2: \(F_0(x) = \arg \min_{\gamma} \sum_{i} L(y_i, \gamma)\)
3: Select the round number \(M\)
4: for \(m = 1, \ldots, M\)
5: Calculate pseudo residual \(r_{im} = -\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\)
6: \(F_m(x) = F_{m-1}(x) + \alpha \cdot h_m(x)\) \((N\) is the number of the dataset)
7: Fit a base regression tree \(h_m\) to \(r_{im}\) using \((x_i, r_{im})\)
8: for \(j = 1, \ldots, J_m\), compute the multiplier \(\gamma_{jm} = \arg \min_{\gamma} \sum_{i} L(y, F_m(x_i) - \gamma h_m(x_i))\) \((j\) is the number of splits in the tree)
9: end for
10: end for
11: Update \(F_m(x) = F_{m-1}(x) + \alpha \cdot h_m(x)\)

Output: \(F_M(x) = F_0(x) + \sum_{m=1}^{M} \alpha \cdot \gamma_{jm} h_m(x)\)

[23] (see Figure 2) is used to optimise the performance of the gradient boosting.

Compared to previous versions, the main property of NSGA-III is a series of reference points that help to extract the Pareto front (groups of optimal answers with no superiorities to each other) more homogenously, specifically in data with high dimensions. These points help to improve the diversity and convergence of NSGA-III and boost its performance for more than two objective problems. The algorithm begins with an initial feasible solution (initial population) and a series of widely separated reference points. These points are placed on a normalised hyperplane using the systematic approach of Das and Danning [26]. This method helps to add every reference point equally inclined on an objective axis and has an intercept for every pair of the objective axis. The number of divisions on every axis is selected by the user, and it is important to compromise between the speed and diversity of the problem. In this work, \(p = 4\) divisions are selected. Therefore, for an \(M = 3\) objective problem, the total number of reference points is

\[
\binom{M+p-1}{p} = \binom{3+4-1}{4} = 15.
\]

In the next step for maintaining the diversity of the problem, every member is associated with reference points evenly. Mutation and crossover are applied, and the elite members are selected and sorted for the remainder of the algorithm. Then, using a nondominated sorting method, members with equal feasibility in the solution are sorted as fronts (\(F_1, F_2, \ldots\)). The new normalised populations are associated with reference points by the shortest perpendicular distance between them and the

FIGURE 2  NSGA-III multi-objective optimisation flowchart
reference points. Finally, using a niche preservation strategy, new reference points are added to existing ones and the reference points with zero associated members are omitted.

The new normalised populations are associated with reference points by their shortest perpendicular distance to these points. Finally, using a niche preservation strategy, new reference points are added to existing ones and the reference points with zero associated members are omitted.

In power systems, it is possible to select various types of measurements as features and define an attribute data space. An optimisation-based feature selection helps to find an attribute space (the best combination of features) where there is maximum discrimination between data of every class. In addition, there is a vast amount of selection of the ensembles. It is essential to define the number of learning rounds (iterations) or the total number of base learners. Furthermore, it is necessary to determine the maximum depth of the regression trees and shrinkage parameters. In selecting these conditions, a compromise between the accuracy of the testing and training results and the complexity of the model (the least complex model) is needed. It is obvious that there are multiple optimum answers that have no superiority to each other. In addition, optimizing one condition contradicts optimizing another condition. Here, a multi-objective-based optimisation is suitable for detecting all optimal models. NSGA-III is a metaheuristic, multi-objective optimisation algorithm selected for reaching these goals.

Gradient boosting and NSGA-III are coded in the MATLAB R2018b script environment, and to construct gradient boosting, MATLAB binary regression tree models are used.

### 3 | PROPOSED METHOD

Local IDMs are advantageous mainly because all measuring, processing and detecting are performed near the utility. They do not need infrastructures for transferring all the DERs data to a central processing system. They are fast and economical, and different groups of measurements may be selected to construct the attribute data space. Then, a proper feature selection (optimisation) method detects the minimum number of features in feature space, where there is maximum discrimination between data of each class (the best feature space).

#### 3.1 | Feature selection

In IBDM, many measured and calculated (processed) features may be selected. In the literature, various combinations of these features are selected [9, 13, 15, 16, 21, 27–28]. However, no explanation is provided about the optimum combination of these features and why their selection is suitable, or if there is dependence between these features and the type of DER. In addition, no explanation is provided about the possibility of boosting the performance of the model or decreasing the NDZ by adding one or more features, and what these features are. This paper demonstrates that, if there is a sufficient number of features, NSGA-III can extract the best combination of them (data spaces) and construct a model with zero NDZ that is robust to overfitting. Between the existing features, NSGA-III can determine which ones are the best combination for minimizing the NDZ. NSGA-III demonstrates that various combinations between the features have no superiority to each other. Therefore, based on the economic condition and the expected reliability and accuracy of the model, selecting between the features is possible. Here, to prove the capability of the method in selecting the appropriate combination between large amounts of the attributes, the NSGA-III feature selection process is performed according to the feature sets in Tables 2-4.

However, in practical applications, the feature selection by this method can be performed in a subset of these features according to the measurement devices of the DERs and the calculated features.

In this work, signals with limited time length are utilised around the disturbance. However, in real-world application, the method is utilised for non-stop continuous electrical signals discretised with a specific sampling frequency. Therefore, for

| **TABLE 2** Feature set 1 |  |
|----------------------------|---|
| \( \Delta P \)             | Active power deviation |
| \( \frac{dP}{dt} \)         | Rate of change of active power |
| \( \Delta Q \)             | Reactive power deviation |
| \( \frac{dQ}{dt} \)         | Rate of change of reactive power |
| \( \Delta V \)             | Voltage deviation |
| \( \frac{dV}{dt} \)         | Rate of change of voltage |
| \( \Delta f \)             | Frequency deviation |
| \( \frac{df}{dt} \)         | Rate of change of frequency |
| \( \frac{dP}{dQ} \)         | Frequency to active power deviation |
| \( \frac{dV}{dQ} \)         | Voltage to reactive power change |
| \( \frac{dP}{dV} \)         | Voltage to active power change |
| \( \frac{dQ}{dV} \)         | Voltage to reactive power change |
| \( \frac{dP}{dt} \)         | Voltage to active power deviation |

| **TABLE 3** Feature set 2 |  |
|----------------------------|---|
| \( \Delta \phi \)          | Voltage phase deviation |
| \( \Delta (\phi - \phi_1) \) | Deviation of voltage current angle difference |
| \( \frac{d\phi}{dt} \)      | Rate of change of voltage angle |
| \( \frac{d(\phi - \phi_1)}{dt} \) | Rate of change of voltage and current angle difference |
| \( \Delta I \)             | Current deviation |
| \( \frac{dI}{dt} \)         | Rate of change of current |
| THD(V)                     | Total harmonic distortion of voltage |
| THD(I)                     | Total harmonic distortion of current |
Algorithm implementation

For accurate anti-islanding, three objective functions (OF) are used in NSGA-III to select the best models, which are robust to overfitting and exhibit a maximum discrimination between classes. The first OF is the minimum number of features ($N_{\text{feature}}$) for selecting minimum measurement devices and considering economic issues. The second OF is for minimizing the complexity of models to programme; the maximum number of splits ($N_{\text{split}}$) (the maximum depth of the tree) in regression trees and the total number of base learners ($N_{\text{base learner}}$) (the number of learning rounds) should be minimised. However, minimizing these values should not deteriorate the gradient boost performance. For this, in the third objective function, the training error ($Train_{\text{err}}\%$) and test data error ($Test_{\text{err}}\%$) must be minimised (to compensate for the performance of the first objective function). In addition, to decrease the data sensitivity of the constructing model, the average loss of the five-fold cross validations ($CV_{\text{fold test}}$) should be minimised (to compensate for the performance of the second objective function).

One of the main properties of the overfitted model is its low training error and high testing error as well as its high cross-validation loss. Hence, by minimizing these properties in the third objective function, the selection of models that are not overfitted is guaranteed. In addition, when the cross-validation loss is minimised, the sensibility of the model to new unseen data is minimised. This, in turn, reinforces the zero NDZ capability of the model.

Thus, the NSGA-III objective functions are

$$\text{OF} (1) = \min \sum N_{\text{feature}}$$

### Table 4 Feature set 3

| P       | Instantaneous active power |
|---------|----------------------------|
| Q       | Instantaneous reactive power |
| V       | Instantaneous Voltage      |
| $\varphi_v$ | Voltage phase              |
| $\varphi_v - \varphi_f$ | Voltage and current phase difference |
| f       | Frequency                  |
| I       | Instantaneous Current      |
| pf      | Power factor               |
Enter selected features from feature set to train gradient boosting

NSGA-III Cost Function

1. If M=nRound % M is learning rounds
2. M=nRound
3. end if
4. If J=Split % J is maximum split of trees
5. J=Split
6. end if
7. for learning rate \[\{0.1; 0.25; 0.5; 0.75; 1\}\]
8. execute Algorithms 2 (gradient boosting)
9. execute 5 fold cross validation
10. calculate accuracy, cross validation loss, error rate
11. end for
12. Select model with maximum accuracy,
13. minimum cross validation loss and error rate
14. Return objective function
15. OF = \{OF(1);OF(2);OF(3)\}

% Initialize algorithm
1. Define number of division in reference points; \(n_{\text{Div}}=10\)
2. Define number of objective function; \(n_{\text{Obj}}=3\)
3. Determine Number of reference points; \(n_{\text{Ref}}=\left(\frac{2^{10}-1}{10}\right)\) = 66
4. Define population size; \(n_{\text{Pop}}=100\)
5. And create normalized reference points
6. Define crossover rate; \(p_{\text{Cross}}=0.7\)
7. Define Mutation rate; \(p_{\text{Mut}}=0.4\)
8. Maximum iteration; \(MaxIt=100\)
9. Define maximum splits of decision tree; \(n_{\text{Split}}=31\)
10. Define maximum training round; \(n_{\text{Round}}=255\)

Enter a measurement based on selected features

FIGURE 3 NSGA-III-based optimisation gradient boosting

\[
\text{OF (2)} = \min \sum N_{\text{base learner}} + N_{\text{split}}
\]

\[
\text{OF (3)} = \min \left[\text{Train}_{\text{err}} + \text{Test}_{\text{err}} + CV_{\text{Fold loss}}\right]
\]

Consequently, the NSGA-III algorithm can find all possible solutions, in which optimum attribute space (features), minimum complexity, minimum error and robustness to overfitting are granted.

Moreover, inputs (members of the population) of the NSGA-III are a binary vector of the selected or non-selected features, the maximum depth of the tree and the learning round. Every member is evaluated in the NSGA-III cost function, and in every evaluation, measured features are selected and extracted to execute gradient boosting, cross-validation and other objective functions. The evaluation process of the NSGA-III finds the Pareto front (the best results of the objective function) for further study to select the appropriate model based on the economic and safety conditions of the power system (see Figure 3).

1 TEST SYSTEM DESCRIPTION

The test system [15] in Figure 4 is used to validate the results. To check the independence of the proposed model from the DER type, various types of DERs are modelled in the network. In this system, two wind farms are in the first feeder. The first DER is a doubly fed induction generator (DFIG_wind) with a rated value of 6.6 MVA and 575 V. It is a rotating machine DER that is partially connected to the network by an inverter through the rotor and directly connected to the power system by a stator. The second DER is a synchronous generator (SG_wind) type 4 full-scale converter with a rated value of 6.6 MVA and 575 V. It is a rotating machine DER with full connection to the power system through an inverter. These models are equipped with a battery energy storage system (BESS) in the DC terminal of the inverters [29]. The feeder is energised with a 12.5-MW rated load and 25-kV line voltage, and the frequency is 60 Hz. Both wind farms are simulated in the voltage and power control mode scenarios, while the BESS is active or inactive. Both turbines are tested in the full loading, half loading and min loading conditions. The second feeder is equipped with a 500-kW, 260-V photovoltaic (PV) farm (an inverter-based nonrotating DER) operating in the power control mode. Finally, a 3.125-MVA, 2.4-kV synchronous diesel generator (Diesel) is modelled in the second feeder. This feeder is energised with a 3.5-MW load and a 25-kV, 60-Hz line voltage. The substation transformer is a 120/25-kV model. All transformer rated capacities and connections are depicted. The test system is simulated in MATLAB R2018b software, and the features are measured by simulation.

4.1 Simulation scenarios

Anti-islanding involves using control systems or relays to prevent the continued occurrence of unintentional islanding. Based
on IEEE Standard 1547-2018 [30], the DER should detect and cease to energise the islanded area within 2 s. In addition, according to IEEE Standard 929-2000 [31], in PV systems, if a mismatch between generation and consumption is more than 50% or the islanding power factor is less than 0.95, the DER should cease islanded EPS in less than 10 cycles. In the first generation of the DERs, when a fault occurs near these DERs, energisation of these DERs must cease regardless of whether they are islanded. This is because the control systems cannot maintain the DERs stability after the clearance fault. There are no control systems for a low-voltage ride through (LVRT) protection embedded in these DERs. However, in the next generation of networks, ever-increasing penetration of the DERs in distribution networks is the motivating idea behind bidirectional power transfer between the main power network and local networks. Disconnecting DERs after any disturbance can lead to instability problems in power systems. Therefore, the LVRT control systems, such as crowbar protection for intense and symmetrical faults and dual vector protection for unsymmetrical protections, are used in the DERs [32]. These novel control systems definitely affect the transient behaviour of the measured signals that are used as features. Furthermore, in contemporary networks, the DERs are operated as parts of the microgrids. Therefore, not only grid connected parts but also islanded parts must continue operating and supplying the loads. In addition, at least one of the DERs must switch to voltage control mode or retain its voltage control mode to have a stable islanded area. Furthermore, every control mode has its own transient behaviour in the post-islanding condition. Thus, modelling the DERs with these details is attempted to have a realistic operation for their simulation.

In the contemporary variable-speed wind turbines that are used in this work, the power is fully or partially injected to the network using a back-to-back converter. Because of the indeterministic characteristic of the wind speed and the necessity of continuously supplying the loads, adding a BESS to the DC bus of the back-to-back converter is suggested. Here, to study these BESS effects in the measured transient signals as features, some of the features are extracted when the battery is active. Both wind turbines are equipped with the BESS in the DC terminal, and the simulations are performed in the active or inactive modes of both batteries. In Figures 5 and 6, two instances of these conditions and their active power and reactive power mismatch are demonstrated. It is seen that when the battery is active and islanding occurs, the rate of change in parameters is smoother and more stable. Consequently, islanding detection using traditional passive methods is deteriorated, and the model stays in the NDZ for a longer time.

According to the scenarios in Table 5, 857 events are simulated for every time window. In most of the islanded scenarios, first, the NDZ of the DERs is extracted [16]. Then, most of the islanded scenarios are simulated inside this NDZ and a few
of them outside the NDZ to check the methods performance outside the NDZ.

In every group of the events, 70% (572 events) of the data is used for training and 30% (285 events) is selected for testing. In addition, to detect the optimum time window (where the minimum number of features with maximum accuracy is preferred), the test system is simulated until 15 cycles after the disturbances. The features in every time window from the first two cycles, first three cycles etc., are extracted for the first 15 cycles. In fact, 14 (number of time window) $\times$ 4 (number of DER) = 56 different feature selection algorithms (NSGA-III) are executed.

5 | SIMULATION RESULTS AND STUDY OF THE NDZ

NSGA-III-based gradient boosting optimisation is tested for the DERs and Pareto fronts (best results of objective function) for all extracted time windows. An example of these Pareto fronts for the DFIG\_wind in a 13-cycle time window is presented as a parallel plot in Figure 7. The results for all time windows are gathered, and only models with 100% accuracy are selected for further processing.

In the present simulation, and specifically in power system operations, the priorities for selecting the best model are as follows. After selecting models with 100% training precision in the time windows, the next priority is minimizing the number of measurement devices. Thus, a model with a minimum use of measurement devices (minimum cost) and maximum number of processed features is more suitable. For example, if possible, it is preferable to select the first- and second-order derivatives of the instantaneous active power instead of the instantaneous active power and voltage.

The next step is the NDZ of the models. Based on the study of the mismatch between generation and consumption and maintaining network safety, a model covering all power mismatches and having zero NDZs is superior. Therefore, to generalise the model to the dataset and reduce the model...
dependency on network topology, an algorithm with minimum cross-validation loss is suitable. To test the NDZ of the proposed algorithm with different feature sets, first, the NDZs of the DERs are detected using the method in [16]. Then, various simulation scenarios, in addition to the testing and training data, are generated inside the NDZ with various active and reactive power mismatches and load quality factors. The trained models with 100% accuracy in the training and testing data are challenged by these data. Zero NDZ is observed in all of these models. Therefore, if NSGA-III models with high accuracy are selected, the system has zero NDZ, but more measurement devices are necessary to extract these features. A compromise between accuracy and economic factors must be made.

None of the selecting models deviate from the time limit of the standards. However, models with a lower response time are preferable. In Table 6, the optimum islanding detection model of the DERs in all time windows and the selected time windows of the DERs are highlighted. Moreover, in Table 7, the selected features of the best models obtained using NSGA-III are presented.

The results show that the fully inverter-based connected DERs need more complicated models to achieve desirable results. In addition, the response time of the models is very short and reliable. Multi-objective-based feature selection provides equally perfect results. Moreover, the backup models can be selected with a minimum difference in features along with the main model. On the other hand, a backup model can be selected with the same features as those of the main islanding detection model if NSGA-III provides a desirable one, even with lower training accuracy.
### TABLE 5  Isolated and non-isolated phenomena of the case study

| Events | Description | Parameters | Number of events |
|--------|-------------|------------|------------------|
| a) Islanded scenarios and the number of events | | | |
| Faults and islanding | One-phase to ground, two-phase to ground, three-phase to ground (fault duration is seven cycles for all scenarios.) | Fault resistance $R_f = 5, 10, 25, 150$ | 300 events |
| During islanding, the loads are mainly passive and have a quality factor $Q_f = 1 - 1.5 - 2 - 2.5$ | Motor loads nominal power 200 HP, 500 HP, 2250 HP | | |
| 10% of scenarios are simulated with motor loads | | | |
| Circuit breaker opens and islanding | One or multiple DERs are islanded | | 72 events |
| During islanding, the loads are mainly passive and have the quality factor $Q_f = 1 - 1.5 - 2 - 2.5$ | | | |
| b) Non-islanded scenarios and the number of events | | | |
| Induction motor start-up | For passive methods, the start-up transient signals are usually detected as islanded scenarios using conventional passive islanding detection methods. | Motor loads nominal power 200 HP, 500 HP, 2250 HP | 36 events |
| Load shedding or circuit breaker opens (non-islanding) | Shedding one or multiple loads | | 120 events |
| Faults and load shedding | One-phase to ground, two-phase to ground, three-phase to ground (fault duration is seven cycles for all scenarios.) | Fault resistance $R_f = 5, 10, 25, 150$ | 260 events |
| Topology change | Circuit breakers 2 to 9 or 18 to 21 opens because of different events and circuit breaker 19 closes at the same time. | | 33 events |
| Capacitor bank switching | For passive methods, capacitor bank switching transient signals are usually detected as islanded scenarios using conventional passive islanding detection methods. These banks are mainly switched near the DERs to check their effect on the signal measured near the DERs as a feature. | Motor loads nominal power 1200 kVar, 600 kVar, 1200 kVar, 600 kVar, 36 events |

#### FIGURE 7  Pareto front of DFIG_wind for time window = 13 cycles

### 6  COMPARING THE RESULTS

In this section, our results are compared with those of traditional passive methods and intelligence-based methods. The main goal of this paper is to use methods that have a simple feature extraction procedure and avoid complex signal processing methods. In total, the compared intelligence-based methods are divided in three groups:

1. Weak classifier: Decision tree.
2. Strong classifier: SVM, ELM, Naïve Bayes and artificial neural networks.
3. Ensemble classifier: Adaboost and gradient boosting.

The results are presented in Table 8. Furthermore, the feature selection process of this paper and the test network are tested and compared with the decision tree, Adaboost and ELM. The
| Time window | Number of measurement devices | Number of features | Cross-validation accuracy | Total number of base learners (total number of splits) | Learning Rate | Response time | Number of measurement devices | Number of features | Cross-validation accuracy | Total number of base learners (total number of splits) | Learning Rate | Response time |
|-------------|-------------------------------|-------------------|---------------------------|------------------------------------------------------|--------------|--------------|-------------------------------|-------------------|---------------------------|------------------------------------------------------|--------------|--------------|
| 1           | 4                            | 10                | 96.32%                    | 21(462)                                              | 0.25         | 0.0795       | 8                            | 15                | 98.36%                    | 77(399)                                              | 0.25         | 0.1345       |
| 2           | 9                            | 15                | 96.33%                    | 249(1494)                                            | 0.25         | 0.1994       | 7                            | 14                | 99.23%                    | 169(845)                                             | 0.25         | 0.1528       |
| 3           | 10                           | 25                | 98%                       | 21(126)                                              | 0.1          | 0.0793       | 6                            | 14                | 98.91%                    | 66(462)                                              | 0.75         | 0.1129       |
| 4           | 9                            | 24                | 96.49%                    | 48(192)                                              | 0.25         | 0.1025       | 8                            | 16                | 99.67%                    | 139(695)                                             | 0.25         | 0.1528       |
| 5           | 9                            | 27                | 99.17%                    | 144(288)                                             | 0.75         | 0.1288       | 6                            | 9                 | 99.45%                    | 105(520)                                             | 0.25         | 0.105        |
| 6           | 10                           | 25                | 97.66%                    | 178(712)                                             | 0.25         | 0.1879       | 7                            | 10                | 98.58%                    | 33(138)                                              | 0.1          | 0.1365       |
| 7           | 10                           | 25                | 97.83%                    | 89(365)                                              | 0.1          | 0.1695       | 7                            | 15                | 100%                      | 105(315)                                             | 0.1          | 0.1645       |
| 8           | 9                            | 31                | 98.66%                    | 11(55)                                               | 0.25         | 0.1555       | 6                            | 9                 | 99.45%                    | 10(30)                                               | 0.25         | 0.155        |
| 9           | 9                            | 26                | 98.5%                     | 133(798)                                             | 0.75         | 0.2465       | 4                            | 8                 | 98.58%                    | 8(32)                                                | 0.25         | 0.1699       |
| 10          | 10                           | 27                | 97.83%                    | 26(130)                                              | 0.75         | 0.1963       | 6                            | 14                | 99.65%                    | 44(176)                                              | 0.1          | 0.209        |
| 11          | 10                           | 27                | 98.66%                    | 142(284)                                             | 1            | 0.2284       | 7                            | 13                | 99.56%                    | 73(292)                                              | 1            | 0.2292       |
| 12          | 10                           | 26                | 98.66%                    | 31(186)                                              | 0.25         | 0.2353       | 4                            | 7                 | 99.12%                    | 80(400)                                              | 1            | 0.2567       |
| 13          | 9                             | 31                | 98.66%                    | 71(355)                                              | 0.5          | 0.2688       | 6                            | 13                | 99.45%                    | 73(318)                                              | 0.5          | 0.2771       |
| 14          | 10                           | 27                | 98.5%                     | 10(60)                                               | 0.5          | 0.2568       | 8                            | 12                | 99.78%                    | 36(180)                                              | 0.1          | 0.268        |
| 15          | 9                             | 21                | 97.3%                     | 17(85)                                               | 0.25         | 0.0962       | 7                            | 13                | 97.82%                    | 38(182)                                              | 0.1          | 0.2561       |

**Synchronous generator wind farm**

**Doubly fed induction generator wind farm**

**Photovoltaic**

**Diesel generator**
TABLE 7  Extracted features of selected time windows for islanding detection

| DER's name | Selected features |
|------------|-------------------|
| Wind farm synchronous generator | mean($P$) energy($\frac{dP}{dt}$) std($P$) $\Delta Q$ $\frac{dQ}{dt}$ mean($Q$) |
| | mean($\frac{d^2Q}{dt^2}$) Max($Q$) - Min($Q$) ($\varphi_v$ - $\varphi_i$) $\Delta f$ $\frac{df}{dt}$ mean($\frac{d}{dt}(\varphi_v - \varphi_i)$) |
| | mean($\frac{d}{dt}(\frac{V}{Q})$) energy($\frac{d^2V}{dt^2}$) std($V$) $\Delta f$ $\frac{df}{dt}$ mean($\frac{d}{dt}(\varphi_v - \varphi_i)$) |
| | $\frac{dV}{\Delta Q}$ Max($f$) - Min($f$) |
| Wind farm DFIG | energy($P$) Std $\left(\frac{dP}{dt}\right)$ $\Delta V$ mean $\left(\frac{d}{dt}(\frac{V}{P})\right)$ energy($\frac{d^2V}{dt^2}$) $\Delta I$ THD$\_f$ |
| Photovoltaic | mean($\frac{d^2P}{dt^2}$) $\frac{dP}{dt}$ energy($\frac{dP}{dt}$) std($\frac{d}{dt}(\varphi_v - \varphi_i)$) $\Delta Q$ mean($\frac{d}{dt}(\varphi_v - \varphi_i)$) $\frac{dQ}{dt}$ |
| | $\frac{d^2V}{dt^2}$ energy($\frac{d}{dt}P$) $\frac{d}{dt}$ energy($\frac{d}{dt}(\varphi_v - \varphi_i)$) $\frac{d}{dt}Q$ |
| Diesel generator | $\frac{dQ}{dt}$ std($Q$) mean($\frac{d}{dt}(\varphi_v - \varphi_i)$) |

TABLE 8  Comparison of different islanding detection methods

| Method | Number of selected features | Accuracy | Extra information |
|--------|-----------------------------|----------|------------------|
| ROCOF [14] | 1 | 72.13–78.65% | The accuracy depends on the setting of threshold values |
| ROCOV [14] | 1 | 68.08% | – |
| Decision Tree [15] | 12 | 98.86% | There is a critical amount of the NDZ, but it is easy to programme and implement. |
| Naïve Bayes [34] | 18 | Synchronous DER = 100% Inverter-based DER = 56% | – |
| Artificial neural network [16] | 2 | Synchronous DER = 100% Inverter-based DER = 97% | Training the algorithm and extracting the optimal parameters of hidden nodes are labourious. |
| Support vector machine [28,34] | 2–18 | Synchronous DER = 99.46% Inverter-based DER = 94% | There are no explicit methods for tuning the coefficients, and there is high risk of overfitting |
| Extreme learning machine [33] | 9–21 | Synchronous DER = 99.7% Inverter-based DER = 98.4% | Selected features depend on the type of DER, and there is a high risk of overfitting |
| Adaboost [27] | 5–35 | Synchronous DER = 98.89% Inverter-based DER = 97.5% | For some data spaces, the model overfits |
| Gradient boosting | 7–31 | 100% for both synchronous and inverter-based DER | – |

decision tree is one of the straightforward machine learning methods, and Adaboost is another sequential ensemble of weak learner methods that enables us to classify data space nonlinearly. In addition, the ELM is a strong classifier with tunable coefficients for nonlinear classification. The Fitensemble and Fitctree functions of MATLAB software are used to test the Adaboost and decision tree algorithms. The ELM is programmed according to the algorithm in [33]. It is seen that, in the decision tree-based classification, there is always a high NDZ that can be fully removed. In addition, the cross-validation loss, even for the best feature space, is below 94%. On the other hand, for some feature sets, the Adaboost algorithm overfits, and its accuracy on test data is very low (below 40%). This phenomenon generally occurs in datasets with many features and few measurement devices. Moreover, in the gradient boosting algorithm, for the same feature set, it is always
This paper has introduced a novel ensemble of intelligent methods for passive islanding detection called gradient boosting. Intelligence-based passive methods are preferable because they decrease the detection time, avoid the manual setting of thresholds and are independent of the DER type. The main goal of ensemble learning methods is to construct a strong model from an appropriate combination of weak models. These methods are advantageous compared to a mono strong model because they have easy to adjust parameters, programme in local processing systems, avoid underfitting, reduce sensitivity to data and enhance robustness to misclassified data. A multi-objective optimisation (the NSGA-III) is used to optimise the model parameters. It is demonstrated that models with zero NDZs and a minimum risk of overfitting can be constructed if there is sufficient data. Moreover, the constructed models have an optimum combination of the feature space and complexity of base learners. NSGA-III determines the most appropriate parameters based on the network topology and type of DER. It is demonstrated that more than one acceptable ensemble topology is in the attribute data space. This, in turn, aids in the selection of a model compatible with the available equipment and economic considerations. Eventually, these methods can provide a reliable, on-time islanding detection with a high accuracy and compliance with the standards.

REFERENCES

1. Manikonda, S.K.G., Gaonkar, D.N.: Comprehensive review of IDMs in DG systems. IET Smart Grid 2(1), 11–24 (2019)
2. Kroposki, B.: Prevention of unintentional islands in power systems with distributed resources. National Renewable Energy Laboratory, Golden, CO (Aug. 2016)
3. Kim, M.S., et al.: Comprehensive review of islanding detection methods for distributed generation systems. Energies 12(5), 837 (2019)
4. Bej, T.: Accurate active islanding detection method for grid-tied inverters in distributed generation. IET Renew. Power Gener. 11(3), 1633–1639, (2017)
5. Babkhi-Jafarabadi, R., et al.: Maximum power point tracking injection method for islanding detection of grid-connected photovoltaic systems in microgrid. IEEE Trans. Power Delivery; (2020)
6. Chen, X., et al.: A novel hybrid islanding detection method for grid-connected microgrids with multiple inverter-based distributed generators based on adaptive reactive power disturbance and passive criteria. IEEE Trans. Power Electron. 34(9), 9342–9356 (2019)
7. Alam, M.R., et al.: Islanding detection of distributed generation using electrical variables in space vector domain. IEEE Trans. Power Delivery 35(2), 861–870 (2020)
8. Alam, M.R., et al.: Assessing the performance of ROCOF relay for anti-islanding protection of distributed generation under subcritical region of power imbalance. IEEE Trans. Ind. Appl. 55(5), 5395–5405 (2019)
9. Dhua, R., et al.: Harmonic filter-based improved islanding detection technique for microgrid. IET Renew. Power Gener. 13(13), 2443–2450 (2019)
10. Malakondiaiah, M., et al.: Second harmonic impendence drift-based islanding detection method. IET Gener. Transm. Distrib. 13(23), 5313–5324 (2019)
11. Bakshi, M., et al.: Novel islanding detection method for multiple DGs based on forced Helmoltz oscillator. IEEE Trans. Smart Grid 9(6), 6448–6460 (2018)
12. Xie, X., et al.: A new passive islanding detection approach considering the dynamic behavior of load in microgrid. Int. J. Electr. Power Energy Syst. 117, 105619 (2020)
13. Hartmann, N.B., et al.: Hardware implementation and real-time evaluation of an ANN-based algorithm for anti-islanding protection of distributed generators. IEEE Trans. Ind. Electron. 65(6), 5051–5059 (2018)
14. Azim, R., et al.: An islanding detection methodology combining decision trees and Sandia frequency shift for inverter-based distributed generations. IET Gener. Transm. Distrib. 11(16), 4104–4113 (2017)
15. Cai, Q., et al.: Islanding detection of hybrid distributed generation under reduced non-detection zone. IEEE Trans. Smart Grid 9(5), 5027–5037 (2018)
16. Sarnet, H., et al.: Minimum non detection zone for islanding detection using an optimal Artificial Neural Network algorithm based on PSO. Renew. Sustainable Energy Rev. 52, 1–18, (2015)
17. Do, H.T., et al.: Passive-islanding detection method using the wavelet packet transform in grid-connected photovoltaic systems. IEEE Trans. Power Electron. 31(10), 6955–6967 (2016)
18. Alsharief, S., et al.: A new approach based on wavelet design and machine learning for islanding detection of distributed generation. IEEE Trans. Smart Grid 5(4), 1575–1583 (2014)
19. Nale, R., et al.: Islanding detection in distributed generation system using intrinsic time decomposition. IET Gener. Transm. Distrib. 13(5), 626–633 (2019)
20. Ahmadianpour, M., et al.: A novel islanding detection technique using modified Sartlet transform in multi-distributed generation. Int. J. Electr. Power Energy Syst. 112, 460–475 (2019)
21. Baghvaei, H.R., et al.: Anti-islanding protection of pv-based microgrids consisting of PHEVs using SVMs. IEEE Trans. Smart Grid 11(1), 483–500 (2020)
22. Khamis, A., et al.: Faster detection of microgrid islanding events using an adaptive ensemble classifier. IEEE Trans. Smart Grid 9(3), 1889–1899 (2018)
23. Deb, K., Jain, H., An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part I: Solving problems with box constraints. IEEE Trans. Evol. Comput. 18(4), 577–601 (2014)
24. Schapire, R.E., Singer, Y.: Improved boosting algorithms using confidence-rated predictions. Mach. Learn. 37(3), 297–326 (1999)
25. Friedman, J.H., Greedy function approximation: A gradient boosting machine. Ann. Statist. 29(5), 1189–1232 (2001)
26. Das, I., Dennis, J.E.: Normal-boundary intersection: A new method for generating the Pareto surface in nonlinear multiobjective optimization problems. SIAM J. Optim. 8(3), 631–657 (1998)
27. Chavoshi, S.A., et al.: Islanding detection of synchronous distributed generation using Adaboost algorithm. Int. Trans. Electr. Energy Syst. 26(8), 1604–1624 (2016)
28. Matić-Cuka, B., Kezunovic, M.: Islanding detection for inverter-based distributed generation using support vector machine method. IEEE Trans. Smart Grid 5(6), 2676–2686 (2014)
29. Qu, L., Qiao, W.: Constant power control of DFIG wind turbines with supercapacitor energy storage. IEEE Trans. Ind. Appl. 47(1), 359–367 (2011)
30. IEEE Standard for Interconnection and Interoperability of Distributed Energy Resources with Associated Electric Power Systems Interfaces. IEEE Standard 1547-2018 (Revision of IEEE Std 1547-2003), 1–138, 2018
31. *IEEE Recommended Practice for Utility Interface of Photovoltaic (PV) Systems*. IEEE Standard 929-2000, i, (2000).

32. Gonzalo, A., et al: Hardware solutions for LVRT. In Doubly Fed Induction Machine: Modeling and Control for Wind Energy Generation Applications. IEEE Press, Piscataway, NJ, (2011).

33. Huang, G.B., et al: Optimization method based extreme learning machine for classification. Neurocomputing 74(1), 155–163 (2010).

34. Faqhruldin, O.N., et al: A universal islanding detection technique for distributed generation using pattern recognition. IEEE Trans. Smart Grid 5(4), 1985–1992 (2014).

How to cite this article: Azizi R, Noroozian R. Islanding detection in distributed energy resources based on gradient boosting algorithm. *IET Renew Power Gen.*, 2021;15:436–451. https://doi.org/10.1049/rpg2.12040