A Machine Learning based Optimized Energy Dispatching Scheme for Restoring a Hybrid Microgrid

Miftah Al Karim*, Jonathan Currie, Tek-Tjing Lie
Dept of Electrical and Electronic Engineering
Auckland University of Technology

September 16, 2017

Abstract

A microgrid operated in stand alone mode is highly vulnerable to instability when the integration of intermittent energy sources are considered. If a short circuit fault occurs in a microgrid while operating at its design limit, often cost effective system recovery becomes a challenging task. Under such contingencies predictive analysis can be used to strengthen the system restoration schemes. In this study, a system based on machine learning algorithm is implemented to forecast the security of a standalone microgrid and based on the forecasting, schedule multiple backup diesel generators under the contingency of loss of a major generating unit. The underlying objective is to maintain the voltage stability with an optimized economic dispatch scheme, right after clearing a critical three phase short circuit fault. Finally, a promising set of outcomes are observed and discussed.

*miftah.alkarim@aut.ac.nz

keywords: Monte Carlo simulation, distributed generation, hybrid microgrid, genetic algorithm, machine learning.

1 Introduction

Maintaining bidirectional power flow often introduces security issues in the modern microgrids [1]. The advent of large scale solar power generation in residential areas as well as integration of large scale wind energy sources contribute even more to this security. A three phase fault in such a stochastic scenario can easily cause protection system failure leading towards cascading outages. Thus a renewable powered microgrid demands novel ways of energy dispatching methods [2]. Traditionally, under these conditions a central station addresses the tertiary regulation that typically has an interval of 24 hours. However, in a modern microgrid integration of distributed generators (DG) is compelling to adopt a more local, decentralized approach [3]. It is because long term planning often is prone to large scale errors leading towards service interruptions. The alternate approach thus, would be to build short term forecasting systems of the non dispatchable energy sources [4]. One of the most sought out methods in this field is the application of machine learning algorithms [5–7]. However, most study limits itself in analyzing the accuracy of the algorithm by comparing actual and forecast data. Thus very few study has been conducted to forecast power system security in order to take service restorative (SR) measures [8]. On the other hand considerable efforts have already been made to improve SR plans by implementing multi agent based systems (MAS), advanced metering infrastructure (AMI), knowledge based systems, linear programming, progressive hedging (PH) etc [4, 9–11]. These methods are computationally complex and often are not suitable for uncertain fault durations, involving integer decision variables such as a binary indication of the presence of rotor angle instability. This issue can be resolved by taking into account two types of uncertainties, i.e., a critical fault during a vulnerable and non-vulnerable periods. These periods then can be further decomposed into multiple scenario based stochastic programs [8]. However, such method would demand accurate forecasting of the system security and integration of that forecasting with the available data in an online basis. The previous studies often ignored this integration of post fault security assessment with the demand and generation data obtained from Distributed Resources (DR). Such integration of data is carried out after the stability is achieved which means this relationship is ignored during the period while addressing a critical fault [12]. This objective is attained mostly by curtailing loads. To minimize the service interruptions understanding the security of the system is a necessity. If the generation data is updated online, system security can be measured dynamically [13, 14].

This study intends to bridge this gap between security forecasting and service restoration plans by implementing a novel architecture. The key argument of this study is to take advantage of heuristic search over the logic reasoning or empirical judgement [15]. The study implements a predictive analytical method that forecasts the vulnerability of a system if a critical three phase short circuit fault occurs at that instance. The prediction is made online, based on the available wind power, solar power and the non critical (controllable) loads. Machine learning driven optimization is used to implement an autonomous restoration scheme after a critical three phase fault followed by loss of a generation. The proposed method takes into account the distributed generations and demands in order to predict the system security. The security assessment is carried out within short intervals considering the possibilities of a major three phase fault takes place in the near future [16, 17].

This analysis explores the idea that under different energy demand and distributed energy generation, the impact of a three phase fault can either be critical or non-critical [16]. The proposed method figures out that criticality and prepares a set of optimized contingent scenarios for restoring the system. The algorithm is based on machine learning that manipulates an optimization platform in order to achieve lowest possible operating cost after ensuring voltage quality throughout the network in a post fault contingency [17]. The goal is achieved by implementing an ensemble of bagged decision tree based system to do the forecasting followed by a genetic algorithm (GA) for the service restoration. Many previous studies have successfully implemented security and reliability indices for system analysis [18]. This study takes a similar approach to measure system security by introducing a binary security index called Probability of Stability (POS). The index considers several scenarios of short circuit faults resulting in isolating a generator bus in the affected area [17]. The database of POS is prepared by a Monte Carlo simulation method. The stability analysis carried out in this study, has a hierarchical structure with a primary goal of restoration and a secondary goal of economic dispatch. The optimality is discussed in terms of stabilizing a system with lowest possible operational cost.

2 The Micro-grid Model

Two different models have been used in this study. For building the method ~~by~~ a small scale system is used. ~~And in~~ order to understand the impact of the proposed method in larger networks an IEEE-39 bus 10 machine system is used. Both the ~~network is~~ shown in Figure-1.

The smaller network has one hydro turbine based synchronous generator, two backup diesel generators (synchronous) of **G1=4MW** and **G2=3MW**, one asynchronous generator representing the wind farm **3MW** and a voltage source converter based solar power plant **5MW**. The loads are lumped on a common transmission grid. This distribution approach is inspired by the microgrids used in [9, 19]. However, this model differs in dividing the loads into two parts; critical and non critical loads [20]. The critical loads are comparable to the base load of a system that has to be met. On the other hand non critical loads are often considered to have flexible levels in any demand response program, specially in an islanded mode [21, 22]. Such flexibility allows load-shedding or curtailment at users discomfort. For simplicity it is assumed that no curtailment cost has to be paid by the service provider. The total demand in this microgrid is higher than the total capacity of the synchronous generator models used as power plants. It signifies that the power quality and stability of the system time to time depends on the wind power plant and the solar power plant. The wind power plant is modelled as an induction generator-based variable speed wind turbine. The solar power plant is modelled as a current source and placed closer to the residential load. The three phase model has diodes, internal resistance and leakage current followed by a voltage source converter (VSC) as presented in [23]. The VSC based solar plant is only implemented as an intermittent energy source. These two distributed and intermittent sources do not have any type of power system stabilizers installed in them thus maintaining stability in the system is carried out through the synchronous generator based models. To capture the full dynamics of bidirectional energy flow both the solar and wind turbine plants are designed not to have any energy storage device.

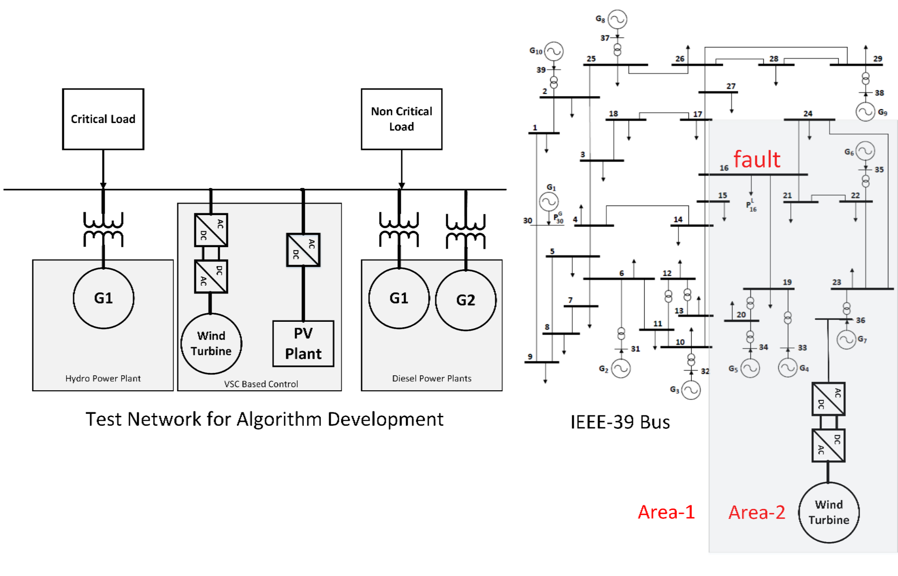

Figure 1: The Micro-grid Model

The generator buses are represented using the typical second order swing-equation;

$$M_i\ddot{\delta}_i + D_i\dot{\delta}_i = P_{mi} - P_{gi}; i \in generator_{1:3} \tag{1}$$

Here, δ_i is the generator rotor angle, P_{mi} is the mechanical power input, P_{gi} is the electrical power output, M_i generator's inertia coefficient and D_i is the generator's damping coefficient. The overall operation is subject to [17];

$$P_{gi} - P_{li} - \sum_{j=1}^{3} U_iU_j(G_{ij}cos\delta_{ij} + B_{ij}sin\delta_{ij}) = 0, \tag{2}$$

$$Q_{gi} - Q_{li} - \sum_{j=1}^{3} U_iU_j(G_{ij}cos\delta_{ij} - B_{ij}sin\delta_{ij}) = 0, \tag{3}$$

In order to test the security of the system and develop the probability of stability (POS) database, a critical three phase fault is placed near the hydro station.

3 Data Model

To build up the forecasting system three data models for wind power, solar power and non critical load have been prepared. **N** number of random data points are generated using those models. Then a Monte Carlo based simulation is used with the data points for developing the Probability of Stability (POS) table. Each scenario has a three phase fault in it.

3.1 Wind Energy Model

For simplicity the wind power generator only considers wind speed as a variable.

$$P_W = \frac{1}{2}\rho A v^3 C_{total} \tag{4}$$

Where, P_W is active power output, C_{total} is overall efficiency of the wind turbine, ρ is air density, *A* is swept area and *v* is the wind velocity. The wind velocity at any certain height is calculated by $v_h = v_r(\frac{h}{h_r})^{\alpha}$ and α is the power law exponent. Where, v_h is the speed at hub height, v_r is the speed at reference height.

3.2 Solar Energy Model (Photovoltaic)

To simulate the characteristics of solar power generation the following irradiance based photovoltaic cell model of a solar cell is used;

$$I = I_{ph} - I_s[e^{\frac{V_{OC}+IR_s}{NV_t}} - 1] - I_{s2}[e^{\frac{V_{OC}+IR_s}{N_2V_t}} - 1] - \frac{V_{OC} + IR_s}{R_p} \tag{5}$$

$$V_{OC}(t, \beta) = V_{OC-STC} - K_V T_C(t) \tag{6}$$

Where V_OC is the open circuit voltage of the PV-module, I_{ph} is the solar-induced current that can be further explained by $I_{ph} = I_{ph0}\frac{Ir}{Ir0}$ I_r is the irradiance in W/m^2, I_{ph0} is the solar current obtained for irradiance I_{r0}; I_s and I_{s2} are the saturation currents of the Diode-1 and Diode-2 inside; N_1 and N_2 are the quality factors diodes; $V_t = \frac{kT}{q}$ is the thermal voltage, (k is Boltzmann constant, T_C is device temperature in Kelvin) and K_V is the open circuit voltage temperature coefficient; $T_C = T_A + (NOCT - 20\deg)\frac{I_r(t,\beta)}{800}$. R_s and R_p are the series and parallel resistances [24]. β is the tilt angle and T_A is the ambient temperature. The overall output power from the plant is given as; $P_array(t, \beta) = \eta_{PV} N_S N_P P_{PV}(t, \beta)$. Here, N_S and N_P are the total number of modules connected in series and parallel, η is the conversion efficiency. $P_{PV} = V_{OC}I$; is the instantaneous power output from each PV-module. The solar power plant is designed to have a maximum of 5 MW power capacity.

3.3 Electrical Load Model

The electrical load model used in this study is a function of the base load as well as the ambient temperature. The ambient temperature influences the residential or controllable loads while it is considered to have no impact on the critical loads. It is observed that the impact of temperature on the residential loads is more prominent [25]. It is a motivation behind considering the controllable load as a stochastic variable. The impact of temperature on the residential load is modelled as a 3^{rd} order polynomial system.

The controllable load model is assumed in MW as;

$$P_{Load} = f(T) = B_{CL} + a_0 + a_1 T + a_2 T^2 + a_3 T^3 \tag{7}$$

Where B_{CL} is the base controllable load that is not influenced by temperature, a = multiplying constants, T = temperature in degree Celsius. The coefficients are $a_0, a_1, a_2, a_3 = -0.5632, -1.185e^{-1}, 1.09e^{-2}, -1.8221e^{-4}$. The overall controllable residential load model is shown in Figure-2.

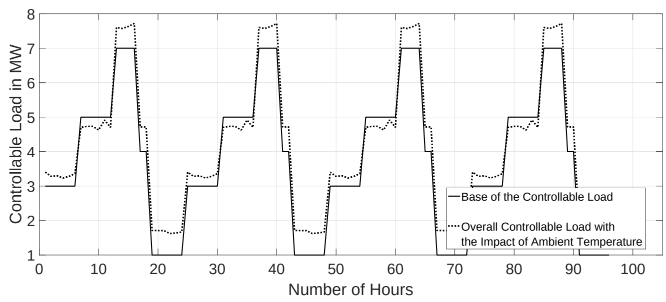

Figure 2: Controllable Load Model

The models are then used to generate random data points for a Monte Carlo simulation.

3.4 Security Index Probability of Stability (POS)

Due to the possibilities of having numerous post disturbance initial conditions, generalizing *service restoration* plans is quite a challenge [15]. This study thus, assumes that the post fault initial conditions only differ from the prefault operational conditions only in the topological layer as a bus isolation at the affected segment. For further simplification, the start up time of the backup diesel generators are considered negligible. A Monte Carlo simulation based approach is taken to prepare the POS database [26] as shown in Table-1.

Wind, solar and consumption data has been randomly selected within a predefined range and fed to the simulation model to create a number of events. Overall **10000** simulation is carried out to develop the POS for the **100** number of clusters [17]. The process of selecting the number of clusters is based on the classification error observed from the classification algorithm. Figure-4(a) shows the process. The POS database indicates which cluster has the probability of being unstable after a critical fault. The first two columns represent generation from wind and solar plants and the third column represents non critical loads right before the three phase fault takes place. Column fourth is prepared using the unsupervised K-means clustering algorithm based on the first three data columns. A total of **100** clusters have been selected for this analysis. The aforementioned **10000** simulations and their associated clusters are also used in training the machine learning platform.

$$Lowest : (\sum_{i=1}^{3}\{Centroid_{ki} - DataColumn_i\}) \tag{8}$$

Here, $Centroid_{ki}$ represents $i - th \mid i \in \{1, 2, 3\}$ centroid for the $k - th$ cluster.

The stochastic data of wind, solar and load can produce infinite combinations, therefore, the unsupervised K-means cluster also serves the purpose of generating finite data segments. The POS is calculated as; $POS_k = \frac{\sum_i^{N_k} S_{ik}}{N_k}$. Here, *k* is the current data cluster, N_k is the total number of simulations in each cluster. S_{ik} is **1** if the system becomes unstable and **0** if the system remains stable.

3.5 Data for Fault Analysis

The proposed microgrid considers four instances of the line voltage; pre-fault line voltage V_{pre}, during fault line voltage V_{flt}, line voltage during the post fault restoration V_{rst}, line voltage after the post fault restoration V_{post}. The aim of this study is to accurately predict system operation strategies that will lead towards achieving a stable and optimal post fault

Table 1: POS Data Table With Clusters

Wind Power MW	Solar Power MW	Controllable Load MW	Cluster	POS_k
1.73	3.47	8.31	1	16.00%
2.77	5.55	9.7	2	64%
...	...	...	...	...
...	...	...	...	...
2.17	4.3	9	45	80%
0.6	1.2	6.8	81	0.0%
2.93	5.87	9.91	96	48%

line voltage. The transition period from the fault state towards the restoration state has been considered similar for all the operational strategies. Incidents such as starting up of a backup generator is considered under the conditions when energy balance cannot be made.

$\sum_{i=1}^{n_g} pg_i + \sum_{i=1}^{n_{ren}} pren_i < P_{dpfcl} + P_{dpfncl} + P_{Tloss}$. Where, P_{dpfcl} is post fault critical load, P_{dpfncl} is the post fault non critical load, P_{Tloss} is the transmission line loss, $pren_i$ power generated from the renewable energy sources, n_g is number of hydro generators, n_{ren} is the number of renewable energy sources.

4 Proposed Algorithm

4.1 Data Preparation for Classification

In the development phase of this study sixteen probable classes have been selected for the algorithm. Each class represents a post fault decision indicating the energy dispatch scheme. These classes become relevant only if a three phase critical fault occurs. If such a fault occurs the system invokes the POS database shown in Table-1. POS database gives an indication of instability based on the measured wind power, solar power and non critical loads at that instance. These three instance data is compared to the cluster centroids for labelling. If the probability of instability is higher than 5% for the identified cluster, it has been considered as a potential candidate for the proposed optimized restoration method. For other events the method is ignored. The sixteen classes or decisions are defined in Table-2. The actions are represented using a set of binary numbers, where for the backup generators **1** means *start* and **0** means *do not start*. ~~And~~ for the non critical loads **1** means *shedding loads* and **0** means *not shedding loads*. The diesel generators are presented as to the swing equation model where the capacities are subject to; $P_{gimin} \leq P_{gi} \leq P_{gimax}$, $Q_{gimin} \leq Q_{gi} \leq Q_{gimax}$, $0 \leq P_{li} \leq P_{di}$, $0 \leq Q_{li} \leq Q_{di}$, $V_{timin} \leq V_{timax}$. Here, P_{li} and Q_{li} are the load after system restoration and P_{di} and Q_{di} are the actual active and reactive demand, V_t is the generator terminal voltage. So the overall load shedding under any of the sixteen actions is $P_{di} - P_{li}$.

Table 2: Sixteen Probable Restoration Schedules

Decision	Start Diesel-1	Start Diesel-2	Shed 1.5 MW	Shed 2.5 MW
1	0	0	0	0
2	0	0	0	1
3	0	0	1	0
..	..	..	..	..
..	..	..	..	..
15	1	1	1	0
16	1	1	1	1

The post fault restoration scheme executes one of these sixteen actions. The primary objective of taking any decision is to maintain as small distance as possible between the pre and post fault voltage states. The objective is achieved by implementing an exhaustive search based optimization technique. This exhaustive search is carried out using the **3482** cases from the **10000** data points where, the system becomes unstable. Which means energy flow from the renewable energy sources alone is not adequate. For each of these cases the optimization routine finds the best decision. The fitness function chosen, is minimizing the sum of squared differences between pre and post fault transmission line voltages.

$$Objective : min(\epsilon_V = \sum_{i=1}^{N}[V_{prefault} - V_{postfault}(i)]^2) \tag{9}$$

~~Here~~ *i* is the data point in consideration. *N* is the total number of data points considered once the fault is cleared. $V_{prefault}$ is the stable line voltage before the fault and $V_{postfault}$ is the post fault stable line voltage. Any contingent scenario represented by one row in Table-1 could have only one decision that yields the lowest voltage deviation. That decision or action is considered to be the optimized target decision for that particular row which belongs to a particular cluster. This cluster based decision mapping is then used to train the classification algorithm where the decisions are set as the target class.

4.2 Preparing the Classification Tree

Once the optimized restoration actions are mapped against the data clusters an ensemble of bagged decision trees has been trained. The tree is trained with four data column as attributes and one as target. The attribute columns are wind power, solar power, demand and the data cluster and the target column is the decision classes. The above **3482** data points including the POS clusters and the target decisions mapped against those clusters have been used to train the tree.

In Figure-3 one limited instance of the ensemble tree is shown;

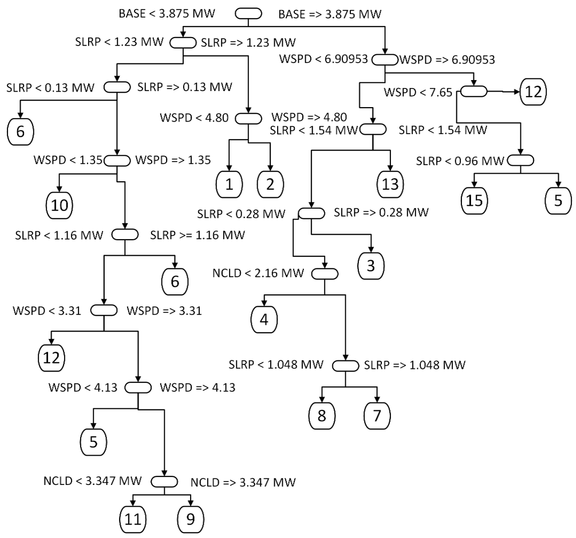

Figure 3: The Decision Tree Classifier (Limited Scenarios)

However, the classification algorithm is not immune to classification errors [27]. A microgrid operated at its design limit is highly vulnerable. Thus any misclassification may lead towards further instability. Therefore, decision ranking system is used for validating the predictions made by the proposed algorithm. The ranking system is a cluster of decision hierarchy exclusively prepared based on the order of optimality. Which means the sixteen decisions have been ranked in an ascending order on a basis of their increasing magnitude of voltage deviation. To prepare this rank, Table-1, Table-2 and the voltage deviations ϵ_V are used. The ranking table only considers the top five decisions yielding the lowest voltage deviation under any scenario. In the previously mentioned **3482** cases a total of thirteen different combinations have been identified. Based on these thirteen observed combinations, an unsupervised hierarchical cluster is prepared. The purpose of having a hierarchical cluster is to validate the predicted decisions made by the ensemble tree. Table-3 shows the ranking using the unsupervised hierarchical cluster. The dendrogram prepared from the hierarchical cluster is shown in Figure-4(b). The hierarchical cluster is also used to find out a decision boundary. It gives an idea which of the additional clusters residing in proximity of the target cluster can also be considered for a particular scenario in order to minimize the cost. For example in Figure-4(b) if cluster-3 fails to restore the system with a lower cost the proposed algorithm will move to cluster-4.

Table 3: The Thirteen Combinations based Hierarchical Cluster

1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}	HCluster
5	4	9	7	11	1
9	7	11	6	10	2
..	..	..	..	..	..
..	..	..	..	..	..
9	5	2	1	6	11
7	3	13	2	15	12
13	15	14	12	16	13

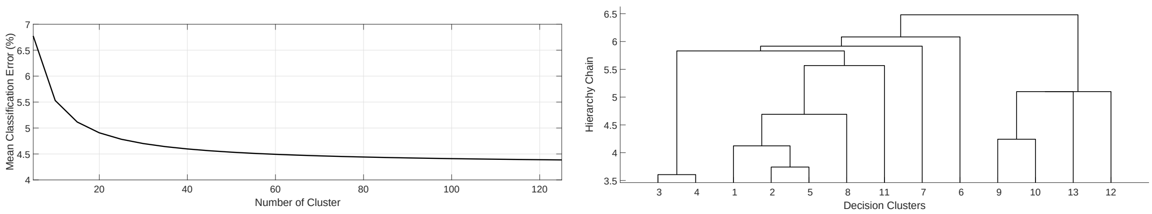

Figure 4: (a) Selection of Number of K-means Clusters (b) Hierarchical Cluster

Furthermore, Two additional ensemble of bagged decision trees have been trained using the identical data set with two different target decisions. These targets are the 2^{nd} and the 3^{rd} ranked decisions. Afterwards a temporary decision set is prepared using the decisions $D1^{st}, D2^{nd}, D3^{rd}$. The decision set is then compared with the hierarchical ranking Table (Table-3).

$$D_{SETi} = \{1^{st}_P \cap 1^{st}_{Ri}, 2^{nd}_P \cap 2^{nd}_{Ri}, 3^{rd}_P \cap 3^{rd}_{Ri}\}; i \in HC_{1:13} \tag{10}$$

Here N_P^{th} stands for the probable predicted decision, N_{Ri}^{th} is the combination of decisions observed in the ranking table from the historical data and HC stands for the hierarchical clusters. D_{SETi} is then used for validation. If validation is failed for any test incident, the training process is repeated.

4.3 Process for Economic Dispatch

The final step of the proposed method is to cater the secondary objective of this study, which is the post contingent economic dispatch (ED). The objective function considers that the aforementioned classification and validation is accurate and therefore the predictions can be used to manipulate the constraint boundaries. The cost function for ED problem is to minimize the total operational cost of the microgrid after clearing the fault and stabilizing the system. The objective is achieved by optimizing the generation from the diesel generators and/or optimizing shedding of non-critical loads. The costs are calculated by using different quadratic functions for the generators.

$$min : \sum_{t} \sum_{i \in G} C_i(pg_{i,t}) + \sum_{t} \sum_{k \in D} C_i(pd_{i,t}); \forall t \in T \tag{11}$$

Where, $C_i(pg_{i,t})$ is the cost function of the *i-th* generator at *t-th* half hour. C_i is formulated as a quadratic function $C_i(pg_Pi, t) = a_{0,i} + a_{1,ipg_{i,t}} + a^2_{2,ipg_{i,t}}$, where, a_0, a_1 and a_2 are the coefficients. A similar approach has also be taken for the cost of shedding load $C_i(pd_Pi, t)$. The ED is subject to an inequality constraint. The inequality constraint considers that the total demand and line losses have to be either equal or less than total generation from the renewable sources $pren_i$ and the diesel generators pg_i after shedding loads if required. The load shedding is represented as p_{Li}

$$P_{dpfcl} + P_{dpfncl} + P_{Tloss} \leq \sum_{i=1}^{n_g} pg_i + \sum_{i=1}^{n_{ren}} pren_i + \sum_{i=1}^{n_{ren}} p_{Li} \tag{12}$$

Where P_{dpfcl} and P_{dpfncl} are the critical and non-critical loads. The post fault demands are kept constant and any large deviation is neglected. Thus $pg_{i,t} - pg_{i,t0}$ is considered zero. Therefore, the upper ramp rate UR_i and lower ramp rate LR_i have also been neglected. The spinning reserve of the generators are considered sufficient.

$$\sum_{i \in G} SR_{i,t} \geq SSR_t; \forall t \in T \tag{13}$$

Where, $SR_{i,t}$ is the available spinning reserve of individual diesel generator, and SSR_t is the system wide required spinning reserve. Each generators also follows the generator output constraints, which means the generation does not exceed its upper limit $pg_{i,t} \leq PG_{i,max}; \forall t \in T$ and also for the renewable energy generators $pgren_{i,t} \leq PGren_{i,max}; \forall t \in T$ [28, 29].

The ED is then solved using genetic algorithm (GA). Here the GA applies **five** standard steps; population initialization, evaluation, selection, crossover and mutation. The GA used maximum **51** generations to reach convergence. The GA is used on the dataset preprocessed by the ensemble of bagged decision trees mentioned earlier. This is the key proposal of this algorithm that shows while optimizing the dispatch in a post contingent scenario, supervised or preprocessed variable vectors yield better outcomes in finding a global minima. The preprocessing is done based the classification process mentioned in the earlier sections. The classifier introduces an upper boundary vector for the design variables namely, the dispatch from two diesel generators and shedding two loads. The upper boundary is prepared based on the decision table (Table-2). For example if the classifier takes *Decision-13 (1100)* then the two diesel generators will be switched on under the condition of $0 \leq pg_{i,t} \leq PG_{i,max}; \forall t \in T$ and no load shedding will take place thus, $pd_{i,t} = 0; \forall t \in T$ and $i \in \{1, 2\}$ [30].

The overall restoration procedure is triggered once a major fault is identified and the POS indicates high probability of instability [12]. The end to end architecture of the proposed algorithm is shown in Figure-5.

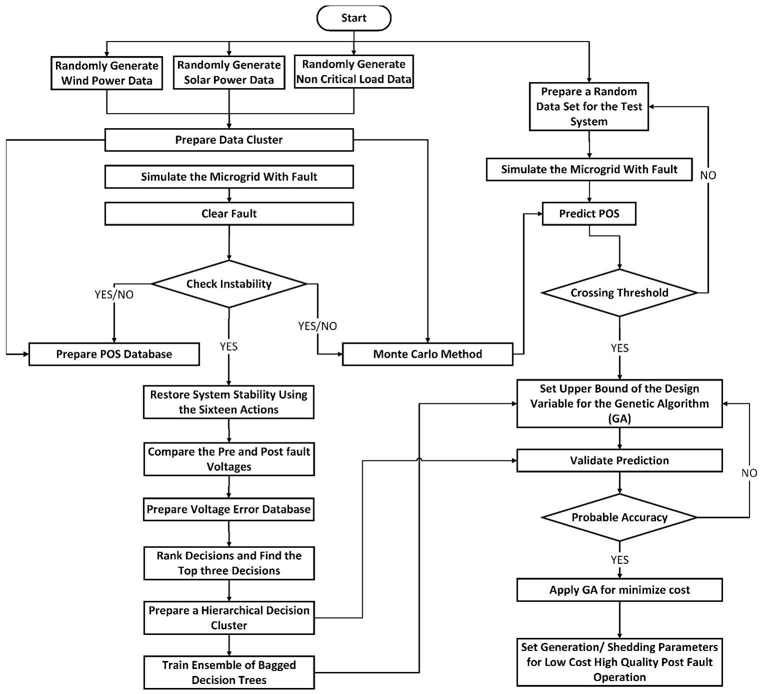

Figure 5: Work-Flow of the Proposed Algorithm

4.4 Exploring Dimensionality

The performance of the proposed algorithm heavily depends on the accuracy of classification. On the other hand the classification relies on the hierarchy of decisions. When the number of decisions and combinations in their hierarchy increases, the risk of classification errors and complexity increase. It is therefore, imperative to understand the role of the proposed method in a larger network. Figure-1 shows an IEEE-39 bus **10**-machine test system, which is used to understand the ramifications of dimensionality. In order to maintain coherency the model is modified by introducing a wind power plant in bus-**36** at the proximity of Generator-**7** [31, 32]. The system like the previous one also considers two types of loads; critical invariant and non-critical variable load. The fault location is represented as *fault* in the figure. To clear the critical fault the transmission line where the fault has occurred is disconnected. When the fault is cleared a rotor angle instability is observed in the network. To eliminate the instability the grid is sectionalized into two areas. The two areas are stabilized and then restored. Area two has a wind power plant of $1000MW$ and area one has two non critical or controllable loads of $300MW$ each. Depending on randomly chosen wind power and non critical load the restoration schemes in different areas of the network vary. Generator-9 in the area-1 and generator-4, generator-5 in area-2 have been considered as the backup diesel generators and also area-1 has the two non critical loads to be shed. For testing the performance of the proposed method *Nine* limited scenarios of a random combinations of wind power and demand have been chosen. The combinations are prepared by dividing wind power and the controllable demands into three segments *High, Medium and Low*. Based on these stochastic scenarios target classes have been prepared for the classification algorithm. All the generators and loads are considered to have a quadratic cost function similar to the functions mentioned in the previous sections. The Table-4 shows some of the distributed control incidents for the Generator-4, 5, 9 and load-1, 2.

Table 4: Stochastic Data Preparation Stage (Selected Scenarios)

Wind Power	Demand	G4	G5	G9	L1	L2
Low	Low	0	0	0	0	0
Low	Low	0	0	0	0	1
Low	Low	1	1	1	1	1
...	...	.	.	.	.	.

5 Results

5.1 Estimation of Instability

The Monte Carlo simulation is carried out with **10000** data points. Out of the **100** data clusters, **6** randomly selected clusters have been chosen to analyze the overall outcome of this study.

Figure-6 shows the POS of all the proposed clusters as well as how simulated data is converging towards obtaining the POS data.

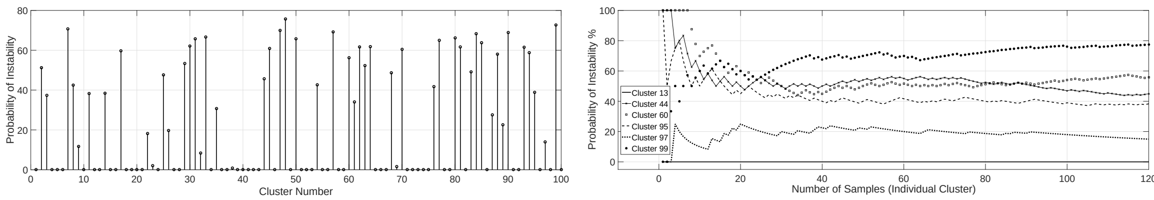

Figure 6: POS of the **100** Clusters and the Monte Carlo method of convergence

Any data cluster having a low POS is ignored in the process of training the ensemble of bagged decision tree. For example *POS cluster-13* has a zero probability of being unstable. Therefore, no restorative measured has been taken for all the instances in *POS cluster-13*.

5.2 Selection of Restorative Action

The proposed ensemble of bagged decision tree method is invoked if the threshold probability of 5% is crossed. Figure-7 demonstrates the result of the classification algorithms in finding out the the 1^{st}, the 2^{nd} and the 3^{rd} top decisions with the lowest possible errors. The top decision database is used in order to control the upper bound of the GA.

The GA is applied to observe the lowest possible operational cost. The results are explained with the previously mentioned **six** POS clusters. The key observation is made in this study is that economic dispatch under duress is not always producing the best possible voltage

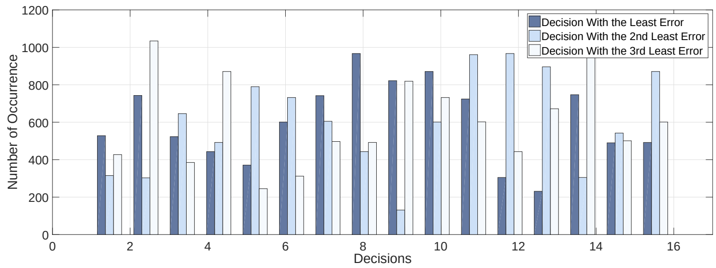

Figure 7: Top Three Decisions after 10k simulations

profile. It means, decisions taken based on the lowest operational cost right after clearing a critical fault may not maintain voltage quality of the system. Thus, taking restorative actions a supervised optimization technique based on previously obtained stability data produces better outcome.

Table-5 shows the three different possibilities while making a decision and Table-6 shows the comparisons between the operational costs. The comparison is made with actions only considering economic dispatch, actions considering the voltage quality and the proposed method.

Table 5: Comparison Between Three Different Decisions

POS Cluster	13	44	60	95	97	99
Decision With the Lowest Cost but poor voltage pofile	NA	11	1	8	10	10
Decision With the Best Voltage Profile	NA	5	3	12	5	7
Acceptable Voltage Profile at Lower Cost	NA	9	3	9	11	9

Depending on the available wind, solar power and controllable loads the above mentioned actions can either be different or can be overlapping. Figure-8 elaborately shows all three decisions and their impacts on the **SIX** clusters. In *cluster-13* the probability of system instability is **zero**. Thus the post fault restorative actions are not invoked. For the other clusters like *cluster-44* the algorithm is applied. In all the cases the decisions with the lowest

Table 6: Normalized Operational Costs

POS Cluster	13	44	60	95	97	99
Decision With the Lowest Cost	NA	0	0	0	0	0
Decision With the Best Voltage Profile	NA	1	1	1	1	1
Acceptable Voltage Profile at Lower Cost	NA	0.96	1	0.62	0.89	0.13

cost has the highest voltage deviation. It is a clear indication that GA for cost optimization alone is not sufficient in order to maintain voltage quality. On the other hand the exhaustive optimization method for maintaining post fault voltage quality based on the decision hierarchy recommends decisions those have the highest operational costs. The proposed algorithm is mitigating these two extrema by taking appropriate actions that produce stable output at a lower cost.

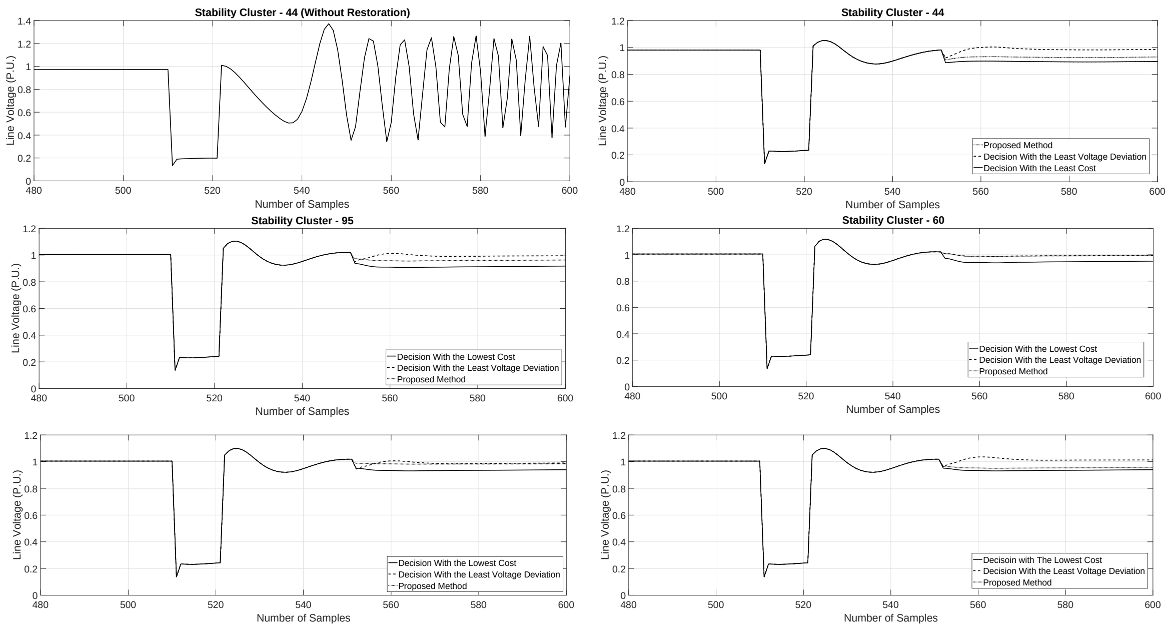

Figure 8: Different Decision Making Processes and Impacts

5.3 Understanding the Combinatorial Explosion

It is quite necessary to understand the ramification of implementing the proposed method in a larger network. For analysing it, one incident with the IEEE-39 bus system has been observed in Figure-9(a). In this event both the wind power and demand is low. The Figure-9(a) shows the impact of dimensionality on the accuracy of the proposed method. In a two area system at first the sectionalized grid has to be stabilized and then restored. In the field of machine learning if dimension of the attributes increases the associated data also increases manifold [27]. Therefore, compare to the smaller network the accuracy of the classification is less in a larger network. Because here the algorithm has to exhibit accurate prediction of two events while in the experimental model shown in the previous sections the prediction has to be for one event. Such expansion leads towards a complex training and testing procedure. This can eventually cause significant decrease in accuracy.

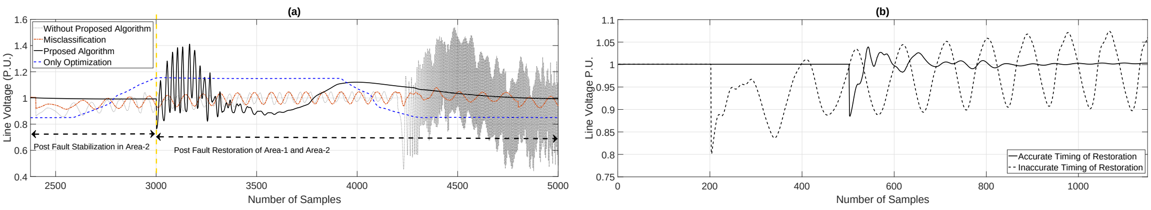

Figure 9: Application of the proposed algorithm on a larger network

The second phenomenon observed is the timing of decision making. In Figure-9(b) two scenarios are shown where the timing of decision making plays two different roles. Due to the increased number of areas the timing of restoring the grid becomes crucial. The restoration should be carried out only if all the areas in the segmented grid are stable. Depending on different stochastic scenarios the periods of achieving stability becomes different.

Thus, the curse of dimensionality would compel to train the machine learning platform with exponentially increased data set in order to reach equal level of accuracy in a larger network [33].

6 Conclusion

This study demonstrates a probabilistic optimization model for restoring a power system after a major three phase fault. The underlying objective of managing resource under a contingent scenario with a machine learning based system is analyzed and presented. The standard GA based optimization system is manipulated using the machine learning based classifier and comprehensive results are achieved.

The control strategies adopted in this study is grid specific. Therefore, data preprocessing plays a vital role in this study. The analyses advocates for considering different approaches for different systems. However, the Monte Carlo simulation based approach is implemented in this study in order to understand the feasibility of developing a data driven generic restoration strategy. The challenge is to bridge between a finite numerical simulation and infinite post fault scenarios.

The proposed classifier also considers classification errors and a simplified rectification technique. The rectification process is an important factor while considering a machine learning algorithm for maintaining a system-wide stable operation under all the stochastic scenarios. The method for rectification implemented in this study was comprehensive. However, such a simplified technique whether is adequate for a larger context or not, need to be evaluated. Also further investigations need to be carried out in order to understand the full ramifications of misclassifying data.

Curse of dimensionality plays a vital role while applying this algorithm. It is observed that for a larger network with multiple areas the training and testing steps increases exponentially. Thus it would be recommended to implement the algorithm in a distributed platform for intra area grid restoration rather than inter area grid restoration.

The performance of the GA is quite satisfactory for the proposed micro grid and the contingencies observed. More scenarios would help comparing and understanding the performance of the GA optimization method not considered in this study. Specially a ~~m~~onte ~~c~~arlo simulation with a larger network would be interesting for future studies. Overall it can be stated that, the proposed Monte Carlo simulation and machine learning algorithm driven optimization model for restoring a stand alone micro grid after a major short circuit fault, is demonstrating a very promising outcome.

References

[1] X. Zhan, T. Xiang, H. Chen, B. Zhou, and Z. Yang, "Vulnerability assessment and reconfiguration of microgrid through search vector artificial physics optimization algorithm," *International Journal of Electrical Power & Energy Systems*, vol. 62, pp. 679–688, 2014.

[2] Q. Ye, T. Ma, Y. Gu, TaoWang, D. Wang, and Y. Bai, "Research on dispatch scheduling model of micro-grid with distributed energy," in *Electricity Distribution (CICED), 2012 China International Conference on*, pp. 1–5, Sept 2012.

[3] Z. Wang and J. Wang, "Self-healing resilient distribution systems based on sectionalization into microgrids," *IEEE Transactions on Power Systems*, vol. 30, pp. 3139–3149, Nov 2015.

[4] J. Saez-Gallego and J. M. Morales, "Short-term forecasting of price-responsive loads using inverse optimization," *IEEE Transactions on Smart Grid*, vol. PP, no. 99, pp. 1–1, 2017.

[5] Z. Liu, W. Gao, Y.-H. Wan, and E. Muljadi, "Wind power plant prediction by using neural networks," in *Energy Conversion Congress and Exposition (ECCE), 2012 IEEE*, pp. 3154–3160, Sept 2012.

[6] A. Jain and D. Zongker, "Feature selection: evaluation, application, and small sample performance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 153–158, Feb 1997.

[7] H. Hippert, C. Pedreira, and R. Souza, "Neural networks for short-term load forecasting: a review and evaluation," *IEEE Transactions on Power Systems*, vol. 16, pp. 44–55, Feb 2001.

[8] B. Zhao, X. Dong, and J. Bornemann, "Service restoration for a renewable-powered microgrid in unscheduled island mode," *IEEE Transactions on Smart Grid*, vol. 6, pp. 1128–1136, May 2015.

[9] H. F. Habib, T. Yossef, M. Cintuglu, and O. Mohammed, "A multi-agent based technique for fault location, isolation, and service restoration," *IEEE Transactions on Industry Applications*, vol. PP, no. 99, pp. 1–1, 2017.

[10] Z. Wang and J. Wang, "Service restoration based on ami and networked mgs under extreme weather events," *IET Generation, Transmission Distribution*, vol. 11, no. 2, pp. 401–408, 2017.

[11] R. F. Sampaio, L. S. Melo, R. P. S. Leão, G. C. Barroso, and J. R. Bezerra, "Automatic restoration system for power distribution networks based on multi-agent systems," *IET Generation, Transmission Distribution*, vol. 11, no. 2, pp. 475–484, 2017.

[12] C. Moreira, F. Resende, and J. P. Lopes, "Using low voltage microgrids for service restoration," *IEEE Transactions on Power Systems*, vol. 22, no. 1, pp. 395–403, 2007.

[13] I. Konstantelos, G. Jamgotchian, S. H. Tindemans, P. Duchesne, S. Cole, C. Merckx, G. Strbac, and P. Panciatici, “Implementation of a massively parallel dynamic security assessment platform for large-scale grids,” *IEEE Transactions on Smart Grid*, vol. 8, pp. 1417–1426, May 2017.

[14] C. Liu, K. Sun, Z. H. Rather, Z. Chen, C. L. Bak, P. Thøgersen, and P. Lund, “A systematic approach for dynamic security assessment and the corresponding preventive control scheme based on decision trees,” in *2014 IEEE PES General Meeting — Conference Exposition*, pp. 1–1, July 2014.

[15] L. H. Fink, K.-L. Liou, and C.-C. Liu, “From generic restoration actions to specific restoration strategies,” *IEEE Transactions on Power Systems*, vol. 10, no. 2, pp. 745–752, 1995.

[16] M. Negnevitsky, N. Tomin, V. Kurbatsky, D. Panasetsky, A. Zhukov, and C. Rehtanz, “A random forest-based approach for voltage security monitoring in a power system,” in *PowerTech, 2015 IEEE Eindhoven*, pp. 1–6, June 2015.

[17] X. Yu and C. Singh, “A practical approach for integrated power system vulnerability analysis with protection failures,” *IEEE Transactions on Power Systems*, vol. 19, no. 4, pp. 1811–1820, 2004.

[18] M. Dehghani, L. Goel, and W. Li, “Pmu based observability reliability evaluation in electric power systems,” *Electric Power Systems Research*, vol. 116, pp. 347–354, 2014.

[19] B. S. Hartono, Y. Budiyanto, and R. Setiabudy, “Review of microgrid technology,” in *QiR (Quality in Research), 2013 International Conference on*, pp. 127–132, June 2013.

[20] A. Parisio, E. Rikos, and L. Glielmo, “A model predictive control approach to microgrid operation optimization,” *Control Systems Technology, IEEE Transactions on*, vol. 22, no. 5, pp. 1813–1827, 2014.

[21] S. Tito, T. Lie, and T. Anderson, “Optimal sizing of a wind-photovoltaic-battery hybrid renewable energy system considering socio-demographic factors,” *Solar Energy*, vol. 136, pp. 525–532, 2016.

[22] H. T. Haider, O. H. See, and W. Elmenreich, “Dynamic residential load scheduling based on adaptive consumption level pricing scheme,” *Electric Power Systems Research*, vol. 133, pp. 27–35, 2016.

[23] P. C. Sekhar, S. Mishra, and R. Sharma, “Data analytics based neuro-fuzzy controller for diesel-photovoltaic hybrid ac microgrid,” *IET Generation, Transmission Distribution*, vol. 9, no. 2, pp. 193–207, 2015.

[24] M. G. Villalva, J. R. Gazoli, *et al.*, “Comprehensive approach to modeling and simulation of photovoltaic arrays,” *IEEE Transactions on Power Electronics*, vol. 24, no. 5, pp. 1198–1208, 2009.

[25] D. Park, M. El-Sharkawi, I. Marks, R.J., L. Atlas, and M. Damborg, “Electric load forecasting using an artificial neural network,” *IEEE Transactions on Power Systems*, vol. 6, pp. 442–449, May 1991.

[26] A. Sankarakrishnan and R. Billinton, “Sequential monte carlo simulation for composite power system reliability analysis with time varying loads,” *IEEE Transactions on Power Systems*, vol. 10, no. 3, pp. 1540–1545, 1995.

[27] N. Qin, C. L. Bak, H. Abildgaard, and Z. Chen, "Multi-stage optimization-based automatic voltage control systems considering wind power forecasting errors," *IEEE Transactions on Power Systems*, vol. 32, pp. 1073–1088, March 2017.

[28] B. M. S. M. Ramadan, T. Logenthiran, R. T. Naayagi, and C. Su, "Hybridization of genetic algorithm and priority list to solve economic dispatch problems," in *2016 IEEE Region 10 Conference (TENCON)*, pp. 1467–1470, Nov 2016.

[29] C. Lin, W. Wu, X. Chen, and W. Zheng, "Decentralized dynamic economic dispatch for integrated transmission and active distribution networks using multi-parametric programming," *IEEE Transactions on Smart Grid*, vol. PP, no. 99, pp. 1–1, 2017.

[30] D. Whitley, "A genetic algorithm tutorial," *Statistics and computing*, vol. 4, no. 2, pp. 65–85, 1994.

[31] T. Athay, R. Podmore, and S. Virmani, "A practical method for the direct analysis of transient stability," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-98, pp. 573–584, March 1979.

[32] A. Rabiee, A. Soroudi, and A. Keane, "Risk-averse preventive voltage control of ac/dc power systems including wind power generation," *IEEE Transactions on Sustainable Energy*, vol. 6, pp. 1494–1505, Oct 2015.

[33] A. Jain and D. Zongker, "Feature selection: evaluation, application, and small sample performance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 153–158, Feb 1997.