IJIE

INTERNATIONAL JOURNAL OF INTEGRATED ENGINEERING
ISSN: 2229-838X e-ISSN: 2600-7916

Vol. 18 No. 1 (2026) 1-18
https://publisher.uthm.edu.my/ojs/index.php/ijie

Predictive Control of Unified Power Quality Conditioner (UPQC) using Random Forest (RF) Algorithm

Muhammad Hasbi Azmi[1], Muhammad Murtadha Othman[1]*, Kerrollenna Martta Madu Anak Kenneth Inggang[1], Norazlan Hashim[1], Ismail Musirin[1], Masoud Ahmadipour[2]

[1] *School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, MALAYSIA*

[2] *Institute of Power Engineering, Department of Electrical and Electronics Engineering, College of Engineering, Universiti Tenaga Nasional (UNITEN), 43000 Kajang, Selangor, MALAYSIA*

*Corresponding Author: mamat505my@yahoo.com
DOI: https://doi.org/10.30880/ijie.2026.18.01.004

Article Info

Received: 12 June 2025
Accepted: 8 February 2026
Available online: 15 April 2026

Keywords

Harmonics, voltage sag, Total Harmonic Distortion (THD), Unified Power Quality Conditioner (UPQC), Proportional gain (P), Integral gain (I), Multiple Time Lag (MTL), Principal Component Analysis (PCA), Random Forest (RF), regression, Root Mean Square Error (RMSE)

Abstract

Power quality is an essential problem that is becoming more significant to energy consumers because non-linear load has been extensively employed in power sectors. Harmonics are produced when nonlinear loads convert AC line voltage to DC, which has detrimental effects on electrical equipment. Conductors and distribution transformers of a power system may overheat if distortion causes an increase in current flow. Temperature increases will reduce the lifespan of electronic gadgets and disrupt power grids if not addressed. Using the Random Forest (RF) algorithm as a machine learning technique, this research predicts the correct value of proportional gain (P) and integral gain (I) to minimize Total Harmonic Distortion (THD) of the source current on the Unified Power Quality Conditioner (UPQC) circuit in the state of voltage sag reduction. There are three case studies on voltage sag reduction at 40%, 30%, and 10%. RMS of source voltage and RMS error in the DC link capacitor voltage regulation circuit that is in the Shunt Controller in the UPQC circuit are recorded. The total number of recorded data is 2,000,000. These data will undergo the multiple time lag process to enhance the data, principal component analysis to reduce the data's dimensionality, and a random forest technique to predict the optimal values of P and I. Using regression analysis and root mean square error (RMSE), the performance of the utilized model will be evaluated. According to the observed results, multiple time lag with a lag value of 24, principal component analysis with a value of 0.000001, tree and leaf values of 100 and 2 for P, and 10 and 5 for I, will create more accurate results. For testing and validating P, the chosen RMSE values are 6.0392e-10 and 7.8365e-13. Meanwhile, the testing and validation RMSE values for I are 1.4738-10 and 1.1427e-13, respectively. The regression result for both instances P and I is equal to 1 considering the optimal setup. Finally, the THD value of the source current was reduced successfully.

1. Introduction

The increased usage of power electronics equipment, such as variable speed drives (VSD) and programmable logic controllers (PLC), has resulted in a complete transformation of the nature of electric loads. These loads are both the principal causes and victims of power quality problems [1]. All of these loads cause waveform disturbances due to their nonlinearity [2]. Electrical equipment is sensitive to voltage sags and swells, interruptions, and waveform distortions, such as harmonics.

Harmonics caused by non-linear loads must be minimized [3]. A load is called non-linear when the shape of the current it absorbs differs from the shape of the voltage it receives. Harmonics are produced by non-linear loads that convert AC line voltage to DC [4]. Electrical equipment is negatively impacted by harmonic distortion. Unwanted distortion may elevate the current in power systems, hence increasing the temperature of neutral conductors and distribution transformers. If left uncontrolled, increased heat and interference can significantly limit the lifespan of electronic equipment and cause damage to power supplies.

UPQC is a common dynamic power shaping tool used to reduce voltage and current harmonics in a polluted power system network, says [5]. UPQC can balance out voltage spikes, compensate for voltage variations, and prevent harmonic load current from affecting the power system. This indicates that UPQC is crucial in addressing issues with power quality. Shunt and series compensation make up UPQC. A series compensator is used to control voltage disturbances, while a shunt compensator is utilized to manage current disturbances, according to [6].

The proportional-integral (PI) controller is located in the DC link capacitor voltage regulation circuit, which is part of the Shunt Controller in the UPQC circuit [7]. The difference, Verror, between the total DC-link voltage signal, Vdc, and its desired reference voltage, Vdc ref, is controlled by a proportional-integral (PI) controller. The DC link voltage regulation circuit then creates an estimated output, Idc, which is intended to serve as the primary control signal for the DC-link voltage level [8]. The voltage of the entire DC link can be altered by adjusting the proportional gain, Kp, and integral gain, Ki. The estimation of P and I values in the UPQC controller leads to difficulties such as a THD value at the source current that is uncertain and above 5% [9]. The performance of the UPQC must be enhanced by guaranteeing that the injected P and I values remain fixed regardless of the load value and in the event of voltage sag. The value of the THD percentage at the source current must always be below 5% [10].

Conventional techniques for tuning PI controllers such as trial-and-error can prove inadequate in dynamic and nonlinear contexts resulting in unsatisfactory performance [11]. This method is difficult because when P and I are large, DC-bus voltage regulation takes over and steady-state dc-bus voltage error is minimal. When P and I are minimal, the real power imbalance does not affect transient performance. Dogruer and Tan [12], employ optimization approaches in the design of PI controllers. This technique has various drawbacks, including the fact that it uses sophisticated algorithms that are difficult to execute effectively and may be subject to numerical noise. Huang [13], said that using a high I value in a PI regulator can lead to increased system responsiveness but may also cause overshoot and oscillations, potentially compromising overall stability. Careful tuning is essential to balance performance and stability in various application scenarios. According to Priya and Prathyusha [14], the PI controller's value is also established by the trial-and-error technique making the PI controller less effective than the fuzzy logic controller. Afrawira et al. [15], also employ trial and error to determine P and I values in order to increase step responsiveness. This approach will be difficult because too much increase in proportional gain leads to control system instability and too much reduction decreases control system response speed. As a result, the RF technique should be employed to address such issues.

The objectives of this project are to predict the right value of proportional gain (P) and integral gain (I) using the RF algorithm and to reduce THD on the UPQC circuit during voltage sag reduction. This study presents the prediction of control parameters P and I during the reduction of voltage sag by 40%, 30%, and 10%, respectively. Multiple time lags, principal component analysis, and the random forest method are used for prediction. Using the above method, a total of 2,000,000 collected data points will be processed. The RMS source voltage and RMS error from the DC-link Capacitor Voltage Regulation circuit are recorded for the accurate prediction of P and I. These values were derived from the UPQC circuit. Multiple time lags assist in improving the raw data and organising it as a matrix [16]. Principal component analysis is utilised to minimise the dimensionality of huge datasets [17]. In this study, random forest is the primary strategy employed. This method is a machine learning technique that predicts the optimal P and I in the state of voltage sag reduction. The performance outcome can be determined by examining the value of regression analysis and root mean square error (RMSE). After determining the optimal P and I values, the DC-link Capacitor Voltage Regulation controller circuit on the UPQC circuit will be set with these values.

This paper will be structured as follows: Section II explains the suggested forecasting model's theory, algorithm, step, and formulation. In Section III, the results and explanation of the random forest simulation for the best P and I forecast values are presented. Section IV will finish with a summary of the findings, while section V will outline future recommendations.

2. Methodology

This research aimed to predict the correct values of control parameters, proportional (P) and Integral (I) in the condition of voltage sag using machine learning techniques, namely the Random Forest method [18]. The control parameters, P and I, are found in the DC-link Capacitor Voltage Regulation circuit, which is one of the controllers in the UPQC circuit. Several techniques, including Multiple Time Lags (MTL), Principal Component Analysis (PCA), and adjusting the number of trees and leaves, have been implemented to ensure that the prediction is accurate. Hence, the procedure will be described in detail in this section. Figure 1 shows the overall flowchart.

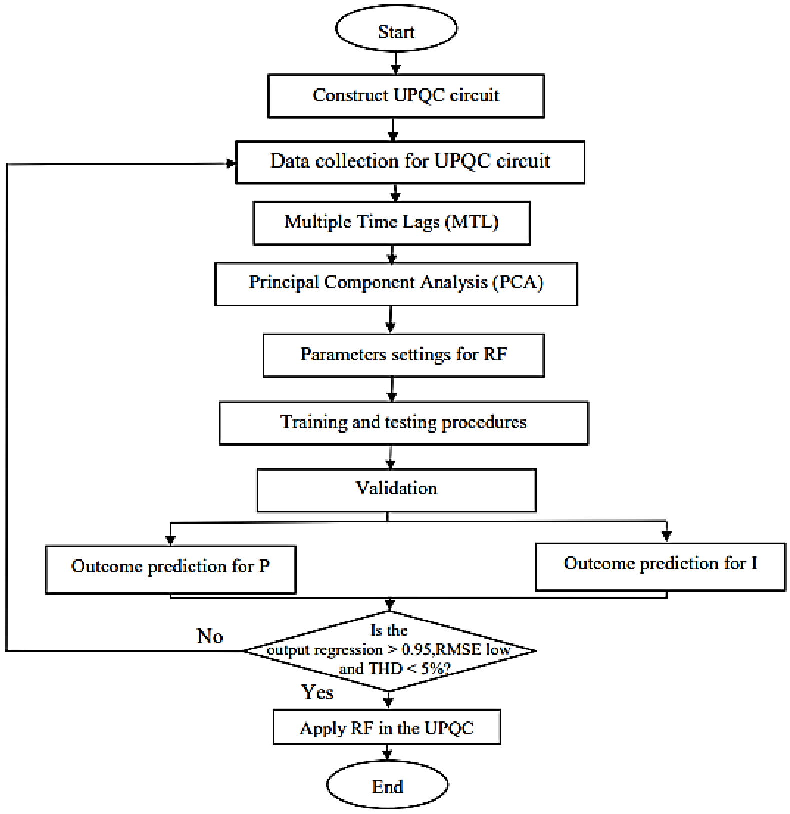


Fig. 1 *Flowchart of Prediction of Non-critical Voltage Sag using Random Forest (RF) based on Multiple Time Lags (MTL) integrated Principal Component Analysis (PCA)*

The operation began with the data collection process. Several data have been acquired from the simulation of the UPQC circuit that was executed on Simulink using MATLAB. Among the values of raw data obtained from the circuit are the RMS value of the supply voltage and the RMS error identified in the DC-link Capacitor Voltage Regulation circuit.

The data is then entered into the Multiple Time Lags (MTL) process, which is transformed and organised data into a matrix [19]. After being organised in the form of a matrix, the data is subjected to principal component analysis. Principal component analysis aims to utilise compressed data in order to minimise the number of variables in the matrix [20]. The random forest technique is then used to anticipate the right values of P and I for the DC-link Capacitor Voltage Regulation circuit, one of the controllers of the UPQC circuit. To apply this RF approach, the first step is to establish the RF parameters, which is done by setting the values of the trees and leaves. The data will then be put into the testing, training, and validation processes. From that, the prediction value for P and I will be created. The resulting values of P and I will be sent into the DC-link Capacitor Voltage Regulation circuit, which is one of the UPQC circuit's controllers.

2.1 Unified Power Quality Conditioner (UPQC)

The participant should sit on the car seat with the The UPQC adjusts the voltage instability on the grid to deliver a proper load voltage. Other than that, the load current is adjusted to reach the necessary source current. Thus, in accordance with the Kirchhoff rules, the fundamental compensation equations are:

$$v_L(t) = v_s(t) - v_{inj}(t) \tag{1}$$

$$i_s(t) = i_L(t) - i_{inj}(t) \tag{2}$$

Where:
$v_L(t)$: mitigated load voltage
$v_s(t)$: disturbed source voltage
$v_{inj}(t)$: voltage injected from UPQC
$i_s(t)$: source current
$i_L(t)$: load current
$i_{inj}(t)$: current injected by a shunt converter

Figure 2 shows a general block schematic of a UPQC circuit [7]. Initially, it contains simply the voltage source and the load. Due to UPQC, a series transformer, series inverter and shunt inverter have been added to the circuit.

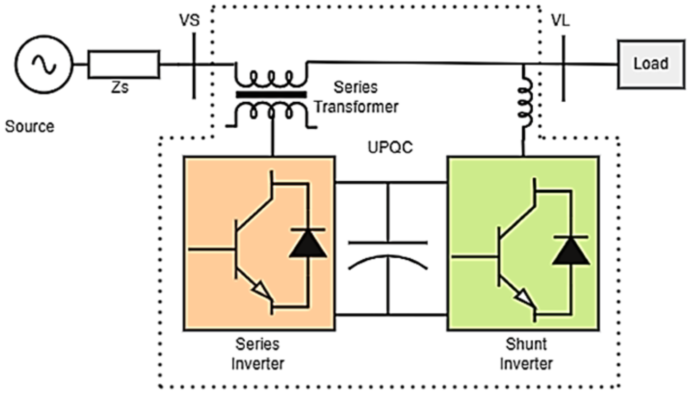


Fig. 2 *Block diagram of UPQC circuit*

Several values have been set on the built-in circuit. A three-phase programmable voltage source with an amplitude of 100V and a frequency of 50Hz has been configured. The inductance value of the line has been set at 10mH. In a non-linear load, a universal bridge has been configured with a snubber resistance of 500, a snubber capacitance of 250nF, a resistor (on) of 1mΩ, and a forward voltage of 0.8V. This block implements a power electronic device bridge, and it must be linked in parallel with the electronic devices. Following that, the series RLC block's R and L values were set to 100Ω and 100mH, respectively. The circuit of non-linear load is depicted in Figure 3.

The transformer has also been configured with a nominal power of 1000VA and a frequency of 50Hz. Winding 1 and 2 parameters have been set to Vrms 100V, with resistance and inductance set at 0.02p.u. for both windings.

UPQC consists of an active series filter and an active shunt filter. At the point of common coupling (PCC), between utility and consumer, the series active filter can provide voltage imbalance adjustment, voltage control, and harmonic correction. The shunt active filter's primary function is to mitigate current harmonics, make up the difference for reactive power and negative-phase sequence current, and control the DC link voltage between the series and shunt active filters.

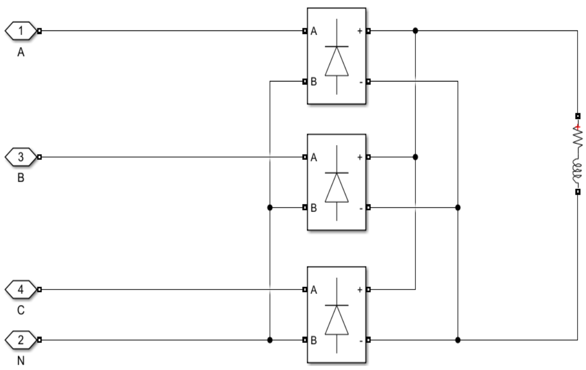


Fig. 3 *Non-linear load circuit*

The Series Active Power Filter is connected between the supply and load terminals by using three single-phase transformers arranged in star across the inverter and in series with the source lines. In addition to providing voltage, these transformers also regulate the switching ripple of the active series filter. The series filter is calibrated to insert voltages that cancel out imbalances in the supply voltages, resulting in perfectly balanced and sinusoidal voltages with the proper amplitude at the PCC. In other words, the needed load terminal voltage is equal to the combination of the supply voltage and the voltage supplied by the series filter.

Next, the Series Active Power Filter circuit was successfully constructed. Figure 4 depicts a series controller developed using Simulink using MATLAB software.

A Shunt Active Power Filter (SAPF) is connected to a harmonically polluted power grid at the point of common coupling (PCC), between the voltage source and a harmonic-generating load. This filter consists of two key elements which are a standard two-level input voltage inverter and controller. The controller's four basic control algorithms are harmonic extraction, DC-link capacitor voltage regulation, current control, and synchronizer. Figure 5 shows the Shunt Active Power Filter controller [7].

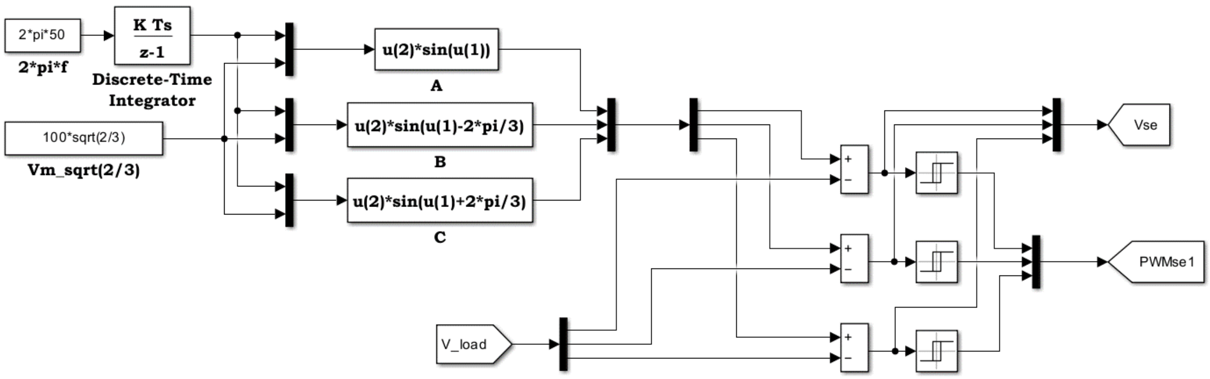


Fig. 4 *Series Active Power Filter circuit that was built using Simulink*

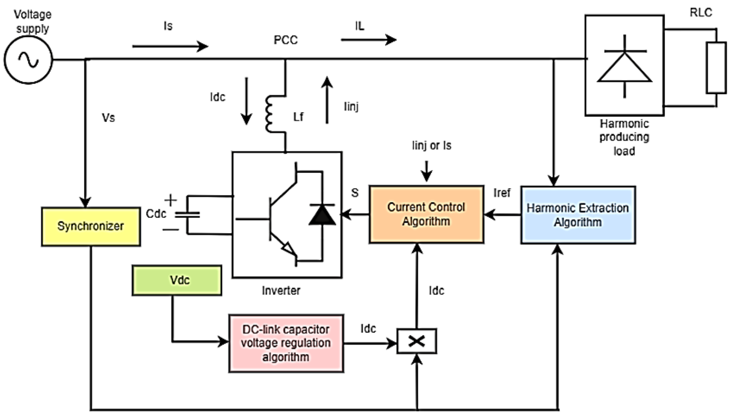


Fig. 5 *Shunt Active Power Filter controller*

2.2 Data Collection

A total of 2,000,000 values of RMS voltage source and RMS error values have been collected and recorded from the UPQC circuit. The data is derived from the Workspace part of the MATLAB software after the circuit has been executed. There have been specified three stages of voltage sag reduction. The first is a 40% decrease in voltage sag, followed by 30% and then 10%.

2.3 Multiple Time Lags

After the data has been saved and recorded, multiple times lag processes are performed to get more accurate predictions [21]. With that, raw data will be sorted according to time intervals. The input data will be placed in a k-by-t matrix throughout this procedure. Each accumulated data will be used to generate the right predicted value for the controller variables Proportional, P, and Integral, I. In this instance, the acquired input data, which are the RMS voltage source and RMS error, has been used in the MTL process. Equation (3) displays the arrangement of input data in the MTL process that will be utilized as training and target data.

$$Z_{k,t} = X_t - X_{t-k} \tag{3}$$

Where:
t : time interval
k : lagging time interval, which is 1, 2, 3,…., K
K: total number of lagging time intervals

2.4 Principal Component Analysis (PCA)

After MTL has been completed, PCA is used to minimize the dimension of the dataset. This is accomplished by changing data originally structured in a matrix to a new coordinate system in which the majority of data variance may be reduced to fewer dimensions than the original data. This is done to ease the process of training, testing, and validation that will be included in the RF method later.

Mathematically, the PCA procedure is separated into multiple phases. The first PCA phase is covariance matrix computation:

$$\widehat{\sum} = n^{-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)(X_i - X_n)^T \tag{4}$$

Next, calculate eigenvalues:

$$\lambda_1 \geq \lambda_2 \geq \ \dots \text{and eigenvectors } e_1, e_2, \dots \text{ of } \widehat{\sum}. \tag{5}$$

The next stage is to determine the dimension, k.
PCA concludes with the definition of dimension-reduced data:

$$Z_i = T_k(X_i) = \bar{X} + \sum_{j=1}^{k} \beta_{ij} e_j \tag{6}$$
$$where\ \beta_{ij} = (X_i - \bar{X}_{,} e_j)$$

2.5 Random Forest (RF) Algorithm

Random Forest is an effective machine-learning technique for handling Classification and Regression problems. RF consists of many decision trees that are tested on different subgroups of a given dataset and averaging their predicted accuracy [22].

Supervised Learning (SL) is the machine learning employed for this research. This form of machine learning enables users to train models with a large amount of data. Training data for SL includes both input and target values [23]. The algorithm identifies a pattern that translates input values to output values and utilises this pattern to predict future values.

Figure 6 depicts the steps of the Random Forest operation. The first step is to randomly choose samples from a specific data or training set. Step 2 of this approach involves the construction of a decision tree for each training set. The third step is to vote by averaging the decision tree. Finally, choose the prediction result with the most votes.

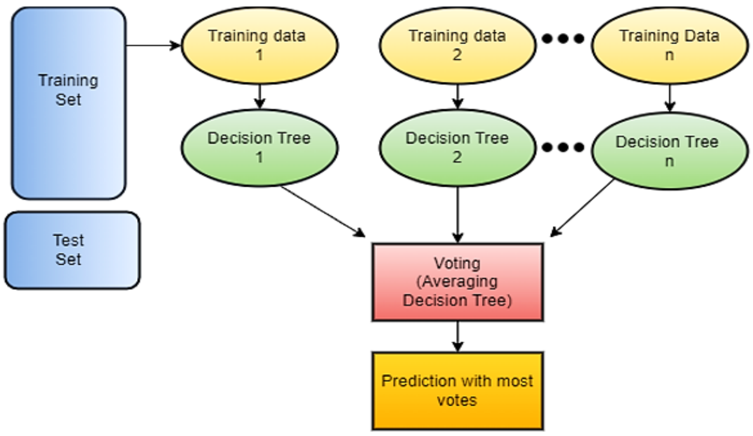


Fig. 6 *Operation of Random Forest algorithm*

This collection of various trees is referred to as an Ensemble. Bagging is one of the ways for Ensemble. Bagging refers to the procedure of generating a distinct training subset from training data samples with replacement. The ultimate result is determined by the majority vote. Bagging is sometimes referred to as Bootstrap Aggregation, which is used by RF. The procedure starts with any random data of origin. After arrangement, it is organised into Bootstrap Sample samples. This is referred to as "bootstrapping." Furthermore, every model is trained separately, generating valued outcomes defined as Aggregation. In the last phase, all results are merged, and the outcome is determined by a majority vote. This process is referred to as Bagging and is performed using an Ensemble Classifier. Figure 7 depicts the general bagging process.

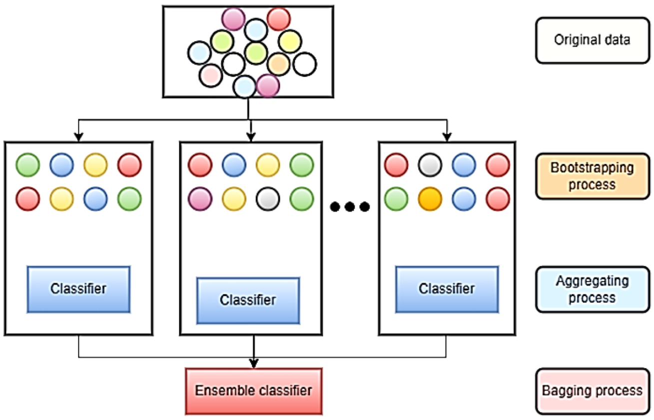


Fig. 7 *Bagging process*

When Bootstrap Aggregating or Bagging is carried out, two separate sets are generated. One set is the bootstrap sample, which consists of the data selected via sampling with replacement to be "In-the-Bag". Another set is the Out-of-Bag (OOB) set, consisting of all data that were not selected during the sampling procedure.

The In-Bag procedure is used to construct a forest where the "Tree-growth" method produces the "great leaves" or can be called as best result. The Out of Bag (OOB) data is utilised to run the prediction error since trees are placed in the forest during the tree growth process, utilising the In-Bag dataset. The key job of OOB in the tree growth approach is to evaluate its estimated value with the anticipated values acquired from the In-Bag. This process is done to identify the best leaves with the lowest prediction error from each tree.

The OOB procedure will be explained in detail in this section. The OOB score is a measure of the number of values successfully predicted on the validation dataset. The validation data is a subset of the bootstrapped sample data that is provided to the bottom models. So, for each bottom model, validation data will be gathered, and each bottom model will be trained using bootstrapped samples. Once all of the bottom models have been trained on the fed selection, the validation samples will be utilised to determine the bottom models' OOB error. Figure 8 gives further context for the Bagging approach.

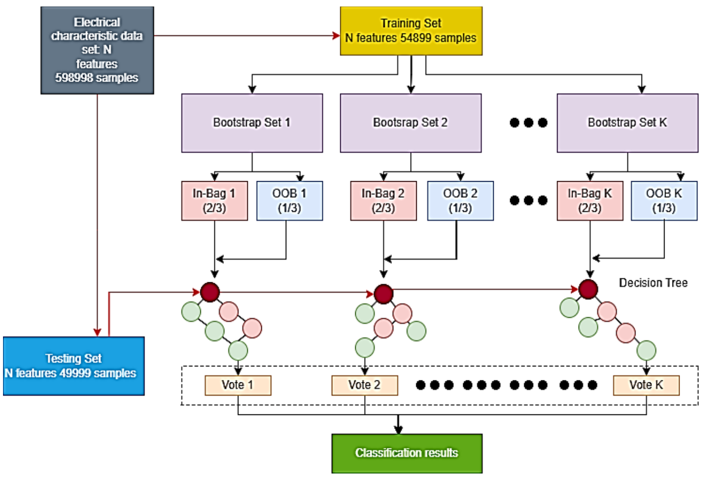


Fig. 8 *Bagging process in detail*

According to the Figure 8, the dataset sample comprises 54899 samples, of which a few bootstrapped samples will be supplied to the bottom model for training. The validation part of the data from the bootstrap samples will now be collected as OOB samples. These bottom models will be trained on the remaining bootstrap data, and once trained, the bottom models will be utilised to predict the OOB samples. Once the bottom models have predicted the OOB samples, the OOB score will be computed. The exact technique will now be followed for all of the bottom models. Hence, depending on the OOB error, the model's performance will improve

The average value of error over all decision trees is the Out-of-Bag error (OOB). OOB error can be calculated as follows:

1. Consider the j^{th} decision tree, DT_j which was fitted to a part of the sample data.
2. For each observation or sample used in training, $z_i$is not included in the sample subset. The equation of z_i is as follows:

$$z_i = (x_i , y_i) \tag{7}$$

3. DT_j is used to predict the outcome, o_i for x_i.
4. The error, e can be calculated as follows:

$$e = |o_i - y_i| \tag{8}$$

5. The average error over all decision trees is thus the OOB error.

Random forests employ hyperparameters to improve the performance and prediction ability of models or to speed up the model. There are two hyperparameters used in this case. The first is the number of trees, often known as the number of estimators. The number of trees constructed by the algorithm prior to averaging the results. The second one is a mini sample leaf. It determines the minimum needed number of leaves to separate an internal node.

Random Forest Algorithm has several advantages, but one of the most significant is that it decreases the danger of overfitting and the needed training time. In addition, it gives a high degree of precision. By approximating missing data, the RF method operates rapidly on big datasets and generates extremely accurate predictions.

2.6 Performance Assessment

To determine the inaccuracy in prediction, the Root Mean Square Error (RMSE) value has been determined. Root Mean Square Error (RMSE) is the prediction error standard deviation. The prediction error is the difference between data points and the regression line. Prediction and regression analysis often use RMSE to validate experimental findings.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \widehat{Y_\iota}\right)^2} \tag{9}$$

Where:
$\hat{Y}$: Predicted value of P and I
Y : Actual P and I value

Regression analysis is a collection of statistical procedures used in statistical modelling to estimate the correlations between the expected value and one or more variables. The most prevalent kind of regression analysis is linear regression, in which the line that best fits the data based on a specified mathematical criterion is determined. A decent linear regression score is 1. A RMSE value of 0 would imply that the model fits the data perfectly.

$$R = \frac{n\,(\sum xy) - \sum x\,\sum y}{\sqrt{[n*(\sum x2 - (\sum x)2)] * [n*(\sum y2 - (\sum y)2)]}} \tag{10}$$

R : Regression
n : total number inside the dataset
x : first variable
y : second variable

Int. Journal of Integrated Engineering Vol. 18 No. 1 (2026) p. 1-18

3. Results

This section will be divided into two parts. The first section shows several waveforms that were collected from the UPQC circuit, and the second section is the result of the accuracy of the anticipated value control parameters, Proportional (P) and Integral (I).

After the simulation, several waveforms were collected from the UPQC circuit. Figure 9 depicts the source, injected, and load voltage waveform. The Vload waveform was effectively stabilised with the assistance of Vinj, as shown by the waveform. There is a voltage sag occur between 0.1s to 0.2s. However, UPQC can mitigate voltage sag and control the grid's voltage instability to provide a suitable load voltage.

In addition, the source, injected, and load current waveforms have been observed. The outcome is shown in Figure 10. The waveform shows that the current is unstable between 0.1s and 0.2s. However, with the aid of the UPQC circuit, the load current is adjusted to obtain the necessary source current.

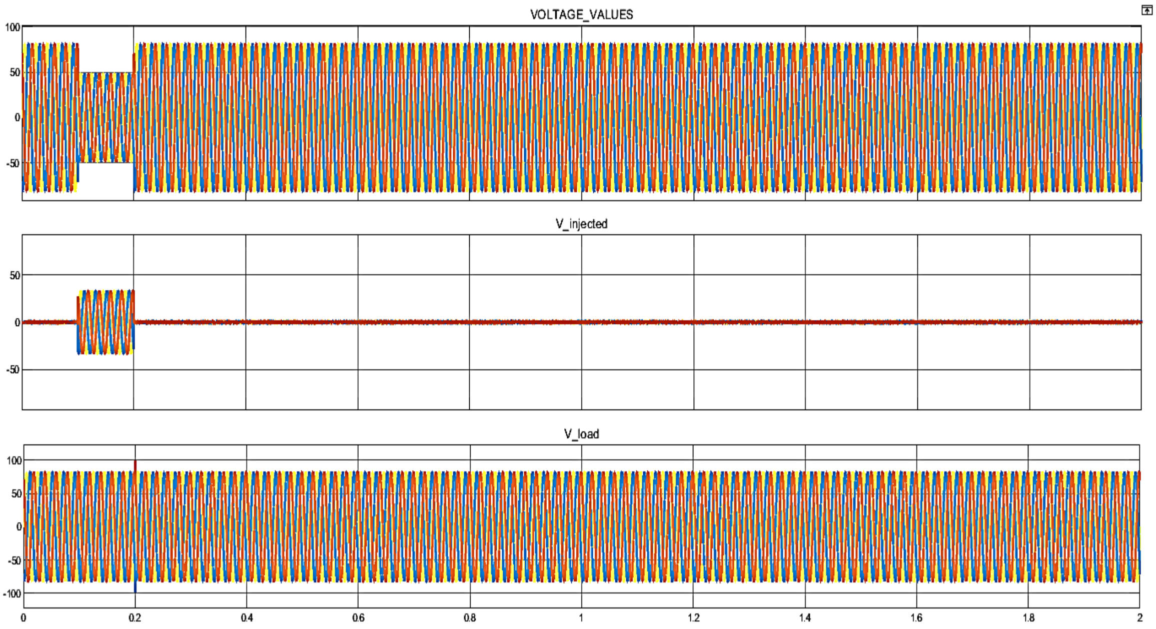


Fig. 9 *Waveform of source, injected and load voltage*

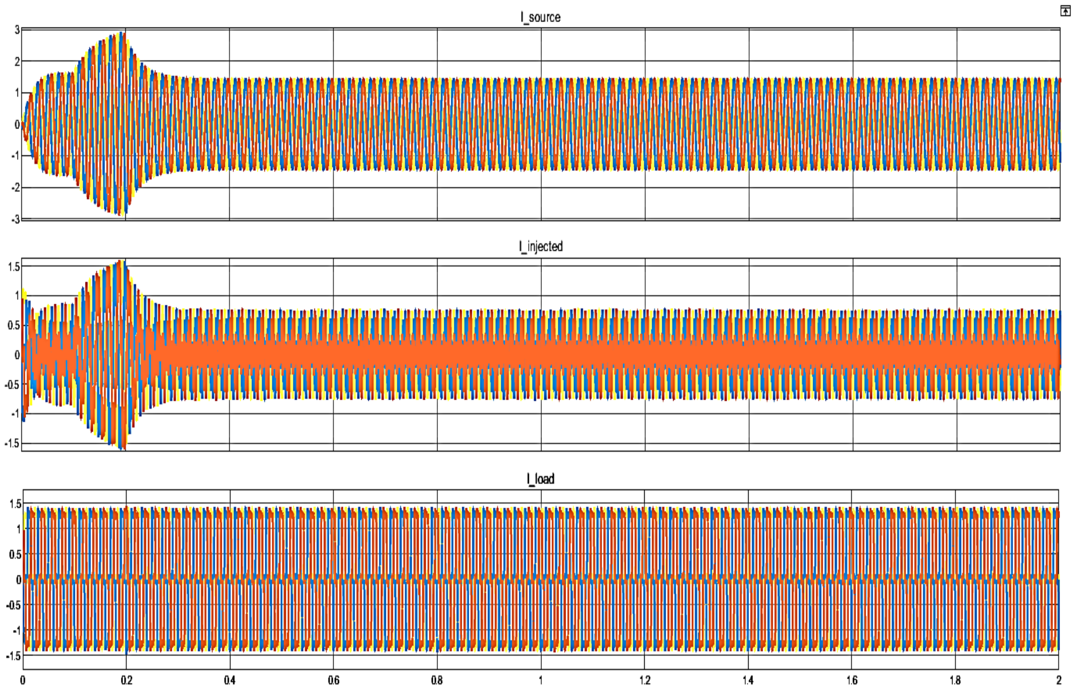


Fig. 10 *Waveform of source, injected and load current*

The RMS source voltage phase 1 waveform as shown in Figure 11 is also recorded since the value will be employed in the next procedure, which is the P and I prediction process using RF. The following waveform demonstrates that the voltage sag happens between 0.1 and 0.2 seconds and recovers to stability with the aid of the UPQC circuit.

Other than that, the waveform of the DC link voltage regulation circuit has been examined as shown in Figure 12. Several values, including the RMS reference value, RMS capacitor value, RMS error value, and output PI value, have been measured. The RMS error value will be utilized in the next process.

Furthermore, this section will also assess the accuracy of the anticipated value control parameters, Proportional (P) and Integral (I). Based on the regression findings and the Root Mean Square Error (RMSE) value, the accuracy will be determined.

There have been specified three levels of voltage sag decrease. The first decrease is 40%, followed by 30%, and lastly 10%. For the voltage sag to be minimized, the anticipated values of P and I must reflect the right values for the three situations of voltage sag reductions.

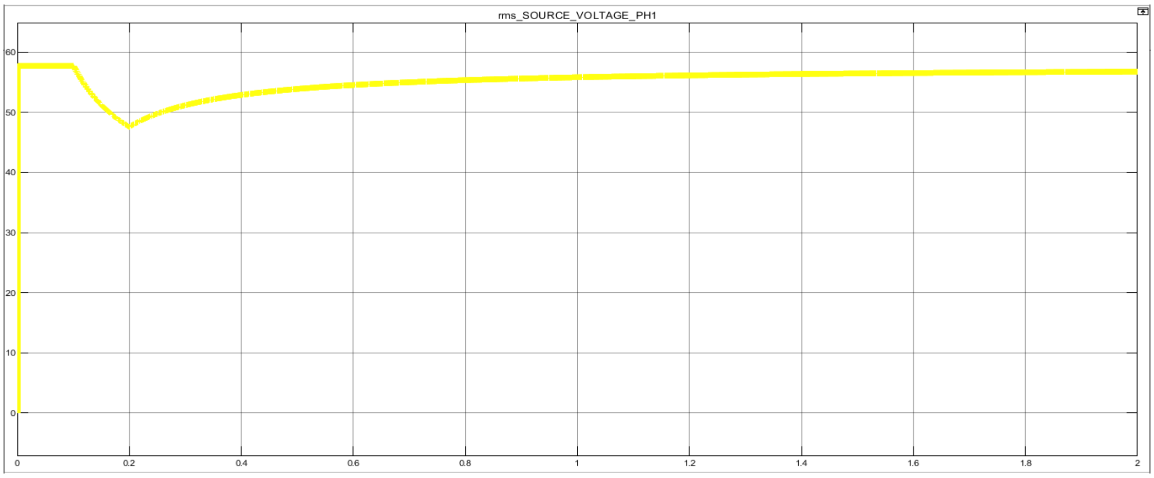


Fig. 11 *Waveform of RMS source voltage phase 1*

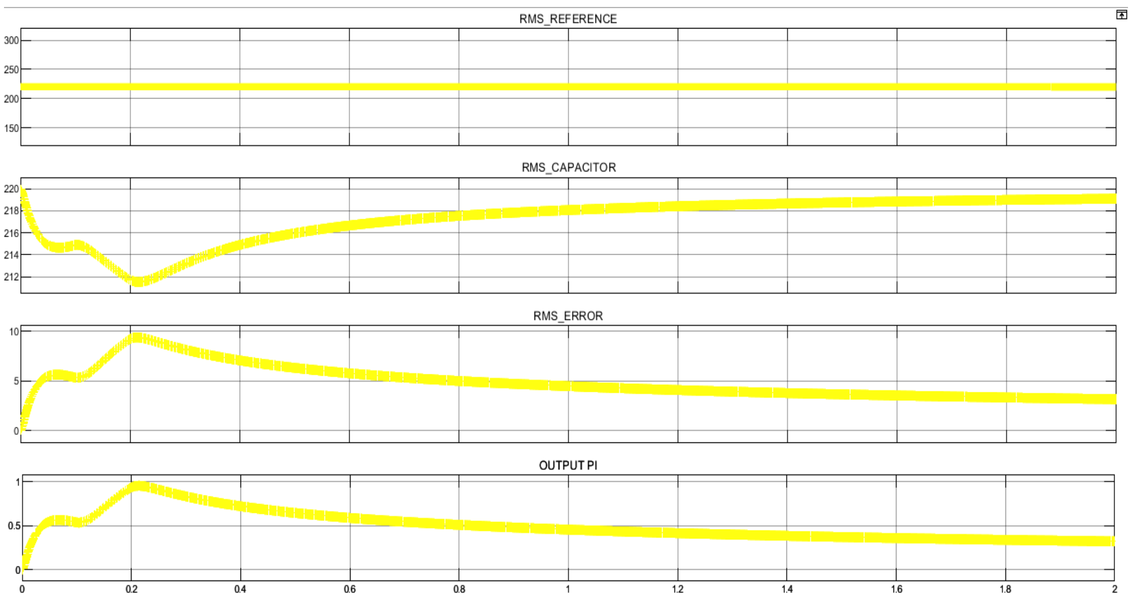


Fig. 12 *DC link voltage regulation circuit waveform*

Table 1 shows the configuration of data set utilized in RF model. Initially, 2,000,000 data of RMS source voltage and 2,000,000 data of RMS error were collected from the UPQC circuit for three levels of voltage sag reduction, which are 40%, 30%, and 10%. The value is then fed into the MTL process. For each of the three cases, MTL generates 1999954 values for RMS source voltage and 1999954 values for RMS error. However, only 600,000 data will be employed from the RMS source voltage and RMS error from the three cases that will be included in the training, testing, and validation procedure using the RF algorithm. The techniques used in this RF process are TreeBagger and Bootstrapping.

Table 1 *Configuration of data set utilized in RF model*

Control Parameters	Proportional (P)	Integral (I)
Total data before MTL	2000000	2000000
Total data after MTL	1999954	1999954
Total no. of data for training, testing and validation using RF	600000	600000
Sets of training data	548999	548999
Sets of testing data	49999	49999
Sets of validation data	999	999
Training function	TreeBagger & Bootstrapping	TreeBagger & Bootstrapping

3.1 Data Collection Through UPQC Circuit and Multiple Time Lag

To enhance the raw data, the RMS source voltage and RMS error from the DC-link Capacitor Voltage Regulation circuit have been extracted from the UPQC circuit and placed into the multiple time lag process with K=24, K=48, and k=72. Table 2 shows the results of RF with various K values in Multiple Time Lags (MTL).

According to Table 2, the lowest value of RMSE during the testing and validation process is when K = 24. For P, the RMSE testing value is 1.1282e-09 and the RMSE validation value is 1.7929e-11. While for I, the testing RMSE value is 4.4721e-07 and the RMSE validation value is 7.4828e-10, respectively.

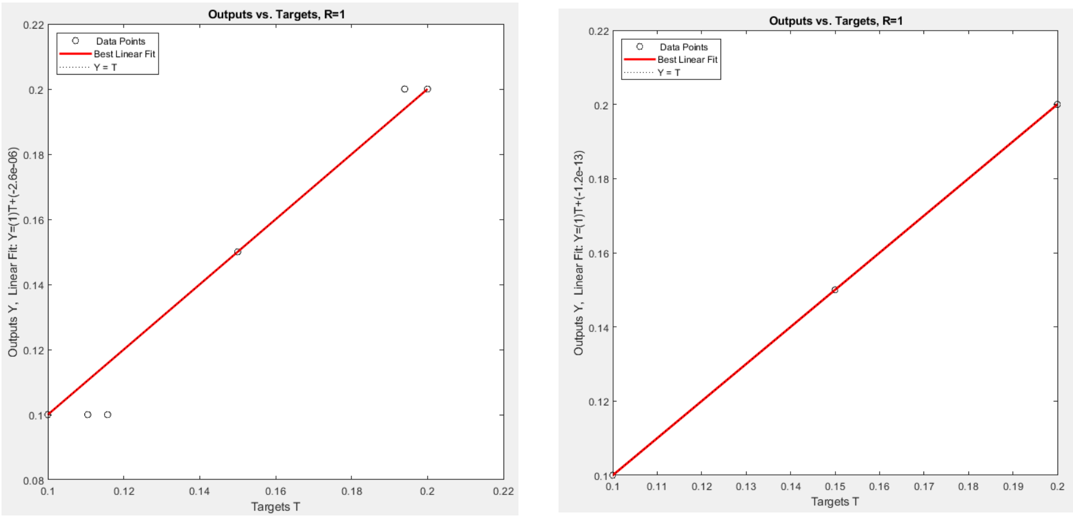
Table 2 *Results of RF with various K values in Multiple Time Lags (MTL)*

MTL	24		48		92	
Control Parameters	P	I	P	I	P	I
Regression Testing	1	1	1	1	1	1
Regression Validation	1	1	1	1	1	1
RMSE Testing (%)	1.1282e-09	4.4721e-07	2.0488e-09	5.3246e-07	1.2698e-09	5.5136e-07
RMSE Validation (%)	1.7929e-11	7.4828e-10	7.7854e-11	8.442e-10	7.6229e-11	8.1625e-10

Figure 13 show the regression result after the testing and validation process for P with K=24 in MTL. The findings reveal that the regression results for both processes are equal to 1.

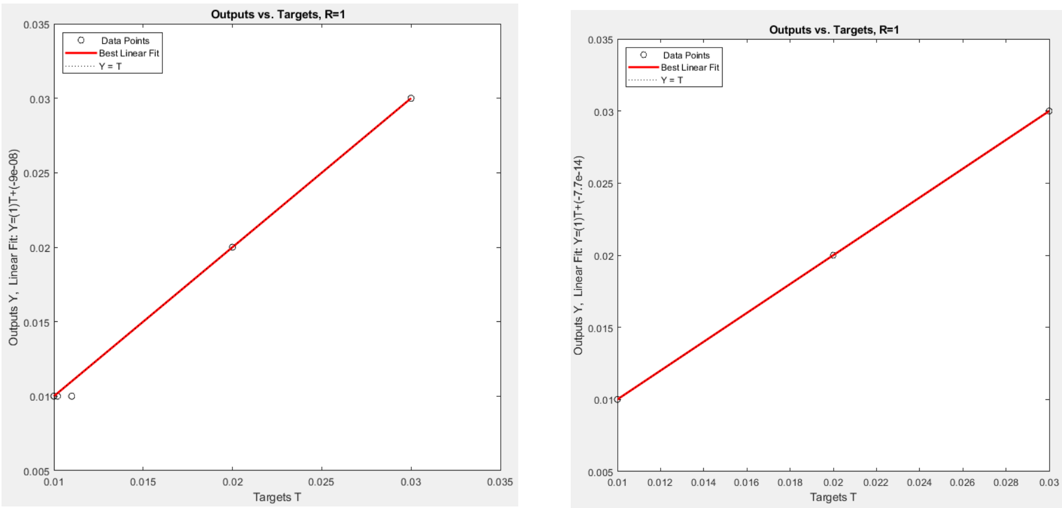
Figure 14 show the regression value after the testing and validation process for I with K=24 in MTL. The outcome reveals that the regression value for both procedures is likewise 1.

Tables 3 and 4 show the results of testing and validating the RF model to anticipate the right value of P and I, given that the dimension of the data has been decreased by methods of principal component analysis. Based on the results of the analysis, the performance is improved by setting a value of 0.0000001% on PCA. The RMSE testing value selected for P is 1.1182e-09, whereas the RMSE validation value is 3.8415e-12. The selected RMSE testing value for I is 1.4721e-08, whereas the RMSE validation value is 1.4420e-11.



(a) **(b)**

Fig. 13 *Regression results for P (a) after testing process; and (b) after validation process with K=24 in MTL*



(a) **(b)**

Fig. 14 *Regression results for I (a) after testing process; and (b) after validation process with K=24 in MTL*

Table 3 *The result of RF for P after dimension reduction using PCA*

Control Parameters	Proportional (P)				
MTL	24				
PCA (%)	0.01	0.001	0.0001	0.00001	0.000001
Regression Testing	1	1	1	1	1
Regression Validation	0.996	0.9999	1	1	1
RMSE Testing (%)	7.3280e-08	7.1866e-08	1.2996e-08	2.0317e-09	1.1182e-09
RMSE Validation (%)	4.0282e-12	3.9831e-12	4.8229e-12	4.83207e-12	3.8415e-12

Table 4 *The result of RF for I after dimension reduction using PCA*

Control Parameters	Integral (I)				
MTL	24				
PCA (%)	0.01	0.001	0.0001	0.00001	0.000001
Regression Testing	1	0.999	1	1	1
Regression Validation	1	1	1	1	1
RMSE Testing (%)	8.9604e-07	1.5619e-08	2.9665e-08	6.3246e-08	1.4721e-08
RMSE Validation (%)	8.6441e-11	2.1698e-11	1.5582e-11	1.5399e-11	1.4420e-11

3.2 Number of Trees and Leaves

The procedure for establishing the value of trees and also leaves in the RF is then executed. The value of trees has increased from 10 to 100, but the value of leaves remains the same. The RMSE value is then evaluated to get the optimal tree value.

According to Tables 5 and 6, the RMSE is smallest when tree=100 for P and tree=10 for I. This indicates that the specified number of trees will be used by the RF model to provide accurate P and I prediction values. For P, the selected value for testing and validation RMSE are 1.1074e-09 and 2.8826e-12. Meanwhile, the RMSE testing and validation values for I are 1.2721e-08 and 1.1242e-12. Next, the value of leaves is also varied to get the best leaves value.

From Tables 7 and 8, the values leaves=2 for P and leaves=5 for I were chosen. This is because this value has the lowest RMSE and may assist in obtaining the right predicted values of P and I for the

RF model. The selected RMSE values for testing and validation of P are 6.0392e-10 and 7.8365e-13. Meanwhile, the RMSE values for testing and validation for I are 1.4738-10 and 1.1427e-13, respectively.

Table 9 displays the ultimate outcome of many analysis methods. Observations indicate that by using K=24 on MTL and 0.000001% on PCA, the number of trees and leaves in the RF model will be 100 and 5 for P and 10 and 5 for I, resulting in a more precise predicted value. RMSE value is also very minimal. For testing and validating P, the chosen RMSE values are 6.0392e-10 and 7.8365e-13. Meanwhile, the testing and validation RMSE values for I are 1.4738-10 and 1.1427e-13, respectively.

Table 5 *The outcome of RF for P by considering various tree values*

Control Parameters	Proportional (P)				
MTL	24				
PCA (%)	0.000001				
Number of Trees	10	30	50	70	100
Number of Leaves	2	2	2	2	2
Regression Testing	1	1	1	1	1
Regression Validation	1	1	1	1	1
RMSE Testing (%)	2.1082e-09	4.7591e-09	2.6773e-09	2.1041e-09	1.1074e-09
RMSE Validation (%)	3.8229e-12	3.9405e-12	4.1028e-12	3.9832e-12	2.8826e-12

Table 6 *The outcome of RF for I by considering various tree values*

Control Parameters	Integral (I)				
MTL	24				
PCA (%)	0.000001				
Number of Trees	10	30	50	70	100
Number of Leaves	2	2	2	2	2
Regression Testing	1	1	1	1	1
Regression Validation	1	1	1	1	1
RMSE Testing (%)	1.2721e-08	4.9316e-08	8.4023e-08	5.5696e-08	6.3989e-08
RMSE Validation (%)	1.1242e-12	1.1369e-12	1.1383e-12	1.1616e-12	1.1476e-12

Table 7 *The outcome of RF for P by considering various leaves values*

Control Parameters	Proportional (P)				
MTL	24				
PCA (%)	0.000001				
Number of Trees	100	100	100	100	100
Number of Leaves	2	5	7	15	20
Regression Testing	1	1	1	1	1
Regression Validation	1	1	1	1	1
RMSE Testing (%)	6.0392e-10	2.3196e-09	1.7591e-09	1.5659e-09	1.7860e-09
RMSE Validation (%)	7.8365e-13	4.0086e-13	3.8318e-13	6.7008e-13	3.9252e-13

Table 8 *The outcome of RF for I by considering various leaves values*

Control Parameters	Integral (I)				
MTL	24				
PCA (%)	0.000001				
Number of Trees	10	10	10	10	10
Number of Leaves	2	5	7	15	20
Regression Testing	1	1	1	1	1
Regression Validation	1	1	1	1	1
RMSE Testing (%)	1.3082e-10	1.4738-10	2.6386e-09	4.7106e-09	2.3920e-09
RMSE Validation (%)	1.1033e-13	1.1427e-13	1.4849e-12	1.1521e-12	1.6827e-12

Table 9 *RF result based on the optimal setup*

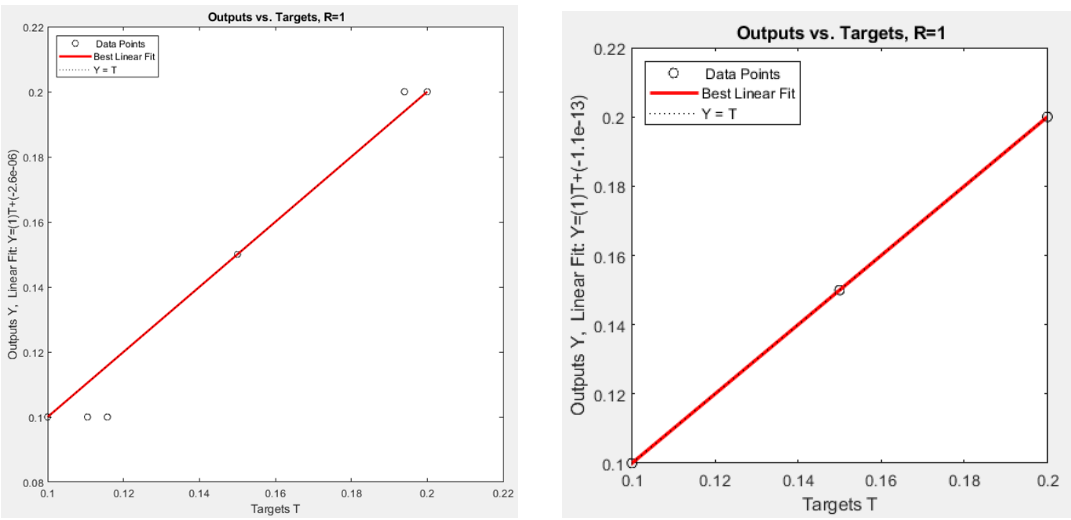
Control Parameters	Proportional (P)	Integral (I)
MTL	24	24
PCA (%)	0.000001	0.000001
Sets of training data	548999	548999
Sets of testing data	49999	49999
Sets of validation data	999	999
Training function	TreeBagger & Bootstrapping	TreeBagger & Bootstrapping
Number of trees	100	10
Numbers of Leaves	2	5
Regression Testing	1	1
Regression Validation	1	1
RMSE Testing (%)	6.0392e-10	1.4738-10
RMSE Validation (%)	7.8365e-13	1.1427e-13

Figure 15 show the regression result after the testing and validation process for P based on the optimal setup of the RF model. The findings reveal that the regression results for both processes are equal to 1.

Figure 16 depict the outcome of the testing and validation procedure for I based on the optimum configuration of the RF model. The results indicate that the regression coefficients for both processes are equal to 1. This shows that the RF approach is able to produce accurate predicted values of P and I to be used in the DC Link Voltage Regulation circuit controller in the UPQC circuit.

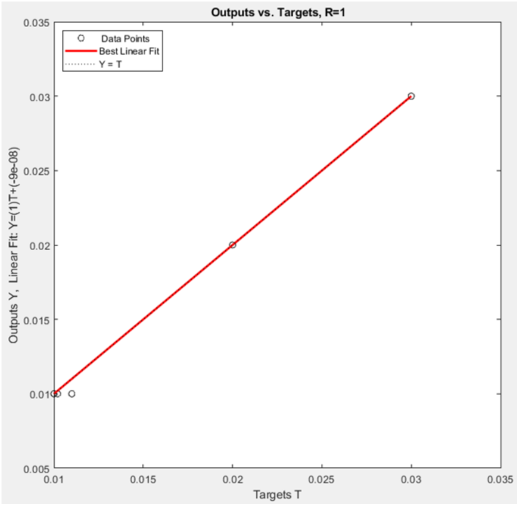
THD has been identified based on the waveform. The table below illustrates the THD value for three different cases of voltage sag reduction which are 40%, 30%, and 10%. THD value of load current, which is the value before mitigation of voltage sag using UPQC, and THD of source current, which is the THD after mitigation of voltage sag using UPQC, are presented below.

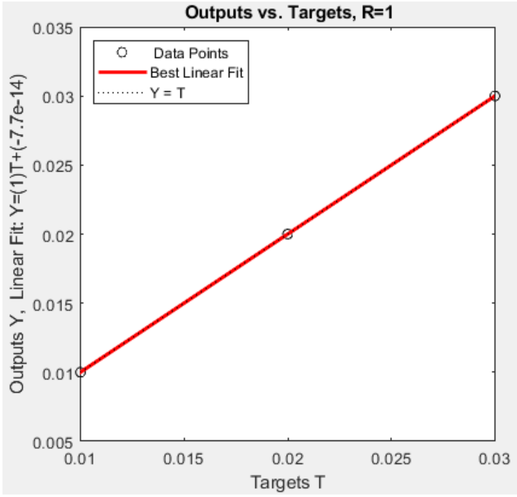
From Table 10, we can see that the THD value for all three reductions of voltage sag is low, falling below 5%, which is the intended THD value. The THD value has lowered from 24.98% to 1.63% in the first case, which represents a 40% reduction in voltage sag. THD dropped from 25% to 1.64% in the second case, which represents a 30% reduction in voltage sag. Finally, for the third case, which is a 10% reduction in voltage sag, the THD value has dropped from 25% to 1.66%. This demonstrates that predicting the P and I values for all three scenarios was successful in lowering the THD value of the source current. Lower THD results in improved power factor, lower maximum currents, and better efficiency.



(a) **(b)**

Fig. 15 *Regression result for P (a) after testing process and (b) after validation process based on the optimal setup of the RF model*





(a) **(b)**

Fig. 16 *Regression result for I (a) after testing process and (b) after validation process based on the optimal setup of the RF model*

Table 10 *The outcome of RF for I by considering various leaves values*

Voltage sag reduction (%)	THD of load current (%)	THD of source current (%)
40	24.98	1.63
30	25.00	1.64
10	25.00	1.66

4. Conclusion

In this research, the RMS voltage source value and RMS error obtained from the UPQC circuit are utilised to predict the right values of P and I in order to lower the THD value of the source current on the UPQC circuit to less than 5%. Several methods, including multiple time lags to improve data, principal component analysis to minimise data dimensionality, and random forest algorithm to predict the accurate values of P and I, have been implemented. In the end, the obtained results are satisfactory for predicting P and I. The selected RMSE values for testing and validating P are 6.0392e-10 and 7.8365e-13. RMSE values for I during testing and validation are 1.4738-10 and 1.1427e-13, respectively. The outcome of regression analysis during the testing and validation process for both cases P and I is 1. Aside from that, the THD value was successfully reduced following the mitigation approach utilising UPQC. For the first case, which is a 40% decrease in voltage sag, the THD of load current before mitigation is 24.98%, while the THD of source current after mitigation is 1.63%. For the second case, which is a 30% reduction in voltage sag, the THD of load current before mitigation is 25.00% while the THD of source current after mitigation is 1.64%. Finally, for the third case, which is a 10% reduction in voltage sag, the THD of the load current is 25.00% before mitigation and the THD of the source current is 1.66% after mitigation. This demonstrates that predicting the P and I values for all three scenarios was successful in lowering the THD value of the source current. Lower THD results in improved power factor, lower maximum currents, and improved efficiency.

Acknowledgement

This research was supported by the Research Management Centre (RMC), Universiti Teknologi MARA (UiTM), Shah Alam, Selangor, Malaysia.

Conflict of Interest

Authors declare that there is no conflict of interests regarding the publication of the paper.

Author Contribution

The authors confirm contribution to the paper as follows: ***study conception and design:*** *Muhammad Hasbi Azmi, Muhammad Murtadha Othman, Kerrollenna Martta Madu Anak Kenneth Inggang;* ***data collection:*** *Muhammad Hasbi Azmi, Muhammad Murtadha Othman, Kerrollenna Martta Madu Anak Kenneth Inggang;* ***analysis and interpretation of results:*** *Muhammad Hasbi Azmi, Muhammad Murtadha Othman, Kerrollenna Martta Madu Anak Kenneth Inggang;* ***draft manuscript preparation:*** *Muhammad Murtadha Othman, Norazlan Hashim, Ismail Musirin, Masoud Ahmadipour. All authors reviewed the results and approved the final version of the manuscript.*

References

[1] Helonde, S., Jha, P., Medhane, C. U., Patil, J. P., Mahto, R. K., & Ramdas, D. (2024). Mathematical Analysis of Power Quality Disturbances in Electrical Distribution Systems: Detection, Classification, and Mitigation Strategies. *Panamerican Mathematical Journal*. https://doi.org/10.52783/pmj.v34.i1.908

[2] Nishad, D. K., Tiwari, A. N., & Khalid, S. (2024). A Comprehensive Analysis of Application Based Linear-Nonlinear loads. In *2024 International Conference on Signal Processing and Advance Research in Computing (SPARC)* (Vol. 1, pp. 1-6). https://doi.org/10.1109/sparc61891.2024.10828719

[3] Pattathurani, L., Dash, S. S., Dwibedi, R. K., Raj, M. D., Kannadasan, R., Savio, M. F. R., Alsharif, M. H., & Kim, J. H. (2022). Harmonics Minimisation in Non-Linear Grid System Using an Intelligent Hysteresis Current Controller Operated from a Solar Powered ZETA Converter. *Sustainability*, *14*(12), 7028. https://doi.org/10.3390/su14127028

[4] Bakar, M. I. A., Magsi, M. H., Uddin, S. M., & Arif, M. A. B. M. (2024). Simulation-Based Study on Harmonic Generation by Non-Linear Loads in 33kV Distribution Systems. In *2024 IEEE 10th International Conference on Smart Instrumentation, Measurement and Applications (ICSIMA)* (pp. 65-70). https://doi.org/10.1109/icsima62563.2024.10675587

[5] Mandapati, N., & K., R. (2022). Power Quality Enhancement using A Unified Power Quality Conditioner (UPQC). *International Journal of Engineering and Advanced Technology*, *11*(5), 168–170. https://doi.org/10.35940/ijeat.e3627.0611522

[6] Rong, S., Xu, M., Chen, X., Ma, J., Li, H., & Li, H. (2022). Voltage Regulation Method of Series Compensator Based on Impedance in Weak Distribution Network. In *2022 4th International Conference on Electrical Engineering and Control Technologies (CEECT)* (pp. 759-764). https://doi.org/10.1109/ceect55960.2022.10030633

[7] Yu, Y., Othman, M. M., & Chu, Y. (2023). Unified Power Quality Conditioner Improved Control Strategy Research Based on Active Disturbance Rejection Controller. In *2023 9th International Conference on Control Science and Systems Engineering (ICCSSE)* (pp. 409-413). https://doi.org/10.1109/iccsse59359.2023.10245415

[8] Luo, W., Shi, T., Vazquez, S., Wang, Z. P., Wu, L., & Franquelo, L. G. (2022). DC-Link Voltage Regulation of Grid-Connected Converters Using Linear Disturbance Observer. *Annual Conference of the IEEE Industrial Electronics Society*, 1–6. https://doi.org/10.1109/IECON49645.2022.9968526

[9] Nagaraju, S., & Bethi, C. (2024). Cga-fopid based upqc for mitigating harmonics and compensate load demand in grid linked hybrid renewable energy sources. *Suranaree Journal of Science and Technology*. https://doi.org/10.55766/sujst-2024-01-e02177

[10] Fidone, G. L., Migliazza, G., Carfagna, E., Immovilli, F., Benatti, D., & Lorenzani, E. (2024). *Power Losses and THD Analysis of Current Source Inverters in Motor Drive Applications: A Comparative Study of Control Schemes*. 1–7. https://doi.org/10.1109/icem60801.2024.10700297

[11] Yadav, S. K., Patel, A., & Mathur, H. D. (2024). PSO-based Online PI Tuning of UPQC-DG in Real-Time. *IEEE Open Journal of Power Electronics*, 1–13. https://doi.org/10.1109/ojpel.2024.3445719

[12] Zhang, X., Li, M., & Ma, J. (2023). Parameter identification and optimization of PI controllers for islanded DC micro-grids. In *2023 7th International Conference on Smart Grid and Smart Cities (ICSGSC)* (pp. 396-401). https://doi.org/10.1109/icsgsc59580.2023.10319161

[13] Huang, H. (2023). PI Control Technology Principal Analysis and Simulation. *Highlights in Science, Engineering and Technology*, *71*, 104-111. https://doi.org/10.54097/hset.v71i.12676

[14] Priya, M., & Prathyusha, C. (2024). *Adaptive Control of Three-Level UPQC via ANFIS for Enhanced Power Quality: A MATLAB / Simulink-Based Evaluation*. *20*(11), 1150–1165. https://doi.org/https://doi.org/10.52783/jes.7393

[15] Afrawira, R., Gumilang, R. F., Amalia, S., & Bandri, S. (2023). Analisa Perbandingan Pengendali PID pada Motor DC Menggunakan Metode Ziegler-Nichols dan Trial and Error. *Ranah Research: Journal of Multidisciplinary Research and Development*, *5*(3), 210-218. https://doi.org/10.38035/rrj.v5i3.758

[16] Billard, L., Douzal-Chouakria, A., & Samadi, S. Y. (2023). Exploring Dynamic Structures in Matrix-Valued Time Series via Principal Component Analysis. *Axioms*, *12*(6), 570. https://doi.org/10.3390/axioms12060570

[17] Mohammed, M. A., Akawee, M. M., Saleh, Z. H., Hasan, R. A., Ali, A., & Sutikno, T. (2023). The effectiveness of big data classification control based on principal component analysis. *Bulletin of Electrical Engineering and Informatics*, *12*(1), 427–434. https://doi.org/10.11591/eei.v12i1.4405

[18] Rohman, F. N., Farikhin, F., & Surarso, B. (2025). Hyperparameter Tuning of Random Forest Algorithm for Diabetes Classification. *International Journal of Current Science Research and Review*, *08*(01). https://doi.org/10.47191/ijcsrr/v8-i1-31

[19] Redan, R. N., Othman, M. M., Hasan, K., & Ahmadipour, M. (2023). Random forest (RF) with daubechies wavelet and multiple time lags (MTL) for solar irradiance forecasting. In *2023 IEEE 3rd International Conference in Power Engineering Applications (ICPEA)* (pp. 374-378). https://doi.org/10.1109/ICPEA56918.2023.10093181

[20] Latif, S. H. A., Alwan, A. S., & Mohamed, A. M. (2022). Principal component analysis as tool for data reduction with an application. *EUREKA: Physics and Engineering*, (5), 184-198. https://doi.org/10.21303/2461-4262.2022.002577

[21] Efendi, A., Mursityo, Y. T., Hidajati, N. W., Andajani, N., Zuraidah, Z., & Handoyo, S. (2024). Multiple Time Series Modeling of Autoregressive Distributed Lags with Forward Variable Selection for Prediction. *Wseas Transactions On Business And Economics*. https://doi.org/10.37394/23207.2024.21.84

[22] Reddy, P. D., & Parvathy, L. R. (2022). Prediction analysis using random forest algorithms to forecast the air pollution level in a particular location. In *2022 3rd International Conference on Smart Electronics and Communication (ICOSEC)* (pp. 1585-1589). https://doi.org/10.1109/ICOSEC54921.2022.9952138

[23] Pereira, R. G., Junior, W., & Guerra, A. A. (2024). The use of Supervised Learning to perform pairwise classification for Record Linkage over real world data. *International Journal for Population Data Science*, *9*(5). https://doi.org/10.23889/ijpds.v9i5.2902