INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

Analysis of 11kV/ 430 V 500 kVA Transformer Dissolved Gas using Pre-processing Techniques through Duval's Triangle

G V S S N Srirama Sarma[1,2]*, B Ravindranath Reddy[3], Pradeep M Nirgude[4], P Vasudeva Naidu[5]

1 Research Scholar, JNTUH, Hyderabad, India
2 Assistant Professor, Department of Electrical and Electronics Engineering, Matrusri Engineering College, Saidabad, Hyderabad, India
3 Deputy Executive Engineer, JNTUH Univerity, Kukatpally, Hyderabad, India
4 Additional Director, UHV Research Laboratory, CPRI, Hyderabad, India
5 Associate Professor, Department of Electrical and Electronics Engineering, Matrusri Engineering College, Saidabad, Hyderabad, India

OPEN ACCESS
Received: 12.10.2021
Accepted: 06.11.2021
Published: 25.11.2021

Citation: Srirama Sarma GVSSN, Reddy BR, Nirgude PM, Vasudeva Naidu P (2021) Analysis of 11kV/ 430 V 500 kVA Transformer Dissolved Gas using Pre-processing Techniques through Duval's Triangle. Indian Journal of Science and Technology 14(39): 2971-2981. https://doi.org/10.17485/IJST/v14i39.1923

* **Corresponding author**.

music.sarma2016@gmail.com

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Abstract

Objectives: Analysis of dissolved gas in transformer oil is very important because the fluctuations in the oil will change the process of content of dissolved gases which will show effect on many factors such as oil temperature and external environment. Hence, it is important to propose a preprocessing technique through Duval's triangle which is important step in the machine learning process for predictive maintenance of transformer. **Methods:** The dissolved gas content for 500 kVA transformer oil based on the training and testing data for fault free condition taken for the period of 5 years (2013-17) and faulty condition taken for 3 years (2018-2021), has been evaluated using Duval's triangle method to express gas concentrations of CH_4, C_2H_2 and C_2H_4 in ppm as percentages of their total. High pass filter, scaling and windowing techniques are used for preprocessing the giant DGA data. **Findings:** The results show that unusual and undesirable intensity in the CH_4, C_2H_2 and C_2H_4 gasses is eliminated by using proposed high pass filter technique. More visible output signal by the way of reducing its dimensionality is obtained by applying proposed scaling and windowing techniques for the filtered DGA data. **Novelty:** Most of the researchers have used limited samples of DGA data for condition monitoring due to which diagnosis of transformer faults is not accurate. To accurately diagnose transformer faults, large set of training and testing data is required. In this paper, the pre-processing of giant DGA data consisting of 960 samples of fault free training with data size of 480000 x 55 and 500 samples of fault free testing with data size of 250000 x 55, faulty training and faulty testing with data size of 960000 x 55 is presented. The preprocessing method adopted will be used in real time application to provide suitable data to train the advanced deep neural network like LSTM through MATLAB interfacing with graphical processor unit. This is required for condition monitoring and accurately predicting the faults in above mentioned transformer.

Keywords: Dissolved Gas Analysis; Duval's Triangle method; Predictive maintenance; Data Ensemble; Pre-processing; Scaling; Windowing

1 Introduction

In the power system, oil-immersed transformer is one of the very important components. When a transformer is subjected to severe overloads, it can cause severe damage to its components. This issue increases the risk of failure. This is because the level of overload can severely affect the useful life of the transformer. Incipient failures in the transformer can be detected by Dissolved Analysis (DGA) technique. Through the use of the DGA method, it is possible to continuously evaluate the condition of a transformer and identify potential issues[1–4]. The distribution of flammable and non-flammable gases is also determined according to the components and failure involved. Aside from being classified according to their type of failure, these gases also have various other characteristics that can affect their data quality. These gases can be classified according to their origin and the material used in their production. In most cases, the data gaps can be filled with the help of pre-processing techniques. The main goal of fault gas analysis is to accurately diagnose the defect that caused the identified gases. Based on a comparison of the stability and precision in anticipating failures, the Duval Triangle approach provides a useful insight DGA applications in transformers[2]. Researchers and field engineers can utilize the findings to make decisions about transformer faults. The majority of DGA diagnosis procedures rely on specialized studies, which may be insensitive to minor flaws that accumulate over time. Unscheduled actions such as oil-tank welding and the electric charge carried by the oil flow might lead to incorrect judgement even in a normal environment.

Over the last few decades, two different types of processes for obtaining dissolved gas analysis data have been presented for predictive maintenance[5] of transformers as an alternative to breakdown remedial maintenance. In comparison with the mathematical approaches, machine learning-based approaches[6] performed improved diagnostic precision due to its outstanding capabilities of classification and predication of machine learning. Therefore, these methods have become prospective means of fault diagnosis in transformers. The majority of researchers have only employed small samples of DGA data for condition monitoring, resulting in an inaccurate identification of transformer problems. A vast set of training and testing data is necessary to accurately diagnose transformer defects.

The primary contributions of this paper are as follows:

- The pre-processing of giant DGA data consisting of 960 samples of fault-free training and 500 samples of fault-free testing, faulty training and faulty testing is done to diagnose transformer faults with precision.
- The proposed high pass filter approach eliminates unexpected and unwanted intensity in the CH_4, C_2H_2, and C_2H_4 gases.
- The proposed scaling and windowing approaches for the filtered giant DGA data result in a more apparent output signal by lowering its dimensionality.

The remainder of this paper is structured as follows: Section II presents Duval's triangle method. Section III explains frame work for preprocessing, section IV details the case study on 500 kVA transformer, and finally, section V discusses the conclusion of this work.

2 Duval's Triangle Method (DTM)

In nearly half of the situations, the Rogers ratio, Dornenburg ratio, IEC 60599, and key gas approaches yield "No result". So they are unable to recognize fault aspects. For dissolved gas analysis of transformer Michael Duval developed a new method called Duval Triangle method in 1974 [7], [8], [9]. Duval's triangle method, does not have any "No Result" in their outputs. In this method concentration of acetylene (C_2H_2), ethylene (C_2H_4) and methane (CH_4) in ppm are expressed as percentages of the total (C_2H_2+C_2H_4+CH_2). This technique recognizes data on dissolved gases using a triangle of relative percentages of CH_4, C_2H_2 and C_2H_4 where each vertex indicates that one of these compounds makes up 100% of the gases analyzed [10]. The fault type which thus produced that blend of gas concentrations is indicated by the fault zone where the spots are found. Because of the relative uncertainty of gas-in-oil concentration measurements at low concentrations, DGA diagnostic techniques such as the Duval Triangle should not be employed unless gas concentrations are much over the detection limit.

Equivalent percentages of each gas are calculated as follows:

$$\%CH_4 = 100x/(x+y+z) \tag{1}$$

$$\%C_2H_4 = 100y/(x+y+z) \tag{2}$$

$$\%C_2H_2 = 100z\,/(x+y+z) \tag{3}$$

Where x = CH_4 PPM, y = C_2H_4 PPM & z = C_2H_2 PPM

In this method, point which is related to percentage of the three specified gases should be located in the triangle based on region of fault to be determined.

The faults which are detected by Duval's triangle are shown in Figure 1 and expressed as follows.

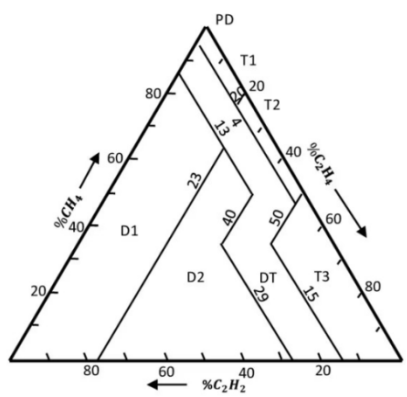

Fig 1. Duval triangle

Partial Discharge (PD) or Corona Discharge (CD): In the gas phase of voids or gas bubbles, PD can arise. DGA can easily identify it. It typically creates enormous amounts of hydrogen because it is created over lengthy periods of time and in vast quantities of paper insulation.

Thermal faults, T <300°C (T1): T1 is shown by the browning of the paper.

Thermal faults, 3000 <T<700 °C (T2): When paper carbonizes, T2 is produced.

Thermal fault, (T3), Hot Spot, T> 700 °C (T3): Oil carbonization, metal colouring, or fusion are all signs of T3.

Discharges of Low energy (D1), Low Energy Arcing and Tracking: Because the gas generation is substantial enough, D1 such as tracking, tiny arcs, and continuous sparking discharges are generally clearly observable by DGA.

Discharges of High energy (D2), High Energy Arcing: Extensive carbonization, metal fusion, and probable equipment tripping are all signs of D2.

Thermal & Electrical fault (DT) or Hot Spot, T>400°C

Table 1 represents the percentage of gases in different zones.

Insufficient test data, lack of comparative evaluation with state-of-the-art, complexity of hybrid techniques and majorly lack of freely available data are the main limitations for developing several methods.

Table 1. Percentage of gases in different zones

Type of Fault Zone	Percentage of gases			
PD	98% CH_4	100% CH_4		
D1	23% C_2H_4	13% C_2H_2	100% C_2H_2	
D2	23% C_2H_4	13% C_2H_2	38% C_2H_4	29% C_2H_2
T1	4% C_2H_2	10% C_2H_4		
T2	4% C_2H_2	10% C_2H_4		
T3	15% C_2H_2	50% C_2H_4	100% C_2H_4	

3 Methodology

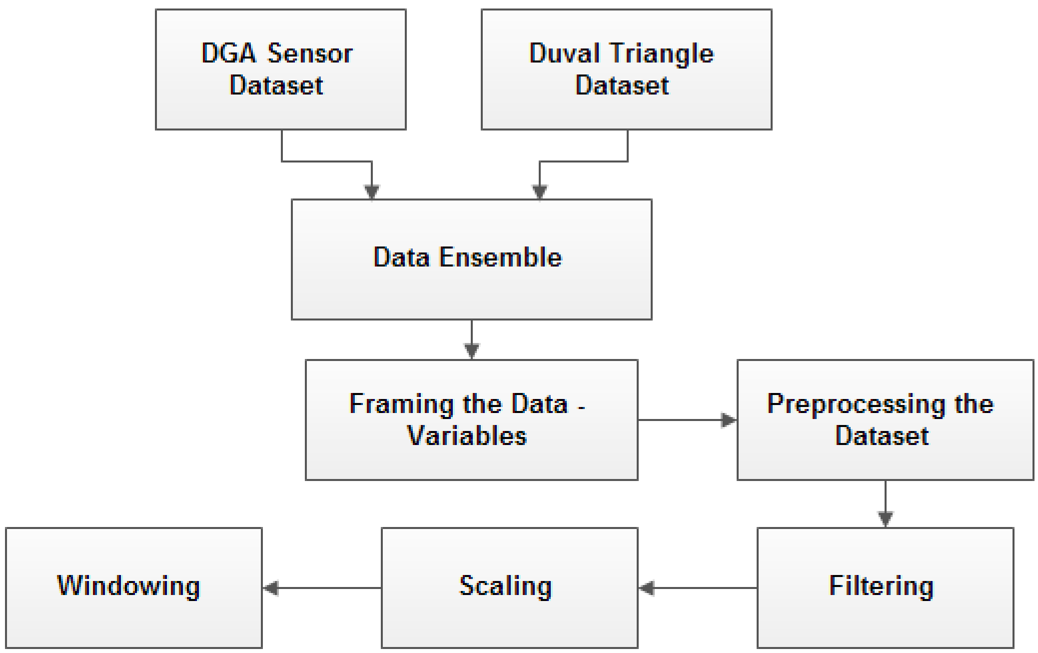

Fig 2. Workflow diagram

From the workflow diagram shown in the Figure 2, the first step in predictive maintenance of transformer is to collect the DGA sensor data under healthy and faulty conditions. The proposed work is based on the time series of three characteristic gases of CH_4, C_2H_2 and C_2H_4. As with most real data with more samples for training and testing, there are a few challenges involved in dealing with this dataset. First, data ensemble is done followed by framing the data variables. The dimensionality of the DGA data in this study is very high. Second, the dimensionality reduction is needed through windowing technique after filtering and scaling the data which is part of preprocessing [5].

An ensemble is a collection of DGA data sets, formed by measuring the data from DGA sensor connected to the transformer under changing conditions. The family of data sets is an ensemble and each row in the ensemble is a member of the ensemble.

Scaling is the method used to sample the data accordingly. In maximum cases, operating with the entire data set can end up too luxurious thinking about the reminiscence and time constraints. Using a sampling set of rules can assist us lessen the scale of the data set to a degree wherein we are able to use a better, however extra luxurious, device gaining knowledge of set of rules. Scaling is the technique that is used horizontally. There are two types of scaling of data that to be considered -normalization and standardization [6]. Normalization is a rescaling of the data from the original range so that all values are within the range of 0 and 1. Precise estimation of the minimum and maximum observable values from available DGA data is required for normalization.

A value is normalized as follows:

$$b = (a - \min)/(\max - \min) \tag{4}$$

Where the minimum and maximum values pertain to the value 'a' being normalized. Fit the scale using available training DGA data, apply the scale to training data and apply the scale to data going forward are the steps involved in scaling. Standardizing a dataset involves rescaling the distribution of values so that the mean of observed values is 0 and the standard deviation is 1. Like normalization, standardization can be useful, and even required in some machine learning algorithms when data has input values with differing scales. The estimation of mean and standard deviation of observable values from the trained DGA data can be obtained as follows.

A value is standardized as follows

$$b = (a - \text{ mean })/ \text{ standard deviation} \tag{5}$$

Where the mean is calculated as

$$\text{Mean } = \operatorname{sum}(a)/\operatorname{count}(a) \tag{6}$$

And the standard deviation is calculated as

$$\text{standard deviation } = \operatorname{sqrt}\left(\operatorname{sum}\left((a - mean)^\wedge 2\right)/\operatorname{count}(a)\right) \tag{7}$$

The mean and standard deviation estimates of a dataset can be more robust to new data than the minimum and maximum. Data scaling is a recommended pre-processing step when working with deep learning neural networks.

Most actual global datasets have a big wide variety of functions. Windowing is the method used to dimensionalise the data. The main advantage of controlling the leakage is an increase in the dynamic range of the analysis, as leakage may swamp signal components of close frequencies and much smaller magnitudes. The use of the Hanning window is nearly forever suggested for signals with narrowly spaced line spectra. The Hamming window may occasionally be more useful due to the much smaller secondary lobe. The Hanning window is useful if the frequency spacing between components is greater than the window's bandwidth. Unless a longer duration signal can be analyzed, a rectangular window would then perform better. The Hanning window is usually a good choice.

The Hann function of length L, used to perform Hann smoothing,[11] is named after the Austrian meteorologist Julius von Hann. It is a window function given by the following expression[12]

$$w_0(x) = \left\{ \begin{array}{ll} \frac{1}{2}\left(1+\cos\left(\frac{2\pi x}{L}\right)\right) = \cos^2\left(\frac{\pi x}{L}\right), & |x| \leq L/2 \\ 0, & |x| \geq L/2 \end{array} \right\} \tag{8}$$

For digital signal processing, the function can be sampled (with spacing L/N) symmetrically as

$$w[n] = w_0\left(\frac{L}{N}(n - N/2)\right) = \frac{1}{2}\left[1 - \cos\left(\frac{2\pi n}{N}\right)\right]\Bigg\}, 0 \leq n \leq N \tag{9}$$

Which is a sequence of N+1 samples, and N can be even or odd.

The following equation generates the coefficients of a Hann window:

$$w(n) = 0.5(1 - \cos(2\pi n/N)), 0 \leq n \leq N \tag{10}$$

the window length L = N + 1.

4 Results & Discussion-Case Study

The data on dissolved gas content of 500 kVA transformer oil, shown in Figure 3 is used in this investigation. The Duval triangle approach has been used to examine the training and testing data sets. The fault free data is taken for the period of 5 years (2013-17) and faulty data is taken for 3 years (2018-2021).

The specifications of the test transformer are shown in the Table 2 given below.

Fig 3. 500 kVA Distribution transformer

Table 2. Specifications of 500 kVA test transformer

KVA		500
Volts at No Load	HV	11000
	LV	430
Amperes	HV	26.24
	LV	671.35
Phases	HV	3 Delta
	LV	3 Star
Type of Cooling		ONAN
Frequency		50HZ
Impedance Volts %		4.29
Vector Group Ref:		Dyn 11

Table 3. Fault free DGA data

Particulars	No. of Samples	Size of Data
Fault free training data	960	480000 x 55
Fault free testing data	500	250000 x 55

Table 4. Faulty DGA data

Particulars	No. of Fault codes	No. of Samples	Size of Data
Faulty training data	20	500	960000 x 55
Faulty testing data	20	500	960000 x 55

4.1 DGA Dataset

The DGA dataset taken by using Hydran DGA sensor on500 kVA, 11000/430 V test transformer from 2013-2021 is divided into fault free testing set, fault free training set, faulty testing set and faulty training set. The details of the data set are mentioned in the Tables 3 and 4 respectively.

Each of the sample is run for 500 times simulation for fault free data.

4.2 Data Input from Duval Triangle

The gas concentrations of three gases from CSV file of 960 samples of huge data are given as input to Duval triangle code in MATALAB. The percentage of three gases along with gas concentrations in PPM is taken input signal for data ensemble. The ”fault type” for different percentage of gases is observed as graphical output.

Figures 4 and 5 depicts the graphical view of the location of relative percentages of acetylene (C_2H_2), ethylene (C_2H_4) and methane (CH_4) for discharges of low energy (D1), discharges of high energy (D2), discharges and thermal faults (DT), thermal faults, T<3000 C (T1), thermal faults 3000 <T<700 °C (T2) and thermal faults T>700 °C (T3) respectively.

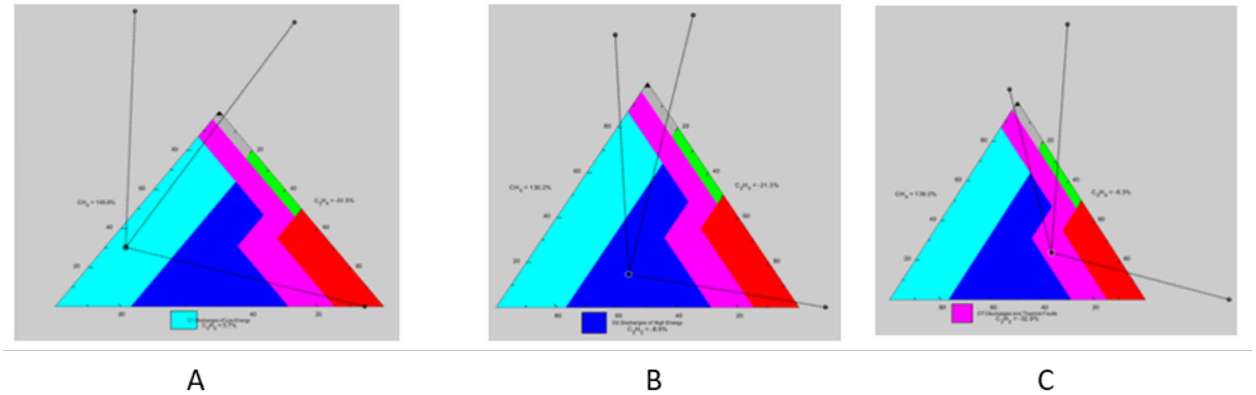

Fig 4. A) Location for D1, B) Location for D2, C) Location for DT

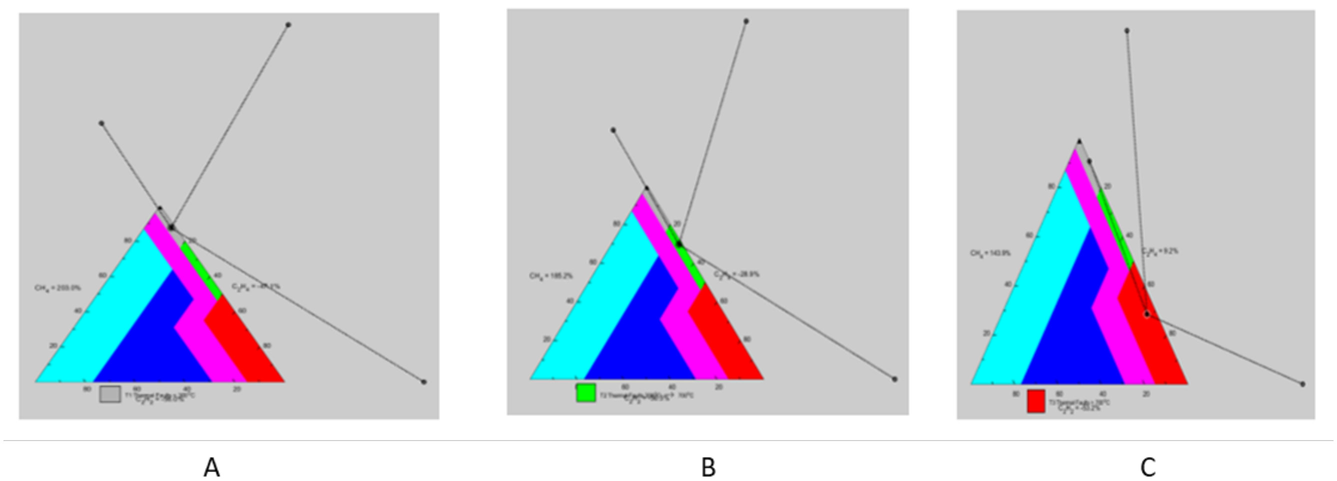

Fig 5. A) Location for T1, B) Location for T2, C) Location for T3

The preprocessing of giant DGA data is done by the MATLAB simulation platform (R2021a, MathWorks) on Dell PC with 16 GB RAM and 3.5 GHz clock, running Windows 7 enterprise operating system (64-bit). Graphical processor unit, Jetson nano hardware is integrated with MATLAB for preprocessing the large DGA data.

4.3 Creating Data Ensemble

This data file is initially loaded in to the MATLAB software and converted into ensemble[13] timetable for understanding how each gas changes with respect to the time. Data ensemble is used to collect the DGA data, converting CSV files to corresponding tables. It is done for fault free testing, fault free training, faulty testing and faulty training data. It is part of data acquisition. The data ensemble data for three gases C_2H_2, C_2H_4 and CH_4 in are shown the Figures 6 and 7 and Figure 8 respectively.

4.4 Data Filtering

High-Pass filter is used to remove inconsistent values. When gas sensor is giving data, there are spike indications at particular row data i.e. 3, 9 and 15 in every set of data. Even after graph is obtained, the curves which are touching reference value i.e. there is no data, the values corresponding to particular row are removed which are shown in the Figure 9. Next step is to fix the sampling the frequency and setting the frequency resolution. The total DGA set data is converted into frequency domain. The output from filter is converted into time domain using scaling.

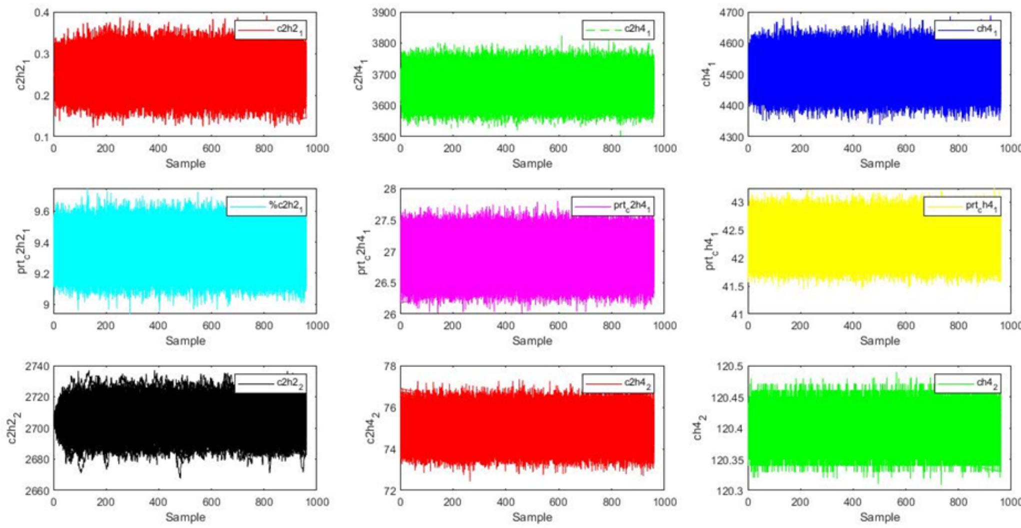

Fig 6. Fault free training C_2H_2 Data

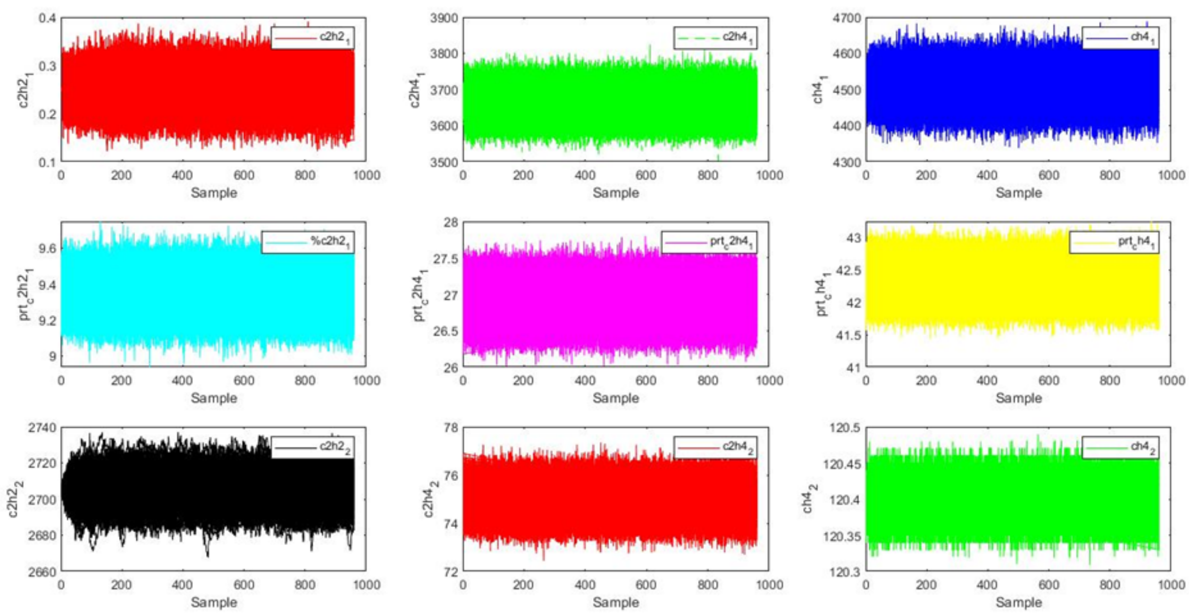

Fig 7. Fault free training C_2H_4 Data

4.5 Data Scaling

In the scaling technique it can be seen that whenever 960 samples are coming there must not be any missing intensity of any gas considered i.e. CH_4, C_2H_2 and C_2H_4. In the scaling process sampling frequency (fs) which is fixed is used to find time resolution, t=1/fs. Based on time resolution, time interval is created depends upon DGA sampling time. Now this time signal data is applied to the total DGA samples data. It is further converted into time domain data for windowing process. The scaling of the data is shown in the Figure 10.

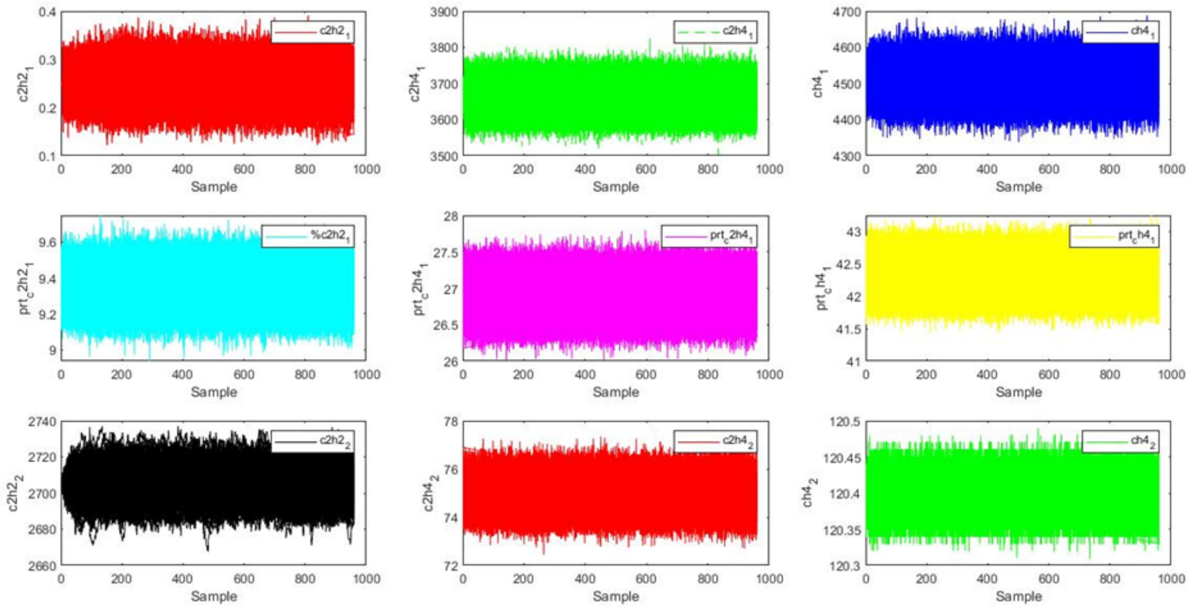

Fig 8. Fault free training CH_4 Data

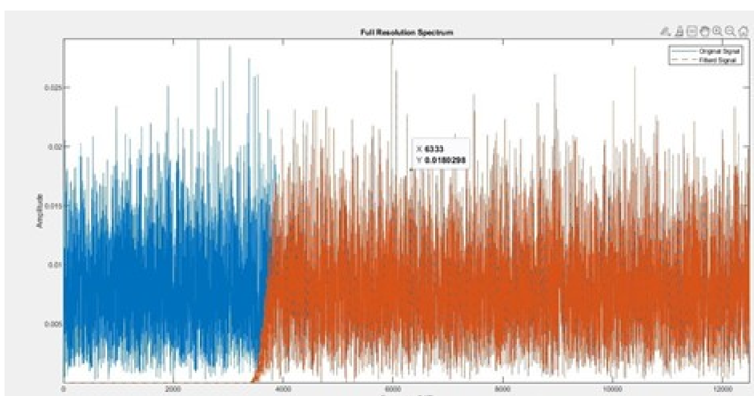

Fig 9. Data filtering using high pass filter

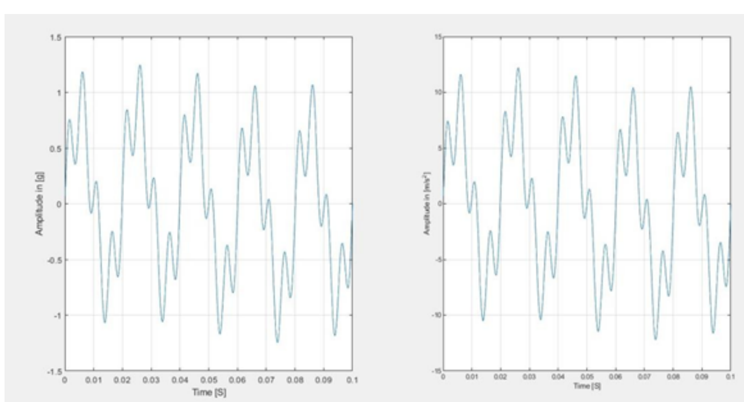

Fig 10. Scaling of DGAdata

4.6 Windowing

Hanning window technique is used in order to estimate the intensity of the gas within the time range of 960 samples vertically. Precaution is taken in windowing and scaling regarding percentage of gases. Envelope is the window frame which indicates the boundary limits of the signal. All the 960 samples in time domain have been converted into frequency domain by using FFT. N is the length of the number of samples in DGA data. Based on the length and time base, the total DGA data is converted into time domain data by using cosine signal followed by hann window MATLAB command. The time domain signal is converted into frequency domain by using FFT analysis as shown in the Figure 11. The total DGA set data is converted into frequency domain and dimensionality is reduced which is required for the large data set. This total process is run in Graphical Processor Unit (GPU) in parallel with MATLAB set mat file.

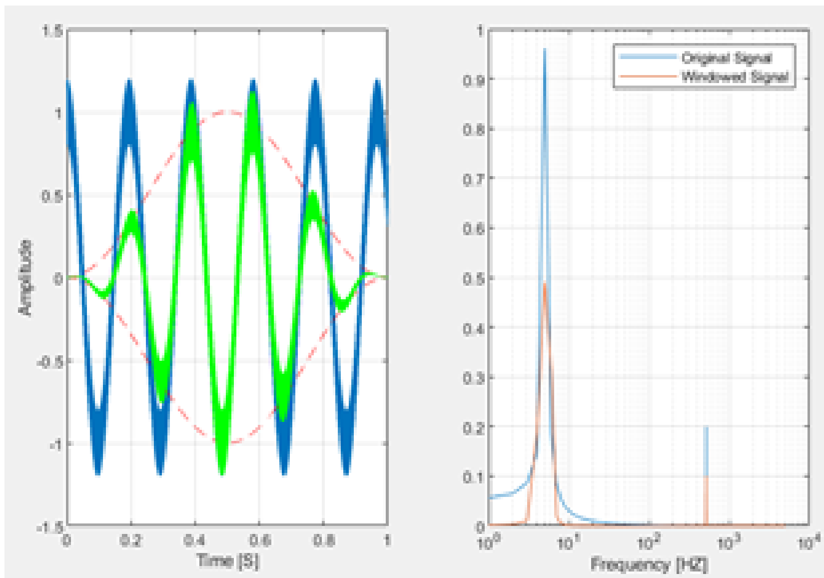

Fig 11. Data Windowing

4.7 Case study analysis

Preprocessing is essential to improve database consistency and completeness. For condition monitoring, most of the researchers have used limited samples of DGA data[4,5,10,14,15] that leads to lack of proper training of advanced data driven techniques, resulting in erroneous diagnosis of transformer faults. In this case study, the unique DGA data set consisting of 960 samples of fault free training with data size of 480000 x 55 and 500 samples of fault free testing with data size of 250000 x 55, faulty training and faulty testing with data size of 960000 x 55 is used for preprocessing. This data is collected for a long period of 8 years. Such a huge data is sufficient for better training of the advanced deep neural networks like LSTM for more accuracy of fault diagnosis evaluation. The results of the case study reveals that the data ensemble for three significant gases C_2H_2, C_2H_4, and CH_4 is made for the preprocessing stage. The suggested high pass filter technology eliminated anomalous and unwanted intensity in the CH_4, C_2H_2, and C_2H_4 gases in the large DGA data consisting of 960 samples of fault free training and 500 samples of fault free testing, faulty training and faulty testing. The proposed scaling and windowing strategies for the filtered DGA data result in a more apparent output signal by lowering its dimensionality so that the data format is suitable for GPU learning.

5 Conclusion

The dissolved gases data for acetylene (C_2H_2), ethylene (C_2H_4) and methane (CH_4) and their percentages through Duval's triangle method makes it best for detailed fault diagnosis as preprocessing in transformer. Hydran sensor is used to extract gases from the oil under fault free and faulty conditions. The pre-processing of huge DGA data, which includes 960 samples of fault-free training with a data size of 480000 x 55 and 500 samples of fault-free testing with a data size of 250000 x 55, as well as faulty training and faulty testing with a data size of 960000 x 55, is discussed in this study. The anomalous and undesired gas intensity is removed using a high pass filter approach during preprocessing. To make output signal very much visible to the Long Short term Memory (LSTM) network, Scaling and Windowing are used to reduce its dimensionality. To train the deep neural network technology, input is made in the form of clean and visible range.

Acknowledgment

Mr. GVSSN Srirama Sarma thanks CITD, Balangar, Hyderabad for providing the huge DGA dataset comprising fault free testing set, fault free training set, faulty testing set and faulty training set of 500 kVA, 11000/430 V test transformer.

References

1) Aciu AM, Nicola CI, Nicola M, Nițu MC. Complementary Analysis for DGA Based on Duval Methods and Furan Compounds Using Artificial Neural Networks. *Energies*. 2021;14(3):588. Available from: https://dx.doi.org/10.3390/en14030588.
2) Mharakurwa ET, Nyakoe GN, Akumu AO. Power Transformer Fault Severity Estimation Based on Dissolved Gas Analysis and Energy of Fault Formation Technique. *Journal of Electrical and Computer Engineering*. 2019;2019:1–10. Available from: https://dx.doi.org/10.1155/2019/9674054.
3) Abu-Siada A. Improved Consistent Interpretation Approach of Fault Type within Power Transformers Using Dissolved Gas Analysis and Gene Expression Programming. *Energies*. 2019;12(4):730. Available from: https://dx.doi.org/10.3390/en12040730.
4) Alqudsi A, El-Hag A. Application of Machine Learning in Transformer Health Index Prediction. *Energies*. 2019;12(14):2694. Available from: https://dx.doi.org/10.3390/en12142694.
5) Kabir F, Foggo B, Yu N. Data Driven Predictive Maintenance of Distribution Transformers. *2018 China International Conference on Electricity Distribution (CICED)*. 2018. doi:10.1109/CICED.2018.8592417.
6) Witten IH, Frank E. Data Mining: Practical Machine Learning Tools and Techniques. 4th ed. Morgan Kaufmann. .
7) Rogers R. IEEE and IEC Codes to Interpret Incipient Faults in Transformers, Using Gas in Oil Analysis. *IEEE Transactions on Electrical Insulation*. 1978;p. 349–354.
8) Data Analytics cases for Asset Awareness. *Electric Power Research Institute* . 2019.
9) Guidelines for operation and maintenance of distribution Transformers. 2018.
10) Poonnoy N, Suwanasri C, Suwanasri T. Fuzzy Logic Approach to Dissolved Gas Analysis for Power Transformer Failure Index and Fault Identification. *Energies*. 2020;14(1):36. Available from: https://dx.doi.org/10.3390/en14010036.
11) Pielawski N, Wählby C. Introducing Hann windows for reducing edge-effects in patch-based image segmentation. *PLOS ONE*. 2020;15(3):e0229839. Available from: https://dx.doi.org/10.1371/journal.pone.0229839.
12) Kumar R, Markam K. Performance Analysis of Kaiser-Hanning Window For Digital Filter Design. *Journal of Emerging Technologies and Innovative Research*. 2018;5(10). Available from: https://www.jetir.org/papers/JETIR1810880.pdf.
13) Ghosh S, Dutta S. Ensemble Machine Learning Methods for better Dynamic Assessment of Transformer Status. *Journal of The Institution of Engineers (India): Series B*. 2021;102(5):1113–1122. Available from: https://dx.doi.org/10.1007/s40031-021-00599-1.
14) Zhang X, Wang S, Jiang Y, Wu F, Sun C. Prediction of dissolved gas in power transformer oil based on LSTM-GA. *IOP Conference Series: Earth and Environmental Science*. 2021;675(1):012099. Available from: https://dx.doi.org/10.1088/1755-1315/675/1/012099.
15) Zeng B, Guo J, Zhang F, Zhu W, Xiao Z, Huang S, et al. Prediction Model for Dissolved Gas Concentration in Transformer Oil Based on Modified Grey Wolf Optimizer and LSSVM with Grey Relational Analysis and Empirical Mode Decomposition. *Energies*. 2020;13(2):422. Available from: https://dx.doi.org/10.3390/en13020422.