

A long short-term memory based prediction model for transformer fault diagnosis using dissolved gas analysis with digital twin technology

Gadepalli Srirama Sarma¹, Bumanapalli Ravindranath Reddy², Pradeep Nirgude³,
Pudi Vasudeva Naidu⁴

¹Department of Electrical and Electronics Engineering, Matrusri Engineering College, Hyderabad, India

²Deputy Executive Engineer, Jawaharlal Nehru Technological University Hyderabad (JNTUH University), Hyderabad, India

³Ultra High Voltage Research Laboratory (UHVRL), Central Power Research Institute (CPRI), Hyderabad, India

⁴Department of Electrical and Electronics Engineering, Matrusri Engineering College, Hyderabad, India

Article Info

Article history:

Received Nov 16, 2021

Revised May 2, 2022

Accepted May 23, 2022

Keywords:

Digital twin

Dissolved gas analysis

Health indicators

LSTM

Predictive maintenance

Pre-processing

ABSTRACT

The most significant tool for defect diagnostics in transformers is dissolved gas analysis (DGA). The time series prediction of dissolved gas levels in oil, when combined with dissolved gas analysis, provides a foundation for transformer fault diagnosis and an early warning. A long short-term memory (LSTM) based prediction model is developed in this paper to train the digital twin for identifying the essential fault in the transformer via DGA. The model is fed with three different gas concentrations as input. This study achieves the performance evaluation in terms of validation accuracy. The suggested model exhibits significant validation accuracy of 99.83%, as indicated by the analyses, thus the early prediction of transformer maintenance is aided. It can be validated that the LSTM model for fault identification and analysis using dissolved gas in the transformer has a lot of research potential.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Gadepalli Srirama Sarma

Research Scholar, JNTUH and Assistant Professor, Department of Electrical and Electronics Engineering

Matrusri Engineering College

Saidabad, Hyderabad, India

Email: music.sarma2016@gmail.com

1. INTRODUCTION

Transformers are the heart of electric power systems, and their operational state decides whether or not the power network is well-regulated. Electrical, mechanical, and thermal stresses cause some gases created during an operation to dissolve in insulating oil [1]-[3]. Dissolved amounts of these gases are being examined as a possible approach to diagnose and anticipate the transformer's performance both inside and out [4]-[7]. The outcome of a dissolved gas analysis (DGA) offers sufficient information to diagnose the state of the transformer's operation. Two different sorts of procedures for getting dissolved gas analysis data have been proposed during the last several decades for predictive maintenance of transformers which is an alternative to breakdown corrective maintenance. Predictive maintenance increases the operational availability of transformers, preclude downtime due to unscheduled maintenance, minimize the costs and maximize safety. Data driven methodologies are superior to model based predictive maintenance elucidations as they attempt to learn predictive models from the data automatically that makes them suitable for a wide range of such problems.

Machine learning has begun to play an important role in predicting transformer failures, and as a result, contributing to predictive maintenance. Artificial neural networks (ANN), support vector machines (SVM), relevance vector machines (RVN), and fuzzy theory [8]-[11] are examples of common machine learning approaches that have been implemented for DGA. All these methods have several specific flaws, such as the need to choose a large number of parameters for accuracy, overfitting, and training duration. Deep learning approaches such as long short-term memory (LSTM) that can extract features automatically have been used, with promising results. To be effective, deep learning techniques require a large amount of data. When compared to other modern methodologies, many deep learning algorithms offered for predictive maintenance problems [12] such as estimating the remaining useful life of transformers showed encouraging results. For greater accuracy LSTM require a significant amount of labelled data to be adequately trained. The following are the paper's main contributions: i) To diagnose transformer defects with precision, huge DGA data consisting of 960 samples of fault-free training and 500 samples of fault-free testing, faulty training, and faulty testing is pre-processed with the proposed high pass filter, scaling and windowing approaches; ii) LSTM network based prediction model is developed to train the digital twin for precisely diagnosing the essential fault in the transformer through the DGA with highest validation accuracy. The rest of this paper is structured as follows: Section 2 describes LSTM network, section 3 explains methodology with the framework for the LSTM-based digital twin training approach, section 4 presents results and discussion, and finally, section 5 discusses the conclusion.

2. LSTM NETWORK

Hochreiter and Schmidhuber proposed LSTM as a recurrent neural network (RNN) based technique [13]. When it comes to processing time series, RNN outperforms other neural networks. RNN training time exists, and LSTM solves RNN based on RNN with only a short-term memory. LSTM adds a memory unit to determine whether the information is helpful and has a sophisticated dynamic structure when compared to RNN. The model can handle the challenge of long-term sequence prediction since it has a long-term memory function. In addition, during long sequence training, LSTM may tackle gradient disappearance and gradient explosion [14], [15]. The classification of sequence data is well suited to long short-term memory (LSTM) networks. For time-series data, LSTM networks are useful because, in order to categorize new signals, they recall the uniqueness of earlier signals. An LSTM network allows users to feed sequence data into it and make predictions based on the discrete time steps in the data.

A gating cell is added to the network architecture in the LSTM model, giving it a “long time memory function” that makes it suited for long-term nonlinear series prognosis issues. Memory cells with a gating mechanism replaces unseen layer neurons in LSTM, as opposed to regular RNN. The forget gate determines which parts of the current state move through to the next, the input gate modifies the current input before it is added to the new state, and the output gate modifies the values that pass from the current state to the output. Figure 1 depicts the basic structure of a memory cell [16], [17].

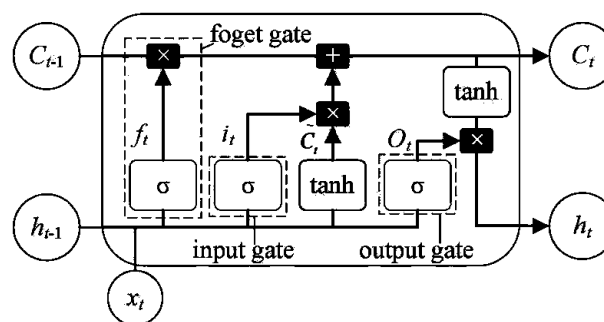


Figure 1. The basic structure of LSTM memory cell

The memory cell is the most important part of the LSTM network. The sequence input x_t at time t , the memory cell state C_{t-1} at time $t - 1$, and the hidden layer cell state h_{t-1} at time $t - 1$ make up the cell input. The memory cell state C_t at t and the hidden layer cell state h_t at t are included in the output, with C_t and h_t containing the model's long-term and short-term memory information, respectively. The reading and modification of the memory cell, as well as the information flow between memory cells, are accomplished by manipulating the above three gates. The formulas are provided by (1) to (3) [18].

$$f_t = \sigma(W_{fx}x_t + W_{fh}h_{t-1} + b_f) \quad (1)$$

$$i_t = \sigma(W_{ix}x_t + W_{ih}h_{t-1} + b_i) \quad (2)$$

$$O_t = \sigma(W_{ox}x_t + W_{oh}h_{t-1} + b_o) \quad (3)$$

Where f_t, i_t and O_t are the state computation results of the forget, input, and output gates respectively. $W_{fx}, W_{fh}, W_{ix}, W_{ih}, W_{ox}, W_{oh}$ and b_f, b_i, b_o respectively, are the weight matrix and offset term of the corresponding gate. Sigmoid activation function is denoted by σ .

Memory cell state C_t and hidden layer state h_t are the output outcomes of memory cell at time t. The following are the formulas:

$$\tilde{C}_t = \tanh(W_c h_{t-1} + W_f x_t + b_c) \quad (4)$$

$$C_t = f_t \circ C_{t-1} + i_t \circ \tilde{C}_t \quad (5)$$

$$h_t = O_t \circ \tanh(C_t) \quad (6)$$

Where \tilde{C}_t denotes the memory cell's state input at time t, and \tanh denotes hyperbolic tangent activation. The state weight matrix and offset term of cell state are represented by W_c and b_c , respectively. The element-by-element multiplication is denoted by the symbol \circ .

3. METHODOLOGY

The LSTM network outperforms well-established models in predicting transformer oil dissolved gases and can successfully deal with the challenge of prediction of nonlinear sequences. Figure 2 depicts the framework for the LSTM-based digital twin training approach, and the actions that follow are carried out.

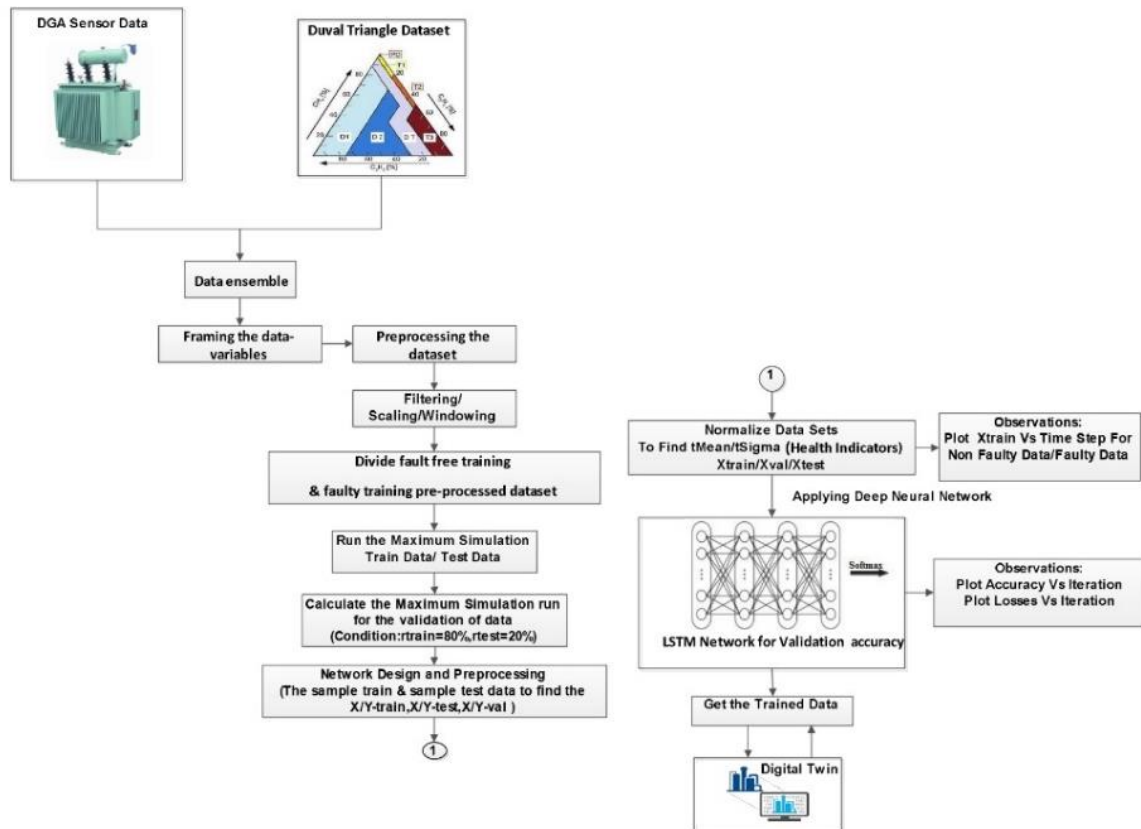


Figure 2. Framework for the LSTM-based digital twin training approach

3.1. DGA data acquisition

The three dissolved gasses (C_2H_2 , C_2H_4 , CH_4) from transformer oil (Grade-1) obtained by using Hydran sensor setup and different gas intensity percentage combinations from Duval triangle has been taken and formed four ensemble files, namely Faultfreetraining.csv, Faultfreetesting.csv, Faultytraining.csv and Faultytesting.csv respectively. In order to affirm the efficacy of the proposed prediction work, the case of this paper analyzes time series data using online monitoring of a 500 kVA transformer as an example shown in Figure 3. The fault-free data is taken for the period of 5 years (2013-17) and the faulty data is taken for 3 years (2018-2021). The specifications of the test transformer are given in the Table 1.



Figure 3. 500 kVA distribution transformer

Table 1. Specifications of 11kV/400V, 500 kVA test transformer

Parameters		Rating
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
Un Tanking Mass in Kgs		1120
Weight of oil in kgs		520
Oil in Liters		590
Total weight in kgs		2265
Maximum temperature rise in oil/WDG		50/55 Deg C

3.2. DGA data set

The DGA dataset taken by using hydran DGA sensor on 500 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 practical DGA data in this case consists of 960 samples of fault-free training with a data size of 480000x55, 500 samples of fault-free testing with a data size of 250000x55, faulty training and faulty testing with a data size of 960000x55, and faulty training and faulty testing with a data size of 960000x55. Table 2 summarizes the data set.

Table 2. Details of the fault-free and faulty data set

Particulars	No. of Samples	Size of Data
Fault free training data	960	480000x55
Fault free testing data	500	250000x55
Faulty training data	500	960000x55
Faulty testing data	500	960000x55

Each simulation's length is determined by the data set. For every three minutes, all simulations are sampled. 960 samples have been trained for 54 hours, 500 samples of 960 have been tested for 48 hours, 200 samples of 960 have been used for validation for 24 hours (fault-free data). Similar procedure for Faulty data set. The columns of apiece data frame includes the following variables:

- The fault type is shown in column 1 (fault code), which ranges from 0 to 20. A fault code of 0 indicates that the system is fault-free, whereas fault numbers 1 through 20 indicate distinct fault categories based on percentage gas intensity combinations obtained from the Duval Triangle.
- The number of times the simulation ran to acquire complete data is mentioned in column 2 (simulationRun). The number of runs in the training and test data sets ranges from 1 to 500, with each value representing a unique random generator state for the simulation across all fault codes.
- The number of times each variable is recorded per simulation is indicated in column 3 (sample).
- The measured variables from both the Hydran Sensor and the Duval Triangle are found in columns 4–55. The Duval triangle method is used in this paper as it has a 96% accuracy rate of determining a transformer defect, according to a review based on the IEC data bank of inspected transformer failures and several other reports [19], [20].

3.3. Data preprocessing

The following steps are carried out during DGA data preprocessing:

- Clean the data: Inconsistent values are removed using a high-pass filter. In both the training and testing data sets, delete data entries with the fault numbers 3, 9, and 15. These fault numbers are unrecognizable, and the simulation findings that go with them are incorrect.
- Scaling: This Method is implemented to bring the large data horizontally on to a scale which can be visualized. The scaling procedure uses a fixed sampling frequency to find time resolution. Time intervals are established based on time resolution and DGA sampling time. This time signal data is now applied to all of the DGA samples. For the windowing procedure, it is then translated into time domain data.
- Windowing: This method is implemented to bring the large data vertically on to a scale which can be visualized. Using a cosine signal and a hanning window [21], [22] the whole DGA data is transformed into time domain data. FFT analysis is used to convert a time domain signal to a frequency domain signal. The data from the entire DGA set is translated to frequency domain, and the dimensionality is lowered, which is necessary for such a big data collection.
- Divide data: By retaining 20% of the training data for validation, the training data is divided into training and validation data. A validation data set allows to assess the model's fit on the training data set while adjusting the hyper parameters of the model. Data splitting is a typical technique for preventing overfitting and under fitting in networks.
- Network design and preprocessing: In this process, the sample train and sample test data are preprocessed to find the network parameters viz. Xtrain, Ytrain, Xtest, Ytest, Xval and Yval.

3.4. Identifying the condition indicators

Normalize data sets: Normalization is a technique for converting numeric values in a data set to a similar level without distorting range disparities. This strategy assures that a variable with a higher value does not dominate the training variables. It also converts numeric data from a larger range to a smaller range without sacrificing any crucial training information. Using data from all simulations in the training data set, the mean and standard deviation for 52 signals are determined.

Visualize data: There are 400 fault-free simulations in the Xtrain data set, followed by 6800 defective simulations. A plot of the fault-free data is created first to visualize the fault-free and defective data. The total 10 signals are labeled in the Xtrain data set for the sake of this plot in order to construct an easy-to-read image. The signals from 1 to 3 are gas concentrations of C_2H_2 , C_2H_4 and CH_4 in ppm respectively. The signals from 4 to 6 are equivalent percentages of C_2H_2 , C_2H_4 and CH_4 respectively obtained from Duval's triangle method. The signals from 7 to 10 are fault code, simulation run, sample number and abnormal condition for the total failure of oil respectively. The visualization of non-faulty data (fault-free) and faulty data for 10 signals of the Xtrain data set is plotted in Figure 4 and Figure 5 respectively. As shown in the picture Figure 5, the LSTM has been completely trained after around 130 samples (time step) and ready for validation.

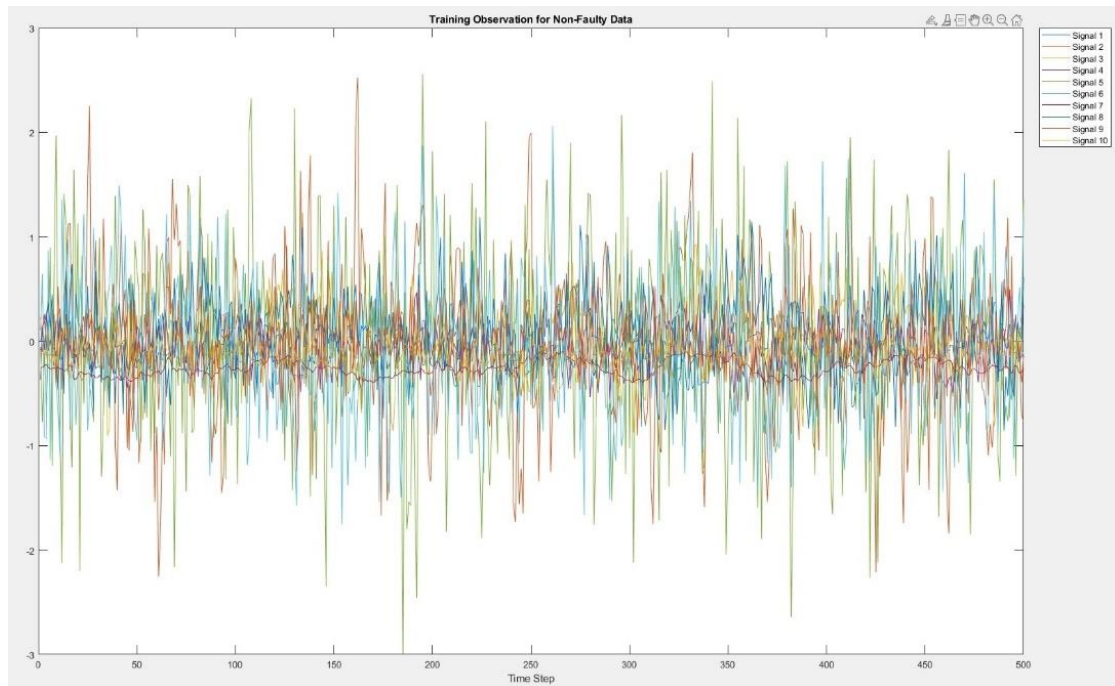


Figure 4. Training observation for non-faulty data

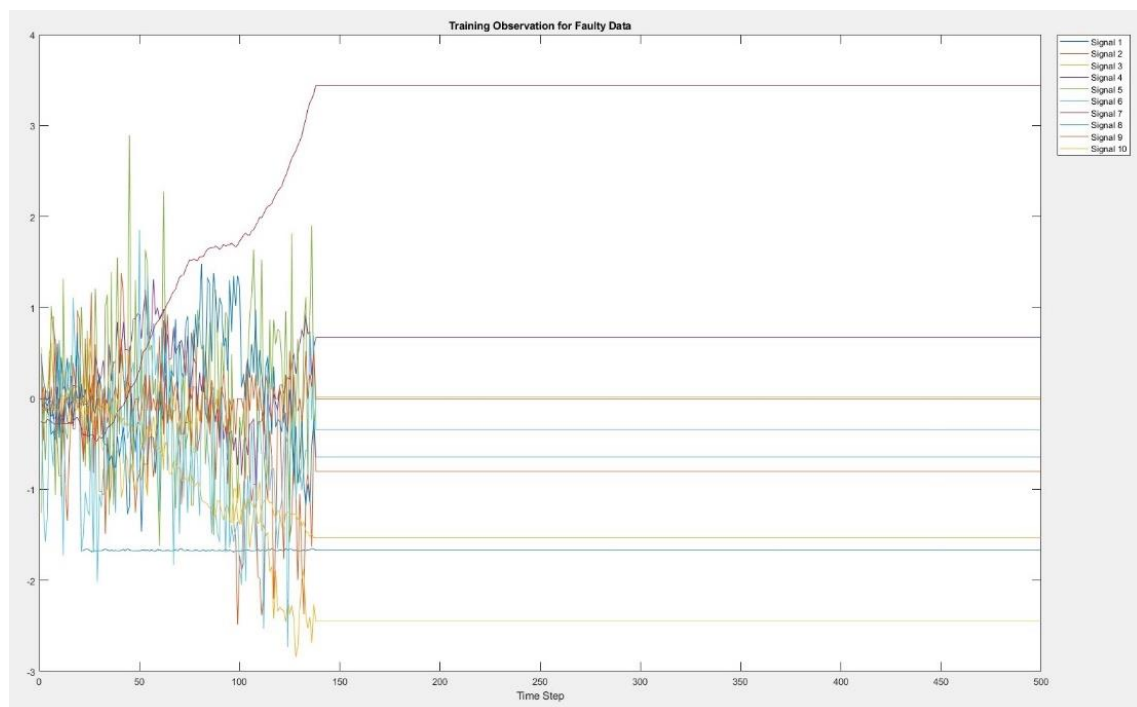


Figure 5. Training observation for faulty data

3.5. Training of the model in the test data set and running the model

In the process of training, the hardware and software used are MATLAB simulation platform (R2021a, MathWorks) on Dell PC with 16 GB RAM, 3.5 GHz clock, running Windows 7 enterprise operating system (64-bit) and integrated with graphical processor unit (GPU), Jetson Nano hardware to increase the speed of computations of deep learning architectures. There are 52 signals with 500 uniform time

steps in the final data set which includes training, validation, and testing data. As a result, the signal or sequence must be categorized to the correct fault number, making it a sequence classification challenge.

A LSTM network, a full-connect linear layer and a softmax layer are used in the proposed prediction model. The input layer size is 52 and number classes are 18. The fully connected layer's outputs are equal to the number of DGA classes to be categorized. In deep neural network models that predict a multinomial probability distribution, the softmax function is utilized as the activation function in the output layer. The number of hidden layers of the network considered are 3 with a unit size of 52, 40 and 25 respectively. The training epochs are 30 with the minimum batch size of 50 and the drop out is 0.2. Figure 6 depicts the LSTM's internal loop structure.

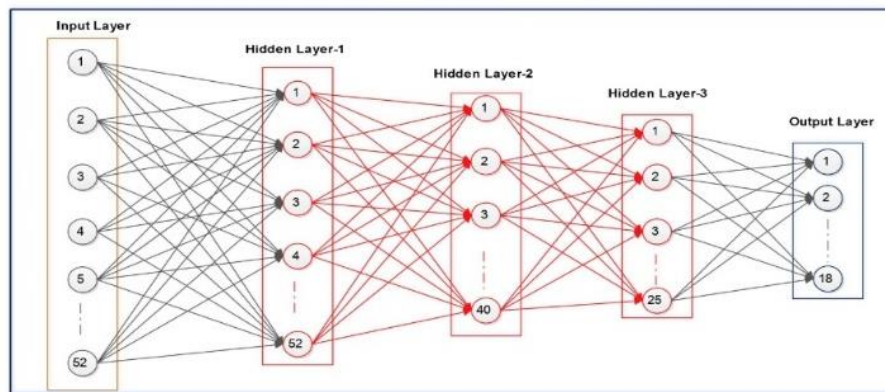


Figure 6. The LSTM's internal loop structure

Algorithm for LSTM network:

- Use fault-free and faulty DGA data (80% for training and 20% for testing)
- Initialize the LSTM network
- Apply both fault-free and faulty training data to the LSTM network
- Apply both the fault-free and the faulty test data to the LSTM network
- Set the training options for both training and testing data
- Run the trained LSTM network on test set and predict the fault type
- Calculate the validation accuracy
- Obtain the confusion matrix to evaluate the efficacy of a categorization network

4. RESULTS AND DISCUSSION

The case study in this work examines time series data utilizing online monitoring of a 500 kVA transformer as a case study to demonstrate the effectiveness of the suggested prediction model. Based on training and testing data for fault free state (2013-17) and faulty condition (2018-21), the dissolved gas content for test transformer oil was calculated, which is elaborately mentioned in section 3.2. The suggested high pass filter approach eliminates unexpected and unwanted intensity in the CH₄, C₂H₂, and C₂H₄ gases. The scaling and windowing strategies applied for the filtered DGA data result in a more apparent output signal by lowering its dimensionality. Dividing the data for training & testing and normalizing the data for identifying health indicators are done to train LSTM model in the given DGA data set. The data pertaining to LSTM network layers and number epochs, including approach to network design and preprocessing are mentioned in section 3.3 to 3.5.

4.1. Accuracy of fault diagnosis evaluation

The performance accuracy ratio (PR) is used to measure fault diagnostic accuracy [23]. The criteria for the overall performance of the model is defined as the ratio of the number of equal predictions (n) to the total number of predictions (N) and is given by:

$$PR = n/N \quad (7)$$

$$\text{Validation accuracy (VA)} = PR * 100\% \quad (8)$$

This criterion demonstrates the neural network's high accuracy in correctly identifying the fault type of unseen signals with few errors. As a result, the better the network, the higher the accuracy. Validation accuracy and loss curves obtained on the test data for LSTM algorithm against the training data are shown in the Figure 7. The efficacy of a categorization network is evaluated using a confusion matrix [24]. The confusion matrix aids in identifying a model's correct predictions as well as its errors for various particular classes. The confusion matrix is made up of columns with the LSTM networks predicted values and rows with true values. As seen in Figure 8, the major diagonal of the LSTM confusion matrix has numerical values, while the other elements in the off diagonal have almost zeros. This indicates that the trained network is efficient, classifying over 99 percent of signals correctly.

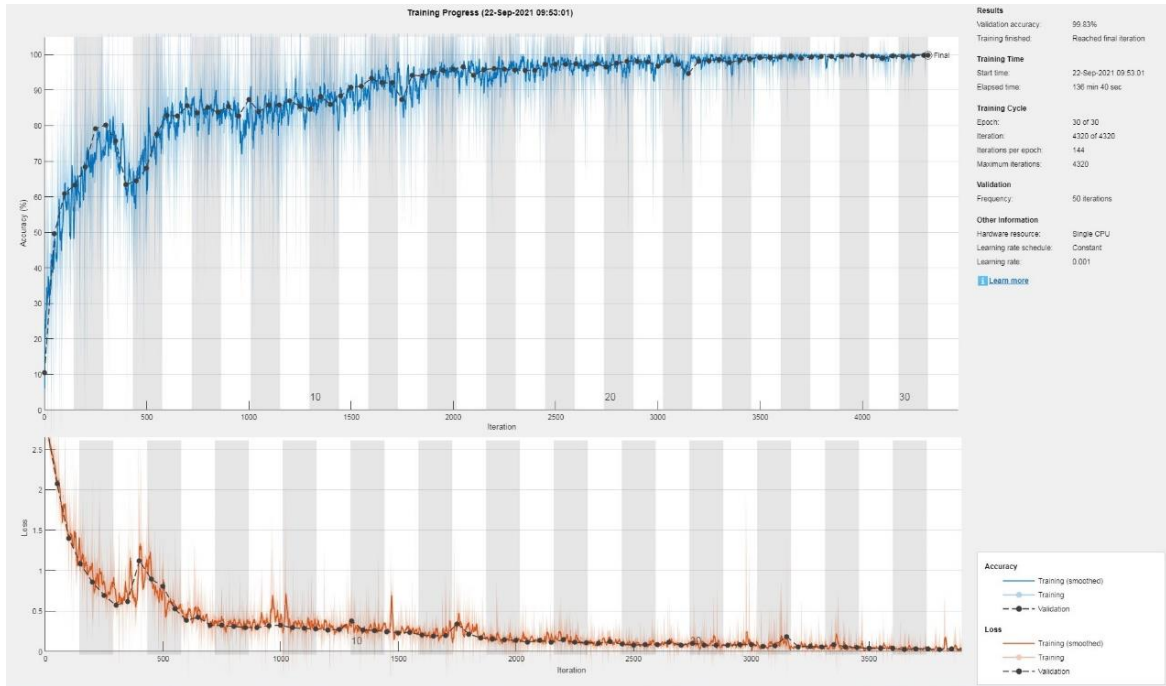


Figure 7. Validation accuracy and loss curves obtained on the test data for LSTM algorithm against the training data

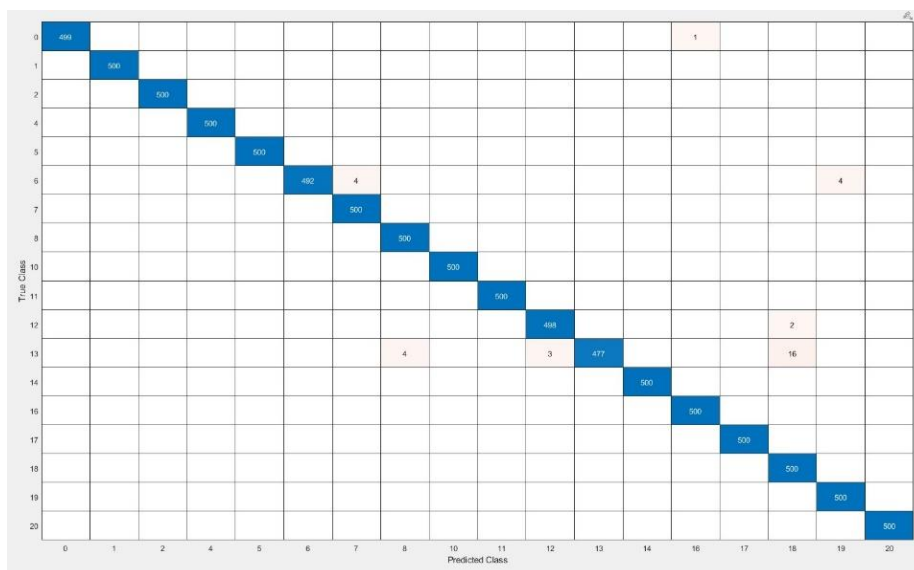


Figure 8. The LSTM confusion matrix

The validation accuracy in terms of precision achieved by the most commonly used convolutional neural network (CNN) model and the machine learning-based support vector machine (SVM) model is 96.04% and 97.11% respectively [25]. For condition monitoring, we employed a unique DGA data set acquired over an 8-year period to properly train the LSTM network so that the validation accuracy attained in this investigation is 99.83%, indicating an effective index for evaluating the fault diagnosis. As a result, the suggested model's performance can be objectively measured for research purposes. A comparison of validation accuracy on different models is shown in the Figure 9.

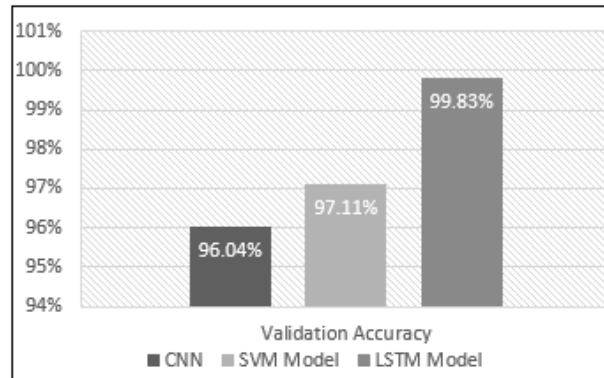


Figure 9. A comparison of validation accuracy on different models

4.2. Digital twin training

The use of a digital twin is an important step towards health management, as it introduces a new paradigm for fault diagnosis [26], [27]. The conceptual model of digital twin is shown in the Figure 10. Data from the functioning asset can be used to tune both data-driven and physics-based models, resulting in a digital twin. The digital twin technology is implemented by using Jetson Nano Hardware, graphical processor unit (GPU), NVIDIA Maxwell architecture with 128 NVIDIA CUDA® cores. It is trained by using LSTM network with the help of CUDA tool kit.

The experimental set up of Jetson Nano Hardware Integrated with MATLAB is shown in the Figure 11. Jetson Nano hardware is used to deploy and compute the DNN for training and testing. High validation accuracy has been achieved by running LSTM network in GPU under parallel computation as a novel approach. The GPU, a digital twin of LSTM network acts as an autonomous device to precisely predict the fault, to monitor the condition and to anticipate the remaining useful life of the transformer.

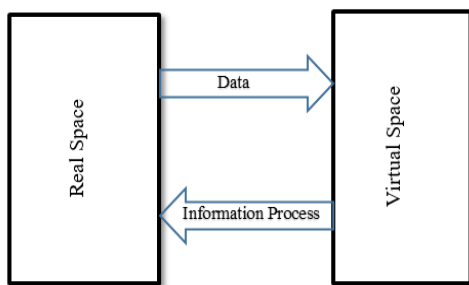


Figure 10. The conceptual model of digital twin



Figure 11. Jetson Nano Hardware integrated with MATLAB

5. CONCLUSION

Based on the current rapid development in testifying the efficacy of the proposed prediction model, this paper analyzed a time series data using a continuous monitoring of 500 kVA transformer. Generally the amount of sampled dissolved gas data is quite limited and the accuracy of multistep prediction is not high. This paper is focused on the application of large scale data using a promising model, LSTM based digital

twin training approach to analyze the dissolved gas concentration in transformer oil. An experiment is conducted to evaluate the proposed model. The analyzed results of the case study show that the projected prediction model has enhanced the validation accuracy, and it can track the effective change in the trend of dissolved gas content in transformer oil. The validation accuracy of 99.83% indicating an effective index for evaluating the fault diagnosis by the developed framework. The trained digital twin on integrating with the test transformer's condition monitoring system, can precisely envisage the transformer's useful life. We can investigate its application in transformer online monitoring using a mobile device.

ACKNOWLEDGMENT

GVSSN Srirama Sarma would like to express his gratitude to CITD, Balangar, Hyderabad, for supplying the massive DGA dataset of a 500 kVA, 11000/430 V test transformer as well as computational resources.




REFERENCES

- [1] N. Lelekakis, D. Martin, W. Guo and J. Wijaya, "Comparison of dissolved gas-in-oil analysis methods using a dissolved gas-in-oil standard," *IEEE Electrical Insulation Magazine*, vol. 27, no. 5, pp. 29-35, 2011, doi: 10.1109/MEI.2011.6025366.
- [2] J. I. Aizpurua *et al.*, "Power transformer dissolved gas analysis through Bayesian networks and hypothesis testing," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 25, no. 2, pp. 494-506, 2018, doi: 10.1109/TDEI.2018.006766.
- [3] J. Fan *et al.*, "Characteristics of Tin Oxide Chromatographic Detector for Dissolved Gases Analysis of Transformer Oil," *IEEE Access*, vol. 7, pp. 94012-94020, 2019, doi: 10.1109/ACCESS.2019.2928869.
- [4] J. Faiz and M. Soleimani, "Dissolved gas analysis evaluation in electric power transformers using conventional methods a review," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 24, no. 2, pp. 1239-1248, 2017, doi: 10.1109/TDEI.2017.005959.
- [5] J. Faiz and M. Soleimani, "Assessment of computational intelligence and conventional dissolved gas analysis methods for transformer fault diagnosis," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 25, no. 5, pp. 1798-1806, 2018, doi: 10.1109/TDEI.2018.007191.
- [6] S. Eeckhoudt, S. Autru and L. Lerat, "Stray gassing of transformer insulating oils: impact of materials, oxygen content, additives, incubation time and temperature, and its relationship to oxidation stability," *IEEE Electrical Insulation Magazine*, vol. 33, no. 6, pp. 27-32, 2017, doi: 10.1109/MEI.2017.8085066.
- [7] J. Li, Q. Zhang, K. Wang, J. Wang, T. Zhou and Y. Zhang, "Optimal dissolved gas ratios selected by genetic algorithm for power transformer fault diagnosis based on support vector machine," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 23, no. 2, pp. 1198-1206, 2016, doi: 10.1109/TDEI.2015.005277.
- [8] S. S. M. Ghoneim, I. B. M. Taha and N. I. Elkalashy, "Integrated ANN-based proactive fault diagnostic scheme for power transformers using dissolved gas analysis," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 23, no. 3, pp. 1838-1845, June 2016, doi: 10.1109/TDEI.2016.005301.
- [9] S. Li, G. Wu, B. Gao, C. Hao, D. Xin and X. Yin, "Interpretation of DGA for transformer fault diagnosis with complementary SAE-ELM and arctangent transform," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 23, no. 1, pp. 586-595, February 2016, doi: 10.1109/TDEI.2015.005410.
- [10] A. Alqudsi and A. E. Hag, "Application of Machine Learning in Transformer Health Index Prediction," *Energies*, vol. 12, no. 4, pp. 1-13, 2019, doi: 10.3390/en12142694.
- [11] B. Zeng *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*, vol. 13, no. 2, pp. 1-20, 2020, doi: 10.3390/en13020422.
- [12] F. Kabir, B. Foggo and N. Yu, "Data Driven Predictive Maintenance of Distribution Transformers," *China International Conference on Electricity Distribution (CICED)*, 2018, pp. 312-316, doi: 10.1109/CICED.2018.8592417.
- [13] X. Zhang, S. Wang, Y. Jiang, F. Wu and C. Sun, "Prediction of dissolved gas in power transformer oil based on LSTM-GA," *IOP Conference Series: Earth and Environmental Science, The Fifth International Conference on Energy Engineering and Environmental Protection 17-19 November 2020, Xiamen, China*, 2021, vol. 675, pp. 1-8, doi: 10.1088/1755-1315/675/1/012099.
- [14] T. Chengshun, S. Dan, T. Wei, C. Shiyu, L. Jianzhao, and J. Jianping, "Surface stress prediction model of steam turbine rotor based on LSTM recurrent neural network[J/OL]," *Proceedings of the Chinese Society of Electrical Engineering*, 2020, pp. 1-15.
- [15] G. Sun, C. Jiang, X. Wang and X. Yang, "Short-term building load forecast based on a data-mining feature selection and LSTM-RNN method," *IEEE Transactions on Electrical and Electronic Engineering*, vol. 15, no. 7, pp. 1002-1010, 2020, doi: 10.1002/tee.23144.
- [16] M. Sundermeyer, H. Ney and R. Schlüter, "From Feedforward to Recurrent LSTM Neural Networks for Language Modeling," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 3, pp. 517-529, 2015, doi: 10.1109/TASLP.2015.2400218.
- [17] Z. Jie and X. Wei, "End-to-end learning of semantic role labeling using recurrent neural networks," *53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, 2015, pp. 1127-1137.
- [18] H. Chuye Y. Zhong, Y. Lu, X. Luo and S. Wang, "A Prediction Model for Time Series of Dissolved Gas Content in Transformer Oil Based on LSTM," *International Conference on Ubiquitous Power Internet of Things (UPIOT 2020) & 4th International Symposium on Green Energy and Smart Grid (SGESG 2020)* vol. 1659, pp. 1-9, 012030, 2020, doi:10.1088/1742-6596/1659/1/012030.
- [19] R. R. Rogers, "IEEE and IEC Codes to Interpret Incipient Faults in Transformers, Using Gas in Oil Analysis," *IEEE Transactions on Electrical Insulation*, vol. EI-13, no. 5, pp. 349-354, 1978, doi: 10.1109/TEI.1978.298141.
- [20] J. Golarz, "Understanding Dissolved Gas Analysis (DGA) techniques and interpretations," *IEEE/PES Transmission and Distribution Conference and Exposition (T&D)*, 2016, pp. 1-5, doi: 10.1109/TDC.2016.7519852.




- [21] P. Nicolas, and W. Carolina, "Introducing Hann windows for reducing edge-effects in patch-based image segmentation," *PLOS ONE*, vol. 15, no. 3, pp. 1-11, 2020, doi: 10.1371/journal.pone.0229839.
- [22] R. Kumar, and K. Markam, "Performance Analysis of Kaiser-Hanning Window For Digital Filter Design," *Journal of Emerging Technologies and Innovative Research*, vol. 5, no. 10, pp. 634-640, 2018.
- [23] Z. Luo *et al.*, "Dissolved Gas Analysis of Insulating Oil in Electric Power Transformers: A Case Study Using SDAE-LSTM," *Mathematical Problems in Engineering*, vol. 2020, no. 2420456, pp. 1-10, 2020, doi: 10.1155/2020/2420456.
- [24] J. Namgung, S. Son, and Y. S. Moon, "Efficient Deep Learning Models for DGA Domain Detection," *Security and Communication Networks*, vol. 2021, no. 8887881, pp. 1-15, 2021, doi: 10.1155/2021/8887881.
- [25] F. Ren, Z. Jiang, X. Wang and J. Liu, "A DGA domain names detection modeling method based on integrating an attention mechanism and deep neural network," *Cybersecur*, vol. 3, no. 4, pp. 1-13, 2020, doi: 10.1186/s42400-020-00046-6.
- [26] Y. Xu, Y. Sun, X. Liu, and Y. Zheng, "A Digital-Twin-Assisted Fault Diagnosis Using Deep Transfer Learning," *IEEE Access*, vol. 7, pp. 19990-19999, 2019 doi: 10.1109/ACCESS.2018.2890566.
- [27] Y. Gao, X. Li, and L. Gao, "A Deep Lifelong Learning Method for Digital-Twin Driven Defect Recognition With Novel Classes," *Journal of Computing and Information Science in Engineering*, vol. 21, no. 3, p. 031004, 2021, doi: 10.1115/1.4049960.

BIOGRAPHIES OF AUTHORS






Gadepalli Srirama Sarma    received his B.E (2000) in Electrical and Electronics Engineering from Andhra University and M.Tech. (2003) in IT in Power Engineering from Jawaharlal Nehru Technological University Hyderabad. He is currently working on his Ph.D. in Electrical Engineering at JNTU, Hyderabad. He is presently working as Assistant Professor in Department of Electrical and Electronics Engineering, Matusri Engineering College, Saidabad, Hyderabad. He can be contacted at email: music.sarma2016@gmail.com.






Bumanapalli Ravindranath Reddy    obtained his B.Tech degree in Electrical and Electronics Engineering from JNTU College of Engineering, Anantapur in 1991. He obtained his M.Tech degree in Energy Systems from the Institute of Post Graduate Studies and Research of JNTUH in 1997. He got the Ph.D. degree in Electrical Power Systems from JNTU, Anantapur. At present he is working as Deputy executive engineer at JNTUH. He has published more than 60 research papers, and his areas of interests are power systems, high voltage engineering, and control systems. His research interests include Simulation studies on transients of different power system equipment. He can be contacted at email: bumanapalli_brreddy@yahoo.co.in



Pradeep Nirgude    obtained his B.E degree in Electrical and Electronics Engineering from Bangalore University. He obtained his M.E degree in High Voltage Engineering from Anna University. He got the Ph.D. degree in Electrical Engineering from Osmania University. He has published 150 research papers and his current research interests include Insulation coordination, Air insulation studies, optimization of transmission line tower window clearances, corona/RIV performance of transmission lines/equipment, electric field analysis of power equipment, high voltage testing & measurements, pollution performance of insulators/surge arrestors and diagnostics & condition assessment of power equipment by frequency response analysis. At present he is working as Additional Director, UHV Research Laboratory, CPRI, Uppal, Hyderabad. He can be contacted at email: pmnirgude@cpri.in.



Pudi Vasudeva Naidu    obtained B. Tech (EEE) degree from Bapatla Engineering College, Nagarjuna University in India, and M. Tech from JNTUCH in India and Doctoral Program (Ph.D.) at Acharya Nagarjuna University, Guntur, A.P, India. Currently he is an Assoc. Professor in EEE Dept. at Matusri Engineering College, Hyderabad. His areas of interest include Power Electronics, Power Quality (PQ) and FACTS. He can be contacted at email: pvdnaidu81@gmail.com.